DESIGN AND DEVELOPMENT OF TWO NOVEL CLASSES OF CYSTEINE

PROTEASE INHIBITORS AGAINST PATHOGENIC TRYPANOSOMES

A Dissertation

by

LINFENG LI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,   Frank M. Raushel
Committee Members, Thomas D. Meek
                             Gary R. Kunkel
                             Tadhg P. Begley
Head of Department, A. Joshua Wand

May 2021

Major Subject: Biochemistry

ABSTRACT

Human diseases caused by the genus *Trypanosoma*, including Chagas disease and African sleeping sickness, affect millions of people and cause enormous socioeconomic burdens in impoverished areas. However, neither vaccines nor well-tolerated therapies are currently available for prevention or treatment of these neglected diseases. The trypanosomal cysteine proteases play key roles in the life cycles of the parasites, and accordingly, become promising targets of drug discovery for these diseases.

The peptidomimetic vinyl sulfone **K11777** is a well characterized covalent inactivator of cruzain, the major cysteine protease of *Trypanosoma cruzi*. However, its irreversible mode of action may be associated with safety issues that impede its progression to clinical trials. We designed and synthesized a novel class of peptidomimetic vinyl heterocyclic inhibitors (PVHIs) which contain less electrophilic bioisosteres in place of the vinyl sulfone warheads. A number of PVHIs exerted potent, time-dependent, but reversible, inhibition of cruzain; and some of them exhibited considerable anti-trypanosomal activity not only in axenic cultures of pathogenic trypanosomes, but also in an infection model with murine cardiomyoblasts. Moreover, the concept of reversible covalent inactivation by vinyl heterocycles is herein embodied, and is potentially applicable to other enzymes containing active-site cysteines.

Cruzain is also effectively inhibited by simple peptidyl aldehydes which also raise concerns about potential toxicity and metabolic instability due to the over-reactive aldehyde group. We introduced a phenol group into the molecules that could form hemiacetal with the aldehyde group and function as a masking strategy. The hemiacetal

proved to be in cyclic form until binding to cruzain which apparently promoted the ring-opening and liberated the aldehyde for reacting with the active-site cysteine. These self-masked aldehyde inhibitors (SMAIs) appeared to be potent, rapidly reversible inhibitors that also showed promising trypanocidal activity. The hemiacetal hydroxyl group of SMAIs was next derivatized to provide potential prodrugs that could be metabolized by host enzymes. Furthermore, the SMAI strategy also enlightened the design of inhibitors for SARS-CoV-2 cysteine protease, leading to a class of potent 2-pyridone-based inhibitors, of which the binding modes with the 3C-like protease were demonstrated by crystallography.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my research advisor, Dr. Thomas D. Meek, for his comprehensive guidance and constant encouragement during my doctoral study.

I would like to thank my committee members, Drs. Frank M. Raushel, Gary R. Kunkel and Tadhg P. Begley. They have provided insightful discussion and advice throughout the course of my research.

I am grateful to all my collaborators, Dr. Jorge Cruz-Reyes's lab, Dr. Josh Wand's lab and Dr. Wenshe Liu's lab, for their professional assistance on cruzain and 3CL^{pro} projects.

Thanks also go to all my labmates, my friends, and the department staff for making my life in college station enjoyable and smooth.

Finally, I would deeply thank my family and my fiancée for their long-lasting love and support.

CONTRIBUTORS AND FUNDING RESOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Dr. Thomas D. Meek [Research Advisor], Dr. Frank M. Raushel [Committee Chair], and Dr. Gary R. Kunkel of the Department of Biochemistry and Biophysics, and Professor Dr. Tadhg P. Begley of the Department of Chemistry.

Dr. Bala Chenna from Dr. Meek's lab performed the glutathione adduct experiment and synthesized nearly half of the compounds. Elizabeth Hernandez, Zachary Goodall, and Jana Gomez from Dr. Cruz-Reyes's lab determined the trypanocidal activity in axenic cultures. Dr. James McKerrow's lab from UCSD determined trypanocidal activity in mice cardiomyoblasts. Dr. Taylor Cole from Dr. Wand's lab helped the analysis of 2D NMR for cruzain and ^{13}C-labeled compound. Drake Mellott from Dr. Meek's lab contributed to the preparation of 3CL^{pro} and its substrates. Dr. Kai Yang from Dr. Liu's lab solved the co-crystal structures of 3CL^{pro} with SMAIs.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

NOMENCLATURE

ACC	7-amino-4-carbamoylmethylcoumarin
AMC	7-amino-4-methylcoumarin
BSF	bloodstream form
CES	carboxylesterase
CHAPS	3-[3-(cholamidopropyl)dimethylammonio]-1-propanesulfonate
CMBP	cyanomethylenetributylphosphorane
CNS	central nervous system
CRP	complement regulatory protein
CYP	cytochrome P450
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCM	dichloromethane
DIPEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
FA	formic acid
FCC	flash column chromatography
GSH	glutathione
hPhe	homophenylalanine
HPLC	high-performance liquid chromatoraphy
HSQC	heteronuclear single quantum coherence
IPMK	inositol polyphosphate multikinase
IPTG	isopropyl β- d-1-thiogalactopyranoside
LAH	lithium aluminium hydride

mCPBA	*meta*-chloroperoxybenzoic acid
MEROPS	an online database for proteases and their inhibitors
MES	2-(*N*-morpholino)ethanesulfonic acid
MMTS	*S*-methyl methanethiosulfonate
MS	mass spectrometry
NECT	nifurtimox-eflornithine combination therapy
NMePip	*N*-methlypiperazinyl
NMR	nuclear magnetic resonance
NTD	neglected tropical disease
Oxz	oxazolyl
PCF	procyclic form
PVHI	peptidomimetic vinylheterocyclic inhibitor
Pyr	pyridinyl
Pyrmd	pyrimidinyl
SMAI	self-masked aldehyde inhibitor
T3P	propylphosphonic anhydride
TAPSO	3-[*N*-tris(hydroxymethyl)methylamino]-2-hydroxypropanesulfonic acid
TBAF	tetra-n-butylammonium fluoride
TBS	*tert*-butyldimethylsilyl
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thz	thiazolyl
VS	vinyl sulfone
VSG	variant surface glycoprotein
WHO	World Health Organization

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

RESEARCH BACKGROUND AND LITERATURE REVIEW

While every year over hundreds of billions of dollars are spent on the research and control of infamous diseases such as cancer, diabetes, cardiovascular disease, HIV/AIDS, etc.,[1] there is a group of infectious diseases affecting over one billion people globally yet receiving little interest of study and funding. These diseases, referred to as neglected tropical diseases (NTDs), are rare in wealthy countries, but quite common in low and middle-income countries of Latin America, Asia, and Africa.[2-3] So far, the World Health Organization (WHO) has categorized twenty communicable diseases as NTDs which are caused by different types of pathogens including viruses, bacteria, protozoa, and helminths. Several NTDs can be controlled or even eradicated through mass drug administration and public health surveillance by efforts from some large pharmaceutical companies, foundations, and government organizations.[4-5] Other NTDs remain unmet medical needs, still requiring the development of effective and inexpensive medicines or interventions.

1.1 Epidemiology, Manifestation, and Control of Human Trypanosomiasis

Of these NTDs, Chagas disease and African sleeping sickness are both infected by parasitic flagellate protozoa that belongs to the genus *Trypanosoma*. The pathogenic protozoan for Chagas disease is *Trypanosoma cruzi* (*T. cruzi*), and that for African sleeping sickness is *Trypanosoma brucei* (*T. brucei*).

Chagas disease, also called American trypanosomiasis, affects an estimated six to seven million people worldwide and puts about 75 million people at risk of being infected.[6] In 2017, approximately 162,500 new cases emerged, and 7,900 deaths were attributed to Chagas disease, rendering it an important public health issue.[7-8] Chagas disease was once only endemic in 21 countries of Latin America for which the transmission of disease is mostly via inoculation with the excreta of infected triatomine bugs (kissing bugs). These vectors could easily breed in rural areas of these countries owing to inadequate sanitation infrastructure. As the most affected country per capita, 18
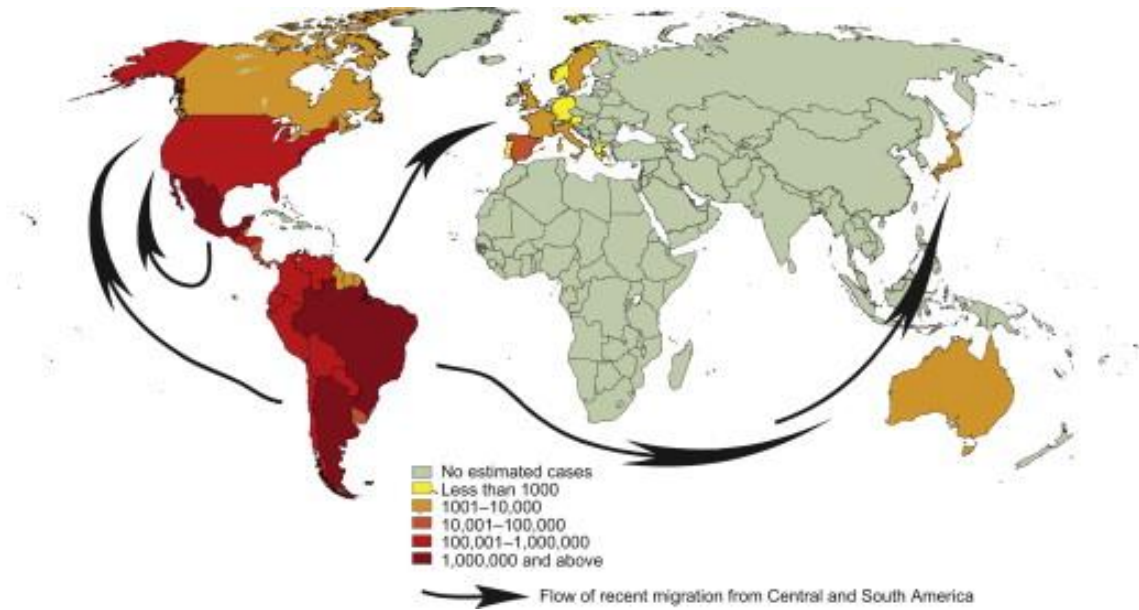
Figure 1.1 Current estimated global population infected by *T. cruzi*. Reprinted from Guhl's work.[235]

– 20% of the population of Bolivia is infected with *T. cruzi*.[9] The lowest prevalence of infection occurs in Brazil (1%) and Mexico (1%), however, as both countries are highly populous, they together account for nearly half of the cases in Latin America.[10] In recent decades, the increasing scale of urbanization and international immigration has expanded the distribution of infection to more developed countries in Europe, North America, and

Western Pacific Region (Japan, Australia) (**Figure 1.1**).[11] According to an estimation

from the Centers for Disease Control and Prevention, more than 300,000 *T. cruzi*-infected

persons reside in the United States.[12] Most of them are immigrants from Latin America

with highest proportion from Mexico (58%) followed by El Salvador (16.4%) and

Guatemala (6.8%). European countries are also heavily represented, and Spain alone has

around 75,000 cases of Chagas disease.[13] Since *T. cruzi* carriers are usually

asymptomatic, it is believed the true prevalence is even higher than these numbers.

Unlike the vector-borne mode in endemic countries, Chagas disease in non-endemic areas

are transmitted mainly through blood transfusion from infected donors, or through

congenital route (mother to child) during pregnancy.[9]

Chagas disease encompasses two clinical phases. After initial infection, an acute

phase which is characterized by high-grade parasitemia occurs and lasts about two

months. With vector-transmitted Chagas disease, the first clinical symptom is dependent

on the site of *T. cruzi* inoculation: either a skin chancre (chagoma) or a unilateral,

painless edema of the eyelids (Romaña sign).[14] As for systemic manifestations, most

patients in acute phase are asymptomatic or present mild symptoms including moderate

fever, malaise, headache, myalgia, edema, lymphadenitis, hepatomegaly, splenomegaly,

etc. After one to two months, production of antibodies and the activation of a host

immune response is commensurate with low level of parasitemia, and the clinical signs

disappear in 90% of the cases even without drug.[11] During the chronic phase, about 30%

of the patients will slowly develop organ dysfunction which often affects cardiac or

digestive system.[9, 15] The most severe and frequent (up to 30%) disease manifestations

include cardiac disorder, such as disturbances of the conduction system, atrial and

ventricular arrhythmias, cardiomyopathy, heart failure and secondary thromboembolism. Dysfunction of the digestive system is also common (10-21%) in the chronic phase, especially the development of dilated esophagus and colon that, respectively, lead to dysphagia and severe constipation.[16-17] Although rare (<5%), the involvement of central nervous system (CNS) can cause dementia or neuritis.[18]

Preventive strategies are different in endemic and non-endemic areas. Vector control is the principal method for prevention of Chagas disease in Latin America.[19] Multiple countries have implemented insecticide-spraying programs by spraying inside housing or at peri-domestic areas with systemic insecticides, typically pyrethroids.[20] This approach has significantly reduced the spread of disease in Brazil, Chile, and Uruguay through elimination of the insect *Triatoma infestans*, a main vector of the disease. However, emerging insecticide resistance has been reported in the last two decades in Argentina and Bolivia. Besides, housing improvement and use of bed-netting also prevent the colonization of houses by triatomine bugs. In addition to vector control, all donated blood and organs for transplantation in endemic countries are screened for *T. cruzi* antibodies. In non-endemic countries where vector control is unneeded and fewer donors are at risk, the control of transfusion-transmitted Chagas disease is usually either by selective donor screening or deferral of donation from risky sources.[9, 21] Nonetheless, universal screening is also adopted by several developed countries with more immigrants from endemic areas including United Kingdom, Spain, the United States, France, Sweden, Switzerland, and Belgium.[22-23]

Currently there is no vaccine available for Chagas disease. For many years, researchers have been concerned that anti-*T. cruzi* immunity triggered by vaccination

would worsen the progression of disease, hence this idea resulted in the stagnation of vaccine development.[24] It is now realized that the key factor of pathogenesis is the persistence of *T. cruzi* in tissues rather than the hyper-responsive immunity.[25] Consequently, a number of preclinical studies have been carried out to investigate different types of vaccine formulations (live-attenuated parasites, recombinant proteins, DNA vaccines) with different adjuvants and carriers (from cytokines, toll-like receptor agonists, nanoparticles).[26-28] Several candidates, either prophylactic or therapeutic, were tested in infected animal models and were able to reduce the parasitemia in heart and blood, suggesting the feasibility of a vaccine against Chagas disease.[29-30] With that being said, no vaccine has hitherto entered clinical trials.

There are only two etiologic drugs for the antiparasitic treatment of Chagas disease: benznidazole (Roche) and nifurtimox (Bayer) which were introduced half a century ago. Both drugs are undoubtedly effective in the acute phase (up to 80% cure) and in congenitally acquired cases (up to 99% cure)[31], yet their efficacy for chronic phase disease remains controversial. Some studies carried out in Argentina and Brazil demonstrated that the efficacy of these drugs in early chronic phase (a few years after acute phase) were close to those with the acute disease based on complete negative result of parasitological and serological tests.[32-34] However, in other research, a group of children in the chronic phase were treated with benznidazole and/or nifurtimox, and the follow-up data after 8 – 20 years showed only 1 out of 12 presented a parasitological and clinical cure.[35] Even more serious is the high proportion (33%) of treated children who progressed to second-degree cardiomyopathy and/or dilated esophagus. Apart from limited efficacy for chronic disease, both drugs have frequent side effects (40% of treated

adults) and required long period (60 days) of treatment.[36] Because the mechanism of

action for benznidazole and nifurtimox is basically to product free radical species to

which the parasites are particularly sensitive,[37] the non-specificity of these radical species

give rise to side effects including anorexia, vomiting, peripheral polyneuropathy and

allergic dermatopathies in treated individuals.[38] These untoward effects frequently lead to

abandonment of treatment, signifying the critical need for development of more effective

and selective chemotherapy, especially for the chronic form of Chagas disease.

1.1.2 African Sleeping Sickness

Compared with Chagas disease, African sleeping sickness, or African

trypanosomiasis, has a lesser influence in terms of population and region. Two subspecies

of *T. brucei* can cause two corresponding forms of disease: *Trypanosoma brucei*

gambiense (*T. b. gambiense*) causes chronic disease in western and central Africa which

is also the predominant form found in 98% of all reported cases; and *Trypanosoma brucei*

rhodesiense (*T. b. rhodesiense*) causes acute disease in eastern and southern Africa.[39]

Sleeping sickness was once a devastating epidemic at the beginning of 20th century and

claimed hundreds of thousands of lives in western and central African countries.[40] Soon

after the outbreak, David Bruce, after whom the *T. brucei* was named, identified the

pathogen and linked the disease to tsetse fly as the vector. The most recent resurgence

occurred in the late 1990s with an estimated 300,000 existing cases, which promoted an

international coordination for controlling the disease, resulting in a steady decrease of

infected cases. As of 2018, the encouraging fact that only 977 cases were recorded

manifested a high possibility to eradicate the disease (defined by interruption of

transmission) by 2030. It is notable that sleeping sickness can occur in vertebrate animals with other subspecies of *T. brucei*. For instance, *T. brucei brucei* (*T. b. brucei*) is a major causative agent for animal trypanosomiasis, which shares fundamental genomic features[41] with *T. b. gambiense* and *T. b. rhodesiense* although it cannot infect humans on account of its susceptibility to lysis by trypanosome lytic factor-1.[42] Additionally, some domestic animals are potential reservoir hosts of *T. b. gambiense*.[43] Therefore, previous and ongoing studies of *T. brucei* may benefit the development of veterinary medicines, but they also serve as a strategic reserve for future outbreaks in human.

Regardless of the different infectious species of *T. brucei*, clinical symptoms of African sleeping sickness undergo two stages: an initial hemolymphatic stage, followed by the neurological stage when *T. brucei* invades the CNS.[44] The acute form of disease caused by *T. b. rhodesiense* progresses to the second stage within a few weeks, and death within half a year; this dissertation does not discuss *T. b. rhodesiense* in detail because it is not the predominant pathogen of sleeping sickness. The chronic form caused by *T. b. gambiense* has a slower onset that is on average 3 years. Typical symptoms in the first stage include intermittent fever, severe headache, pruritus, and posterior cervical lymphadenopathy (Winterbottom's sign).[45] In the second stage, neurological symptoms become more obvious such as changes in behavior, sensory disturbances, and poor coordination; the most characteristic disturbance of sleep cycle, from which the name of disease is derived, comprises daytime somnolence, episodes of sudden sleepiness and nighttime insomnia. African sleeping sickness is considered inevitably fatal if left untreated, although cases of asymptomatic carriers have been reported.[46]

No vaccine exists for prophylaxis against African sleeping sickness. Moreover, the prospect for a conventional vaccine is dim in the foreseeable future because the *T. brucei* genome contains 806 variant surface glycoprotein (VSG) genes so that the parasite is capable of switching its antigenic type and evading the immune system.[47] Since the main reservoir of *T. b. gambiense* is human, which is different from zoonotic *T. b. rhodesiense*, the best approach to control *T. b. gambiense* outbreaks is through active and passive case detection followed by treatment, and vector control to reduce the transmission.[48] Like triatomine bugs in America, tsetse flies are also vulnerable to pyrethroid insecticides, thus aerial/ground spraying of insecticide, insecticide-impregnated nets, and insecticide-treated cattle are all low-cost methods for vector control. Besides, introducing sterile male flies also proved effective in Zanzibar and Senegal albeit this is less practical due to high cost.[49]

Treatment of African sleeping sickness relies on a handful of drugs, all of which are replete with drawbacks such as non-oral administration, intolerable adverse effects, and drug resistance.[50] The first-line treatment for first-stage *T. b. gambiense* infection is pentamidine by injection into vein or muscle. It is generally well tolerated, despite some undesirable effects like hypoglycaemia (5-40%), hypotension, and abdominal pain.[51] The first-line treatment for second-stage *T. b. gambiense* is nifurtimox-eflornithine combination therapy (NECT) which was included in WHO Essential Medicines List.[52] The benefits of NECT are higher cure rates (95-98%), less severe side effects, easier administration, and lower chance of inducing resistance, compared to using melarsoprol (a third-line treatment) or using eflornithine alone. In 2019, a new oral drug, fexinidazole, was included in WHO Essential Medicines list for *T. b. gambiense* infection while it

8

showed activity against *T. cruzi* as well.[52] It is considered a first-line treatment for both

first stage (91% efficacy), and non-severe second stage.

1.2 Pathogenic *Trypanosoma*

As noted above, *T. cruzi* and *T. brucei* are the only human pathogenic members of

the genus *Trypanosoma*; however, they are categorized into different clades, implying

they are not as closely related to each other as thought before.[53] *T. cruzi* belongs to a

clade named *Stercoraria* while *T. brucei* belongs to a clade named *Salivaria*.[54]

Stercorarian trypanosomes are carried in the posterior gut of blood-feeding insects from

the subfamily *Triatominae*. During feeding, the infected vector intakes plenty of blood

that forces the formation and elimination of excreta (feces/urine). The bite site or intact

mucosa is thereby contaminated by these excreta which contain infective forms of

parasites. Transmission is eventually completed when these trypanosomes penetrate the

skin and disseminate in the body.[55] In contrast, salivarian trypanosomes are passed to the

recipient via the saliva of tsetse flies from the genus *Glossina*. These parasites not only

colonize the intestinal track of their vector, but also occur in the salivary gland or biting

mouthpart where they develop into infective forms. Therefore, the host is inoculated with

salivarian trypanosomes by the tsetse fly bite even before it sucks blood. Both triatomine

and tsetse fly become infected when taking a blood meal from the mammalian host.

T. cruzi passes its life cycle in two intermediate hosts, i.e., invertebrate vector

(triatomine bug) and vertebrate host (human), and undergoes complex morphological

changes. Specifically, there are three main developmental stages of *T. cruzi*:

trypomastigote, amastigote, and epimastigote. As shown in **Figure 1.2A**, the cycle in

vector (left) arbitrarily starts with the triatomine bug taking a blood meal from human

host infected with trypomastigotes which circulate in the bloodstream. Most of the

ingested trypomastigotes are lysed in the stomach of the insect while the surviving

trypomastigotes move to midgut where they transform into epimastigotes, also known as

pro-cyclic forms (PCFs), in the insect's midgut after a few days.[56] Epimastigotes then

proliferate through binary fission, and migrate to the microvilli of the insect's hindgut

and rectum,[57] an event thought to trigger the differentiation of non-infective

epimastigotes into non-dividing, highly infective trypomastigotes (or metacyclic

trypomastigotes). These metacyclic trypomastigotes are excreted with the feces and urine,

which subsequently enter the vertebrate host cycle (right) as described in the last

paragraph. Once transmitted, the metacyclic trypomastigotes invade the host cells near

the inoculation site by phagocytosis involving sophisticated interactions between both

cells. In the cytoplasm, the long and thin flagellate parasites transform into short, non-

infective amastigotes that can replicate intracellularly. Upon completion of a replication

cycle, the amastigotes transform back to trypomastigotes which cause cell rupture and

become bloodstream forms (BSFs) of trypomastigotes. Bloodstream trypomastigotes

infect adjacent cells or distribute via bloodstream to infect cells of other tissues where
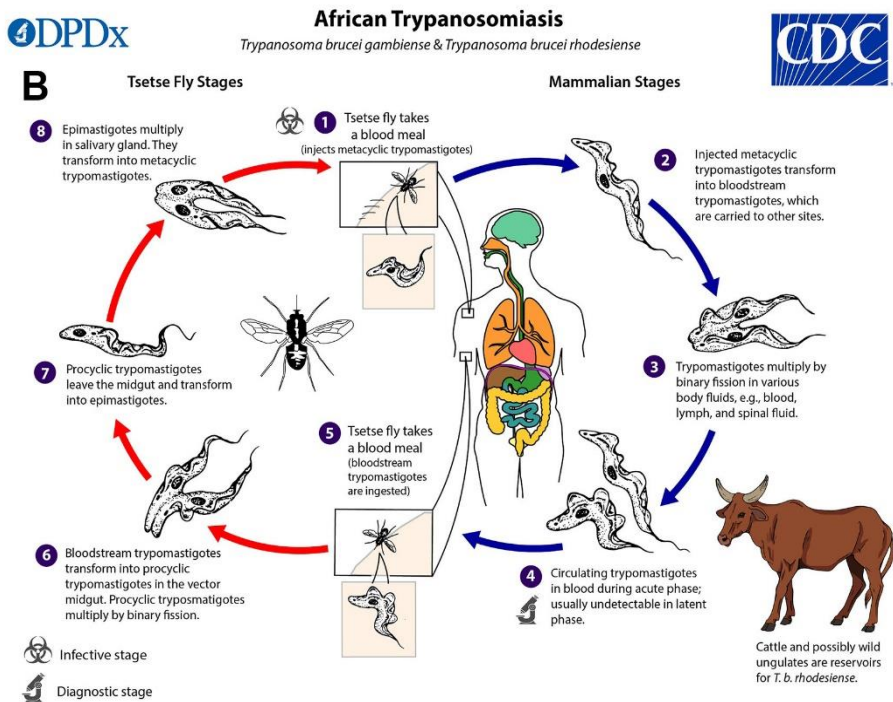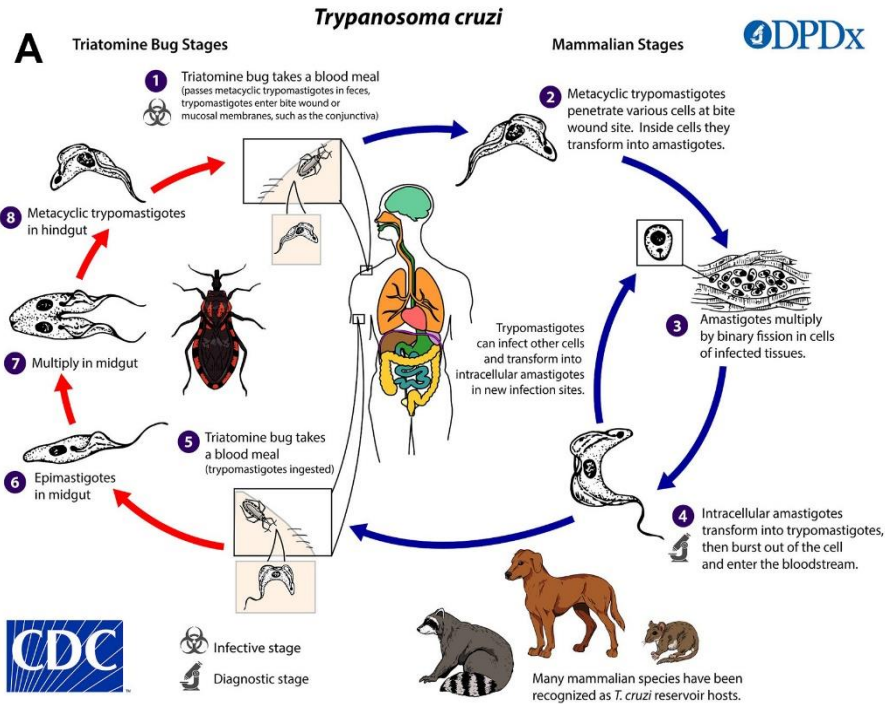
Figure 1.2 Life cycles of *T. cruzi* (A) and *T. brucei* (B) in human hosts and insect vectors. Reprinted from CDC's website.[3]

they transform into amastigotes again and repeat the cycle. This infective cycle,

commonly occurring in cardiomyocytes, peripheral skeletal/muscle cells, and endothelial

cells, can result in clinical symptoms.

Compared with that of *T. cruzi*, the life cycle of *T. brucei* is similar but has several distinct features as shown in **Figure 1.2B**. During the insect cycle, a tsetse fly becomes infected with bloodstream trypomastigotes when it bites an infected mammalian host. The parasites differentiate into procyclic trypomastigotes that multiply by binary fission and establish a midgut infection. Then procyclic trypomastigotes depart from the midgut, travel a long way during which they transform to epimastigotes, and finally arrive at salivary gland where they adhere to the epithelium and divide once again to produce metacyclic trypomastigotes.[58] These infective forms are injected into another mammalian host and directly delivered to its bloodstream instead of invading cells, which is remarkably different from the intracellular *T. cruzi*. Metacyclic trypomastigotes differentiate to bloodstream trypomastigotes which spread throughout the body with circulation system and invade extravascular tissues, including the CNS. Meanwhile, these trypomastigotes are able to replicate either in blood or in lymphatic and spinal fluids.

The differences between *T. cruzi* and *T. brucei* are also reflected in their immune evasion strategies for vertebrate hosts.[59] *T. cruzi* amastigotes readily get away from the humoral immunity as they are sequestered inside host cells. A group of mucins on amastigote surface are involved in the protection of parasite as well as assistance of host cell invasion.[60] The primary evasion mechanism of bloodstream trypomastigotes is via maneuvering the host complement system using several surface glycoproteins, also known as complement regulatory proteins (CRPs).[61] For example, the most studied protein, namely 160 kDa CRP, can bind to C3b and C4b, thereby impeding the activation of both the classical and alternative pathways of complement.[62] In contrast, bloodstream

trypomastigotes of extracellular forms *T. brucei* evade the host immune system mainly by expression of many new VSGs as mentioned in the previous section. VSGs are so numerous that they form a compact coat, which prevents the immune system from accessing the invariant surface antigens. Although this VSG coat can be recognized by host antibodies, frequent antigenic variations of VSGs result in a new coat thus escaping the immune response. Studies also demonstrated other roles of the VSG coat such as inhibition of complement-mediated lysis and induing the production of cytokines.[63] It is worth noting that there are many other mechanisms accountable for immune evasion of the parasites, including modulation of host immunity,[64] regulation of cytokines gene expression,[65-67] etc., which are not elaborated herein yet are still being intensely investigated.

      T. cruzi has a complex life cycle consisting of four life stages. First, trypomastigotes circulating in the blood of an infected mammalian host is ingested by the feeding insect vector. The trypomastigotes then transform first into epimastigotes that divide by binary fission and then into non-dividing, infectious metacyclic trypomastigotes in the hindgut of the insect; they are next deposited within the vector feces during subsequent blood meals. Natural transmission to a new mammalian host occurs when the parasite laden feces contaminate oral or nasal mucous membranes, the conjunctivae, or wounds in the skin, including vector bites. Once in the mammalian host, the trypomastigotes enter host cells, and transform into the multiplying intracellular forms or amastigotes, which then transform into bloodform trypomastigotes. These forms are released into the bloodstream as the host cell ruptures, and are then ready to invade healthy cells.

Much effort has been made to develop anti-trypanosomal drugs employing different approaches including phenotypic screening, target-based drug discovery, and repurposing of approved drugs.[68-69] A small number of new drug candidates have been discovered and are in late-phase clinical trials, however, their application to chronic-stage trypanosomiasis remains a challenge. Posaconazole, a triazole antifungal agent, was previously reported to have promising trypanocidal activity in both stages of Chagas disease in murine models.[70-71] Two separate trials showed it was not as efficacious as benznidazole in chronic patients with even more adverse reactions.[72-73] Another triazole compound E1224 was shown to eradicate *T. cruzi* at a rate on par with benznidazole but was unable to maintain the clearance of *T. cruzi* after months of treatment.[74] Besides, the antiarrhythmic drug amiodarone has anti-*T. cruzi* activity[75] and is undergoing a phase 3 clinical trial[76] to assess its effect among chronic patients with mild-to-moderate cardiomyopathy. Phenotypic screening for African trypanosomiasis identified a promising compound acoziborole (SCYX-7158) which is trypanocidal against many strains of *T. brucei* (*T. b. gambiense*, *T. b. rhodesiense*, *T. b. brucei*, etc.).[77] It is now in phase 2b/3 trials being carried out in the Democratic Republic of the Congo.[78-80]

Advances in genomics technology and biological imaging have smoothed the path for target-based drug discovery. A number of pathways in trypanosomes are considered druggable because they are either exclusive to parasites or appreciably different from those in mammalian host. Inositol polyphosphate multikinase (IPMK) of the inositol phosphate pathway was chemically validated as an anti-trypanosomal target because it is an essential enzyme for growth of bloodstream trypomastigotes and infection of mice.[81-82] Two IPMK inhibitors of different kinetic modes were identified and shown to be

14

effective against *T. cruzi* amastigotes.[82] Ergosterol is a unique and essential sterol for protozoa and fungi, and is absent in animals.[83] As a result, key enzymes in biosynthetic pathway of ergosterol become potential druggable targets. A validated enzyme is C14-demethylase (CYP51) of which the synthesis is inhibited by azole derivatives including posaconazole, ravuconazole (active form of E1224) and so on.[84] Other druggable targets in this pathway include squalene synthase and farnesyldiphosphate synthase.[85] Enzymes involved in metabolism, morphological change, host cell invasion and immune evasion have also been identified as anti-trypanosomal targets. Protozoal cysteine proteases–cruzain, brucipain, and rhodesain–are long-standing, important examples of drug targets for trypanosomal diseases, the former of which is the core topic of this dissertation, and is extensively discussed in following sections. Additionally, trans-sialidase,[86-87] trypanothione reductase,[88-89] and superoxide dismutase[90-91] are all studied targets to date. The unique mitochondria in *T. cruzi* are also organelles to be targeted. Several molecules including nitrobenzaldehyde thiosemicarbazone,[92] azalactone,[93] and triphenylphosphonium derivatives[94] are able to selectively decrease membrane potential of trypanosomal mitochondria, thus killing the parasites. As a large group of enzymes, trypanosomal protein kinases are of particular interest due to differences with human kinome.[95] For example, in trypanosomal kinome, there are much more STE, CMGC and NEK kinases but no tyrosine-like receptor kinase at all.[96-97] A few studies for identification of druggable kinases have been reported for *T. brucei*,[98-99] while little is known about essential kinome of *T. cruzi* and more researches in this field are ongoing.

1.3 Cysteine Proteases as Anti-trypanosomal Drug Targets

1.3.1 Biological Significance of Cysteine Proteases for Trypanosomes

Drug development by targeting proteases in infectious pathogens has gained

substantial success as demonstrated by ten HIV-1 aspartic protease inhibitors and seven

HCV NS3/4A serine protease inhibitors all of which are now used in clinical practice.[100]

Although currently there is no cysteine protease inhibitor approved by FDA, a few

candidates are under clinical development against various cysteine proteases of human

and microorganism. The genomes of trypanosomes encode several cysteine proteases

affiliated to different families according to the MEROPS, an online database for

peptidases and their inhibitors.[101] The most common and well characterized trypanosomal

cysteine proteases include cathepsin L-like (CatL-like) proteases and cathepsin B-like

(CatB-like) proteases, which are all papain-like enzymes assigned to clan CA family C1,

on the basis of similarities in their amino acid sequences. CatL-like proteases, namely

cruzain of *T. cruzi* and brucipain or rhodesain (TbCatL) of *T. brucei* have been

structurally and biochemically characterized more so than CatB-like proteases, namely

TcCatB of *T. cruzi* and TbCatB of *T. brucei*.

Compared with other eukaryotes, the genomes of trypanosomal parasites are

clustered to produce polycistronic mRNA.[47, 102] The genome of *T. cruzi* contains many

gene clusters each made up of 3-5 paralogues that are tandemly arrayed on different

chromosomes. This thus implies the presence of other cruzain isoforms, one of which has

been studied and termed as cruzain-2,[103] which shares 86% sequence identity with

cruzain. In like fashion, the CatL-like proteases of *T. brucei* are encoded by over 20

tandemly-arrayed genes. Cruzain is ubiquitously present in all life stages of *T. cruzi* with greater expression levels in epimastigotes and amastigotes than in the bloodstream trypomastigotes.[104-105] On the contrary, cruzain-2 is mostly found in the trypomastigotes despite its high sequence identity with cruzain.[106] Nonetheless, the RNA levels of both proteases are similar in all stages,[107] suggesting that gene regulation is through translational or posttranslational control. On the other hand, the organization of genes for CatB-like proteases in *T. cruzi* and *T. brucei* seems less complicated, as they are present just as single-copy genes.[108-109]

As a member of the papain family, cruzain is initially synthesized as a precursor protein (pro-cruzain, or zymogen) which has an *N*-terminal signal sequence, a pro-domain, and a catalytic domain (**Figure 1.3A**). Unlike other papain-like enzymes, there is a unique extension at the *C*-terminal of the protein, which also is found in the *T. brucei* subspecies. Native cruzain is a sulfated glycoprotein with glycans attached to some residues around the active site and in the *C*-terminal extension.[110-111] The signal sequence is responsible for transporting pro-cruzain to the endoplasmic reticulum (ER) for *N*-linked glycosylation and is cleaved off after transportation.[112] Once in the ER, the pro-domain is able to implement the correct folding of the nascent protein and relocation of the mature protein to the lysosome/endosome compartments.[113] In addition, the pro-domain constitutes a potent cruzain inhibitor to suppress unwanted proteolytic activity.[114] The pro-domain is removed from the catalytic domain via autoproteolysis in an acidified Golgi compartment, which is a necessary process for final intracellular sorting.[115] The catalytic domain consists of ~210 residues with a Cys_{25}-His_{162} catalytic dyad. Studies suggest the *C*-terminal extension is an immunogenic part that is recognized by serum

antibodies in many chronic patients, yet its exact function remains vague.[116] It is noted

that "cruzipain" was the original term for native parasite-generated protease while

"cruzain" only denoted the recombinant protease of which the *C*-terminal extension is
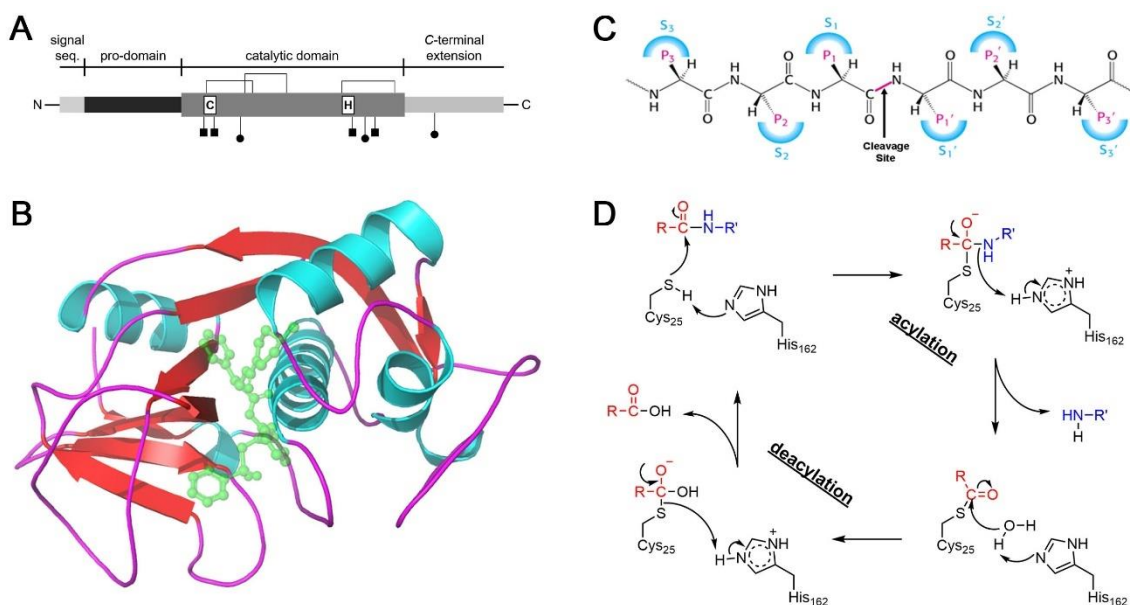
Figure 1.3 Structure and catalytic mechanism of cruzain. (A) Domain diagram of native pro-cruzain. Squares above the catalytic domain delineate disulfide bonds; C and H are catalytic Cys_{25} and His_{162}; the sites of *N*-glycosylation (■) and *O*-glycosylation (●) are marked below the catalytic domain; recombinant cruzain usually has no glycosylation and *C*-terminal extension. (B) 3D structure overview of cruzain (extracted monomer, PDB accession code: 2OZ2). The α-helices (cyan) and β-sheets (red) are connected by loops (magenta); ligand inhibitor (green) is shown as a translucent ball-and-stick model. (C) Schechter-Berger nomenclature for proteases. Depending on the distance and direction from cleavage site, the sub-sites on the protease are named S1/S2/S3/…(towards *N*-terminus) and S1'/S2'/S3'/…(towards *C*-terminus); likewise, peptidyl substrate/inhibitor residues are named P1/P2/P3/… and P1'/P2'/P3'/… (D) A brief scheme of the catalytic cycle of cruzain.

truncated in most cases, and hardly affects any biochemical properties and structural

studies. Nowadays the former term has become obsolete, and the latter is used as a

general name. The intracellular trafficking of TbCatL appears to be similar to that of

cruzain as indicated by RNAi experiments.[117] Little information about the processing and

sorting of trypanosomal CatB-like proteases is available, calling for more related studies.

Being the major cysteine protease of *T. cruzi*, cruzain has been, for many years, serving as a paradigmatic target of inhibitor design and biological research for all trypanosomal cysteine proteases. Cruzain plays multifaceted roles in the life cycle of *T. cruzi*, generally divided into physiological and pathological roles. Like other enzymes in lysosomal organelles, cruzain is able to digest external and parasite protein under acidic conditions, either to produce small molecule nutrients or to dispose of intracellular wastes. Studies indicated overexpression of cruzain was associated with an augmented level of the nutrient-consuming metacyclogenesis;[118] on the other hand, inhibition of cruzain in epimastigotes caused accumulation of toxins inside the Golgi apparatus.[115] Besides, a variety of inhibitors were employed in several studies to elucidate the roles of cruzain in the differentiation of trypomastigotes into amastigotes, transformation of amastigotes back to trypomastigotes, and proliferation of epimastigotes and amastigotes.[119] Apart from its physiological roles, cruzain participates in pathological events such as host cell invasion and evasion from immune response. Araujo-Jorge *et al.* first found that the addition of alpha-2-macroglobulin, a plasma protease inhibitor, inhibited the infection of bloodstream trypomastigotes of *T. cruzi* into murine macrophages and fibroblasts.[120-121] Souto-Padron *et al.* further showed the uptake of trypomastigotes by macrophages was significantly reduced by $F(ab')_2$ fragments of anti-cruzain antibodies,[122] convincingly suggesting the involvement of cruzain in *T. cruzi*-host cell interaction. Invasion of trypomastigotes into mammalian cells was also linked to cruzain because of its ability to aid in releasing kallidins which induce inflammation and promotes uptake of *T. cruzi*.[123-125] To generate kallidins, cruzain either directly proteolyzes the kininogens, or processes the prekallikreins to make more products

19

indirectly. Besides, cruzain also enhances invasive activity of trypomastigotes by activating latent transforming growth factor beta, an important and versatile cytokine.[126] The story of cruzain with host cell invasion may be even more complex than above noted, so it is with immune evasion. Cruzain is able to cleave human IgG at the hinge region and hydrolyze Fc fragment to small peptides with minimal degradation of Fab fragment.[127-128] Such a "fabulation mechanism" leaves an intact Fab fragment attached to the antigen on *T. cruzi* surface as a protection but impedes the activation of complement system or opsonization for phagocytosis. Doyle *et al.* claimed, according to their unpublished data, *T. cruzi* parasites expressing negligible cruzain activity were unable to infect wild-type mice even at doses of 10^6 trypomastigotes and were only lethal in RAG1 knockout mice which were deleted mature T and B cells and were thus severely immunodeficient. The authors later demonstrated that cruzain, by degradation of NF-κB P65, could thwart the activation of macrophages during early infection (<1 h).[129] This unresponsiveness of phagocytosis at the bite site would support the survival of parasites in early infection and allow their subsequent dissemination to distant tissues. This also explains why AIDS patients with concurrent *T. cruzi* infection often develop fatal Chagas disease.[130] Lastly, attempts to knock out the cruzain gene ended in failure, likely resulting from the lethality of this process itself. Taken together, cruzain is definitely an indispensable enzyme for *T. cruzi*.

Although CatL-like proteases in *T. brucei* have similar functions as cruzain, there are several studies regarding their *T. brucei*-specific roles. By activating the host protease-activated receptor 2 (PAR-2) and hence generating calcium activation signals,[131-132] brucipain is able to drive *T. brucei* to traverse the blood-brain barrier

(BBB). It also confers the parasite's resistance to lysis by components found in human serum, and to the suppression of surface coat exchange by host anti-VSG IgG.[133] Trypanosomal CatB-like proteases are also involved in virulence and are potential drug targets. TbCatB is of importance to host protein digestion, especially iron acquisition by degradation of transferrin.[134] In *T. cruzi*, a 30 kDa CatB-like protease (presumably TcCatB) has been identified and sequenced,[135] probably working as an alternative enzyme when cruzain availability is decreased in parasites.

The rationale of papain-like proteases, especially cruzain, to be drug targets not only relies on the essentiality of their biological roles in parasites, but also takes into account any potential off-target toxicity. One is likely to be concerned that inhibitors targeting papain-like proteases may impair host cells or tissues through unspecific binding to homologous and analogous host proteins. Actually, such concerns may be mitigated considering the following facts. In mammalian hosts, trypsin-like serine proteases are the major family of all proteases, while in protozoa, papain-like cysteine proteases greatly outnumber other classes of proteases.[136-137] This contrast is the most obvious when it comes to comparing the levels of the two protease families involved in gastrointestinal protein digestion. This is because many trypsin-like proteases differentiated to digestive enzymes with the evolution of the pancreas in vertebrates, but papain-like proteases in protozoa were not supplanted by serine proteases during evolution.[138] On the other hand, thiolate is a stronger nucleophile than alkoxide owing to the nature of the more polarizable sulfur atom, making cysteine proteases usually more efficient than serine proteases for executing their biological functions, especially at acidic pH. These protozoan cysteine proteases, in turn, are often more vulnerable to oxidative

21

stress or exogenous electrophiles. Therefore, cysteine protease inhibitors can easily achieve selectivity over many of the host serine proteases. Although serine proteases are dominant in mammalian hosts, cysteine proteases are not unique to parasites so that off-target effects for host homologues should be noted. Fortunately, since the typical concentrations of host cathepsins in lysosomes may be as high as 1 mM,[139] such high, local concentrations may protect them from an inhibitor which is incapable of matching this concentration in the lysosome to a fair extent. In addition, the human proteome presents a functional redundancy pertaining to cysteine proteases, which is not observed in the simpler parasitic proteome. In fact, gene knockouts of some cathepsins in murine models caused no embryonic lethality with minimal phenotypic consequences.[140] Moreover, as cruzain is abundantly expressed on the surface of *T. cruzi* amastigotes, it is more vulnerable to inhibitors relative to human cathepsins which are sheltered in organelles.

1.3.2 Structural Basis of Cruzain Inhibition

From a homological perspective, it is more precise to describe cruzain as CatF-like rather than CatL-like because its sequence identity with CatF reaches ca. 50%, higher than other homologous human cathepsins. The first crystal structure of recombinant structure was solved as a complex with Cbz-Phe-Ala-fluoromethylketone by McGrath *et al.* in 1995,[141] since then more than 20 structures with different substrates, inhibitors or mutations have been reported employing both X-ray crystallography and protein NMR. The spatial structure of a cruzain monomer comprises two domains in common with proteases in papain family (**Figure 1.3B**).[142] One of the domains mainly consists of

several alpha-helices and the other contains extensive antiparallel beta-sheets. Located at the interfaces between the two domains is the cleft-shaped active site, housing the catalytic dyad Cys_{25}-His_{162} and several pockets for substrate binding designated using the Schechter and Berger nomenclature (**Figure 1.3C**).[143] The S_1' and S_1 subsites are both solvent-exposed, shallow pockets that virtually contribute little interaction to ligand binding. In papain-like proteases, the S_2 pocket is the primary determinant for substrate specificity. Substrates with aliphatic or aromatic side chains at P_2 position are commonly preferred because S_2 pockets of many CatL-like proteases are lined with hydrophobic residues.[144-145] While cruzain also favors bulky non-polar P_2 group (e.g. Leu or Phe), it is able to accommodate cationic residues (e.g. Arg) owing to the presence of an unusual Glu_{208} at the base of its S_2 pocket which can form a critical ionic interaction with basic residues.[146] Due to this interaction, substrate specificity is pH-dependent. In particular, cruzain prefers hydrophobic groups to Arg by nearly two orders of magnitude at pH 5.5. Structural analysis revealed that, the carboxylate moiety of Glu_{208} was oriented to the basin of S_2 pocket at neutral pH to get closer to the incoming charged moiety, but at acidic pH it swung away from the pocket so that this ionic interaction was undermined. Harris *et al.* carried out a specificity profiling using a combinatorial library of fluorogenic substrates with a fixed P_1 Arg or Leu at pH 5.5.[145] The first-tier P_2 residues in descending order of preference are Leu>Phe~Tyr>Val; the second-tier P_2 residues include Ile, Trp and Nle (norleucine) which exhibited comparable binding affinities for S_2 pocket yet significantly lower than the first-tier ones; besides, substrates with other P_2 residues are essentially not hydrolyzed by cruzain. While P_3 and P_4 positions generally show broad specificity, P_3 does have a slight preference for Arg and Lys and P_4 is repulsive to bulky

aliphatic or aromatic amino acids such as Leu and Phe. It is noteworthy that all the specificity profiles of P_2, P_3 and P_4 are independent of P_1 structure, consistent with the poorly-defined S_1 pocket as described earlier.

The catalytic mechanism of cruzain is by and large similar to papain with subtle differences. A complete catalytic cycle generally consists of an acylation stage followed by a deacylation stage (**Figure 1.3D**). In the acylation stage, the nucleophilic Cys_{25} attacks the carbonyl carbon of the peptide bond in substrate, followed by protonation and scission of the C-N bond, which leads to release of the amine product and formation of an acylated intermediate. In the following deacylation stage, the thioester intermediate is hydrolyzed to produce the carboxylate product. However, different details with regard to this mechanism have been proposed in two recent publications. Based on a pure QM/MM study,[147] in the acylation stage, the N1 atom of the substrate is first protonated to produce a transient intermediate, making the carbonyl C1 more susceptible to the nucleophilic attack by Cys_{25}. It is believed the Cys_{25} thiolate is stabilized by Trp_{26}, His_{162} and Gly_{163} rather than a simple ion pair of the dyad. In the deacylation stage, His_{162} abstracts the proton from a water molecule to make a hydroxide as the base for hydrolytic reaction through a concerted mechanism. Residues Gln_{19}, Asn_{182}, Trp_{184} appears to be critical for modulating the pK_a of His_{162}, which acts as an acid in the acylation stage and as a base in the deacylation stage. The calculated activation energy of the one-step deacylation stage is highest in the reaction diagram, indicating the deacylation is the rate-limiting step. Our group later published a paper on cruzain mechanism using experimental approaches including pH-rate profiles and solvent kinetic isotope effect (sKIE).[148] In our proposed mechanism, the key feature is that the initial protonation states of dyad are both neutral.

Substrate binding may trigger a conformational change, and brings the dyad closer for proton transfer. The deacylation is also proposed as a general base-catalyzed reaction and turns out to be rate-limiting which agrees with the former study. The real scenario may be a combination of both, or even more complex depending on substrate structure and developmental stage, nonetheless, these studies have laid a firm foundation for mechanism-based drug design.

1.3.3 Development Status of Cruzain Inhibitors

The first proof-of-principle inhibitor/inactivator used as the cruzain-targeted treatment for Chagas disease was **K11777**, which since then has undergone extensive preclinical evaluations and optimizations. It showed potent, apparent inhibition against cruzain (IC_{50}^{app} = 3.6 nM) as well as moderate trypanocidal activity (EC_{50} = 3 − 4 µM) for different strains of *T. cruzi*.[149] This compound could kill the intracellular amastigotes by arresting the processing of cruzain at the level of Golgi complex.[150] Treated mice in chronic infection had repeatedly negative hemocultures indicative of parasitological cure. It was also efficacious in an immunocompromised murine model of acute *T. cruzi* infection as well as in an immunocompetent canine model.[130, 151] Although **K11777** demonstrated no histological abnormalities or auto-immune phenomena, its progress to clinical trial was halted due to hepatotoxicity and tolerability issues (measured by the incidence of certain adverse event among a given population of tested animals) at low dose in primates and dogs.[152]

The structure of **K11777** is a peptidomimetic vinyl sulfone as shown in **Figure 1.4A**. The β-carbon of the vinyl sulfone is attacked by Cys_{25}, leading to the formation of

a carbanion intermediate in a conjugated fashion (Michael addition); then α-protonation of the intermediate produces a stable thioether, irreversibly inactivating the enzyme.[153] The crystal structure of cruzain complexed with **K11777** has confirmed the formation of a covalent adduct.[142] In addition, the scaffold of **K11777** snugly spans S_3, S_2, S_1 and S_1' pockets of the enzyme by virtue of a hydrogen bonding network involving Gln_{19}, Gly_{66}, Asp_{161}, His_{162} and Trp_{184}. These interactions are fairly conserved as found in cruzain structures complexed with other analogous peptidyl backbones. Glu_{208} is directed away from the hydrophobic P_2 phenyl group which is in accordance with aforementioned substrate specificity.
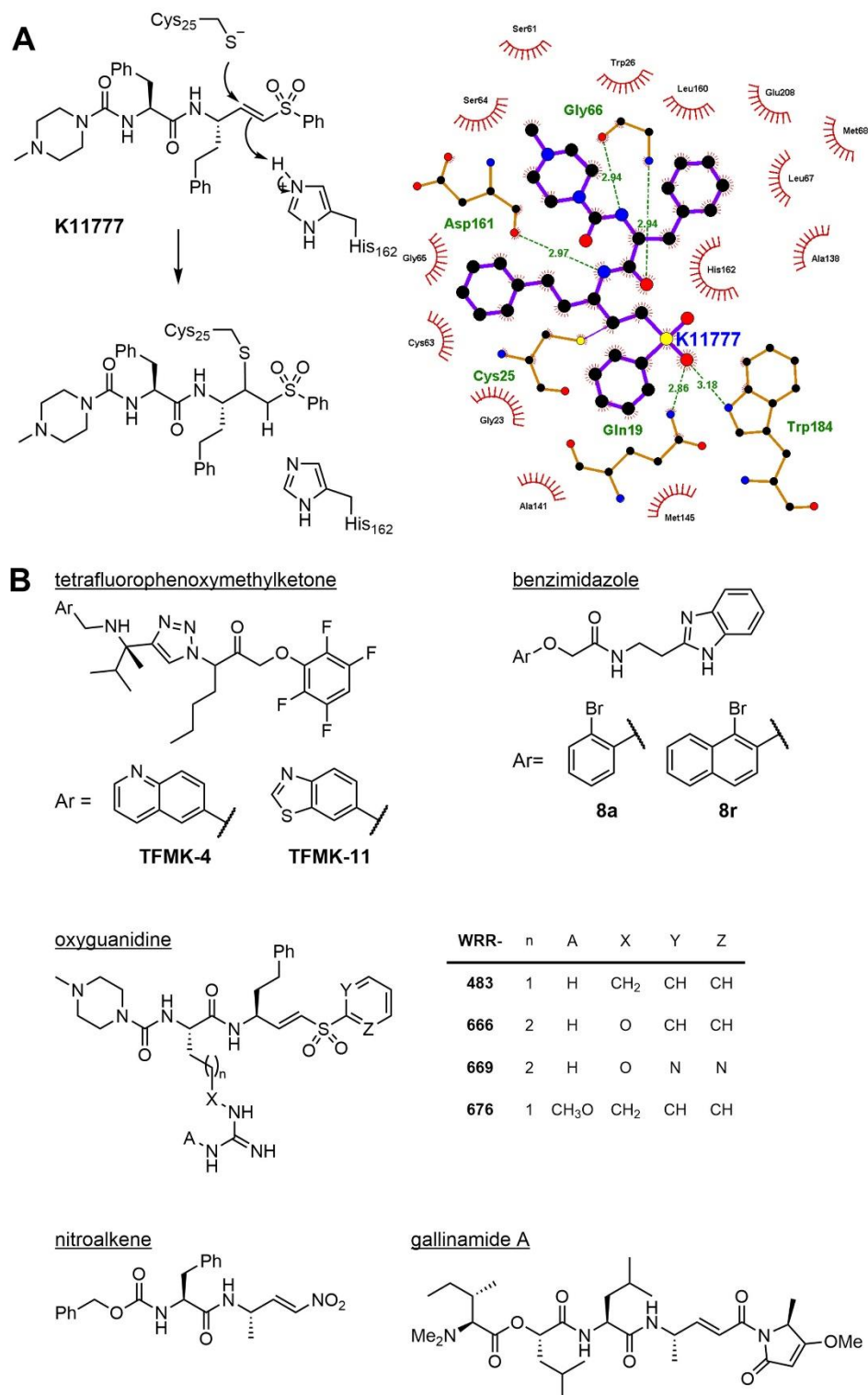
Figure 1.4 Current inhibitors/inactivators for cruzain under preclinical development. (A) Structure and inhibition mechanism of **K11777**. (B) Several classes of cruzain inhibitors including non-peptidyl or peptidomimetic inhibitors and natural product, with representative compound(s) of each class.

Besides **K11777** and its derivatives, a diversity of molecules have been studied as cruzain inhibitors and inactivators (**Figure 1.4B**). A class of tetrafluorophenoxymethyl ketone-based compounds are nonpeptidic, irreversible cruzain inactivators that utilize tetrafluorophenol as the leaving group. The first compound of this class, **TFMK-4**, was developed by a screening approach,[154] and exerted potent inactivation against cruzain with a second-order inactivation constant (k_{inact}/K_I) of 147,000 s^{-1}M^{-1} as well as complete elimination of *T. cruzi* in cell culture at 10 µM. This compound also mitigated symptoms of acute-phase infection in a mouse model with good tolerability. The authors obtained the structure of cruzain-inhibitor complex based on which Neitz *et al.* further optimized the *in vitro* potency and *in vivo* pharmacokinetic profile of this series, leading to the identification of compound **TFMK-11** with improved pharmacokinetic and ~10-fold increase of trypanocidal activity (IC$_{50}$ = 2 nM in *T. cruzi* infected C2C12 host cells).[155]

Benzimidazoles represent a class of nonpeptidic, noncovalent cruzain inhibitors. From a campaign combining high-throughput and virtual screenings, Ferreira *et al.* discovered a hit compound **8a** with an K_i of 800 nM against cruzain.[156] In the co-crystal structure, the hydrophobic bromophenyl ring fits into S$_2$ pocket while the benzimidazole interacts with protein indirectly through water-bridged hydrogen bonds. Medicinal chemistry efforts on **8a** resulted in a large group of analogs which fulfilled a comprehensive SAR analysis. Replacement of the bromophenyl with a naphthyl moiety yielded the most potent compound **8r** with 10-fold improvement (K_i = 82 nM) likely due to full shape complementarity of the S$_2$ pocket.[157] Moreover, compound **8r** showed moderate trypanocidal activity (EC$_{50}$ = 16.2 µM) and no acute toxicity in mice.

Efforts for target-based drug design principally focused on the modification of **K11777** either to optimize the peptidomimetic scaffold or to replace the vinyl sulfone with another electrophilic warhead. The design of oxyguanidine derivatives was inspired by the fact that Arg could be accommodated by S_2 pocket of cruzain but not of other CatL-like proteases. **WRR-483**, an analog of **K11777** with Arg at P_2, proved to be as effective as **K11777** at eliminating *T. cruzi* in cell culture and animal model despite mediocre potency against cruzain.[158] Oxyguanidine derivatives of **WRR-483** remarkably inhibited cruzain with a second-order rate constant up to 320,400 $M^{-1}s^{-1}$.[159] **WRR-676** demonstrated favorable metabolic stability and a promising trypanocidal activity (EC_{50} = 269nM) in *T. cruzi*-infected macrophages. Interestingly, crystallographic analysis revealed that **WRR-666** was a covalent inactivator similar with **K11777** while **WRR-669** adopted an unexpected, noncovalent binding mode, presumably as a result of the different P_1' moiety.

Nitroalkenes are another class of peptidomimetic cruzain inhibitors currently being studied. Many of them showed high potency not only against cruzain and rhodesain, but also against cathepsin L and, to a lesser extent cathepsin B. On the basis of molecular modeling result, the most active compound (K_i = 0.44 nM) bearing a simple Ala at P_1 position was believed to bind to cruzain in a similar fashion to vinyl sulfones.[160] However, these nitroalkene inhibitors were reversible following protein dialysis, implying a distinctive chemical nature of the warhead. Computational studies and NMR data suggested the establishment of a reaction equilibrium between cruzain Cys_{25} and the double bond in inactivator.

A natural product called gallinamide A was first identified as a potent irreversible inactivator of human cathepsin L. Boudreau *et al.* found this molecule had a sub-nanomolar value of IC_{50} (0.26 nM) for cruzain and was exceptionally toxic to *T. cruzi* amastigotes (LD_{50} = 14.7 nM) which outperformed the positive control benznidazole (LD_{50} = 1.5 µM).[161] The chemical mechanism of covalent inactivation was elucidated for cathepsin L as a Michael addition to the acrylamide moiety, which was likely applicable to cruzain as well. Although this double bond is a weaker electrophile than that in **K11777**, gallinamide A turned out to be irreversible evidenced by rapid dilution assay.

CHAPTER 2

PEPTIDOMIMETIC VINYL HETEROCYCLIC INHIBITORS OF CRUZAIN EFFECT
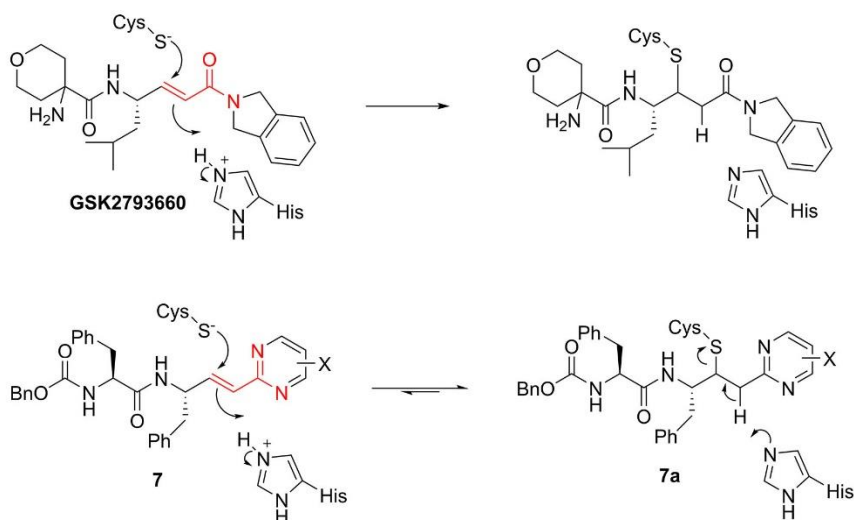
ANTITRYPANOSOMAL ACTIVITY*

2.1 Introduction

Inactivators that form reversible covalent adducts with cysteine groups on

enzymes have received recent attention.[162-164] Such reversible covalent inactivators

demonstrate time-dependent inactivation, like irreversible inactivators, but may exert

greater selectivity for the intended target rather than homologous "off-target" enzymes.

This is because while the initial-collision complexes of irreversible inactivators with a

panel of related enzymes may have variable affinities, over time the establishment of a

permanent covalent bond may render this initial selectivity inconsequential. However, in

the case of a reversible covalent inhibitor, their residence times on these enzymes are

likely to be variable, leading to the ultimate "relief" from covalent inactivation for off-

targets.

The design of dipeptide vinyl-heterocyclic inhibitors is predicated on existing

irreversible covalent inactivators of cysteine proteases such as the vinyl sulfone of

K11777 and the acrylamide of **GSK27993660**, an inactivator of human cathepsin C that

reached Phase I clinical trials.[165] Both compounds form irreversible covalent adducts with

the active-site cysteines of the respective enzymes via a thia-Michael reaction (**Figure

1.4A**, **Figure 2.1**), and both have encountered either toxicity issues or adverse events in,

Figure 2.1 Structures of GSK2793660 and **7** with thia-Michael addition of Cys_{25} to the vinyl groups in these compounds. The putative adduct **7a** reverts to the fully conjugated **7** upon the reverse of adduct formation. The common bioisosteric atoms of GSK2793660 and **7** are highlighted in red. X is an electron-donating or electron-withdrawing substituent.

respectively, either animals or humans.[165-166] Accordingly, we sought replacements of the

vinyl sulfone and acrylamide "warheads" with less electrophilic moieties that would

undergo *reversible* thia-Michael addition, in order to develop inactivators of high potency

for cruzain, but also with suitable selectivity for trypanosomal over human cysteine

proteases. One approach is the replacement of the vinyl sulfone and carboxamide group

of the acrylamide with a bioisosteric heterocyclic group which is conjugated to the

reactive vinyl group. A 2-vinyl pyrimidine **7** is one such bioisosteric replacement (**Figure**

2.1), in which one of the ring nitrogens mimics the amide carbonyl and sulfone oxygen

while the other substitutes for the amide nitrogen (red atoms). As the pyrimidine is

conjugated to the vinyl group, the expected addition of the thiol group of the cysteine to

the β-carbon of the vinyl group and attending protonation of the α-carbon would

eliminate this conjugation. Subsequent re-establishment of conjugation via reverse of the

thia-Michael reaction would provide reversible covalent inactivation of the enzyme.

Another feature afforded by this vinyl-heterocycle is the ability to modify the reactivity of the vinyl group by the substitution of the heterocycle with electron-donating or electron-withdrawing groups. In this chapter, we designed, synthesized, and evaluated a panel of dipeptide compounds containing a vinyl group replacing the scissile amide group of the substrate, which is conjugated to a phenyl group or a collection of heterocycles. For some, we investigated the ability of glutathione to form covalent adducts with their vinyl groups, in order to explore the electronic nature of the heterocycle required for facile addition of thiols. Many of these compounds displayed potent, time-dependent inhibition of cruzain, as well as anti-trypanosomal activity in cell culture.

2.2 Results and Discussion

2.2.1 Computer-assisted Inhibitor Design

To aid in the rational design of our vinyl-heterocyclic inhibitors we employed molecular docking of these compounds to a model constructed from the crystal structure of **K11777**-cruzain (**Figure 1.3B**) which contains a covalent bond between the inactivator and Cys_{25}.[142] Owing to our hypothesis that the vinyl-heterocyclic inhibitors have the ability to undergo a reversible thia-Michael addition with the active-site Cys_{25} of cruzain, it is necessary to consider scenarios of both non-covalent and covalent binding. To this end, we first predicted the binding patterns for NMePip-Phe-hPhe-vinyl-2Pyrmd (**9**) which has the same scaffold as **K11777** using Glide[167-169] and CovDock[170] modules embedded in the Schrodinger software package. In the covalent model, the binding of **9** with cruzain was highly conserved when compared to that of **K11777** (**Figure 2.2A**). The

33

N-1 of **9** was within hydrogen bonding distance of Gln_{19} and Trp_{184}, allowing the stabilization of the vinyl-heterocycle in a nearly analogous fashion to the sulfone moiety in **K11777**. In addition, the α-carbon of the inhibitor is positioned within 2.4 Å of His_{162}, an interatomic distance that would easily allow facile proton transfer between this carbon and the imidazole nitrogen, supporting our hypothesis that a reversible adduct could be formed with cruzain. The non-covalent model (**Figure 2.2B**) shared similar shape complementarity with the covalent binding pose, except that it was slightly shifted away from the binding site as compelled by the docking algorithm to avoid clashing with Cys_{25}. This suggested that covalent bond formation would only slightly perturb the non-covalent binding conformation. Overall, these data suggested that the binding of our newly designed compounds containing a vinyl-heterocyclic warhead have the ability to interact with cruzain in a very similar fashion to the characterized, irreversible inactivators of the enzyme. Similarly, we carried out docking for five other modified structures (**7**, **11, 12, 13,** and **15**) and their covalent-docking affinity values (Cdock affinity) are summarized in **Appendix A**. The corresponding inhibition constants (predicted K_i) converted from these affinity values ranged from 0.79 to 6.1 μM, and with the exception of compound **7**, were similar with the experimental values found in **Table 2.2**. In agreement with our dipeptide substrate kinetic data, we observed an increased Cdock affinity for **13** and **15** which each contained a hPhe in the P_1 position as compared to Phe. In addition, the substitution of a pyridine ring at the P_1' positions of **11, 12, 13,** and **15** may subtly improve the binding compared to the pyrimidine substituent of **7**. Further, the *N*-methylation of pyridine resulted in a fairly large shift in Cdock affinity, possibly resulting from an additional ion-ion/dipole interaction. On the basis of these
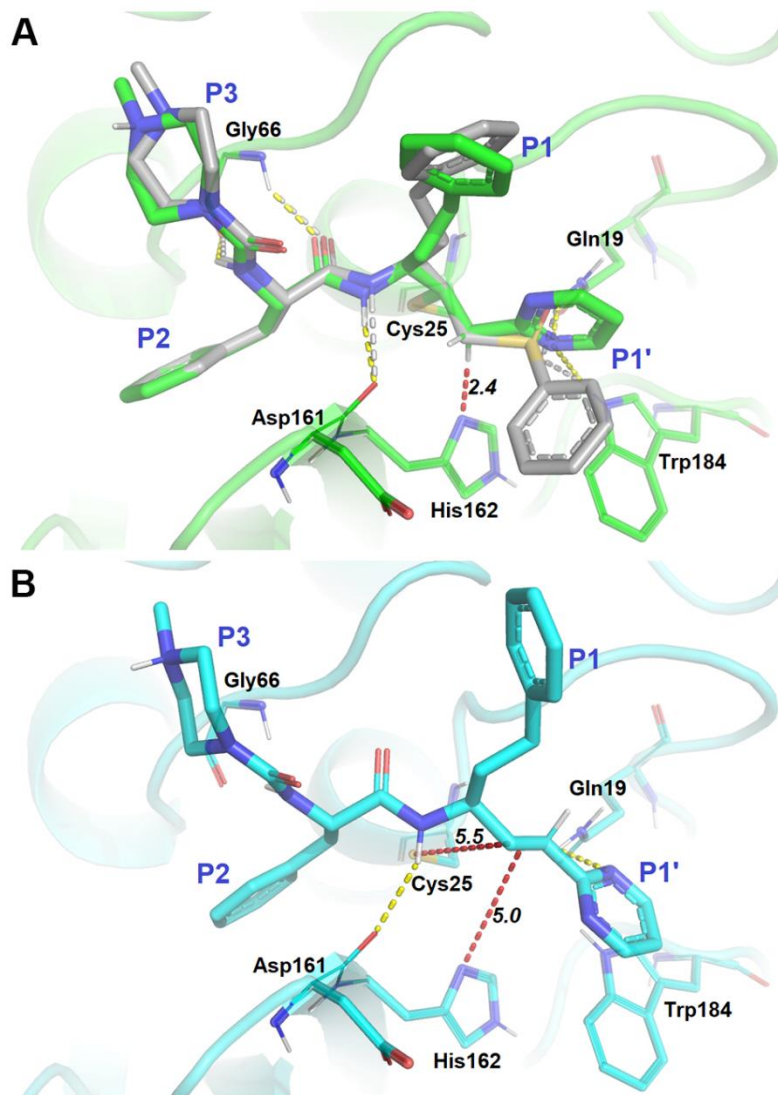
Figure 2.2 Molecular models of compound **9** bound to cruzain. (A) **K11777** (gray) is superimposed with a binding pose (green) in which a covalent bond is formed between the β-carbon of the vinyl group of **9**; (B) Binding pose (cyan) of **9** in which no covalent bond is formed with Cys_{25}. Yellow and white dashed lines represent hydrogen bonds with surrounding residues for **9** and **K11777**, respectively. Red dashed lines are measurements between catalytic dyad and vinyl moiety of **9**.

docking analyses, our inhibitor design focused on using a dipeptidic scaffold containing

Phe in the P_2 position and either Phe or hPhe in the P_1 position. We varied the identity of

the P_1' substituent in order to modify the electrophilicity of the olefin bond, but generally

maintained functional groups that possibly afford hydrogen bonding with Gln_{19} to

stabilize the binding of the compounds near Cys_{25} of cruzain.

