# Mapping Inhibitor Binding Modes on an Active Cysteine Protease via Nuclear Magnetic Resonance Spectroscopy

Gregory M. Lee,[†] Eaman Balouch,[†] David H. Goetz,[†] Ana Lazic,[†] James H. McKerrow,[‡] and Charles S. Craik*,[†]

[†]Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94158-2280, United States

[‡]Department of Pathology and Center for Discovery and Innovation in Parasitic Diseases, University of California, San Francisco, California 94158-2250, United States

S *Supporting Information*
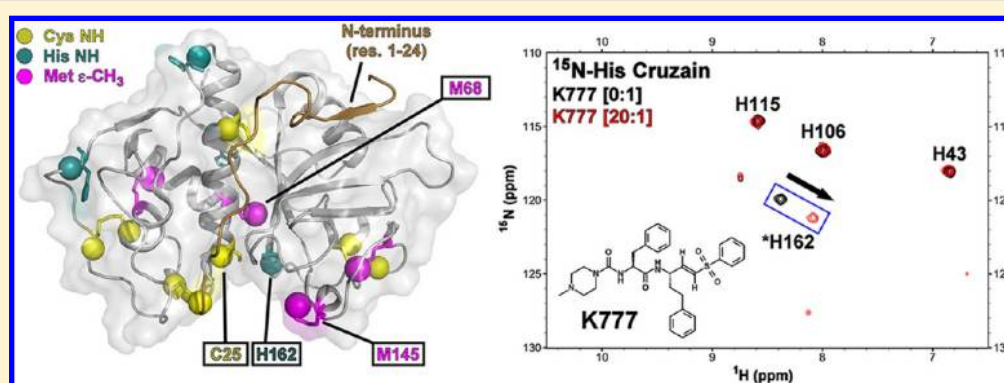
ABSTRACT: Cruzain is a member of the papain/cathepsin L family of cysteine proteases, and the major cysteine protease of the protozoan *Trypanosoma cruzi*, the causative agent of Chagas disease. We report an autoinduction methodology that provides soluble cruzain in high yields (>30 mg/L in minimal medium). These increased yields provide sufficient quantities of active enzyme for use in nuclear magnetic resonance (NMR)-based ligand mapping. Using circular dichroism and NMR spectroscopy, we also examined the solution-state structural dynamics of the enzyme in complex with a covalently bound vinyl sulfone inhibitor (K777). We report the backbone amide and side chain carbon chemical shift assignments of cruzain in complex with K777. These resonance assignments were used to identify and map residues located in the substrate binding pocket, including the catalytic Cys25 and His162. Selective [15N]Cys, [15N]His, and [13C]Met labeling was performed to quickly assess cruzain–ligand interactions for a set of eight low-molecular weight compounds exhibiting micromolar binding or inhibition. Chemical shift perturbation mapping verified that six of the eight compounds bind to cruzain at the active site. Three different binding modes were delineated for the compounds, namely, covalent, noncovalent, and noninteracting. These results provide examples of how NMR spectroscopy can be used to screen compounds for fast evaluation of enzyme–inhibitor interactions to facilitate lead compound identification and subsequent structural studies.

Chagas disease is the leading cause of heart disease in Latin America, with an estimated 8−11 million people infected and an additional 100 million at risk.[1] An estimated 50000 deaths per year can be attributed to cardiomyopathy and gastrointestinal damage caused by chronic Chagas disease.[2,3] This neglected tropical disease is caused by the protozoan *Trypanosoma cruzi* and is transmitted by the introduction of fecal matter of reduviid bugs into the bloodstream of mammalian hosts after they have been bitten by the insect. Although normally confined to Latin America, cases of Chagas disease are appearing in Europe and the southern United States.[1,4] The only currently approved treatments for Chagas disease, nifurtimox and benznidazole, have numerous deleterious side effects and are available in North America in only limited quantities from the Centers for Disease Control and Prevention. Both drugs are generally prescribed during the early