2.2.2 Synthesis of PVHIs

The general synthetic routes employing either Wittig[171] or Horner-Wadsworth-Emmons (HWE)[172] reactions shown in **Scheme 2.1** were used to synthesize peptidomimetic vinyl-heterocyclic compounds from aldehydes and halo-methyl heterocycles.

Commercially-available Boc-protected L-amino acids phenylalanine, homophenylalanine and alanine (*a*) were converted to Weinreb amides[173] by T3P-catalyzed coupling to *N,O*-dimethylhydroxylamine hydrochloride to afford *b* (GP1,
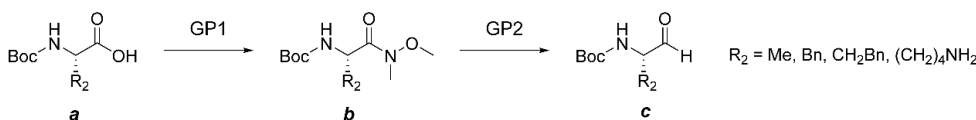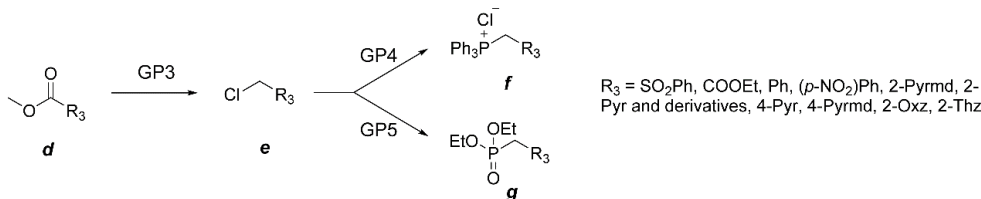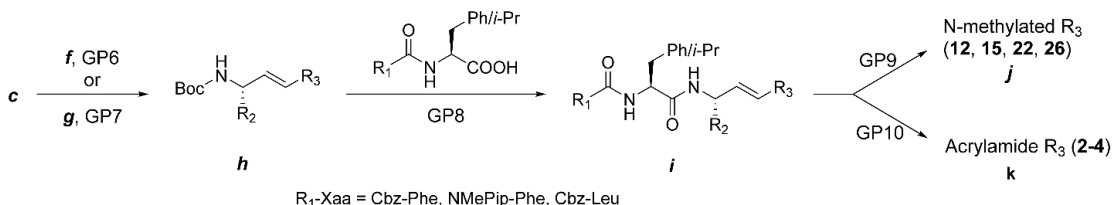

General Procedure 1 in Section 2.3.2). Reduction of the Weinreb amide using LAH at -

Scheme 2.1 General synthetic route to PVHIs. GP1. *N,O*-dimethylhydroxylamine, T3P, DIPEA, DCM, 0 °C; GP2. LAH, THF, -10 °C; GP3. 1) NaBH$_4$, EtOH, 0 °C, 2) SOCl$_2$, DCM; GP4. PPh$_3$, benzene, reflux; GP5. P(OEt)$_3$, 150 °C; GP6. LHMDS, THF, -70 °C – 0 °C; GP7. LHMDS, THF, -70 °C – 0 °C; GP8. 1) TFA, DCM, 0 °C, 2) R$_1$-Xaa-OH, T3P, DIPEA, DCM, 0 °C; GP9: MeI, MeCN, reflux; GP10: 1) LiOH, H$_2$O, 2)ClCOOEt, NH$_4$Cl.

10 °C in anhydrous THF provided the Boc-amino acid aldehyde (***c***, GP2), generally in overall yields of ~80% (***a-c***).

Phosphonium salts of methylheterocycles were in general prepared by derivatization of either the 2-methylcarboxy- or 2-hydroxymethyl-heterocycle (***d*** to ***f***, **Scheme 2.1**). Methyl 2-carboxy-pyrimidine (or pyridine, oxazole, and thiazole) was reduced using sodium borohydride to the primary alcohol, followed by conversion of the alcohol to the 2-chloromethyl-pyrimidine (***e***) using $SOCl_2$ or $POCl_3$ in DCM or $CHCl_3$ (GP3). Reaction of ***e*** with triphenylphosphine provided the Wittig reagent phosphonium salt (***f***) at overall yields of 28 – 80% (GP4). Wittig coupling of ***f*** with a peptide aldehyde (***c***) using LHMDS in anhydrous THF or sodium methoxide in benzene as base provided the peptide vinyl-heterocyclic product ***h*** (GP6), with general overall yields of 13 – 54%. Typically, the ratio of *E:Z* was 4:1, and separation of these regioisomers was readily achieved using silica gel column chromatography.

Alternatively, 2-chloromethyl-heterocycle ***e*** was converted to its phosphonate ***g*** by use of the Arbuzov reaction with triethyl-phosphite (~80% yields, GP5). The resulting phosphonate was de-protonated with LHDMS in THF, and then coupled with aldehyde ***c*** to provide the peptide vinyl-heterocycle ***h*** at 20% – 80% yield (GP7). The Boc group was removed quantitatively by treatment with TFA in DCM, then the free amine was coupled with P_3-P_2 fragment (R_1-Xaa-OH) using T3P to give the inhibitor ***i*** (GP8).

In addition, some of the PVHIs underwent *N*-methylation of the heterocycle (***j***, GP9). Further, we also prepared several acrylamides (***k***) through hydrolysis of corresponding acrylate ester and subsequent treatment with ethyl chloroformate and NH_4Cl (GP10). Final products were confirmed structurally by NMR and LCMS as

described in Section 2.3.2. It is important to note that proton NMR analysis of the products (*i*, *j* and *k*) indicated negligible epimerization at the α-carbon in these products, as evidenced by the absence of diastereomers.

### *2.2.3 Electrophilicity of Vinyl-heterocycles*

To evaluate the chemical reactivity of the vinyl group in our PVHIs and **K11777** we treated selected compounds with glutathione (GSH) at pH 8.0 to determine their reactivity in a thia-Michael addition of the sulfhydryl group of GSH with the vinyl group of the inhibitors. Normally, the addition of glutathione to an enzyme inhibitor is to be avoided, but here this serves as a means to evaluate the electrophilicity of these inhibitors. **K11777** and compounds **7**, **11, 12, 15, 17**, **25** and **26** (**Table 2.1**), which respectively contain a vinyl sulfone (**R₃ = I**), a vinyl-2-pyrimidine (**R₃ = IV**), a vinyl-2-pyridine (**R₃ = V**), a vinyl-2-N-methylpyridium (**R₃ = V**, **R₄ = Me**), a vinyl-2-(4-trifluoromethyl)-pyridine (**R₃ = V**, **R₅ = CF₃**), a vinyl-2-thiazole (**R₃ = IX**), and a vinyl-2-N-methylthiazolium (**R₃ = IX**, **R₄ = Me**). As previously reported,[174] the formation of a glutathione adduct with **K11777** was very slow ($k = 0.00028$ mM^{-1}min^{-1}, **Table 2.1**, **Figure 2.3**), and we were unable to ascertain an equilibrium constant for the **K11777**-GSH adduct. For the PVHIs, the reaction between GSH and the vinyl-2-pyrimidine (**7**), the vinyl-pyridine (**11)**, the vinyl-2-(4-trifluoromethyl)-pyridine (**17**), and the vinyl-thiazole (**25**), was negligible as no adduct was observed after 90 minutes incubation with either a 2:1 or 10:1 molar ratio of GSH:inhibitor. The electron-withdrawing 4-trifluoromethyl group on the pyridine of **17** had no effect on the electrophilicity of **11**. In contrast, addition of GSH to the vinyl group of vinyl-2-N-methylpyridinium (**12**, **15**) and

38

Table 2.1 Kinetic constants of thiolation of cruzain inhibitors[a]

Compound	Rate of Thiolation, k (mM^{-1}min^{-1})	K_{eq} (M^{-1})
K11777	0.00028 ± 0.0004	NA
7	Negligible	NA
11	Negligible	NA
12	0.037 ± 0.002	7400
15	0.054 ± 0.004	2400
17[b]	Negligible	NA
25	Negligible	NA
26	0.015 ± 0.005	930

[a]1 mM glutathione was mixed with 0.5 mM K11777 and compounds in Tris (pH 8.0), 10% DMSO (v/v) at room temperature. Aliquots were analyzed by LCMS as described; [b]5mM glutathione was used for **17**; NA, not applicable.

vinyl-2-N-methylthiazolium (**26**) in a 2:1 molar ratio resulted in the rapid formation of GSH adducts at respective rates of 0.037 mM^{-1}min^{-1}, 0.054 mM^{-1}min^{-1}, and 0.015 mM^{-1}min^{-1}, and apparent equilibrium was achieved for these compounds in 90 min (K_{eq} = 7400 M^{-1}, 2400 M^{-1}, and 930 M^{-1}, respectively). This demonstrated that the *N*-methylation of the PVHIs afforded a significant increase in the electrophilicity of the vinylic position, enabling rapid addition to thiols, owing to the strong electron-withdrawing effect of the methylpyridinium moiety. For example, compounds **11** and **12** are identical except for the *N*-methylpyridine group of compound **12**; compound **12** readily forms an adduct with GSH (97% conversion of **12** to its GSH adduct in 20 min at a 10:1 molar ratio of GSH to compound)**,** whereas compound **11** is unreactive towards GSH. Interestingly, the rate of thiolation of compound **12** is 50% that of **15**, while the values of K_{eq} indicated that the **12**-GSH adduct is three times more abundant than that of **15**-GSH. This suggested that the phenylalanyl sidechain of **12** may retard the addition of GSH to its vinyl group, and also slowed the presumed base-catalyzed elimination of GSH from its adduct with **12**. Overall, these results demonstrated that the reactivity of vinyl-
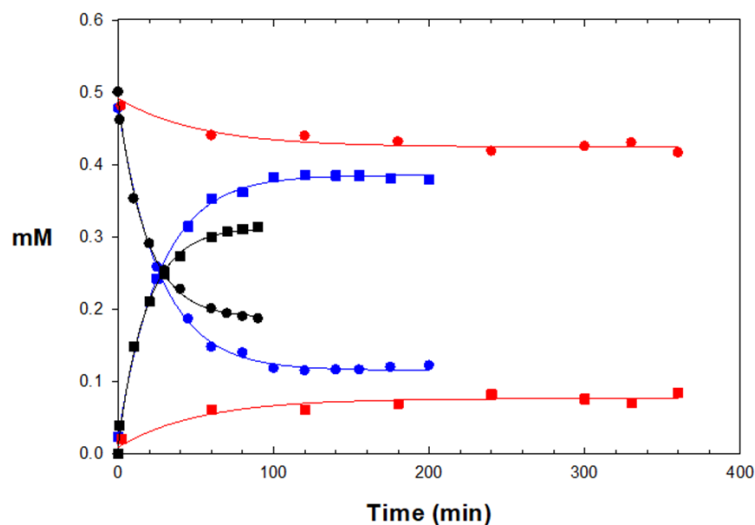
Figure 2.3 Time courses of depletion of **K11777** (red circles), **12** (blue circles), and **15** (black circles) upon formation of adducts (red, blue, and black squares, respectively) with glutathione. Lines drawn through for substrate depletion and adduct formation, were respectively, [Substrate] = (0.5 mM -A)(1-exp(-k*t)) + C and [Adduct] = A*(1-exp(-k*t)) + C, with resulting kinetic parameters found in **Table 2.1**.

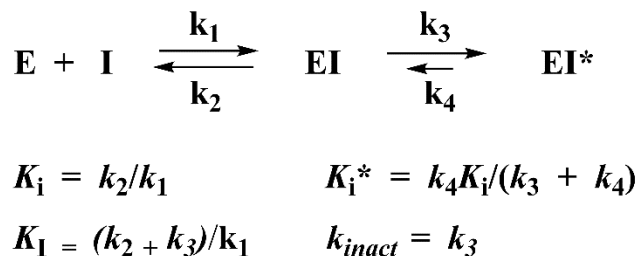heterocycles with GSH and presumably Cys_{25} vary with the nature of the heterocycles. Hence, it is possible to tune the electronic properties of the vinyl bond in the PVHIs, thereby allowing for development of modifiable electrophilic inhibitors of other enzymes that have an active-site cysteine or other nucleophile.

2.2.4 Kinetic Analysis of Cruzain Inhibitors and Inactivators

Scheme 2.2 is a kinetic depiction of inhibition and inactivation of cruzain and the relevant kinetic parameters.[175] The initial, and usually rapid, formation of EI is characterized by the inhibition constant K_i. For time-dependent inhibitors EI progresses to a second, tighter complex EI*, generally over the course of minutes, characterized by K_i^*, for which $K_i^* < K_i$ when $k_4 < k_3$. For irreversible covalent inactivators, k_4 and $K_i^* \sim$ 0, and the kinetic parameter k_{inact}/K_I is generally reported. For reversible time-dependent

$$E + I \xrightleftharpoons[k_2]{k_1} EI \xrightleftharpoons[k_4]{k_3} EI*$$

$$K_i = k_2/k_1 \qquad\qquad K_i* = k_4 K_i/(k_3 + k_4)$$
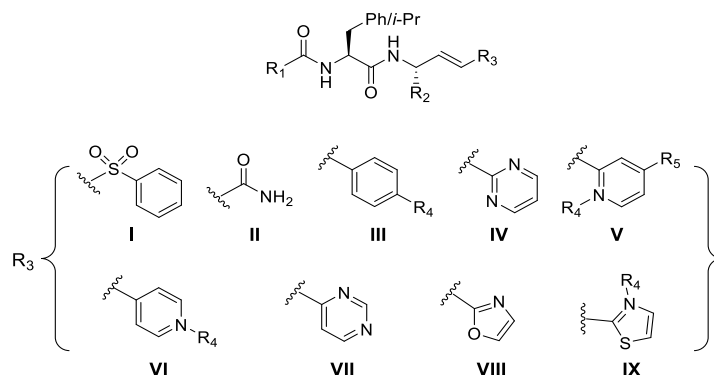
$$K_I = (k_2 + k_3)/k_1 \qquad k_{inact} = k_3$$

Scheme 2.2 Kinetic depiction of inhibition and inactivation.

inhibitors, initiation of reaction by adding enzyme to substrate and inhibitor leads to

concave-downward, curvilinear time courses of product formation in which reaction rates

demonstrably decrease as the EI* complex forms. Typical data, as exemplified for

compound **15**, are shown in **Figure 2.4**. Alternatively, extended pre-incubation of

enzyme and inhibitor, followed by dilution of the inhibitor and initiation of reaction with

high concentrations of substrate, leads to concave-upward curvilinear plots of product

formation as E reforms from EI* (Compound **15**, **Figure 2.4B**). Results of this analysis

for cruzain inhibitors and inactivators are collected in **Table 2.2.**

       K11777 comprises a useful benchmark compound despite the fact that it is an

irreversible inactivator of cruzain (reported kinetic data: apparent IC_{50} of 2 nM, $k_{inact}/K_I =$

234,000 $M^{-1}s^{-1}$).[158-159] We replaced the P_1 hPhe group of **K11777** with a Phe sidechain to

provide vinyl sulfone **1,** which had apparently equivalent potency ($K_i* = 3.6$ nM) to that

of **K11777**, but which, interestingly, exhibited kinetically reversible inhibition of cruzain.

However, a crystal structure we obtained for **1** bound to cruzain indicated the formation

of a C-S bound between Cys_{25} and **1** (**Appendix B**). This may indicate that a

phenylalanyl group at the P_1 position partly impedes the ability of an adjacent vinyl

electrophile to access Cys_{25}, as was observed with the solution phase GSH addition

studies to our PVH compounds.

41

Table 2.2 Kinetic data of PVHIs of cruzain[a]

Compound	Structure	R_1	R_2	R_3	Cruzain Inhibition or Inactivation		
					K_i (μM)	K_i* (μM)	k_{inact}/K_{inact} (M^{-1}s^{-1})
K11777	NMePip-Phe-hPhe-VSPh	NMePip	CH$_2$Bn	I	NA	0.002[b]	234,000[b,c]
1	Cbz-Phe-Phe-VSPh	BnO	Bn	I	ND	0.0036 ± 0.0001	ND
2	Cbz-Phe-Phe-vinyl-CONH$_2$	BnO	Bn	II	37 ± 2	NA	21.7 ± 0.8
3	Cbz-Phe-hPhe-vinyl-CONH$_2$	BnO	CH$_2$Bn	II	3 ± 1	NA	1700 ± 500
4	NMePip-Phe-hPhe-vinyl-CONH$_2$	NMePip	CH$_2$Bn	II	3.4 ± 0.4	NA	1900 ± 200
5	Cbz-Phe-Phe-vinyl-Ph	BnO	Bn	III	1.8 ± 0.1	0.87 ± 0.05	NA
6	Cbz-Phe-Phe-vinyl-(4-NO$_2$)Ph	BnO	Bn	III, R_4 = NO$_2$	ND	0.37 ± 0.02	NA
7	Cbz-Phe-Phe-vinyl-2Pyrmd	BnO	Bn	IV	28 ± 1	0.364 ± 0.004	NA
8	Cbz-Phe-hPhe-vinyl-2Pyrmd	BnO	CH$_2$Bn	IV	>35	NA	NA
9	NMePip-Phe-hPhe-vinyl-2Pyrmd	NMePip	CH$_2$Bn	IV	>10	2.2 ± 0.1	NA
10	Cbz-Phe-Ala-vinyl-2Pyrmd	BnO	Me	IV	58 ± 6	25 ± 1	NA
11	Cbz-Phe-Phe-vinyl-2Pyr	BnO	Bn	V	5.5 ± 0.4	0.31 ± 0.01	NA
12	Cbz-Phe-Phe-vinyl-2PyrNMe	BnO	Bn	V, R_4 = Me	3.8 ± 0.4	0.28 ± 0.08	NA
13	Cbz-Phe-hPhe-vinyl-2Pyr	BnO	CH$_2$Bn	V	1.06 ± 0.07	0.171 ± 0.004	NA
14	NMePip-Phe-hPhe-vinyl-2Pyr	NMePip	CH$_2$Bn	V	ND	3.4 ± 0.1	NA
15	Cbz-Phe-hPhe-vinyl-2PyrNMe	BnO	CH$_2$Bn	V, R_4 = Me	0.76 ± 0.04	0.126 ± 0.004	NA
16	Cbz-Phe-hPhe-vinyl-2-(4-OMe)-Pyr	BnO	CH$_2$Bn	V, R_5 = OMe	>5	NA	NA
17	Cbz-Phe-hPhe-vinyl-2-(4-CF$_3$)-Pyr	BnO	CH$_2$Bn	V, R_5 = CF$_3$	NA	0.57 ± 0.05	NA
18	Cbz-Leu-hPhe-vinyl-2Pyr	BnO	CH$_2$Bn	V	7.8 ± 0.6	1.42 ± 0.09	NA
19	Cbz-Phe-Ala-vinyl-2Pyr	BnO	Me	V	ND	4.8 ± 0.2	NA
20	Cbz-Phe-Lys-vinyl-2Pyr	BnO	(CH$_2$)$_4$NH$_2$	V	17.3 ± 0.3	0.87 ± 0.02	NA
21	Cbz-Phe-Phe-vinyl-4Pyr	BnO	Bn	VI	ND	5.5 ± 0.2	NA
22	Cbz-Phe-Phe-vinyl-4PyrNMe	BnO	Bn	VI, R_4 = Me	92 ± 5	4.0 ± 0.1	NA
23	Cbz-Phe-Phe-vinyl-4Pyrmd	BnO	Bn	VII	10.8 ± 1.4	1.14 ± 0.07	NA
24	Cbz-Phe-Phe-vinyl-2Oxz	BnO	Bn	VIII	10 ± 1	0.71 ± 0.01	NA
25	Cbz-Phe-Phe-vinyl-2Thz	BnO	Bn	IX	ND	1.71 ± 0.09	NA
26	Cbz-Phe-Phe-vinyl-2ThzNMe	BnO	Bn	IX, R_4 = Me	ND	0.94 ± 0.06	NA

[a]Data obtained at 25°C, pH 7.5; [b]Reported as apparent IC$_{50}$ in ref 158; [c]Reported as 32,500 M^{-1}s^{-1} (pH 8.0) in ref 159; NA, not applicable; ND, not determined; K_i and K_i* are respectively, the initial and tight-binding inhibition constants.

We next evaluated three C-terminal acrylamides (R_3 = II) within the Cbz-Phe-Phe, Cbz-Phe-hPhe, and NMePip-Phe-hPhe scaffolds (2 – 4). The acrylamides within the
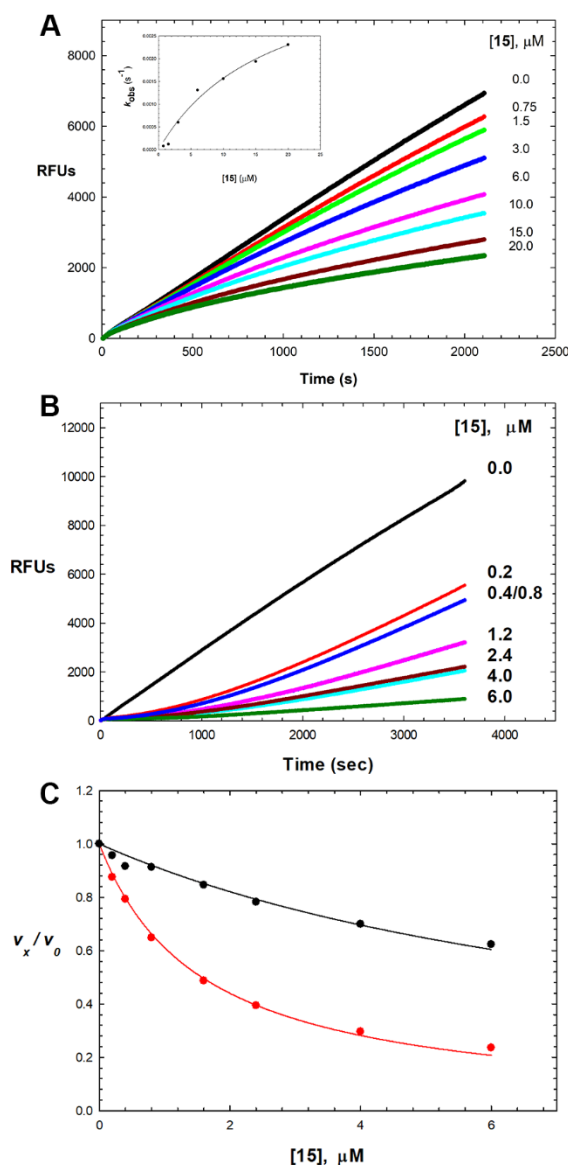
Figure 2.4 Time-dependent inhibition of cruzain by **15**. (A) Reaction initiated by addition of cruzain (0.1 nM) with Cbz-Phe-Arg-AMC (10 μM) and 0-20 μM **15** (pH 7.5). Lines drawn through the experimental data points were from fitting of each inhibitor concentration to eq. 2-3, from which the replot of k_{obs} vs. [**15**] is shown in the inset (fitting to eq. 2-4: $K_i = 2.0 \pm 0.9$ μM, $k_3 = 0.004 \pm 0.001$ s^{-1}, and $k_4 \sim 0$); (B) Following 1 h pre-incubation of cruzain (0.1 nM) with 0-6 μM **15**, reaction was initiated by addition of Cbz-Phe-Arg-AMC (10 μM). (C) Fitting of cruzain inhibition by compound **15** for v_i/v_0 (black) and v_s/v_0 (red) using eq. 2-5 with results of this found in **Table 2.2**.

Cbz-Phe-hPhe and NMePip-Phe-hPhe scaffolds afforded apparently irreversible covalent inactivation (k_{inact}/K_I = 1700-1900 M^{-1}s^{-1}), while Cbz-Phe-Phe-acrylamide (**2**) was less effective (k_{inact}/K_I = 22 M^{-1}s^{-1}). Comparing the values of k_{inact}/K_I for **K11777** and **4**

indicated that the vinyl sulfone is overwhelmingly more effective as a covalent inactivator than its acrylamide counterpart, possibly owing to hydrogen bond contacts of the sulfone oxygen with Gln_{19}, which position the vinyl group proximal to Cys_{25} of cruzain. As with **1**, a Phe rather than a hPhe group at the P_1 position, may retard covalent formation over the time course of kinetic analysis when one compares the rates of apparent inactivation of **2** vs. **3** and **4**, as seen with peptide substrates.

We therefore sought to explore the effects of replacement of both the vinyl-phenylsulfone and acrylamide groups with a phenyl and heterocyclic groups conjugated to the vinyl group. The Cbz-Phe-Phe-vinyl-benzene compound **5** is a time-dependent inhibitor of cruzain ($K_i^* = 0.87$ μM), but substitution of the *para* position of the phenyl ring with an electron-withdrawing nitro group (compound **6**) led to a nearly 3-fold improvement in potency ($K_i^* = 0.34$ μM), suggesting that the vinyl group of **6** is more capable of thiolation by the cruzain. As seen with **1**, these compounds also demonstrated reversible inhibition of cruzain, possibly due to the P_1 phenylalanine. Due to poor aqueous solubility (solubility of **5** and **6** ≤ 2 μM in 10% DMSO), the inhibitors containing vinyl-benzene were not explored further.

Subsequently, six heterocyclic groups ($R_3 = $ **IV-IX**) conjugated to the presumed electrophilic vinyl group were evaluated within several dipeptide scaffolds. The vinyl-2-pyrimidine ($R_3 = $ **IV**), vinyl-2-pyridine ($R_3 = $ **V**), vinyl-2-oxazole ($R_3 = $ **VIII**), and vinyl-2-thiazole ($R_3 = $ **IX**) groups, unlike the vinyl-4-pyridine ($R_3 = $ **VI**) and vinyl-4-pyrimidine ($R_3 = $ **VII**), maintain bioisosteric similarity to the reactive acrylamides and vinyl sulfones, which is reflected in their more potent inhibition of cruzain as detailed

below. Most of these compounds induced time-dependent inhibition on cruzain and were found to be kinetically reversible with residence time (τ) of 6-20 min.

The vinyl-2-pyrimidine moiety ($\mathbf{R_3} = \mathbf{IV}$) in the Cbz-Phe-Phe scaffold afforded compound **7**, which exerted time-dependent inhibition of cruzain with an initial value of $K_i = 5$ μM and subsequent tight-binding inhibition of $K_i^* = 0.38$ μM. Substitution of the phenyl group of **5** by a pyrimidine group greatly improved the solubility of **7** (≥ 100 μM in 10% DMSO). Extended pre-incubation with **7**, followed by dilution, and addition of an excess of substrate, resulted in slow recovery of cruzain activity, indicating that any covalent reaction between cruzain and **7** was kinetically reversible ($k_4 = 0.0018 \pm 0.0003$ s^{-1}; $\tau = 9$ min). Interestingly, when the 2-pyrimidinyl moiety is appended to Cbz-Phe-hPhe (**8**), the resulting compound is a poor inhibitor of cruzain ($K_i > 35$ μM); however, when the 2-pyrimidinyl group is attached to afford the same scaffold as **K11777,** we obtained an inhibitor of low micro-molar potency (**9**, $K_i^* = 2.2$ μM). Substitution of the P$_1$ Phe with Ala (**10**, $K_i = 25$ μM) produced a poor inhibitor of cruzain, indicating the essentiality of a larger sidechain in the P$_1$ position, as was observed with dipeptide substrates. To probe the importance of the vinyl group for the inhibition of cruzain, we prepared an analogue in which the vinyl group of **7** was reduced (compound **27**). This inhibitor lacked time-dependent behavior ($K_i = 22$ μM), and was 100-fold less potent than its vinyl analogue **7**, which demonstrated the importance of the vinyl group for the inhibition of cruzain. We prepared inhibitor **23** which contains a vinyl-4-pyrimidinyl ($\mathbf{R_3} = \mathbf{VII}$) group that does not maintain bioisosteric similarity to the acrylamides. **23** exhibited three-fold less potency than the bioisosteric vinyl-2-pyrimidine (**7**). Similarly,

inhibitors containing the vinyl-4-pyridyl ($\textbf{R}_3 = \textbf{VI}$) (**21**, **22**) lack bioisosteric equivalence to the acrylamides and were found to be only modest inhibitors of cruzain.

Inhibitors containing a vinyl-2-pyridinyl group ($\textbf{R}_3 = \textbf{V}$) were explored more widely. Cbz-Phe-Phe-vinyl-2-pyridine **11** exhibited time-dependent inhibition of cruzain with an initial value of $K_i = 5.5$ μM and subsequent tight-binding inhibition of $K_i^* = 0.31$ μM ($k_4 = 0.0012 \pm 0.0002$ s^{-1}; $\tau = 13$ min), and solubility of **11** was ≤ 30 μM in 10% DMSO. Unlike the vinyl-pyrimidinyl group of **7**, placement of the vinyl-2-pyridinyl group in the Cbz-Phe-hPhe scaffold improved inhibition by 3-fold (**13**, $K_i^* = 0.17$ μM), while the vinyl-2-pyridinyl group was much less effective in the NMePip-Phe-Phe scaffold (**14**, $K_i^* = 3.4$ μM). Substitution of an electron-donating methoxy group on the pyridine ring (**16**) of the Cbz-Phe-hPhe scaffold diminished the inhibitory activity of the vinyl-2-pyridinyl heterocycle compared to its unsubstituted counterpart **13** by >50-fold, suggesting that the methoxy group is large enough to create a steric barrier to inhibitor binding. In contrast, the substitution at C-4 of the pyridine with the electron-withdrawing trifluoromethyl group resulted in modest inhibition (**17**, $K_i^* = 0.57$ μM), but nonetheless was less potent than the unsubstituted pyridine **13**. Apparently, this result arises from steric crowding as OMe > CF$_3$ > H, implicating that substitution at the C-4 position of the heterocycles are not well tolerated.

We next investigated how the P$_1$ and P$_2$ sidechains of these vinyl-2-pyridinyl inhibitors effect inhibition. The replacement of the P$_2$ Phe with Leu resulted in diminished potency (**18**, $K_i^* = 1.42$ μM) compared to the Cbz-Phe-Phe and Cbz-Phe-hPhe scaffolds, overall demonstrating that inhibitors with bulky hydrophobic substituents in P$_1$ and P$_2$ enhanced binding to cruzain. To analyze how short alkyl and charged groups

effected inhibition, we prepared Cbz-Phe-Ala-vinyl-2-pyridine (**19**) and Cbz-Phe-Lys-vinyl-2-pyridine (**20**). We found that the Cbz-Phe-Lys scaffold, which mimics our most optimal substrate, Cbz-Phe-Arg-AMC, exhibited good inhibition ($K_i^* = 0.87$ μM), whereas **19** was a poor inhibitor ($K_i^* = 4.8$ μM), in concert with the poor substrate activity of Cbz-Phe-Ala-AMC.

Seeking to improve the electrophilicity of the vinyl-2-pyridinyl group, we prepared N-methylated analogues **12** and **15**. This modification resulted in improved aqueous solubility (≥ 50 μM in 10% DMSO), and provided potent time-dependent inhibition of cruzain (**12**, $K_i^* = 0.28$ μM; **15**, $K_i^* = 0.126$ μM) comparable to, or exceeding, the inhibition exerted by their un-methylated counterparts (**11** and **13**). Inhibition data for compound **15** were fitted by all methods outlined in Section 2.3.10. We fitted each curve in **Figure 2.4A** to eq. 2-3, and the resulting values of k_{obs} were re-plotted vs. [**15**] (inset), which demonstrated a hyperbolic dependence of the inhibitor (fitting to eq. 2-4: $K_i = 2.0 \pm 0.9$ μM, $k_3 = 0.004 \pm 0.001$ s^{-1}, and $k_4 \sim 0$). Pre-incubation of cruzain and variable concentrations of **15**, followed by initiation of reaction by the addition of substrate, produced time courses like that shown in **Figure 2.4B**. These data demonstrated a significant lag phase for recovery of cruzain activity, indicative of the slow desorption of the inhibitor, with or without the formation of a covalent bond with Cys$_{25}$. Finally, analysis of inhibition of cruzain by **15** at early and late phases of the time courses by fitting to eq. 2-5 provided values of $K_i = 0.76 \pm 0.04$ μM and $K_i^* = 0.126 \pm 0.004$ μM (**Table 2.2**). Of note, in pre-incubation studies, all PVHIs which contain the vinyl-pyridinyl substituent displayed kinetic reversibility.

We investigated 5-membered ring heterocycles that are bioisosteric with acrylamide inactivators. The syntheses of the vinyl-2-oxazole (**24**), the vinyl-2-thiazole (**25**) and its N-methylated counterpart (**26**) into the Cbz-Phe-Phe scaffold proved facile, and provided useful inhibitors. Vinyl-2-oxazole **24** was a sub-micromolar inhibitor of cruzain ($K_i^* = 0.71$ μM). Vinyl-2-thiazole inhibitors **25** and **26** were cruzain inhibition of similar potency ($K_i^* = 1.71$ and 0.94 μM, respectively), for which N-methylation of the thiazole improved potency by nearly two-fold.

Cruzain inhibitors **5** – **26** allowed the evaluation of six heterocyclic groups (R_3 = **IV-IX**) appended to the presumed electrophilic vinyl group within several dipeptide scaffolds. The vinyl-2-pyrimidine, vinyl-2-pyridine, vinyl-2-N-methylpyridinium, vinyl-2-oxazole, and vinyl-2-thiazole substituents, unlike the vinyl-4-pyrimidine and vinyl-4-pyridine heterocycles, maintain bioisosteric similarity to the reactive acrylamides, and provided potent, time-dependent inhibitors in accord with our hypothesis. Of these PVHIs, the 2-pyridine, the charged 2-N-methylpyridinium, and the vinyl-2-pyrimidine present the most interesting heterocycles for further exploration. The inhibition of cruzain displayed by these PVHIs may be due to the reversible formation of an adduct with active-site Cys_{25}, as is supported by the loss of time dependent inhibition when the vinyl group is saturated. Importantly, we have no evidence that such a reversible covalent bond is formed, and ongoing studies are underway to address this point.

2.2.5 Selectivity of PVHIs for Cruzain over Homologous Human Cathepsins

Cruzain has, respectively, 25%, 15%, and 23% amino acid identity with human cathepsins L, B, and S.[142, 161] It is preferable to proceed with cruzain inhibitors that do not readily inhibit these human lysosomal cathepsins which might engender cellular toxicity. We evaluated selected cruzain inhibitors vs. the human cysteine proteases cathepsins L,

Table 2.3 Enzymatic selectivity of cruzain inhibitors[a]

Compound	Structure	K_i* (μM)			
		cruzain	human Cathepsin L	human Cathepsin B	human Cathepsin S
K11777	NMePip-Phe-hPhe-VSPh	IC_{50} = 0.2 nM[b]	IC_{50} = 0.2 nM[b]	IC_{50} = 5.7 nM[b]	IC_{50} = 0.6 nM[b]
7	Cbz-Phe-Phe-vinyl-2Pyrmd	0.29 ± 0.01	1.1 ± 0.1	32 ± 3	0.37 ± 0.02
11	Cbz-Phe-Phe-vinyl-2Pyr	0.29 ± 0.02	4.3 ± 0.5	28 ± 4	1.8 ± 0.3
12	Cbz-Phe-Phe-vinyl-2PyrNMe	0.31 ± 0.02	0.70 ± 0.04	19 ± 4	0.87 ± 0.07
13	Cbz-Phe-hPhe-vinyl-2Pyr	0.123 ± 0.0004	1.9 ± 0.2	6.5 ± 0.9	1.41 ± 0.06
15	Cbz-Phe-hPhe-vinyl-2PyrNMe	0.089 ± 0.002	0.88 ± 0.06	37 ± 6	0.32 ± 0.04
24	Cbz-Phe-Phe-vinyl-2Oxz	0.71 ± 0.01	0.29 ± 0.03	90 ± 14	1.51 ± 0.05

[a]Inhibition data obtained at pH 5.5, 25°C in 10% DMSO (v/v); [b]Reported as apparent IC_{50} in Ref. 158.

B, and S (**Table 2.3**). For this selectivity comparison all inhibition data were obtained at pH 5.5, for which K_i* values were invariant for all inhibitors except compound **15** (K_i* = 89 nM). The cruzain inhibitors demonstrated moderate selectivity vs. cathepsins L and S (generally, 3-fold or greater), while all of these inhibitors displayed 40-fold or higher selectivity vs. cathepsin B. The vinyl-2-pyridine inhibitors **13** and **15** are particularly selective as their K_i values are over 10-fold lower than the corresponding values with the three human cathepsins. In contrast, **K11777** showed potent inactivation at nanomolar concentrations for all three human cathepsins; this apparent lack of selectivity possibly arising from its irreversible mode of inactivation. These results suggest that suitable selectivity for reversible cruzain inhibitors may be more easily attained than for irreversible ones.

*2.2.6 Effects of PVHIs in Axenic Cultures of T. cruzi and in a Cell Model of T. cruzi*

Infection

Initially, we tested selected compounds against epimastigotes of *T. cruzi* (strain Y, ATCC 50832GFP) in axenic culture. As is observed here (**Table 2.4**), and has been shown previously, **K11777** weakly inhibited the growth of *T. cruzi* epimastigotes (EC_{50} ~ 60 μM).[176] PVHIs **7, 12,** and **15** inhibited the growth of epimastigotes of *T. cruzi* (EC_{50} = 2-20 μM) while **11, 13,** and **24** were poorly effective. Compounds **12** and **15** were comparably potent against cultures of *T. cruzi* (EC_{50} = 8.6 and 2.1 μM, respectively), and were, at a minimum, 10-fold more active than **K11777**.

Table 2.4 Effects of cruzain inhibitors on trypanosome and human cell growth[a]

Compound	cruzain K_i* (μM)	*T. cruzi* axenic culture EC_{50} (μM)	*T. cruzi*-infected cardiomyoblasts (C2C12) EC_{50} (μM)	*T. b. brucei* PCFs EC_{50} (μM)	*T. b. brucei* BSFs EC_{50} (μM)	human cell cytotoxicity CC_{50} (μM)	C2C12 cytotoxicity CC_{50} (μM)	selectivity index CC_{50}/EC_{50}
K11777	IC_{50} = 2 nM[b]	>20	0.7 ± 0.2	1.7 ± 0.5	0.09 ± 0.06	60-100	>10	140
7	0.364 ± 0.004	20	9.0 ± 0.5	7.1 ± 0.9	10.4 ± 0.2	>100	>10	>10
9	2.2 ± 0.1	20	ND	15 ± 2	>20	ND	ND	ND
11	0.31 ± 0.01	>20	4.9 ± 0.2	5 ± 1	5 ± 4	>100	>10	>20
12	0.28 ± 0.08	8.7 ± 0.1	9.9 ± 0.5	13 ± 3	6.6 ± 0.6	>100	>10	>10
13	0.171 ± 0.004	>20	5.9 ± 0.3	>10	4 ± 2	>100	>10	>20
15	0.126 ± 0.004	2.1 ± 0.1	5.4 ± 0.9	5.9 ± 0.2	2.8 ± 0.1	>100	>10	>20
24	0.71 ± 0.01	>20	ND	27 ± 5	ND	>100	ND	ND

[a]Effects of inhibitors were evaluated as EC_{50} for axenic *T. cruzi*, PCFs/BSFs of *T. b. brucei* and *T. cruzi*-infected murine cardiomyoblasts. The selectivity index is the ratio of inhibitor cytotoxicity in human dermal fibroblasts (CC_{50})/ trypanosomacidal activity (EC_{50}) in infected cardiomyoblasts. [b]Ref. 158.

Selected cruzain inhibitors were further evaluated in a more relevant cellular model of Chagas disease: *T. cruzi*-infected murine cardiomyoblasts (C2C12 cells) (**Table 2.4**, **Figure 2.5**). Inhibitors **7, 11, 12, 13,** and **15** exhibited antiparasitic efficacy at values of EC_{50} = 5 – 10 μM, while displaying no cytotoxicity against the host cardiomyoblasts (CC_{50} > 10 μM). These EC_{50} values demonstrated that the anti-trypanosomal activities of the reversible PVHIs are within an order of magnitude of potency of the irreversible
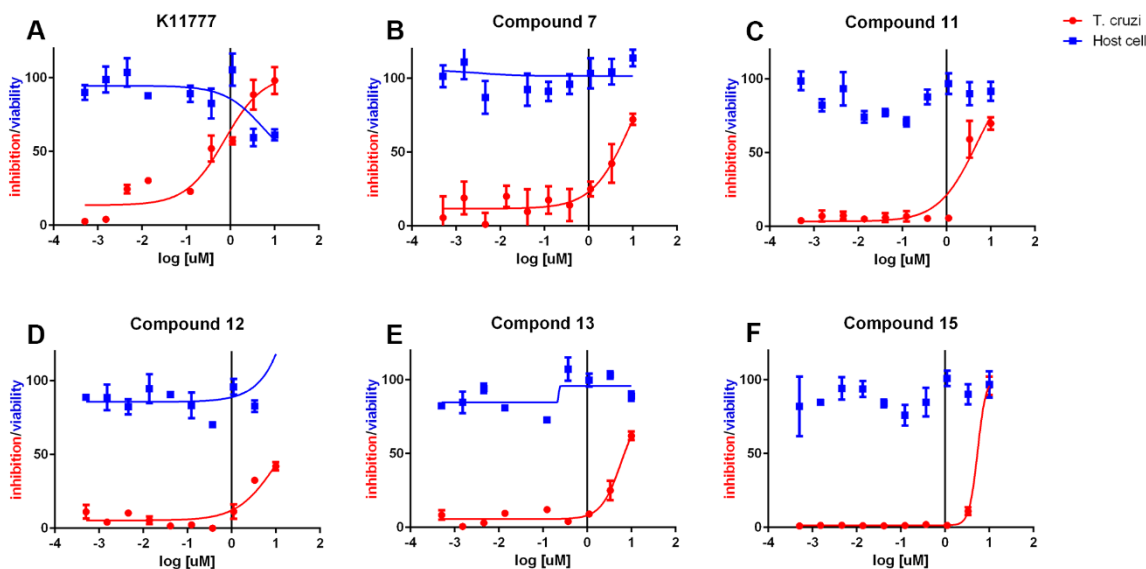
Figure 2.5 Effects of cruzain inhibitors on growth of *T. cruzi*-infected murine cardiomyoblasts. Growth of inhibition of *T. cruzi* (red) is superimposed with the viability of the cardiomyoblasts (blue).

inactivator, **K11777** (EC_{50} = 0.7 μM), despite the large difference in activity vs. cruzain. Accordingly, the PVHIs, while reversible in action, and with no apparent mammalian or human cytotoxicity, are nearly as effective as the potent, irreversible inactivator **K11777**. Further, the best of the PVHIs are less than 3-fold less potent than the currently-used anti-chagasic drug benznidazole (LD_{50} = 1.5 μM)[161] suggesting that a second generation of PVHIs may provide clinical candidates.

2.2.7 Effects of PVHIs in Axenic Cultures of T. b. brucei

We additionally tested our cruzain inhibitors in axenic cultures of the related protozoan *T. b. brucei*, owing to the high structural similarity and reported essentiality of the cysteine proteases brucipain (TbCatL) in *T. b. brucei*.[177-179] It has been demonstrated that the cruzain inhibitor **K11777** is active in cellular cultures of both *T. b. brucei* and *T.*

cruzi, supporting the notion that our PVHIs could be effective in growth inhibition of both species of parasite. For insect PCFs of *T. b. brucei* (ATCC PRA-381), compounds **7, 9, 11, 12,** and **15** demonstrated growth inhibition at EC_{50} values of 5-15 μM (**Table 2.4,**
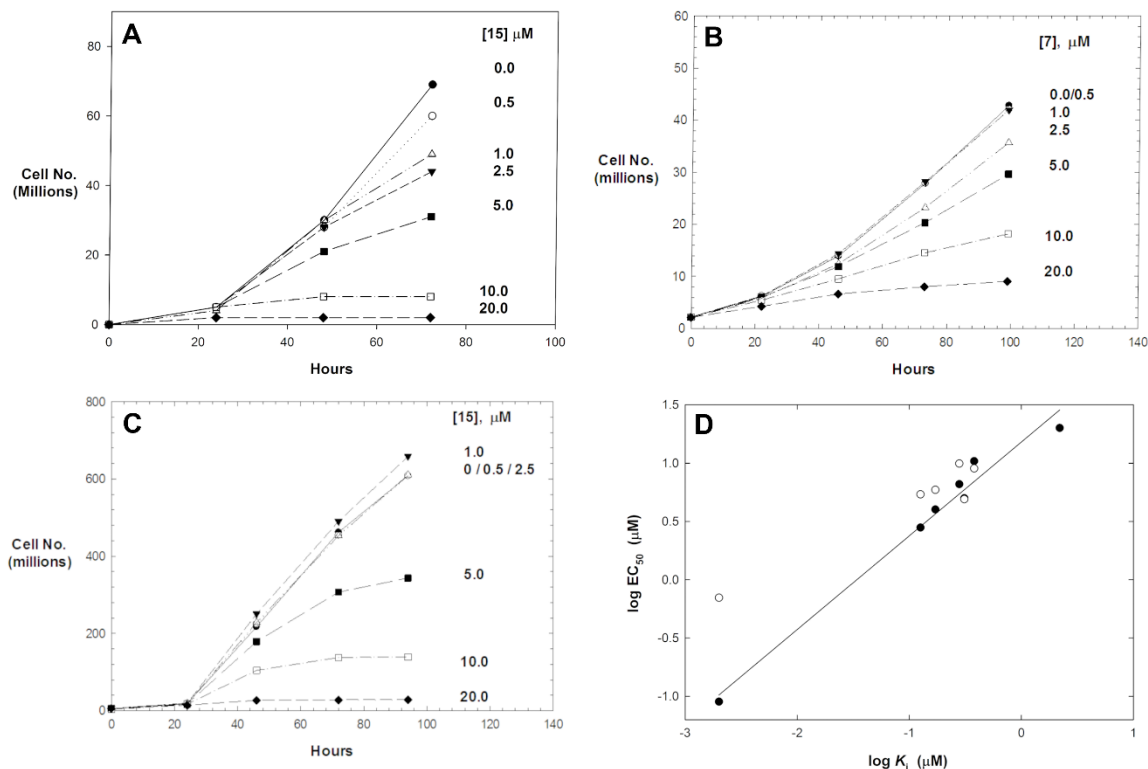
Figure 2.6 Cell-growth inhibition of *T. b. brucei*. (A) Inhibition of bloodstream forms by **15**; (B) Inhibition of bloodstream forms by **7**; (C) Inhibition of procyclic forms by **15**; (D) A correlation plot of values of EC_{50} for trypanocidal activity vs. *T. b. brucei* BSFs (closed circles, $r^2 = 0.979$, slope = 0.80) and *T. cruzi* in murine cardiomyoblasts (open circles).

Figure 2.6). When compared to **K11777** ($EC_{50} = 1.7$ μM), these PVHIs exhibited potent cell-growth inhibition. For example, compound **15** ($EC_{50} = 5.9$ μM) was only 3-fold less potent vs. *T. b. brucei* than **K11777**. Values of EC_{50} for these PVHIs roughly correlated with their values of K_i*, with the exception of compound **13**.

We next evaluated these inhibitors in axenic cultures of human bloodstream forms (BSFs) of *T. b. brucei* (ATCC PRA-383). All PVHIs that were active vs. procyclic forms

of *T. b. brucei* were also trypanocidal vs. the bloodstream forms, but with equal or lower

EC_{50} values compared to the procyclic forms (**Table 2.4**). Compared to PVHIs that had

similar potencies in both PCFs and BSFs, **K11777** was nearly 20-fold more potent in *T.*

b. brucei BSFs than in PCFs. These results suggested that a cathepsin L-like cysteine

protease in *T. b. brucei,* such as brucipain (or *Tb*CatL),[177] is essential for growth of

procyclic and bloodstream *T. b. brucei*, but perhaps an additional cysteine protease, such

as TbCatB, is also essential in BSFs of *T. b. brucei*, as this enzyme is sensitive to **K11777**

but not to the PVHIs. This is similar to the findings of Yang *et al.* who showed using an

activity-based protein probe of **K11777** that TbCatB and brucipain (TbCatL) are both

labeled in BSFs of *T. b. brucei* while only brucipain is labeled in PCFs.[180] This could

explain the exceptional trypanocidal activity of **K11777** in BSFs. This will be the focus

of our future studies. Nonetheless, the activity of the PVHIs vs. *T. b. brucei* BSFs may

hold promise for progression to their evaluation in models of African trypanosomiasis.

Interestingly, the values of EC_{50} obtained for PCFs of *T. b. brucei* and amastigotes

of *T. cruzi* were nearly identical for most PVHIs, despite their more modest inhibition of

cruzain. Shown in **Figure 2.6D** is a correlation plot of log EC_{50} for anti-trypanosomal

activity for bloodstream forms of *T. b. brucei* and the amastigote forms of *T. cruzi* from

the murine cardiomyoblast infection model. For the former, the correlation is excellent (r^2

$= 0.979$, slope $= 0.80$), and the activity against parasites is nearly a 1:1 correlation with

logK_i with these inhibitors. This result provided support that our PVHIs are targeting a

cruzain-like protease in *T. b. brucei.* For *T. cruzi*, this correlation is not as strong, in part

due to an absence of a sufficient range of data. We have also compared the cytotoxicity

of selected inhibitors in human dermal fibroblasts vs. *T. cruzi*-infected cardiomyoblasts

(selectivity index in **Table 2.4**), which demonstrates the PVHIs are more than 10-fold

selective for trypanosomes vs. human cells.

2.3 Materials and Methods

2.3.1 General Information of Synthetic Chemistry

All reagents and starting materials were obtained from commercial suppliers and

used without further purification unless otherwise stated. Reactions were run under an

atmosphere of nitrogen or argon and at ambient temperature unless otherwise noted.

Reaction progress was monitored using thin layer chromatography and by analysis

employing an HPLC-MS (UltiMate 3000 equipped with a diode array coupled to a MSQ

Plus Single Quadrupole Mass Spectrometer, ThermoFisher Scientific) using electrospray

positive and negative ionization detectors. Reported liquid chromatography retention

times (Rt) were established using the following conditions: column: Phenomenex Luna 5

μm C18(2) 100 Å, 4.6 mm, 50 mm, Mobile phase A: water with 0.1% formic acid (v/v).

Mobile phase B: MeCN with 0.1% formic acid (v/v). Temperature: 25 °C. Gradient:

0−100% B over 6 min, then a 2 min hold at 100% B. Flow: 1 mL/min. Detection: MS and

UV at 254, 280, 214, and 350 nm.

Semi-preparative HPLC purification of compounds was performed on a Thermo

Fisher Scientific UltiMate 3000 with a single wavelength detector coupled to a fraction

collector. Purifications were conducted using the following conditions: column:

Phenomenex Luna 5 um C18(2) 100 Å, 21.2 mm, 250 mm, Mobile phase A: water with

0.1% formic acid (v/v). Mobile phase B: MeCN with 0.1% formic acid (v/v).

Temperature: Room temperature. Gradient: 0−100% B over 30 min, then a 5-min hold at 100% B. Flow: 20 mL/min. Detection: UV (254 nm).

$^1H/^{13}C$ NMR magnetic resonance spectra were obtained in CDCl$_3$, CD$_3$OD, or DMSO-d_6 at 400MHz/100MHz at 298 K on a Bruker Avance III NanoBay console with an Ascend magnet unless otherwise noted. The following abbreviations were utilized to describe peak patterns when appropriate: br = broad, s = singlet, d = doublet, q = quartet, t = triplet, and m = multiplet. All final compounds used for testing in assays and biological studies had purities that were determined to be >95% as evaluated by their proton NMR spectra and their HPLC/MS based on ultraviolet detection at 254 nm. Similar RP-HPLC conditions were used for the experiments of GSH addition to vinyl-heterocycles. Masses detected were in the range 100 − 1000 Da and were detected in positive or negative mode depending on the ionization of the molecule.

2.3.2 Synthetic Procedures and Compound Characterization

General procedures (GP1-GP10 in **Scheme 2.1**) of synthesizing PVHIs were detailed below. Each GP described the synthesis of one representative compound.

*GP1. Synthesis of Weinreb amides (**a** to **b**).* A solution of Boc-L-homophenylalanine (12.02 g, 43.03 mmol) in anh. DCM (200 mL) was cooled to 0 °C under an N$_2$ atmosphere. Et$_3$N (18.1 mL, 129.09 mmol, 3 eq.) was added slowly, followed by addition of *N,O*-dimethylhydroxylamine hydrochloride (6.3 g, 64.5 mmol, 1.5 eq.) and dropwise addition of T3P (50% (w/v) in MeCN, 41.1 mL, 64.55 mmol, 1.5 eq.). The resulting mixture was stirred at 0°C for 30 min to 1h until TLC analysis (EtOAc/hexane=1:1, v/v) showed the disappearance of starting material. The reaction

mixture was diluted with DCM and washed with H_2O. The organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product. Purification of the crude product by silica gel column chromatography using a gradient of 5% – 50% of EtOAc in hexane as eluent yielded the pure Weinreb amide *tert*-butyl (*S*)-(1-(methoxy(methyl)amino)-1-oxo-4-phenylbutan-2-yl)carbamate (***b***, 13.3 g, 41.31 mmol, 96% yield) as a colorless gum.

*GP2. LAH reduction of Weinreb amides (**b** to **c**).* To a solution of *tert*-butyl (*S*)-(1-(methoxy(methyl)amino)-1-oxo-4-phenylbutan-2-yl)carbamate (***b***, 6.7 g, 20.78 mmol) in anh. THF (120 mL) at -10°C under a N_2 atmosphere was added dropwise LAH (2.0 M in THF, 12.5 mL, 24.93 mmol, 1.2 eq.). The resulting mixture was stirred at -10°C for 30 min. Upon completion of reaction as shown by TLC analysis (EtOAc/Hexane=1:1, v/v), the reaction was quenched at the same temperature by adding dropwise 1N HCl, followed by removal of THF by rotary evaporation. Diethyl ether (500 mL) was added to the solid residue, and the solution was washed with aq. $NaHCO_3$ (1 X 50 mL) and brine (1 X 50 mL). The organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product. Purification of the crude material by silica gel column chromatography using a gradient of 10% – 60% of EtOAc in hexane as eluent yielded the pure aldehyde *tert*-butyl (*S*)-(1-oxo-4-phenylbutan-2-yl)carbamate (***c***, 4.89 g, 18.57 mmol, 89% yield) as a white solid.

*GP3. Preparation of chloromethyl-heterocycles (**d** to **e**).* To a suspension of methyl pyrimidine-2-carboxylate (***d***, 1.156 g, 8.37 mmol) in anh. EtOH (20 mL) at 0°C under N_2 atmosphere, was added portion-wise $NaBH_4$ (0.443 g, 11.72 mmol, 5 eq.). The reaction mixture was stirred at 25°C for 2h. Upon the completion of the reaction as

56

shown by TLC analysis (EtOAc/Hexane=1:1, v/v), the reaction solvents were removed by rotary evaporation. To the resultant colorless gummy residue was added ice cold H_2O (20 mL) and extracted with DCM (5 X 50 mL). The organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product pyrimidin-2-yl-methanol (0.900 g, 8.17 mmol). To this pyrimidin-2-yl-methanol in $CHCl_3$ (20 mL) at 0°C under N_2 atmosphere was added dropwise $POCl_3$ (1.95 mL, 3.21 g, 2.5 eq.). The reaction mixture was stirred at 25°C for 1h, followed by refluxing for an additional 3h under gentle heating until TLC analysis (EtOAc/Hexane=3:1, v/v) showed the completion of the reaction. The reaction was quenched by a careful addition of aq. $NaHCO_3$ and further addition of solid $NaHCO_3$ to afford a basic pH. The aqueous layer was extracted with $CHCl_3$ (3 X 50 mL), and the organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the pure product 2-(chloromethyl)pyrimidine (*e*, 0.948 g, 7.43 mmol, 63% yield) as a light yellow semi-solid, which was used further without any purification.

*GP4. Preparation of heterocyclic phosphonium ylides (**e** to **f**, Wittig reagents).* A mixture of 2-(chloromethyl)pyrimidine (*e*, 0.92 g, 7.22 mmol) and triphenyl phosphine (2.1 g, 7.94 mmol, 1.1 eq.) in anh. benzene (25 mL) was refluxed under N_2 atmosphere for 24h until TLC analysis (MeOH/DCM=1:19, v/v) showed the completion of the reaction. The reaction mixture was concentrated via rotary evaporation, and the gummy residue was triturated with diethyl ether (3 X 10 mL). The solid obtained was purified by silica gel column chromatography using a gradient of 1% – 10% of MeOH in DCM as eluent to afford the pure product triphenyl(pyrimidin-2-ylmethyl)phosphonium chloride (*g*, 0.797 g, 2.039 mmol, 28% yield).

*GP5. Preparation of heterocyclic phosphonates (**e** to **g**, HWE reagents).* 2-(chloromethyl)pyridine hydrochloride (**e**, 16.5 g, 100.6 mmol) in DCM (100 mL) was treated with aq. NaHCO$_3$ (20 mL), and the DCM layer was dried over anh. Na$_2$SO$_4$. The filtrate was concentrated by rotary evaporation. The alkyl halide thus obtained along with triethyl phosphite (35 mL, 201.2 mmol, 2.0 eq.) were heated at 150°C under N$_2$ atmosphere for 5h until TLC analysis (MeOH/DCM=1:19, v/v) showed the completion of the reaction. The reaction mixture was purified by silica gel column chromatography using a gradient of 10% – 100% of EtOAc in hexane and later 1% – 10% of MeOH in DCM as eluent to yield the pure product 2-pyridyl methyl phosphonate (**g**, 18.26 g, 79.66 mmol, 79% yield).

*GP6. Wittig reaction (**c** + **f** to **h**).* To a suspension of the Wittig reagent triphenyl(pyrimidin-2-ylmethyl)phosphonium chloride (**f**, 0.719 g, 1.839 mmol) in anh. THF (40 mL) at -70°C under an N$_2$ atmosphere, was added dropwise LHMDS (1.0 M in THF, 2.03 mL, 2.024 mmol, 1.1 eq.), which was stirred at the same temperature for 15 min. To this mixture a solution of Boc-Phe-H (**c**, 0.321 g, 1.287 mmol, 0.7 eq.) in THF (10 mL) was added, and stirred over 2 h until the temperature reached -40°C. Upon completion of reaction as revealed by TLC analysis (EtOAc/hexane=1:1, v/v), the reaction was quenched by addition of 0.1 mL of glacial acetic acid, followed by aq. NaHCO$_3$. Most of the THF was removed carefully using a rotary evaporator, and the residue was extracted with EtOAc (2X). The organic layer was dried over anh. Na$_2$SO$_4$ and filtered. The filtrate was concentrated *in vacuo* to afford the crude material, which was purified by silica gel column chromatography using a gradient of 5% – 30% of EtOAc in hexane as eluent, yielding the pure olefin *tert*-butyl (*S,E*)-(1-phenyl-4-

58

(pyrimidin-2-yl)but-3-en-2-yl)carbamate (**h,** E-isomer, 0.060 g, 14%). The other *Z*-isomer (0.014 g) was isolated as a side product and the ratio of *E* to *Z* isomers was typically 4:1.

GP7. Horner–Wadsworth–Emmons reaction (c + g to h). To a solution of the 2-pyridyl methyl phosphonate ester (**g**, 1.30 g, 5.65 mmol) in anh. THF (25 mL) at -70°C under N_2 atmosphere, was added dropwise LHMDS (1.0 M in THF, 6.22 mL, 6.22 mmol, 1.1 eq.). The reaction was stirred at the same temperature for 15 min, followed by dropwise addition of a solution of Boc-hPhe-H (**c**, 1.34 g in 10 mL THF, 5.09 mmol, 0.9 eq.). The reaction was stirred until it reached the temperature -20°C over 2 h. Upon completion of reaction revealed by TLC analysis (EtOAc/hexane=1:1, v/v), to the reaction mixture at 0°C was added glacial acetic acid (0.5 mL), followed by addition of 20 mL of saturated $NaHCO_3$. The aqueous layer was extracted with EtOAc (3 X 100 mL). Extracts were washed with brine (1 X 50 mL), and the organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product, which was purified by silica gel column chromatography using a gradient of 10% – 50% of EtOAc in hexane as eluent to yield the pure product *tert*-butyl (*S,E*)-(5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)carbamate (**h**, 0.344 g, 1.016 mmol, 20% yield).

GP8. Amide coupling with P_3-P_2 fragment (h to j). To a solution of *tert*-butyl (*S,E*)-(5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)carbamate (**h**, 0.143 g, 0.423 mmol) in anh DCM (5 mL) at 0°C, was added dropwise TFA (1.5 mL in 1 mL DCM) with stirring at the same temperature for 1h. Upon the completion of the reaction as revealed by TLC analysis (EtOAc/hexane=1:1, v/v), the reaction solvent was removed by a rotary evaporator. The resulting oil was co-evaporated on a rotary evaporator with $CHCl_3$ (3X)

and ether (3X). The solid product was dried on high vacuum to yield the TFA salt (*S*,*E*)-5-phenyl-1-(pyridin-2-yl)pent-1-en-3-aminium trifluoroacetate (0.149 g, 0.423 mmol), which was used in subsequent synthetic steps without further purification. To a solution of above TFA salt in anh. DCM (5 mL) at -10°C under N_2 atmosphere, was added dropwise DIPEA (0.6 mL, 0.3.44 mmol, 8 eq.), followed by addition of Cbz-Phe-OH (0.13 g, 0.43 mmol, 1 eq.) and T3P (50% in EtOAc, 0.41 mL, 1.5 eq.). The reaction was stirred at 0°C for an additional 1 h. Upon the completion of reaction revealed by TLC analysis (EtOAc/hexane=1:1, v/v), the reaction mixture was diluted with DCM (50 mL), and then washed with H_2O (3X) and brine (3X). The organic layer was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product, which was purified by silica gel column chromatography using a gradient of 10% – 50% of EtOAc in hexane as eluent to yield the pure product benzyl ((*S*)-1-oxo-3-phenyl-1-((((*S*,*E*)-5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)carbamate (*i*, 0.113 g, 0.217 mmol, 51%).

*GP9. N-methylation using methyl iodide (**i** to **j**).* To a suspension of benzyl ((*S*)-1-oxo-3-phenyl-1-((((*S*,*E*)-1-phenyl-4-(pyridin-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (*i*, 0.049 g, 0.098 mmol) in anh. MeCN (5 mL) and under N_2 atmosphere, was added MeI (0.03 mL, 0.490 mmol, 5 eq.), and the reaction mixture was heated under reflux for 9h. Upon the completion of the reaction revealed by TLC analysis (EtOAc/hexane=1:1, v/v), the solvents were removed by rotary evaporation. The resulting gummy residue was dissolved in $CHCl_3$ (1 mL), and precipitated with ether (5 mL). The solvents were decanted, and this procedure was repeated twice. The solid obtained was dried under high vacuum to give pure product 2-((*S*,*E*)-3-((*S*)-2-

(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-4-phenylbut-1-en-1-yl)-1-methylpyridin-1-ium iodide as a yellow solid (*j*, 0.039 g, 61%).

*GP10. Preparation of peptide acrylamide (**i to k**).* A solution of ethyl (S,E)-4-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-5-phenylpent-2-enoate (*i*, 0.346 g, 0.69 mmol) in THF (6mL) at 0°C was treated with LiOH (1N in H_2O, 0.83mL, 0.83mmol, 1.2 eq) and stirred overnight. The reaction was concentrated by rotary evaporation and the aqueous layer was added water and acidified to pH 1-2, and extracted with EtOAc (3X). The combined organic layers were dried and concentrated to yield the crude acrylic acid. To a solution of this acrylic acid (0.124 g, 0.262 mmol) in THF (6mL) at -15°C was added Et_3N (0.11 mL, 0.787 mmol, 3 eq.) and dropwise addition of $ClCO_2Et$ (0.035 mL, 0.367 mmol), which resulted in a white precipitate. The reaction mixture was stirred at the same temperature for an additional 30 min, then aq. 1M NH_4Cl (0.4 mL) was added dropwise with continuous stirring over 3h until a temperature of 25°C was attained. Upon completion of reaction as revealed by TLC analysis (EtOAc/hexane=1:1, v/v), most of the reaction solvent was removed using a rotary evaporator, and the solid residue was extracted with EtOAc. The organic layer was washed with aq. $NaHCO_3$ (2X), H_2O (1X) and brine (1X), and was dried over anh. Na_2SO_4 and filtered. The filtrate was concentrated *in vacuo* to afford the crude product. Purification of the crude product by silica gel chromatography using a gradient of 20% – 100% of EtOAc in hexane as eluent, yielded the pure product benzyl ((S)-1-(((S,E)-5-amino-5-oxo-1-phenylpent-3-en-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (0.027 g, 0.057 mmol, 22% yield).

*4-Methyl-N-((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(phenylsulfonyl)but-3-en-2-yl)amino)propan-2-yl)piperazine-1-carboxamide (**1**, Cbz-Phe-Phe-VSPh).* White solid, 0.115 g, 0.202 mmol, 56% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.79 (d, 2H, J = 6.8 Hz), 2.85 – 3.12 (m, 2H), 4.26 (q, 1H, J = 7.3 Hz), 4.79 – 4.95 (m, 1H), 5.04 (s, 2H), 5.13 (s, 1H), 5.75 (s, 1H), 5.96 (dd, 1H, J_1 = 1.8 Hz, J_2 = 15.1 Hz), 6.78 (dd, 1H, J_1 = 4.8 Hz, J_2 = 15.1 Hz), 6.95 – 7.03 (m, 2H), 7.05 – 7.11 (m, 2H), 7.12 – 7.23 (m, 6H), 7.27 – 7.39 (m, 5H), 7.47 – 7.56 (m, 2H), 7.57 – 7.67 (m, 1H), 7.72 – 7.84 (m, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.4, 40.3, 50.4, 56.7, 67.4, 127.3, 127.4, 127.8, 128.2, 128.5, 128.7 (2C), 128.8, 129.0, 129.3, 129.4, 131.1, 133.6, 135.5, 136.1, 136.2, 140.2, 144.6, 156.0, 170.5; LC-MS t$_R$ 7.27 min, m/z 569.31 [M+H]$^+$, (C$_{33}$H$_{32}$N$_2$O$_5$S$^+$ Calcd 569.21).

*Benzyl ((S)-1-(((S,E)-5-amino-5-oxo-1-phenylpent-3-en-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**2**, Cbz-Phe-Phe-vinyl-CONH$_2$).* White solid, 0.027 g, 0.057 mmol, 22% yield. ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 2.64 – 2.77 (m, 1H), 2.84 (d, 2H, J = 7.2 Hz), 2.96 (dd, 1H, J_1 = 3.9 Hz, J_2 = 13.7 Hz), 4.19 – 4.31 (m, 1H), 4.62 (pentet, 1H, J = 6.8 Hz), 4.95 (s, 1H), 5.85 (d, 1H, J = 15.5 Hz), 6.56 (dd, 1H, J_1 = 5.9 Hz, J_2 = 15.5 Hz), 6.93 (s, 1H), 7.12 – 7.46 (m, 17H), 8.25 (d, 1H, J = 8.2 Hz); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 28.7, 37.6, 50.9, 56.07, 65.1, 124.1, 126.2, 126.4, 127.4, 127.6, 127.9, 128.1, 128.2, 129.2, 129.5, 137.0, 137.8, 138.0, 142.0, 155.6, 166.2, 170.7; LC-MS t$_R$ 4.71 min, m/z 472.46 [M+H]$^+$, (C$_{28}$H$_{29}$N$_3$O$_4$$^+$ Calcd 472.22).

*Benzyl ((S)-1-(((S,E)-6-amino-6-oxo-1-phenylhex-4-en-3-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**3**, Cbz-Phe-hPhe-vinyl-CONH$_2$).* White solid, 0.013 g, 0.027 mmol, 15% yield. ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 1.64 – 1.92 (m, 2H), 2.54 – 2.72 (m, 2H), 2.81 (dd, 1H, J_1 = 10.6 Hz, J_2 = 13.6 Hz), 3.03 (dd, 1H, J_1 = 4.0 Hz, J_2 =

13.6 Hz), 4.25 – 4.33 (m, 1H), 4.35 – 4.43 (m, 1H), 4.85 – 5.04 (m, 2H), 5.90 (d, 1H, J = 15.5 Hz), 6.54 (dd, 1H, J_1 = 5.7 Hz, J_2 = 15.5 Hz), 6.94 (s, 1H), 7.15 – 7.35 (m, 15H), 7.41 (s, 1H), 7.49 (d, 1H, J = 8.5 Hz), 8.22 (d, 1H, J = 8.2 Hz); LC-MS t_R 4.89 min, m/z 486.24 [M+H]$^+$, ($C_{29}H_{31}N_3O_4^+$ Calcd 486.24).

*N-((S)-1-(((S,E)-6-amino-6-oxo-1-phenylhex-4-en-3-yl)amino)-1-oxo-3-phenylpropan-2-yl)-4-methylpiperazine-1-carboxamide (**4**, NMePip-Phe-hPhe-vinyl-CONH$_2$).* White solid, 0.022 g, 0.048 mmol, 18% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.66 – 1.91 (m, 2H), 2.23 (s, 3H), 2.25 – 2.34 (m, 4H), 2.50 – 2.64 (m, 2H), 2.99 – 3.15 (m, 2H), 3.25 – 3.41 (m, 4H), 4.44 – 4.57 (m, 1H), 4.62 (q, 1H, J = 7.4 Hz), 5.37 (d, 1H, J = 7.6 Hz), 5.60 (dd, 1H, J_1 = 1.3 Hz, J_2 = 15.3 Hz), 5.75 (s, 1H), 6.01 (s, 1H), 6.62 (dd, 1H, J_1 = 5.5 Hz, J_2 = 15.3 Hz), 6.93 (d, 1H, J = 8.2 Hz), 7.04 – 7.31 (m, 11H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 32.0, 36.1, 38.5, 43.9, 46.1, 49.9, 54.6, 56.2, 122.9, 126.2, 126.9, 128.5, 128.6, 128.8, 129.7, 137.3, 141.1, 144.4, 157.3, 167.5, 172.0; LC-MS t_R 2.54 min, m/z 478.36 [M+H]$^+$, ($C_{27}H_{35}N_5O_3^+$ Calcd 478.28).

*Benzyl ((S)-1-(((S,E)-1,4-diphenylbut-3-en-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**5**, Cbz-Phe-Phe-vinyl-Ph).* Off-white solid, 0.054 g, 0.107 mmol, 35% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.87 (dt, J = 2.8, 6.3 Hz, 2H), 2.96 – 3.16 (m, 2H), 4.35 (q, J = 7.6 Hz, 1H), 4.80 – 4.96 (m, 1H), 5.29 (d, J = 9.6 Hz, 1H), 5.69 (d, J = 9.9 Hz, 1H), 5.94 (ddt, J = 3.0, 6.3, 15.9 Hz, 1H), 6.27 (d, J = 15.9 Hz, 1H), 7.03 – 7.14 (m, 2H), 7.14 – 7.28 (m, 10H), 7.28 – 7.40 (m, 9H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 29.7, 38.6, 41.3, 51.8, 67.1, 126.4, 126.6, 127.1, 127.6, 128.0, 128.2, 128.3, 128.4, 128.5, 128.7, 129.3, 129.4, 130.9, 136.5, 136.8, 169.8; LC-MS t_R 6.20 min, m/z 505.31 [M+H]$^+$, ($C_{33}H_{32}N_2O_3^+$ Calcd 505.25).

Benzyl ((S)-1-(((S,E)-4-(4-nitrophenyl)-1-phenylbut-3-en-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (6, Cbz-Phe-Phe-vinyl-(4-NO₂)Ph). White fluffy solid, 0.530 g, 0.964 mmol, 55% yield. ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 2.76 (dd, 1H, $J_1 = 9.7$ Hz, $J_2 = 13.4$ Hz), 2.85 – 2.97 (m, 3H), 4.18 – 4.35 (m, 1H), 4.68 (pentet, 1H, $J = 6.7$ Hz), 4.99 (s, 2H), 6.41 (d, 1H, $J = 16.1$ Hz), 6.50 (dd, 1H, $J_1 = 5.4$ Hz, $J_2 = 16.1$ Hz), 7.13 – 7.37 (m, 15H), 7.42 (d, 1H, $J = 8.5$ Hz), 7.58 (d, 2H, $J = 8.5$ Hz), 8.14 – 8.25 (m, 3H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 37.7, 39.0, 51.9, 56.2, 65.2, 123.9, 126.1, 126.2, 127.0, 127.4, 127.6, 127.9, 128.0, 128.1, 128.2, 129.2, 129.3, 135.6, 137.0, 137.8, 138.0, 143.3, 146.2, 155.6, 170.6; LC-MS t$_R$ 6.12 min, *m/z* 550.28 [M+H]$^+$, (C$_{33}$H$_{31}$N$_3$O$_5$$^+$ Calcd 550.23).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(pyrimidin-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (7, Cbz-Phe-Phe-vinyl-2Pyrmd). White solid, 0.026 g, 0.0513 mmol, 33% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.73 – 2.96 (m, 2H), 3.01 (d, 2H, $J = 4.8$ Hz), 4.23 – 4.50 (m, 1H), 4.91 – 5.02 (m, 1H), 5.05 (s, 2H), 5.30 (s, 1H), 6.05 (s, 1H), 6.48 (d, 1H, $J = 15.7$ Hz), 7.00 (dd, 1H, $J_1 = 5.6$ Hz, $J_2 = 15.7$ Hz), 7.05 – 7.37 (m, 16H), 8.63 (d, 2H, $J = 4.0$ Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.6, 41.0, 51.6, 56.5, 67.3, 119.0, 126.9, 127.2, 128.2, 128.3, 128.6, 128.7, 128.9, 129.4, 129.5, 130.3, 136.3, 136.5, 136.7, 139.2, 156.1, 157.1, 164.2, 170.3; LC-MS t$_R$ 5.14 min, *m/z* 507.26 [M+H]$^+$, (C$_{31}$H$_{30}$N$_4$O$_3$$^+$ Calcd 507.24).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-5-phenyl-1-(pyrimidin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)carbamate (8, Cbz-Phe-hPhe-vinyl-2Pyrmd). Off-white solid, 0.280 g, 0.538 mmol, 29% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.76 – 2.03 (m, 2H), 2.63 (t, $J = 7.9$ Hz, 2H), 3.10 (t, $J = 8.4$ Hz, 2H), 4.47 (s, 1H), 4.74 (h, $J = 7.3$ Hz, 1H),

5.10 (d, J = 7.9 Hz, 2H), 5.53 (s, 1H), 6.27 (s, 1H), 6.56 – 6.72 (m, 1H), 6.96 – 7.05 (m, 1H), 7.06 – 7.14 (m, 3H), 7.19 (d, J = 7.3 Hz, 3H), 7.21 – 7.34 (m, 10H), 8.66 (dd, J = 4.9, 15.8 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.0, 36.2, 38.6, 50.5, 56.6, 67.1, 118.9, 126.0, 127.0, 128.0, 128.1, 128.3, 128.4, 128.5, 128.7, 128.8, 129.3, 129.4, 130.1, 136.4, 139.6, 141.2, 156.9, 164.1, 170.3; LC-MS t_R = 5.56 min, m/z 521.24 [M+H]$^+$, (C$_{32}$H$_{32}$N$_4$O$_3^+$ Calcd 521.26).

4-Methyl-N-((S)-1-oxo-3-phenyl-1-(((S,E)-5-phenyl-1-(pyrimidin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)piperazine-1-carboxamide (9, NMePip-Phe-hPhe-vinyl-2Pyrmd). Off-white gum, 0.054 g, 0.105 mmol, 54% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.82 – 1.99 (m, 2H) ,2.29 (s, 3H), 2.34 – 2.46 (m, 4H), 2.61 (t, 1H, J = 7.5 Hz), 3.09 (d, 2H, J = 7.5 Hz), 3.39 (s, 4H), 4.58 – 4.74 (m, 2H), 5.47 (d, 1H, J = 6.3 Hz), 6.61 (d, 1H, J = 15.7 Hz), 6.87 (d, 1H, J = 8.2 Hz), 6.98 (dd, 1H, J_1 = 6.3 Hz, J_2 = 15.7 Hz), 7.06 – 7.31 (m, 11H), 8.66 (d, 2H, J = 4.8 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.1, 36.5, 38.8, 43.5, 45.6, 50.6, 54.3, 56.1, 118.9, 126.0, 126.9, 128.4, 128.5 (2C), 128.6, 129.6, 130.1, 137.2, 140.1, 141.3, 157.0, 164.3, 171.8; LC-MS t_R 2.96 min, m/z 513.17 [M+H]$^+$, (C$_{30}$H$_{36}$N$_6$O$_2^+$ Calcd 513.30).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-4-(pyrimidin-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (10, Cbz-Phe-Ala-vinyl-2Pyrmd). White fluffy solid, 0.017 g, 0.039 mmol, 14% yield. ^{1}H NMR (400 MHz, CDCl$_3$ + MeOD) δ 1.32 (d, 3H, J = 6.6 Hz), 2.96 (dd, 1H, J_1 = 8.0 Hz, J_2 = 13.6 Hz), 3.10 (dd, 1H, J_1 = 6.4 Hz, J_2 = 13.6 Hz), 4.36 – 4.44 (m, 4H), 4.71 (pentet, 1H, J = 6.4 Hz), 4.97 – 5.12 (m, 2H), 6.52 (d, 1H, J = 15.7 Hz), 6.98 (dd, 1H, J_1 = 5.8 Hz, J_2 = 15.7 Hz), 7.13 – 7.36 (m, 11H), 8.70 (d, 2H, J = 4.9 Hz); ^{13}C NMR (100 MHz, CDCl$_3$ + MeOD) δ 19.6, 38.7, 46.1, 56.1, 66.7, 119.0,

126.7, 127.6, 127.9, 128.1, 128.3 (2C), 129.2, 136.3, 141.6, 156.9 (2C), 157.0, 163.9, 171.0; LC-MS t$_R$ 4.68 min, m/z 430.91 [M+H]$^+$, (C$_{25}$H$_{26}$N$_4$O$_3^+$ Calcd 431.21).

*Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(pyridin-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (**11**, Cbz-Phe-Phe-vinyl-2Pyr).* White solid, 0.554 g, 1.096 mmol, 81% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.87 (dq, 1H, J_1 = 6.8 Hz, J_2 = 13.7 Hz), 3.01 (d, 2H, J = 7.1 Hz), 4.35 (q, 1H, J = 7.0 Hz), 4.91 (pentet, 1H, J = 6.8 Hz), 5.06 (s, 2H), 5.24 (s, 1H), 5.88 (d, 1H, J = 8.4 Hz), 6.32 (d, 1H, J = 15.7 Hz), 6.58 (dd, 1H, J_1 = 6.1 Hz, J_2 = 15.7 Hz), 7.08 – 7.36 (m, 17H), 7.59 (dt, 1H, J_1 = 1.6 Hz, J_2 = 7.7 Hz), 8.53 (d, 1H, J = 4.3 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.4, 41.1, 51.7, 56.4, 67.1, 122.1, 122.2, 126.7, 127.0, 128.0, 128.2, 128.4, 128.5 (2C), 128.8, 129.4 (2C), 130.5, 133.0, 136.1, 136.4, 136.8, 149.5, 154.8, 155.9, 170.0; LC-MS t$_R$ 4.39 min, m/z 506.24 [M+H]$^+$, (C$_{32}$H$_{31}$N$_3$O$_3^+$ Calcd 506.24).

*2-((S,E)-3-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-4-phenylbut-1-en-1-yl)-1-methylpyridin-1-ium iodide (**12**, Cbz-Phe-Phe-vinyl-2PyrNMe).* Yellow solid, 0.039 g, 0.060 mmol, 61% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 3.01 – 3.15 (m, 3H), 3.19 (dd, 1H, J_1 = 7.7 Hz, J_2 = 13.6 Hz), 4.19 (s, 3H), 4.51 (q, 1H, J = 7.0 Hz), 4.90 – 5.02 (m, 2H), 5.06 (s, 1H), 5.81 (s, 1H), 6.69 (d, 1H, J = 15.7 Hz), 6.86 (dd, 1H, J_1 = 4.6 Hz, J_2 = 15.7 Hz), 7.08 – 7.28 (m, 14H), 7.65 – 7.79 (m, 2H), 7.89 – 8.01 (m, 1H), 8.21 – 8.31 (m, 1H), 8.97 – 9.09 (m, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.1, 39.8, 47.5, 52.6, 57.1, 66.8, 119.9, 126.1, 126.3, 126.9, 127.1, 127.5, 127.6, 127.9, 128.5, 128.6, 128.8, 129.6, 129.7, 136.7, 136.9, 144.9, 146.0, 147.6, 152.8, 156.2, 171.5; LC-MS t$_R$ 3.25 min, m/z 520.32 [M+H]$^+$, (C$_{33}$H$_{34}$N$_3$O$_3^+$ Calcd 520.26).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)carbamate (13, Cbz-Phe-hPhe-vinyl-2Pyr). White solid, 0.113 g, 0.217 mmol, 51% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.73 – 1.95 (m, 2H), 2.59 (t, 2H, *J* = 7.9 Hz), 3.04 (d, 2H, *J* = 7.0 Hz), 4.38 – 4.52 (m, 1H), 4.65 (pentet, 1H, *J* = 7.1 Hz), 5.03 (s, 2H), 5.59 (d, 1H, *J* = 6.4 Hz), 6.34 (d, 1H, *J* = 5.1 Hz), 6.43 (d, 1H, *J* = 15.7 Hz), 6.53 (dd, 1H, *J$_1$* = 6.2 Hz, *J$_2$* = 15.7 Hz), 7.02 – 7.32 (m, 17H), 7.57 (dt, 1H, *J$_1$* = 1.4 Hz, *J$_2$* = 7.7 Hz), 8.51 (d, 1H, *J* = 4.4 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.1, 36.4, 38.7, 50.8, 56.6, 67.1, 122.1, 122.3, 126.0, 127.0, 128.0, 128.2, 128.4, 128.5 (2C), 128.7, 129.5, 130.5, 133.8, 136.2, 136.5 (2C), 141.4, 149.5, 155.0, 156.1, 170.4; LC-MS t$_R$ 4.68 min, *m/z* 518.74, 520.41 [M+H]$^+$, (C$_{33}$H$_{33}$N$_3$O$_3$$^+$ Calcd 520.26).

4-Methyl-N-((S)-1-oxo-3-phenyl-1-(((S,E)-5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)piperazine-1-carboxamide (14, NMePip-Phe-hPhe-vinyl-2Pyr). Off-white solid, 0.090 g, 0.176 mmol, 44% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.77 – 1.94 (m, 2H), 2.21 (s, 3H), 2.22 – 2.27 (m, 4H), 2.59 (t, 2H, *J* = 8.0 Hz), 3.08 (d, 2H, *J* = 7.0 Hz), 3.28 (m, 4H), 4.61 (pentet, 1H, *J* = 7.3 Hz), 4.7 (q, 1H, *J* = 7.3 Hz), 5.39 (d, 1H, *J* = 7.7 Hz), 6.46 (d, 1H, *J* = 15.8 Hz), 6.56 (dd, 1H, *J$_1$* = 6.2 Hz, *J$_2$* = 15.8 Hz), 6.98 (d, 1H, *J* = 8.4 Hz), 7.06 – 7.24 (m, 12H), 7.59 (dt, 1H, *J$_1$* = 1.8 Hz, *J$_2$* = 7.7 Hz), 8.53 (d, 1H, *J* = 4.3 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.1, 36.6, 39.0, 43.8, 46.1, 50.7, 54.6, 55.9, 122.0, 122.2, 125.9, 126.8, 128.4 (2C), 128.5, 129.6, 130.3, 134.2, 136.4, 137.2, 141.4, 149.5, 155.1, 157.0, 171.6; LC-MS t$_R$ 2.60 min, *m/z* 512.28 [M+H]$^+$, (C$_{31}$H$_{37}$N$_5$O$_2$$^+$ Calcd 512.30).

2-((S,E)-3-((S)-2-((((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-5-phenylpent-1-en-1-yl)-1-methylpyridin-1-ium (15, Cbz-Phe-hPhe-vinyl-2PyrNMe).

Yellow solid, 0.029 g, 0.044 mmol, 89% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.98 – 2.10 (m, 1H), 2.11 – 2.25 (m, 1H), 2.56 – 2.85 (m, 2H), 3.04 – 3.35(m, 2H), 4.22 (s, 3H), 4.62 (d, 1H, J = 5.5 Hz), 4.79 (s, 1H), 4.92 – 5.09 (m, 2H), 5.88 (s, 1H), 6.70 (d, 1H, J = 15.8 Hz), 6.77 (dd, 1H, J_1 = 3.6 Hz, J_2 = 15.8 Hz), 7.09 – 7.34 (m, 14H), 7.66 – 7.79 (m, 2H), 7.90 (d, 1H, J = 6.0 Hz), 8.17 – 8.30 (m, 1H), 8.98 (d, 1H, J = 4.5 Hz); LC-MS t$_R$ 3.48 min, m/z 534.25 [M+H]$^+$, (C$_{34}$H$_{36}$N$_3$O$_3^+$ Calcd 534.28).

Benzyl ((S)-1-(((S,E)-1-(4-methoxypyridin-2-yl)-5-phenylpent-1-en-3-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (16, Cbz-Phe-hPhe-vinyl-2-(4-OMe)-Pyr). Pale-yellow solid, 0.072 g, 0.131 mmol, 33% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.73 – 2.01 (m, 2H), 2.59 (dt, J = 7.9, 37.9 Hz, 2H), 3.11 (dd, J = 7.0, 11.7 Hz, 2H), 3.87 (d, J = 9.2 Hz, 3H), 4.50 (dd, J = 7.6, 23.1 Hz, 1H), 4.68 (t, J = 7.1 Hz, 1H), 5.08 (d, J = 4.6 Hz, 2H), 5.74 (dd, J = 8.1, 65.8 Hz, 1H), 6.39 – 6.58 (m, 1H), 6.67 – 6.81 (m, 2H), 7.11 – 7.15 (m, 2H), 7.16 – 7.26 (m, 8H), 7.27 – 7.34 (m, 7H), 8.37 (dd, J = 5.8, 23.3 Hz, 1H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.0, 36.3, 38.7, 50.6, 55.2, 56.5, 67.1, 108.3, 108.5, 126.0, 126.6, 127.0, 127.9, 128.0, 128.1, 128.1, 128.3, 128.4, 128.4, 128.4, 128.5, 128.5, 128.7, 129.4, 129.4, 134.7, 135.1, 136.5, 141.3, 150.0, 156.1, 166.6, 170.3; LC-MS t$_R$ = 3.57 min, m/z 550.16 [M+H]$^+$, (C$_{34}$H$_{35}$N$_3$O$_4^+$ Calcd 549.26).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-5-phenyl-1-(4-(trifluoromethyl)pyridin-2-yl)pent-1-en-3-yl)amino)propan-2-yl)carbamate (17, Cbz-Phe-hPhe-vinyl-2-(4-CF$_3$)-Pyr). White solid, 0.260 g, 0.442 mmol, 79% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.76 – 1.95 (m, 2H), 2.59 (t, 2H, J = 7.8 Hz), 3.06 (d, 2H, J = 7.1 Hz), 4.36 – 4.51 (m, 1H), 4.65 (pentet, 1H, J = 7.1 Hz), 5.05 (s, 2H), 5.49 (s, 1H), 6.13 (s, 1H), 6.35 (d, 1H, J = 15.7 Hz), 6.59 (dd, 1H, J_1 = 6.0 Hz, J_2 = 15.7 Hz), 7.06 – 7.11 (m, 2 H), 7.14 – 7.33 (m, 15 H),

68

8.67 (d, 1H, J = 4.9 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 32.1, 36.4, 38.7, 50.7, 56.8, 67.3, 117.5, 121.6, 124.3, 126.2, 127.0 (C-F), 127.2, 128.1, 128.3, 128.5, 128.6 (3C), 128.8, 129.2, 129.5, 136.2, 136.6, 138.7, 139.1, 141.2, 150.5, 156.2 (C-F), 156.4, 170.4; LC-MS t$_R$ 5.16 min, m/z 587.95 [M+H]$^+$, (C$_{34}$H$_{32}$F$_3$N$_3$O$_3^+$ Calcd 588.25).

Benzyl ((S)-4-methyl-1-oxo-1-(((S,E)-5-phenyl-1-(pyridin-2-yl)pent-1-en-3-yl)amino)pentan-2-yl)carbamate (18, Cbz-Leu-hPhe-vinyl-2Pyr). White solid, 0.038 g, 0.078 mmol, 26% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 0.92 (t, 6H, J = 6.4 Hz), 1.45 – 1.57 (m, 1H) , 1.59 -1.73 (m, 2H), 1.79 (s, 1H), 1.86 – 2.05 (m, 2H), 2.68 (t, 2H, J = 7.9 Hz), 4.03 – 4.27 (m, 1H), 4.71 (pentet, 1H, J = 6.9 Hz), 5.10 (s, 2H), 5.14 (s, 1H), 6.26 (d, 1H, J = 6.2 Hz), 6.58 (d, 1H, J = 15.8 Hz), 6.69 (dd, 1H, J_1 = 6.0 Hz, J_2 = 15.8 Hz), 7.10 – 7.34 (m, 12H), 7.60 (dt, 1H, J_1 = 1.7 Hz, J_2 = 7.7 Hz), 8.54 (d, 1H, J = 4.6 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 23.1, 24.9, 32.3, 36.7, 41.2, 50.9, 54.0, 67.3, 122.3, 122.4, 126.1, 128.2, 128.4, 128.6 (2C), 128.7, 130.6, 134.0, 136.3, 136.7, 141.6, 149.7, 155.1, 156.5, 171.6; LC-MS t$_R$ 4.58 min, m/z 484.46, 485.49, 486.38 [M+H]$^+$, (C$_{30}$H$_{35}$N$_3$O$_3^+$ Calcd 486.28).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-4-(pyridin-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (19, Cbz-Phe-Ala-vinyl-2Pyr). White solid, 0.590 g, 1.374 mmol, 75% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.58 (s, 3H), 2.96 – 3.23 (m, 2H), 4.36 (q, J = 7.4 Hz, 1H), 4.63 – 4.80 (m, 1H), 5.10 (d, J = 1.4 Hz, 2H), 5.33 (s, 1H), 5.67 (d, J = 8.3 Hz, 1H), 6.40 (dd, J = 1.3, 15.8 Hz, 1H), 6.51 (dd, J = 5.5, 15.8 Hz, 1H), 7.13 (ddd, J = 1.1, 4.8, 7.4 Hz, 1H), 7.16 – 7.21 (m, 3H), 7.21 – 7.37 (m, 8H), 7.62 (td, J = 1.8, 7.7 Hz, 1H), 8.45 – 8.61 (m, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 20.0, 38.8, 46.4, 56.2, 66.8, 121.8,

122.3, 126.7, 127.7, 128.0, 128.3, 128.7, 129.3, 135.6, 136.4, 137.0, 148.8, 155.0, 170.8; LC-MS t_R = 3.49 min, *m/z* 430.35 [M+H]$^+$, (C$_{26}$H$_{27}$F$_3$N$_3$O$_3$$^+$ Calcd 430.21).

*(S,E)-5-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-7-(pyridin-2-yl)hept-6-en-1-aminium chloride (**20**, Cbz-Phe-Lys-vinyl-2Pyr).* White solid, 0.027 g, 0.048 mmol, 51% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.44 – 1.59 (m, 2H), 1.65 – 1.84 (m, 4H), 2.89 – 3.02 (m, 3H), 3.09 – 3.19 (m, 1H), 4.43 (t, 1H, *J* = 7.5 Hz), 4.64 – 4.70 (m, 1H), 5.01 – 5.11 (m, 2H), 6.54 (d, 1H, *J* = 16.1 Hz), 6.92 (dd, 1H, *J$_1$* = 5.7 Hz, *J$_2$* = 16.1 Hz), 7.11 (t, 1H, *J* = 7.5 Hz), 7.21 (t, 2H, *J* = 7.5 Hz), 7.25 – 7.40 (m, 7H), 7.91 (t, 1H, *J* = 6.8 Hz), 8.10 (d, 1H, *J* = 8.0 Hz), 8.54 (t, 1H, *J* = 8.0 Hz), 8.69 (d, 1H, *J* = 5.2 Hz); ^{13}C NMR (100 MHz, MeOD) δ 23.8, 27.9, 34.0, 39.1, 40.6, 51.9, 58.2, 67.6, 121.6, 125.9, 126.5, 127.7, 128.5, 128.9, 129.5, 129.7, 130.5, 138.1, 138.4, 142.0, 146.3, 147.8, 151.2, 158.3, 174.1; LC-MS t_R 3.16 min, *m/z* 486.98 [M+H]$^+$, 508.94 [M+Na]$^+$ (C$_{29}$H$_{35}$N$_4$O$_3$$^+$ Calcd 487.27, C$_{29}$H$_{35}$N$_4$O$_3$Na$^+$ Calcd 509.25).

*Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(pyridin-4-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (**21**, Cbz-Phe-Phe-vinyl-4Pyr).* Off-white solid, 0.051 g, 0.109 mmol, 50% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.83 (d, *J* = 6.8 Hz, 2H), 3.02 (d, *J* = 7.8 Hz, 2H), 4.36 (s, 1H), 4.86 (s, 1H), 5.08 (s, 2H), 5.90 (s, 1H), 6.13 (s, 1H), 7.10 (dd, *J* = 6.5, 31.0 Hz, 6H), 7.18 – 7.27 (m, 6H), 7.31 (dd, *J* = 3.6, 6.4 Hz, 5H), 8.50 (d, *J* = 6.0 Hz, 2H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.4, 41.0, 51.5, 52.1, 67.1, 105.0, 120.9, 126.9, 127.1, 127.9, 128.3, 128.4, 128.5, 128.6, 128.8, 129.3, 129.4, 133.4, 136.4, 150.0, 170.1; LC-MS t_R = 3.49 min, *m/z* 506.29 [M+H]$^+$, (C$_{32}$H$_{31}$N$_3$O$_3$$^+$ Calcd 506.24).

*4-((S,E)-3-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-4-phenylbut-1-en-1-yl)-1-methylpyridin-1-ium iodide (**22**, Cbz-Phe-Phe-vinyl-4PyrNMe).*

Yellow solid, 0.020 g, 0.030 mmol, 87% yield. ^{1}H NMR (400 MHz, MeOD + CDCl$_3$) δ 2.75 – 3.09 (m, 4H), 4.28 – 4.43 (m,5H), 4.87 (q, 1H, J = 6.3 Hz), 5.08 (s, 2H), 6.28 (d, 1H, J = 15.9 Hz), 6.83 (dd, 1H, J_1 = 5.0 Hz, J_2 = 15.9 Hz), 7.16 – 7.35 (m, 15H), 7.82 (d, 2H, J = 6.5 Hz), 8.71 (d, 2H, J = 6.5 Hz); ^{13}C NMR (400 MHz, CDCl$_3$ + MeOD) δ 38.2, 40.1, 47.8, 52.2, 56.3, 66.8, 124.5, 126.8 (2C), 127.6, 128.0, 128.3, 128.4 (3C), 128.5, 129.1, 129.2, 136.1, 136.4, 143.9, 144.7, 153.1, 156.2, 171.3; LC-MS t$_R$ 3.36 min, m/z 521.33 [M+H]$^+$, (C$_{33}$H$_{34}$N$_3$O$_3^+$ Calcd 521.27).

*Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(pyrimidin-4-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (**23**, Cbz-Phe-Phe-vinyl-4Pyrmd).* Off-white solid, 0.800 g, 1.579 mmol, 64% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.79 – 2.93 (m, 2H), 3.01 (d, 2H, J = 6.9 Hz), 4.35 – 4.48 (m, 1H), 4.93 (pentet, 1H, J = 6.7 Hz), 5.04 (s, 2H), 5.49 (s, 1H), 6.16 (d, 1H, J = 15.6 Hz), 6.29 (s, 1H), 6.88 (dd, 1H, J_1 = 5.8 Hz, J_2 = 15.6 Hz), 6.97 (d, 1H, J = 4.6 Hz), 7.05 – 7.33 (m, 15H), 8.56 (d, 1H, J = 5.2 Hz), 9.06 (s, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.5, 40.9, 51.6, 56.6, 67.2, 118.8, 126.9, 127.1, 128.0, 128.2, 128.3, 128.6 (3C), 128.8, 129.4 (2C), 136.2, 136.5, 138.6, 156.1, 157.4, 158.8, 161.5, 170.4; LC-MS t$_R$ 5.27 min, m/z 507.35 [M+H]$^+$, (C$_{31}$H$_{30}$N$_4$O$_3^+$ Calcd 507.24).

*Benzyl ((S)-1-(((S,E)-4-(oxazol-2-yl)-1-phenylbut-3-en-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (**24**, Cbz-Phe-Phe-vinyl-2Oxz).* White solid, 0.020 g, 0.040 mmol, 29% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.72 – 2.88 (m, 2H), 2.93 – 3.11 (m, 2H), 4.35 (q, 1H, J = 7.0 Hz), 4.88 (pentet, 1H, J = 6.7 Hz), 5.05 (s, 2H), 5.30 (s, 1H), 5.99 (d, 1H, J = 5.8 Hz), 6.14 (d, 1H, J = 16.1 Hz), 6.48 (dd, 1H, J_1 = 5.9 Hz, J_2 = 16.1 Hz), 7.04 – 7.35 (m, 16H), 7.54 (s, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.6, 40.9, 51.5,

56.7, 67.3, 117.3, 127.1, 127.4, 128.2, 128.4, 128.7 (3C), 129.0, 129.4 (2C), 136.2, 136.4 (2C), 137.1, 138.3, 156.1, 160.8, 170.3; LC-MS t_R 5.41 min, m/z 496.25 [M+H]$^+$, ($C_{30}H_{29}N_3O_4^+$ Calcd 496.22).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S,E)-1-phenyl-4-(thiazol-2-yl)but-3-en-2-yl)amino)propan-2-yl)carbamate (25, Cbz-Phe-Phe-vinyl-2Thz). Light yellow solid, 0.206 g, 0.403 mmol, 54% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.75 – 2.89 (m, 2H), 2.99 (d, 2H, J = 7.1 Hz), 4.28 – 4.46 (m, 1H), 4.87 (d, 1H, J = 6.4 Hz), 4.96 – 5.11 (m, 2H), 5.44 (s, 1H), 6.19 (s, 1H), 6.39 (dd, 1H, J_1 = 5.1 Hz, J_2 = 16.0 Hz), 6.45 (d, 1H, J = 16.0 Hz), 7.01 – 7.39 (m, 16H), 7.71 (d, 1H, J = 3.2 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.7, 41.0, 51.5, 56.6, 67.2, 118.5, 123.9, 126.9, 127.2, 128.1, 128.3, 128.6 (2C), 128.8, 129.4 (2C), 135.0, 136.2, 136.5 (2C), 143.4, 156.0, 166.0, 170.4; LC-MS t_R 5.64 min, m/z 512.18 [M+H]$^+$, ($C_{30}H_{29}N_3O_3^+$ Calcd 512.20).

2-((S,E)-3-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-4-phenylbut-1-en-1-yl)-3-methylthiazol-3-ium iodide (26, Cbz-Phe-Phe-vinyl-2ThzNMe). Yellow solid, 0.003 g, 0.005 mmol, 15% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 2.51 (s, 1H), 2.81 – 3.96 (m, 1H), 2.98 – 3.18 (m, 3H), 3.71 (s, 3H), 3.93 (s, 2H), 4.36 – 4.49 (m, 1H), 4.99 (s, 2H), 5.85 – 6.11 (m, 1H), 6.81 (d, 1H, J = 15.6 Hz), 6.91 (dd, 1H, J_1 = 3.4 Hz, J_2 = 15.6 Hz), 7.06 – 7.33 (m, 15H), 7.76 (s, 1H), 7.96 (m, 1H); ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.1, 39.8, 52.5, 54.7, 57.3, 66.7, 115.1, 121.6, 126.9, 127.1, 127.6, 128.0 (2C), 128.5, 128.7, 128.8, 129.5 (2C), 136.9, 137.0, 138.8, 150.6, 156.4, 168.7, 171.9; LC-MS t_R 3.31 min, m/z 526.18 [M+H]$^+$, ($C_{31}H_{32}N_3O_3S^+$ Calcd 526.22).

Benzyl ((S)-1-oxo-3-phenyl-1-(((S)-1-phenyl-4-(pyrimidin-2-yl)butan-2-yl)amino)propan-2-yl)carbamate (27, Cbz-Phe-Phe-(CH$_2$)$_2$-2Pyrmd). To a solution of

Boc-Phe-vinyl-2Pyrmd (**h**, prepared following GP1-GP7, 0.05g, 0.15mmol) in anh. EtOAc (8mL) was added Pd/C (10%wt, 0.016mg) under H_2 atmosphere and stirred overnight. The reaction was filtered, and the filtrate was concentrated. The product was coupled with P_3-P_2 fragment following GP8 to give compound **27**. White solid, 0.031 g, 0.060 mmol, 48% yield. ^{1}H NMR (400 MHz, CDCl$_3$) δ 1.63 – 1.82 (m, 1H), 1.85 – 2.01 (m, 1H), 2.63 – 2.77 (m, 2H), 2.78 – 2.94 (m, 2H), 2.99 (d, 2H, J = 6.8 Hz), 4.05 – 4.23 (m, 1H), 4.31 (pentet, 1H, J = 7.1 Hz), 5.07 (s, 2H), 5.34 (s, 1H), 6.43 (d, 1H, J = 6.4 Hz), 7.04 – 7.38 (m, 16H), 8.59 (d, 2H, J = 4.9 Hz); ^{13}C NMR (100 MHz, CDCl$_3$) δ 31.5, 35.8, 38.8, 41.1, 51.1, 56.8, 67.2, 118.6, 126.6, 127.0, 128.2, 128.3, 128.5, 128.7, 128.8, 129.4, 129.6, 136.4, 136.7, 137.9, 155.9, 157.0, 170.5, 170.9; LC-MS t$_R$ 5.10 min, m/z 509.15, 509.28 [M+H]$^+$, (C$_{31}$H$_{32}$N$_4$O$_3^+$ Calcd 509.26).

2.3.3 Evaluation of Covalent Adducts of Glutathione and PVHIs

The compounds **7**, **11**, **12**, **15, 25, 26**, and **K1777** (0.5 mM) were added to 100 mM Tris (pH 8.0), 10% (v/v) DMSO, and 1 mM or 5 mM reduced glutathione to a final volume of 0.2 mL at room temperature. Samples were analyzed by HPLC-MS (as described above) by injecting 0.01-mL aliquots onto a Luna 5 mm C18(2) 100 Å, 4.6 mm, 50 mm column (Phenomenex) using the HPLC method prescribed in the Section 2.3.1 at 0-6 h time points. The chromatographic peaks for each cruzain inhibitor and its covalent adduct with glutathione were characterized by their values of m/z using electrospray positive-ionization detection and UV absorbance at 254 nm: **K11777**, retention time: 4.75 min, m/z: 575.05; **K11777**-GSH, retention time: 4.64 min, m/z: 882.30; **7**, retention time: 6.71 min, m/z: 507.11; **11**, retention time: 5.48 min, m/z:

506.06; **12**, retention time: 4.77 min, m/z: 520.10; **15**, retention time: 4.81 min, m/z:

535.23; **15**-GSH, retention time: 4.72 min, m/z: 842.42; **17**, retention time: 5.03 min,

m/z: 587.95; **25**, retention time: 4.43 min, m/z: 511.87; **26**, retention time: 4.42 min, m/z:

525.90; **26**-GSH, retention time: 3.61 min, m/z: 833.97. Integration of the

chromatographic peaks of the inhibitors and their GSH-adducts at each time point was

used to determine the rate of GSH-adduct formation. Integration of the chromatographic

peaks of the inhibitors and their GSH-adducts was used at each time point to calculate the

concentration of remaining inhibitor and its GSH-adduct.

2.3.4 Enzyme Preparation

Recombinant human cathepsins B, L, and S were purchased from Millipore

Sigma and used without further treatment. Based on a published protocol[181] with

modifications,[148] detailed procedures for cruzain expression, purification and activation

were described below.

Craik group at UCSF gifted us with the plasmid encoding the *N*-terminally His$_6$-

tagged procruzain with the *C*-terminal domain being truncated (GenBank code:

M84342.1). The plasmid was transformed into ArcticExpress™ (DE3) competent cells

(Agilent, 230192) which were then inoculated into LB media containing 100 µg/mL of

carbenicillin and 20 µg/mL of gentamicin. This starting culture was grown overnight to

saturation at 37 °C. On the next day, 3 mL of starting culture was added to 600 mL of

ZYM-5052 autoinduction media containing carbenicillin and gentamicin inside a 2 L

baffled flask. After being grown until OD$_{600}$ reaching 0.6 (typically 3 – 4 h), this big

culture was moved to 20 °C and incubated for up to 3 d to enable autoinduction without

addition of IPTG. Cells were harvested by centrifugation (5,000 × g for 30 min) and stored at -20 ℃.

Frozen cell pellets were thawed and resuspended in lysis buffer (also called buffer A: 50 mM Tris·HCl, 300 mM NaCl, 10 mM imidazole, pH 10.0) at 4 ℃. Protease inhibitors, including 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM S-Methyl methanethiosulfonate (MMTS), were immediately added to prevent degradation. After being stirred for 30 min, the suspension was added with lysozyme (0.2 mg/mL), DNase I (1 μM), $MgSO_4$ (1 mM), and $CaCl_2$ (1 mM, hygroscopic) and continued to stir for 30 min. The slurry sample was sonicated at 60% amplitude with 20 s pulse every 1 min for 15 times in total to lyse the cells. The resulting lysate was centrifuged at 17, 000 × g for 45 min, and the supernatant was rapidly filtered through a 5 μm filter followed by being transferred into a Superloop™ 150 mL. The filtration was recommended but not necessary. Once the Superloop™ was appropriately connected to the FPLC instrument (ÄKTA™ pure), the sample was loaded onto two pre-equilibrated HisTrap™ FF 5 mL crude columns arranged in a row. The columns were first washed with 100 mL (10 CV) of 100% buffer A at a flowrate of 4.5 mL/min, then eluted with increasing buffer B (50 mM Tris·HCl, 300 mM NaCl, 500 mM imidazole, pH 10.0) with following gradient: 0-40% B over 25 CV; 40-100% B over 5 CV; 100%B for 15 CV. The product peak typically occurred at 25-30% B and was collected to fraction tubes which were immediately added with 1 mM of MMTS. The combined fractions were concentrated to a volume < 50 mL and dialyzed against 3 × 2 L of activation buffer (50 mM NaOAc, 100 mM NaCl, 0.1 mM EDTA-Na₂, pH 5.0) at 4 ℃ overnight. On the next day, the solution in dialysis tubing turned cloudy and was transferred to a capped container. The activation

was initiated by addition of 5 mM DTT and incubated at 37 °C for up to 1 – 2 h until the cloudy solution became apparently clear. Prolonged incubation might lead to further degradation of cruzain, hence the activated sample should be placed on ice and inhibited by 1 mM of MMTS as soon as possible. The purity of cruzain was determined by SDS-PAGE and was usually high enough (> 90%) for enzymatic assays. The clarified solution was buffer-exchanged to general assay buffer (pH 7.5), flash-frozen and stored at -80 °C with 10-20% glycerol (v/v). In some case when purity was not satisfactory, the activated solution was buffer-exchanged to 10X PBS (pH 5.0) and further purified by size exclusion chromatography. The sample must be concentrated to a small volume (3 – 4 mL) and passed through a 0.2 µm filter prior to being transferred to a Superloop™ 10 mL. It was then slowly loaded onto HiPrep™ 26/60 Sephacryl® S-100 HR (GE17-1194-01) at a flowrate of < 0.5 mL/min. The column was washed isocratically with 320 mL (1 CV) of the same PBS buffer in absence of any reducing agent at 1.0 – 1.5 mL/min. Similarly, 1 mM MMTS was added to each of the major fractions upon elution from the column which were stored at -80 °C. Prior to use, the MMTS in the frozen sample should be removed by at least four rounds of centrifugal filtration (Amicon® Ultra, 10 kDa NMWL) using an assay buffer containing 5 mM DTT. The concentration of cruzain was roughly calculated from absorbance at 280 nm using an extinction coefficient (ε_{280}) of 59930 $M^{-1} \cdot cm^{-1}$ (reduced Cys); then it was accurately titrated by varied concentrations of E-64 protease inhibitor.

2.3.5 Enzyme Assays and Evaluation of Inhibitors

All enzyme assays were performed at 25 °C. Initial rates of the peptidolytic

reaction catalyzed by cruzain were measured by monitoring the fluorescence generated

by cleavage of the dipeptide-AMC bond. Assays were conducted in 96-well plates

(Greiner, flat-bottom, clear black plates) in a total volume of 250 µL, containing either 50

mM MES (pH 7.5), 50 mM TAPSO, 100 mM DEA, 1 mM CHAPS, 1 mM Na_2EDTA, 5

mM DTT and 10% DMSO (v/v) or 50 mM sodium acetate (pH 5.5), 50 mM MES, 100

mM TEA, 1 mM CHAPS, 1 mM Na_2EDTA, 5 mM DTT and 10% DMSO (v/v).

Substrates were dissolved in 100% DMSO, and were then diluted 10-fold such that when

added to reaction mixtures, final DMSO concentration were 10% (v/v). Reactions were

initiated with addition of 1-10 µL of cruzain (final concentrations: 0.1 – 3.0 nM (pre-

incubation studies)). Fluorescence was measured on either a SpectraMax M5 (Molecular

Devices) or a Synergy HTX (Biotek, Wisnooki, VT) microplate reader (λ_{ex} = 360 nm,

λ_{em} = 460 nm). Initial rates were determined from continuous kinetic time courses, and

calculated from the earliest time points, typically at less than 10 min.

Compounds were evaluated as inhibitors or inactivators of cruzain in two ways:

(1) enzyme was added to reaction mixtures containing substrate (typically, 10 µM Cbz-

Phe-Arg-AMC) and inhibitor, and reaction time courses were measured for 0-40 min. In

addition to other methods, the effects of all inhibitors on reaction rates were determined

at t = 0-200 s; v_i) and at longer incubation times (t > 1000 s; v_s), to ascertain the

respective inhibition constants K_i and K_i*. (2) Enzyme and compound were pre-incubated

over extended periods of time, and then aliquots were removed and diluted 50- to 100-

fold into reaction mixtures containing 10 µM of Cbz-Phe-Arg-AMC, followed by the assessment of the resulting time courses.

For assays of cathepsin L, B, and S, cruzain inhibitors were evaluated in reaction mixtures containing a buffer of sodium acetate (pH 5.5), 1 mM CHAPS, 1 mM Na$_2$EDTA, and 5 mM DTT at 25 °C. The substrate Cbz-Phe-Arg-AMC was dissolved in 100% DMSO, as were all inhibitors, and aliquots of both substrates and inhibitors were added to 0.25 mL reaction mixtures to final concentrations of 10% DMSO (v/v). Michaelis constants for Cbz-Phe-Arg-AMC were determined for all three human cathepsins as: cathepsin L (2.9 µM), cathepsin S (60 µM), and cathepsin B (150 µM), and fixed concentrations of Cbz-Phe-Arg-AMC of 1 or 2 K_m were used to evaluate inhibitors. Cruzain inhibitors were added at seven concentrations and one fixed concentration of Cbz-Phe-Arg-AMC, and time courses of AMC formation were analyzed as with cruzain.

2.3.6 Evaluation of Cruzain Inhibitors in Axenic Cell Cultures of T. b. brucei and T. cruzi

Selected cruzain inhibitors were evaluated in axenic cell cultures of *T. b. brucei* (procyclic trypomastigotes; ATCC PRA-381) and *T. cruzi* (epimastigote forms; strain Y, ATCC 50832GFP). *T. b. brucei* was grown in SDM-79 medium and *T. cruzi* was grown in ATCC medium (1029 LIT medium). Both media included fetal calf serum (10%) and penicillin/streptomycin (50 U/mL). Test compounds, including **K11777**, were dissolved in 100% DMSO, and added to cell cultures at final concentrations of 0.5-20 µM (maximum DMSO = 1% (v/v)). Control samples contained equal amounts of DMSO. *T. b. brucei* and *T. cruzi* (5 mL in flask cultures) at 26°C were seeded at ~3 x 10^6 cells, and

diluted daily maintaining a mid-log growth phase for up to 120 h. Treated cells were

typically grown for 4 days (*T. b. brucei*) or 5 days (*T. cruzi*). After each cell dilution,

fresh compound or an equal volume of DMSO (control samples), was supplemented into

the cultures, while maintaining a constant concentration of each inhibitor. Cell counts

were scored using a Z2 Coulter Counter.

2.3.7 Evaluation of Cruzain Inhibitors in T. cruzi-infected Murine Cardiomyoblasts

For the evaluation of the anti-trypanocidal activity of cruzain inhibitors, we

infected a C2C12 mouse cardiomyoblast cell line (ATCC CRL-1772) with *T. cruzi* strain

Ca-I/72 (a gift from James Dvorak, National Institutes of Health) in 1536-microwell

plates. In each well was added 10^3 cells and 10^4 parasites in a total volume of 10 µl

including the test compounds in 10-point dose-response dilutions starting at 10 µM (3-

fold dilutions). The plates were incubated at 37ºC for 48h, and the wells were fixed with

2% paraformaldehyde in PBS, and stained with 5 µg/ml of 4′,6-diamidino-2-

phenylindole. After at least 30 min of incubation at room temperature in the dark, the

plates were read in an automated microscope, ImageXpress MicroXL (Molecular

Devices), and the images were analyzed by custom-built software to quantify and assess

viability of the parasites, as well as the host cells independently. The compilation of data

was used to calculate the antiparasitic activity (EC_{50}) and host cytotoxicity (CC_{50}).

2.3.8 Evaluation of Human Cell Toxicity

Primary human dermal fibroblast (HDF) cells were used to evaluate human cell

toxicity of cruzain inhibitors. HDF cells were plated in a 384-well plate at 2400 cells/well

(62,000 cells/mL). Inhibitors in 100% DMSO were added in duplicate to final concentrations of 0.001 – 0.1 mM and 1% DMSO (v/v) with 1% DMSO as a control sample), and cells were cultured at 37°C for 48 h, followed by addition of resazurin. Cell viability was then assessed by reading of fluorescence (λex/λem: 544 nm/590 nm) after an additional 24 h of incubation.

2.3.9 Molecular Modeling

Molecular docking for selected cruzain inhibitors was performed in Schrödinger Suite software package. The receptor (PDB accession code: 2OZ2) was prepared by Protein Preparation Wizard and the ligands were translated and optimized by LigPrep modul. Glide module was used for a conventional non-covalent docking while CovDock module was used for a covalent docking, **K11777** in the crystal structure was set as a reference ligand and placed in the center of a cubic enclosing box of 18 $\mathring{A}^3$. Cys_{25} was selected as the reactive residue. The SMARTS pattern of ligands was defined on the basis of molecular structure. The binding affinity was scored and ranked using Glide. The best three poses were generated as .pdb files and manually inspected.

2.3.10 Analysis of Kinetic Data

Initial velocity data for cruzain-catalyzed reactions of fluorogenic peptide substrates were determined by fitting to eq. 2-1 using GraphPad Prism or SigmaPlot. For eq. 2-1, k_{cat} is the turnover number, $[E]_t$ is the concentration of active sites of cruzain, and K_M is the Michaelis constant for the substrate. Cruzain concentrations were determined by spectrophotometric analysis of purified sample solutions.

$$\frac{v}{[E]_t} = \frac{k_{cat}[S]}{K_M + [S]} \quad\quad (2\text{-}1)$$

Competitive inhibition was fitted to eq. 2-2, in which [S] and [I] are concentrations of substrate and inhibitor, respectively, V_{max} is the maximal velocity, and K_{is} is the slope inhibition constant.

$$v = \frac{V_{max}[S]}{K_M(1 + [I]/K_{is}) + [S]} \quad\quad (2\text{-}2)$$

Data for time-dependent inhibition were fitted by several methods. All time-course data were fitted to eq. 2-3 for studies in which reaction was initiated by the addition of enzyme, wherein P is the fluorescence generated by AMC formation, C is a non-zero constant, v_s and v_i are respectively the steady-state and initial enzymatic rates, t is time, and k_{obs} is the observed rate of conversion of the initial inhibited rate to the final inhibited rate.[175] In cases for which reaction was initiated with an excess of substrate, following pre-incubation of enzyme and inhibitor, for eq. 2-3, $v_i = 0$.

$$P = v_s t + \left[\frac{v_i - v_s}{k_{obs}}\right][1 - e^{(-k_{obs}*t)}] + C \quad\quad (2\text{-}3)$$

Values of k_{obs} vs. [inhibitor] were then re-plotted and fitted to eq. 2-4, for which k_3 and k_4 represent the respective rates of formation and dissolution of the EI* complex as depicted in **Scheme 2.2**.

$$k_{obs} = k_4 + \frac{k_3[I]}{K_i\left(1 + \frac{[S]}{K_M}\right) + [I]} \quad\quad (2\text{-}4)$$

Inhibition constants were also obtained by fitting v_i and v_s data to eq. 2-5, in which v_x is the rate in the presence of inhibitor for either early (v_i) or late (v_s) phases of each time course, v_0 is the rate in the absence of inhibitor, K_M is the Michaelis constant of the substrate, and K_{ix} is the apparent inhibition constant, K_i or K_i*, obtained from fitting of v_i or v_s, respectively.

$$\frac{v_x}{v_0} = \frac{1}{1+[I]/[K_{ix}\left(1+\frac{[S]}{K_M}\right)]} \tag{2-5}$$

2.4 Conclusions

We have developed a novel class of reversible inhibitors for the essential cysteine protease of *Trypanosoma cruzi*, cruzain. These compounds, the peptidomimetic vinyl heterocycles, contain bioisosteric replacements for the acrylamide and vinyl sulfone warheads present in irreversible, covalent inactivators such as **K11777**. We also demonstrated that PVHIs containing vinyl-2-N-methylpyridinium or vinyl-2-N-methylthiazolium groups, unlike other inhibitors, readily form Michael adducts with glutathione. Our survey demonstrated that the most optimal cruzain inhibitors contained vinyl-2-pyrimidine, vinyl-2-pyridine, and vinyl-2-N-methylpyridinium groups. These PVHIs proved to be potent, time-dependent inhibitors of cruzain, albeit, fully reversible in terms of mode of action. These PVHIs are significantly active in both axenic cultures of *T. b. brucei* and in a cell infection model of *T. cruzi*, and further optimization may produce more potent anti-trypanosomal agents. Importantly, the concept of reversible covalent inactivation by vinyl-heterocycles is potentially expandable to other enzymes which contain active-site cysteines, such as EGFR, G12C K-ras and other protein kinases for which irreversible acrylamide inactivators comprise effective drugs.[182]

CHAPTER 3

SELF-MASKED ALDEHYDES AS A NOVEL CLASS OF CYSTEINE PROTEASE

INHIBITORS FOR *TRYPANOSOMA CRUZI* AND SARS-COV-2

3.1 Introduction

Among the reversible covalent warheads explored for cruzain to date, an aldehyde

group often, if not always, affords outstanding inhibitory potency, provided that the

peptide scaffold can be accommodated by the binding site. A simple peptide aldehyde

KS5 (Cbz-Phe-Ala-H) showed nanomolar potency (IC_{50} = 7.2 nM) vs. cruzain, on par

with a vinyl sulfone inactivator **KS48** that is structurally analogous to **K11777** (**Figure

3.1A**).[183] All efforts to install a conformationally-constrained scaffold, either the γ-lactam

(**KS12**) or the α-methyl Phe at P$_2$ (**KS30**), failed to maintain the original geometry and

thus caused drastic loss of binding affinity. Using the same Cbz-Phe-Ala- or similar Cbz-

Phe-Val- peptidomimetic scaffolds, another research effort aiming to hunt for assorted α-

keto-based warheads, including α-ketoesters, α-ketoamides, and α-ketoacids, were

inferior inhibitors to their aldehyde counterparts (**Figure 3.1B**).[184] This fact could be

attributed to the higher electrophilicity of an aldehyde, or its smaller size might also assist

with accessing the catalytic center. Later studies further expanded the scope of warheads

to nitriles, oximes, azanitriles, and some heterocycles (**Figure 3.1C**).[185-186] Except for

azanitriles, aldehyde inhibitors not only had superior potency for cruzain, but also

showed significant trypanocidal activity in mammalian cells infected with a strain of *T.

cruzi* amastigotes. Azanitriles did show improved inhibition of cruzain, in accordance
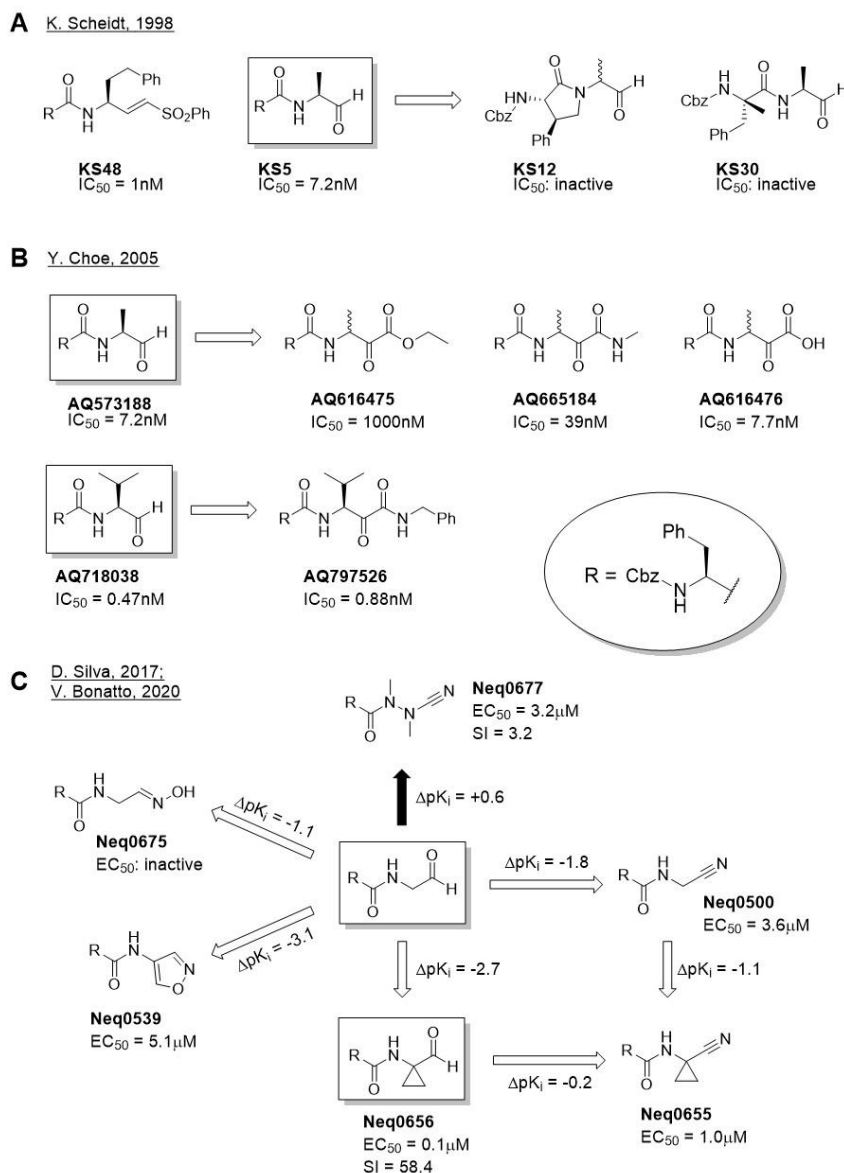
Figure 3.1 Peptidomimetic inhibitors equipped with an aldehyde group displayed superior inhibitory activity against cruzain (IC_{50}) and *T. cruzi* (EC_{50}) over many other warheads. R group is invariant for all compounds in this figure. SI represents selectivity index for inhibition of *T. cruzi* over host cell. ΔpK_i values in (C) indicate the differences in potency between the reference aldehyde ($K_i = 7.9$ nM) and compounds bearing various warheads.

with the computation study and reaction kinetics which implied that the intrinsic

reactivity of an azanitrile is superior to an aldehyde. Nonetheless, they did not surpass the

activity of the aldehyde at cellular level, and were more toxic to the host cells. As for the

P_1 position, these studies collectively suggested that, while a strained structure

(cyclopropyl, **Figure 3.1C**) was not a good choice, lipophilic residues like Phe, Cha, or

Val (**Figure 3.1B**) contributed, perhaps entropically, to binding albeit no apparent

protein-ligand interaction was observed. In brief, many pieces of evidence have proven

that an aldehyde group comprises an excellent warhead for cruzain and other cysteine

proteases, although it has been primarily used as a starting point for lead optimization or

as a chemical probe.

Despite foregoing facts, an aldehyde is commonly deemed a double-edged sword

in drug design. The major concern of using aldehydes in therapeutic molecules is linked

to their inherently high reactivity that may give rise to immunotoxicity and poor

pharmacokinetic properties.[187-189] However, apart from providing only tool molecules,[190-192] aldehydes indeed have been used as drugs for over half a century in compounds

including natural products and synthetic molecules.[193] Generally speaking, aldehyde

substituents in natural products are well-tolerated, and much less susceptible to

metabolism. For example, the aldehyde-containing streptomycin is a WHO medicine and

is considered to be extremely safe;[52] moreover, its aldehyde group is not even involved in

its mechanism of action.[194] Besides, many natural products are quite hydrophilic so that

their non-specific interactions with off-target biomolecules are minimized.[195] On the

contrary, aldehyde groups in synthetic medicines likely have more systemic exposure to

both targets and off-target protein species.[193] These aldehydes can react with a number of

nucleophiles in biological systems: they can form Schiff bases with free amines like

lysines, hydrazides, hydrazines, protein *N*-termini and nucleobases; also, they can form

hemiacetals with alcohols, and thiohemiacetals with thiols. Therefore, they are typically

of great importance to the biological activity; meanwhile they are frequently subject to

aldehyde metabolism and induce immunotoxicities, liver injuries, and undesired drug-drug interactions. Theoretically, confining drug exposure to the germane protein target and eschewing off-target biomolecules are two principal approaches to address the safety and pharmacokinetic stability issues associated with aldehydes. In practice, however, such approaches have been challenging to implement.[162]

An alternative approach to incorporate aldehydes into drug design is to mask the aldehyde using another functional group. The anti-malarial aldehyde inhibitor of falcipain **PG3b** was "masked" by an 1,2,4-trioxolane structure, resulting in the prodrug **PG4b** from which the active aldehyde is liberated upon reaction with Fe(II) inside the vacuoles of the plasmodial parasite (**Figure 3.2A**).[196-197] Unfortunately, this clever masked aldehyde resulted in toxicity to the host cells, apparently from the action of free radical byproduct(s). Its utility is also limited by the variable availability of Fe(II) in target cells. In a study of a calpain inhibitor, Nakamura *et al.* synthesized a peptidyl aldehyde with a homoserine at P_1 position which spontaneously formed a cyclic hemiacetal **SNJ1715** (**Figure 3.2B**).[198-199] Compared to the free aldehyde **SJA6017** with P_1 Leu, this self-masked aldehyde acquired improved physicochemical properties and excellent
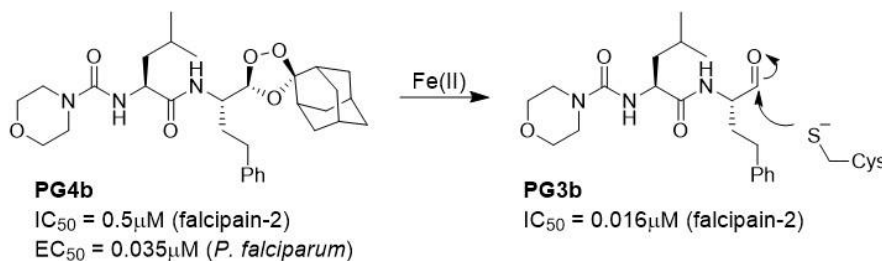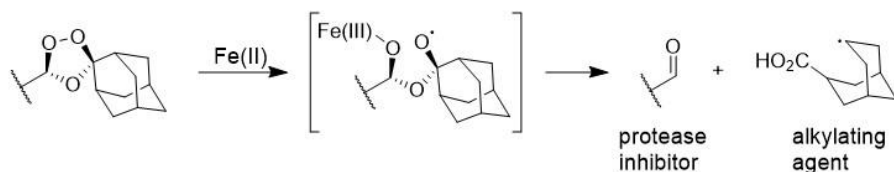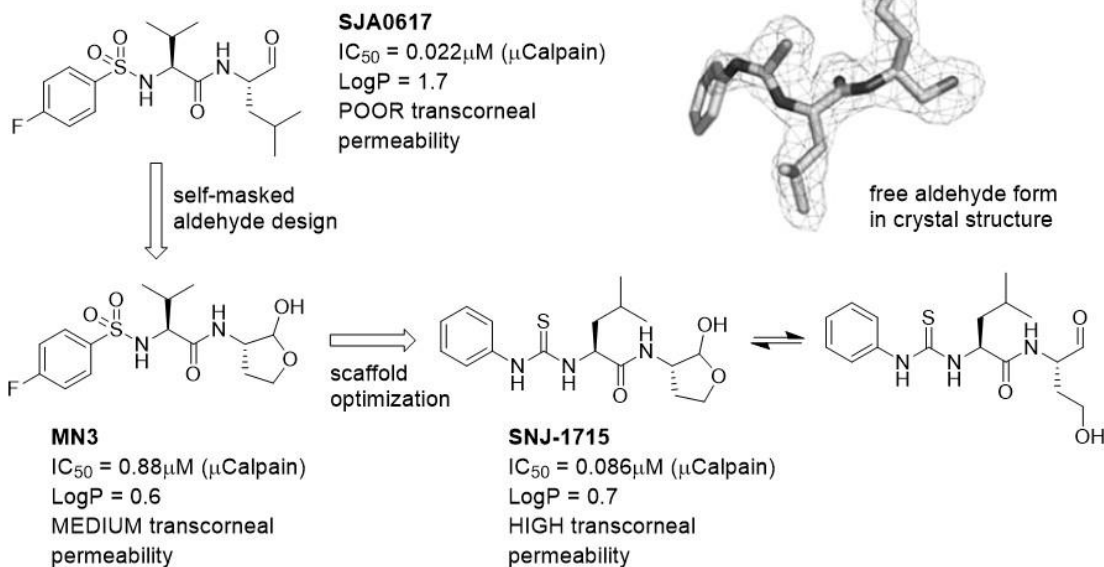
Figure 3.2 Examples of masked aldehyde strategy in the design of protease inhibitors. (A) 1,2,4-trioxolane inhibitor breaks down to free aldehyde and radical to effect antimalaria activity. (B) γ-lactol inhibitor releases its free aldehyde to inhibit μ-calpain by forming thiohemiacetal with the catalytic cysteine.

transcorneal permeability, although it showed less activity against calpain.

Crystallographic study (PDB accession: 2G8E) revealed the authentic binding form to be the free aldehyde which formed a thiohemiacetal with the catalytic Cys_{115}. This interesting finding prompted us to scrutinize the practice of self-masked aldehyde

inhibitors (SMAIs) to cruzain inhibitors. To my knowledge, this strategy has not been reviewed and studied systematically, although it has been applied occasionally to other highly reactive electrophiles recently.[200]

In the end of this chapter, we also attempted to make SMAIs for inhibiting a cysteine protease of SARS-CoV-2 which causes the formidable COVID-19.[201] The ongoing pandemic not only has deprived us millions of lives, it also resulted in the largest global recession since World War II.[202] Immediately after the outbreak, thousands of research programs have been focusing on the development of treatments and vaccines.[203-204] Promising drug targets of SARS-CoV-2 include the coronaviral spike protein, RNA-dependent RNA polymerase (RdRp), 3C-like protease (3CL[pro]) and papain-like protease (PL[pro]).[205-206] The 3CL[pro] is also a major viral cysteine protease belonging to clan PA family C3, and shares a high sequence identity (~96%) with that of SARS-CoV. Thanks to plentiful research on SARS-CoV in past years, a vast number of compounds have been accumulated against its 3CL[pro] and may be effective for the new virus. Among those compounds, peptidomimetic aldehydes also displayed exceptional inhibitory activity.[207-209] The P_1 position of peptide substrates, which are nearly invariably glutamine residues, is critical to the binding affinity, and may be substituted by γ- or δ-lactams, which mimics a Gln in endogenous substrates. We have taken advantage of this moiety to make SMAIs, of which more details regarding the design and effect are described later in this dissertation.

3.2 Results and Discussion

3.2.1 Rationale of SMAIs

In Chapter 2, the dipeptide aldehyde **1** (Cbz-Phe-Phe-H) was synthesized as an intermediate for preparation of a series of vinyl heterocyclic inhibitors, and proved to be an extraordinarily potent inhibitor of cruzain, as discussed below. Compound **1** is also a non-natural, fairly lipophilic molecule. As stated in the introductory section, such a molecule is likely vulnerable to aldehyde metabolism and is also inclined to be immunotropic. In order to modify its P_1 structure to afford a self-masked aldehyde, we added a hydroxyl group to the 2' position of the phenyl ring, converting the P_1 group to an *ortho*-tyrosine (**2**, Cbz-Phe-*o*-Tyr-H) (**Figure 3.3A**). The oxygen of the phenol group is in close proximity to the aldehydic carbon permitting a nucleophilic addition reaction to occur, producing a cyclic hemiacetal (δ-lactol) similar to the γ-lactol found in calpain inhibitor **MN3**.[198] We anticipated that a SMAI would remain "locked" in its δ-lactol form before binding to cruzain, after which enzyme catalysis would elaborate the free aldehyde, followed by formation of a hemithioacetal adduct with cruzain (**Figure 3.3A**). If true, then the potential advantages of SMAI include: (a) the intramolecular nature of the δ-lactol likely provides sustained protection of the aldehyde outside of an enzyme active site; (b) compared to the 1,2,3-trioxolane masked aldehyde,[196] the cleavage of the hemiacetal does not produce reactive byproducts; and (c) the introduced hydroxyl group on the *o*-tyrosine sidechain is small enough so as to impose minimal perturbation of the original binding mode of the parent aldehyde **1**.

3.2.2 Computer-aided Inhibitor Design

The first question arising from our hypothesis is: could the lactol form of **2** bind

to cruzain and remain closed? To preliminarily address this question, we employed

molecular modeling using docking methods conforming to the binding of compound **2** in

both its noncovalent, lactol form (**Figure 3.3B**), or as its free aldehyde form which was
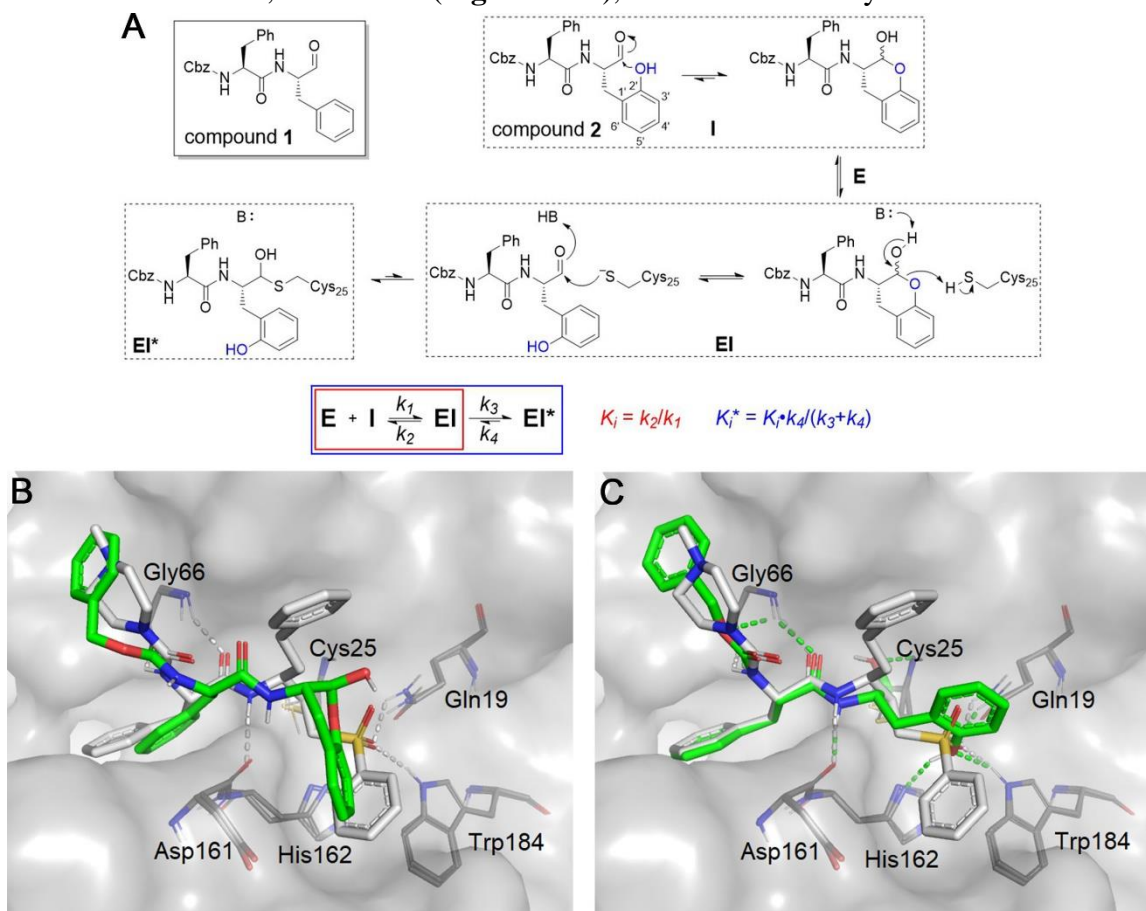
Figure 3.3 Rationale of the SMAI design. (A) The aldehyde group of **2** is expected to be masked by the 2'-phenol group via spontaneous formation of a lactol. It is anticipated that the SMAI will undergo enzyme-catalyzed opening of the lactol ring, and subsequently form the hemithioacetal adduct with Cys25. The scheme describes a two-step inhibition mechanism in which rapid formation of an **EI** complex precedes isomerization to **EI***, which slowly converts back to **EI**. (B) Lactol form of **2** (green) non-covalently docked to cruzain. (C) Opened form of **2** covalently docked to form a hemithioacetal with Cys25. Both structures are superimposed with a covalently-bound **K11777** (white) from crystal structure (PDB ID: 2OZ2). Colored dashed lines represent corresponding cruzain-inhibitor interactions.

docked as a hemithioacetal adduct with the catalytic Cys$_{25}$ (**Figure 3.3C**). The covalent

adduct of **2** is predicted to bind in a manner similar to that of the covalent inactivator

K11777 as seen in the cruzain-**K11777** co-crystal structure. As expected, the phenoxy

substituent of the covalently-bound inhibitor **2** is well tolerated in the cruzain active site,

and may form hydrogen bonds with Gln$_{19}$, His$_{162}$, or Trp$_{184}$, similar to the sulfone

oxygens of **K11777**. Although the Cbz-Phe group in **Figure 3.3B** adopts a similar

orientation as the opened form (**Figure 3.3C**), the lactol displays a puckered

conformation in the active site. The bound lactol has a poorer binding free energy

($\Delta G_{predicted} = -5.08$ kcal/mol) than that of the hemithioacetal ($\Delta G_{predicted} = -7.24$ kcal/mol),

apparently owing to the bicyclic lactol moiety. One may infer that the recognition of the

Cbz-Phe scaffold of **2** by cruzain assists in the binding of the lactol group, and orients it

for enzyme-catalyzed ring-opening to yield the high-affinity aldehyde.

3.2.3 Kinetic Analysis of SMAIs

A series of SMAIs and related compounds **1 – 12** were prepared and evaluated as

inhibitors of cruzain (**Table 3.1**). Time-course data for inhibition of cruzain by many of

these compounds conformed to the kinetic scheme shown in **Figure 3.3A**, that is,

initiation of reaction by adding enzyme to substrate and inhibitor led to curvilinear time

courses, in which reaction rates demonstrably decreased as the **EI** complex (characterized

by K_i) progressed to the tighter **EI*** complex (characterized by K_i*). The observation of

time-dependence may or may not indicate that covalent bond formation has occurred, but

it does reflect that K_i* $< K_i$, arising from either the formation of a hemithioacetal between
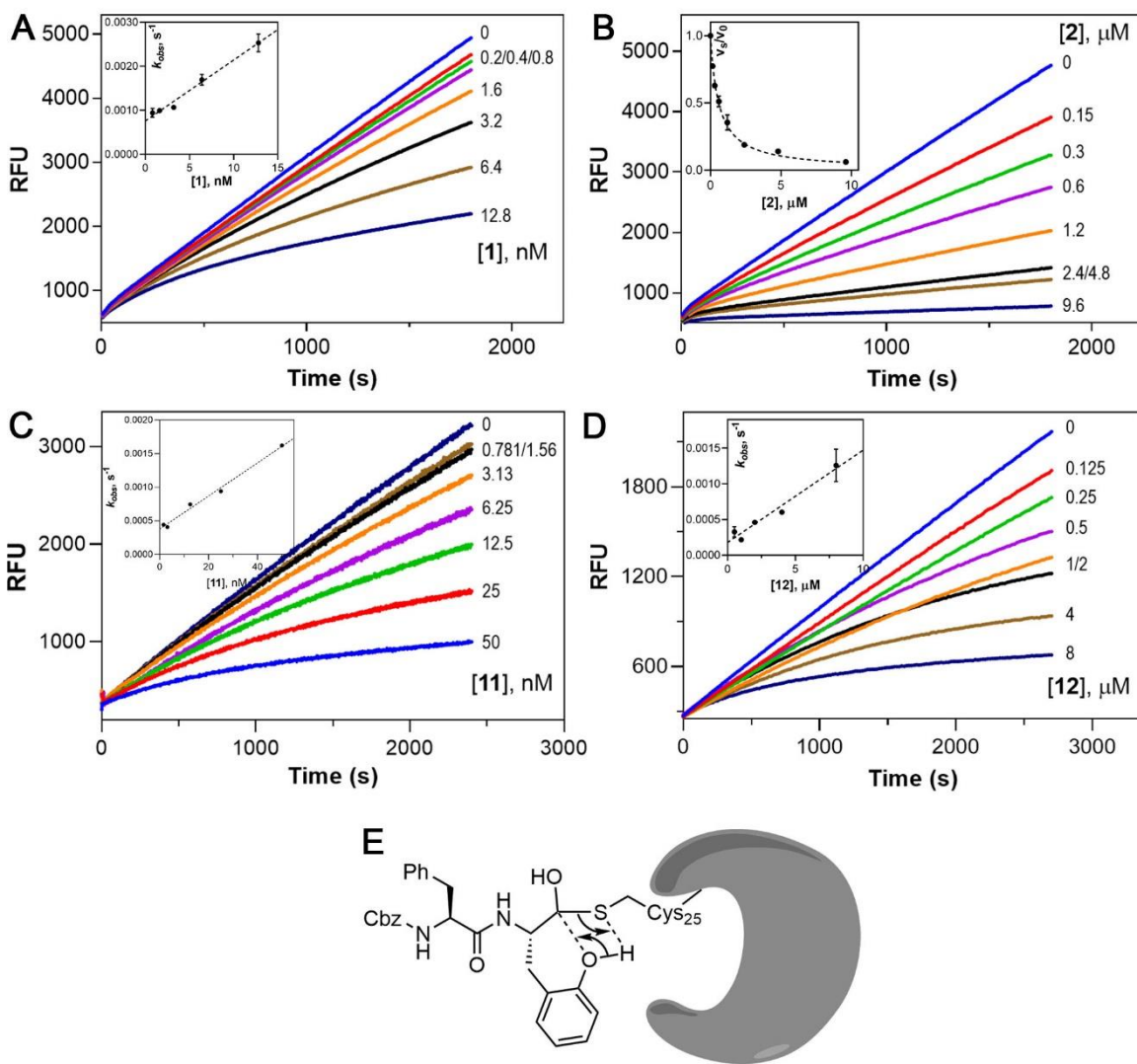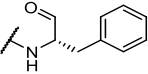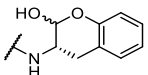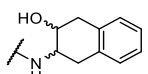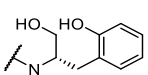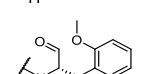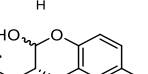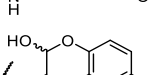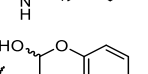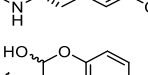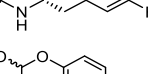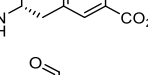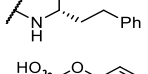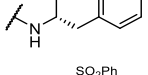
Figure 3.4 Kinetic analysis of inhibition of cruzain by **1**, **2**, **11**, and **12**. (A-D) Time-courses of cruzain inhibition. Insets for (A), (C) and (D): the k_{obs} values were obtained by fitting progress curves to eq. 2-3. Replot of k_{obs} vs. [I] with the line drawn through data points from fitting to eq. 3-2. **1**: $k_4 = (7.6 \pm 0.6) \times 10^{-4}$ s^{-1}; **11**: $k_4 = (3.8 \pm 0.4) \times 10^{-4}$ s^{-1}; **12**: $k_4 = (1.8 \pm 0.7) \times 10^{-4}$ s^{-1}. Inset for (B): the steady-state rates with (v_s) and without inhibitor (v_0) were obtained at t ≥ 20 mins. Plot of v_s/v_0 vs. [**2**] with the line drawn through data points from fitting to eq. 2-5. (E) A proposed mechanism of phenoxy-assisted conversion of **EI*** back to **EI**.

enzyme and inhibitor, or a slow isomerization step of **EI** to **EI*** not involving covalent

bond formation.