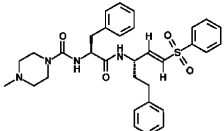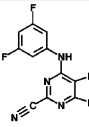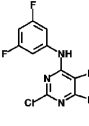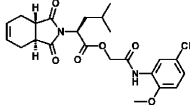acute stages of the disease and are less effective for the treatment of the chronic condition.[5]
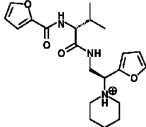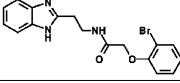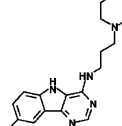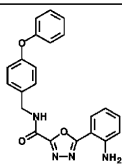
A promising target for potential therapeutics against Chagas disease is the major cysteine protease expressed by *T. cruzi*, cruzain. Cruzain is a member of the papain superfamily of cysteine proteases and is most sequentially and structurally homologous to cathepsin L. In the case of Chagas disease, cruzain is present in all life stages of the trypanosome and is implicated in cellular entry and digestion of immunoglobins.[1] The active site of cruzain contains two catalytic residues (Cys25 and His162), with possibly a third (Asn182) residue contributing to proteolysis (sequential residue numbering,[6]

Table 1. Cruzain Inhibitors Used for NMR Titration Experiments

Compound Structure		^{a}ZINC/SMDC Number	Inhibitor Class	Inhibitory/Binding Constant against Cruzain (µM)	Ref.
	1	K777 (55865535)	covalent, irreversible	$K_d = 1.9$	6
	2	36379291	covalent, reversible	$IC_{50} = 0.01$	24 (cmpd 32)
	3	746694	non-covalent; intermediate of cmpd **2**	n/a	24 (cmpd 32 intermediate)
	4	8691187	non-covalent	$K_i = 2.0$ $IC_{50} = 1.0$	13 (cmpd 4)
	5	3363859	non-covalent	$K_i = 6.0$ $IC_{50} = 7.0$	13 (cmpd 5)
	6	943080	non-covalent	$K_i = 2.0$	13 (cmpd 27)
	7	2236859	non-covalent	$K_i = 2.0$	13 (cmpd 29)
	8	43120874	non-interacting / aggregator	n/a	12

^{a}ZINC database;[46] SMDC is the University of California, San Francisco, Small Molecule Discovery Center.

instead of residue numbering matching that of papain,[7] is used for cruzain throughout this work). A fourth residue (Gln19) serves to stabilize the oxyanion.[8] Similar to many other cysteine protease family members, cruzain is initially expressed as an inactive zymogen, which undergoes a pH-, temperature-, and concentration-dependent self-activation, transforming into the mature, catalytically active protease.[8,9]

One lead compound that targets the mature form of cruzain and is currently in preclinical trials is the vinyl sulfone K777 (Table 1, previously designated K11777).[10] K777 inhibits cruzain by irreversibly binding to the thiol group of the catalytic Cys25. K777 is reported to cure *T. cruzi*-infected mice and has exhibited efficacy against multiple strains of nifurtimox- and benznidazole-resistant *T. cruzi*, validating cruzain as a therapeutic target for Chagas disease.[11] However, the limited potency of K777 has generated interest in identifying new chemical scaffolds that are more effective and are noncovalent competitive inhibitors of the enzyme.

Previous structure−activity relationship (SAR) studies focusing on cruzain−inhibitor complexes have historically relied on X-ray crystallography. However, because of concentration-dependent self-proteolysis, no apo structures of active cruzain are currently available. Although high-throughput screening efforts have identified potential lead compounds, not all cruzain−inhibitor complexes yielded well-behaved crystals.[12,13] In addition, the previous expression protocol[14,15] involved production of cruzain via inclusion bodies and the subsequent unfolding and refolding of the protease, severely limiting overall yields. With these factors in mind, we (i) developed an expression and purification protocol that provided sufficient quantities of cruzain for NMR studies, (ii) used NMR-based methods to quickly verify and map binding modes of newly discovered inhibitors, especially those for which there are no available crystallographic data, and (iii) initiated solution-state biophysical studies to gain further insights into apo and inhibited cruzain structure−function relationships. Importantly, cruzain serves as a model for the

cysteine cathepsin proteases, a large family of important enzymes.

This report represents the first NMR study that examines the solution behavior of an active cathepsin L-like cysteine protease. Using autoinduction methods,[16] we were able to increase the overall yields of activated, mature cruzain by >20-fold versus that of the original inclusion body preparation.[15] This new optimized protocol also permitted the expression of isotopically labeled samples for NMR studies, allowing us to obtain backbone and side chain resonance assignments of a covalently inhibited form of cruzain. These assignments, along with selective [^{15}N]Cys, [^{15}N]His, and [^{13}C]Met cruzain labeling, were critical in mapping and characterizing inhibitor binding modes of eight small-molecule inhibitors, including K777. The overall results described herein provide a methodology for rapidly enhancing our understanding of structure−function relationships within the cysteine cathepsin family.