We first evaluated cruzain inhibition by time-course data containing fixed

concentrations of substrate and variable concentrations of inhibitors, for which reactions

Table 3.1 Inhibition data of SMAIs and related compounds for cruzain, human cathepsin L and B.

Compound	Structure P3-P2	P1	Overall inhibition constant K_i* (nM)[a] cruzain	human cathepsin L	human cathepsin B
1	Cbz-Phe-	(aldehyde, Ph)	0.44 ± 0.02	ND	ND
2	Cbz-Phe-	(HO, O, benzopyran)	49 ± 2	28 ± 0.9	4500 ± 100
3	NMePip-Phe-	(HO, tetrahydronaphthalene)	>100,000	ND	ND
4	Cbz-Phe-	(HO, HO, Ph)	>10,000	ND	ND
5	Cbz-Phe-	(O, OMe, Ph)	22 ± 2	ND	ND
6	Cbz-Phe-	(HO, O, OMe)	350 ± 32	38 ± 1.6	5500 ± 400
7	Cbz-Phe-	(HO, O, methyl)	103 ± 5	58 ± 2.2	6100 ± 900
8	Cbz-Phe-	(HO, O, Cl)	74 ± 10	27 ± 0.7	1400 ± 100
9	Cbz-Phe-	(HO, O, F)	48 ± 2	23 ± 1.3	2300 ± 200
10	Cbz-Phe-	(HO, O, CO_2Me)	18 ± 0.5	10.8 ± 0.8	670 ± 30
11	NMePip-Phe-	(aldehyde, Ph)	0.5 ± 0.2	ND	ND
12	NMePip-Phe-	(HO, O, benzopyran)	47 ± 2	20 ± 0.9	1300 ± 100
K11777	NMePip-Phe-	(SO_2Ph, Ph)	0.2[b]	0.2[b]	5.7[b]

[a]Cruzain was assayed at pH 7.5 while hCatL and hCatB were assayed at pH 5.5. [b]Reported as apparent IC_{50} in ref 158.

were initiated by the addition of enzyme. Inhibition of cruzain by the free aldehyde **1** was

characterized by downward-concave curvilinear time courses, often referred to as "burst

kinetics" (**Figure 3.4A**), indicative of time-dependent inhibition for which equilibrium

between cruzain and **1** was slowly established over 30 minutes. Each curve was fitted to

eq. 2-3 and the resulting values of k_{obs} were replotted vs. inhibitor concentration [**1**] (**Figure 3.4A**, inset). The replot was best fitted to eq. 3-2 (refer to Section 3.3.5), implying a lack of saturation of the **EI** complex by **1**, for which $K_i \gg K_i^*$. The unimolecular rate constant for conversion of **EI*** back to **EI** was extremely slow [$k_4 = (7.6 \pm 0.6) \times 10^{-4}$ s^{-1}]. The low value of k_4 is likely the reason for the slow onset of inhibition and potency of aldehyde **1**.

Unlike aldehyde **1,** inhibition of cruzain by **2** exhibited almost linear time courses with only slight curvature observed at early stages (**Figure 3.4B**), as was also observed for SMAIs **6** – **10** (all containing the Cbz-Phe- scaffold). The nominal burst phase was only observed within the first 1-2 min, after which an apparent steady-state reaction was established, indicating a significantly faster rate (greater k_4) for conversion of **EI*** back to **EI** than for compound **1**. This difference suggests that, compared to a relatively stable hemithioacetal intermediate cruzain-**1** species, the reverse reaction of this intermediate for a SMAI is likely facilitated by the attack of its phenoxy group on the hemithioacetal (**Figure 3.4E**). To characterize these inhibitors, we obtained steady-state rates of v_s and v_0 at assay times $\geq$ 20 min. For instance, upon plotting v_s/v_0 vs. [**2**] (**Figure 3.4B**, inset) and fitting using eq. 2-5, we obtained a value of the overall inhibition constant (K_i^*) of 49 $\pm$ 2 nM. For the free aldehyde **11** and the SMAI **12**, both of which incorporated the N-methylpiperazinyl-Phe scaffold (NMePip-Phe) as found in **K11777**, the ratio of their activities [K_i^*(**12**)/K_i^*(**11**) = 94] as cruzain inhibitors was almost identical to that of their Cbz-Phe-containing counterparts **1** and **2** [K_i^*(**2**)/K_i^*(**1**) = 110], and was also comparable to that of aforementioned **SJA0617** and its SMAI **MN3** (**Figure 3.2B**). Similar to aldehyde **1**, slow onset of inhibition was observed for **11** (**Figure 3.4C**). Surprisingly,
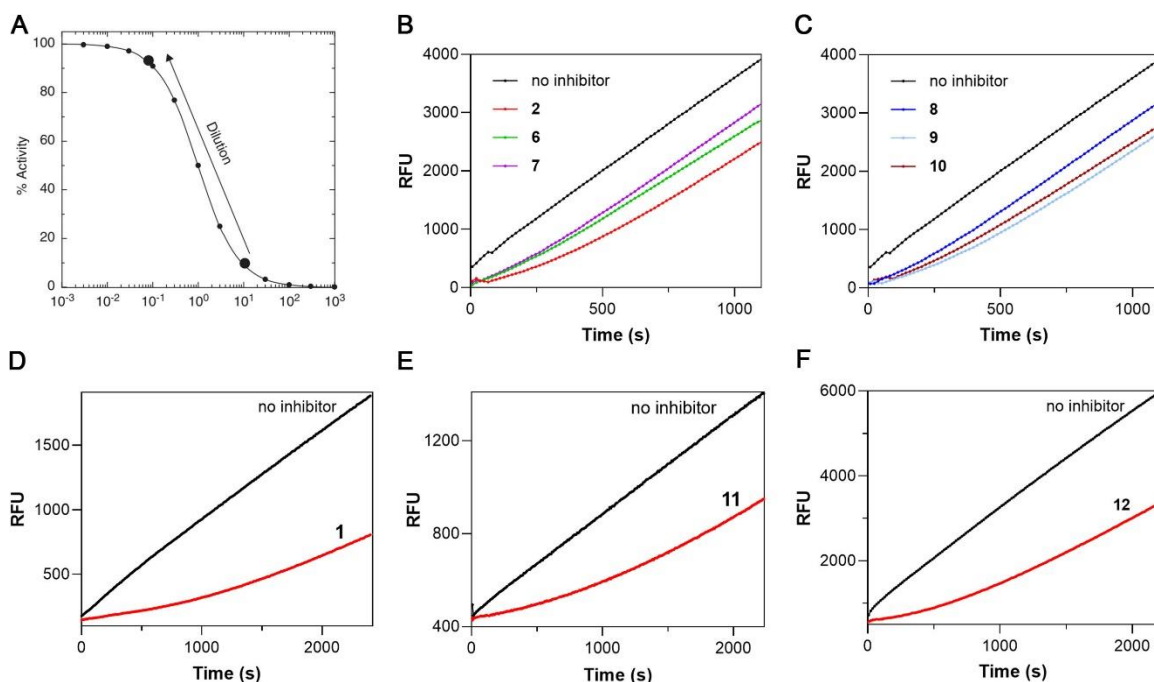
Figure 3.5 Rapid dilution assay for SMAIs. (A) Dilution scheme for testing the reversibility of a SMAI. Upon 100-fold dilution, the inhibitor concentration decreases from 10-fold > K_i^{*app} (91% inhibition) to 10-fold < K_i^{*app} (9% inhibition). (B, C) Cruzain activity rapidly recovered from inhibition by **2** and **6 – 10**. (D-F) Cruzain activity recovery showed a significant lag for **1**, **11**, and **12**. All curves were fitted to eq. 2-3 to provide k_{obs} values listed in **Table 3.2**.

unlike SMAIs containing Cbz-Phe- scaffold, **12** also exhibited time-dependent inhibition

(**Figure 3.4D**).

To explore the reversibility of cruzain inhibition by these compounds, we first

pre-incubated each inhibitor with cruzain, followed by a 100-fold rapid dilution

accompanying the addition of substrate (**Figure 3.5A**). For inhibitors **2** and **6 – 10**

(**Figure 3.5B, C**), 91% recovery of cruzain activity was observed over the course of

minutes, wherein the residual concentration of inhibitor was $0.1 \times K_i^{*app}$, capable of

affording 9% inhibition, thereby demonstrating that binding of these inhibitors was fully

reversible. In comparison, the activity of cruzain pre-incubated with aldehydes **1**, **11**, or

Table 3.2 Kinetic parameters obtained from rapid dilution assays

Compound	k_{obs} ($\times 10^{-4}$ s^{-1})	Residence time τ (s)^a	Estimated k_4 ($\times 10^{-4}$ s^{-1})^b
1	3.18 ± 0.24	3141	2.89
2	30.2 ± 1.9	331	27.5
6	67.4 ± 4.8	148	61.3
7	54.3 ± 3.5	184	49.4
8	62.4 ± 2.5	160	56.8
9	32.1 ± 1.3	312	29.2
10	53.1 ± 5.7	188	48.3
11	3.71 ± 0.21	2695	3.37
12	6.33 ± 0.42	1580	5.75

^{a}Residence time $\tau = 1/k_{obs}$; $^b k_4$ values were estimated from k_{obs} based on eq. 3-2.

SMAI **12** was restored at much slower rates (**Figure 3.5D-F**). After fitting these "lag" time courses to eq. 2-3, the calculated residence times ($\tau = 1/k_{obs}$) for compounds **2** and **6** – **10** ranged from 2.5 to 5.5 min, whereas **1**, **11**, and **12** exhibited significantly longer residence times of 52, 45, and 26 min, respectively (**Table 3.2**). While longer residence times were expected for aldehydes **1** and **11**, compound **12** was the only SMAI to effect significant time-dependent inhibition as evidenced by the observation of both burst and lag kinetic time courses, and the rate of conversion of the **EI*** complex to **EI** was comparable to that of aldehyde **11**. Considering that compound **12** was equipotent to **2**, one could infer that the peptidomimetic scaffold of **12** might also affect the adduct formation so that k_3/k_4 remained nearly unchanged for both the SMAI and the free aldehyde.

3.2.4 Structure-Activity Relationships of SMAIs

Inhibition constants for all compounds prepared for this study are shown in **Table 3.1**. The equipotency of SMAIs **2** and **12** at ~50 nM suggests that their different P_3 sidechains have no impact on their inhibition. We prepared compound **3**, an analog of SMAI **12** in which its ether oxygen was replaced with a methylene group, and the resulting compound containing the stable tetrahydro-naphthol did not inhibit cruzain at ≤ 100 µM, despite its similarity to the chroman-2-ol group of the SMAIs. Substitution of the aldehyde of compound **2** with a primary alcohol also resulted in the poor inhibitor **4**, which corroborates the importance of the formation of carbon-sulfur bond in the cruzain-SMAI complex. In the free aldehyde **5**, the phenoxy group was methylated, affording an inhibitor that is 50-fold less active than aldehyde **1**, implying that a bulkier 2' substituent sterically hinders the formation of the hemithioacetal.

Another potential feature of this SMAI structure is the ability to tune the reactivity of phenol group by installing substituents on the benzene ring. We introduced several 5'-substituents including electron-donating groups (e.g., -OMe, -Me), weak electron-withdrawing groups (e.g., -F, -Cl), and moderate-to-strong electron withdrawing groups (e.g., $-CO_2Me$). As shown in **Table 3.1**, the electron-donating substituents on **6** ($K_i^* = 350$ nM) and **7** ($K_i^* = 100$ nM) appear to undermine their potency, as these inhibitors bind, respectively, 7- and 20-fold more weakly that un-substituted SMAI **2**. The stronger electron-withdrawing capability of the ester substituent (**10**; $K_i^* = 18$ nM) provided a 3-fold increase in potency vs. **2**, suggesting that the 5'-methyl ester facilitates the opening of the lactol ring. In addition to the electronic effect, a steric effect is likely operative. The methoxy group of compound **6** is bulkier than the methyl group of **7**,

which may contribute to its over 3-fold lower activity. Chlorine (0.79 Å, **8**) has a slightly

larger radius than fluorine (0.42 Å, **9**), and fluorine is similar in size to a hydrogen (0.53

Å, **2**) (note: the C-F is longer than the C-H bond so that, at that increased length, the

fluorine is effectively the same radius as hydrogen),[210] consistent with the differences

between their K_i* values. It is interesting that the steric effect apparently dictates the

differences in potency between compounds **8** and **9**, probably because fluorine and

chlorine are only weakly electron-withdrawing. The methyl ester of **10**, while comprising

the bulkiest substituent among this series of compounds, provided an inhibitor of nearly

three times the potency of un-substituted **2**. Apparently, the "positive" electron-

withdrawing effect afforded by this large substituent overcomes any "negative" steric

effect, resulting in the SMAI of highest potency in this study. These results suggested that

appropriate substitution of the P_1 phenyl ring in SMAIs could optimize their potencies,

and further, may affect the stability of enzyme-bound lactol.

3.2.5 Mechanistic Study of SMAI Inhibition

Although a kinetic mechanism of SMAI is proposed above, the exact chemical

mechanism of inhibition has not yet been corroborated directly by experimental evidence.

To this end, 2D NMR is an ideal technique because it is applicable to sample in aqueous

solution, and is able to provide dynamic structural information that is more authentic than

static structural information afforded by X-ray crystallography. The most obvious

drawback of 2D NMR is the poor detection limit due to low abundance of ^{13}C, especially

when the molecule of interest is not very soluble in aqueous solutions. Therefore, it was

necessary for us to make a ^{13}C-enriched inhibitor. More specifically, the aldehydic carbon

of compound **12** was ^{13}C-labeled because this compound has higher aqueous solubility

than other compounds with Cbz-Phe- scaffold.
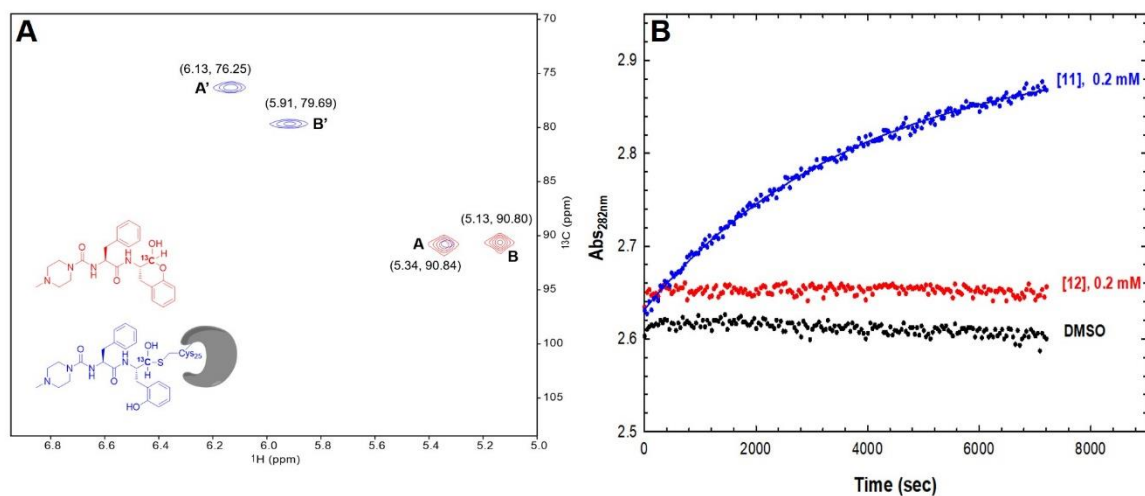
Figure 3.6 Mechanistic study of cruzain inhibition by SMAI. (A) Expansion of
the superposed ^{1}H-^{13}C HSQC NMR spectra of ^{13}C-labeled compound **12** with (blue) and
without (red) an approximately equimolar concentration of cruzain, which were obtained
at 800 MHz (^{1}H) at 25 °C. (B) Time-course of phenylhydrazone formation by treating 0.2
mM **11** (aldehyde) or **12** (lactol) with 1 mM phenylhydrazine, with a control sample
containing DMSO. Data for the curve with compound **11** were fitted to $OD_{282nm} = a[1 - \exp(-k_{obs}t)] + C$, from which $a = 0.279$ OD_{282nm}, $k_{obs} = (2.61 \pm 0.05) \times 10^{-4}$ s[-1], and $C = 2.632$ s[-1], while fitting the other data to this expression led to negligible values of k_{obs}.

1D NMR of this compound in organic solvent ($CDCl_3$ and CD_3OD) demonstrated

that the ^{13}C-labeled carbon was virtually 100% in the δ-lactol form, as no aldehyde

proton peak was observed, with its very characteristic chemical shift around 9.0 – 10.5

ppm.[211] We then analyzed 0.4 mM of ^{13}C-labeled **12** in phosphate buffer (pH 7.5) by

HSQC NMR. No discernable peak was present at $\delta(^{13}C) > 180$ ppm, ruling out the

existence of minute concentrations of free aldehyde in this aqueous solution (See

Appendix H for full spectrum). The two salient peaks, A and B, occurring near $\delta(^{13}C)$

90.8 ppm, were consistent with the signal of a carbon occurring in a hemiacetal (**Figure**

3.6A, red). These signals could not be assigned as an aldehyde hydrate in view of the fact

99

that analysis by LC-MS displayed a molecular ion peak corresponding to the hemiacetal but not the hydrate (**Appendix C**). We propose that peaks A and B are associated with two δ-lactol anomers spontaneously generated during lactol formation. The ratio of anomer A to anomer B is 1:1.32 based on peak volumes (**Appendix D**), yet the exact stereochemistry cannot be assigned at this point.

To the inhibitor sample was added an approximately stoichiometric amount of cruzain, and a new spectrum was acquired after 1 h (**Figure 3.6A**, blue). The apparent hemiacetal peak B was eliminated while a trace of peak A remained. Since the original peak B was 32% greater than peak A, this result suggested a slight preference of cruzain for anomer B over A. The ^{13}C signals of peaks A and B were shifted to higher field occurring at 76.25 ppm (A') and 79.69 ppm (B') upon the addition of cruzain, which was a strong indication of hemithioacetal formation. Such upfield shifts were described elsewhere, and attributed to a less efficient deshielding effect by a sulfur atom than by an oxygen atom.[212] The broadening of peaks A' and B' also suggested a protein-bound ligand, due to reduced tumbling and the increased relaxation time of the cruzain-bound ^{13}C-labeled **12**.

As phenylhydrazine readily forms phenylhydrazone adducts with aldehydes, we treated 0.2 mM concentrations of aldehyde **11** and SMAI **12** with 1 mM phenylhydrazine in the same buffer (lacking DTT) used in the NMR study (**Figure 3.6B**) to determine the fraction of free aldehyde found in **12**. While aldehyde **11** was rapidly, and apparently, completely converted to phenylhydrazone, **12** remained intact after 2 hours. This finding showed that even an excess amount of phenylhydrazine cannot drive the equilibrium of **12** towards the formation of open-form aldehyde, which not only corroborated the

chemical stability of lactol, but also indicated that the ring-opening is likely an enzyme-catalyzed process. Therefore, we conclude that SMAI **2** predominantly maintains its lactol form in aqueous solution in the absence of cruzain, while its binding to cruzain apparently promotes opening of its ring followed by the formation of a covalent bond with Cys_{25}.

3.2.6 Selectivity of SMAIs for Cruzain over Homologous Human Cathepsins

Whether or not our cruzain inhibitors will need to be highly selective over mammalian cysteine proteases as discussed in Section 1.3.1, selectivity for the intended enzyme target always eliminates concerns about toxicity. Accordingly, SMAI were tested as potential inhibitors of human CatB and CatL at lysosomal pH (5.5). Generally, the inhibition constants of representative compounds for human CatB are above 1 µM except for compound **10**, which exerts considerable inhibition for human CatB at submicromolar concentration. Because the substrate specificity of cruzain is very similar to that of human CatL, these inhibitors might be expected to effectively inhibit human CatL, and, indeed, where studied, SMAIs were typically over two-fold more potent for human cathepsin L than for cruzain. Considering the apparent abundance of CatL in mammalian tissues and cells,[139] the *in vivo* toxicity of these inhibitors may be negligible, although this needs to be demonstrated.

3.2.7 Effects of SMAIs in Trypanosomes

Given the anti-cruzain potency of these SMAIs, we continued to assess their trypanocidal activity in axenic cultures of *T. b. brucei* in both PCFs and BSFs (**Table**

Table 3.3 Effects of SMAIs in trypanosomes

Compound	Cruzain Inhibition K_i* (nM)	T. b. brucei PCFs EC$_{50}$ (µM)[a]	T. b. brucei BSFs EC$_{50}$ (µM)[a]	T. cruzi -infected cardiomyoblasts (C2C12) EC$_{50}$ (µM)[a]
1	0.44 ± 0.02	8 ± 2	3.3 ± 2.1	-
2	49 ± 2	17 ± 10	6.8 ± 1.1	5.4 ± 0.2
6	350 ± 30	11 ± 1	2.6 ± 1.1	-
7	103 ± 5	7 ± 2	0.5 ± 0.2	-
8	74 ± 10	10 ± 2	5.8 ± 3.4	-
9	48 ± 2	11 ± 0.4	2.7 ± 0.2	3.7 ± 0.5
10	18 ± 0.5	6.7 ± 0.5	4.0 ± 0.2	-
12	47 ± 2	4 ± 0.2	0.6 ± 0.1	0.5 ± 0.2
13[b]	~1,000	-	3.7 ± 0.2	-
14[b]	-	-	3.1 ± 0.2	-
15[b]	-	-	4.0 ± 0.2	-
16[b]	-	-	12 ± 0.1	-
17[b]	-	-	14 ± 0.9	-
K11777	0.2[c]	1.7 ± 0.5	0.09 ± 0.06	0.7 ± 0.2

[a]Trypanocidal activities of compounds in axenic cultures of *T. b. brucei* PCFs/BSFs and in cardiomyoblasts with *T. cruzi* infection were all measured as EC$_{50}$ value which is the half maximal effective concentration; [b]Structures of **13-17** are shown in Figure 3.7B; [c]Reported as apparent IC$_{50}$ in reference 158.

3.3). For the PCFs, these inhibitors were all effective in eliminating parasites at a low micromolar concentration. The least potent compound **2** had an EC$_{50}$ of 17 ± 10 µM, while the most potent antitrypanosomal compound **12** exhibited an EC$_{50}$ of 4.0 ± 0.2 µM. While the "benchmark" anti-trypanosomal inactivator **K11777** exhibited a slightly lower value of EC$_{50}$ of 1.7 ± 0.5 µM, SMAI **12** is only two-fold less potent than this inactivator. For the BSFs of *T. b. brucei* (**Appendix E**), SMAIs generally showed higher trypanocidal activity than for PCFs which is encouraging as BSF is the disease-related form. The most active inhibitors for BSFs, **7** and **12**, exhibited respective EC$_{50}$ values of 0.5 and 0.6 µM that outperformed the anti-trypanosomal drug, diminazene (Berenil®, EC$_{50}$ = 0.99 ± 0.12 µM in *T. b. brucei* BSFs).[213] Distinct from the low potency of **6** vs. cruzain, it showed comparable trypanocidal activity in *T. b. brucei* with other SMAIs, which is indicative of either a higher cell membrane permeability of **6** owing to its methoxy group, or a subtle

difference of binding sites between cruzain and the target inside *T. b. brucei*. Notably, the free aldehyde **1** was equally or less potent in killing both forms of axenic *T. b. brucei* compared to our SMAIs. Considering that aldehyde **1** is nearly four orders of magnitude less effective vs. *T. b. brucei* than vs. purified cruzain ($EC_{50}/K_i^* = 7,500$), while the average value of EC_{50}/K_i^* for all SMAIs is 74, one may speculate that the SMAIs are more accessible to its cellular target(s) than free aldehyde **1** in cell culture media. In this sense, masking of the active aldehyde within the SMAIs affords significant protection of its electrophilic group.

Additionally, compounds **2**, **9** and **12** were analyzed in a murine cardiomyoblast model of *T. cruzi* infection, in which the disease-relevant amastigote forms of the parasite were evaluated (**Table 3.3**, **Appendix E**). Compounds **2** and **9** killed *T. cruzi* at respective EC_{50} values of 5.4 and 3.7 µM, while **12** exhibited an EC_{50} of 0.5 ± 0.2 µM, which was superior to the benchmark inhibitor **K11777** ($EC_{50} = 0.7 \pm 0.2$ µM). These SMAIs were essentially equipotent in axenic cultures of *T. b. brucei* BSFs and *T. cruzi*-infected cardiomyoblasts, suggesting they target a cysteine protease homologue of cruzain in *T. b. brucei*. Additionally, they exerted no apparent toxicity against the host cardiomyoblasts at up to 10 µM where **K11777** exhibited cytotoxicity. Accordingly, the self-masking of the aldehyde group in SMAIs provides the apparent delivery of these otherwise reactive compounds to trypanosomes harbored within mammalian cardiomyoblasts, with no apparent untoward effects on the host cells. These results encouraged progressing the SMAIs to pre-clinical analysis, as well as the development of prodrug forms of SMAIs.

3.2.8 Design and Evaluation of O-derivatized SMAIs as Prodrugs

The pharmacokinetic evaluation of compound **12** in mice by intravenous administration indicated a short half-life arising from apparent first-pass metabolism (data not shown). Its analogue **K11777** has superior pharmacokinetic properties to **12**, and it was shown that the sites of oxidative metabolism on **K11777** were largely confined to the *N*-methylpiperazine ring, common to both inhibitors, and the homophenylalanyl sidechain.[214-215] Therefore, metabolism of the cyclic hemiacetal of **12** may be responsible for its rapid clearance in mice.

Pralnacasan (VX-740) is a cyclic α-keto-acetal prodrug for caspase-1, allowing for selective intervention in the pro-inflammatory cytokine cascade.[216] This prodrug can be regarded as a SMAI, as it is rapidly converted to the active aldehyde inhibitor (VRT-18858) by plasma esterases via formation of a hemiacetal intermediate (**Figure 3.7A**).[217] Pralnacasan has a greatly improved oral bioavailability (F = 50%) compared to VRT-18858 (F = 4%). In kind, we designed two types of derivatization on the hemiacetal hydroxyl group, i.e., *O*-acylation and *O*-alkylation (**Figure 3.7B**). The *O*-acylated compounds should be hydrolyzed to release compound **12** by the action of cellular esterases, which are ubiquitous in ER lumen of mammalian tissues. The *O*-alkylated compounds structurally resemble pralnacasan, yet their conversion to active inhibitors are
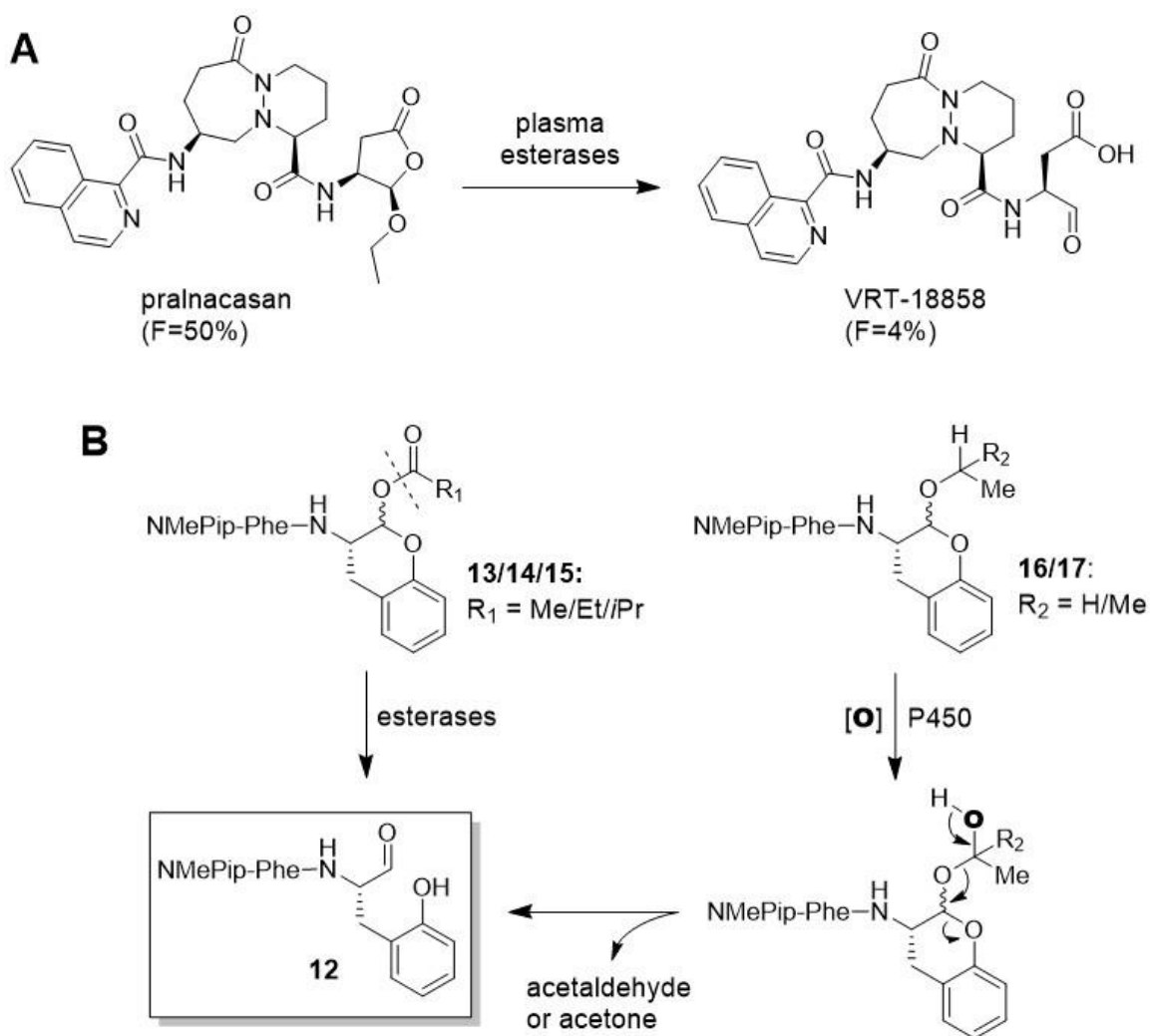
Figure 3.7 Rationale of SMAI prodrugs. (A) Pralnacasan is a SMAI prodrug for VRT-18858, an inhibitor for caspase-1. (B) Proposed metabolic routes of *O*-derivatized SMAIs to compound **12**.

different. From *in silico* prediction by SMARTCyp,[218] hydroxylation of the carbon marked in the scheme may occur with certain liver P450 enzymes, forming a new hemiacetal that subsequently collapses to form compound **12**.

To explore this prodrug approach, we prepared *O*-acylated compounds **13 – 15** and *O*-alkylated compounds **16** and **17** (**Figure 3.7B**) These compounds exhibited negligible inhibition against cruzain, with the exception of **13**, which weakly inhibited cruzain in a time-dependent fashion ($K_i* = 1$ µM, **Appendix F**). Compounds **13 – 15**
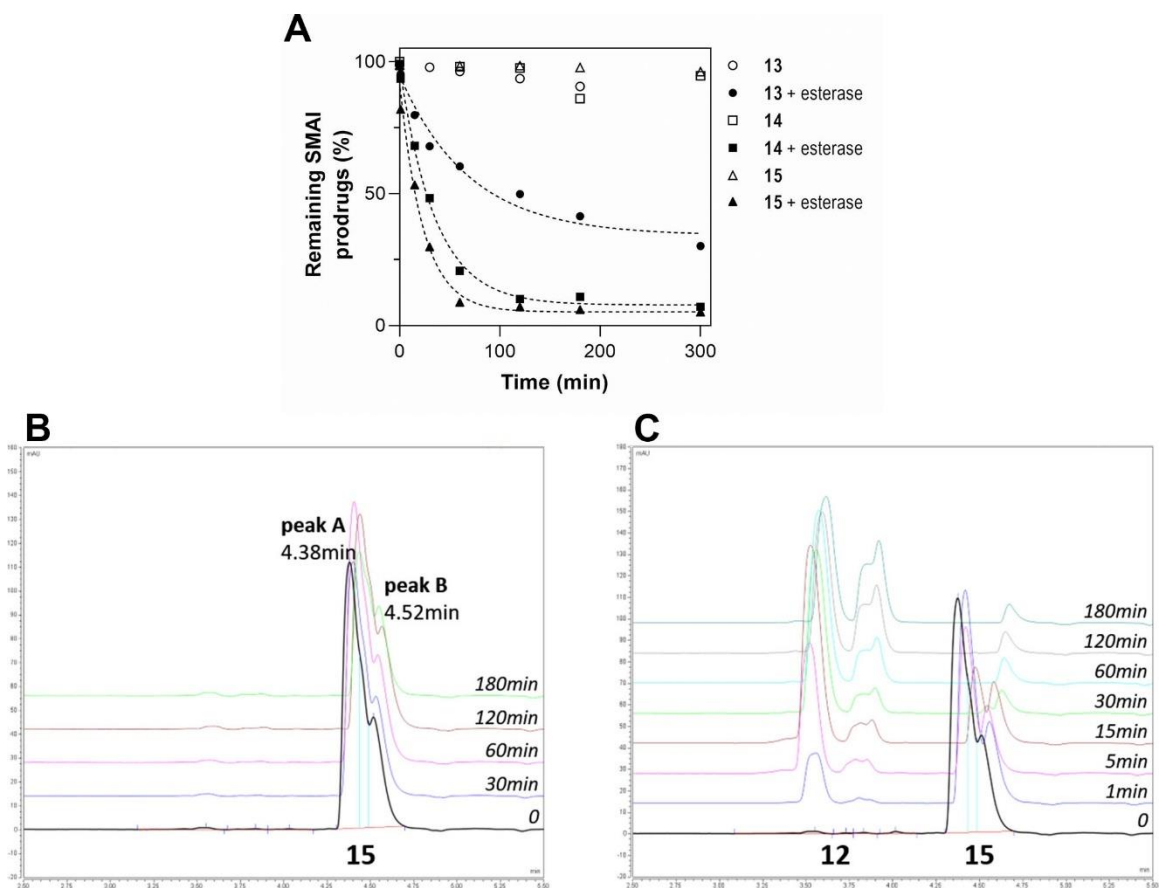
Figure 3.8 Hydrolysis of *O*-acylated SMAIs by esterase. (A) Time-course of remaining compounds **13** – **15** (structures shown in **Figure 3.7B**) in reactions with or without addition of esterase. (B) HPLC traces of **15** in buffer (control). (B) HPLC traces of **15** treated with esterase.

were treated with porcine esterase in buffer (pH 7.5), and the hydrolysis of these prodrugs were monitored by LC-MS (**Figure 3.8A**). In control samples without esterase, these compounds largely remained intact. Compound **13** was an exception and degraded about 10% in buffer over 3 h, which likely explains the observed inhibition of cruzain by **13** due to the formation of **12**. Upon addition of the esterase, compounds **13** – **15** were all converted to compound **12** at variable rates of reaction with respective half-lives of 48, 24, and 16 min, and displayed a trend of increased hydrolysis with the increasing steric bulk of the acyl groups. A reasonable interpretation is that the main mammalian esterase, carboxylesterase-1 (CES1), has a preference for larger acyl groups.[219-220] This is also

106

likely the reason for incomplete conversion of **13** to **12**, as it may not be the preferred

substrate for CES1. Accordingly, modifying the acyl group of *O*-acylated compounds to

obtain a suitable half-life can potentially overcome the first-pass metabolism of **12**. As

observed for compound **15** (**Figure 3.8B, C**), there are two peaks in the chromatograph

that presumably correspond to the two anomers, but the conversion of one peak (A, t_R =

4.38 min) to **12** is faster than the other (B, t_R = 4.52 min). Peak A is largely eliminated

after 30 min, while peak B, though it constitutes a smaller proportion of untreated

compound, does not decompose completely even after 3 h, suggesting that the anomer in

peak A is the more specific substrate for esterase. These results provide proof of concept

that use of *O*-acylated prodrugs of SMAIs will provide a means to deliver these inhibitors

in vivo.

Compound **12** was also modified to yield two mixed acetals **16** and **17** (**Figure

3.7B**) that did not inhibit cruzain. Because cytochrome P450 (CYP) forms 3A4, 2D6 and

2C9 together account for over 60% of drug-metabolizing P450 isoforms,[221] we used these

enzymes to conduct *in vitro* assays to determine if **16** and **17** could be transformed to

compound **12** as we expected. Both compounds were incubated with different CYPs in

the presence of a NADPH-regenerating system for up to 4 h. Unfortunately, no

transformation to **12** was observed for either compound by any of the CYPs. However,

the apparent stability vs. these purified P450s does not mean that (an)other microsomal

oxidase(s) may not release the active SMAI from these prodrug forms. *In vivo* analysis of

these prodrugs may as yet point to their value as prodrugs.

These compounds were also tested in axenic cultures of *T. b. brucei* BSFs (**Table

3.3, Appendix E**). Although less active than their parent compound **12**, compounds **13** –

15 showed promising trypanocidal activity comparable to other SMAIs. These encouraging results demonstrated that the *O*-acylated prodrug forms could well be converted to **12,** presumably by a parasitic enzyme with esterase functionality. Interestingly, compounds **16** and **17** were able to kill the parasites in spite of lower activity ($K_i^* > 5$ µM). Since treatment with P450s did not lead to transformation of these prodrugs, the data implied the existence of either a drug-metabolizing enzyme or another potential activating enzyme in *T. b. brucei*.

3.2.9 Application of SMAI Strategy to Inhibitor Design for SARS-CoV-2 3CL^pro

The SMAI strategy is not limited to cruzain, but it may also find application in other protease categories like cathepsins as long as the S_1 binding pocket can accommodate an aromatic ring or alkyl groups of similar size. As mentioned in the introductory section, SARS-CoV-2 main protease 3CL^pro is a potential drug target for this approach. Numerous inhibitors for homologous cysteine proteases in SARS-CoV, MERS-CoV, and some picornaviruses have shown different levels of inhibition against SARS-CoV-2. Peptidomimetic aldehyde inhibitors of 3CL^pro contain a nearly-invariant 2-oxo-pyrrolidin-2-yl group (γ-lactam) as the P_1 side chain (**Table 3.4**, header). We propose that 2-pyridone can act as a surrogate for the γ-lactam. Apart from their similarity in size and heteroatom substitution, the tautomerization of 2-pyridone towards 2-hydroxypyridine has been well characterized. Although aqueous solution tends to favor the 2-pyridone, the presence of the *C*-terminal aldehyde could well form a SMAI with the 2-hydroxypyridine.

Table 3.4 Inhibition data of potential SMAIs for SARS-CoV-2 3CL^{pro} and human cathepsin L.
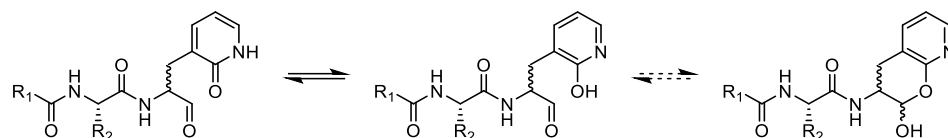
Compound	Structure R_1	Structure R_2	Overall inhibition constant K_i^* (nM)[a] SARS-CoV-2 3CL^{pro}	Overall inhibition constant K_i^* (nM)[a] human cathepsin L
18			860 ± 90	2.3 ± 0.1
19[b]	Cbz		9 ± 2	53 ± 13
20			61 ± 16	141 ± 30
21			261 ± 40	ND
22			187 ± 30	39 ± 3
23			57 ± 11	45 ± 9
24			402 ± 21	ND

[a]3CL^{pro} was assayed at pH 7.5 while hCatL was assayed at pH 5.5; [b]Reported K_i^* of **19** for 3CL^{pro} is an average number of K_i^* values obtained using eq. 3-1 with a floated or a fixed concentration of enzyme.

We have synthesized several 2-pyrdione compounds **18 – 24** with different peptidomimetic scaffolds as listed in **Table 3.4**. Their potency for 3CL^{pro} ranged from submicromolar to nanomolar values of K_i^*. From the perspective of peptidomimetic scaffolds in the P$_3$ position, a 2-indolyl group was not as effective as Ac-Val (**20** vs. **23**, or **21** vs. **24**); for the P$_2$ position, there was a clear preference for a cyclohexylalanyl (Cha) sidechain compared to a leucyl and 4-nitrophenylalanyl (4-nitroPhe) sidechain, i.e.

109
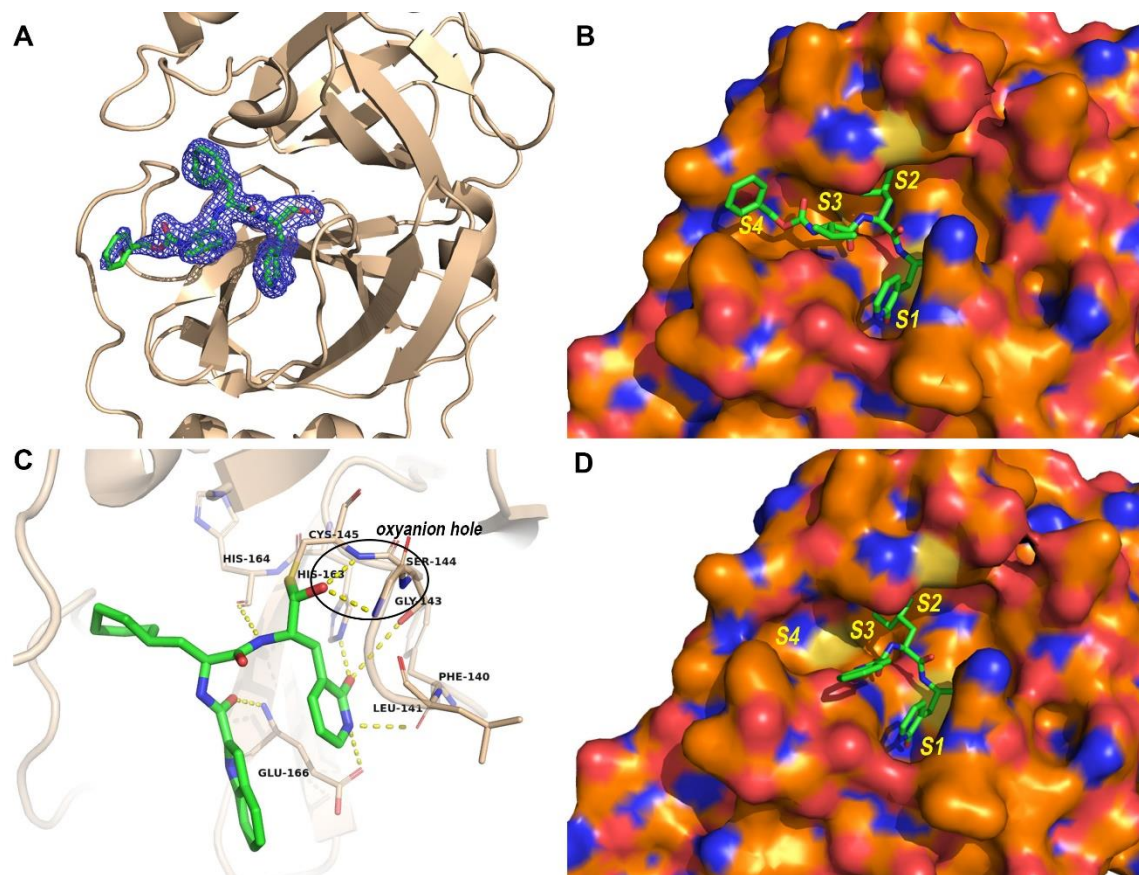
Figure 3.9 Crystal structures of SARS-CoV-2 3CL^pro complexed with **19** and **23**. (A) Ribbon representation of 3CL^pro-**19** complex with $2F_o\text{-}F_c$ electron density map contoured at 0.7σ. (B) Surface representation of 3CL^pro-**19** complex. (C) Interactions between 3CL^pro and **23**. Dashed lines depict hydrogen bonds; oxyanion hole is circled. (D) Surface representation of 3CL^pro-**23** complex.

Cha > Leu > 4-nitroPhe (**19**, **20**, **23** > **22** > **21**, **24**); and the nature of the P_4 group effects important contributions to inhibitor binding, as **19**, the most potent inhibitor in our survey, was superior to **20**. Compound **19** slightly differs from **20** in that the former has a Cbz group compared to the *N*-acetyl group of **20** at P_4 position. Proton NMR of **19** in organic solvent (10% CD_3OD in $CDCl_3$) confirmed the apparent absence of aldehyde, suggesting it forms a δ-lactol, though it is unclear what the exact species is in aqueous buffers. Compound **19** was such a potent inhibitor of 3CL^pro that its K_i^* (9 nM) was a result of apparent titration of the enzyme (40 nM used in the assay). Like other 2-

pyridone inhibitors, compound **19** also inhibited human CatL, which was recently shown to be essential to the penetrance of SARS-CoV-2 into mammalian cells, in that it catalyzes essential cleavage of the coronaviral spike protein.[222] As a result, compound **19** has the potential to be a dual-acting inhibitor for two enzymes which are critical to the infection of human cells by SARS-CoV-2.

We obtained high-resolution (1.70 Å) crystal structures of 3CL[pro] in complex with **19** or **23** (**Appendix G**). In both structures, the well-defined electron density (**Figure 3.9A, C**) near the active site confirmed the formation of a hemithioacetal with active-site Cys_{145}, and the resulting hydroxyl group of the hemithioacetal is stabilized by an oxyanion hole provided by Gly_{143} and Cys_{145} . The P_1 2-pyridone establishes essential interactions with 3CL[pro] (**Figure 3.9C**), comparable to those inhibitors that have a 2-oxopyrrolidine moiety at the P_1 side chain:[223] the carbonyl oxygen of the 2-pyridone accepts two hydrogen bonds from His_{163} and Ser_{144}, while the amide nitrogen acts as a hydrogen bond donor to Glu_{166} and Phe_{140}. The Cha group at P_2 inserted cozily in a deeply-buried S_2 pocket where no water is present, indicating Cha contributes to binding affinity via a likely hydrophobic effect. Unlike the Cbz-Val moiety of **19** that was well situated in the S_4-S_3 binding cleft (**Figure 3.9B**), the 2-indolyl group of **23** partially swayed outward to the solvent (**Figure 3.9D**) which was in agreement with their difference in potency. These crystal structures will guide us in the optimization of 2-pyridone inhibitors and, as with cruzain, 3CL[pro] is apparently capable of catalyzing ring-opening of the putative masked aldehyde.
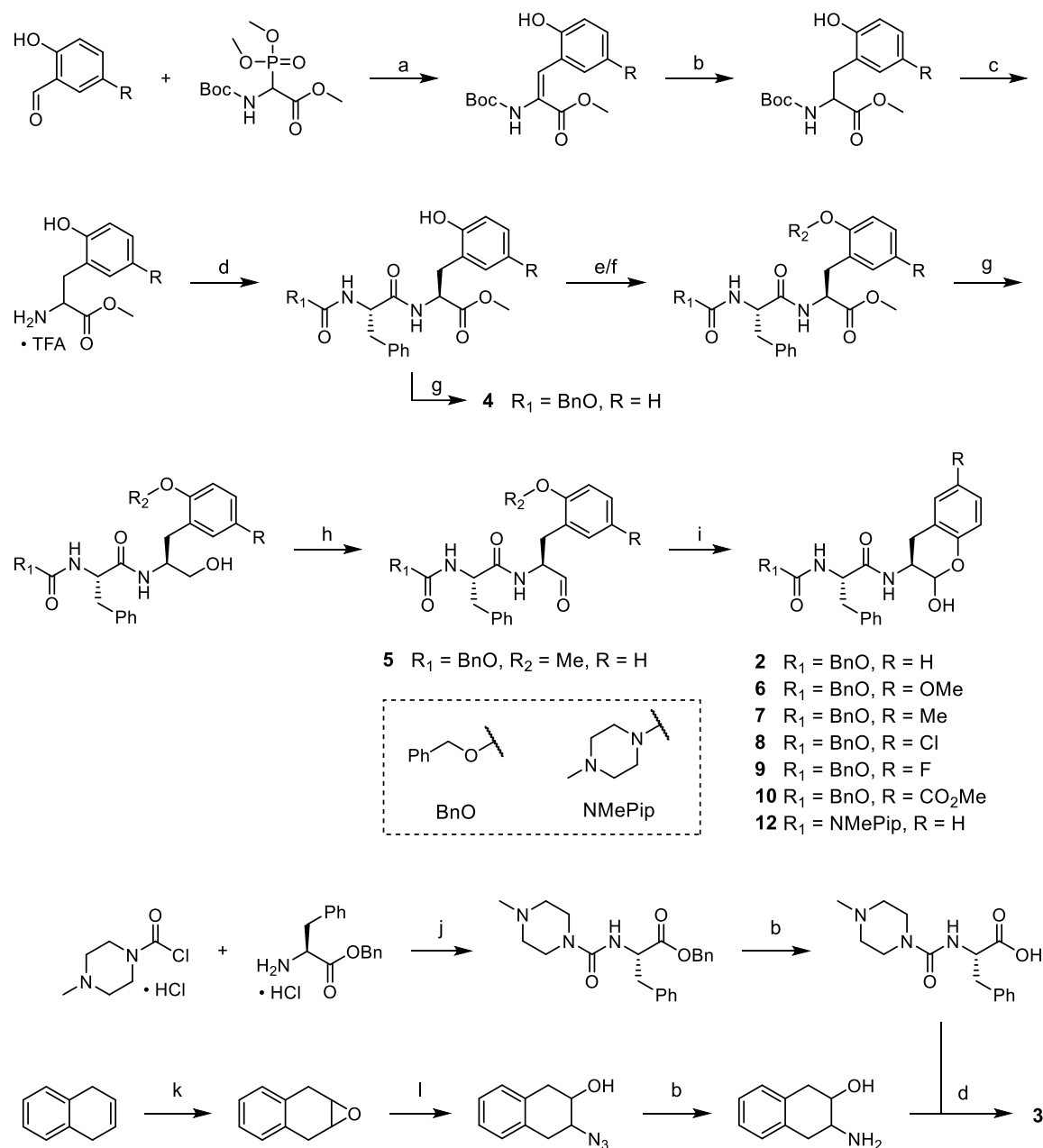
Scheme 3.1 Synthesis of SMAIs and their analogs. (a) DBU, DCM. (b) Pd/C, H$_2$, MeOH. (c) TFA, DCM. (d) Cbz-Phe-OH or NMePip-Phe-OH or 2-hydroxy-3-aminotetralin, DIPEA, T3P, DCM. (e) TBSCl, imidazole, DCM. (f) K$_2$CO$_3$, CH$_3$I, DMF. (g) NaBH$_4$, MeOH. (h) Dess-Martin periodinane, NaHCO$_3$, DCM, 0 ^{o}C. (i) TBAF, THF, 0 ^{o}C. (j) Et$_3$N, THF, 0 ^{o}C. (k) *m*CPBA, chloroform. (l) NaN$_3$, MeOH/H$_2$O, 60 ^{o}C.

3.2.10 Synthesis of SMAIs

Scheme 3.1 delineated the synthesis of common SMAIs and their analogs starting

from the HWE coupling reaction between salicylic aldehydes bearing different 5-

112

substituents and (±)-Boc-α-phosphonoglycine trimethyl ester under basic conditions (step a). The double bond of product is reduced by Pd/C-catalyzed hydrogenation (step b), followed by removal of the Boc protecting group to give the key intermediate (step c), a substituted *ortho*-tyrosine methyl ester. This intermediate was coupled to different peptidomimetic scaffolds using T3P, yielding the methyl ester of the "full" inhibitor (step d). At this stage, the phenol group was protected as it was relatively acidic, and might affect subsequent reactions. The *tert*-butyldimethylsilyl (TBS) group was selected as the protecting group (step e) due to its stability and mild deprotection condition. Methoxy group was used for making compound **5** but not for protection as its removal was inefficient and demanded harsh conditions (step f). The TBS protection proceeded smoothly with the use of imidazole as base and catalyst. The protected product was treated with sodium borohydride to reduce the methyl ester to a primary alcohol (step g), which was further oxidized using Dess-Martin periodinane to form the aldehyde (step h). Finally, removal of the TBS group by tetra-*n*-butylammonium fluoride (TBAF) rapidly generated the cyclic lactol compound (step i). While the scaffold Cbz-Phe-OH was commercially available, NMePip-Phe-OH was synthesized in lab. The amide coupling product between phenylalanine benzyl ester with 4-methylpiperazine-1-carbonyl chloride (step j) underwent Pd/C-catalyzed hydrogenolysis to produce the NMePip-Phe-OH. This acid was also used to prepare the "locked" cyclic compound **3**. The route to its P_1 moiety started with the epoxidation of 1,4-dihydronaphthalene using *m*CPBA (step k). The

epoxide ring was opened by nucleophilic attack of sodium azide (step l) which was

converted to free amine via catalytic hydrogenation.

13	R_1 = Me
14	R_1 = Et
15	R_1 = i-Pr
16	R_2 = Et
17	R_2 = i-Pr
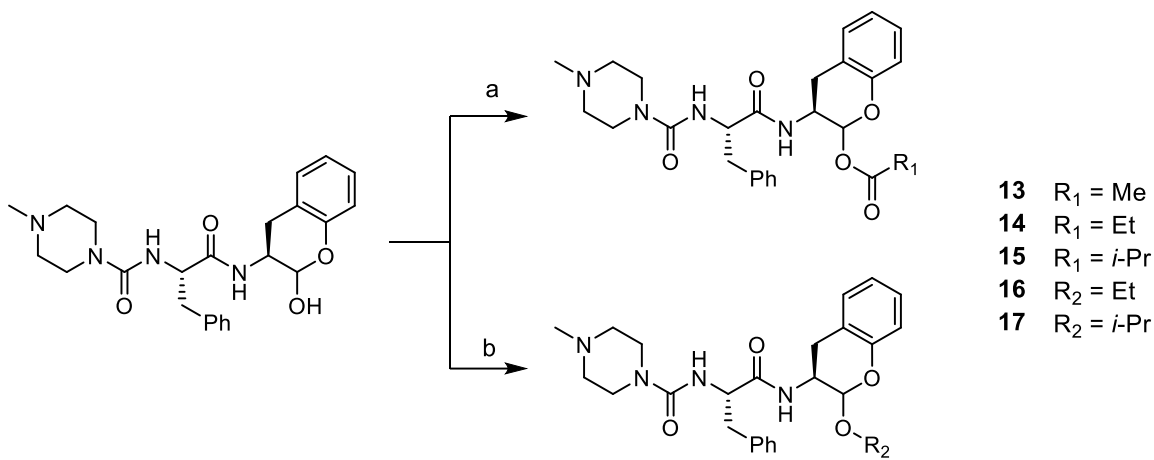
Scheme 3.2 Synthesis of *O*-derivatized SMAIs. (a) acetic/propionic/isobutyric anhydride, Et₃N, DMAP, DCM. (b) BF₃OEt₂, EtOH/*i*PrOH

Derivatization of the hydroxyl group of compound **12** included *O*-acylation and

O-alkylation (**Scheme 3.2**). The *O*-acylation was performed by reacting **12** with the

corresponding acetic/propionic/isobutyric anhydride using catalytic DMAP (step a). The

O-alkylation was carried out by reacting **12** with corresponding ethanol/isopropanol using

boron trifluoride etherate (BF₃OEt₂) as the Lewis acid catalyst (step b). On account of the

anomeric carbon, SMAIs and their derivatives have two diastereomers which exist in

equilibrium and cannot be isolated by flash column chromatography or preparative

HPLC.

It is necessary to adopt a new approach (**Scheme 3.3**) to prepare the ^{13}C-labeled compound **12** as no ^{13}C-labeled starting material for **Scheme 3.1** was commercially available. Since glycine-1-^{13}C is an affordable chemical, it was used to build the P_1 *ortho*-tyrosine employing a nickel(II)-based chiral auxiliary. To synthesize this auxiliary, benzyl-L-proline was converted to a more reactive acyl chloride intermediate using methanesulfonyl chloride (MsCl) which was immediately coupled with 2-aminobenzophenone (step a). The resulting compound (BPB) was reacted with Gly-1-^{13}C and nickel (II) nitrate under a strongly base environment, generating a complex named Ni-BPB-[13]Gly of which the α-carbon of glycine was activated by the coordinated nickel (step b). The complexed glycine thereby formed a bond with the side chain moiety in a
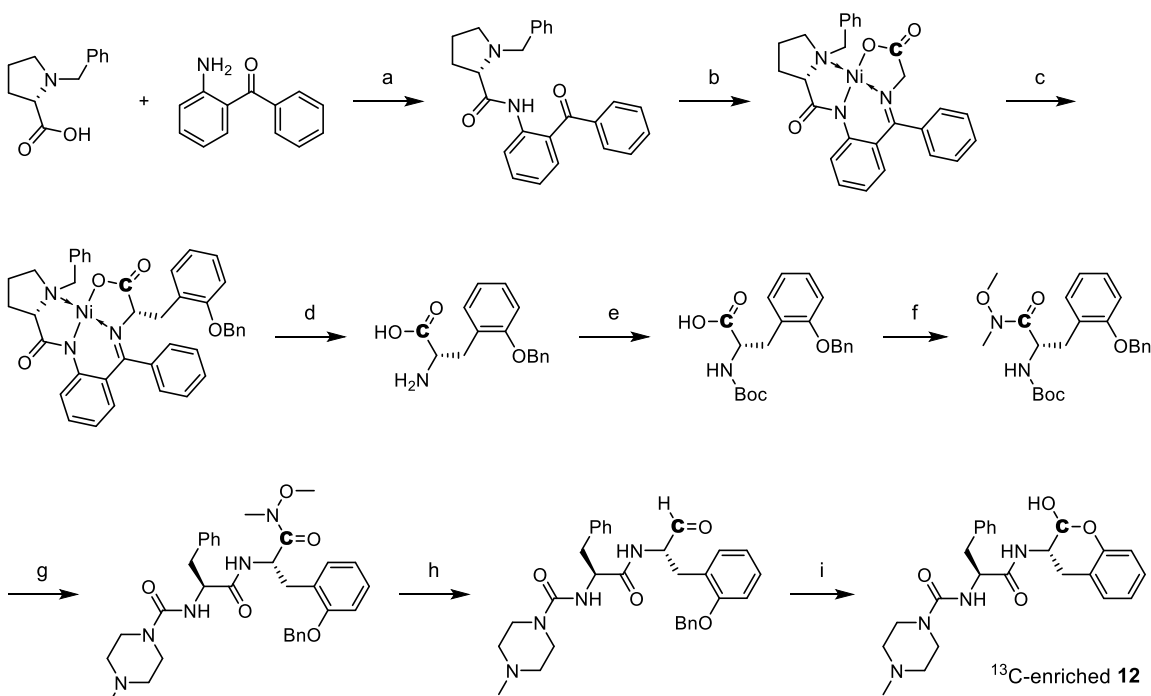
Scheme 3.3 Synthesis of ^{13}C-enriched 12. (a) 1-methylimidazole, MsCl, DCM, 45 ºC. (b) glycine-1-^{13}C, Ni(NO$_3$)$_2$·6H$_2$O, KOH, MeOH, 60 ºC. (c) (2-(benzyloxy)phenyl)methanol, CMBP, toluene, 120 ºC. (d) 8-quinolinol, MeCN/H$_2$O, 40 ºC. (e) Et$_3$N, Boc anhydride, dioxane/H$_2$O. (f) *N,O*-dimethylhydroxylamine, DIPEA, T3P, DCM. (g) TFA, DCM; NMePip-Phe-OH, DIPEA, T3P, DCM. (h) LAH, THF, 0 ºC. (i) Pd/C, H$_2$, MeOH.

Mitsunobu-Tsunoda reaction by use of cyanomethylene tributylphosphorane (CMBP) at elevated temperature (step c).[224] This type of reaction could exclusively generate an *S*-configured amino acid because of the hindrance from the benzyl group of proline. Subsequent decomplexation was performed using 8-quniolinol by extracting Ni(II) from the complex (step d)[225], liberating the unnatural amino acid *o*Tyr(OBn) which was first protected with a Boc group (step e), which was then converted to a Weinreb amide (step f). Next, this compound was deprotected and coupled to the NMePip-Phe-OH (step g). The Weinreb amide was reduced to aldehyde with an equivalent amount of LiAlH$_4$ (step h). The benzyl protecting group was eventually cleaved off by Pd/C hydrogenolysis to yield the ^{13}C-labeled product (step i).

The preparation of 2-pyridone-based inhibitors for 3CL^{pro} began with 2-oxo-1,2-dihydropyridine-3-carbaldehyde instead of salicylaldehyde (**Scheme 3.4**). Because the 2-pyridone was fixed as P$_1$ moiety, its Weinreb amide was prepared on a relatively large
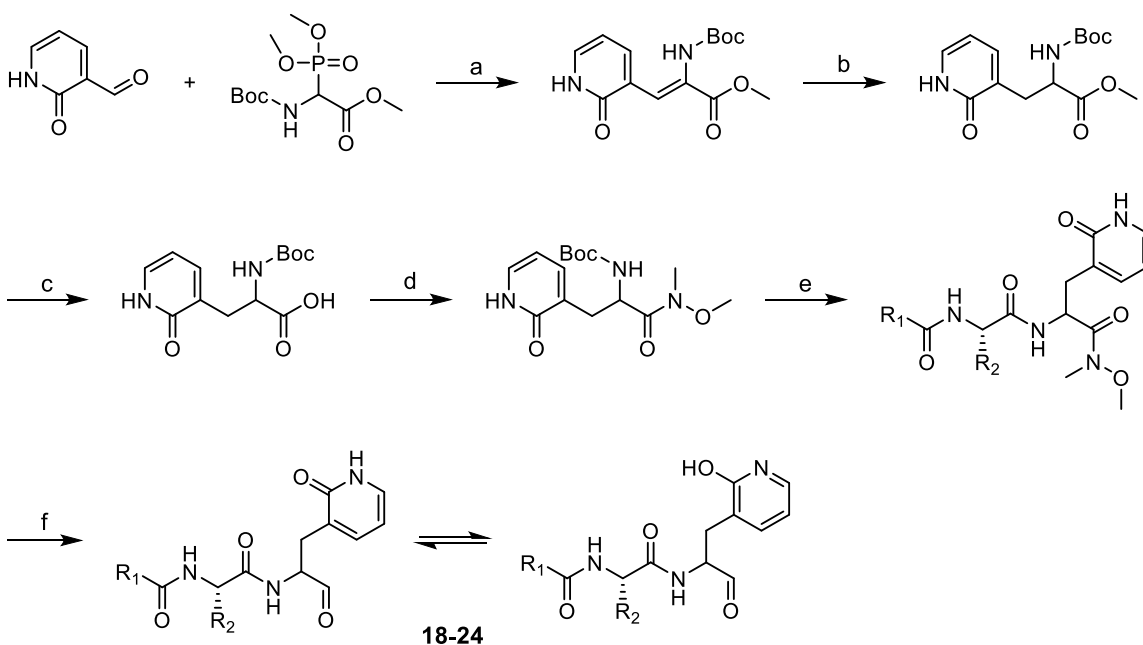
Scheme 3.4 Synthesis of 2-pyridone-based SMAIs. (a) DBU, DCM. (b) Pd/C, H$_2$, MeOH. (c) LiOH, MeOH/H$_2$O. (d) *N,O*-dimethylhydroxylamine, DIPEA, T3P, DCM. (e) TFA, DCM; NMePip-Phe-OH, DIPEA, T3P, DCM. (f) LAH, THF, 0 °C.

scale prior to coupling to various scaffolds, allowing for an economical synthesis. To this end, after the HWE reaction (step a) and olefin hydrogenation (step b), the methyl ester was hydrolyzed under basic conditions (step c), and following reactions (step d-f) were similar to **Scheme 3.3**. Notably, this route does not require protection of the 2-hydroxyl group probably because 2-pyridone is the predominant form in reaction. In fact, we attempted to protect it with a TBS or benzyl group, but neither was successful.

3.3 Materials and Methods

3.3.1 General Information of Synthetic Chemistry

Unless otherwise noted, all starting materials, reagents and solvents were obtained commercially and used without further purification and distillation. Reactions were conducted under an inert atmosphere (argon or nitrogen gas), and monitored by TLC or HPLC-MS. TLC experiments were performed with silica gel plates on aluminum foil (Sigma, 60778) and visualized under UV light (254 and 365 nm) or using specific stains (ninhydrin for oTyr(OBn), $KMnO_4$ for UV-insensitive compounds, 2,4-DNP for various aldehydes, etc.). HPLC analysis was implemented by an UltiMate 3000 HPLC system equipped with a diode array detector and was coupled to MS analysis by an ISQ™ EM single quadrupole mass spectrometer utilizing positive or negative electrospray ionization. The typical settings of HPLC-MS were as follows. Column: Phenomenex Luna 5 μm C18(2) 100 Å, 4.6 mm × 50 mm; mobile phase A: water containing 0.1% formic acid (v/v); mobile phase B: MeCN containing 0.1% formic acid (v/v); temperature: 25 °C; elution: pre-equilibration of column for 2 mins at 10% B, gradient

elution at 10%-100% B over 6 mins, then isocratic elution at 100% B for 2 mins; flow

rate: 1 mL/min; UV detector: four channels at 254, 280, 214, and 350 nm; MS

parameters: HESI, scan range m/z 100-1000, vaporizer temperature 350 ^{o}C, ion transfer

tube temperature 300 ^{o}C, source CID voltage 20 V. Most compounds were purified by

flash column chromatography (FCC) on silica gel (200-300 mesh) with different solvent

systems. Some compounds were purified by semi-preparative HPLC (Prep-HPLC) on the

same UltiMate 3000 HPLC system which was connected to a single wavelength detector

and a fraction collector sequentially. The typical settings of Prep-HPLC were as follows:

Column: Phenomenex Luna 5 μm C18(2) 100 Å, 21.2 mm × 250 mm; mobile phase A:

water containing 0.1% formic acid (v/v); mobile phase B: MeCN containing 0.1% formic

acid (v/ v); elution method: pre-equilibration of column for 5 mins at 10% B, gradient

elution at 10%-100% B over 25 mins, then isocratic elution at 100% B for 5 mins; flow

rate: 21.2 mL/min; UV detector: single channel, typically at 254 nm. ^{1}H and ^{13}C NMR

spectra of compounds in CDCl$_3$, CD$_3$OD, or (CD$_3$)$_2$SO were recorded on a Bruker

AVANCE III 400 MHz using tetramethylsilane (TMS, 0.00 ppm) or residual solvent

(CDCl$_3$, 7.26 ppm; CD$_3$OD, 3.31 ppm; (CD$_3$)$_2$SO, 2.50 ppm) as the internal standard.

Spectral processing and analysis were performed with MestreNova.

3.3.2 Synthetic Procedures and Compound Characterization

Detailed procedures and characterization of compounds for all reactions in

Scheme 3.1-**3.4** were as follows. Some of them were general procedures which were

described by representative compounds.

*HWE reaction of substituted salicylaldehyde with (±)-Boc-α-phosphonoglycine trimethyl ester (**Scheme 3.1**, a).* To a solution of (±)-Boc-α-phosphonoglycine trimethyl ester (4.22 g, 14.2 mmol, 1.2 eq) in DCM (20 mL) was added DBU (2.12 mL, 14.2 mmol, 1.2 eq) dropwise at -10 °C and the resulting mixture was stirred for 20 min. Then 4-methylsalicylaldehyde (1.61 g, 11.8 mmol, 1.0 eq) in DCM (10 mL) was slowly added to the mixture over 10 min. The reaction was stirred at room temperature overnight. The mixture was concentrated under reduced pressure, diluted with EtOAc (100 mL), and washed successively with saturated aqueous NH_4Cl (20 mL), saturated aqueous $NaHCO_3$ (20 mL) and brine (20 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (30-50% EtOAc in hexane, v/v) to give methyl (S/E)-2-((*tert*-butoxycarbonyl)amino)-3-(2-hydroxy-5-methylphenyl)acrylate (2.91g, 80%). ^{1}H NMR (400 MHz, Methanol-d_4) δ 1.45 (s, 9H), 2.25 (s, 3H), 3.33 (q, J = 1.7 Hz, 1H), 3.83 (s, 3H), 6.77 (d, J = 8.3 Hz, 1H), 7.02 (dd, J = 2.2, 8.4 Hz, 1H), 7.41 (d, J = 2.2 Hz, 1H), 7.53 (s, 1H). ^{13}C NMR (100 MHz, MeOD) δ 19.22, 27.18, 51.36, 79.97, 115.23, 120.46, 124.58, 128.22, 129.63, 131.02, 153.42, 154.76, 166.73. LC-MS: t_R = 4.93 min; $C_{11}H_{14}NO_3^+$ [M + H - Boc]$^+$, m/z calcd 208.10, found 208.08.

*Catalytic hydrogenation of the double bond (**Scheme 3.1**, b).* Methyl (S/E)-2-((*tert*-butoxycarbonyl)amino)-3-(2-hydroxy-5-methylphenyl)acrylate (1.15 g, 3.74 mmol, 1.0 eq) was placed in a two-necked round bottom and charged with N_2 gas. 10% palladium on carbon powder (Pd/C, 110 mg, cat.) was quickly added to the flask, followed by addition of MeOH (20 mL). The flask was degassed and backfilled with H_2 for three cycles. The reaction mixture was stirred at room temperature overnight with a

119

balloon of H_2 for replenishment. The balloon was removed, and the mixture was filtered under reduced pressure. Notice that the operation should be rapid, and the Pd/C powder must be kept wet to avoid catching fire, and was appropriately disposed of in a water-filled, cap-closed container. The filtrate was concentrated under reduced pressure and purified by FCC (30% EtOAc in hexane, then 6% MeOH in DCM) to give methyl 2-((*tert*-butoxycarbonyl)amino)-3-(2-hydroxy-5-methylphenyl)propanoate (1.1 g, 95%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 0.25 (s, 6H), 1.05 (s, 9H), 1.39 (s, 9H), 2.79 – 3.02 (m, 1H), 3.07 (dd, J = 4.8, 13.5 Hz, 1H), 2.36 (s, 3H), 3.75 (s, 3H), 4.43 (d, J = 6.2 Hz, 1H), 5.54 (d, J = 7.4 Hz, 1H), 6.65 – 6.73 (m, 2H), 6.73 – 6.79 (m, 1H). ^{13}C NMR (100 MHz, CDCl$_3$) δ -4.15, 18.19, 20.57, 25.82, 28.24, 32.84, 51.98, 54.97, 79.50, 113.30, 116.22, 119.28, 127.84, 147.51, 153.95, 155.33, 172.70. LC-MS: t_R = 4.98 min; $C_{16}H_{24}NO_5^+$ [M + H]$^+$, m/z calcd 310.16, found 310.14.

*Removal of the Boc protecting group (**Scheme 3.1**, c).* To a suspension of methyl 2-((*tert*-butoxycarbonyl)amino)-3-(2-hydroxy-5-methylphenyl)propanoate (1.1 g, 3.56 mmol, 1.0 eq) in DCM (6 mL) was added TFA (3 mL) at 0 °C, and the resulting mixture was stirred for 30 min. The mixture was then co-concentrated with toluene three times to remove most of the residual TFA. The obtained crude product 3-(2-hydroxy-5-methylphenyl)-1-methoxy-1-oxopropan-2-aminium trifluoroacetate was used without further purification.

*Amide coupling reaction of P3-P2 acid with P1 amine (**Scheme 3.1**, d).* A mixed suspension of 3-(2-hydroxy-5-methylphenyl)-1-methoxy-1-oxopropan-2-aminium trifluoroacetate (~1.15 g, 3.56 mmol, 1.0 eq), Cbz-Phe-OH (1.07g, 3.56 mmol, 1.0 eq) and DIPEA (2.17 mL, 12.46 mmol, 3.5 eq) in DCM (20 mL) was cooled to 0 °C followed

by dropwise addition of T3P (>50% in MeCN, 3.53 mL, 5.34 mmol, 1.5 eq). The resulting mixture was stirred at room temperature for 2 h. The reaction was concentrated under reduced pressure, diluted with EtOAc (75 mL), and washed successively with 5% citric acid (15 mL), saturated aqueous $NaHCO_3$ (15 mL) and brine (15 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (40% EtOAc in hexane) to give methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-hydroxy-5-methylphenyl)propanoate (1.11 g, 63% for two steps). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.30 (s, 3H), 2.83 – 3.16 (m, 4H), 3.69 (s, 3H), 4.50 (s, 1H), 4.75 (s, 1H), 4.98 – 5.14 (m, 2H), 5.70 (s, 1H), 6.68 (t, *J* = 8.8 Hz, 1H), 6.91 – 7.10 (m, 3H), 7.22 (dd, *J* = 20.7, 26.3 Hz, 7H), 7.33 (s, 3H), 8.08 (s, 1H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 20.66, 33.10, 38.37, 52.47, 53.19, 56.10, 67.24, 117.11, 124.49, 127.01, 127.88, 128.22, 128.32, 128.42, 128.53, 128.61, 129.25, 130.87, 135.98, 136.14, 153.57, 156.27, 171.72, 172.08. LC-MS: t_R = 5.39 min; $C_{28}H_{31}N_2O_6^+$ [M + H]$^+$, m/z calcd 491.22, found 491.19.

*Protection of the phenol with a TBS protecting group (**Scheme 3.1**, e).* A mixed solution of methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-hydroxy-5-methylphenyl)propanoate (990 mg, 2.02 mmol, 1.0 eq), TBSCl (609 mg, 4.04 mmol, 2.0 eq), and imidazole (412 mg, 6.06 mmol, 3.0 eq) was stirred at room temperature overnight. The reaction was quenched by addition of 0.5M HCl (10 mL), and was stirred for another 15 min. The mixture was concentrated under reduced pressure, and was partitioned between EtOAc (50 mL) and water (10 mL). The organic layer was washed with saturated aqueous $NaHCO_3$ (10 mL) and brine (10 mL), and was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product

was purified by FCC (20% EtOAc in hexane) to give methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)propanoate (897 mg, 73%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 0.16 (s, 6H), 0.93 (s, 9H), 2.37 (s, 3H) 2.59 – 2.87 (m, 2H), 2.89 – 3.05 (m, 2H), 3.54 (d, *J* = 4.3 Hz, 3H), 4.36 (d, *J* = 22.6 Hz, 1H), 4.64 (q, *J* = 8.3 Hz, 1H), 4.95 (p, *J* = 12.1 Hz, 2H), 5.32 – 5.55 (m, 1H), 6.56 – 6.65 (m, 1H), 6.83 – 7.02 (m, 3H), 7.05 – 7.25 (m, 9H). ^{13}C NMR (101 MHz, CDCl$_3$) δ -3.50, 18.27, 20.53, 25.81, 32.94, 38.57, 52.20, 52.57, 56.08, 66.96, 119.79, 125.91, 126.90, 127.89, 127.95, 128.02, 128.04, 128.07, 128.14, 128.44, 128.47, 128.57, 128.71, 129.28, 129.35, 130.82, 136.25, 136.47, 152.62, 155.90, 170.76, 171.62, 172.00. LC-MS: t_R = 7.23 min; C$_{34}$H$_{45}$N$_2$O$_6$Si$^+$ [M + H]$^+$, m/z calcd 605.30, found 605.24.

*Protection of the phenol with a methyl group (**Scheme 3.1**, f).* To a solution of methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-hydroxyphenyl)propanoate (172 mg, 0.36 mmol, 1.0 eq) in DMF (2 mL) was successively added K$_2$CO$_3$ (100 mg, 0.72 mmol, 2.0 eq) and iodomethane (67 µL, 1.08 mmol, 3.0 eq) at 0 °C. The resulting mixture was stirred at room temperature for 20 h during which the system should be kept securely sealed to avoid evaporation of iodomethane. The mixture was diluted with EtOAc (50 mL), and washed with water extensively (5 × 10 mL). The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (30% EtOAc in hexane) to give methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-methoxyphenyl)propanoate (167 mg, 95%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.02 (qd, *J* = 5.7, 13.6, 16.9 Hz, 4H), 3.61 (s, 3H), 3.93 (s, 3H),

4.48 (s, 1H), 4.72 (s, 1H), 4.88 – 5.12 (m, 2H), 5.69 (d, J = 8.2 Hz, 1H), 6.64 – 6.88 (m, 2H), 6.92 – 7.06 (m, 3H), 7.08 – 7.18 (m, 4H), 7.24 (d, J = 18.6 Hz, 6H), 7.77 (d, J = 43.8 Hz, 1H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 32.65, 32.96, 38.40, 52.33, 52.40, 53.70, 56.04, 67.11, 115.84, 120.35, 122.73, 126.88, 126.90, 127.88, 128.16, 128.52, 128.72, 129.30, 131.30, 136.14, 154.71, 156.21, 156.23, 171.98, 172.34. LC-MS: t_R = 5.71 min; C$_{28}$H$_{31}$N$_2$O$_6^+$ [M + H]$^+$, m/z calcd 491.22, found 491.33.

*Reduction of the methyl ester to alcohol (**Scheme 3.1**, g).* To a solution of methyl (S)-2-((S)-2-(((benzyloxy)carbonyl)amino)-3-phenylpropanamido)-3-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)propanoate (897 mg, 1.48 mmol, 1.0 eq) in MeOH (10 mL) was added NaBH$_4$ (1.2 g, 31.7 mmol, >20 eq) in multiple portions every 30 min. The reaction was stirred at room temperature for another 2 h, and then quenched by addition of saturated aqueous NH$_4$Cl (10 mL). The mixture was concentrated under reduced pressure, and diluted with EtOAc (50 mL). The organic layer was washed with saturated aqueous NaHCO$_3$ (10 mL) and brine (10 mL), and was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (35% EtOAc in hexane) to give benzyl ((S)-1-(((S)-1-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)-3-hydroxypropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (450 mg, 53%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 0.26 (s, 6H), 1.03 (s, 9H), 2.40 (s, 3H), 2.74 (dt, J = 7.0, 13.1 Hz, 2H), 2.97 (dd, J = 6.8, 40.5 Hz, 2H), 3.25 – 3.60 (m, 2H), 4.02 – 4.15 (m, 1H), 4.39 (s, 1H), 4.97 – 5.15 (m, 2H), 5.69 (dd, J = 7.8, 66.8 Hz, 1H), 6.54 (dd, J = 6.8, 121.3 Hz, 1H), 6.72 – 6.81 (m, 1H), 7.04 – 7.14 (m, 2H), 7.15 – 7.30 (m, 6H), 7.34 (q, J = 6.2, 7.1 Hz, 4H). ^{13}C NMR (101 MHz, CDCl$_3$) δ -4.05, 18.27, 21.02, 25.87, 31.17, 39.00, 52.34, 56.69, 63.66, 67.03, 119.96,

126.28, 127.00, 127.56, 127.93, 128.03, 128.16, 128.47, 128.65, 129.29, 130.27, 131.00, 136.24, 136.53, 152.38, 155.94, 170.98, 171.35. LC-MS: t_R = 6.90 min; $C_{33}H_{45}N_2O_5Si^+$ [M + H]$^+$, m/z calcd 577.31, found 577.37.

*Oxidation of the alcohol to aldehyde (**Scheme 3.1**, h).* To a solution of benzyl ((S)-1-(((S)-1-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)-3-hydroxypropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (151 mg, 0.26 mmol, 1.0 eq) in DCM (8 mL) at 0 °C was added Dess-Martin periodinane (133 mg, 0.31 mmol, 1.2 eq) and NaHCO$_3$ powder (55 mg, 0.65 mmol, 2.5 eq). The resulting mixture was stirred at 0 °C for 1 h, and then quenched by addition of saturated aqueous Na$_2$S$_2$O$_3$ (2 mL). The reaction was concentrated under reduced pressure, diluted with EtOAc (50 mL), and washed with brine (3 × 10 mL). The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (25% EtOAc in hexane) to give benzyl ((S)-1-(((S)-1-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)-3-oxopropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (68 mg, 45%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 0.24 (d, *J* = 4.6 Hz, 6H), 1.01 (s, 9H), 2.24 (s, 3H), 2.77 – 3.20 (m, 4H), 4.51 (q, *J* = 6.7 Hz, 1H), 5.02 – 5.15 (m, 2H), 5.35 (s, 1H), 6.52 (d, *J* = 5.5 Hz, 1H), 6.71 (d, *J* = 8.2 Hz, 1H), 6.85 (d, *J* = 2.2 Hz, 1H), 6.94 (dd, *J* = 2.3, 8.2 Hz, 1H), 7.16 (d, *J* = 7.1 Hz, 2H), 7.21 – 7.28 (m, 3H), 7.35 (dt, *J* = 4.7, 6.9 Hz, 5H), 9.41 (s, 1H). ^{13}C NMR (100 MHz, CDCl$_3$) δ -4.14, -4.01, 14.19, 18.28, 20.43, 20.99, 25.90, 29.92, 38.80, 53.40, 56.11, 59.57, 60.35, 67.01, 118.77, 125.68, 127.09, 128.01, 128.16, 128.50, 128.67, 128.91, 129.26, 130.94, 131.91, 136.13, 151.28, 170.96, 198.75. LC-MS: t_R = 6.81 min; $C_{33}H_{43}N_2O_5Si^+$ [M + H]$^+$, m/z calcd 575.29, found 575.36.

*Removal of the TBS protecting group (**Scheme 3.1**, i).* To a solution of benzyl ((S)-1-(((S)-1-(2-((*tert*-butyldimethylsilyl)oxy)-5-methylphenyl)-3-oxopropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate (68 mg, 0.12 mmol, 1.0 eq) in THF (3 mL) was slowly added 1.0M TBAF in THF (131 µL, 0.13 mmol, 1.1 eq) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h, and concentrated under reduced pressure. The residue was diluted with EtOAc (50 mL) and washed with saturated aqueous NH_4Cl (10 mL) and brine (10 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (35% EtOAc in hexane) to give benzyl ((2S)-1-(((3S)-2-hydroxy-6-methylchroman-3-yl)amino)-1-oxo-3-phenylpropan-2-yl)carbamate, i.e. compound 7 (20 mg, 36%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.10 – 2.18 (m, 3H), 2.37 – 3.10 (m, 4H), 4.10 – 4.45 (m, 2H), 4.71 – 5.10 (m, 3H), 5.53 (dd, *J* = 7.2, 51.9 Hz, 1H), 5.97 – 6.41 (m, 1H), 6.59 (td, *J* = 10.3, 31.2, 32.4 Hz, 2H), 6.78 (dd, *J* = 8.2, 18.7 Hz, 1H), 6.98 – 7.28 (m, 10H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 20.46, 25.96, 39.10, 46.02, 56.24, 67.20, 91.04, 91.93, 116.68, 118.34, 119.01, 119.23, 128.04, 128.12, 128.20, 128.52, 128.67, 128.77, 129.19, 129.32, 129.64, 130.50, 130.63, 136.02, 136.26, 148.52, 156.15, 171.23. LC-MS: t_R = 5.32 min; $C_{27}H_{29}N_2O_5^+$ [M + H]$^+$, m/z calcd 461.21, found 461.27.

*Preparation of benzyl (4-methylpiperazine-1-carbonyl)-L-phenylalaninate (**Scheme 3.1**, j).* To a solution of 4-methylpiperazine-1-carbonyl chloride hydrochloride (1.365 g, 6.86 mmol, 1.1 eq) in THF (20 mL) at -10 °C was added Et$_3$N (2.08 mL, 14.96 mmol, 2.4 eq) dropwise. The resulting mixture was stirred for 15 min and then was added a solution of benzyl L-phenylalaninate hydrochloride (1.82 g, 6.23 mmol, 1.0 eq) in THF (20 mL) dropwise. The reaction was stirred at room temperature overnight, quenched by

addition of water (10 mL), and concentrated under reduced pressure. The residue was diluted with EtOAc (100 mL) and washed successively with saturated aqueous NH$_4$Cl (20 mL), saturated aqueous NaHCO$_3$ (20 mL) and brine (20 mL). The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (5-10% MeOH in DCM) to give benzyl (4-methylpiperazine-1-carbonyl)-L-phenylalaninate (1.55 g, 60%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.31 (s, 3H), 2.37 (t, *J* = 5.1 Hz, 4H), 3.14 (d, *J* = 4.2 Hz, 2H), 3.37 (q, *J* = 4.9 Hz, 4H), 4.87 (d, *J* = 5.1 Hz, 2H), 5.05 – 5.28 (m, 2H), 7.02 (dd, *J* = 2.9, 6.5 Hz, 2H), 7.20 – 7.27 (m, 3H), 7.30 – 7.42 (m, 5H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 38.31, 43.69, 46.09, 54.32, 54.59, 67.11, 126.93, 128.44, 128.53, 128.57, 129.38, 135.26, 136.11, 156.46, 172.45. LC-MS: t_R = 3.42 min; C$_{22}$H$_{28}$N$_3$O$_3^+$ [M + H]$^+$, m/z calcd 382.21, found 382.2.

*Epoxidation of 1,4-dihydronaphthalene (**Scheme 3.1**, k).* To a solution of 1,4-dihydronaphthalene (2.09 mg, 16.1 mmol, 1.0 eq) in chloroform (40 mL) was slowly added 70% *m*CPBA (4.74 g, 19.3 mmol, 1.2 eq) at 0 °C. The resulting mixture was stirred at room temperature overnight, and quenched by addition of 2M KOH (60 mL). The reaction was concentrated under reduced pressure, and diluted with EtOAc (100 mL). The organic layer was washed with brine (3 × 20 mL), and was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (10% EtOAc in hexane) to give 1a,2,7,7a-tetrahydronaphtho[2,3-b]oxirene (2.03 g, 86%) ^{1}H NMR (400 MHz, Chloroform-*d*) δ 3.05 (d, *J* = 17.7 Hz, 2H), 3.18 (d, *J* = 16.8 Hz, 2H), 3.29 – 3.36 (m, 2H), 6.95 (dd, *J* = 3.5, 5.6 Hz, 2H), 7.06 (dd, *J* = 3.4, 5.7 Hz, 2H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 29.84, 51.71, 126.57, 129.33, 131.74.

Epoxide ring-opening of 1a,2,7,7a-tetrahydronaphtho[2,3-b]oxirene (Scheme 3.1, l). To a solution of 1a,2,7,7a-tetrahydronaphtho[2,3-b]oxirene (2.02 g, 13.8 mmol, 1.0 eq) in a mixture of MeOH (30 mL) and H_2O (10 mL) was added NaN_3 (1.8 g, 27.7 mmol, 2.0 eq) and NH_4Cl (1.11 g, 20.8 mmol, 1.5 eq). The resulting mixture was heated to 60 °C and stirred overnight. The reaction was cooled to room temperature, concentrated under reduced pressure, and diluted with EtOAc (100 mL). The organic layer was washed with brine (3×20 mL), and was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product 3-azido-1,2,3,4-tetrahydronaphthalen-2-ol (2.22 g, 85%) was used without further purification. ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.77 – 2.86 (m, 2H), 3.17 (ddd, *J* = 2.9, 5.8, 16.5 Hz, 2H), 3.66 (td, 1H), 3.87 (td, *J* = 5.8, 9.4 Hz, 1H), 7.04 – 7.09 (m, 2H), 7.09 – 7.16 (m, 2H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 33.64, 36.57, 63.62, 70.47, 126.50, 126.68, 128.59, 128.99, 132.62, 133.41.

Acylation of the hydroxyl group of compound 12 (Scheme 3.2, a). To a solution of compound **12** (50 mg, 0.114 mmol, 1.0 eq) in DCM (3 mL) was added acetic anhydride (32 µL, 0.342 mmol, 3.0 eq), Et_3N (48 µL, 0.342 mmol, 3.0 eq) and DMAP (2.8 mg, 0.023 mmol, 0.2 eq). The resulting mixture was stirred at room temperature overnight, and concentrated under reduced pressure. The residue was diluted with EtOAc (50 mL) and washed successively with saturated aqueous NH_4Cl (10 mL), saturated aqueous $NaHCO_3$ (10 mL) and brine (10 mL). The organic layer was dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (10% MeOH in DCM) to give (3S)-3-((S)-2-(4-methylpiperazine-1-carboxamido)-3-phenylpropanamido)chroman-2-yl acetate, i.e. compound 13 (31 mg, 57%). ^{1}H NMR

(400 MHz, Chloroform-*d*) δ 2.01 (d, *J* = 23.4 Hz, 3H), 2.29 (d, *J* = 3.5 Hz, 3H), 2.30 –

2.38 (m, 4H), 2.69 – 2.97 (m, 2H), 2.98 – 3.15 (m, 2H), 3.25 – 3.41 (m, 4H), 4.49 (dt, *J* =

6.9, 13.8 Hz, 1H), 5.18 (dd, *J* = 7.4, 40.0 Hz, 1H), 6.12 (dd, *J* = 2.5, 17.4 Hz, 1H), 6.31 –

6.54 (m, 1H), 6.82 – 6.91 (m, 1H), 6.91 – 7.07 (m, 2H), 7.07 – 7.19 (m, 2H), 7.19 (s, 3H),

7.26 – 7.34 (m, 2H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 20.99, 26.61, 29.67, 30.86, 38.69,

43.71, 45.99, 53.39, 54.48, 56.23, 89.33, 117.07, 119.40, 121.89, 127.09, 128.06, 128.70,

128.74, 129.15, 129.20, 136.94, 150.22, 157.02, 169.38, 171.96. LC-MS: t_R = 3.29 min;

C$_{26}$H$_{33}$N$_4$O$_5^+$ [M + H]$^+$, m/z calcd 481.24, found 481.3.

*Alkylation of the hydroxyl group of compound 12 (**Scheme 3.2**, b).* To a solution

of compound **12** (40 mg, 0.091 mmol, 1.0 eq) in EtOH (2 mL) was added BF$_3$OEt$_2$ (300

μL, 2.43 mmol, >20 eq) dropwise at 0 °C. The resulting mixture was stirred at room

temperature overnight, and quenched by addition of saturated aqueous NH$_4$Cl (2 mL).

The reaction was concentrated under reduced pressure and then partitioned between

DCM (20 mL) and water (20 mL). The water layer was further washed with DCM (2 ×

20 mL). The combined organic layers were dried over Na$_2$SO$_4$, filtered, and concentrated

under reduced pressure. The obtained crude product was purified by FCC (8% MeOH in

DCM) to give *N*-((2S)-1-(((3S)-2-ethoxychroman-3-yl)amino)-1-oxo-3-phenylpropan-2-

yl)-4-methylpiperazine-1-carboxamide, i.e. compound 16 (18.3 mg, 43%). ^{1}H NMR (400

MHz, Chloroform-*d*) δ 1.00 (t, 3H), 2.16 – 2.36 (m, 7H), 2.69 (dd, *J* = 9.2, 14.2 Hz, 1H),

2.83 – 2.95 (m, 1H), 3.10 (dt, *J* = 7.2, 13.4 Hz, 1H), 3.21 (dd, *J* = 4.0, 6.1 Hz, 1H), 3.31

(dt, *J* = 6.0, 11.1 Hz, 3H), 3.58 – 3.76 (m, 1H), 4.13 – 4.27 (m, 1H), 4.41 (ddd, *J* = 8.0,

14.1, 38.9 Hz, 1H), 4.68 (dd, *J* = 2.3, 49.7 Hz, 1H), 4.90 – 5.15 (m, 1H), 5.78 (dd, *J* =

8.9, 24.8 Hz, 1H), 6.73 (d, *J* = 8.1 Hz, 1H), 6.80 (q, *J* = 6.8, 7.4 Hz, 1H), 6.90 (d, *J* = 7.2

Hz, 1H), 7.03 (t, $J = 7.2$ Hz, 1H), 7.07 – 7.30 (m, 5H). ^{13}C NMR (101 MHz, CDCl$_3$) δ

14.94, 14.98, 27.08, 39.59, 43.72, 45.10, 46.06, 54.58, 56.30, 64.01, 96.02, 96.34, 116.79,

120.48, 121.16, 126.99, 127.67, 128.65, 128.75, 129.20, 129.33, 136.84, 137.17, 150.58,

150.70, 156.56, 156.68, 171.26, 171.66. LC-MS: t_R = 3.53 min; C$_{26}$H$_{35}$N$_4$O$_4^+$ [M + H]$^+$,

m/z calcd 467.27, found 467.3.

Preparation of (S)-N-(2-benzoylphenyl)-1-benzylpyrrolidine-2-carboxamide

(BPB, Scheme 3.3, a). To a mixed solution of benzyl-L-proline (1.03 g, 5.0 mmol, 1.0

eq) and 1-methylimidazole (877 µL, 11.0 mmol, 2.2 eq) in DCM (10 mL) was added

MsCl (387 µL, 5.0 mmol, 1.0 eq) dropwise at 0 ºC. The resulting mixture was stirred at

room temperature for 10 min and then was added a solution of 2-aminobenzophenone

(888 mg, 4.5 mmol, 0.9 eq) in DCM (10 mL). The reaction was heated to 45 ºC and

stirred overnight. The reaction was quenched by addition of NH$_4$Cl (15 mL), and

concentrated under reduced pressure. The residue was diluted with EtOAc (75 mL),

washed with saturated aqueous NaHCO$_3$ (15 mL) and brine (15 mL). The organic layer

was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained

crude product was purified by FCC (15-25% EtOAc in hexane) to give (S)-N-(2-

benzoylphenyl)-1-benzylpyrrolidine-2-carboxamide, i.e. BPB (1.344 g, 70%). ^{1}H NMR

(400 MHz, Chloroform-*d*) δ 1.72 – 1.85 (m, 2H), 1.91 – 2.01 (m, 1H), 2.17 – 2.31 (m,

1H), 2.40 (td, $J = 6.8, 9.5$ Hz, 1H), 3.20 (ddd, $J = 2.4, 6.4, 9.1$ Hz, 1H), 3.31 (dd, $J = 4.8,$

10.1 Hz, 1H), 3.58 (d, $J = 12.9$ Hz, 1H), 3.91 (d, $J = 12.9$ Hz, 1H), 7.07 (td, $J = 1.1, 7.6$

Hz, 1H), 7.10 – 7.17 (m, 3H), 7.24 (s, 1H), 7.33 – 7.40 (m, 2H), 7.49 (qd, $J = 6.5, 7.9$ Hz,

4H), 7.56 – 7.61 (m, 1H), 7.73 – 7.83 (m, 2H), 8.57 (dd, $J = 1.0, 8.4$ Hz, 1H), 11.50 (s,

1H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 24.18, 31.03, 53.89, 59.87, 68.32, 121.53, 122.19,

125.36, 127.06, 128.16, 128.31, 129.13, 130.10, 132.45, 132.54, 133.35, 138.15, 138.59, 139.22, 174.60, 197.99. LC-MS: t_R = 3.50 min; $C_{25}H_{25}N_2O_2^+$ [M + H]$^+$, m/z calcd 385.19, found 384.85.

Preparation of the tetracoordinate nickel (II) complex of (S,E)-2-(((2-(1-benzylpyrrolidine-2-carboxamido)phenyl)(phenyl)methylene)amino)acetic-1-^{13}C acid (Ni-BPB-Gly, **Scheme 3.3**, b).* To a mixed suspension of BPB (860 mg, 2.24 mmol, 1.0 eq), glycine-1-^{13}C (425 mg, 5.59 mmol, 2.5 eq), and Ni(NO$_3$)$_2$·6H$_2$O (1.30 g, 4.47 mmol, 2.0 eq) in MeOH (7 mL) at 45 °C was added a solution of ground KOH (752 mg, 13.4 mmol, 6.0 eq) in MeOH (3 mL). The resulting mixture was stirred at 60 °C for 1 h (note that prolonged heating might lead to racemization).The reaction was neutralized by addition of acetic acid (800 µL), and diluted with water to a volume of 50 mL. The reaction was concentrated under reduced pressure and extracted with DCM (75 mL). The organic layer was dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The obtained crude product was purified by FCC (20-30% acetone in DCM) to give the tetracoordinate nickel (II) complex of (S,E)-2-(((2-(1-benzylpyrrolidine-2-carboxamido)phenyl)(phenyl)methylene)amino)acetic-1-^{13}C acid, i.e. Ni-BPB-Gly* (715 mg, 64%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.07 (dddd, *J* = 2.4, 5.6, 8.5, 15.1 Hz, 1H), 2.12 – 2.21 (m, 1H), 2.42 (dddd, *J* = 8.3, 9.5, 10.7, 13.4 Hz, 1H), 2.57 (dddd, *J* = 2.6, 6.3, 9.2, 14.4 Hz, 1H), 3.25 – 3.40 (m, 1H), 3.46 (dd, *J* = 5.4, 10.7 Hz, 1H), 3.61 – 3.74 (m, 3H), 3.77 (dd, *J* = 5.5, 20.1 Hz, 1H), 4.48 (d, *J* = 12.7 Hz, 1H), 6.69 (ddd, *J* = 1.2, 6.9, 8.1 Hz, 1H), 6.80 (dd, *J* = 1.7, 8.3 Hz, 1H), 6.92 – 7.04 (m, 1H), 7.05 – 7.13 (m, 1H), 7.20 (ddd, *J* = 1.7, 6.9, 8.7 Hz, 1H), 7.28 – 7.33 (m, 1H), 7.42 (t, *J* = 7.6 Hz, 2H), 7.52 (ddd, *J* = 3.4, 7.7, 18.9 Hz, 3H), 8.01 – 8.12 (m, 2H), 8.31 (dd, *J* = 1.1, 8.6 Hz, 1H).

^{13}C NMR (100 MHz, CDCl$_3$) δ 23.69, 30.76, 57.56, 61.57, 63.16, 69.94, 124.26, 125.16, 126.27, 128.91, 129.11, 129.34, 129.59, 131.73, 132.22, 132.77, 133.17, 133.37, 134.67, 142.62, 143.81, 177.21, 179.27. LC-MS: t_R = 3.71 min; C$_{26}$^{13}CH$_{26}$N$_3$NiO$_3^+$ [M + H]$^+$, m/z calcd 499.14, found 498.83.

Preparation of the tetracoordinate nickel (II) complex of (S)-3-(2-(benzyloxy)phenyl)-2-(((E)-(2-((S)-1-benzylpyrrolidine-2-carboxamido)phenyl)(phenyl)methylene)amino)propanoic-1-^{13}C acid (Ni-BPB-oTyr(OBn), **Scheme 3.3**, c).* A mixed slurry of Ni-BPB-Gly* (715 mg, 1.43 mmol, 1.0 eq), (2-(benzyloxy)phenyl)methanol (613 mg, 2.86 mmol, 2.0 eq), and CMBP (750 µL, 2.86 mmol, 2.0 eq) in toluene (5 mL) was heated to 120 °C and stirred overnight. The reaction was cooled to room temperature and concentrated under reduced pressure. The obtained crude product was purified by FCC (10% acetone in DCM) to give the tetracoordinate nickel (II) complex of (S)-3-(2-(benzyloxy)phenyl)-2-(((E)-(2-((S)-1-benzylpyrrolidine-2-carboxamido)phenyl)(phenyl)methylene)amino)propanoic-1-^{13}C acid, i.e. Ni-BPB-^{13}oTyr(OBn), (417 mg, 40%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 1.48 – 1.59 (m, 1H), 1.99 (ddd, *J* = 6.5, 8.4, 11.1 Hz, 1H), 2.04 – 2.14 (m, 1H), 2.31 (tdd, *J* = 2.9, 6.7, 9.7 Hz, 2H), 2.76 (ddd, *J* = 1.6, 4.4, 13.6 Hz, 1H), 2.90 – 3.05 (m, 1H), 3.22 (ddd, *J* = 5.0, 7.2, 13.6 Hz, 1H), 3.33 (dd, *J* = 8.0, 9.3 Hz, 1H), 3.44 (d, *J* = 12.6 Hz, 1H), 4.18 (q, *J* = 4.4 Hz, 1H), 4.23 (d, *J* = 12.6 Hz, 1H), 4.55 (d, *J* = 10.8 Hz, 1H), 4.90 (d, *J* = 10.8 Hz, 1H), 6.14 (dt, *J* = 1.5, 7.7 Hz, 1H), 6.41 (dd, *J* = 1.7, 8.2 Hz, 1H), 6.63 (ddd, *J* = 1.2, 6.9, 8.2 Hz, 1H), 6.83 – 6.94 (m, 2H), 6.96 – 7.22 (m, 9H), 7.28 (d, *J* = 7.6 Hz, 2H), 7.33 – 7.46 (m, 4H), 7.95 – 8.08 (m, 2H), 8.35 (dd, *J* = 1.1, 8.7 Hz, 1H). ^{13}C NMR (100 MHz, Chloroform-*d*) δ 22.35, 25.60, 28.10, 59.19, 63.38, 68.12, 69.68, 71.86, 112.85,

122.38, 123.33, 123.36, 124.10, 127.39, 127.74, 127.80, 128.06, 128.35, 128.51, 128.73, 128.88, 129.20, 129.27, 129.61, 130.06, 130.30, 130.32, 134.10, 134.82, 137.07, 145.11, 155.82, 158.83, 170.83, 175.71. LC-MS: t_R = 4.70 min; $C_{40}{}^{13}CH_{38}N_3NiO_4{}^+$ [M + H]$^+$, m/z calcd 695.22, found 694.96.

Preparation of (S)-2-amino-3-(2-(benzyloxy)phenyl)propanoic-1-^{13}C acid (H-oTyr(OBn)*-OH, Scheme 3.3, d). A mixed solution of Ni-BPB-*o*Tyr(OBn)* (522 mg, 0,75 mmol, 1.0 eq) and 8-quinolinol (272 mg, 1.88 mmol, 2.5 eq) in MeCN (10 mL) and H_2O (1 mL) was heated to 40 ^{o}C and stirred overnight. The reaction was filtered, and the filtrate was concentrated. The residue was diluted with DCM (30 mL), and washed with water (3 × 30 mL). The organic layer containing the retrieved BPB was concentrated and saved for future use. The combined water layers were concentrated to a small volume that was subsequently purified by Prep-HPLC to give (S)-2-amino-3-(2-(benzyloxy)phenyl)propanoic-1-^{13}C acid, i.e. H-*o*Tyr(OBn)*-OH (87 mg, 43%). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.96 (ddd, *J* = 2.5, 9.4, 14.3 Hz, 1H), 3.49 (ddd, *J* = 2.9, 4.5, 14.3 Hz, 1H), 3.93 (dt, *J* = 4.7, 9.4 Hz, 1H), 5.20 (s, 2H), 6.92 (td, *J* = 1.1, 7.4 Hz, 1H), 7.05 (d, *J* = 8.1 Hz, 1H), 7.24 (dd, *J* = 6.7, 8.2 Hz, 2H), 7.28 – 7.34 (m, 1H), 7.38 (dd, *J* = 6.6, 8.3 Hz, 2H), 7.45 – 7.53 (m, 2H). ^{13}C NMR (100 MHz, CDCl$_3$) δ 33.52, 55.91, 71.50, 114.97, 120.74, 124.67, 127.65, 127.97, 128.07, 128.32, 130.64, 139.00, 157.44, 173.03. LC-MS: t_R = 2.70 min; $C_{15}{}^{13}CH_{18}NO_3{}^+$ [M + H]$^+$, m/z calcd 273.13, found 272.75.

Protection of the amino group of (S)-2-amino-3-(2-(benzyloxy)phenyl)propanoic-1-^{13}C acid with Boc protecting group (Boc-oTyr(OBn)*-OH, Scheme 3.3, e). To a solution of H-*o*Tyr(OBn)*-OH (87 mg, 0.32 mmol, 1.0 eq) in dioxane (5 mL) and H_2O (5

mL) was added Et₃N (134 μL, 0.96 mmol, 3.0 eq) and Boc anhydride (84 mg, 0.38 mmol, 1.2 eq) at 0 °C. The resulting mixture was stirred at room temperature for 2.5 h, and then concentrated under reduced pressure. The residue was acidified by addition of 0.1M HCl to pH 2-3 and then diluted with EtOAc (50 mL). The organic layer was washed with brine (3 × 10 mL), dried over Na₂SO₄, filtered, and concentrated under reduced pressure. The obtained crude product (S)-3-(2-(benzyloxy)phenyl)-2-((*tert*-butoxycarbonyl)amino)propanoic-1-¹³C acid, i.e. Boc-*o*Tyr(OBn)*-OH (113 mg, 94%) was used without further purification. ¹H NMR (400 MHz, Chloroform-*d*) δ 1.28 (s, 9H), 2.74 – 3.06 (m, 1H), 3.15 (d, *J* = 14.5 Hz, 1H), 4.33 – 4.55 (m, 1H), 5.00 (d, *J* = 6.7 Hz, 2H), 6.82 (d, *J* = 8.0 Hz, 2H), 7.04 – 7.15 (m, 2H), 7.21 (t, *J* = 7.3 Hz, 1H), 7.28 (d, *J* = 14.8 Hz, 2H), 7.36 (d, *J* = 7.4 Hz, 2H), 9.86 (s, 1H). ¹³C NMR (100 MHz, CDCl₃) δ 28.30, 29.71, 32.44, 54.81, 70.26, 111.95, 121.12, 127.22, 127.93, 128.46, 128.64, 131.40, 176.93. LC-MS: t_R = 4.28 min; $C_{20}{}^{13}CH_{25}NNaO_5{}^+$ [M + Na]⁺, m/z calcd 395.17, found 394.83; $C_{15}{}^{13}CH_{18}NO_3{}^+$ [M + H - Boc]⁺, m/z calcd 273.13, found 272.82.

Preparation of the Weinreb amide tert-butyl (S)-(3-(2-(benzyloxy)phenyl)-1-(methoxy(methyl)amino)-1-oxopropan-2-yl-1-¹³C)carbamate (Boc-oTyr(OBn)-N(Me)OMe, Scheme 3.3, f).* The procedure for amide coupling reaction of Boc-*o*Tyr(OBn)*-OH (113 mg, 0.30 mmol, 1.0 eq) with *N,O*-dimethylhydroxylamine hydrochloride (44.4 mg, 0.45 mmol, 1.5 eq) was similar to that for step d in **Scheme 3.1**. The obtained crude product was purified by FCC (25% EtOAc in hexane) to give *tert*-butyl (S)-(3-(2-(benzyloxy)phenyl)-1-(methoxy(methyl)amino)-1-oxopropan-2-yl-1-¹³C)carbamate, i.e. Boc-*o*Tyr(OBn)*-N(Me)OMe (136 mg, 100%). ¹H NMR (400 MHz, Chloroform-*d*) δ 1.25 (s, 9H), 2.79 – 2.96 (m, 2H), 2.99 (d, *J* = 5.6 Hz, 3H), 3.44 (s, 3H),

4.90 (s, 1H), 4.99 (d, J = 4.1 Hz, 2H), 5.21 (d, J = 8.9 Hz, 1H), 6.79 (t, J = 7.7 Hz, 2H),

7.00 – 7.14 (m, 2H), 7.18 – 7.32 (m, 3H), 7.38 (d, J = 7.5 Hz, 2H). ^{13}C NMR (100 MHz,

CDCl$_3$) δ 28.30, 29.66, 32.03, 45.85, 61.26, 70.23, 79.09, 111.68, 120.70, 125.50, 127.52,

127.87, 128.15, 128.54, 131.42, 137.08, 155.20, 157.10, 172.86. LC-MS: t_R = 4.60 min;

$C_{22}{}^{13}CH_{30}N_2NaO_5{}^+$ [M + Na]$^+$, m/z calcd 438.21, found 437.81; $C_{17}{}^{13}CH_{22}N_2O_5{}^+$ [M + H

- Boc]$^+$, m/z calcd 316.17, found 315.73.

Preparation of N-((S)-1-(((S)-3-(2-(benzyloxy)phenyl)-1-(methoxy(methyl)amino)-
1-oxopropan-2-yl-1-^{13}C)amino)-1-oxo-3-phenylpropan-2-yl)-4-methylpiperazine-1-
carboxamide (NMePip-Phe-oTyr(OBn)*-N(Me)OMe, Scheme 3.3, g). The procedure for

removal of Boc protecting group of Boc-*o*Tyr(OBn)*-N(Me)OMe (136 mg, 0.33 mmol,

1.0 eq) was similar to that for step c in **Scheme 3.1**. Subsequently, the procedure for

amide coupling reaction of the crude product with NMePip-Phe-OH (114 mg, 0.39 mmol,

1.2 eq) was similar to that for step d in **Scheme 3.1**. The obtained crude product was

purified by FCC (8% MeOH in DCM) to give *N*-((S)-1-(((S)-3-(2-(benzyloxy)phenyl)-1-

(methoxy(methyl)amino)-1-oxopropan-2-yl-1-^{13}C)amino)-1-oxo-3-phenylpropan-2-yl)-4-

methylpiperazine-1-carboxamide, i.e. NMePip-Phe-*o*Tyr(OBn)*-N(Me)OMe (107 mg,

55%).^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.27 (s, 3H), 2.30 (dt, J = 3.7, 6.5 Hz, 4H),

2.96 (dd, J = 2.9, 6.3 Hz, 3H), 3.07 (s, 4H), 3.28 (ddd, J = 7.7, 13.5, 17.9 Hz, 4H), 3.48

(s, 3H), 4.49 (q, J = 6.5 Hz, 1H), 4.99 (d, J = 7.0 Hz, 1H), 5.05 (s, 2H), 5.19 (s, 1H), 6.62

(s, 1H), 6.79 (t, J = 7.3 Hz, 1H), 6.90 (dd, J = 7.3, 19.1 Hz, 2H), 7.12 – 7.25 (m, 6H),

7.30 (t, J = 7.3 Hz, 1H), 7.37 (t, J = 7.4 Hz, 2H), 7.46 (d, J = 7.2 Hz, 2H). ^{13}C NMR (100

MHz, CDCl$_3$) δ 32.03, 32.92, 38.74, 43.62, 46.03, 49.23, 54.53, 55.10, 61.22, 70.22,

111.82, 120.68, 125.09, 126.70, 127.57, 127.97, 128.30, 128.36, 128.59, 129.68, 131.43,

136.95, 137.02, 171.76. LC-MS: t_R = 3.72 min; $C_{32}{}^{13}CH_{42}N_5O_5{}^+$ [M + H]$^+$, m/z calcd 589.32, found 589.02.

Preparation of N-((S)-1-(((S)-1-(2-(benzyloxy)phenyl)-3-oxopropan-2-yl-3-
^{13}C)amino)-1-oxo-3-phenylpropan-2-yl)-4-methylpiperazine-1-carboxamide (NMePip-
Phe-oTyr(OBn)*-H, Scheme 3.3, h). To a solution of NMePip-Phe-oTyr(OBn)*-
N(Me)OMe (107 mg, 0.182 mmol, 1.0 eq) in THF (8 mL) at -10 ^{o}C was added 2.0 M
LAH in THF (218 µL, 0.218 mmol, 1.2 eq) dropwise. The resulting mixture was stirred
at 0 ^{o}C for 1 h, and then quenched by tiny pieces of ice. The reaction was concentrated
under reduced pressure, and diluted with EtOAc (50 mL). The organic layer was washed
successively with a saturated solution of Rochelle salt (15 mL), NaHCO$_3$ (15 mL) and
brine (15 mL), dried over Na$_2$SO$_4$, filtered, and concentrated under reduced pressure. The
obtained crude product was purified by FCC (10% MeOH in DCM) to give the *N*-((S)-1-
(((S)-1-(2-(benzyloxy)phenyl)-3-oxopropan-2-yl-3-^{13}C)amino)-1-oxo-3-phenylpropan-2-
yl)-4-methylpiperazine-1-carboxamide, i.e. NMePip-Phe-*o*Tyr(OBn)*-H (81 mg, 84%).
^{1}H NMR (400 MHz, Chloroform-*d*) δ 2.28 (s, 3H), 2.31 (t, *J* = 4.2 Hz, 4H), 2.95 – 3.13
(m, 4H), 3.29 (dt, *J* = 5.2, 9.7 Hz, 4H), 4.52 (q, *J* = 7.0 Hz, 2H), 4.95 (d, *J* = 7.3 Hz, 1H),
5.04 (dd, *J* = 3.6, 10.5 Hz, 2H), 6.82 (s, 3H), 7.10 – 7.24 (m, 6H), 7.29 – 7.50 (m, 6H),
9.36 (s, 1H). ^{13}C NMR (100 MHz, CDCl3) δ 29.66, 38.59, 43.67, 46.04, 54.52, 55.31,
59.49, 70.25, 97.58, 97.62, 97.73, 99.08, 99.24, 111.96, 121.04, 126.93, 127.55, 128.19,
128.55, 128.60, 128.73, 129.33, 129.43, 131.61, 136.52, 136.76, 156.39, 156.60, 171.89,
198.37. LC-MS: t_R = 3.76 min; $C_{30}{}^{13}CH_{36}N_4O_4{}^+$ [M + H]$^+$, m/z calcd 530.28, found
530.03.

Preparation of ^{13}C-enriched compound 12 (Scheme 3.3, i). The procedure for

reductive hydrogenolysis of the benzyl group of NMePip-Phe-*o*Tyr(OBn)*-H (81 mg,

0.152 mmol, 1.0 eq) was similar to that for step b in **Scheme 3.1**. The obtained crude

product was purified by Prep-HPLC to give compound *N*-((S)-1-(((S)-2-hydroxy-2l3-

chroman-3-yl-2-^{13}C)amino)-1-oxo-3-phenylpropan-2-yl)-4-methylpiperazine-1-

carboxamide, i.e. ^{13}C-labeled compound **12** (45 mg, 67%). ^{1}H NMR (400 MHz,

Chloroform-*d*) δ 1.28 (s, 1H), 2.45 (s, 3H), 2.65 (p, *J* = 6.5 Hz, 4H), 3.03 (dt, *J* = 7.6,

21.2 Hz, 2H), 3.40 – 3.60 (m, 4H), 4.21 – 4.40 (m, 1H), 4.64 (dq, *J* = 7.7, 46.1 Hz, 1H),

5.38 (dd, *J* = 30.9, 170.9 Hz, 1H), 6.16 (d, *J* = 8.0 Hz, 1H), 6.50 (dd, *J* = 8.0, 141.5 Hz,

1H), 6.86 – 7.08 (m, 3H), 7.13 (dd, *J* = 3.0, 6.9 Hz, 1H), 7.20 (dt, *J* = 2.3, 6.5 Hz, 2H),

7.25 (d, *J* = 6.7 Hz, 2H), 7.71 (s, 2H), 8.27 (s, 1H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 26.88,

30.87, 39.00, 43.55, 45.87, 54.40, 56.02, 91.14, 91.98, 116.82, 118.97, 119.93, 121.18,

126.92, 127.75, 128.63, 129.23, 129.37, 136.96, 151.02, 156.87, 172.19. LC-MS: t_R =

2.77 min; C$_{23}$^{13}CH$_{31}$N$_4$O$_4^+$ [M + H]$^+$, m/z calcd 440.24, found 439.94.

The procedures for steps a and d in **Scheme 3.4** were similar to those for steps a

and b in **Scheme 3.1**, respectively. The procedures for steps d, e and f were similar to

those for steps f, g and h in **Scheme 3.3**, respectively. Therefore, only step c was

described in detail as below.

*Preparation of 2-((tert-butoxycarbonyl)amino)-3-(2-oxo-1,2-dihydropyridin-3-

yl)propanoic acid (Scheme 3.4, c).* A mixed solution of methyl 2-((*tert*-

butoxycarbonyl)amino)-3-(2-oxo-1,2-dihydropyridin-3-yl)propanoate (1.18 g, 4.0 mmol,

1.0 eq) and LiOH (192 mg, 8.0 mmol, 2.0 eq) in MeOH (10 mL) and H$_2$O (5 mL) was

stirred at room temperature for 2 h. The reaction was acidified by addition of 0.1M HCl

to pH 2-3, and then concentrated under reduced pressure. The residue was extracted with DCM (3×50 mL). The organic layers were combined, dried over Na_2SO_4, filtered, and concentrated under reduced pressure. The obtained crude product 2-(($tert$-butoxycarbonyl)amino)-3-(2-oxo-1,2-dihydropyridin-3-yl)propanoic acid was used without further purification. ^{1}H NMR (400 MHz,) δ 1.39 (s, 9H), 2.86 (dd, $J = 8.9$, 14.0 Hz, 1H), 3.11 (dd, $J = 4.3$, 13.9 Hz, 1H), 4.42 (dd, $J = 4.3$, 8.8 Hz, 1H), 6.32 (t, $J = 6.7$ Hz, 1H), 7.30 (dd, $J = 2.0$, 6.6 Hz, 1H), 7.39 – 7.54 (m, 1H).

3.3.3 Enzyme Preparation

The information about human cathepsins L and B, and the experimental details of cruzain expression, purification and activation can be found in Section 2.3.4.

The gene expression and purification of SARS-CoV-2 3CL^{pro} were described in our recent publication.[226] Briefly, the 3CL^{pro}-encoding gene (ORF1ab nucleotides 10055-10972, GenBank code: MN988668.1) was *E. coli*-optimized and ligated into the vector pGEX-6P-1 plasmid (Genscript). At the *N*-terminus with regards to the 3CL^{pro} gene, the construct encoded a *GST* sequence preceding a 3CL^{pro} cleavage sequence (SAVLQ↓SGF) which resided in between Nsp4 and Nsp5 of the viral polyprotein. At the *C*-terminus, the construct contained a modified PreScission protease cleavage sequence (SGVTFQ↓GP) followed by a His$_6$-tag sequence. The authentic *N*-terminus (SGF) was produced by removal of the GST tag during autoproteolysis while the authentic *C*-terminus was exposed after purification by immobilized metal ion affinity chromatography (IMAC), with subsequent treatment with PreScission protease.

The plasmid was transformed into BL21(DE3) competent cells (NEB) and the clones were inoculated into LB media containing 100 µg/mL of ampicillin which was grown overnight at 37 °C. On the next day, 1 L of big culture was inoculated with the starting culture and grown at 37 °C until OD_{600} reaching 0.6 – 0.8, at which time 1.0 mM of isopropyl β-D-1-thiogalactoside (IPTG) was added for induction. After being incubated for another 4 – 5 h, cells were harvested by centrifugation ($6,000 \times g$ for 30 min) and stored at -80 °C.

Frozen cell pellets were thawed and resuspended in buffer A (12 mM Tris·HCl, 120 mM NaCl, 0.1 mM EDTA-Na_2, 2 mM DTT, pH 7.5). The cells were lysed by sonication or use of a French press (25,000 psi), with the lysates then clarified by centrifugation at $26,000 \times g$. The supernatant was passed through a 0.45-µm filter and loaded on to a HisTrap™ HP 5 mL column equilibrated with buffer A. The column was washed with buffer A, and eluted using increasing buffer B (12 mM Tris·HCl, 120 mM NaCl, 500 mM imidazole, 0.1 mM EDTA, 2 mM DTT, pH 7.5) with a linear gradient to 35% B over 25 column volumes (CV). The fractions containing target protein were pooled and dialyzed against buffer A to remove imidazole. The resulting solution was treated with PreScission protease (3.5 U/mg of 3CL^{pro}) overnight, and loaded onto a GSTrap™ HP 5 mL column and HisTrap™ HP column arranged in series, to remove the GST-fused PreScission protease, the His$_6$-tag, and undigested His$_6$-tagged protein. The His$_6$-tag-free 3CL^{pro} was collected from the flow-through, characterized by SDS-PAGE, concentrated by centrifugal filtration (Amicon® Ultra, 10 kDa NMWL). The apparently homogenous 3CL^{pro} was stored at -80 °C with 50% glycerol (v/v).

3.3.4 Enzyme Assays and Evaluation of Inhibitors

The experimental details of cruzain inhibition assays, including rapid dilution assays, can be found in Section 2.3.5.

Kinetic assays of purified recombinant SARS-CoV 3CL^{pro} were performed at 25 ^{o}C in 96-well half-area flat-bottom plates (Greiner, 675076). The reaction buffer consisted of 20 mM Tris·HCl, 150 mM NaCl, 0.1 mM EDTA-Na$_2$, and 2 mM DTT (pH 7.5). Varied concentrations of a FRET-based decapeptide substrate ACC-SAVLQSGFRK(DNP)-NH$_2$ was first dissolved in pure DMSO and then diluted by 10-fold into the reaction mixture (10% DMSO). Reactions were initiated by addition of 3CL^{pro} (30 – 50 nM). The total volume of each reaction was taken up to 80 µL. The fluorescence generated by hydrolyzed substrate was immediately measured using a Synergy HTX (Biotek) microplate reader at λ_{ex} of 360 nm and λ_{em} of 460 nm. Initial rates were calculated as fluorescence change per unit time within <10% consumption of substrate, and were fitted to eq. 2-1 to produce a K$_M$ of 26 µM for this substrate. Evaluation of inhibitors **18** – **24** was performed using the burst inhibition kinetic method. Varied concentrations of a given inhibitor were pre-mixed with substrate in pure DMSO. Mixtures were then added to the reaction buffer containing 50 nM of 3CL^{pro}. Steady-state reaction rates (v$_s$) were measured typically between 10 – 20 min.

3.3.5 Analysis of Kinetic Data

Most of the data processing methods can be found in Section 2.3.10. However, there are some differences described as follows.

As a tight-binding inhibitor, the overall inhibition constants (K_i^*) of compound **1** for cruzain and compound **19** for 3CL^{pro} were obtained by fitting v_i data to the quadratic Morrison equation (eq. 3-1) using either a floated or fixed concentration of enzyme.[227] Notably, eq. 2-5 also gave a very similar value of K_i^* for compound **1**.

$$\frac{v_s}{v_0} = \frac{[E]-[I]-K_i^*(1+\frac{[S]}{K_M}) + \sqrt{([E]+[I]+K_i^*(1+\frac{[S]}{K_M}))^2-4[E][I]}}{2[E]}$$ (3-1)

As previously described, the linear relation between k_{obs} and [I] of compound **1** and **12** were due to $K_i \gg K_i^*$. Under this condition, the concentration of inhibitor required to observe slow binding inhibition would be much less than the value of K_i for the EI complex. In this case, eq. 2-4 was reduced to following equation.

$$k_{obs} = k_4[1 + \frac{[I]}{K_i^*\left(1+\frac{[S]}{K_M}\right)}]$$ (3-2)

Eq. 3-2 is a linear function with slope $= k_6/K_i^{*app}$ and y-intercept $= k_6$. For rapid dilution assay, the inhibitor concentration was fixed at $0.1 \times K_i^{*app}$ after dilution, so k_4 could be estimated as $k_{obs}/1.1$ according to eq. 3-2.

The half-lives of hydrolytic reactions of O-acylated SMAIs were obtained from following the single-phase decay equation.

$$A = (A_0 - C)e^{-kt} + C$$ (3-3)

In eq. 3-3, A_0 and C are the initial and final fraction of unhydrolyzed compound, respectively, k is the pseudo-first order rate constant which can be converted to half-life ($t_{1/2} = \ln2/k$).

3.3.6 2D NMR in Aqueous Solution

Wilmad precision NMR tubes (541-PP-8, thin wall, 800 MHz, 5 mm × 8 inch) were used for this experiment. The total volume of sample in each tube was 660 µL, and the final concentrations of the components in the NMR buffer were as follows: 18 mM phosphate (K_2HPO_4/KH_2PO_4), 45 mM NaCl, 4.5 mM DTT, pH 7.5, 10% DMSO-d_6 (v/v). A 10 mM stock of ^{13}C-labeled 12 in DMSO-d_6 was diluted into the NMR buffer to specific concentration. The control sample contained 0.6 mM of ^{13}C-labeled 12, while the experimental sample contained 0.4 mM of ^{13}C-labeled 12 and 0.44 mM of cruzain freshly purified by size exclusion chromatography from a -80 °C stock to minimize the self-degradation. Both samples were submitted to a Bruker AVANCE III HD 800 MHz (18.8 Tesla) equipped with 5 mm triple resonance cryoprobe (TCI). ^{1}H-^{13}C HSQC NMR spectra were acquired at 25 °C with 72 scans for control sample and 128 scans for experimental sample. Spectral widths were 9615 Hz and 28179 Hz in the ^{13}C and ^{1}H dimensions, respectively. Spectra were processed using NMRpipe,[228] and cross-peak intensities were analyzed using Sparky and MestreNova.[229-230]

3.3.7 Detection of Aldehyde Content

In a 96-well clear microplate, compounds **11** and **12** in DMSO solutions, or pure DMSO as a control, were diluted into the above NMR buffer without addition of DTT to make 250 µL-samples, each containing 0.2 mM of compounds and 10% DMSO (v/v). Upon addition of 1M phenylhydrazine (0.25 µL), the formation of phenylhydrazone in these wells were continuously monitored for 2 hours by measuring UV absorbance at the maximal wavelength of 282 nm.

3.3.8 Evaluation of O-derivatized SMAIs as Prodrugs

Esterase from porcine liver was purchased as an ammonium sulfate suspension (Sigma, E2884, 10 U/µL). *O*-acylated compounds were first prepared as 20 mM stock solutions in DMSO. Inside a microcentrifuge tube, 100 µL of compound stock was diluted into 900 µL of reaction buffer to make 1 mL of solution containing 2 mM compound, 1X cruzain buffer (50 mM MES, 50 mM TAPSO, 1 mM EDTA-Na$_2$, 1mM CHAPS), and 10% DMSO (v/v), at pH 7.5. For the experimental sample, 4 µL of esterase suspension was added to the reaction tube (40 U/mL) which was immediately placed on a low speed orbital shaker in a 37 °C incubator. Besides, a negative control without addition of esterase and a blank control with only buffer were set up in the meantime. Each aliquot (140 µL) was taken from the reaction solution at 0, 15, 30, 60, 120, 180, and 300 min, and was quenched with 260 µL of MeCN. The resulting solution was vortexed and centrifuged to precipitate the protein, and the supernatant was analyzed on HPLC-MS under UV 270 nm. The time course of the fraction of intact *O*-acylated compound conformed to an integrated equation of a first-order reaction (eq. 3-3), from which the half-life of hydrolysis by esterase was deduced.

CypExpress™ 3A4, 2D6 and 2C9 were commercially obtained P450 products (Sigma catalogue numbers: MTOXCE3A4, MTOXCE3D6, and MTOXCE2C9) which already included glucose-6-phosphate dehydrogenase (G6PDH) and Mg^{2+} within the system. Inside a small glass vial equipped with a magnetic stirring bar, the following components were mixed in an 800 µL of solution: 0.5 mM *O*-alkylated compound, 5 mM G6P-Na, 2 mM NADP-Na$_2$, 1X cruzain buffer (50 mM MES, 50 mM TAPSO, 1 mM EDTA-Na$_2$, 1mM CHAPS), 2.5% DMSO (v/v), pH 7.5. For the experimental samples, 64

mg of CypExpress™ P450 lyophilized powder was added to the reaction vial (80 mg/mL), which was gently shaken to homogenize the solid, and immediately placed on a magnetic mixer in a 30 °C incubator. The vial was capped loosely with foil to allow aeration as well as to reduce evaporation. Additionally, a negative control without addition of CypExpress™ P450, and a blank control with only buffer were set up in the meantime. Each aliquot (65 µL) was taken from the reaction solution at 0, 30, 60, 120, 180, and 240 min, and was quenched with 65 µL of MeCN. The resulting solution was vortexed and centrifuged to precipitate the protein, and the supernatant was analyzed on HPLC-MS under UV 270 nm. The fraction of intact *O*-alkylated compound was plotted versus reaction time.

3.3.9 X-ray Crystallography

SARS-CoV-2 3CL^{pro} was recombinantly expressed, purified and concentrated to 10 mg/mL.[226, 231] Crystallization conditions were initially screened against several commercially available kits (Index™, Crystal Screen™, Crystal Screen™ 2, PEGRx™ 1, PEGRx™ 2, PEG/Ion™, and PEG/Ion™ 2) at 18 °C with the sitting-drop method. A droplet containing 1.0 µL of 3CL^{pro} and 1.0 µL of reservoir solution was allowed to equilibrate against 100 µL of reservoir solution. The most promising condition yielding crystals (PEG/Ion™ tube 44: 0.2 M ammonium phosphate dibasic, 20% PEG3350 (w/v), pH 8.0) was further optimized in terms of temperature and concentration of precipitant and protein. The best crystals were obtained with 14 mg/ml of 3CL^{pro} in 0.2 M ammonium phosphate dibasic, 17% PEG3350 (w/v), pH 8.0, at 25 °C, and were grown overnight under this condition. Compound **19** stock was diluted into a specific volume of

reservoir solution to prepare a 0.5 mM inhibitor solution containing 2% DMSO (v/v). On the next day, these crystals were washed successively with reservoir solution three times and with the prepared inhibitor solution three times. The resulting mixture was incubated at 25°C for 2 d. The cryoprotectant solution contained reservoir solution plus 30% glycerol, 0.5 mM inhibitor and 2% DMSO. Cryo-protected crystals were harvested for data collection.

The 3CL^{pro}-**19/23** data were collected to 1.70 Å on a Rigaku R-AXIS IV++ image plate detector and at the Advanced Light Source beamline 5.0.2 using a PITALUS3 S 6M detector. Reflections were indexed, integrated, scaled, and merged in space group C121 using iMosflm.[232] The structures were solved by molecular replacement in the Phaser module of the Phenix package using a high-resolution structure of the apo SARS-CoV-2 3CL^{pro} (PDB accession code: 6Y2E) as the search model.[231, 233] The atomic coordinates and geometric restraints for the inhibitors were generated in CCP4 suite.[232] The inhibitors were built into the F_o-F_c electron density omit maps using Coot.[234] Structural refinement was performed with Real-space Refinement module of the Phenix package.

3.3.10 Others

The evaluation method of SMAIs in axenic culture of *T. b. brucei*, and in *T. cruzi*-infected murine cardiomyoblasts, and the experimental information of molecular modeling, can be entirely found in Section 2.3.6, 2.3.7 and 2.3.9, respectively.

3.4 Conclusions

Taking advantage of a simple peptidyl aldehyde that is highly potent, we designed and elaborated a self-masking strategy in hope of overcoming the obvious drawbacks of the aldehyde. A series of SMAIs were synthesized with a promising inhibitor profile in terms of anti-cruzain activity, trypanocidal activity, and reversibility. The substituents on the phenol ring also offered infinite opportunities for fine-tuning of the lactol open/closure. By means of 2D NMR and synthetically incorporation of ^{13}C into compound **12**, we demonstrated that SMAIs functioned as "quiescent affinity labels" in that the lactol ring remained closed in buffer but readily opened up and reacted with the catalytic cysteine upon addition of target cruzain. This feature might keep the free aldehyde from unwanted metabolism, but also prevent its unspecific binding to induce toxicity or immune response. The attempt to derivatize the lactol hydroxyl group resulted in *O*-acylated SMAIs that demonstrated potential to be prodrugs. We also explored the use of this strategy for inhibiting 3CL^{pro} of SARS-CoV-2. A class of 2-pyridone compounds were prepared, one of which (**19**) showed excellent inhibitory activity. Its binding mode was also illustrated by virtue of X-ray crystallography. These results collectively showed the promise and versatility of SMAI in the field of protease inhibitor development.

CHAPTER 4

SUMMARY AND OUTLOOK

Human trypanosomiasis, especially Chagas disease, are threatening millions of people living in poverty. The cysteine proteases of the pathogenic parasites have been investigated extensively and are promising therapeutic targets of anti-trypanosomal drugs. Our group has been committed to the development of cruzain inhibitor as potential treatment for Chagas disease.

Classical Michael acceptors are commonly used as inactivators of proteases. **K11777** is such a Michael acceptor that effectively inactivates cruzain and shows potent anti-*T. cruzi* activity. The high reactivity of activated double bond can raise concern about modification of unspecific targets. By replacing the sulfonyl group with a diversity of heterocycles, a class of peptidomimetic inhibitors with less electrophilic warheads (PVHIs) were proposed to allow reversible, covalent inhibition of cruzain, which could result in relief from covalent inactivation. Covalent molecular modeling was employed to serve as a rapid validation of the design in terms of basic shape complementarity. Next, the electrophilicity of vinyl-heterocycles were evaluated in model reactions with glutathione, preliminarily demonstrating the possibility to tune the electronic properties of the double bond. Kinetic evaluation and analysis of a number of PVHIs proved many of them to be potent, time-dependent, and fully reversible cruzain inhibitors with a complex structure-activity relationship. Some of the PVHIs presented considerable anti-trypanosomal activity not only in axenic cultures of *T. cruzi* or *T. b. brucei*, but also in an infection model with murine cardiomyoblasts. On the other hand, the inherent complexity

of heterocycles also brings about multiple challenges. The fine-tuning of PVHIs is quite intractable because a minor change in the substituent can result in different electronic and steric properties, as well as altered metabolic site or route of the heterocycle.

The other class of cruzain inhibitors in this dissertation was inspired by the extraordinary potency of peptidomimetic aldehydes vs. cruzain. Masking the highly reactive aldehyde with well-positioned P_1 phenol group produced the hemiacetal SMAIs. In like fashion with the development of PVHIs, the SMAIs were preliminarily validated by computational study, and then kinetically characterized. These molecules appeared to be potent, rapidly reversible inhibitors. The phenol group might play a critical role in the expedited reversibility relative to a free aldehyde. 2D NMR analysis in aqueous solution demonstrated the "quiescent affinity labeling" property of a ^{13}C-enriched SMAI. Furthermore, acylation of the lactol hydroxyl group yielded a class of "pro-prodrug" molecules, which released the free aldehyde upon addition of esterase. These derivatized SMAIs were also effective in axenic *T. b. brucei* culture, suggesting the possibility of their *in vivo* efficacy. Such intramolecular machinery was also applied to the main cysteine protease of SARS-CoV-2 by making a 2-pyridone-based P_1 position. Crystallographic study indicated the equivalence of the 2-pyridone and the γ/δ-lactam, the latter being almost an invariant moiety in 3C/3CL^{pro} inhibitors. For future work, whether or not the 2-pyridone aldehyde fulfills a closed structure of SMAI, the *O*-acylation of the tautomer of 2-pyridone inhibitors can be easily attempted. Considering the high potency of various peptidyl aldehydes for 3CL^{pro}, these inhibitors may be further developed as promising prodrugs for COVID-19.

REFERENCES

1.   Global spending on health: a world in transition. Geneva: World Health Organization; 2019 (WHO/HIS/HGF/HFWorkingPaper/19.4). Licence: CC BY-NC-SA 3.0 IGO.

2.   Hotez, P. J.; Aksoy, S.; Brindley, P. J.; Kamhawi, S., What constitutes a neglected tropical disease? *PLoS Negl. Trop. Dis.* **2020,** *14* (1), e0008001.

3.   Control of Neglected Tropical Diseases. https://www.who.int/teams/control-of-neglected-tropical-diseases (accessed Oct 10, 2020).

4.   Keenan, J. D.; Hotez, P. J.; Amza, A.; Stoller, N. E.; Gaynor, B. D.*, et al.*, Elimination and eradication of neglected tropical diseases with mass drug administrations: a survey of experts. *PLoS Negl. Trop. Dis.* **2013,** *7* (12), e2562.

5.   Taleo, F.; Taleo, G.; Graves, P. M.; Wood, P.; Kim, S. H.*, et al.*, Surveillance efforts after mass drug administration to validate elimination of lymphatic filariasis as a public health problem in Vanuatu. *Trop. Med. Health* **2017,** *45*, 18.

6.   Chagas disease (American trypanosomiasis). https://www.who.int/health-topics/chagas-disease#tab=tab_1 (accessed Aug 22, 2020).

7.   Disease, G. B. D.; Injury, I.; Prevalence, C., Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018,** *392* (10159), 1789-1858.

8.   Collaborators, G. B. D. C. o. D., Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018,** *392* (10159), 1736-1788.

9.   Angheben, A.; Boix, L.; Buonfrate, D.; Gobbi, F.; Bisoffi, Z.*, et al.*, Chagas disease and transfusion medicine: a perspective from non-endemic countries. *Blood Transfus.* **2015,** *13* (4), 540-550.

10.   Rassi, A.; Rassi, A.; Marin-Neto, J. A., Chagas disease. *The Lancet* **2010,** *375* (9723), 1388-1402.

11.   Lidani, K. C. F.; Andrade, F. A.; Bavia, L.; Damasceno, F. S.; Beltrame, M. H.*, et al.*, Chagas Disease: From Discovery to a Worldwide Health Problem. *Frontiers in Public Health* **2019,** *7* (166).

12.   Bern, C.; Montgomery, S. P., An Estimate of the Burden of Chagas Disease in the United States. *Clin. Infect. Dis.* **2009,** *49* (5), e52-e54.

13.   Strasen, J.; Williams, T.; Ertl, G.; Zoller, T.; Stich, A.*, et al.*, Epidemiology of Chagas disease in Europe: many calculations, little knowledge. *Clin. Res. Cardiol.* **2014,** *103* (1), 1-10.

14.   Prata, A., Clinical and epidemiological aspects of Chagas disease. *The Lancet Infectious Diseases* **2001,** *1* (2), 92-100.

15.   Vago, A. R.; Andrade, L. O.; Leite, A. A.; d'Ávila Reis, D.; Macedo, A. M.*, et al.*, Genetic Characterization of Trypanosoma cruzi Directly from Tissues of Patients with Chronic Chagas Disease: Differential Distribution of Genetic Types into Diverse Organs. *The American Journal of Pathology* **2000,** *156* (5), 1805-1809.

16.   Pinazo, M. J.; Cañas, E.; Elizalde, J. I.; García, M.; Gascón, J.*, et al.*, Diagnosis, management and treatment of chronic Chagas' gastrointestinal disease in areas where Trypanosoma cruzi infection is not endemic. *Gastroenterol. Hepatol.* **2010,** *33* (3), 191-200.

17.   Matsuda, N. M.; Oliveira, R. B.; Dantas, R. O.; Iazigi, N., Effect of isosorbide dinitrate on gastroesophageal reflux in healthy volunteers and patients with Chagas' disease. *Dig. Dis. Sci.* **1995,** *40* (1), 177-182.

18.   Carod-Artal, F. J.; Gascon, J., Chagas disease and stroke. *The Lancet Neurology* **2010,** *9* (5), 533-542.

19.   Mougabure-Cueto, G.; Picollo, M. I., Insecticide resistance in vector Chagas disease: evolution, mechanisms and management. *Acta Trop.* **2015,** *149*, 70-85.

20.   Guhl, F.; Pinto, N.; Aguilera, G., Sylvatic triatominae: a new challenge in vector control transmission. *Mem. Inst. Oswaldo Cruz* **2009,** *104 Suppl 1*, 71-75.

21.   Garraud, O.; Andreu, G.; Elghouzzi, M. H.; Laperche, S.; Lefrere, J. J., Measures to prevent transfusion-associated protozoal infections in non-endemic countries. *Travel Med. Infect. Dis.* **2007,** *5* (2), 110-112.

22.   Benjamin, R. J.; Stramer, S. L.; Leiby, D. A.; Dodd, R. Y.; Fearon, M*., et al.*, Trypanosoma cruzi infection in North America and Spain: evidence in support of transfusion transmission. *Transfusion* **2012,** *52* (9), 1913-1921; quiz 1912.

23.   Castro, E., Chagas' disease: lessons from routine donation testing. *Transfus. Med.* **2009,** *19* (1), 16-23.

24.   Dumonteil, E.; Herrera, C.; Buekens, P., A therapeutic preconceptional vaccine against Chagas disease: A novel indication that could reduce congenital transmission and accelerate vaccine development. *PLoS Negl. Trop. Dis.* **2019,** *13* (1), e0006985.

25.   Dumonteil, E.; Bottazzi, M. E.; Zhan, B.; Heffernan, M. J.; Jones, K*., et al.*, Accelerating the development of a therapeutic vaccine for human Chagas disease: rationale and prospects. *Expert Review of Vaccines* **2012,** *11* (9), 1043-1055.

26.   Dumonteil, E., Vaccine development against Trypanosoma cruzi and Leishmania species in the post-genomic era. *Infect. Genet. Evol.* **2009,** *9* (6), 1075-1082.

27.   Pereira, I. R.; Vilar-Pereira, G.; Marques, V.; da Silva, A. A.; Caetano, B*., et al.*, A Human Type 5 Adenovirus-Based Trypanosoma cruzi Therapeutic Vaccine Re-programs Immune Response and Reverses Chronic Cardiomyopathy. *PLoS Pathog.* **2015,** *11* (1), e1004594.

28.   Dumonteil, E.; Escobedo-Ortegon, J.; Reyes-Rodriguez, N.; Arjona-Torres, A.; Ramirez-Sierra, M. J., Immunotherapy of Trypanosoma cruzi Infection with DNA Vaccines in Mice. *Infect. Immun.* **2004,** *72* (1), 46-53.

29.   Sanchez-Valdez, F. J.; Perez Brandan, C.; Ferreira, A.; Basombrio, M. A., Gene-deleted live-attenuated Trypanosoma cruzi parasites as vaccines to protect against Chagas disease. *Expert Rev Vaccines* **2015,** *14* (5), 681-697.

30.   Gupta, S.; Garg, N. J., A Two-Component DNA-Prime/Protein-Boost Vaccination Strategy for Eliciting Long-Term, Protective T Cell Immunity against Trypanosoma cruzi. *PLoS Pathog.* **2015,** *11* (5), e1004828.

31.   Bern, C.; Messenger, L. A.; Whitman, J. D.; Maguire, J. H., Chagas Disease in the United States: a Public Health Approach. *Clin. Microbiol. Rev.* **2019,** *33* (1), e00023-00019.

32.   Sosa Estani, S.; Segura, E. L., Treatment of Trypanosoma cruzi infection in the undetermined phase. Experience and current guidelines of treatment in Argentina. *Mem. Inst. Oswaldo Cruz* **1999,** *94 Suppl 1*, 363-365.

33.   Sosa Estani, S.; Segura, E. L.; Ruiz, A. M.; Velazquez, E.; Porcel, B. M.*, et al.*, Efficacy of chemotherapy with benznidazole in children in the indeterminate phase of Chagas' disease. *Am. J. Trop. Med. Hyg.* **1998,** *59* (4), 526-529.