■ MATERIALS AND METHODS

Design of the His$_6$-Tagged Zymogen. The gene sequence encoding the C-terminally truncated procruzain (Δc; GenBank entry M84342.1)[14,15] was used as the template for the construct described in this study. The forward primer was designed to encode an NdeI restriction site, followed by a His$_6$ sequence, a BamHI restriction site, and an N-terminal procruzain sequence starting at Cys−104. The reverse primer encodes the cruzain Δc terminal residue Gly215, followed by stop codons and a HindIII restriction site. Primer sequences (Integrated DNA Technologies, Inc.) were as follows: 5′-GACGCCATATGCATCACCATCACCATCACGGATCCT-GCCTCGTCCCCGCGGCGACG-3′ (forward) and 5′-CGT-GCAAGCTTCTATTAACCGACCACCGCGGAGCT-3′ (reverse). The resulting oligonucleotide encoding the His$_6$-tagged procruzain sequence was ligated into a pET21a vector (Novagen/EMD) and then transformed into ArcticExpress (DE3)RIL Competent Cells (Agilent). The genes were verified through DNA sequencing. The resulting protein sequence is displayed in Figure 1a.

Protein Expression. Starter cultures used to express His$_6$-tagged procruzain consisted of 5 mL of Luria broth containing carbenicillin (100 μg/mL) and gentamycin (20 μg/mL). After being shaken for >15 h at 37 °C, the overnight LB starter culture was added directly to 2 × 1 L of autoinduction minimal medium contained in baffled 4 L shaker flasks. Modified versions of autoinduction medium recipes N-5052 and PA-5052[16] were used for uniform ^{15}N labeling and selective ^{15}N or ^{13}C labeling, respectively. In the case of uniform ^{15}N labeling, 2.7 g of ^{15}NH$_4$Cl (Cambridge Isotope Laboratories or Sigma-Aldrich) per liter of medium was added. In the case of selective labeling, three separate protease samples were produced by adding one of the following to 1 L of autoinduction minimal medium: (i) 100 mg of [^{15}N]Cys (Cambridge Isotope Laboratories), (ii) 100 mg of [^{15}N$_3$]His (Ajinomoto), or (iii) 250 mg of [^{13}C-methyl]Met (Cambridge Isotope Laboratories). For uniform ^{13}C, ^{15}N, and ^{2}H labeling, a modified version of autoinduction medium recipe C-750501 was used.[16] Here, 1 L of uniformly ^{13}C-, ^{15}N-, and ^{2}H-labeled rich medium (*Escherichia coli* OD2 CDN, Silantes GmbH) was diluted with 1 L of D$_2$O (Cambridge Isotope Laboratories) and split into 2 × 1 L aliquots. All autoinduction minimal medium solutions contained carbenicillin (100 μg/mL) and gentamycin (20 μg/mL) as antibiotics. Further specific details regarding the modifications to the standard autoinduction recipes are

reported in the Supporting Information. After the sample had been shaken at 37 °C for approximately 5 h (until the OD$_{600}$ was >0.5), the temperature was reduced to 20 °C, and the culture was allowed to shake for an additional 72 h. The cultures were then centrifuged, and the cell pellets were stored at −20 °C until they were ready for lysis. Preconditioning of LB starter cultures into small volumes (50 mL) of autoinduction minimal medium overnight was tried and demonstrated to be ineffective and is therefore not recommended.

Protein Purification and Zymogen Activation. Frozen cell pellets were thawed, resuspended in 50 mL of lysis buffer per liter of expression medium, and lysed via sonication: 20 s pulse followed by a 120 s recovery time for a 2 min total pulsing time. The lysis buffer consisted of 50 mM Tris (pH 10), 300 mM NaCl, 10 mM imidazole, 1 mM CaCl$_2$, 1 mM MgSO$_4$, 1 μM DNase I (Roche), 1 mM phenylmethanesulfonyl fluoride (PMSF), and 2 mM methyl methanethiosulfonate (MMTS). PMSF and MMTS are used as covalent cysteine protease inhibitors and can be reversed by the addition of DTT. The resulting lysate was centrifuged, and the supernatant was passed through a 0.22 μm filter prior to FPLC purification. The initial purification was performed at 4 °C using two 5 mL HisTrapFF crude nickel columns (GE Healthcare LifeSciences) attached in series. The following buffers were used: wash/binding, 50 mM Tris (pH 10), 300 mM NaCl, and 10 mM imidazole; elution, 50 mM Tris (pH 10), 300 mM NaCl, and 500 mM imidazole. Elution occurs with 30% elution buffer. MMTS (final concentration of 5 mM) was added to fractions containing His$_6$-tagged procruzain upon elution from the column.