34.   de Andrade, A. L.; Zicker, F.; de Oliveira, R. M.; Almeida Silva, S.; Luquetti, A.*, et al.*, Randomised trial of efficacy of benznidazole in treatment of early Trypanosoma cruzi infection. *Lancet* **1996,** *348* (9039), 1407-1413.

35.   Silveira, C. A. N.; Castillo, E.; Castro, C., Avaliação do tratamento específico para o Trypanosoma cruzi em crianças, na evolução da fase indeterminada. *Rev. Soc. Bras. Med. Trop.* **2000,** *33*, 191-196.

36.   Carrilero, B.; Murcia, L.; Martinez-Lage, L.; Segovia, M., Side effects of benznidazole treatment in a cohort of patients with Chagas disease in non-endemic country. *Rev. Esp. Quimioter.* **2011,** *24* (3), 123-126.

37.   Docampo, R., Sensitivity of parasites to free radical damage by antiparasitic drugs. *Chem. Biol. Interact.* **1990,** *73* (1), 1-27.

38.   Olivera, M. J.; Cucunubá, Z. M.; Álvarez, C. A.; Nicholls, R. S., Safety Profile of Nifurtimox and Treatment Interruption for Chronic Chagas Disease in Colombian Adults. *The American Journal of Tropical Medicine and Hygiene* **2015,** *93* (6), 1224-1230.

39.   Trypanosomiasis,   human   African   (sleeping   sickness). https://www.who.int/en/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness) (accessed Feb 17, 2020).

40.   Fevre, E. M.; Coleman, P. G.; Welburn, S. C.; Maudlin, I., Reanalyzing the 1900-1920 sleeping sickness epidemic in Uganda. *Emerg. Infect. Dis.* **2004,** *10* (4), 567-573.

41.   Cooper, A. In *Linkage mapping and genetic analysis of Trypanosoma brucei*, 2010.

42.	Stephens, N. A.; Kieft, R.; MacLeod, A.; Hajduk, S. L., Trypanosome resistance to human innate immunity: targeting Achilles' heel. *Trends in Parasitology* **2012,** *28* (12), 539-545.

43.	Njiokou, F.; Nimpaye, H.; Simo, G.; Njitchouang, G. R.; Asonganyi, T.*, et al.*, Domestic animals as potential reservoir hosts of Trypanosoma brucei gambiense in sleeping sickness foci in Cameroon. *Parasite* **2010,** *17* (1), 61-66.

44.	Sternberg, J. M., Human African trypanosomiasis: clinical presentation and immune response. *Parasite Immunol.* **2004,** *26* (11‑12), 469-476.

45.	Ormerod, W. E., Hypothesis: the significance of Winterbottom's sign. *J. Trop. Med. Hyg.* **1991,** *94* (5), 338-340.

46.	Jamonneau, V.; Ravel, S.; Garcia, A.; Koffi, M.; Truc, P.*, et al.*, Characterization of Trypanosoma brucei s.l. infecting asymptomatic sleeping-sickness patients in Côte d'Ivoire: a new genetic group? *Ann. Trop. Med. Parasitol.* **2004,** *98* (4), 329-337.

47.	Berriman, M.; Ghedin, E.; Hertz-Fowler, C.; Blandin, G.; Renauld, H.*, et al.*, The genome of the African trypanosome Trypanosoma brucei. *Science* **2005,** *309* (5733), 416-422.

48.	Franco, J. R.; Simarro, P. P.; Diarra, A.; Jannin, J. G., Epidemiology of human African trypanosomiasis. *Clin. Epidemiol.* **2014,** *6*, 257-275.

49.	Vreysen, M. J. B.; Saleh, K. M.; Ali, M. Y.; Abdulla, A. M.; Zhu, Z.-R.*, et al.*, Glossina austeni (Diptera: Glossinidae) Eradicated on the Island of Unguja, Zanzibar, Using the Sterile Insect Technique. *J. Econ. Entomol.* **2000,** *93* (1), 123-135.

50.	Büscher, P.; Cecchi, G.; Jamonneau, V.; Priotto, G., Human African trypanosomiasis. *The Lancet* **2017,** *390* (10110), 2397-2409.

51.	Pohlig, G.; Bernhard, S. C.; Blum, J.; Burri, C.; Mpanya, A.*, et al.*, Efficacy and Safety of Pafuramidine versus Pentamidine Maleate for Treatment of First Stage Sleeping Sickness in a Randomized, Comparator-Controlled, International Phase 3 Clinical Trial. *PLoS Negl. Trop. Dis.* **2016,** *10* (2), e0004363.

52.	World Health Organization Model List of Essential Medicines, 21st List, 2019. Geneva: World Health Organization; 2019. Licence: CC BY-NC-SA 3.0 IGO.

53.   Piontkivska, H.; Hughes, A. L., Environmental kinetoplastid-like 18S rRNA sequences and phylogenetic relationships among Trypanosomatidae: Paraphyly of the genus Trypanosoma. *Mol. Biochem. Parasitol.* **2005,** *144* (1), 94-99.

54.   Hoare, C. A., The classification of mammalian trypanosomes. In *Ergebnisse der Mikrobiologie Immunitätsforschung und Experimentellen Therapie: Fortsetzung der Ergebnisse der Hygiene Bakteriologie · Immunitätsforschung und Experimentellen Therapie Begründet von Wolfgang Weichardt*, Henle, W.; Kikuth, W.; Meyer, K. F.; Nauck, E. G.; Tomcsik, J., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 1966; pp 43-57.

55.   Tyler, K. M.; Engman, D. M., The life cycle of Trypanosoma cruzi revisited. *Int. J. Parasitol.* **2001,** *31* (5), 472-481.

56.   Castro, D. P.; Seabra, S. H.; Garcia, E. S.; Souza, W. d.; Azambuja, P., Trypanosoma cruzi: Ultrastructural studies of adhesion, lysis and biofilm formation by Serratia marcescens. *Exp. Parasitol.* **2007,** *117* (2), 201-207.

57.   Nogueira, N. F. S.; Gonzalez, M. S.; Gomes, J. E.; de Souza, W.; Garcia, E. S., *et al.*, Trypanosoma cruzi: Involvement of glycoinositolphospholipids in the attachment to the luminal midgut surface of Rhodnius prolixus. *Exp. Parasitol.* **2007,** *116* (2), 120-128.

58.   Rotureau, B.; Subota, I.; Buisson, J.; Bastin, P., A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. *Development* **2012,** *139* (10), 1842-1850.

59.   Oladiran, A.; Belosevic, M., Immune evasion strategies of trypanosomes: a review. *J. Parasitol.* **2012,** *98* (2), 284-292.

60.   Buscaglia, C. A.; Campo, V. A.; Frasch, A. C. C.; Di Noia, J. M., Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology* **2006,** *4* (3), 229-236.

61.   Norris, K. A.; Schrimpf, J. E.; Szabo, M. J., Identification of the gene family encoding the 160-kilodalton Trypanosoma cruzi complement regulatory protein. *Infect. Immun.* **1997,** *65* (2), 349-357.

62.   Norris, K. A.; Bradt, B.; Cooper, N. R.; So, M., Characterization of a Trypanosoma cruzi C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. *The Journal of Immunology* **1991,** *147* (7), 2240-2247.

63.   Taylor, J. E.; Rudenko, G., Switching trypanosome coats: what's in the wardrobe? *Trends Genet.* **2006,** *22* (11), 614-620.

64.   Barrett, M. P.; Burchmore, R. J.; Stich, A.; Lazzari, J. O.; Frasch, A. C*., et al.*, The trypanosomiases. *Lancet* **2003,** *362* (9394), 1469-1480.

65.   Vaidya, T.; Bakhiet, M.; Hill, K. L.; Olsson, T.; Kristensson, K*., et al.*, The gene for a T lymphocyte triggering factor from African trypanosomes. *J. Exp. Med.* **1997,** *186* (3), 433-438.

66.   Olsson, T.; Bakhiet, M.; Edlund, C.; Hojeberg, B.; Van der Meide, P. H*., et al.*, Bidirectional activating signals between Trypanosoma brucei and CD8+ T cells: a trypanosome-released factor triggers interferon-gamma production that stimulates parasite growth. *Eur. J. Immunol.* **1991,** *21* (10), 2447-2454.

67.   Samudio, M.; Montenegro-James, S.; Cabral, M.; Martinez, J.; Rojas de Arias, A*., et al.*, Cytokine responses in Trypanosoma cruzi-infected children in Paraguay. *Am. J. Trop. Med. Hyg.* **1998,** *58* (1), 119-121.

68.   McKerrow JH, S.-N. J., McCall L-I, Otrubova K, Twenty Five Years of Drug Discovery at the University of California Targeting Kinetoplastid Parasites. *J Pharm Drug Devel* **2016,** *4* (1), 101.

69.   Altamura, F.; Rajesh, R.; Catta-Preta, C. M. C.; Moretti, N. S.; Cestari, I., The current drug discovery landscape for trypanosomiasis and leishmaniasis: Challenges and strategies to identify drug targets. *Drug Dev Res* **2020,** *n/a* (n/a).

70.   Urbina, J. A.; Payares, G.; Contreras, L. M.; Liendo, A.; Sanoja, C*., et al.*, Antiproliferative effects and mechanism of action of SCH 56592 against Trypanosoma (Schizotrypanum) cruzi: in vitro and in vivo studies. *Antimicrob. Agents Chemother.* **1998,** *42* (7), 1771-1777.

71.   Molina, J.; Martins-Filho, O.; Brener, Z.; Romanha, A. J.; Loebenberg, D*., et al.*, Activities of the Triazole Derivative SCH 56592 (Posaconazole) against Drug-Resistant Strains of the Protozoan ParasiteTrypanosoma (Schizotrypanum) cruzi in Immunocompetent and Immunosuppressed Murine Hosts. *Antimicrob. Agents Chemother.* **2000,** *44* (1), 150-155.

72.	Molina, I.; Gómez i Prat, J.; Salvador, F.; Treviño, B.; Sulleiro, E.*, et al.*, Randomized Trial of Posaconazole and Benznidazole for Chronic Chagas' Disease. *N. Engl. J. Med.* **2014,** *370* (20), 1899-1908.

73.	Morillo, C. A.; Waskin, H.; Sosa-Estani, S.; del Carmen Bangher, M.; Cuneo, C.*, et al.*, Benznidazole and Posaconazole in Eliminating Parasites in Asymptomatic T. Cruzi Carriers: The STOP-CHAGAS Trial. *J. Am. Coll. Cardiol.* **2017,** *69* (8), 939-947.

74.	Azoles E1224. https://dndi.org/research-development/portfolio/azoles-e1224/ (accessed Sep 30, 2014).

75.	Lourenço, A. M.; Faccini, C. C.; Costa, C. A. d. J.; Mendes, G. B.; Fragata Filho, A. A., Evaluation of in vitro anti-Trypanosoma cruzi activity of medications benznidazole, amiodarone hydrochloride, and their combination. *Rev. Soc. Bras. Med. Trop.* **2018,** *51*, 52-56.

76.	A Trial Testing Amiodarone in Chagas Cardiomiopathy. https://ClinicalTrials.gov/show/NCT03193749.

77.	Ding, D.; Zhao, Y.; Meng, Q.; Xie, D.; Nare, B.*, et al.*, Discovery of Novel Benzoxaborole-Based Potent Antitrypanosomal Agents. *ACS Med. Chem. Lett.* **2010,** *1* (4), 165-169.

78.	Wall, R. J.; Rico, E.; Lukac, I.; Zuccotto, F.; Elg, S.*, et al.*, Clinical and veterinary trypanocidal benzoxaboroles target CPSF3. *Proceedings of the National Academy of Sciences* **2018,** *115* (38), 9616-9621.

79.	Nagle, A. S.; Khare, S.; Kumar, A. B.; Supek, F.; Buchynskyy, A.*, et al.*, Recent Developments in Drug Discovery for Leishmaniasis and Human African Trypanosomiasis. *Chem. Rev.* **2014,** *114* (22), 11305-11347.

80.	Jacobs, R. T.; Nare, B.; Wring, S. A.; Orr, M. D.; Chen, D.*, et al.*, SCYX-7158, an Orally-Active Benzoxaborole for the Treatment of Stage 2 Human African Trypanosomiasis. *PLoS Negl. Trop. Dis.* **2011,** *5* (6), e1151.

81.	Cestari, I.; Anupama, A.; Stuart, K., Inositol polyphosphate multikinase regulation of Trypanosoma brucei life stage development. *Mol. Biol. Cell* **2018,** *29* (9), 1137-1152.

82.   Cestari, I.; Haas, P.; Moretti, Nilmar S.; Schenkman, S.; Stuart, K., Chemogenetic Characterization of Inositol Phosphate Metabolic Pathway Reveals Druggable Enzymes for Targeting Kinetoplastid Parasites. *Cell Chemical Biology* **2016,** *23* (5), 608-617.

83.   Sueth-Santiago, V.; Decote-Ricardo, D.; Morrot, A.; Freire-de-Lima, C. G.; Lima, M. E., Challenges in the chemotherapy of Chagas disease: Looking for possibilities related to the differences and similarities between the parasite and host. *World J. Biol. Chem.* **2017,** *8* (1), 57-80.

84.   Lepesheva, G. I.; Ott, R. D.; Hargrove, T. Y.; Kleshchenko, Y. Y.; Schuster, I.*, et al.*, Sterol 14α-Demethylase as a Potential Target for Antitrypanosomal Therapy: Enzyme Inhibition and Parasite Cell Growth. *Chem. Biol.* **2007,** *14* (11), 1283-1293.

85.   Urbina, J. A.; Concepcion, J. L.; Rangel, S.; Visbal, G.; Lira, R., Squalene synthase as a chemotherapeutic target in Trypanosoma cruzi and Leishmania mexicana. *Mol. Biochem. Parasitol.* **2002,** *125* (1), 35-45.

86.   Agustí, R.; París, G.; Ratier, L.; Frasch, A. C. C.; de Lederkremer, R. M., Lactose derivatives are inhibitors of Trypanosoma cruzi trans-sialidase activity toward conventional substrates in vitro and in vivo. *Glycobiology* **2004,** *14* (7), 659-670.

87.   Buschiazzo, A.; Amaya, M. F.; Cremona, M. L.; Frasch, A. C.; Alzari, P. M., The crystal structure and mode of action of trans-sialidase, a key enzyme in Trypanosoma cruzi pathogenesis. *Mol. Cell* **2002,** *10* (4), 757-768.

88.   Lo Presti, M. S.; Bazán, P. C.; Strauss, M.; Báez, A. L.; Rivarola, H. W.*, et al.*, Trypanothione reductase inhibitors: Overview of the action of thioridazine in different stages of Chagas disease. *Acta Trop.* **2015,** *145*, 79-87.

89.   Benson, T. J.; McKie, J. H.; Garforth, J.; Borges, A.; Fairlamb, A. H.*, et al.*, Rationally designed selective inhibitors of trypanothione reductase. Phenothiazines and related tricyclics as lead structures. *Biochem. J.* **1992,** *286* (1), 9-11.

90.   Phan, I. Q. H.; Davies, D. R.; Moretti, N. S.; Shanmugam, D.; Cestari, I.*, et al.*, Iron superoxide dismutases in eukaryotic pathogens: new insights from Apicomplexa and Trypanosoma structures. *Acta Crystallographica Section F* **2015,** *71* (5), 615-621.

91.   Moreno-Viguri, E.; Jiménez-Montes, C.; Martín-Escolano, R.; Santivañez-Veliz, M.; Martin-Montes, A.*, et al.*, In Vitro and in Vivo Anti-Trypanosoma cruzi Activity of

New Arylamine Mannich Base-Type Derivatives. *J. Med. Chem.* **2016,** *59* (24), 10929-10945.

92.   Britta, E. A.; Scariot, D. B.; Falzirolli, H.; Da Silva, C. C.; Ueda-Nakamura, T.*, et al.*, 4-Nitrobenzaldehyde thiosemicarbazone: a new compound derived from S-(-)-limonene that induces mitochondrial alterations in epimastigotes and trypomastigotes of Trypanosoma cruzi. *Parasitology* **2015,** *142* (7), 978-988.

93.   de Azeredo, C. M. O.; Ávila, E. P.; Pinheiro, D. L. J.; Amarante, G. W.; Soares, M. J., Biological activity of the azlactone derivative EPA-35 against Trypanosoma cruzi. *FEMS Microbiol. Lett.* **2017,** *364* (4).

94.   Cortes, L. A.; Castro, L.; Pesce, B.; Maya, J. D.; Ferreira, J.*, et al.*, Novel Gallate Triphenylphosphonium Derivatives with Potent Antichagasic Activity. *PLoS One* **2015,** *10* (8), e0136852.

95.   Merritt, C.; Silva, L. E.; Tanner, A. L.; Stuart, K.; Pollastri, M. P., Kinases as Druggable Targets in Trypanosomatid Protozoan Parasites. *Chem. Rev.* **2014,** *114* (22), 11280-11304.

96.   Parsons, M.; Worthey, E. A.; Ward, P. N.; Mottram, J. C., Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics* **2005,** *6* (1), 127.

97.   Parsons, M.; Valentine, M.; Deans, J.; Schieven, G. L.; Ledbetter, J. A., Distinct patterns of tyrosine phosphorylation during the life cycle of Trypanosoma brucei. *Mol. Biochem. Parasitol.* **1991,** *45* (2), 241-248.

98.   Fernandez-Cortes, F.; Serafim, T. D.; Wilkes, J. M.; Jones, N. G.; Ritchie, R.*, et al.*, RNAi screening identifies Trypanosoma brucei stress response protein kinases required for survival in the mouse. *Sci. Rep.* **2017,** *7* (1), 6156.

99.   Urbaniak, M. D.; Mathieson, T.; Bantscheff, M.; Eberhard, D.; Grimaldi, R.*, et al.*, Chemical Proteomic Analysis Reveals the Drugability of the Kinome of Trypanosoma brucei. *ACS Chem. Biol.* **2012,** *7* (11), 1858-1865.

100.   Abbenante, G.; Fairlie, D. P., Protease inhibitors in the clinic. *Med. Chem.* **2005,** *1* (1), 71-104.

101.   Rawlings, N. D.; Barrett, A. J.; Thomas, P. D.; Huang, X.; Bateman, A.*, et al.*, The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2017,** *46* (D1), D624-D632.

102.   El-Sayed, N. M.; Myler, P. J.; Bartholomeu, D. C.; Nilsson, D.; Aggarwal, G.*, et al.*, The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. *Science* **2005,** *309* (5733), 409-415.

103.   de A. Lima, A. P. C.; Tessier, D. C.; Thomas, D. Y.; Scharfstein, J.; Storer, A. C.*, et al.*, Identification of new cysteine protease gene isoforms in Trypanosoma cruzi. *Mol. Biochem. Parasitol.* **1994,** *67* (2), 333-338.

104.   Eakin, A. E.; Mills, A. A.; Harth, G.; McKerrow, J. H.; Craik, C. S., The sequence, organization, and expression of the major cysteine protease (cruzain) from Trypanosoma cruzi. *J. Biol. Chem.* **1992,** *267* (11), 7411-7420.

105.   Fampa, P.; Lisboa, C. V.; Jansen, A. M.; Santos, A. L.; Ramirez, M. I., Protease expression analysis in recently field-isolated strains of Trypanosoma cruzi: a heterogeneous profile of cysteine protease activities between TC I and TC II major phylogenetic groups. *Parasitology* **2008,** *135* (9), 1093-1100.

106.   Lima, A. P.; dos Reis, F. C.; Serveau, C.; Lalmanach, G.; Juliano, L.*, et al.*, Cysteine protease isoforms from Trypanosoma cruzi, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. *Mol. Biochem. Parasitol.* **2001,** *114* (1), 41-52.

107.   Tomás, A. M.; Kelly, J. M., Stage-regulated expression of cruzipain, the major cysteine protease of Trypanosoma cruzi is independent of the level of RNA1. *Mol. Biochem. Parasitol.* **1996,** *76* (1-2), 91-103.

108.   Mackey, Z. B.; O'Brien, T. C.; Greenbaum, D. C.; Blank, R. B.; McKerrow, J. H., A Cathepsin B-like Protease Is Required for Host Protein Degradation in Trypanosoma brucei. *J. Biol. Chem.* **2004,** *279* (46), 48426-48433.

109.   Nóbrega, O. T.; Santos Silva, M. A.; Teixeira, A. R. L.; Santana, J. M., Cloning and sequencing of tccb, a gene encoding a Trypanosoma cruzi cathepsin B-like protease1Note: Nucleotide sequence data reported in this paper are available in the GenBank™ data base under the accession number AF043246.1. *Mol. Biochem. Parasitol.* **1998,** *97* (1), 235-240.

110. Duschak, V. G.; Couto, A. S., Cruzipain, the major cysteine protease of Trypanosoma cruzi: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. A review. *Curr. Med. Chem.* **2009,** *16* (24), 3174-3202.

111. Barboza, M.; Duschak, V. G.; Cazzulo, J. J.; de Lederkremer, R. M.; Couto, A. S., Presence of sialic acid in N-linked oligosaccharide chains and O-linked N-acetylglucosamine in cruzipain, the major cysteine proteinase of Trypanosoma cruzi. *Mol. Biochem. Parasitol.* **2003,** *126* (2), 293-296.

112. Turk, V.; Stoka, V.; Vasiljeva, O.; Renko, M.; Sun, T.*, et al.*, Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim. Biophys. Acta* **2012,** *1824* (1), 68-88.

113. Huete-Pérez, J. A.; Engel, J. C.; Brinen, L. S.; Mottram, J. C.; McKerrow, J. H., Protease Trafficking in Two Primitive Eukaryotes Is Mediated by a Prodomain Protein Motif. *J. Biol. Chem.* **1999,** *274* (23), 16249-16256.

114. Lalmanach, G.; Lecaille, F.; Chagas, J. R.; Authié, E.; Scharfstein, J.*, et al.*, Inhibition of trypanosomal cysteine proteinases by their propeptides. *J. Biol. Chem.* **1998,** *273* (39), 25112-25116.

115. Engel, J. C.; Doyle, P. S.; Palmer, J.; Hsieh, I.; Bainton, D. F.*, et al.*, Cysteine protease inhibitors alter Golgi complex ultrastructure and function in Trypanosoma cruzi. *J. Cell Sci.* **1998,** *111* (5), 597-606.

116. Martinez, J.; Campetella, O.; Frasch, A. C.; Cazzulo, J. J., The major cysteine proteinase (cruzipain) from Trypanosoma cruzi is antigenic in human infections. *Infect. Immun.* **1991,** *59* (11), 4275-4277.

117. Tazeh, N. N.; Silverman, J. S.; Schwartz, K. J.; Sevova, E. S.; Sutterwala, S. S.*, et al.*, Role of AP-1 in Developmentally Regulated Lysosomal Trafficking in Trypanosoma brucei. *Eukaryotic Cell* **2009,** *8* (9), 1352-1361.

118. Tomas, A. M.; Miles, M. A.; Kelly, J. M., Overexpression of cruzipain, the major cysteine proteinase of Trypanosoma cruzi, is associated with enhanced metacyclogenesis. *Eur. J. Biochem.* **1997,** *244* (2), 596-603.

119. Franke de Cazzulo, B. M.; Martínez, J.; North, M. J.; Coombs, G. H.; Cazzulo, J. J., Effects of proteinase inhibitors on the growth and differentiation of Trypanosoma cruzi. *FEMS Microbiol. Lett.* **1994,** *124* (1), 81-86.

120.   Araujo-Jorge, T. C.; Lage, M.-J. F.; Rivera, M. T.; Carlier, Y.; Van Leuven, F., Trypanosoma cruzi: Enhanced alpha-macroglobulin levels correlate with the resistance of BALB/cj mice to acute infection. *Parasitol. Res.* **1992,** *78* (3), 215-221.

121.   de Araujo-Jorge, T. C.; Sampaio, E. P.; de Souza, W., Trypanosoma cruzi: Inhibition of host cell uptake of infective bloodstream forms by alpha-2-macroglobulin. *Z. Parasitenkd.* **1986,** *72* (3), 323-329.

122.   Souto-Padrón, T.; Campetella, O. E.; Cazzulo, J. J.; de Souza, W., Cysteine proteinase in Trypanosoma cruzi: immunocytochemical localization and involvement in parasite-host cell interaction. *J. Cell Sci.* **1990,** *96 (Pt 3)*, 485-490.

123.   Del Nery, E.; Juliano, M. A.; Lima, A. P. C. A.; Scharfstein, J.; Juliano, L., Kininogenase Activity by the Major Cysteinyl Proteinase (Cruzipain) from Trypanosoma cruzi. *J. Biol. Chem.* **1997,** *272* (41), 25713-25718.

124.   Scharfstein, J.; Monteiro, A. C.; Schmitz, V.; Svensjö, E., Angiotensin-converting enzyme limits inflammation elicited by Trypanosoma cruzi cysteine proteases: a peripheral mechanism regulating adaptive immunity via the innate kinin pathway. *Biol. Chem.* **2008,** *389* (8), 1015-1024.

125.   Scharfstein, J.; Schmitz, V.; Morandi, V.; Capella, M. M.; Lima, A. P*., et al.*, Host cell invasion by Trypanosoma cruzi is potentiated by activation of bradykinin B(2) receptors. *J. Exp. Med.* **2000,** *192* (9), 1289-1300.

126.   Ferrão, P. M.; d'Avila-Levy, C. M.; Araujo-Jorge, T. C.; Degrave, W. M.; Gonçalves, A. d. S*., et al.*, Cruzipain Activates Latent TGF-β from Host Cells during T. cruzi Invasion. *PLoS One* **2015,** *10* (5), e0124832.

127.   Bontempi, E.; Cazzulo, J. J., Digestion of human immunoglobulin G by the major cysteine proteinase (cruzipain) from Trypanosoma cruzi. *FEMS Microbiol. Lett.* **1990,** *58* (3), 337-341.

128.   Berasain, P.; Carmona, C.; Frangione, B.; Cazzulo, J. J.; Goñi, F., Specific cleavage sites on human IgG subclasses by cruzipain, the major cysteine proteinase from Trypanosoma cruzi. *Mol. Biochem. Parasitol.* **2003,** *130* (1), 23-29.

129.   Doyle, P. S.; Zhou, Y. M.; Hsieh, I.; Greenbaum, D. C.; McKerrow, J. H*., et al.*, The Trypanosoma cruzi Protease Cruzain Mediates Immune Evasion. *PLoS Pathog.* **2011,** *7* (9), e1002139.

130.   Doyle, P. S.; Zhou, Y. M.; Engel, J. C.; McKerrow, J. H., A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immunodeficient-Mouse Model of Infection. *Antimicrob. Agents Chemother.* **2007,** *51* (11), 3932-3939.

131.   Grab, D. J.; Garcia-Garcia, J. C.; Nikolskaia, O. V.; Kim, Y. V.; Brown, A*., et al.*, Protease Activated Receptor Signaling Is Required for African Trypanosome Traversal of Human Brain Microvascular Endothelial Cells. *PLoS Negl. Trop. Dis.* **2009,** *3* (7), e479.

132.   Nikolskaia, O. V.; de A. Lima, A. P. C.; Kim, Y. V.; Lonsdale-Eccles, J. D.; Fukuma, T*., et al.*, Blood-brain barrier traversal by African trypanosomes requires calcium signaling induced by parasite cysteine protease. *The Journal of Clinical Investigation* **2006,** *116* (10), 2739-2747.

133.   Santos, C. C.; Coombs, G. H.; Lima, A. P. C. A.; Mottram, J. C., Role of the Trypanosoma brucei natural cysteine peptidase inhibitor ICP in differentiation and virulence. *Mol. Microbiol.* **2007,** *66* (4), 991-1002.

134.   O'Brien, T. C.; Mackey, Z. B.; Fetter, R. D.; Choe, Y.; O'Donoghue, A. J*., et al.*, A Parasite Cysteine Protease Is Key to Host Protein Degradation and Iron Acquisition. *J. Biol. Chem.* **2008,** *283* (43), 28934-28943.

135.   Garcia, M. P.; Nóbrega, O. T.; Teixeira, A. R. L.; Sousa, M. V.; Santana, J. M., Characterisation of a Trypanosoma cruzi acidic 30 kDa cysteine protease. *Mol. Biochem. Parasitol.* **1998,** *91* (2), 263-272.

136.   Puente, X. S.; Sánchez, L. M.; Overall, C. M.; López-Otín, C., Human and mouse proteases: a comparative genomic approach. *Nature Reviews Genetics* **2003,** *4* (7), 544-558.

137.   Yan, H.-B.; Lou, Z.-Z.; Li, L.; Brindley, P. J.; Zheng, Y*., et al.*, Genome-wide analysis of regulatory proteases sequences identified through bioinformatics data mining in Taenia solium. *BMC Genomics* **2014,** *15* (1), 428.

138.   McKerrow, J. H., The diverse roles of cysteine proteases in parasites and their suitability as drug targets. *PLoS Negl. Trop. Dis.* **2018,** *12* (8), e0005639.

139.   Xing, R.; Addington, A. K.; Mason, R. W., Quantification of cathepsins B and L in cells. *Biochem. J.* **1998,** *332 (Pt 2)*, 499-505.

140.   Kindy, M. S.; Yu, J.; Zhu, H.; El-Amouri, S. S.; Hook, V*., et al.*, Deletion of the cathepsin B gene improves memory deficits in a transgenic ALZHeimer's disease mouse model expressing AβPP containing the wild-type β-secretase site sequence. *J. Alzheimers Dis.* **2012,** *29* (4), 827-840.

141.   McGrath, M. E.; Eakin, A. E.; Engel, J. C.; McKerrow, J. H.; Craik, C. S*., et al.*, The Crystal Structure of Cruzain: A Therapeutic Target for Chagas' Disease. *J. Mol. Biol.* **1995,** *247* (2), 251-259.

142.   Kerr, I. D.; Lee, J. H.; Farady, C. J.; Marion, R.; Rickert, M*., et al.*, Vinyl Sulfones as Antiparasitic Agents and a Structural Basis for Drug Design. *J. Biol. Chem.* **2009,** *284* (38), 25697-25703.

143.   Schechter, I.; Berger, A., On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967,** *27* (2), 157-162.

144.   Choe, Y.; Leonetti, F.; Greenbaum, D. C.; Lecaille, F.; Bogyo, M*., et al.*, Substrate Profiling of Cysteine Proteases Using a Combinatorial Peptide Library Identifies Functionally Unique Specificities. *J. Biol. Chem.* **2006,** *281* (18), 12824-12832.

145.   Harris, J. L.; Backes, B. J.; Leonetti, F.; Mahrus, S.; Ellman, J. A*., et al.*, Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proceedings of the National Academy of Sciences* **2000,** *97* (14), 7754-7759.

146.   Gillmor, S. A.; Craik, C. S.; Fletterick, R. J., Structural determinants of specificity in the cysteine protease cruzain. *Protein Sci.* **1997,** *6* (8), 1603-1611.

147.   Arafet, K.; Ferrer, S.; Moliner, V., Computational Study of the Catalytic Mechanism of the Cruzain Cysteine Protease. *ACS Catalysis* **2017,** *7* (2), 1207-1215.

148.   Zhai, X.; Meek, T. D., Catalytic Mechanism of Cruzain from Trypanosoma cruzi As Determined from Solvent Kinetic Isotope Effects of Steady-State and Pre-Steady-State Kinetics. *Biochemistry* **2018,** *57* (22), 3176-3190.

149.   Engel, J. C.; Ang, K. K. H.; Chen, S.; Arkin, M. R.; McKerrow, J. H*., et al.*, Image-Based High-Throughput Drug Screening Targeting the Intracellular Stage of Trypanosoma cruzi, the Agent of Chagas' Disease. *Antimicrob. Agents Chemother.* **2010,** *54* (8), 3326-3334.

150.	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* (4), 725-734.

151.	Barr, S. C.; Warner, K. L.; Kornreic, B. G.; Piscitelli, J.; Wolfe, A*., et al.*, A Cysteine Protease Inhibitor Protects Dogs from Cardiac Damage during Infection by Trypanosoma cruzi. *Antimicrob. Agents Chemother.* **2005,** *49* (12), 5160-5161.

152.	K777 (Chagas). https://dndi.org/research-development/portfolio/k777/ (accessed Sep 30, 2014).

153.	Schneider, T. H.; Rieger, M.; Ansorg, K.; Sobolev, A. N.; Schirmeister, T*., et al.*, Vinyl sulfone building blocks in covalently reversible reactions with thiols. *New Journal of Chemistry* **2015,** *39* (7), 5841-5853.

154.	Brak, K.; Doyle, P. S.; McKerrow, J. H.; Ellman, J. A., Identification of a New Class of Nonpeptidic Inhibitors of Cruzain. *J. Am. Chem. Soc.* **2008,** *130* (20), 6404-6410.

155.	Neitz, R. J.; Bryant, C.; Chen, S.; Gut, J.; Hugo Caselli, E*., et al.*, Tetrafluorophenoxymethyl ketone cruzain inhibitors with improved pharmacokinetic properties as therapeutic leads for Chagas' disease. *Bioorg. Med. Chem. Lett.* **2015,** *25* (21), 4834-4837.

156.	Ferreira, R. S.; Simeonov, A.; Jadhav, A.; Eidam, O.; Mott, B. T*., et al.*, Complementarity Between a Docking and a High-Throughput Screen in Discovering New Cruzain Inhibitors. *J. Med. Chem.* **2010,** *53* (13), 4891-4905.

157.	Ferreira, R. S.; Dessoy, M. A.; Pauli, I.; Souza, M. L.; Krogh, R*., et al.*, Synthesis, Biological Evaluation, and Structure–Activity Relationships of Potent Noncovalent and Nonpeptidic Cruzain Inhibitors as Anti-Trypanosoma cruzi Agents. *J. Med. Chem.* **2014,** *57* (6), 2380-2392.

158.	Chen, Y. T.; Brinen, L. S.; Kerr, I. D.; Hansell, E.; Doyle, P. S*., et al.*, In Vitro and In Vivo Studies of the Trypanocidal Properties of WRR-483 against Trypanosoma cruzi. *PLoS Negl. Trop. Dis.* **2010,** *4* (9), e825.

159.	Jones, B. D.; Tochowicz, A.; Tang, Y.; Cameron, M. D.; McCall, L.-I*., et al.*, Synthesis and Evaluation of Oxyguanidine Analogues of the Cysteine Protease Inhibitor WRR-483 against Cruzain. *ACS Med. Chem. Lett.* **2016,** *7* (1), 77-82.

160.   Latorre, A.; Schirmeister, T.; Kesselring, J.; Jung, S.; Johé, P*., et al.*, Dipeptidyl Nitroalkenes as Potent Reversible Inhibitors of Cysteine Proteases Rhodesain and Cruzain. *ACS Med. Chem. Lett.* **2016,** *7* (12), 1073-1076.

161.   Boudreau, P. D.; Miller, B. W.; McCall, L.-I.; Almaliti, J.; Reher, R*., et al.*, Design of Gallinamide A Analogs as Potent Inhibitors of the Cysteine Proteases Human Cathepsin L and Trypanosoma cruzi Cruzain. *J. Med. Chem.* **2019,** *62* (20), 9026-9044.

162.   Copeland, R. A.; Pompliano, D. L.; Meek, T. D., Drug–target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery* **2006,** *5* (9), 730-739.

163.   Bradshaw, J. M.; McFarland, J. M.; Paavilainen, V. O.; Bisconte, A.; Tam, D*., et al.*, Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nat. Chem. Biol.* **2015,** *11* (7), 525-531.

164.   Krishnan, S.; Miller, R. M.; Tian, B.; Mullins, R. D.; Jacobson, M. P*., et al.*, Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *J. Am. Chem. Soc.* **2014,** *136* (36), 12624-12630.

165.   Miller, B. E.; Mayer, R. J.; Goyal, N.; Bal, J.; Dallow, N*., et al.*, Epithelial desquamation observed in a phase I study of an oral cathepsin C inhibitor (GSK2793660). *Br. J. Clin. Pharmacol.* **2017,** *83* (12), 2813-2820.

166.   Barr, S. C.; Warner, K. L.; Kornreic, B. G.; Piscitelli, J.; Wolfe, A*., et al.*, A Cysteine Protease Inhibitor Protects Dogs from Cardiac Damage during Infection by Trypanosoma cruzi. *Antimicrob. Agents Chemother.* **2005,** *49* (12), 5160.

167.   Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J*., et al.*, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-1749.

168.   Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L*., et al.*, Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004,** *47* (7), 1750-1759.

169.   Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R*., et al.*, Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes. *J. Med. Chem.* **2006,** *49* (21), 6177-6196.

170.   Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R*., et al.*, Docking Covalent Inhibitors: A Parameter Free Approach To Pose Prediction and Scoring. *J. Chem. Inf. Model.* **2014,** *54* (7), 1932-1940.

171.   Wittig, G.; Schöllkopf, U., Über Triphenyl-phosphin-methylene als olefinbildende Reagenzien (I. Mitteil. *Chem. Ber.* **1954,** *87* (9), 1318-1330.

172.   Wadsworth, W. S.; Emmons, W. D., The Utility of Phosphonate Carbanions in Olefin Synthesis. *J. Am. Chem. Soc.* **1961,** *83* (7), 1733-1738.

173.   Nahm, S.; Weinreb, S. M., N-methoxy-n-methylamides as effective acylating agents. *Tetrahedron Lett.* **1981,** *22* (39), 3815-3818.

174.   Palmer, J. T.; Rasnick, D.; Klaus, J. L. Irreversible cysteine protease inhibitors containing vinyl groups conjugated to electron withdrawing groups. 2001.

175.   Morrison, J. F.; Walsh, C. T., The Behavior and Significance of Slow-Binding Enzyme Inhibitors. In *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1988; pp 201-301.

176.   Ndao, M.; Beaulieu, C.; Black, W. C.; Isabel, E.; Vasquez-Camargo, F*., et al.*, Reversible Cysteine Protease Inhibitors Show Promise for a Chagas Disease Cure. *Antimicrob. Agents Chemother.* **2014,** *58* (2), 1167.

177.   Steverding, D.; Sexton, D. W.; Wang, X.; Gehrke, S. S.; Wagner, G. K*., et al.*, Trypanosoma brucei: Chemical evidence that cathepsin L is essential for survival and a relevant drug target. *Int. J. Parasitol.* **2012,** *42* (5), 481-488.

178.   McKerrow, J.; Doyle, P.; Engel, J.; Podust, L.; Robertson, S*., et al.*, Two approaches to discovering and developing new drugs for Chagas disease. *Mem. Inst. Oswaldo Cruz* **2009,** *104*, 263-269.

179.   Nkemgu, N. J.; Grande, R.; Hansell, E.; McKerrow, J. H.; Caffrey, C. R*., et al.*, Improved trypanocidal activities of cathepsin L inhibitors. *Int. J. Antimicrob. Agents* **2003,** *22* (2), 155-159.

180.   Yang, P.-Y.; Wang, M.; He, C. Y.; Yao, S. Q., Proteomic profiling and potential cellular target identification of K11777, a clinical cysteine protease inhibitor, in Trypanosoma brucei. *Chemical Communications* **2012,** *48* (6), 835-837.

181.   Lee, G. M.; Balouch, E.; Goetz, D. H.; Lazic, A.; McKerrow, J. H*., et al.*, Mapping Inhibitor Binding Modes on an Active Cysteine Protease via Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2012,** *51* (50), 10087-10098.

182.   Zhao, Z.; Bourne, P. E., Progress with covalent small-molecule kinase inhibitors. *Drug Discovery Today* **2018,** *23* (3), 727-735.

183.   Scheidt, K. A.; Roush, W. R.; McKerrow, J. H.; Selzer, P. M.; Hansell, E*., et al.*, Structure-based design, synthesis and evaluation of conformationally constrained cysteine protease inhibitors. *Bioorg. Med. Chem.* **1998,** *6* (12), 2477-2494.

184.   Choe, Y.; Brinen, L. S.; Price, M. S.; Engel, J. C.; Lange, M*., et al.*, Development of alpha-keto-based inhibitors of cruzain, a cysteine protease implicated in Chagas disease. *Bioorg. Med. Chem.* **2005,** *13* (6), 2141-2156.

185.   Silva, D. G.; Ribeiro, J. F. R.; De Vita, D.; Cianni, L.; Franco, C. H*., et al.*, A comparative study of warheads for design of cysteine protease inhibitors. *Bioorg. Med. Chem. Lett.* **2017,** *27* (22), 5031-5035.

186.   Bonatto, V.; Batista, P. H. J.; Cianni, L.; De Vita, D.; Silva, D. G*., et al.*, On the intrinsic reactivity of highly potent trypanocidal cruzain inhibitors. *RSC Med. Chem.* **2020,** *11* (11), 1275-1284.

187.   Nelson, S. D., Metabolic activation and drug toxicity. *J. Med. Chem.* **1982,** *25* (7), 753-765.

188.   Dalvie, D. K.; Kalgutkar, A. S.; Khojasteh-Bakht, S. C.; Obach, R. S.; O'Donnell, J. P., Biotransformation Reactions of Five-Membered Aromatic Heterocyclic Rings. *Chem. Res. Toxicol.* **2002,** *15* (3), 269-299.

189.   O'Brien, P. J.; Siraki, A. G.; Shangari, N., Aldehyde Sources, Metabolism, Molecular Toxicity Mechanisms, and Possible Effects on Human Health. *Crit. Rev. Toxicol.* **2005,** *35* (7), 609-662.

190.   Ekert, P. G.; Silke, J.; Vaux, D. L., Caspase inhibitors. *Cell Death Differ.* **1999,** *6* (11), 1081-1086.

191.   Kisselev, A. F.; Goldberg, A. L., Proteasome inhibitors: from research tools to drug candidates. *Chem. Biol.* **2001,** *8* (8), 739-758.

192.	Gehringer, M.; Laufer, S. A., Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **2019,** *62* (12), 5673-5724.

193.	Gampe, C.; Verma, V. A., Curse or Cure? A Perspective on the Developability of Aldehydes as Active Pharmaceutical Ingredients. *J. Med. Chem.* **2020**.

194.	Holdiness, M. R., Clinical Pharmacokinetics of the Antituberculosis Drugs. *Clin. Pharmacokinet.* **1984,** *9* (6), 511-544.

195.	Meanwell, N. A., Improving Drug Design: An Update on Recent Applications of Efficiency Metrics, Strategies for Replacing Problematic Elements, and Compounds in Nontraditional Drug Space. *Chem. Res. Toxicol.* **2016,** *29* (4), 564-616.

196.	Gibbons, P.; Verissimo, E.; Araujo, N. C.; Barton, V.; Nixon, G. L.*, et al.*, Endoperoxide Carbonyl Falcipain 2/3 Inhibitor Hybrids: Toward Combination Chemotherapy of Malaria through a Single Chemical Entity. *J. Med. Chem.* **2010,** *53* (22), 8202-8206.

197.	Lopes, F.; Santos, M. M. M.; Moreira, R., In *1.2 Designing Covalent Inhibitors: A Medicinal Chemistry Challenge*, De Gruyter: 2015; pp 44-60.

198.	Nakamura, M.; Yamaguchi, M.; Sakai, O.; Inoue, J., Exploration of cornea permeable calpain inhibitors as anticataract agents. *Bioorg. Med. Chem.* **2003,** *11* (7), 1371-1379.

199.	Cuerrier, D.; Moldoveanu, T.; Inoue, J.; Davies, P. L.; Campbell, R. L., Calpain Inhibition by α-Ketoamide and Cyclic Hemiacetal Inhibitors Revealed by X-ray Crystallography. *Biochemistry* **2006,** *45* (24), 7446-7452.

200.	Eaton, J. K.; Furst, L.; Ruberto, R. A.; Moosmayer, D.; Hilpmann, A.*, et al.*, Selective covalent targeting of GPX4 using masked nitrile-oxide electrophiles. *Nat. Chem. Biol.* **2020,** *16* (5), 497-506.

201.	Cucinotta, D.; Vanelli, M., WHO Declares COVID-19 a Pandemic. *Acta biomedica : Atenei Parmensis* **2020,** *91* (1), 157-160.

202.   COVID-19 to Plunge Global Economy into Worst Recession since World War II. https://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-into-worst-recession-since-world-war-ii (accessed June 8, 2020).

203.   Haghani, M.; Bliemer, M. C. J.; Goerlandt, F.; Li, J., The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. *Safety Science* **2020,** *129*, 104806.

204.   Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P*., et al.*, Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science* **2020,** *6* (3), 315-331.

205.   Sohag, A. A. M.; Hannan, M. A.; Rahman, S.; Hossain, M.; Hasan, M*., et al.*, Revisiting potential druggable targets against SARS-CoV-2 and repurposing therapeutics under preclinical study and clinical trials: A comprehensive review. *Drug Development Research n/a* (n/a).

206.   Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R., Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem* **2020,** *21* (5), 730-738.

207.   Al-Gharabli, S. I.; Shah, S. T. A.; Weik, S.; Schmidt, M. F.; Mesters, J. R*., et al.*, An Efficient Method for the Synthesis of Peptide Aldehyde Libraries Employed in the Discovery of Reversible SARS Coronavirus Main Protease (SARS-CoV Mpro) Inhibitors. *ChemBioChem* **2006,** *7* (7), 1048-1055.

208.   Zhu, L.; George, S.; Schmidt, M. F.; Al-Gharabli, S. I.; Rademann, J*., et al.*, Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral Res.* **2011,** *92* (2), 204-212.

209.   Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J*., et al.*, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020,** *368* (6497), 1331-1335.

210.   Clementi, E.; Raimondi, D. L.; Reinhardt, W. P., Atomic Screening Constants from SCF Functions. II. Atoms with 37 to 86 Electrons. *J. Chem. Phys* **1967,** *47* (4), 1300-1307.

211.   Jonathan Clayden, N. G., and Stuart Warren, *Organic Chemistry*. Second ed.; Oxford University Press: 2012.

212.   Mackenzie, N. E.; Grant, S. K.; Scott, A. I.; Malthouse, J. P. G., Carbon-13 NMR study of the stereospecificity of the thiohemiacetals formed on inhibition of papain by specific enantiomeric aldehydes. *Biochemistry* **1986,** *25* (8), 2293-2298.

213.   Enanga, B.; Ariyanayagam, M. R.; Stewart, M. L.; Barrett, M. P., Activity of Megazol, a Trypanocidal Nitroimidazole, Is Associated with DNA Damage. *Antimicrob. Agents Chemother.* **2003,** *47* (10), 3368-3370.

214.   Jacobsen, W.; Christians, U.; Benet, L. Z., In vitro evaluation of the disposition of A novel cysteine protease inhibitor. *Drug Metab. Dispos.* **2000,** *28* (11), 1343-1351.

215.   Vermeire, J. J.; Suzuki, B. M.; Caffrey, C. R., Odanacatib, a Cathepsin K Cysteine Protease Inhibitor, Kills Hookworm In Vivo. *Pharmaceuticals (Basel)* **2016,** *9* (3).

216.   Randle, J. C. R.; Harding, M. W.; Ku, G.; Schönharting, M.; Kurrle, R., ICE/Caspase-1 inhibitors as novel anti-inflammatory drugs. *Expert Opinion on Investigational Drugs* **2001,** *10* (7), 1207-1209.

217.   Charrier, J.-D.; Durrant, S. J.; Studley, J.; Lawes, L.; Weber, P., Synthesis and evaluation of novel prodrugs of caspase inhibitors. *Bioorg. Med. Chem. Lett.* **2012,** *22* (1), 485-488.

218.   Olsen, L.; Montefiori, M.; Tran, K. P.; Jørgensen, F. S., SMARTCyp 3.0: enhanced cytochrome P450 site-of-metabolism prediction server. *Bioinformatics* **2019,** *35* (17), 3174-3175.

219.   Bencharit, S.; Morton, C. L.; Xue, Y.; Potter, P. M.; Redinbo, M. R., Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. *Nat. Struct. Mol. Biol.* **2003,** *10* (5), 349-356.

220.   Wang, D.; Zou, L.; Jin, Q.; Hou, J.; Ge, G.*, et al.*, Human carboxylesterases: a comprehensive review. *Acta Pharm. Sin. B* **2018,** *8* (5), 699-712.

221.   Zanger, U. M.; Schwab, M., Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* **2013,** *138* (1), 103-141.

222.   Zhao, M.-M.; Yang, W.-L.; Yang, F.-Y.; Zhang, L.; Huang, W.*, et al.*, Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising target for new drug development. *medRxiv* **2020**, 2020.2010.2025.20218990.

223.   Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.*, et al.*, Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020,** *11* (1), 4417.

224.   Noisier, A. F. M.; Harris, C. S.; Brimble, M. A., Novel preparation of chiral α-amino acids using the Mitsunobu–Tsunoda reaction. *Chemical Communications* **2013,** *49* (70), 7744-7746.

225.   Mahindra, A.; Millard, C. J.; Black, I.; Archibald, L. J.; Schwabe, J. W. R.*, et al.*, Synthesis of HDAC Substrate Peptidomimetic Inhibitors Using Fmoc Amino Acids Incorporating Zinc-Binding Groups. *Organic Letters* **2019,** *21* (9), 3178-3182.

226.   Mellott, D. M.; Tseng, C.-T.; Drelich, A.; Fajtová, P.; Chenna, B. C.*, et al.*, A cysteine protease inhibitor blocks SARS-CoV-2 infection of human and monkey cells. *bioRxiv* **2020**, 2020.2010.2023.347534.

227.   Tight Binding Inhibitors. In *Enzymes*, 2000; pp 305-317.

228.   Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.*, et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995,** *6* (3), 277-293.

229.   Willcott, M. R., MestRe Nova. *J. Am. Chem. Soc.* **2009,** *131* (36), 13180-13180.

230.   Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2014,** *31* (8), 1325-1327.

231.   Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.*, et al.*, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020,** *368* (6489), 409-412.

232.   Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.*, et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **2011,** *67* (4), 235-242.

233.   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W*., et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **2010,** *66* (2), 213-221.

234.   Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D* **2010,** *66* (4), 486-501.

235.   Patterson, J. S.; Guhl, F., 5 - Geographical distribution of Chagas disease. In *American Trypanosomiasis: Chagas Disease One Hundred Years of Research (Second Edition)*, Telleria, J.; Tibayrenc, M., Eds. Elsevier: London, 2017; pp 89-112.

APPENDIX A. MOLECULAR MODELING OF PVHI COMPOUNDS

In addition to compound **9**, we also utilized Schrodinger to develop both covalent and non-covalent models of compounds **7**, **11**-**13** and **15** bound to cruzain, also based on the crystal structure of **K11777** covalently bound to cruzain as described in Experimental Section. The covalent and non-covalent models are in general similar in terms of the orientation of the peptide sidechains. The modeled structures differ, however, by virtue of the nature of how the molecular modeling was performed, leading to the distal positioning of the vinyl group in the non-covalent structures from the active-site residues Cys_{25} and His_{162}. Compounds **12** and **15** which each contain an N-methyl-pyridine, indicate that the methyl group with its attending positive charge is directed to His_{162}, suggesting that the latter is neutral.

Table A1. Predicted affinity of covalent docking for select structures

Compound	Structure	Cdock Affinity (kcal/mol)	Predicted K_i (μM)
7	Cbz-Phe-Phe-vinyl-2Pyrmd	-7.11	6.11
9	NMePip-Phe-hPhe-vinyl-2Pyrmd	-7.01	7.23
11	Cbz-Phe-Phe-vinyl-2Pyr	-7.17	5.52
12	Cbz-Phe-Phe-vinyl-2PyrNMe	-8.00	1.36
13	Cbz-Phe-hPhe-vinyl-2Pyr	-7.63	2.54
15	Cbz-Phe-hPhe-vinyl-2PyrNMe	-8.32	0.79
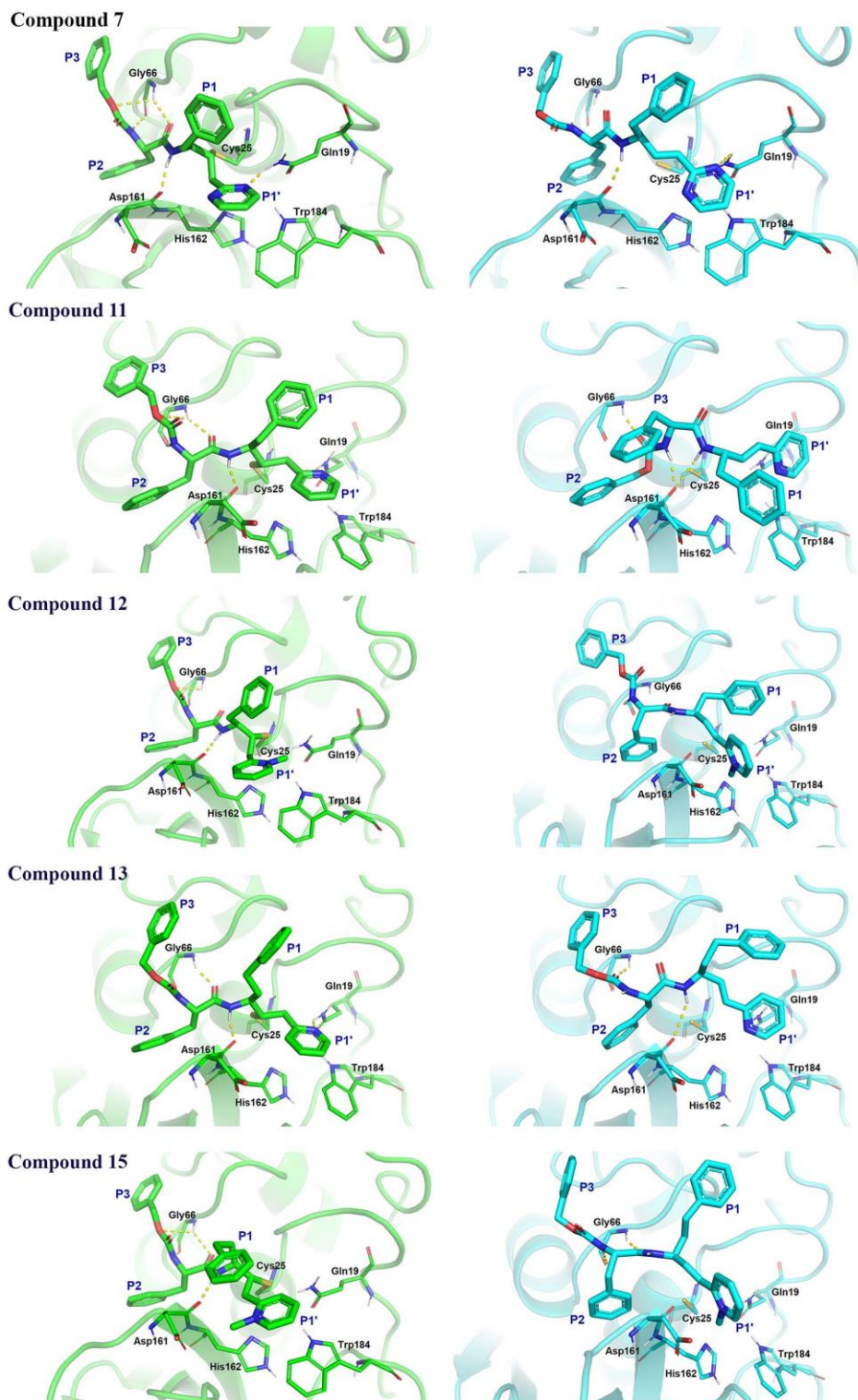
Figure A1. Molecular models of compound **9**, **11-13** and **15** bound to cruzain created using Schrodinger. Left panels: binding poses in which a covalent bond is formed between the β-carbon of the vinyl group of the inhibitor. Right panels: binding poses in which no covalent bond is formed with Cys$_{25}$.

173
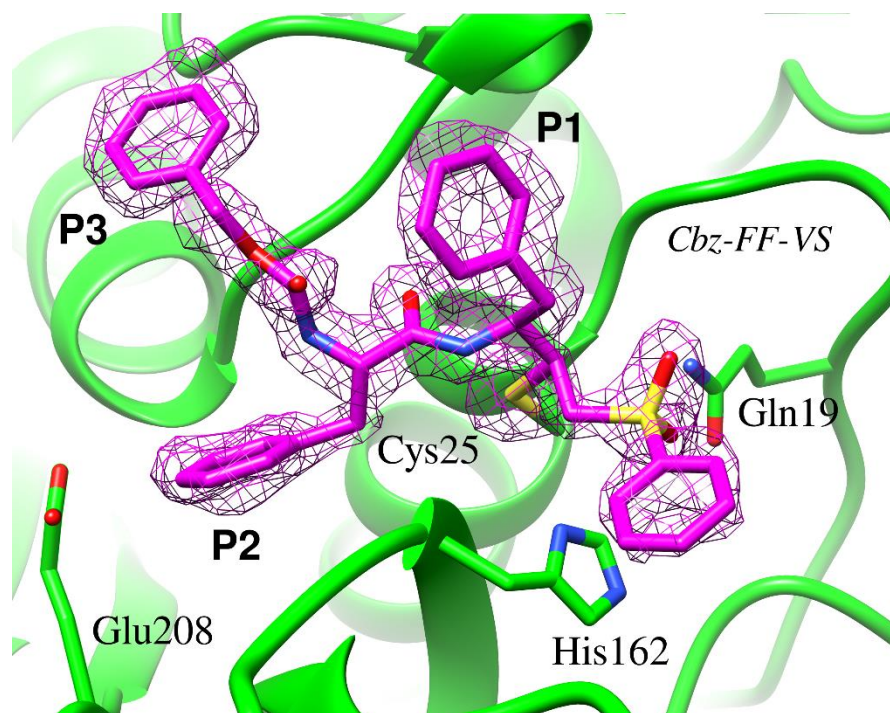
Figure A2. Crystal structure of cruzain bound to Cbz-Phe-Phe-VSPh (PVHI **1**). A covalent bond is evidently formed between the sulfur of Cys$_{25}$ and the β-carbon of the former olefin bond (Tang, *et al.*).

Figure A3. LC-MS of ^{13}C-labeled **12**. Calculated m/z for molecular ion is 440.24 for hemiacetal and 458.25 for hydrate.

Table A2. The details of peaks in ¹H-¹³C HSQC NMR

peak	¹H (ppm)	¹³C (ppm)	Width 1	Width 2	Volume
12 only					
A	5.34	90.84	192.6	31.4	40225129.1
B	5.13	90.80	192.59	26.24	52937692.9
12 + *cruzain*					
A'	6.13	76.25	137.58	49.49	5030193.12
B'	5.91	79.69	165.06	63.22	6245466.52
A (residual)	5.33	90.81	192.6	40.23	1725522.38
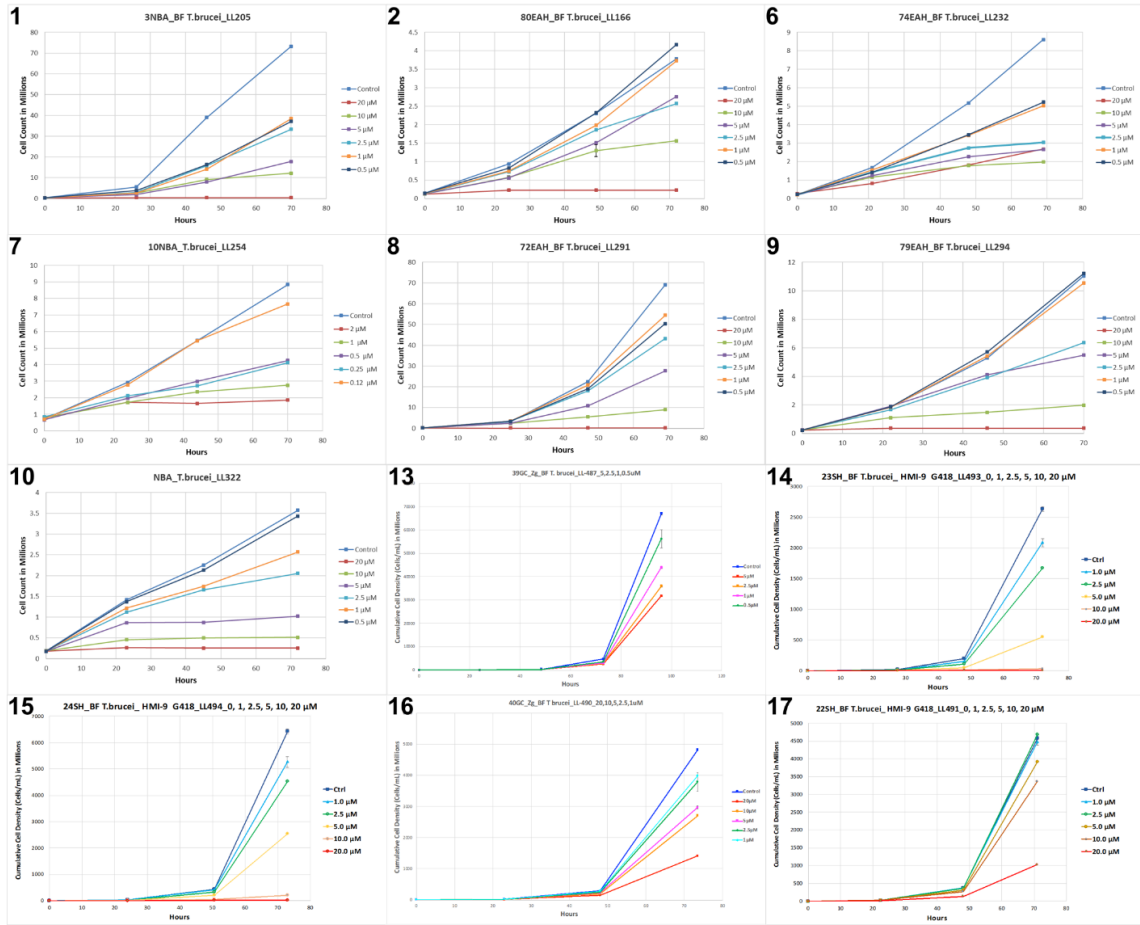
Figure A4. Growth curves for *T. b. brucei* BSFs treated with SMAIs and their prodrugs. Each diagram only represents one of the duplicate experiments.
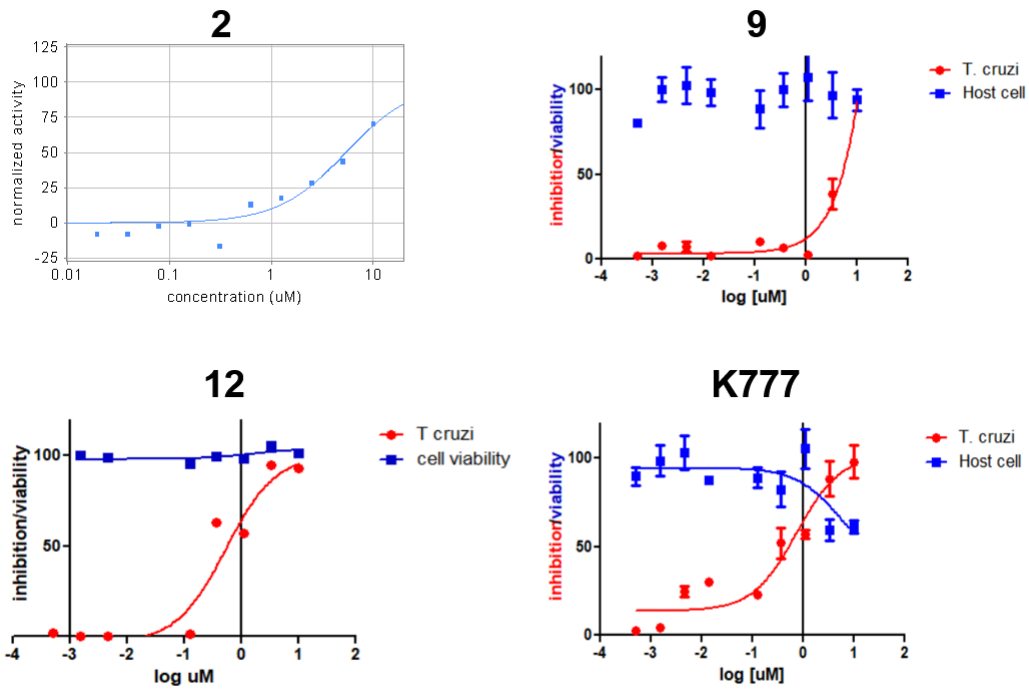
Figure A5. Cell viability of *T. cruzi*-infected murine cardiomyoblasts in presence of **2**, **9**, **12**, or **K11777**.
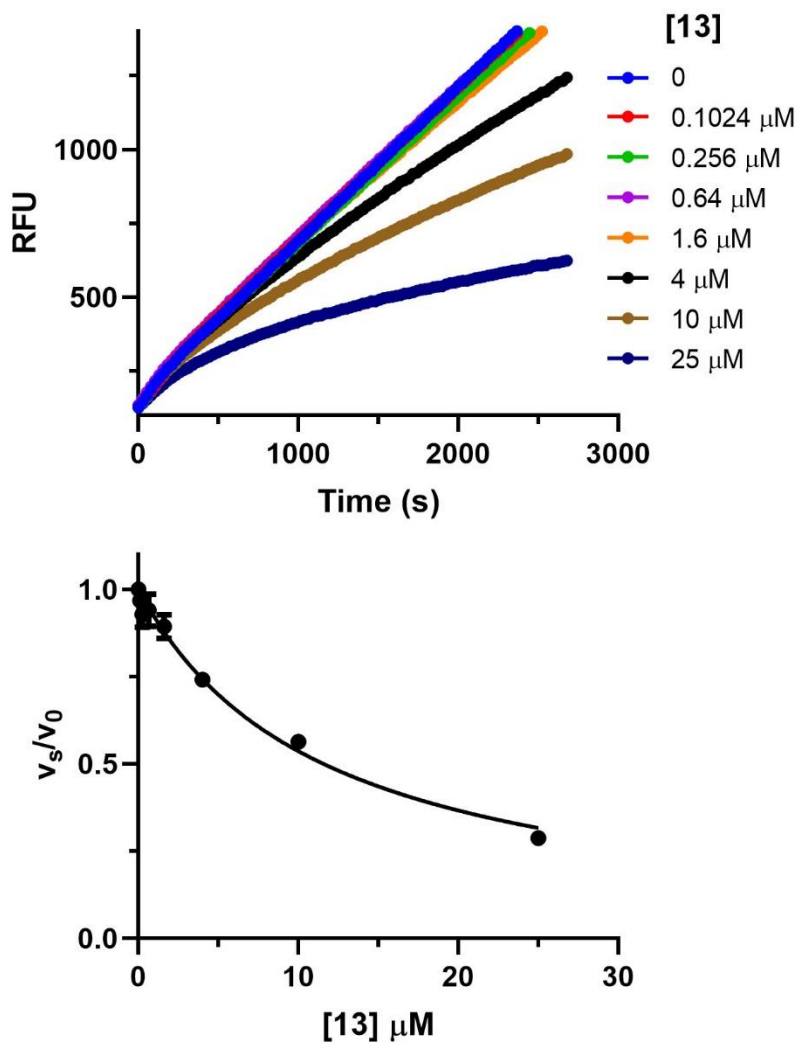
Figure A6. Time-dependent cruzain inhibition by **13**.

Table A3. Statistic summary of co-crystal structures of 3CL^{pro} complexed with 19 and 23.

	3CL^{pro}-**19**	3CL^{pro}-**23**
Data collection		
Space group	C 1 2 1	C 1 2 1
Cell dimensions		
a, b, c (Å)	54.20, 80.74, 85.76	54.22, 80.95, 85.53
α, β, γ (°)	90.00, 97.12, 90.00	90.00, 97.34, 90.00
Resolution (Å)	48.25-1.70 (1.76-1.70)	48.29-1.70 (1.73-1.70)
R_{merge}	0.084 (1.632)	0.073 (0.856)
$<I/\sigma I>$	10.2 (0.8)	11.5 (0.9)
CC$_{1/2}$	0.996 (0.475)	0.998 (0.824)
Completeness (%)	99.8 (99.3)	100.0 (99.7)
Redundancy	6.4 (5.6)	6.5 (6.0)
Refinement		
Resolution (Å)	48.25-1.70	48.29-1.70
No. Reflections	40267 (4001)	40244 (3999)
R_{work}/R_{free}	0.1857/0.2058	0.2030/0.2328
RMSD in bond length	0.008	0.009
RMSD in bond angles	1.17	1.3
No. atoms		
Protein	2395	2395
Ligand	40	34
Water	173	186
B factors		
Protein	41.6	38.1
Ligand	43.4	39.6
Water	44.9	41.7

PVHI compounds.

1:

2:

8.259 8.239 7.392 7.380 7.358 7.349 7.331 7.313 7.307 7.303 7.285 7.262 7.250 7.211 7.199 7.191 7.181 6.927 6.587 6.572 6.548 6.534 5.876 5.873 5.837 5.834 5.944 5.950 4.615 4.255 3.765 3.298 3.246 2.943 2.933 2.849 2.831 2.755 2.728 2.721 2.674 2.504 2.500 2.495

0.95 18.66 1.08 1.00 0.96 1.96 0.99 1.06 0.22 0.23 0.99 1.73 1.36 0.21 0.36

170.74 166.15 155.64 142.03 138.01 137.00 129.54 129.11 128.20 128.09 127.91 127.58 127.35 126.40 126.17 124.11 65.11 56.07 50.91 40.14 40.93 39.72 39.52 39.31 39.10 38.89 37.76 28.74

182

3

BC-481C, 1H NMR, DMSO-d6

8.225
8.205
7.503
7.481
7.414
7.326
7.308
7.297
7.279
7.273
7.260
7.254
7.224
7.214
7.208
7.198
7.185
7.180
7.162
6.943
6.560
6.546
6.522
6.507
5.916
5.878
4.972
4.961
4.929
4.308
4.302
3.296
3.015
3.005
2.813
2.805
2.613
2.594
2.517
2.512
2.508
2.503
2.499
1.800
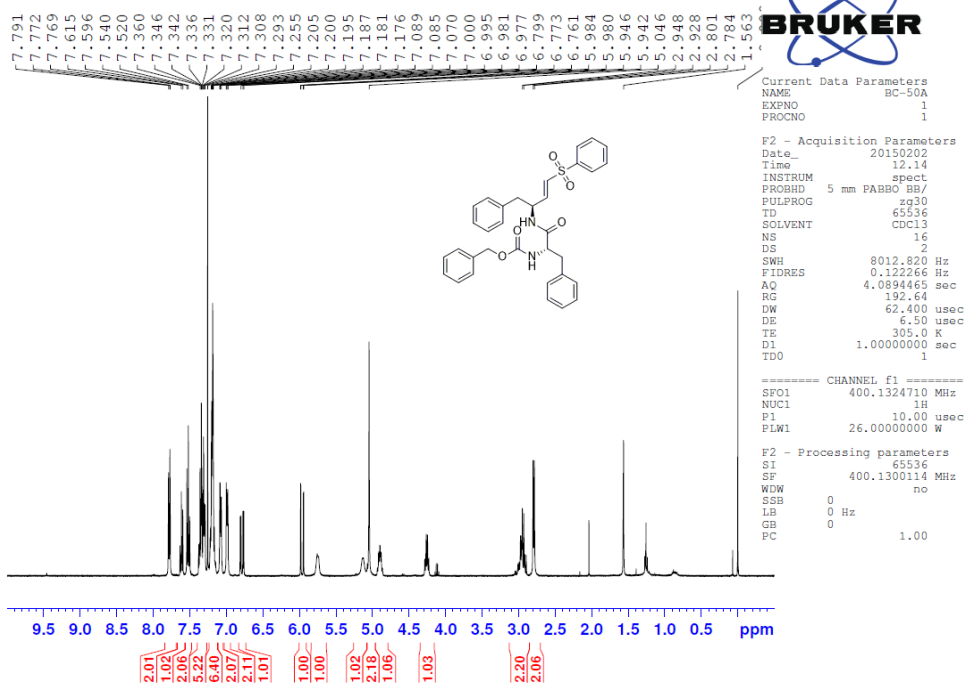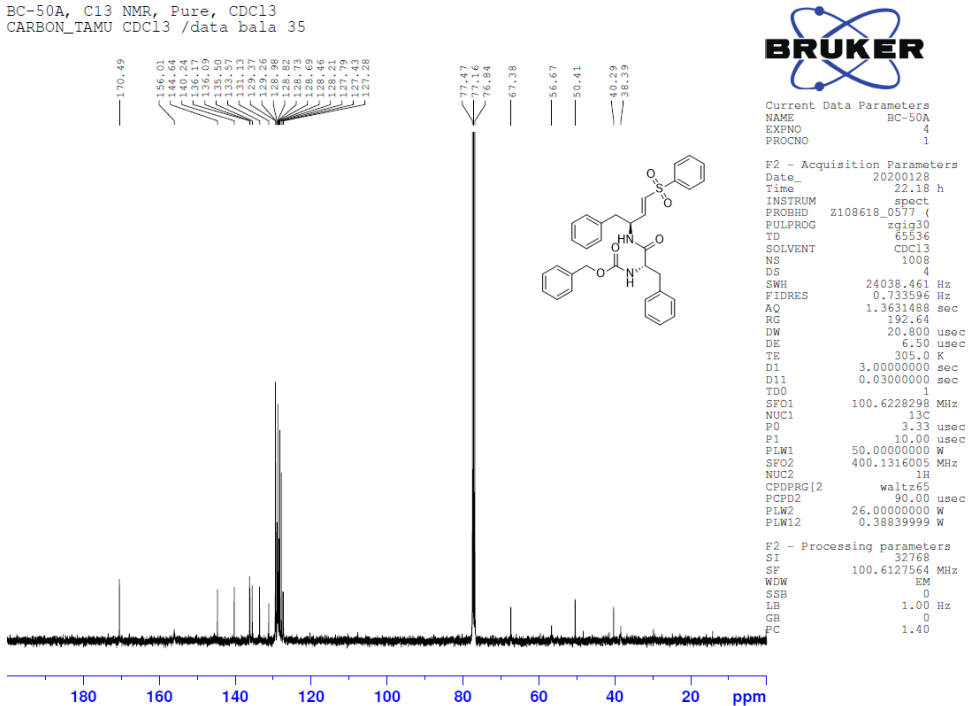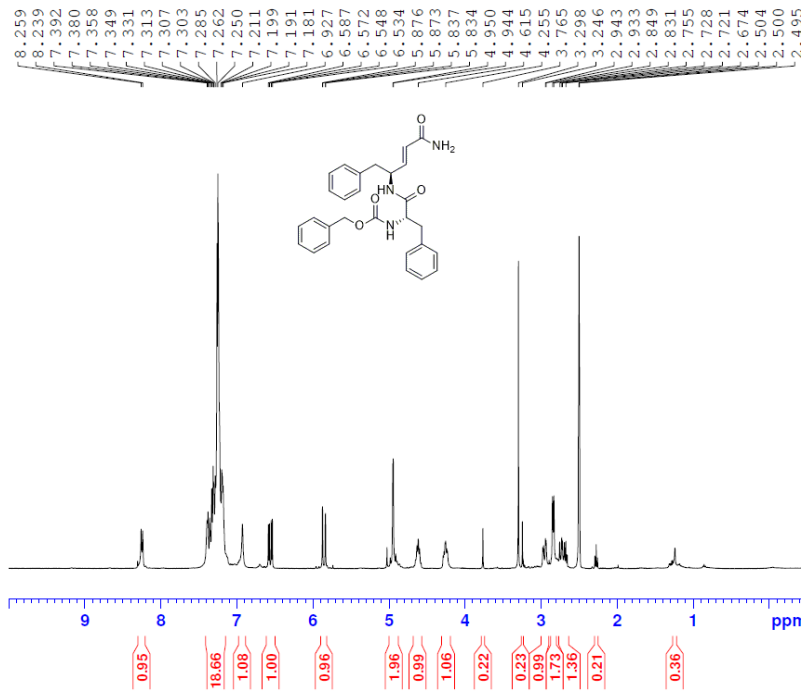
Current Data Parameters
NAME   BC-481C
EXPNO   1
PROCNO   1
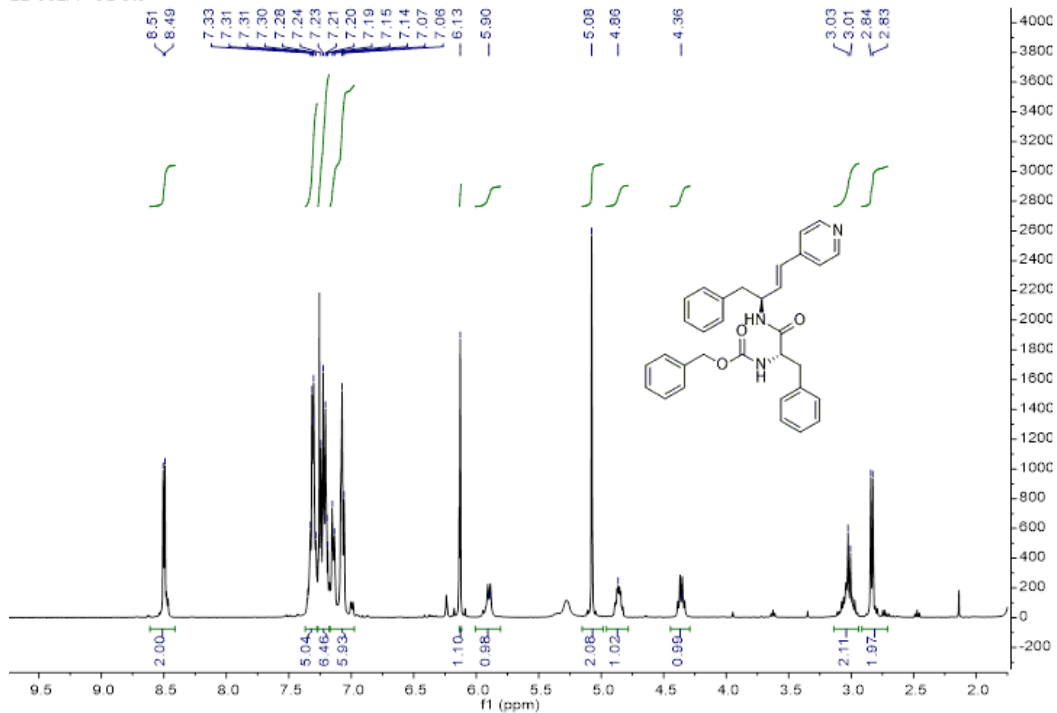
F2 - Acquisition Parameters
Date_   20180717
Time   11.11 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zg30
TD   65536
SOLVENT   DMSO
NS   16
DS   2
SWH   7211.539 Hz
FIDRES   0.220079 Hz
AQ   4.5438294 sec
RG   192.64
DW   69.333 usec
DE   6.50 usec
TE   305.0 K
D1   1.00000000 sec
TD0   1
SFO1   400.1322007 MHz
NUC1   1H
P1   9.70 usec
PLW1   26.00000000 W

F2 - Processing parameters
SI   65536
SF   400.1300000 MHz
WDW   EM
SSB   0
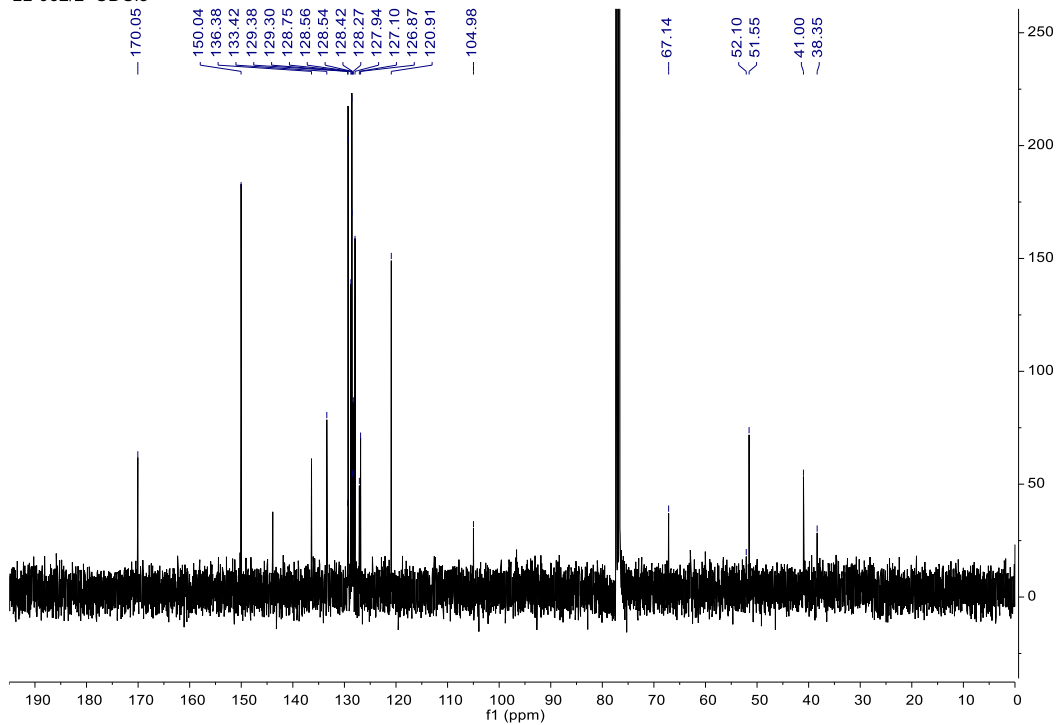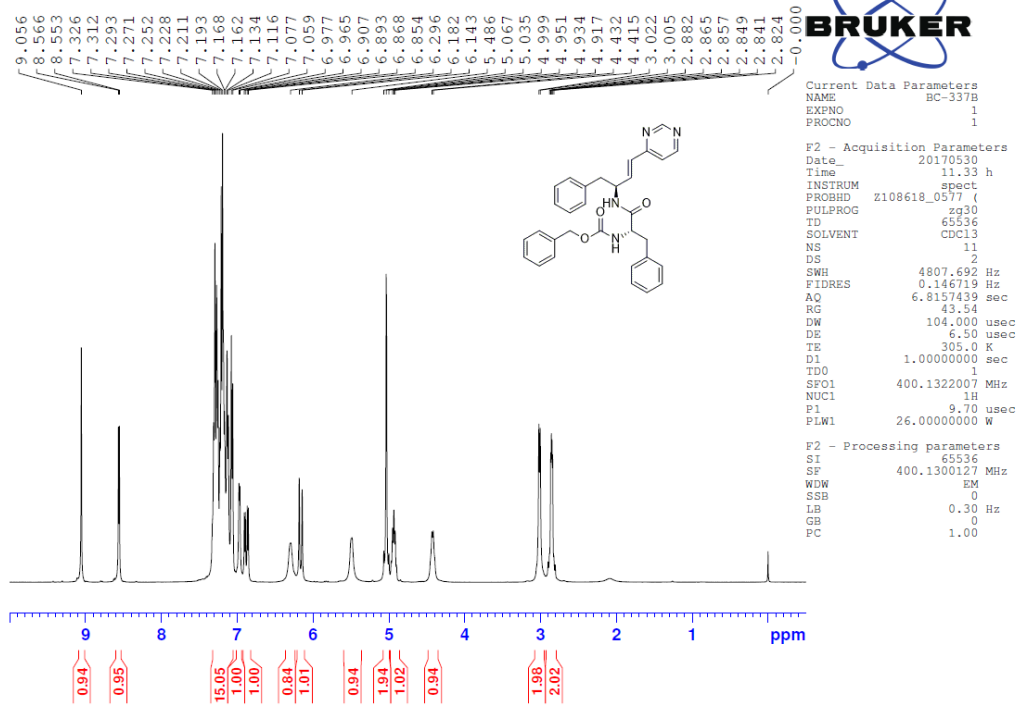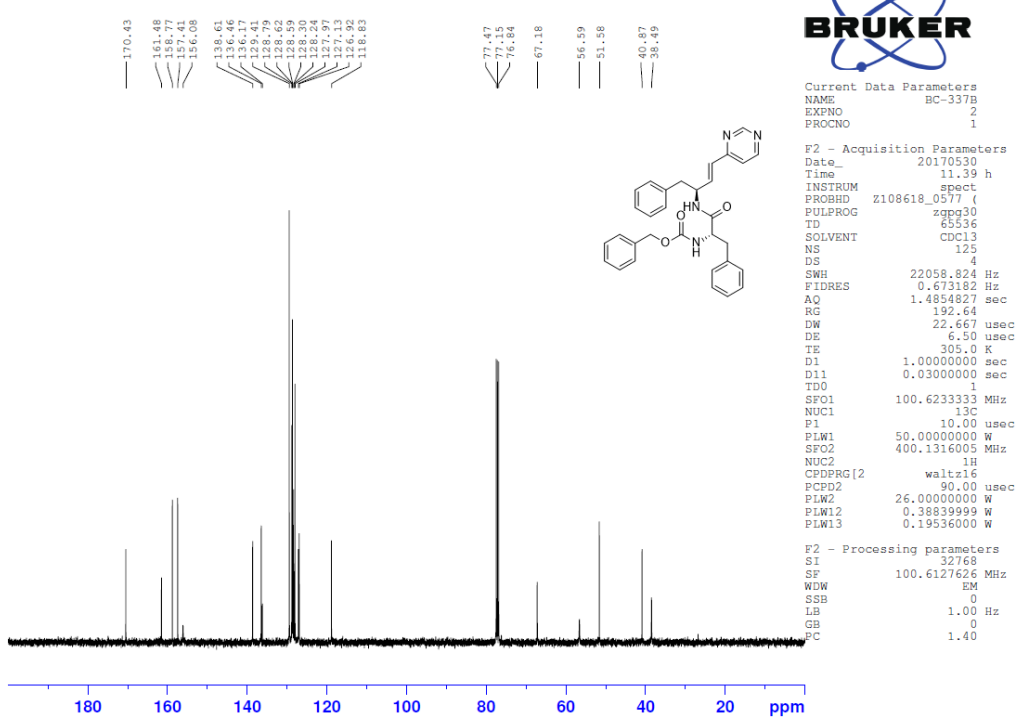LB   0.30 Hz
GB   0
PC   1.00

1.01 0.97 0.96 15.23 0.93 1.00 0.99 2.06 0.97 1.07 1.06 1.05 2.03 2.09

4:

5:

LL-29 fr41+/1 CDCl3

LL-29 fr41+/3 CDCl3
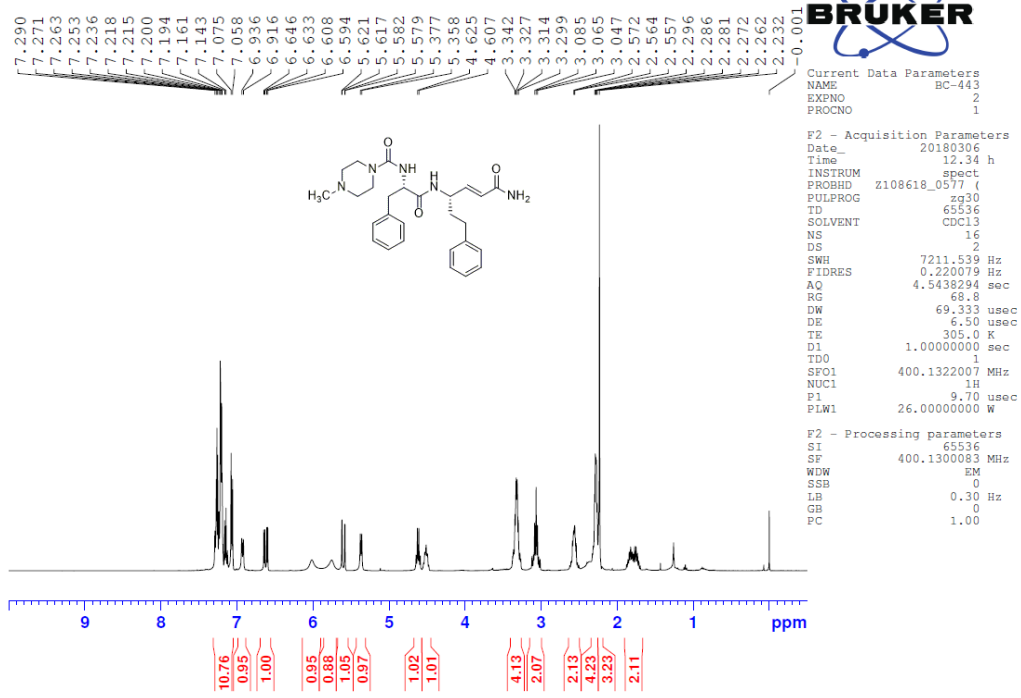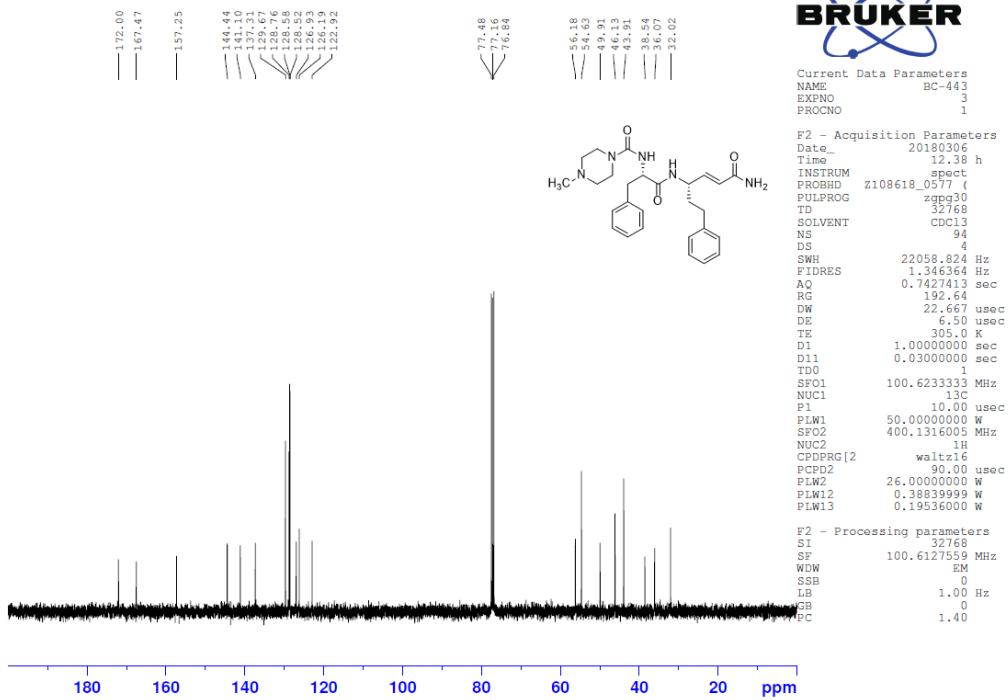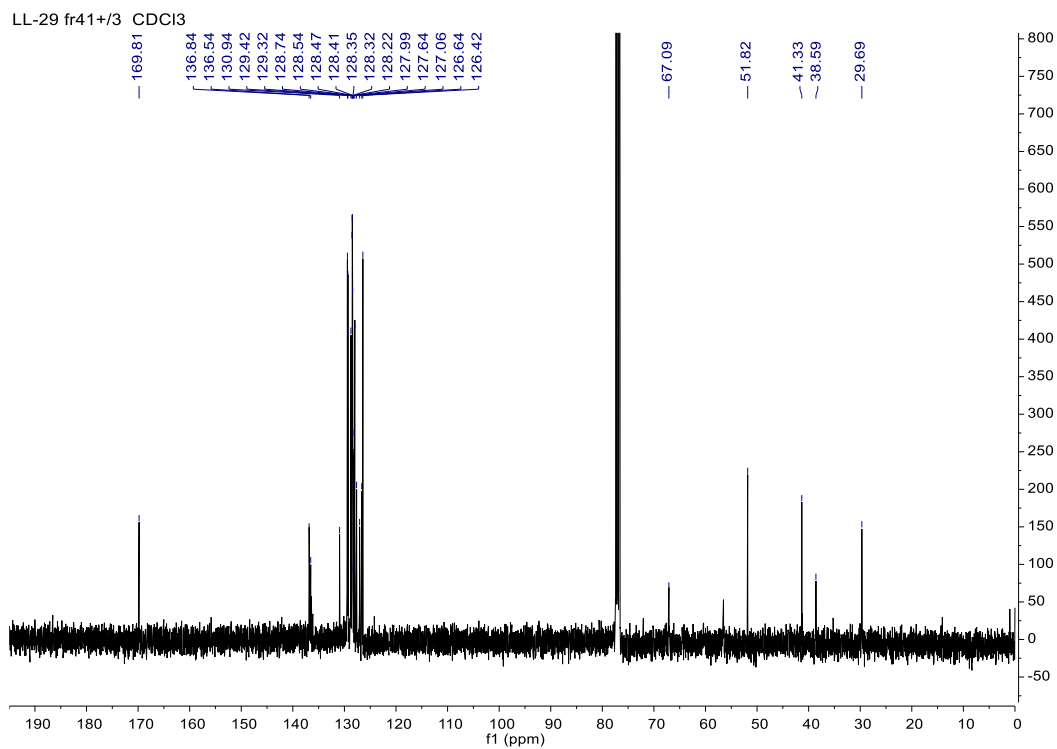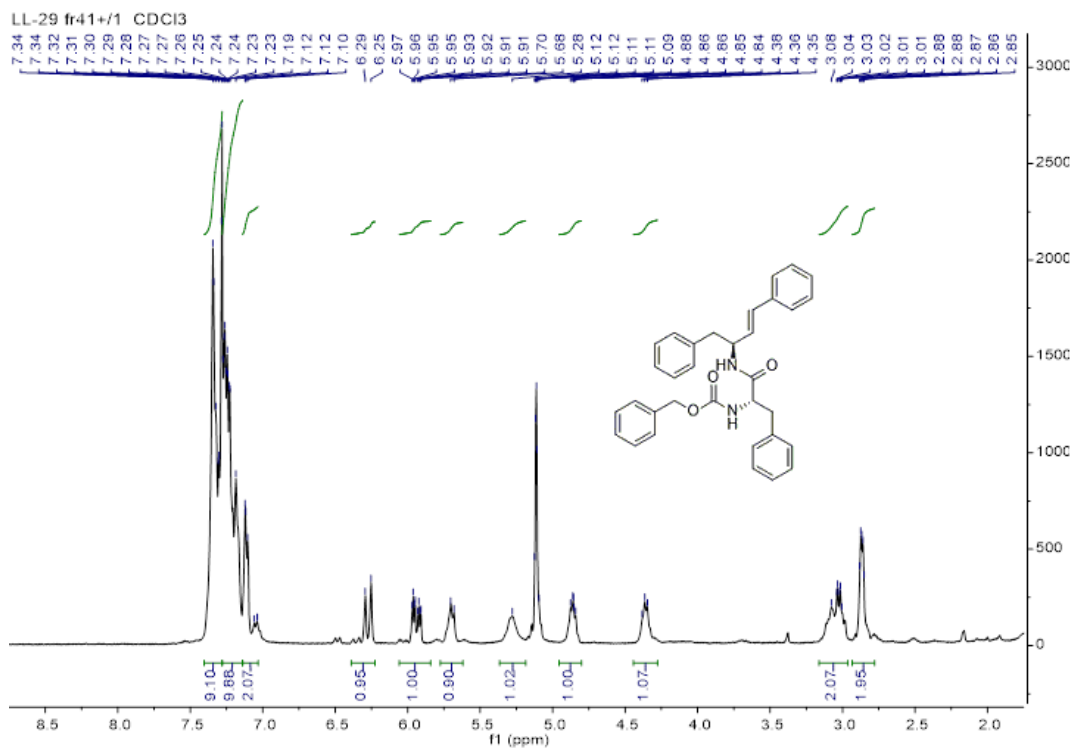
6:

BC-250C, 1HNMR, purified-twice, DMSO-d6

BC-250C, C13NMR, purified-twice, DMSO-d6

7:

BC-239A 1H NMR Purified-crystallized CDCl3

BC-239A C13 NMR Purified-crystallized CDCl3

187

8:

LL-307_fr25/1 CDCl3

LL-307_fr25/2 CDCl3

188
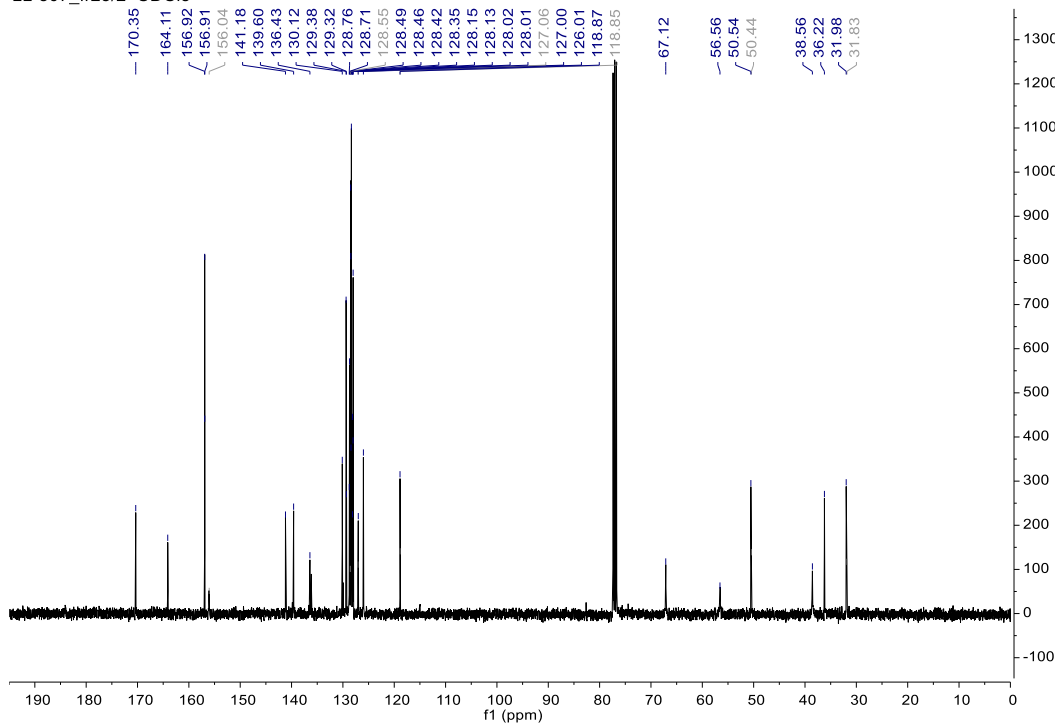
9:

BC-415A, 1H NMR, Purified, CDCl3
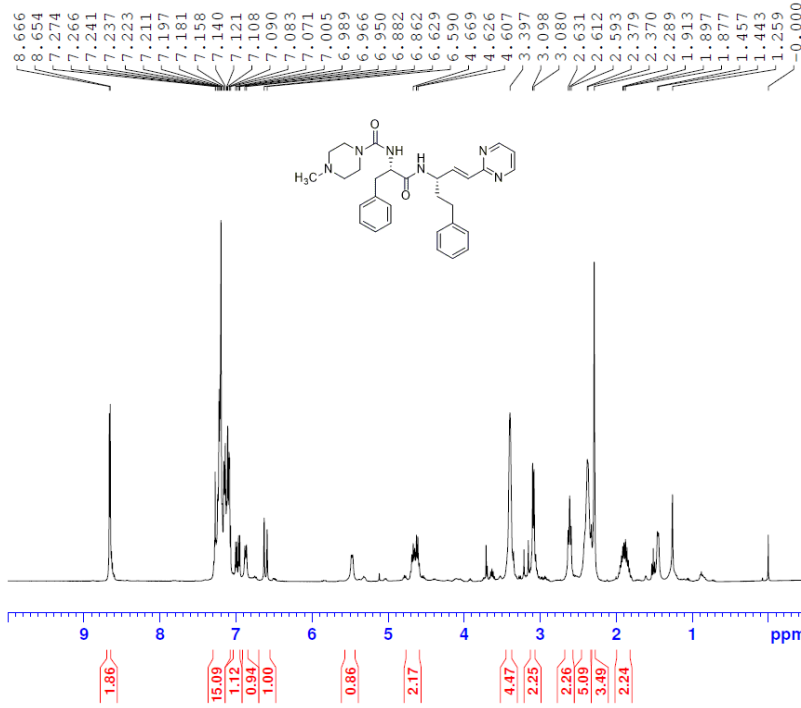
BC-415A, C13 NMR, Purified, CDCl3

10:

BC-308A, 1H NMR, Purified, CDCl3+MeOD
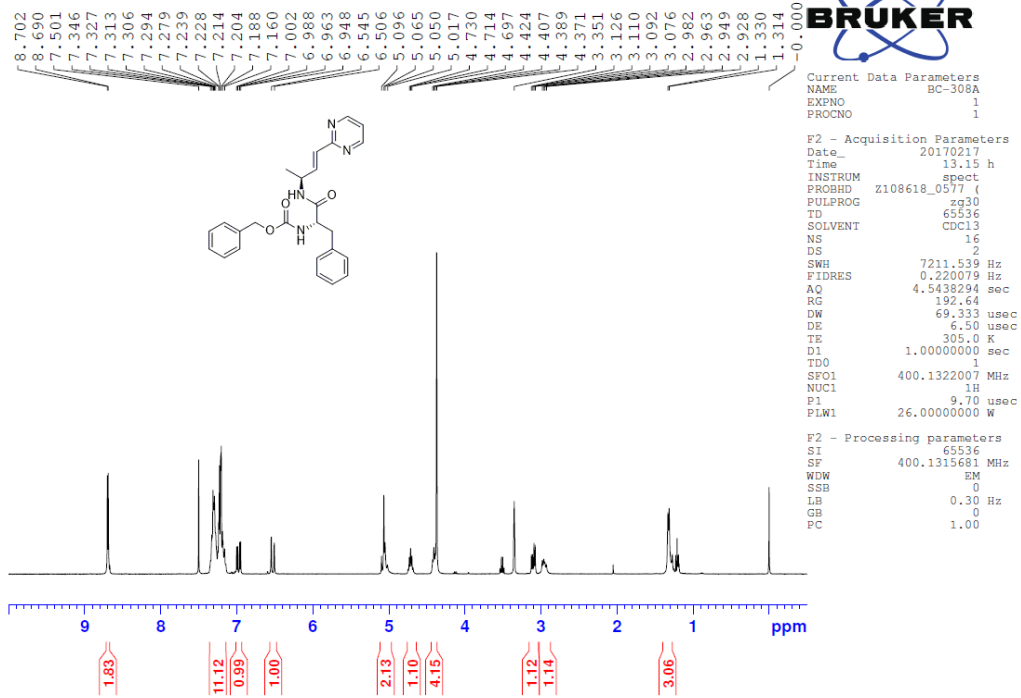
BC-308A, C13 NMR, Purified, CDCl3+MeOD

11:

8.531
8.520
7.603
7.588
7.584
7.569
7.565
7.343
7.324
7.311
7.297
7.280
7.252
7.234
7.217
7.199
7.185
7.182
7.169
7.145
7.125
7.097
7.081
6.610
6.595
6.571
6.556
6.340
6.301
5.891
5.870
5.064
4.928
4.911
4.358
4.340
3.019
3.001
2.900
2.883
2.855
2.838
1.716
-0.001
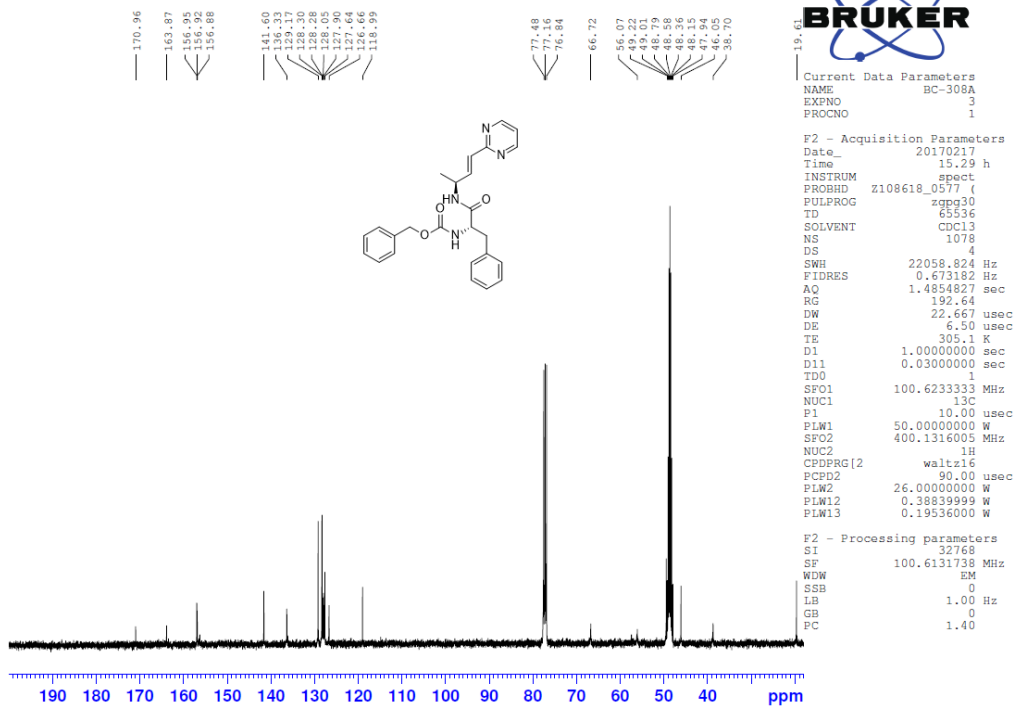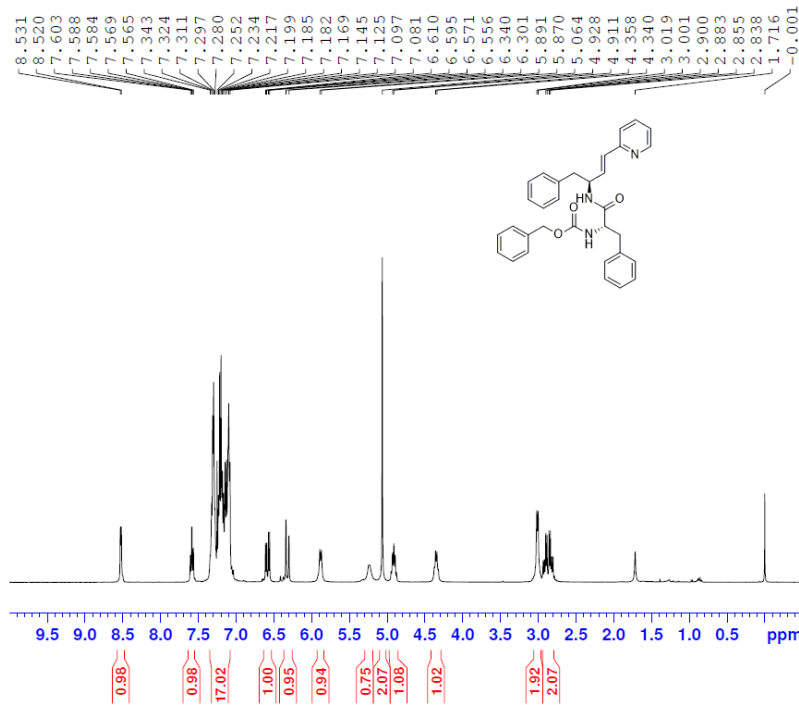
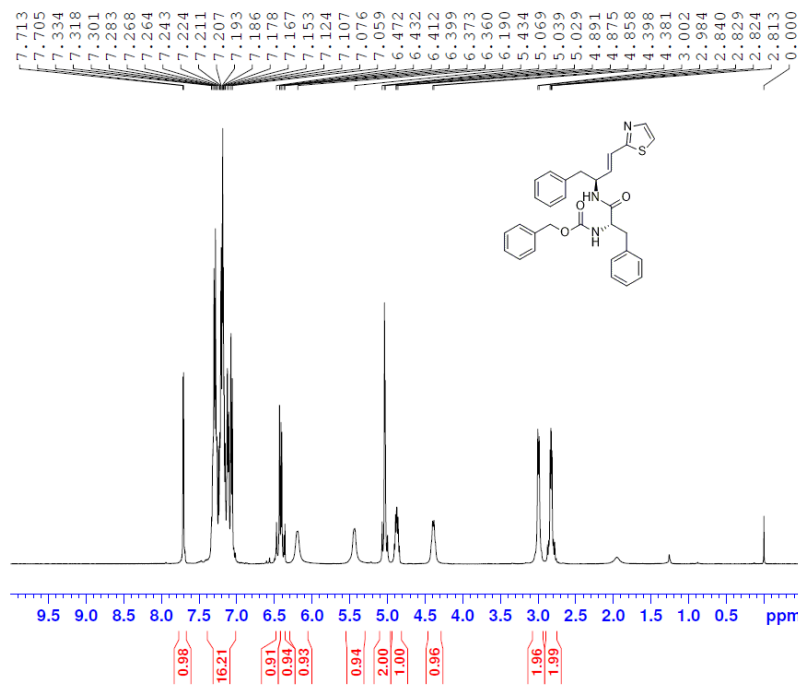Current Data Parameters
NAME   BC-278A
EXPNO   1
PROCNO   1

F2 - Acquisition Parameters
Date_   20161205
Time   10.39 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   16
DS   2
SWH   7211.539 Hz
FIDRES   0.220079 Hz
AQ   4.5438294 sec
RG   100.88
DW   69.333 usec
DE   6.50 usec
TE   305.0 K
D1   1.00000000 sec
TDO   1
SFO1   400.1322007 MHz
NUC1   1H
P1   9.70 usec
PLW1   26.00000000 W

F2 - Processing parameters
SI   65536
SF   400.1300127 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   1.00
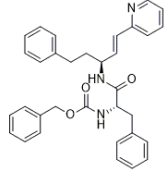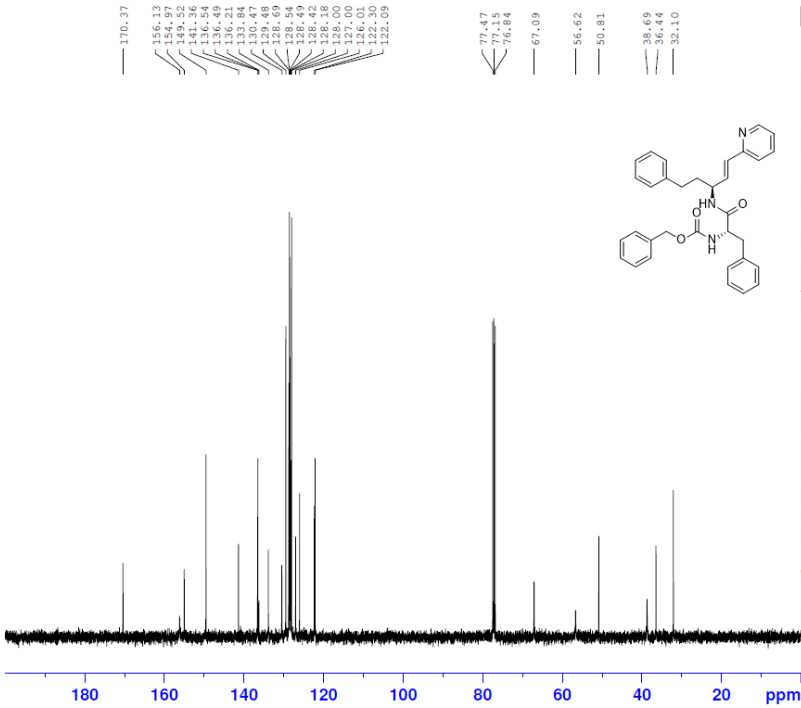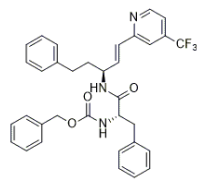
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

0.98 0.98 17.02 1.00 0.95 0.94 0.75 2.07 1.08 1.02   1.92 2.07

169.97
155.92
154.79
149.51
146.79
136.41
136.12
133.03
130.46
129.40
129.35
128.75
128.54
128.48
128.21
128.04
128.00
126.67
122.24
122.05
77.33
77.01
76.69
67.09
56.41
51.65
41.09
38.39
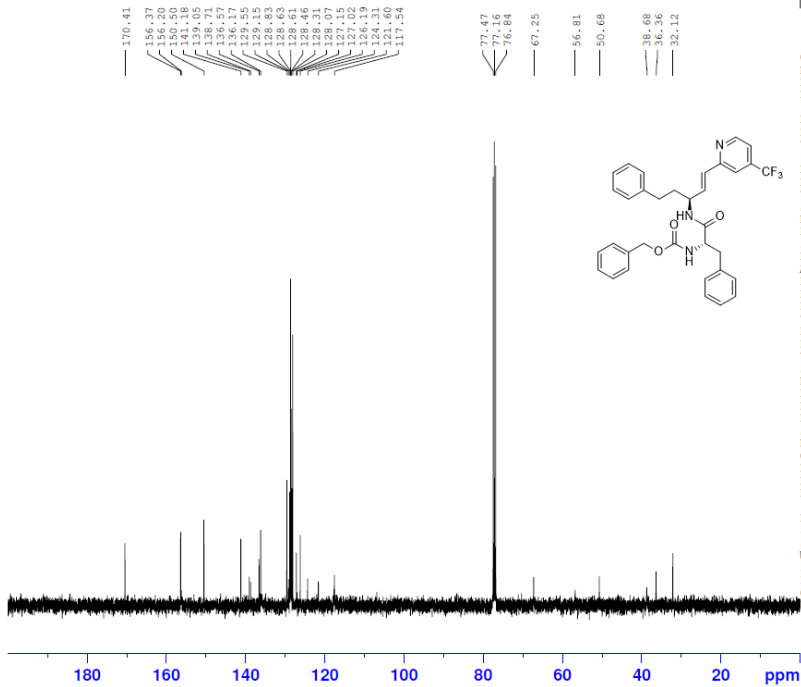
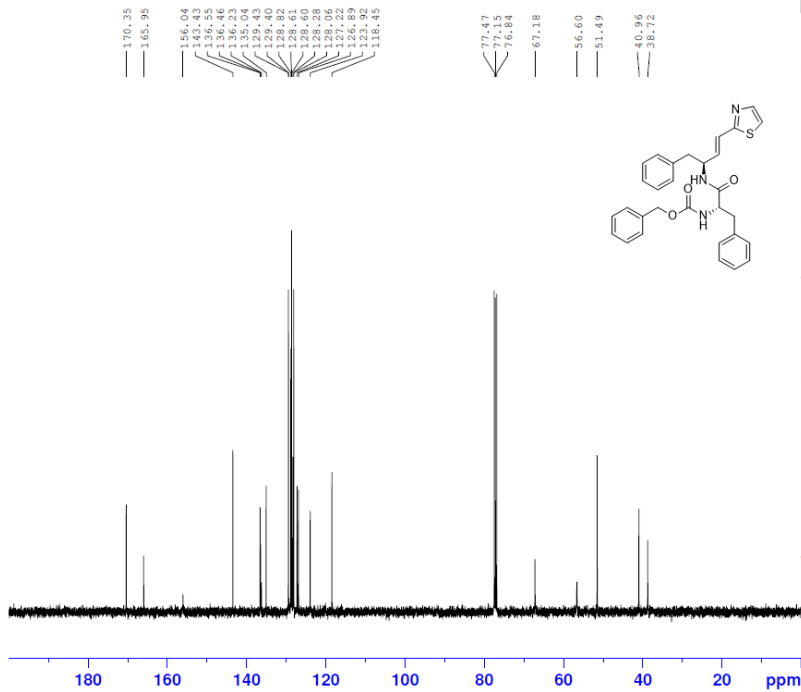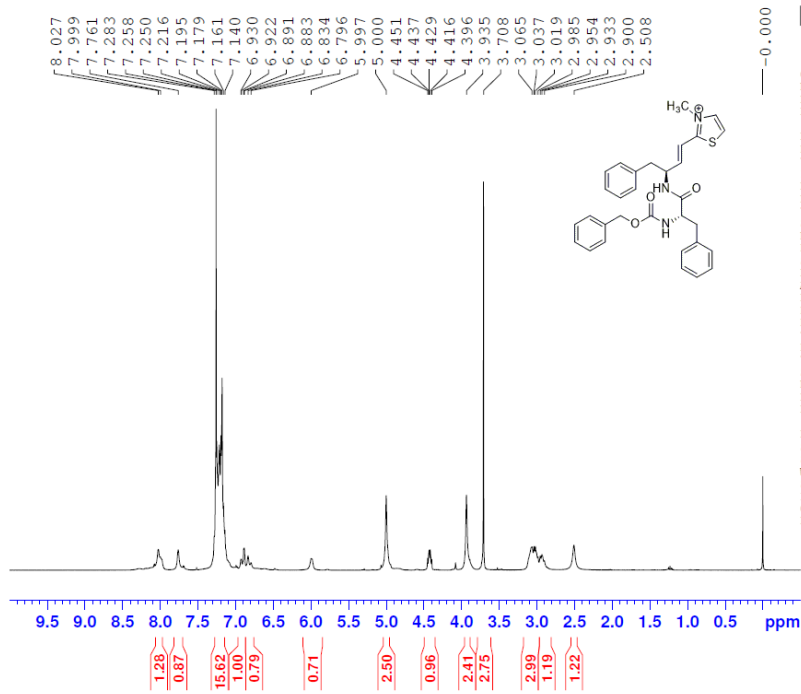Current Data Parameters
NAME   BC-260D
EXPNO   3
PROCNO   1

F2 - Acquisition Parameters
Date_   20161017
Time   16.01 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zgpg30
TD   65536
SOLVENT   CDCl3
NS   353
DS   4
SWH   22058.824 Hz
FIDRES   0.673182 Hz
AQ   1.4854827 sec
RG   192.64
DW   22.667 usec
DE   6.50 usec
TE   304.1 K
D1   1.00000000 sec
D11   0.03000000 sec
TDO   1
SFO1   100.6233333 MHz
NUC1   13C
P1   10.00 usec
PLW1   50.00000000 W
SFO2   400.1316005 MHz
NUC2   1H
CPDPRG[2   waltz16
PCPD2   90.00 usec
PLW2   26.00000000 W
PLW12   0.38839999 W
PLW13   0.19536000 W

F2 - Processing parameters
SI   32768
SF   100.6127696 MHz
WDW   EM
SSB   0
LB   1.00 Hz
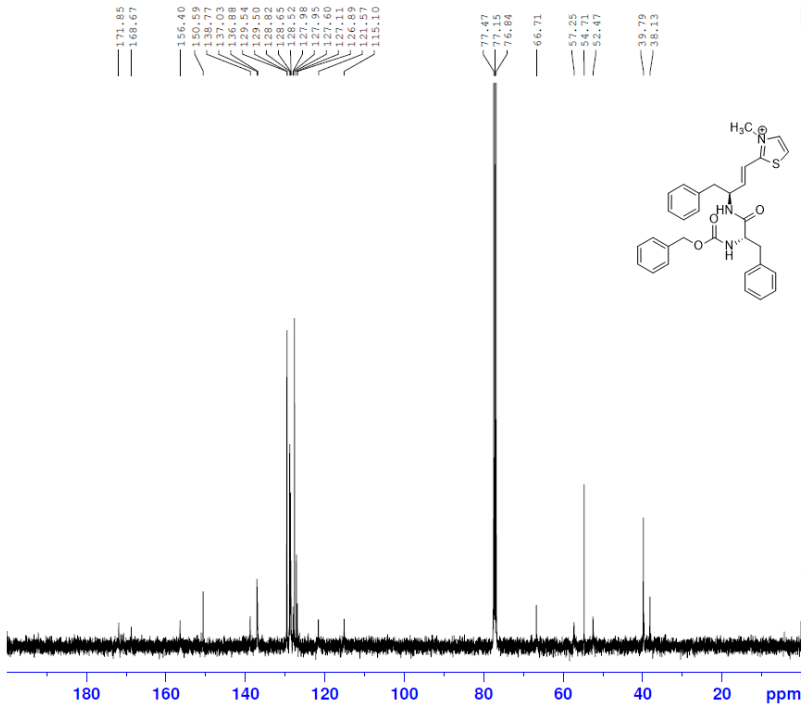GB   0
PC   1.40

180 160 140 120 100 80 60 40 20 ppm

12:

BC-279_ppt_pure, 1HNMR, CDCl3

BC-279_ppt_pure, C13NMR, CDCl3

13

8.515
8.504
7.592
7.589
7.573
7.570
7.554
7.551
7.252
7.248
7.232
7.225
7.206
7.187
7.173
7.158
7.143
7.131
7.104
7.084
7.065
6.562
6.546
6.523
6.507
6.447
6.408
5.600
5.583
5.027
4.667
4.649
4.631
4.456
3.052
3.035
2.605
2.586
2.566
1.890
1.871
1.855
1.839
1.820

0.93 1.00 17.89 1.00 0.98 0.81 0.87 2.00 1.00 0.92 2.06 1.94 2.01

170.37
156.13
154.97
149.52
141.36
136.54
136.49
136.21
133.84
130.47
129.48
128.69
128.54
128.49
128.42
128.18
128.00
127.00
126.01
122.30
122.09
77.47
77.15
76.84
67.09
56.62
50.81
38.69
36.44
32.10

14

194

15

16:

17:

18:

18

19:

LL-106_frl-L_re.1.fid CDCl3

LL-106_forpaper.2.fid CDCl3

20:

BC-566, HPLC purified, MeOD, 1H NMR
PROTON_TAMU MeOD /data bala 4
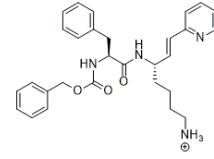
BC-566, HPLC purified, MeOD, C13 NMR
CARBON_TAMU MeOD /data bala 4

21:

LL-062/1 CDCl3

8.51, 8.49, 7.33, 7.31, 7.30, 7.28, 7.24, 7.23, 7.21, 7.20, 7.19, 7.15, 7.14, 7.07, 7.06, 6.13, 5.90, 5.08, 4.86, 4.36, 3.03, 3.01, 2.84, 2.83

2.00, 5.04, 6.46, 5.93, 1.10, 0.98, 2.08, 1.02, 0.99, 2.11, 1.97

LL-062/2 CDCl3

170.05, 150.04, 136.38, 133.42, 129.38, 129.30, 128.75, 128.56, 128.54, 128.42, 128.27, 127.94, 127.10, 126.87, 120.91, 104.98, 67.14, 52.10, 51.55, 41.00, 38.35

201

22:

BC-266, 1HNMR, Pure, MeOD+CDCl3

BC-266, C13 NMR, CDCl3 + MeOD
CARBON_TAMU CDCl3 /data bala 54

23:

203

24:

BC-218A Purified-crystallized 1HNMR, CDCl3

7.541 7.328 7.316 7.307 7.300 7.283 7.251 7.245 7.239 7.229 7.223 7.211 7.204 7.190 7.184 7.173 7.132 7.115 7.101 7.071 7.054 6.515 6.500 6.475 6.460 6.159 6.119 6.006 5.991 5.054 4.893 4.876 4.860 4.363 4.345 3.001 2.982 2.967 2.831 2.814 2.805 1.763 0.000
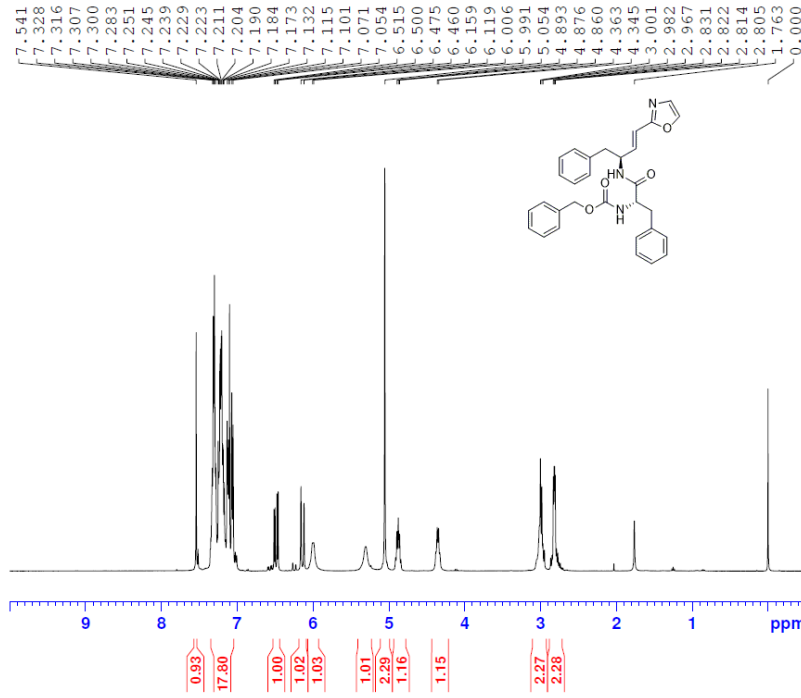
Current Data Parameters
NAME   BC-218A
EXPNO   1
PROCNO   1

F2 - Acquisition Parameters
Date_   20160506
Time   16.29 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   16
DS   2
SWH   7211.539 Hz
FIDRES   0.220079 Hz
AQ   4.5438294 sec
RG   86.84
DW   69.333 usec
DE   6.50 usec
TE   305.0 K
D1   1.00000000 sec
TD0   1
SFO1   400.1322007 MHz
NUC1   1H
P1   9.70 usec
PLW1   26.00000000 W

F2 - Processing parameters
SI   65536
SF   400.1300129 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   1.00

0.93 17.80 1.00 1.02 1.03 1.01 2.29 1.16 1.15 2.27 2.28
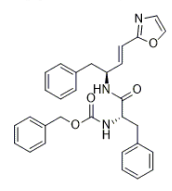
BC-218A, C13 NMR, CDCl3
CARBON_TAMU CDCl3 /data bala 14

170.28 160.75 156.05 138.31 137.11 136.42 136.35 136.23 129.44 129.40 128.73 128.71 128.41 128.22 127.05 127.03 117.27 77.47 77.15 76.84 67.33 56.68 51.46 40.87 38.55
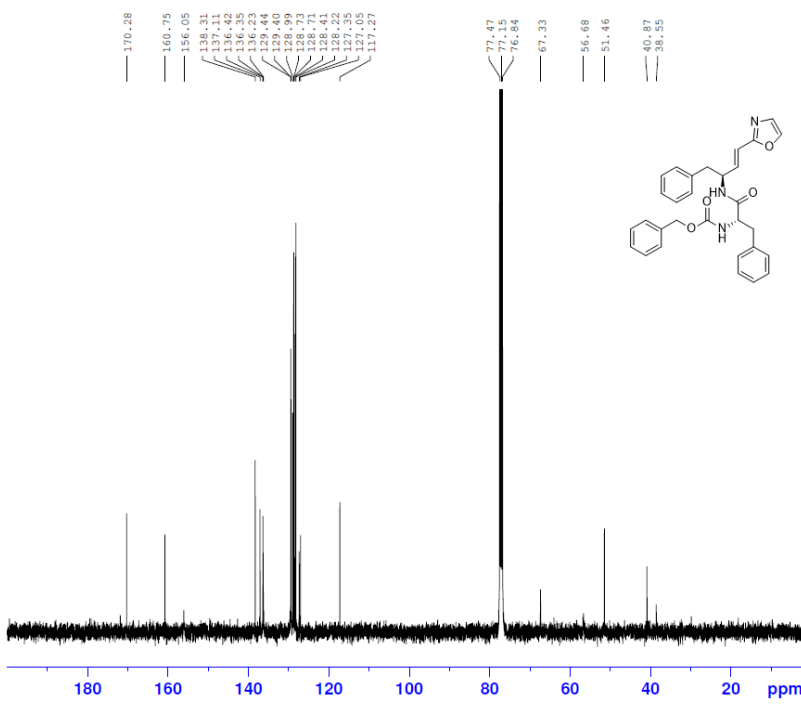
Current Data Parameters
NAME   BC-218A
EXPNO   4
PROCNO   1

F2 - Acquisition Parameters
Date_   20200202
Time   2.06 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zgig30
TD   65536
SOLVENT   CDCl3
NS   4108
DS   4
SWH   24038.461 Hz
FIDRES   0.733596 Hz
AQ   1.3631488 sec
RG   192.64
DW   20.800 usec
DE   6.50 usec
TE   305.0 K
D1   3.00000000 sec
D11   0.03000000 sec
TD0   1
SFO1   100.6228298 MHz
NUC1   13C
P0   3.33 usec
P1   10.00 usec
PLW1   50.00000000 W
SFO2   400.1316005 MHz
NUC2   1H
CPDPRG[2   waltz65
PCPD2   90.00 usec
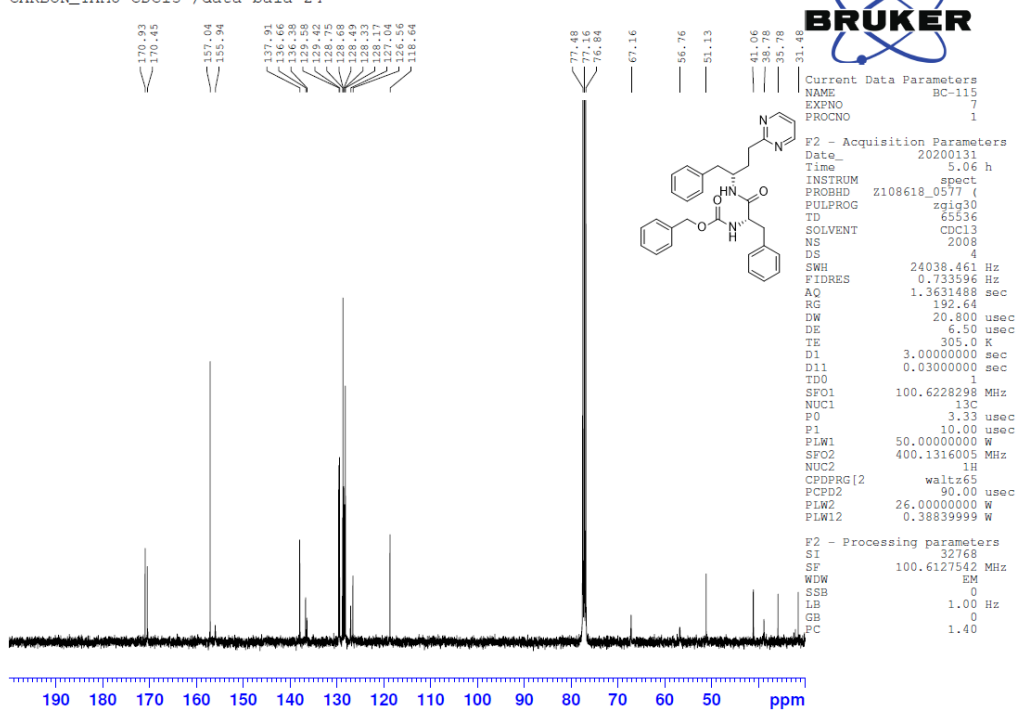PLW2   26.00000000 W
PLW12   0.38839999 W

F2 - Processing parameters
SI   32768
SF   100.6127535 MHz
WDW   EM
SSB   0
LB   1.00 Hz
GB   0
PC   1.40

204

25:

BC-174C, 1H NMR, HPLC Purified, CDCl3

BC-174C, C13 NMR, HPLC Purified, CDCl3

26:

BC-512B, 1H NMR, solid after 3 ether washes, CDCl3

BC-512B, C13 NMR, solid after 3 ether washes, CDCl3

27:

BC-115 1HNMR Purified CDCl3

8.594
8.582
7.330
7.316
7.308
7.254
7.232
7.216
7.198
7.171
7.158
7.152
7.131
7.107
7.104
7.085
7.072
7.059
6.440
6.420
5.066
4.313
4.294
4.132
2.996
2.979
2.883
2.872
2.864
2.858
2.844
2.838
2.824
2.725
2.709
2.689
2.672
1.957
1.945
1.940
1.764
1.750
1.743
1.727

Current Data Parameters
NAME   BC-115
EXPNO   1
PROCNO   1

F2 - Acquisition Parameters
Date_   20150707
Time   13.17
INSTRUM   spect
PROBHD   5 mm PABBO BB/
PULPROG   zg30
TD   65536
SOLVENT   CDCl3
NS   16
DS   2
SWH   7211.539 Hz
FIDRES   0.110039 Hz
AQ   4.5438294 sec
RG   55.4
DW   69.333 usec
DE   6.50 usec
TE   305.0 K
D1   1.00000000 sec
TD0   1

======== CHANNEL f1 ========
SFO1   400.1322007 MHz
NUC1   1H
P1   9.70 usec
PLW1   26.00000000 W

F2 - Processing parameters
SI   65536
SF   400.1300120 MHz
WDW   EM
SSB   0
LB   0.30 Hz
GB   0
PC   1.00

2.00 15.75 0.87 0.98 1.91 1.05 1.10 1.99 2.39 1.77 1.23 1.44

BC-115, C13 NMR, CDCl3
CARBON_TAMU CDCl3 /data bala 24

170.93
170.45
157.04
155.94
137.91
136.66
136.38
129.58
129.42
128.75
128.68
128.49
128.33
128.17
127.04
126.56
118.64
77.48
77.16
76.84
67.16
56.76
51.13
41.06
38.78
35.78
31.48

Current Data Parameters
NAME   BC-115
EXPNO   7
PROCNO   1

F2 - Acquisition Parameters
Date_   20200131
Time   5.06 h
INSTRUM   spect
PROBHD   Z108618_0577 (
PULPROG   zgig30
TD   65536
SOLVENT   CDCl3
NS   2008
DS   4
SWH   24038.461 Hz
FIDRES   0.733596 Hz
AQ   1.3631488 sec
RG   192.64
DW   20.800 usec
DE   6.50 usec
TE   305.0 K
D1   3.00000000 sec
D11   0.03000000 sec
TD0   1
SFO1   100.6228298 MHz
NUC1   13C
P0   3.33 usec
P1   10.00 usec
PLW1   50.00000000 W
SFO2   400.1316005 MHz
NUC2   1H
CPDPRG[2   waltz65
PCPD2   90.00 usec
PLW2   26.00000000 W
PLW12   0.38839999 W

F2 - Processing parameters
SI   32768
SF   100.6127542 MHz
WDW   EM
SSB   0
LB   1.00 Hz
GB   0
PC   1.40

207

SMAI compounds.

Note: The presence of lactol anomers in some of the SMAIs led to complicated splitting and low-quality NMR spectra; therefore, the NMR spectra of corresponding precursors (the second last products) were provided.

1:

2:

3:

4:

6:

212

7:

8:

9:

10:

12:

13:

14:

15:

16:

17:

222

18:

aldehyde proton range
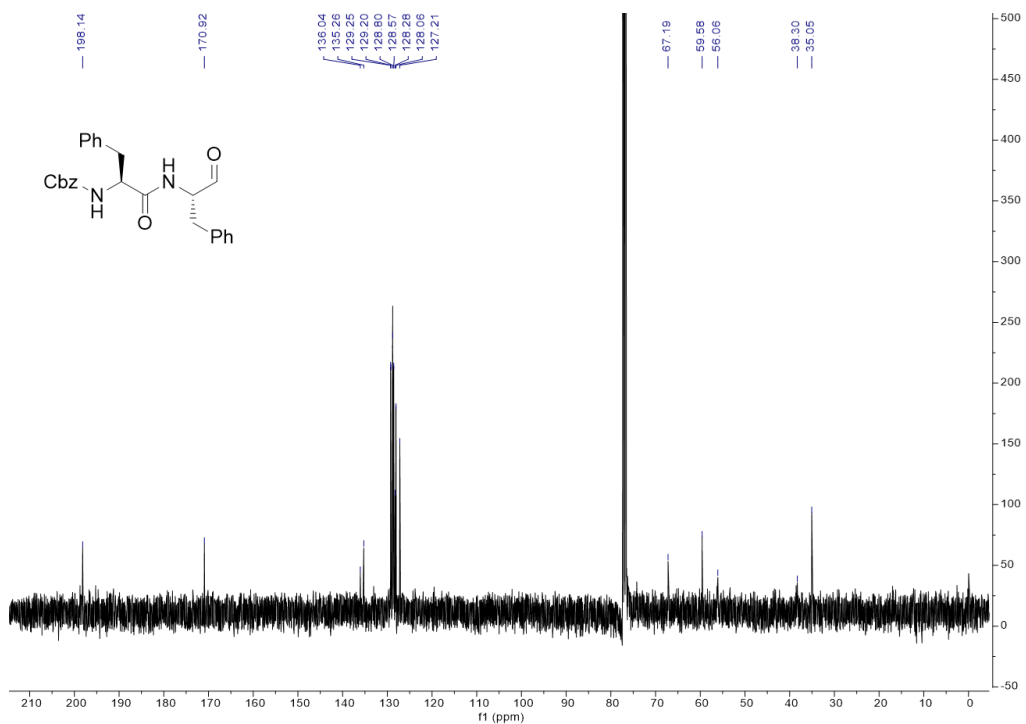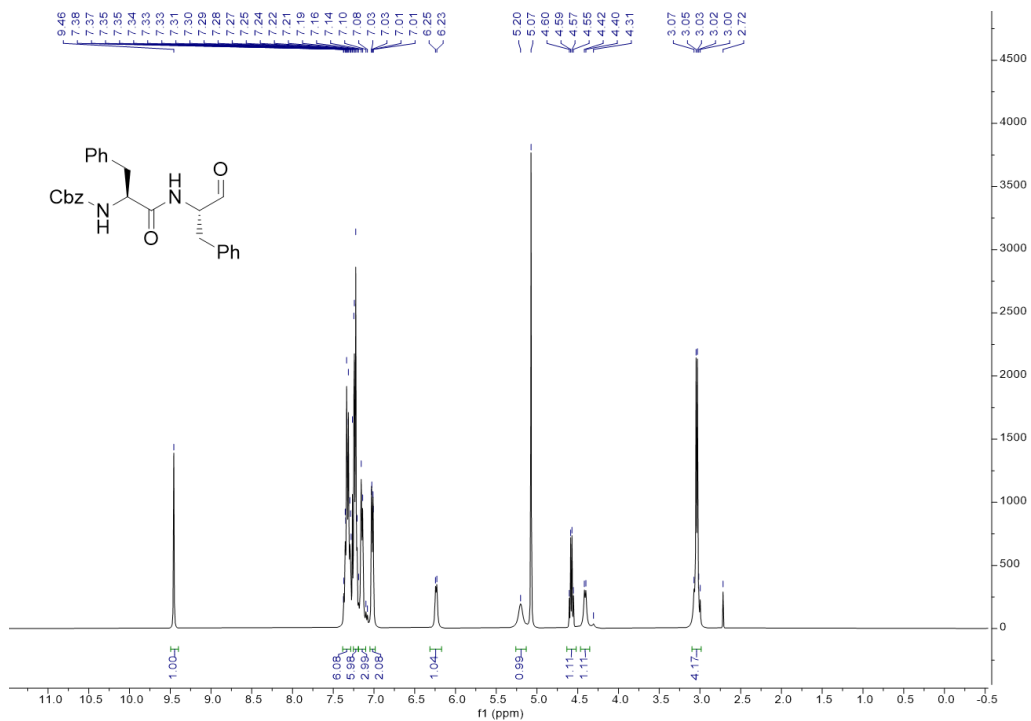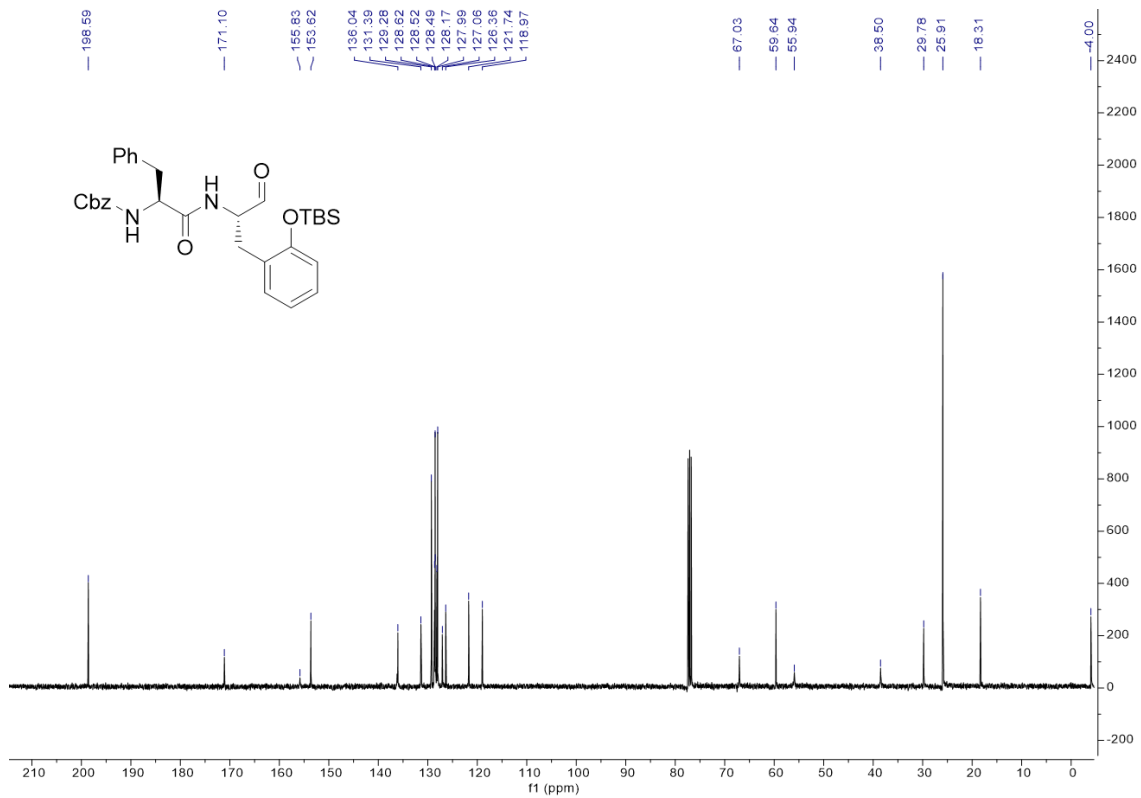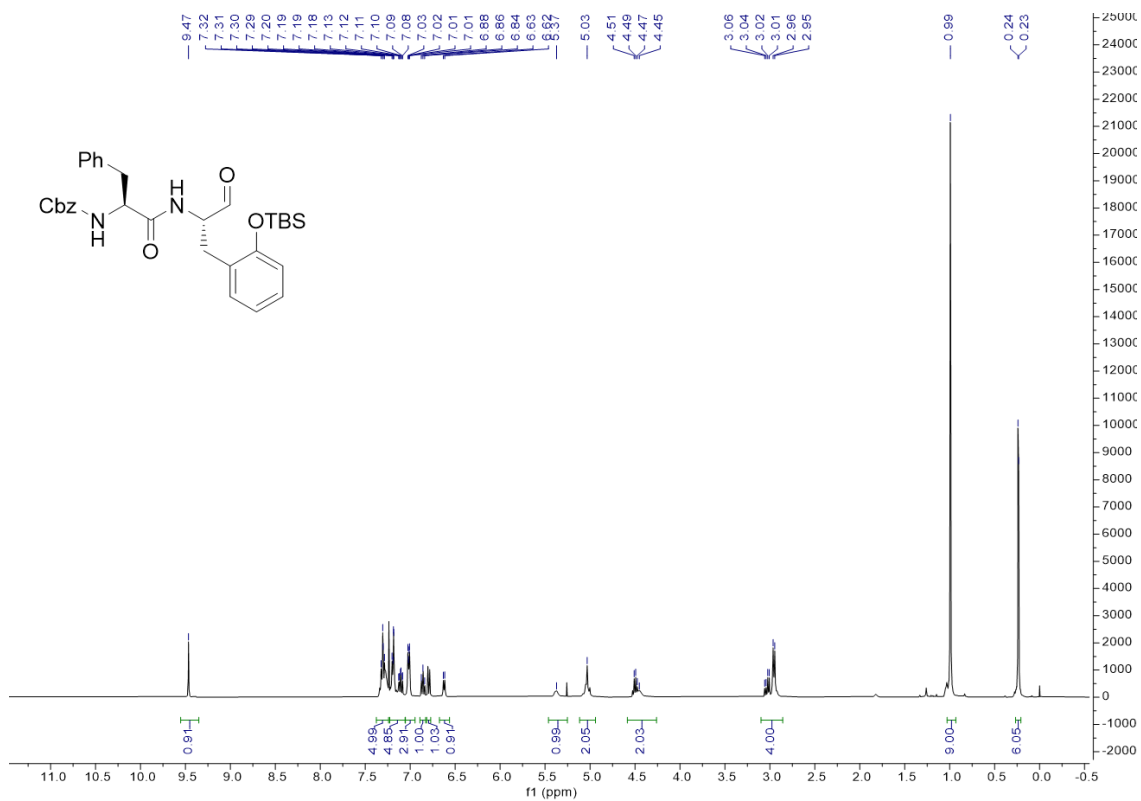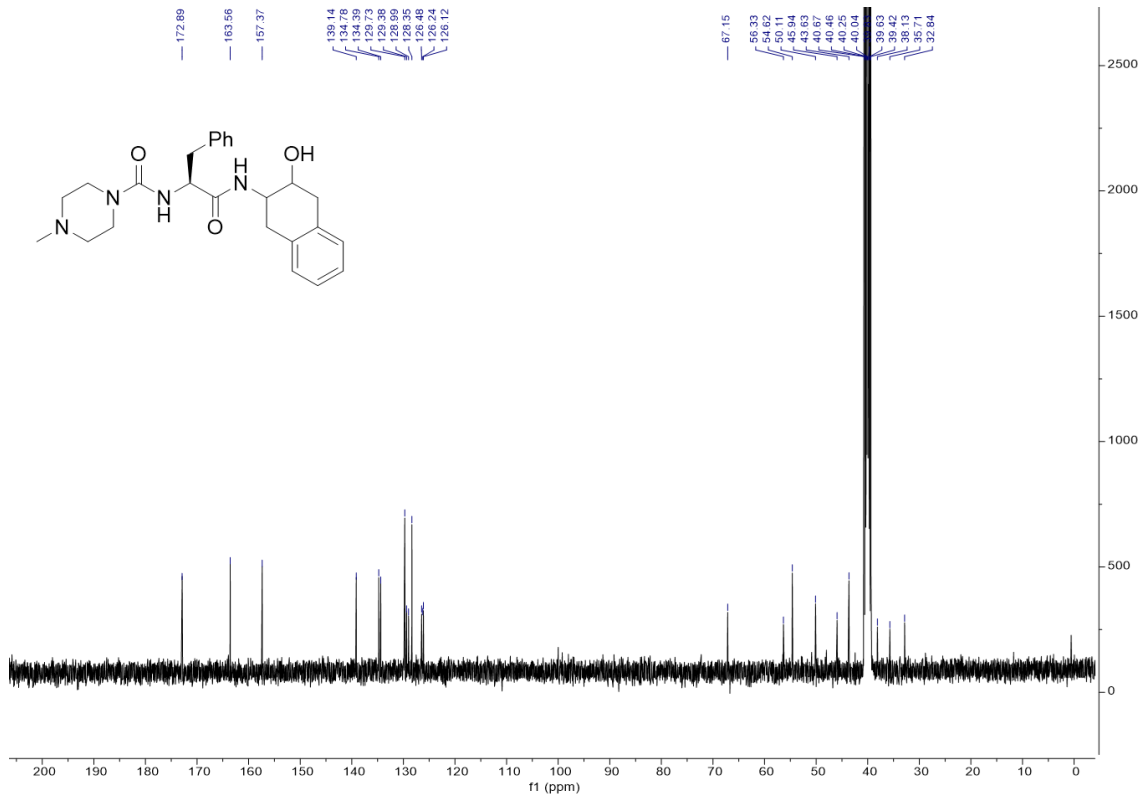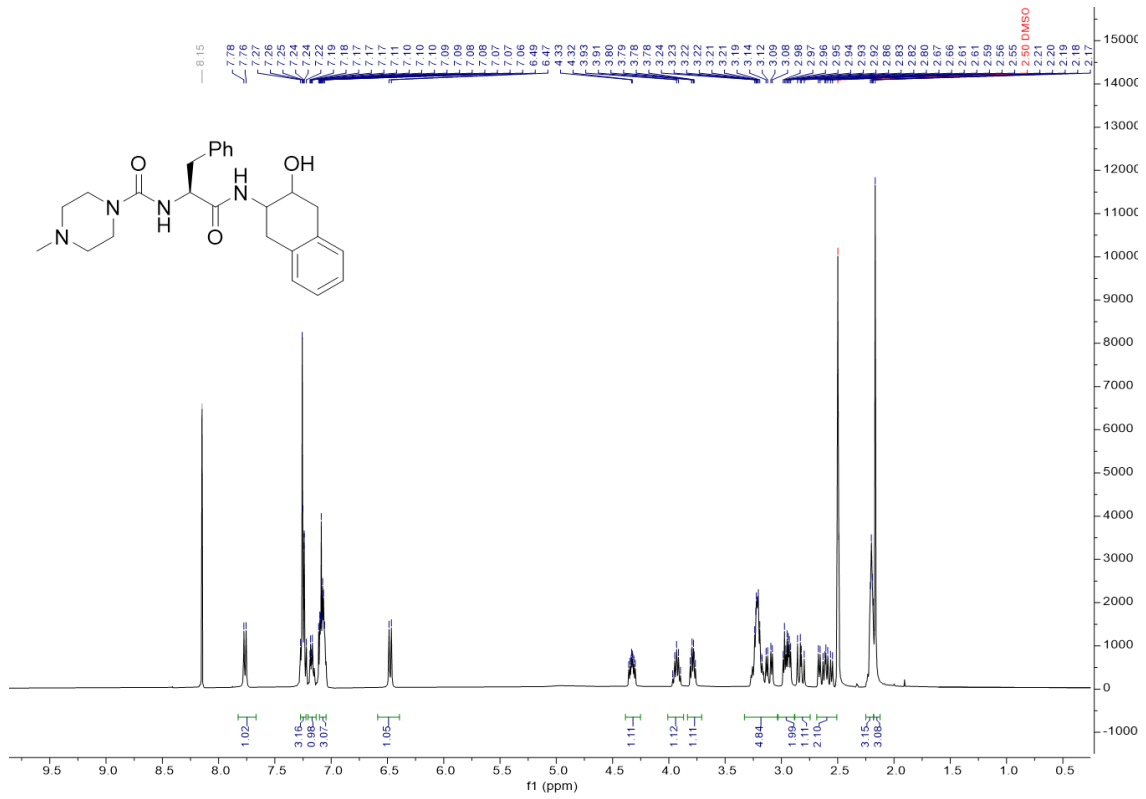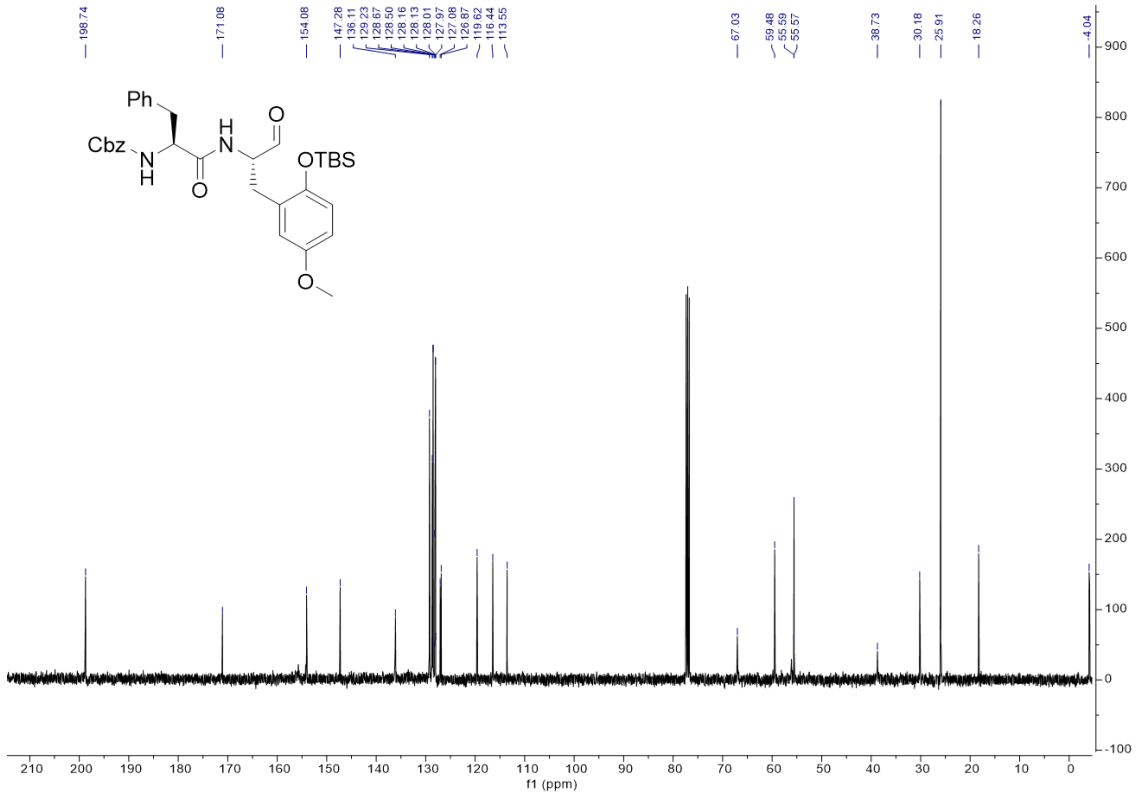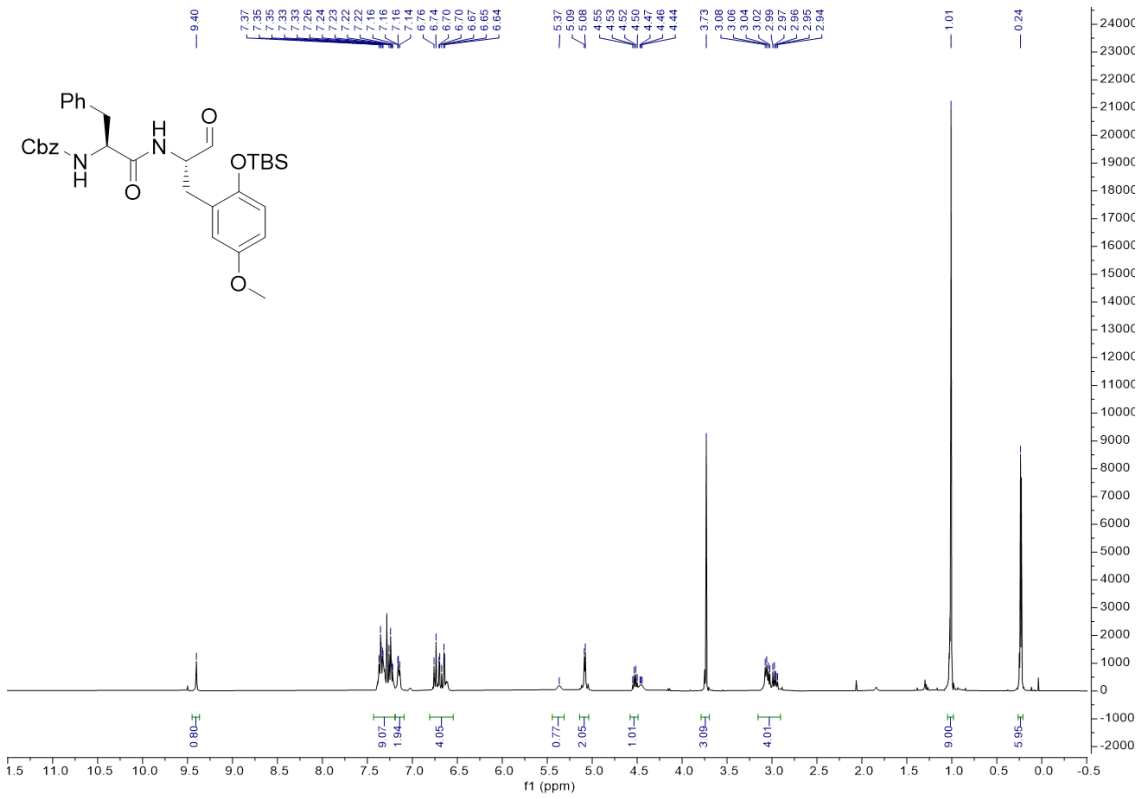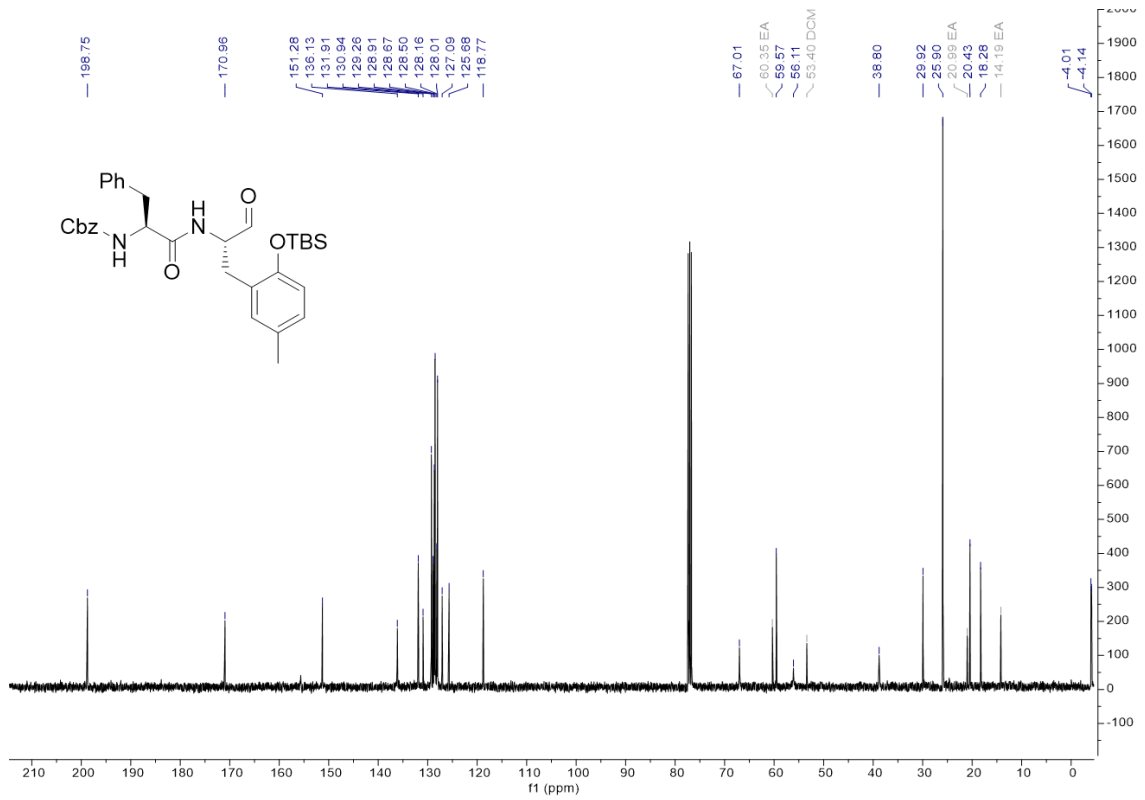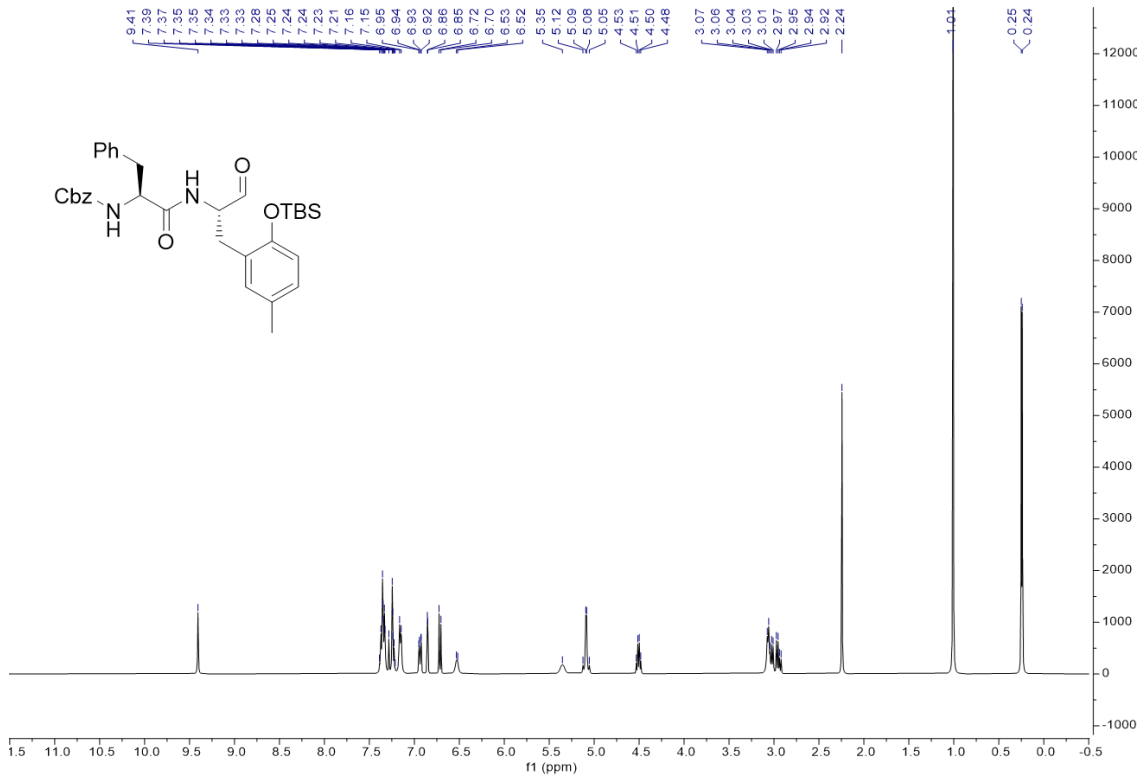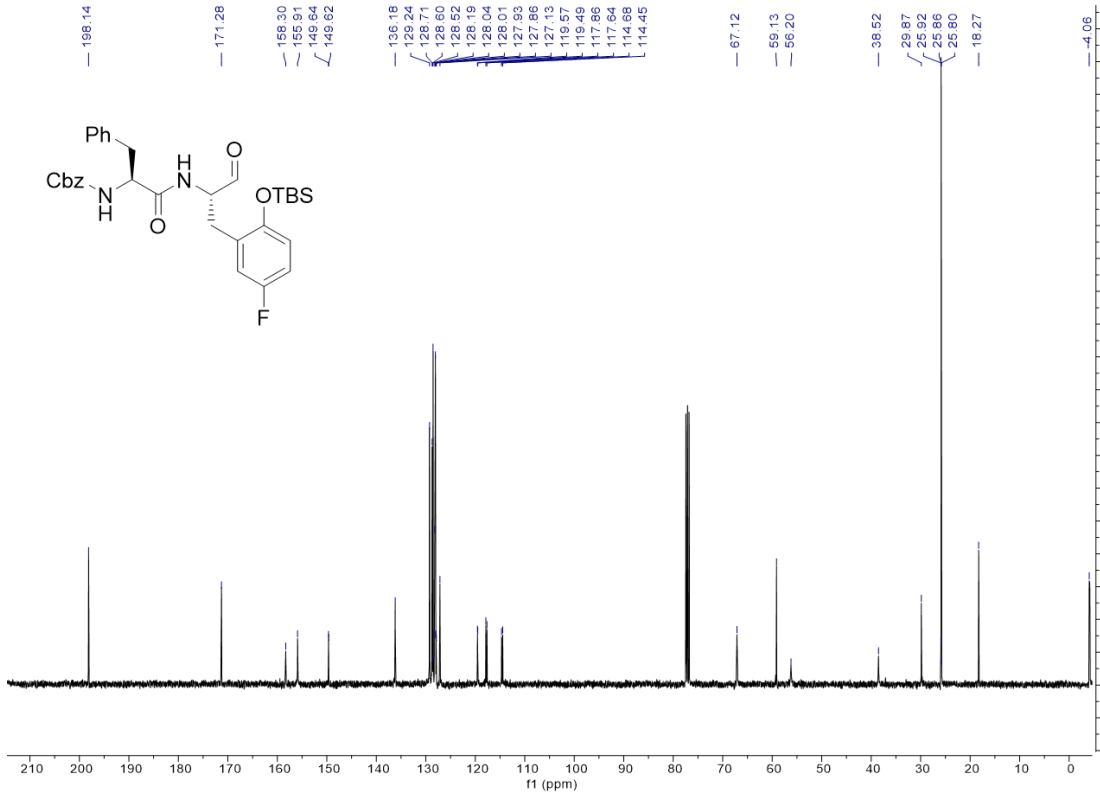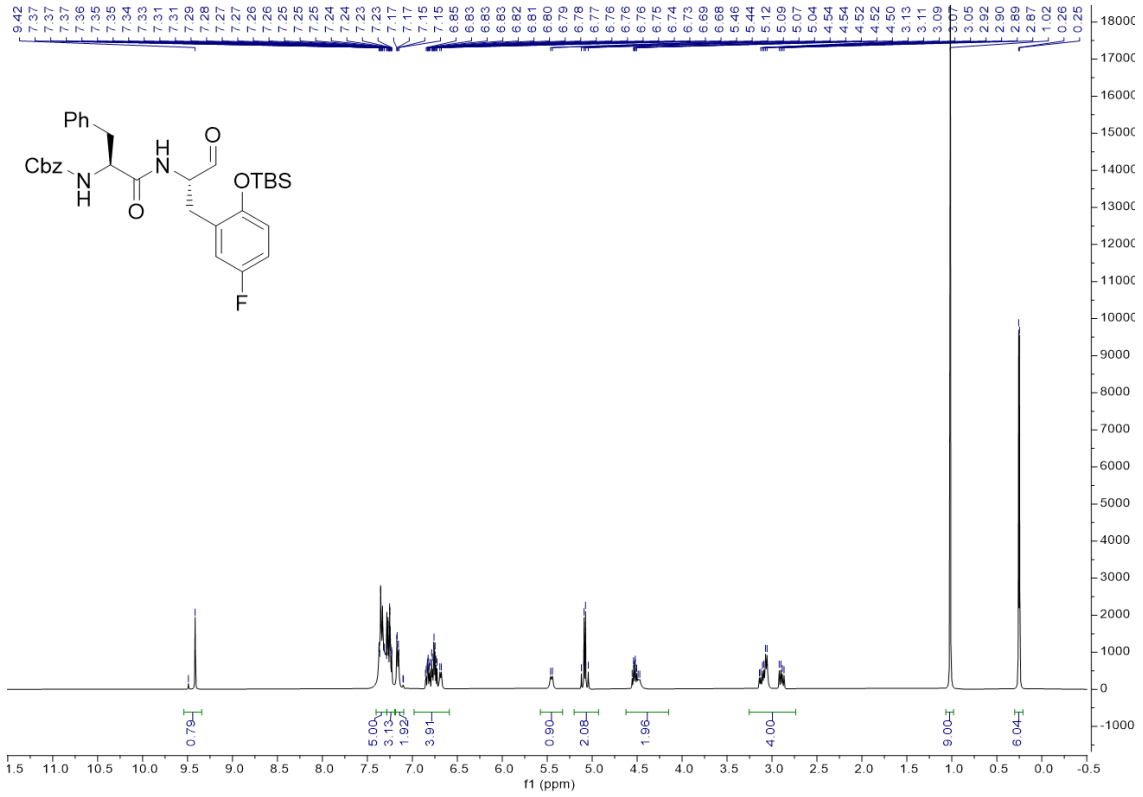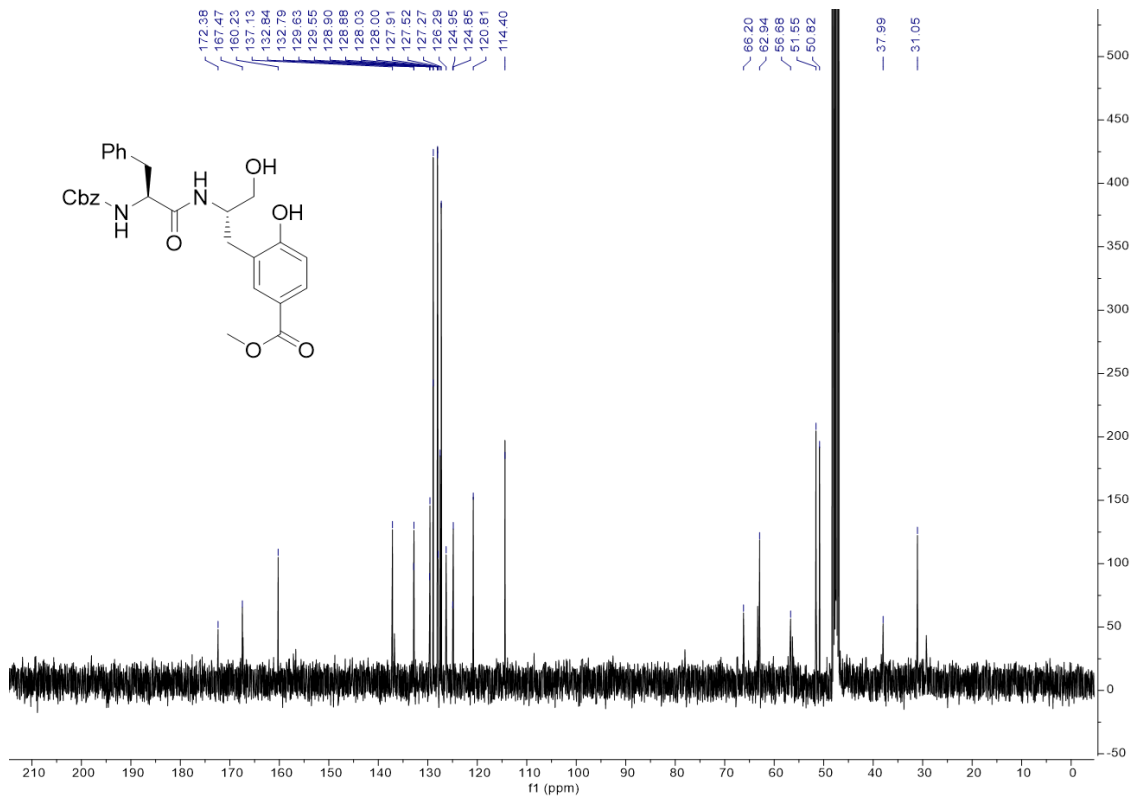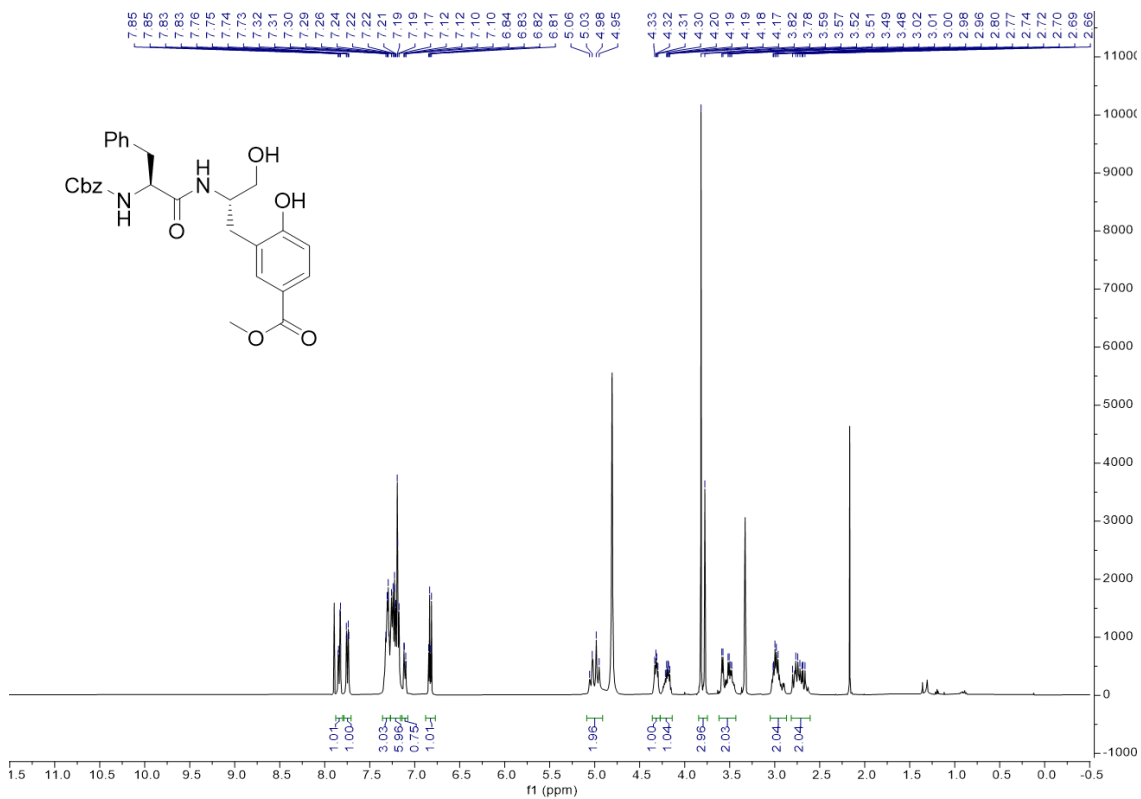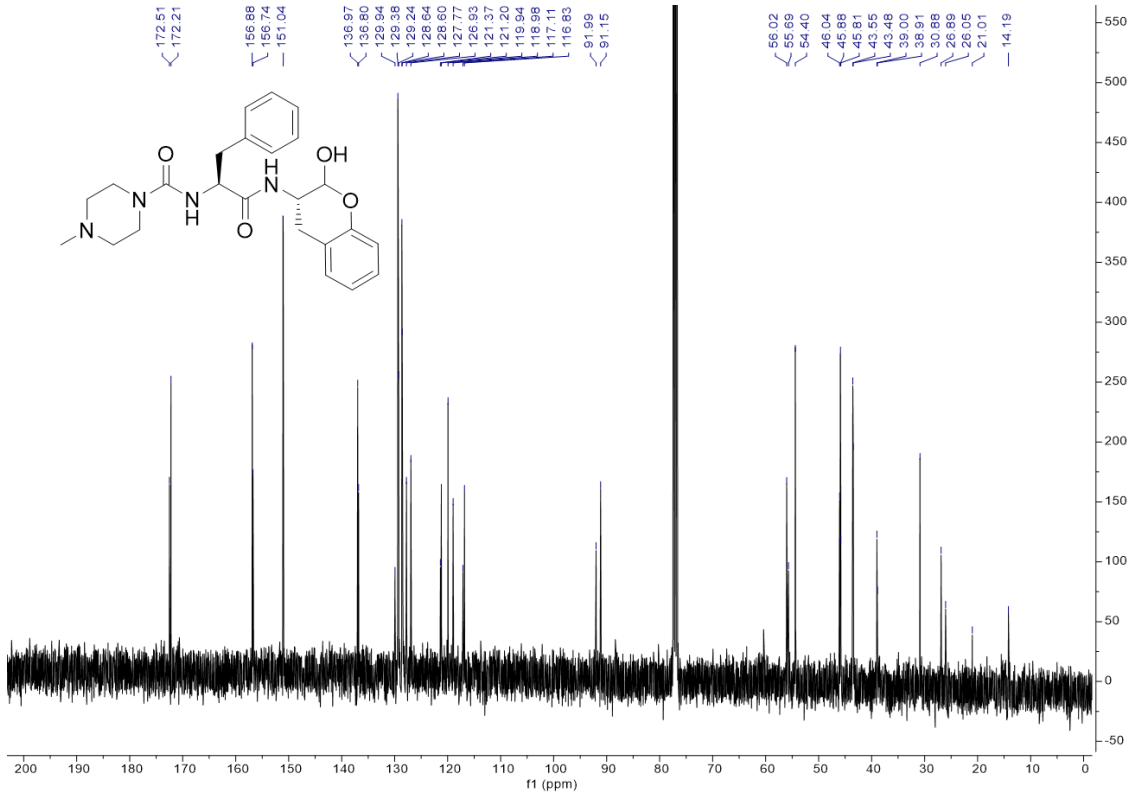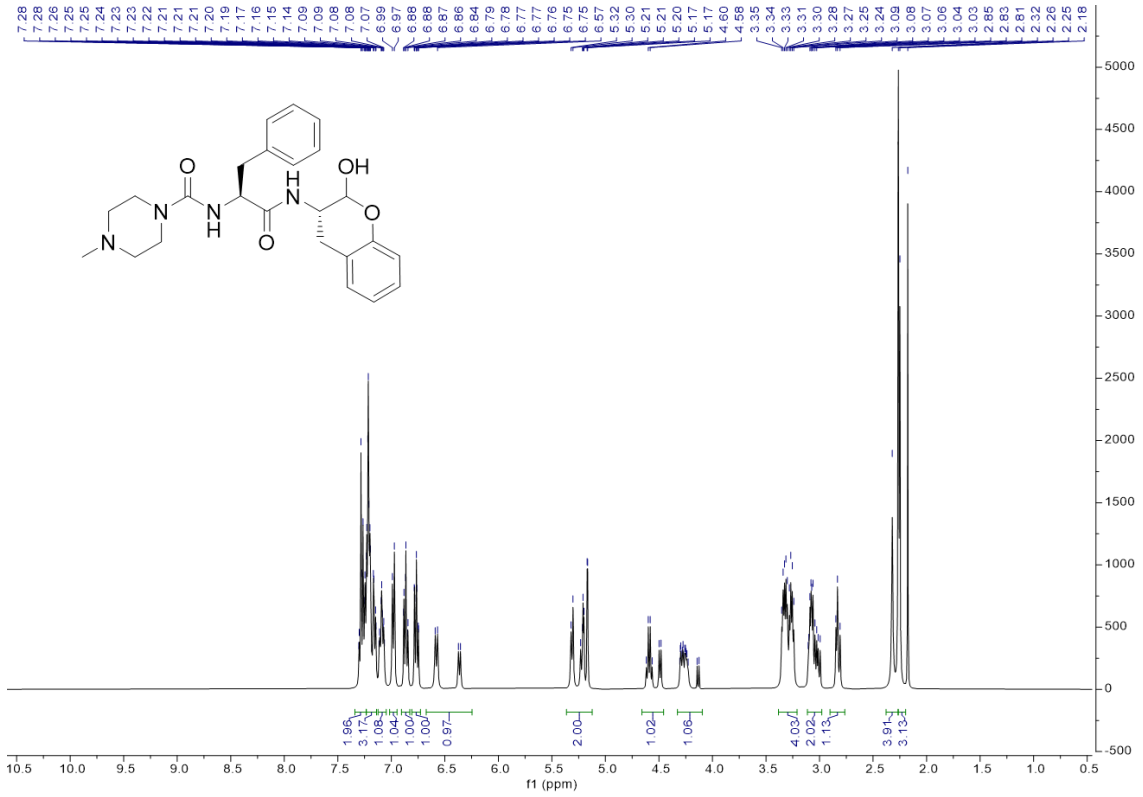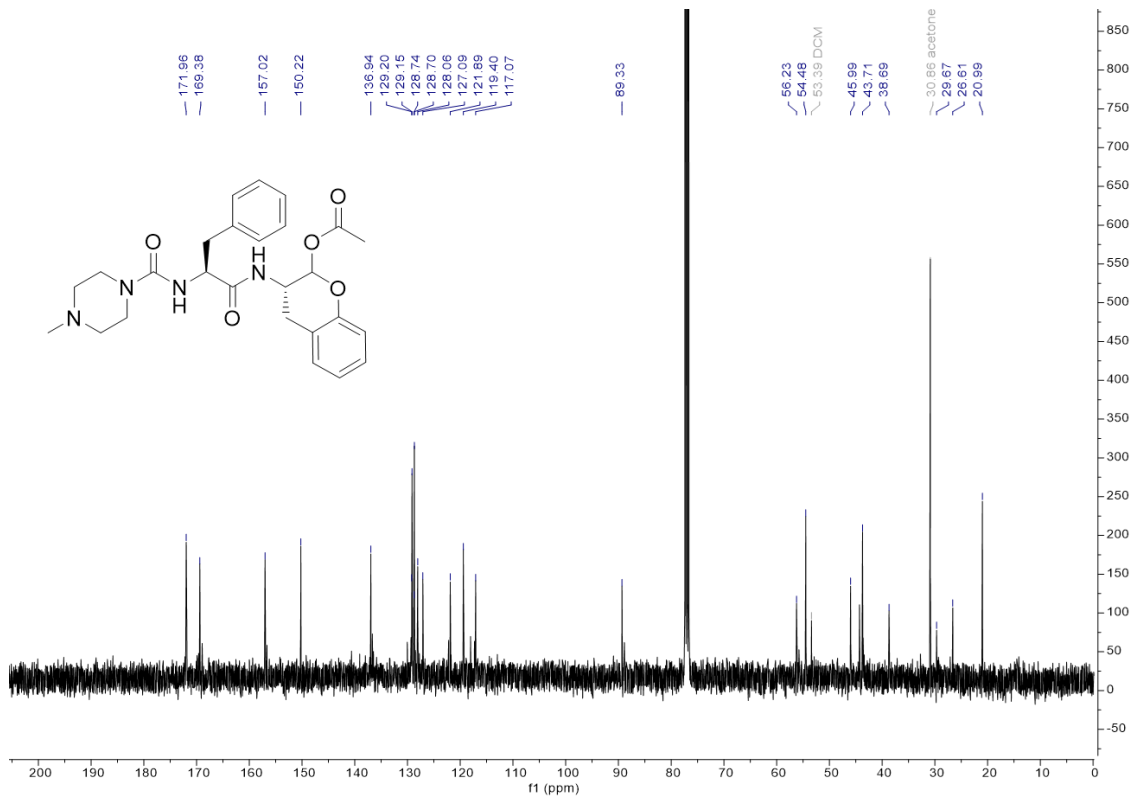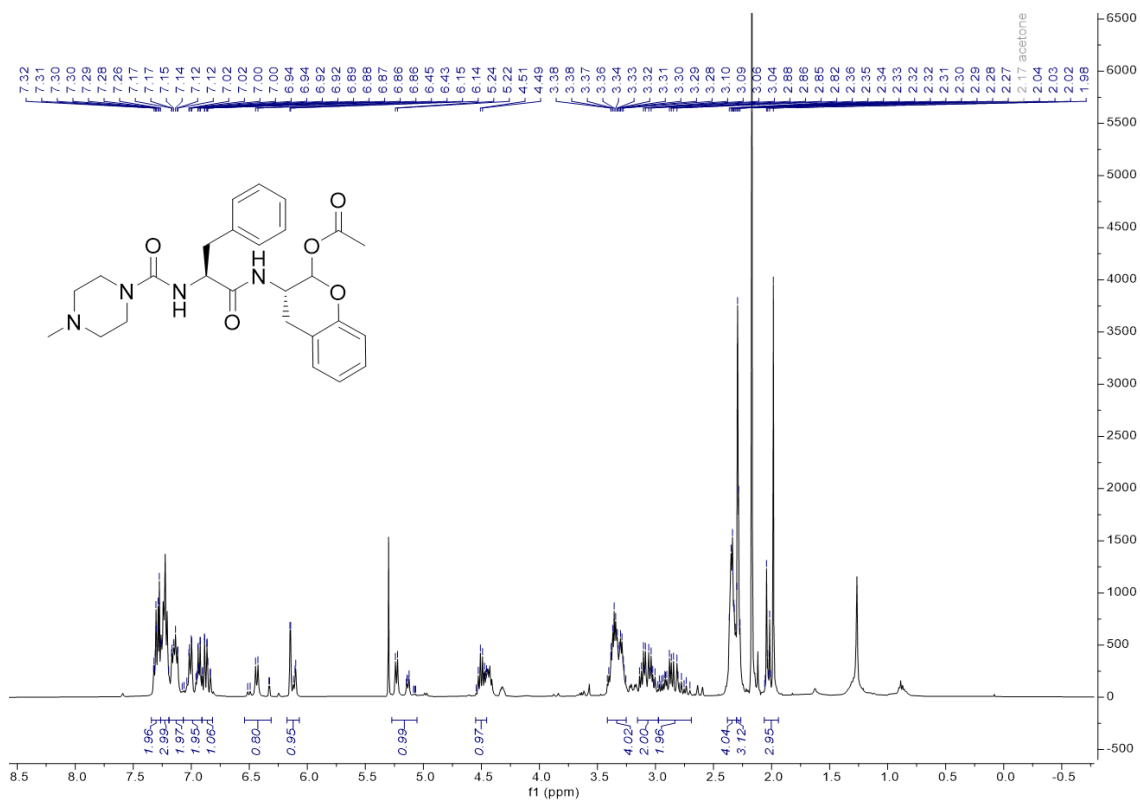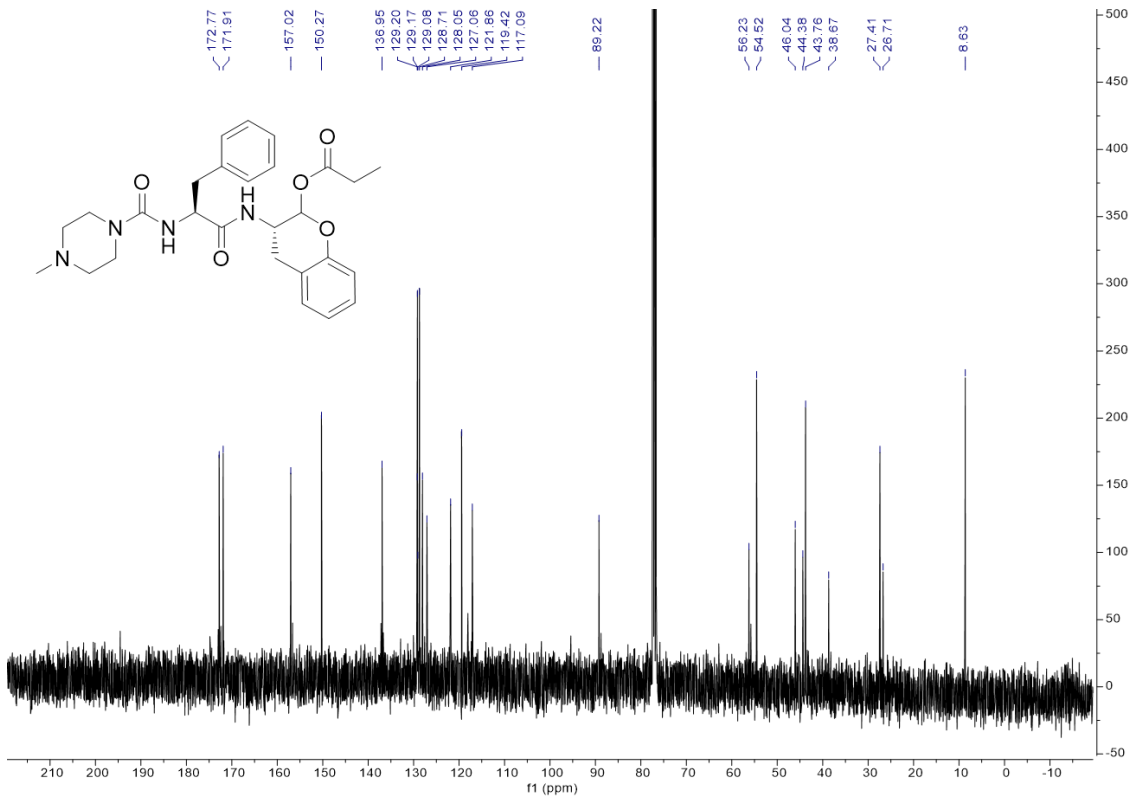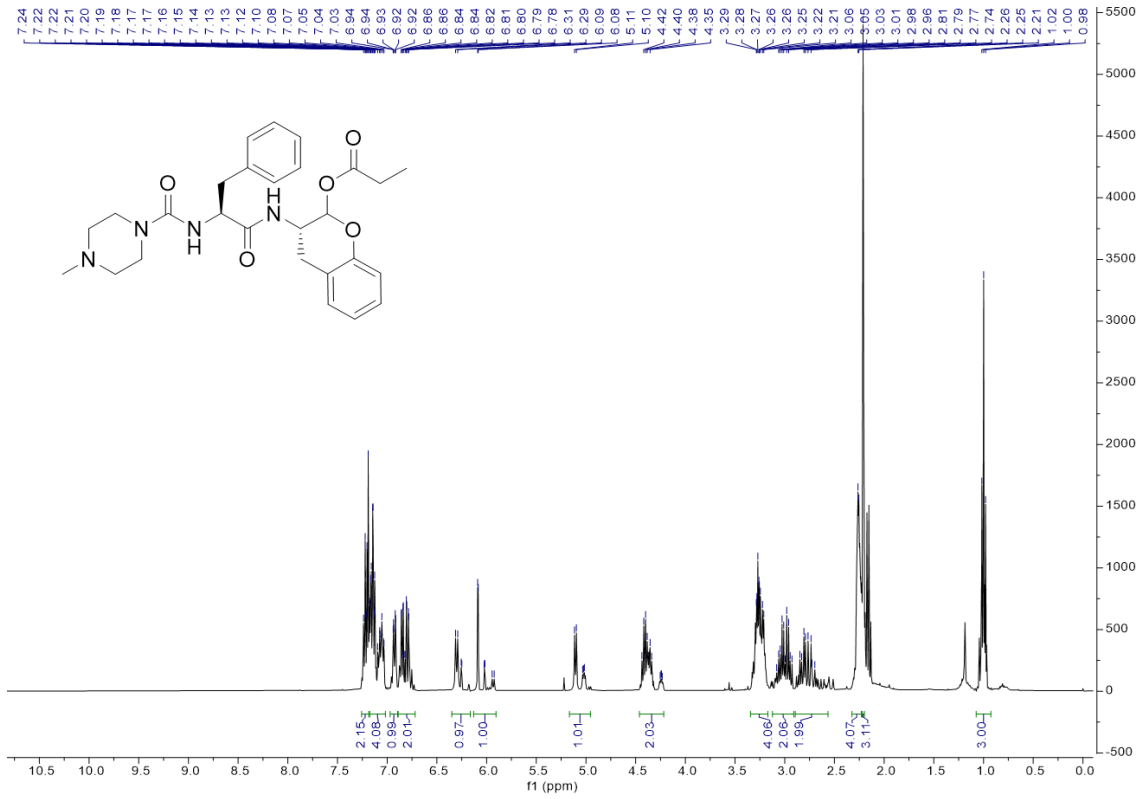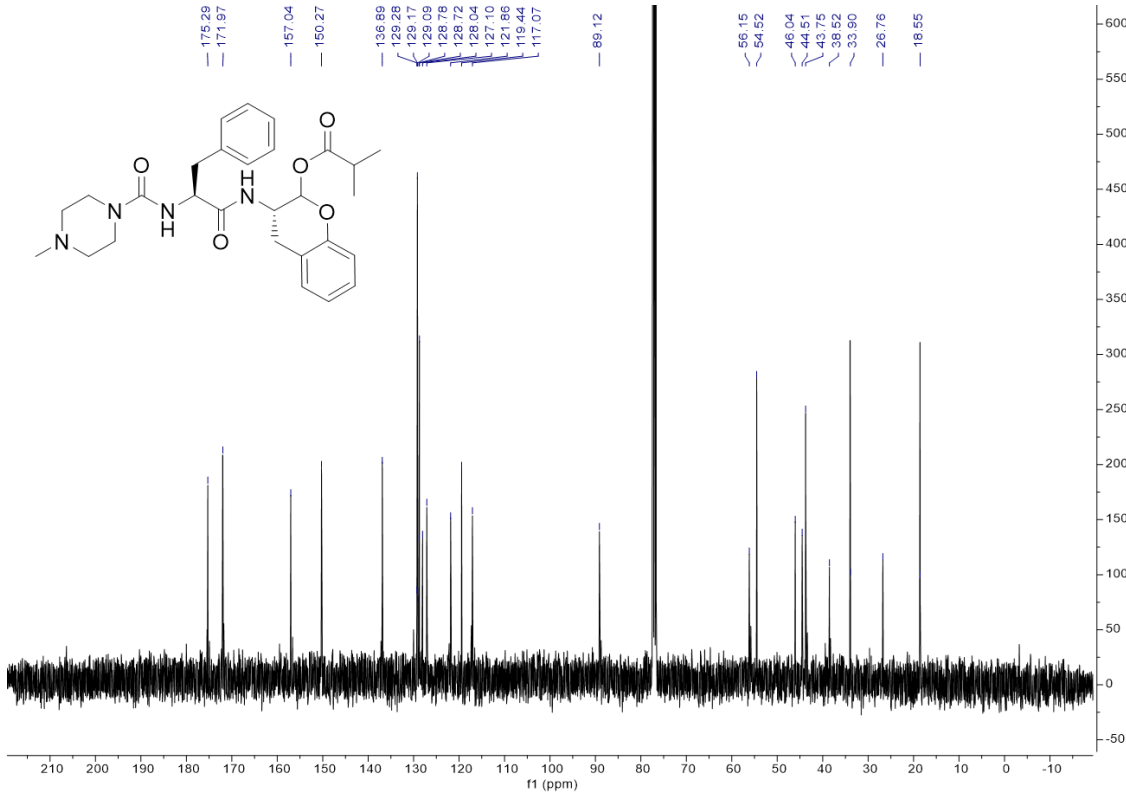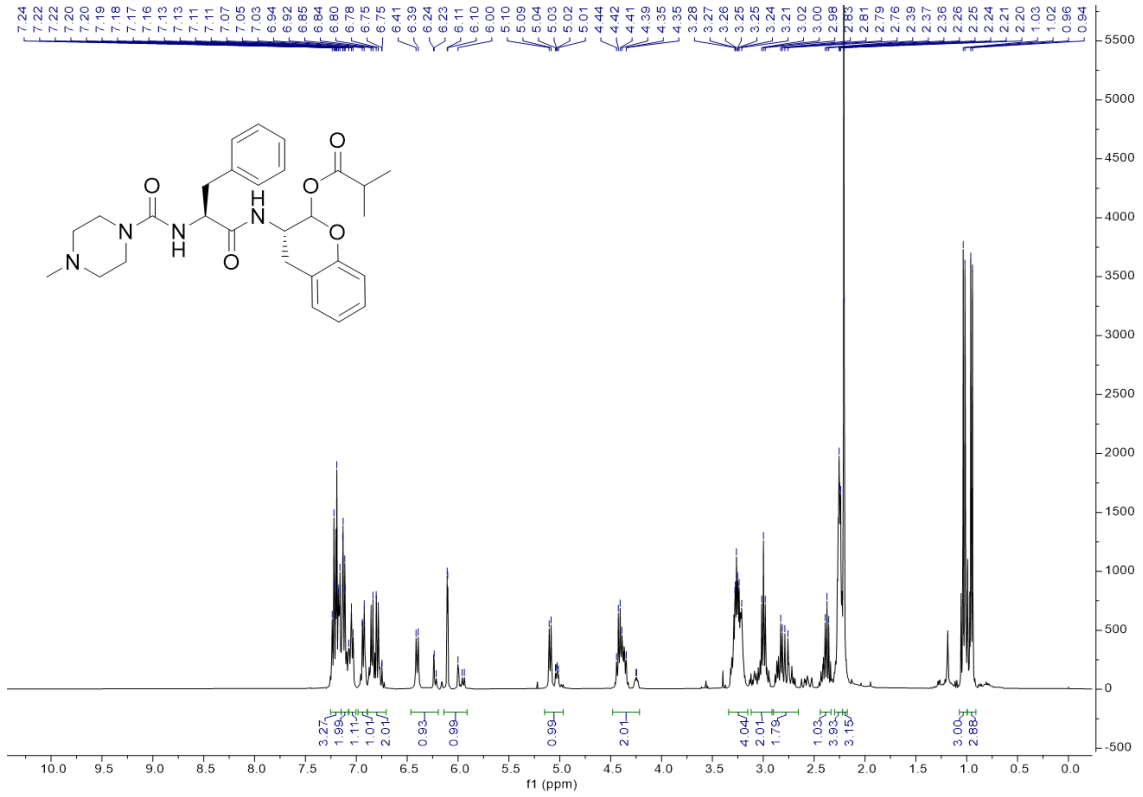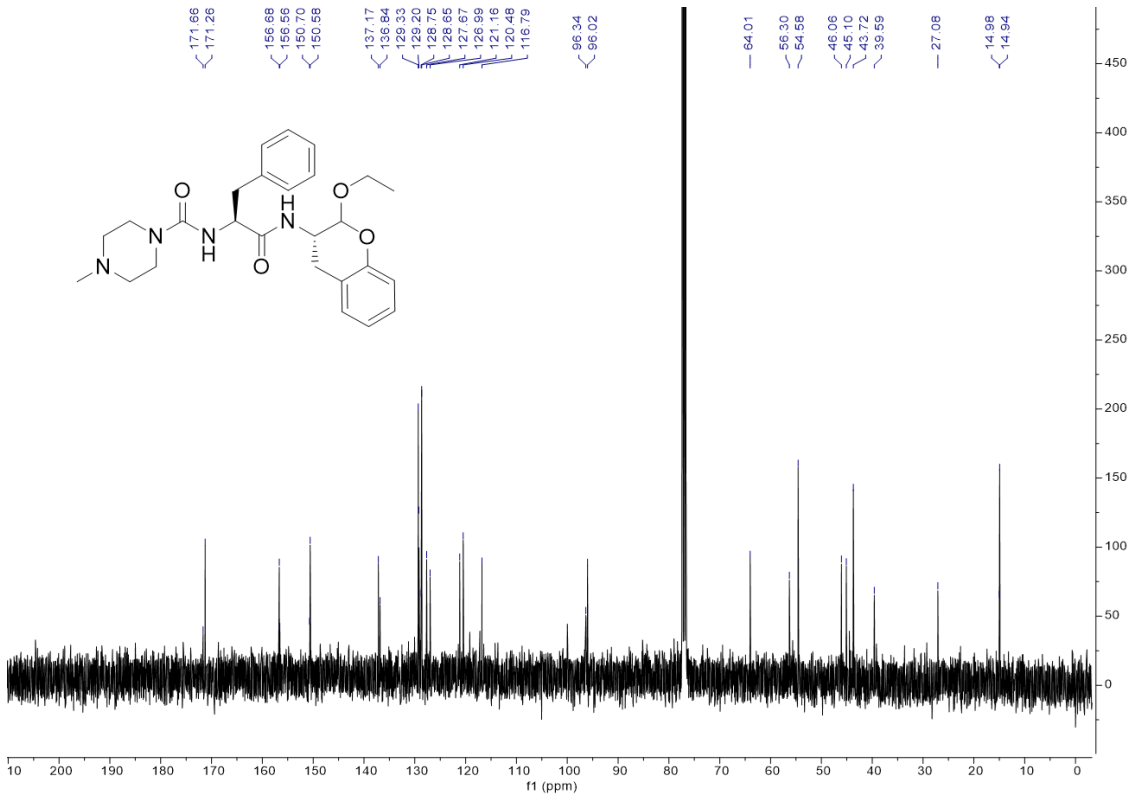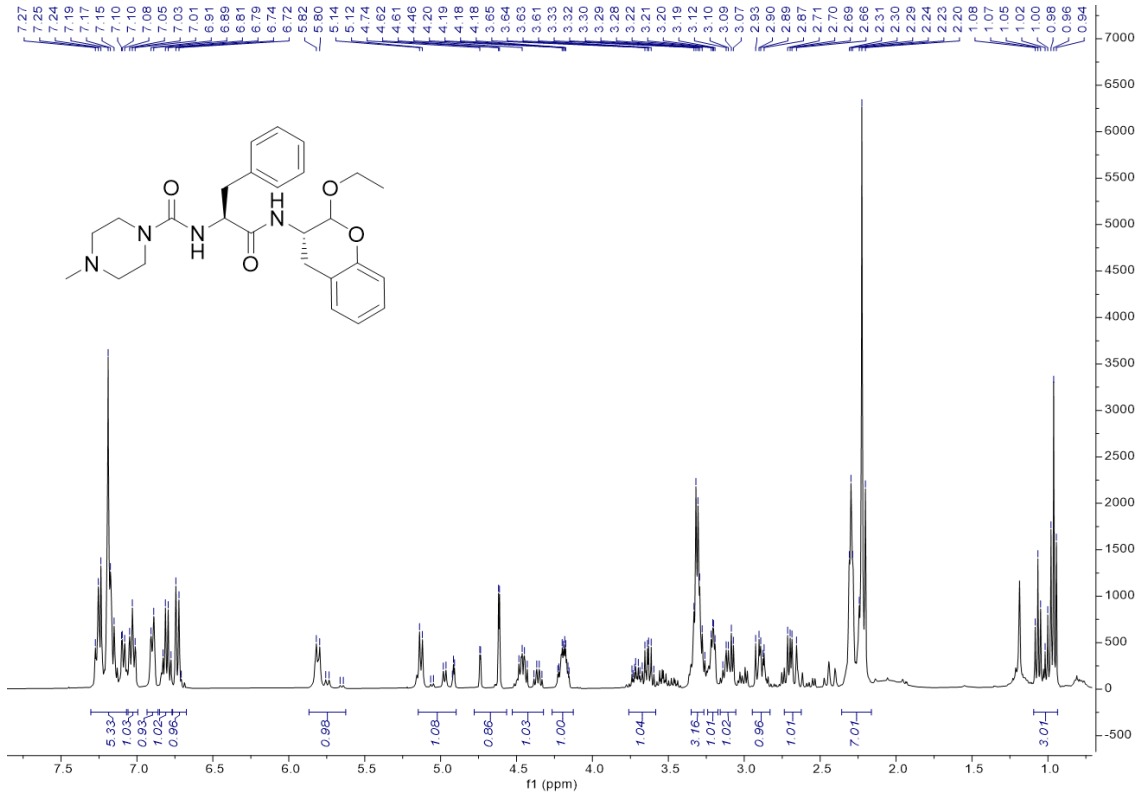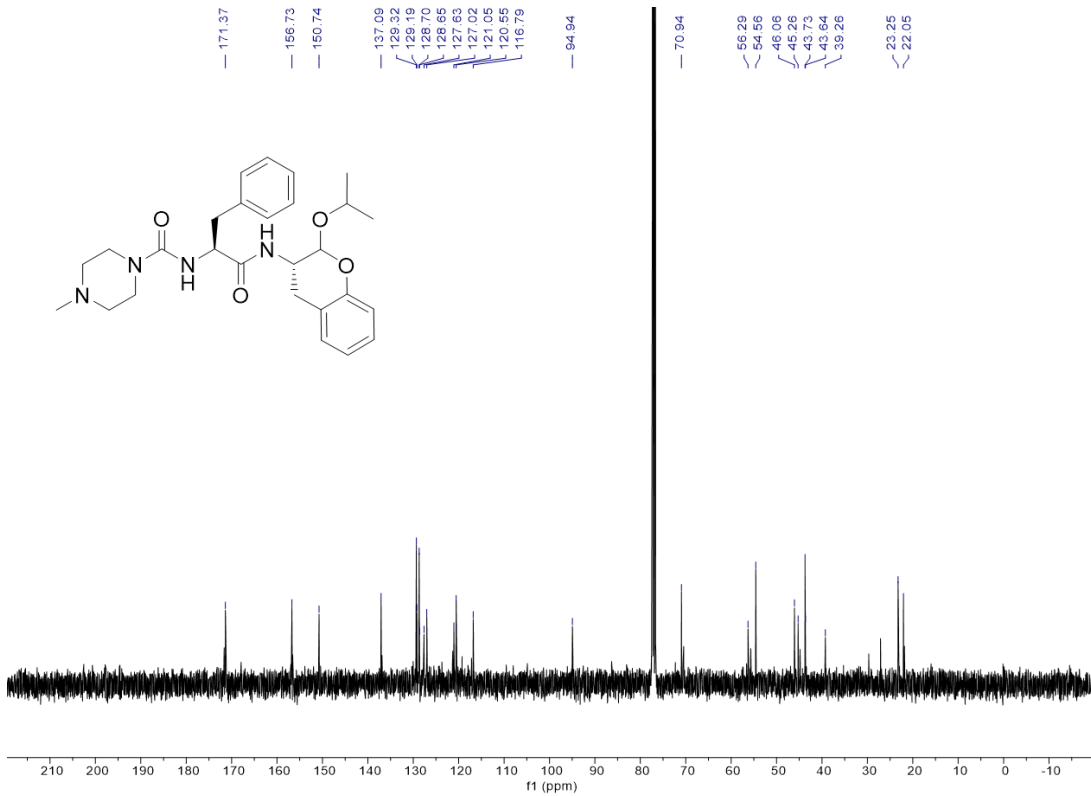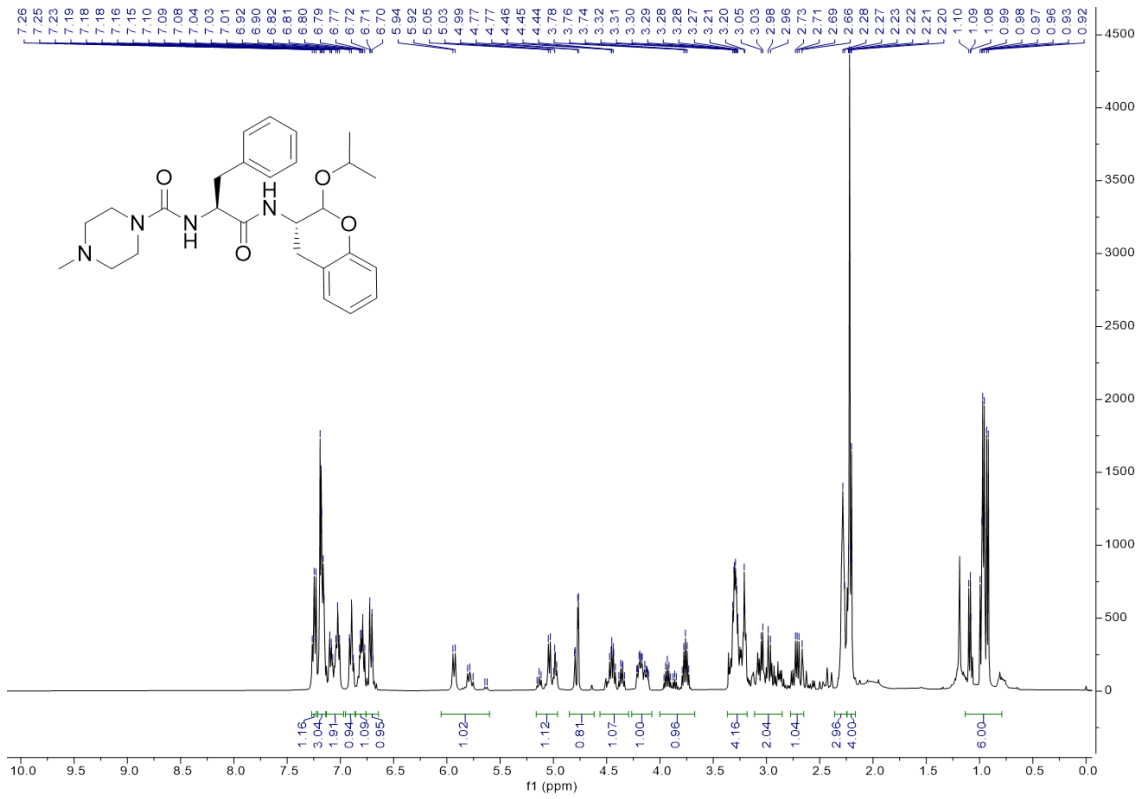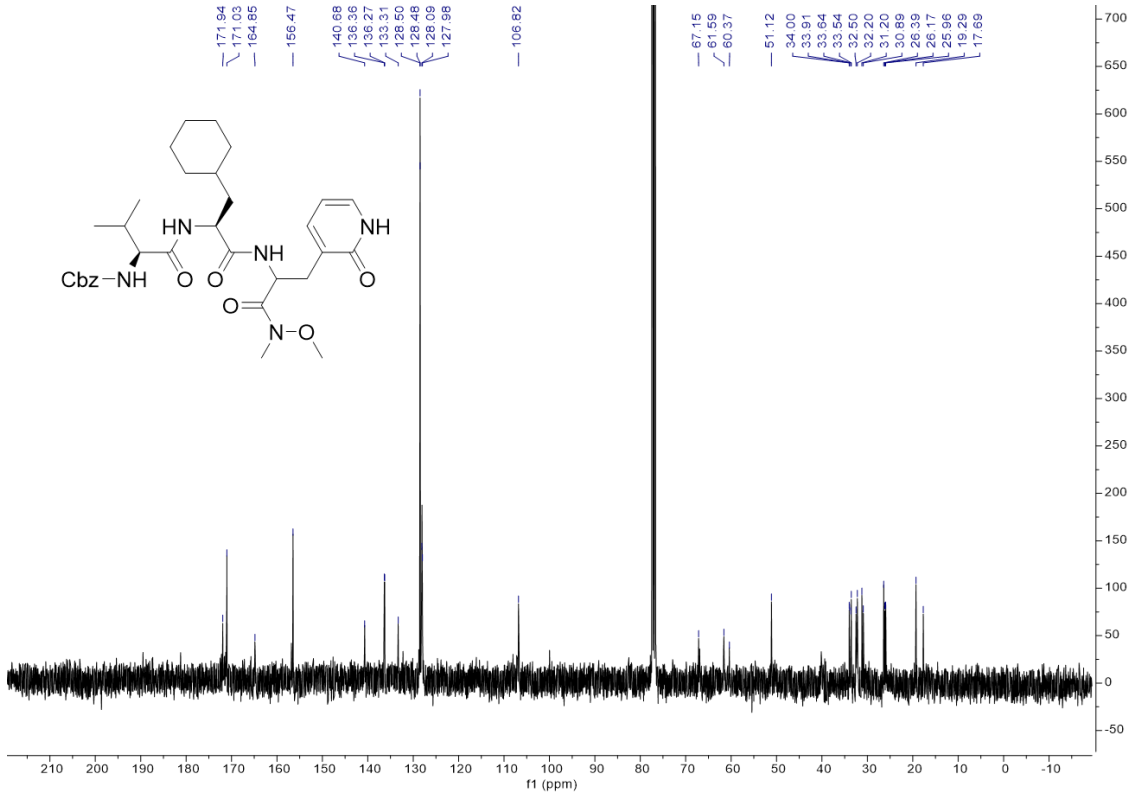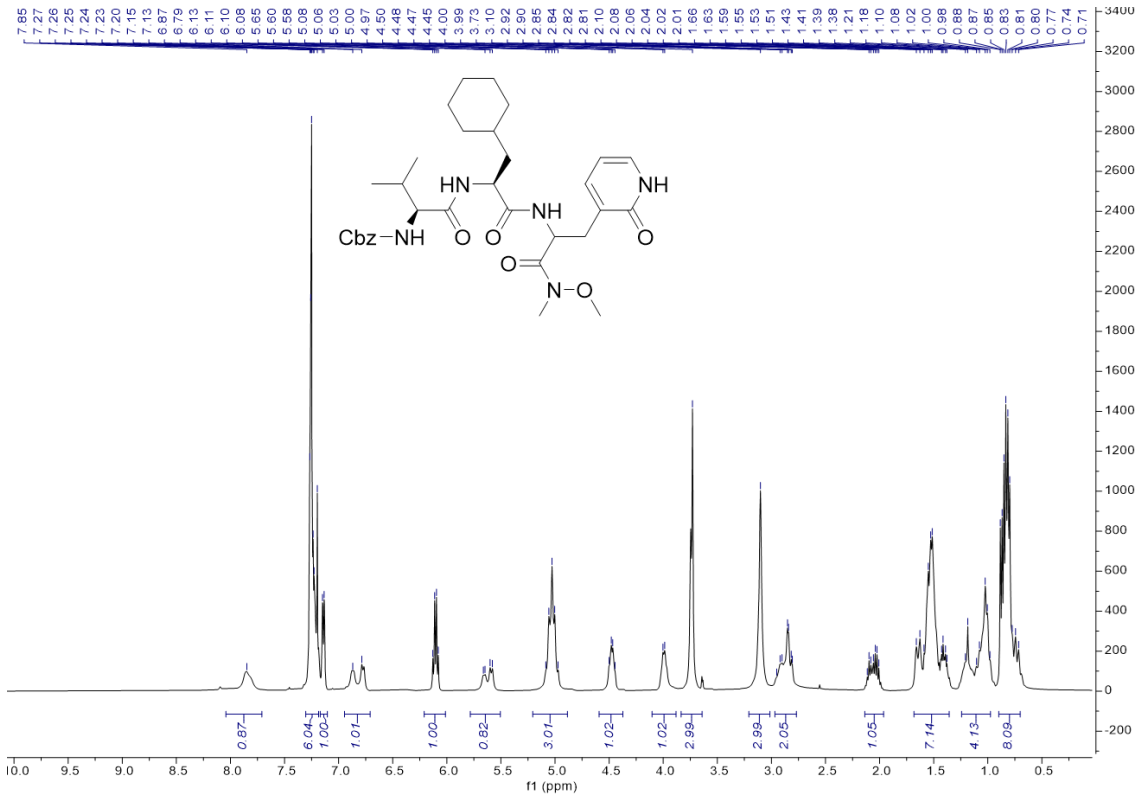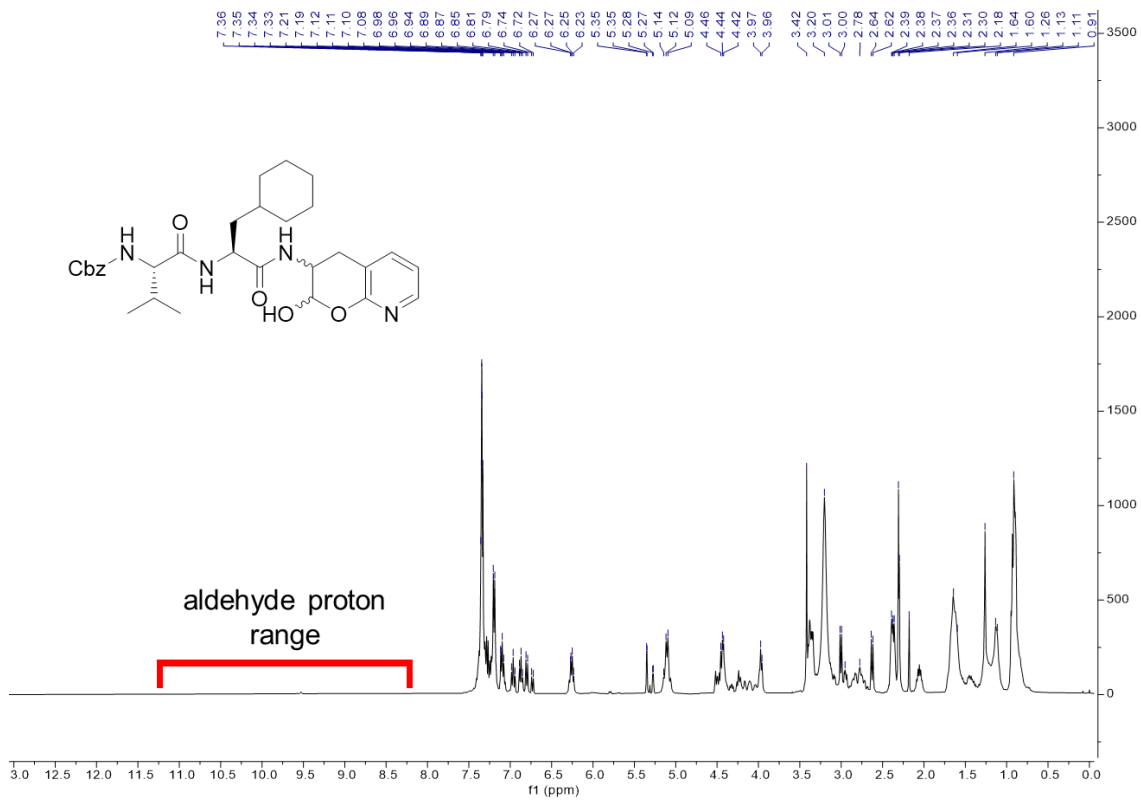
^{1}H-^{13}C HSQC NMR of ^{13}C-labeled **12** (full view):