Major fractions were combined and dialyzed at 4 °C in activation buffer [50 mM acetate (pH 5.0), 100 mM NaCl, and 0.1 mM EDTA] containing no reducing agents. After overnight dialysis, the resulting cloudy solution was transferred to 50 mL conical centrifuge tubes. Autoproteolysis of the pro-region (~14 kDa) from the zymogen (~37 kDa) was initiated by addition of DTT (final concentration of 1 mM) to the solution and incubation in a 37 °C water bath. Transformation of the inactive zymogen into the catalytically active domain (cruzain, ~23 kDa) was monitored by removing 50 μL aliquots of the solution at selected time points. These samples were analyzed either through sodium dodecyl sulfate−polyacrylamide gel electrophoresis under denaturing conditions (Figure 1b) or via a fluorescence-based activity assay using the peptide substrate Z-Phe-Arg-AMC (Biotium, Inc.).[14] The total reaction time for processing procruzain into active cruzain was generally 20−40 min, as indicated by the solution becoming clear. Longer activation times of up to 2 h were required to completely proteolyze the smaller pro-domain fragments. To prevent further degradation of the mature cruzain, the reaction was quenched by placement of the sample tubes on ice and addition of MMTS to a final concentration of 1 mM.

The solution was next buffer exchanged to 10× PBS (pH 5.0), and a second round of FPLC purification was performed at 4 °C using a preparative-scale Superdex75 (26/60) size exclusion column (GE Healthcare LifeSciences). The FPLC buffers used in this purification round did not contain reducing agents. Unless otherwise specified below, MMTS (final concentration of 1 mM) was added to each of the major fractions immediately upon elution from the column. The fractions were then collected, flash-frozen, and stored at −80 °C until they were ready for use.

Estimated final enzyme concentrations were determined using a NanoDrop 2000c UV spectrophotometer (Thermo

Scientific) with the following extinction coefficients: ε_{280} = 68910 M^{-1} cm^{-1} (oxidized Cys) and 68410 M^{-1} cm^{-1} (reduced Cys) for His$_6$-GS-procruzain; ε_{280} = 60430 M^{-1} cm^{-1} (oxidized Cys) and 59930 M^{-1} cm^{-1} (reduced Cys) for cruzain. Final estimated yields of the inactive zymogen procruzain and the purified active cruzain are reported in Results.

NMR Spectroscopy. All samples were buffer exchanged from 10× PBS (pH 5.0) to NMR buffer prior to data acquisition. The standard NMR buffer consisted of 20 mM potassium phosphate (pH 5.0), no salt, 1 mM DTT, and 10% (v/v) ^{2}H$_2$O. The apo or inhibited states of the active site Cys25 and protease concentrations of the NMR samples varied depending on the experiment performed and are indicated below. All data was acquired at 27 °C on Bruker Avance 500 and 800 MHz instruments equipped with triple-resonance cryoprobes. External ^{13}C, ^{15}N, and ^{1}H chemical shift referencing was performed as previously reported.[17] All spectral data were processed with NMRpipe[18] and analyzed with Sparky.[19]

NMR Resonance Assignments. To prevent further self-proteolysis of the uniformly ^{13}C-, ^{15}N-, and ^{2}H-labeled apo cruzain, major fractions eluting during the size exclusion purification step were immediately inhibited with greater than stoichiometric ratios of the vinyl sulfone K777. Standard K777 stocks consisted of 10 mM unlabeled inhibitor dissolved in d_6-DMSO (Cambridge Isotope Laboratories). NMR samples consisted of ~0.5 mM uniformly ^{13}C-, ^{15}N-, and ^{2}H-labeled cruzain−K777 complex in 0.45−0.50 mL of NMR buffer. The suite of three-dimensional triple-resonance experiments used to acquire backbone amide and side chain carbon resonance assignments consisted of HNCACB, HN(CO)CACB, HNCO, and CC(CO)NH.[20] A three-dimensional ^{15}N−^{1}H NOESY-HSQC spectrum acquired on a uniformly ^{15}N- and ^{1}H-labeled cruzain−K777 complex was used to assign the methionine ε-methyl groups. General spectral parameters and acquisition times are listed in the Supporting Information. Chemical shift assignments are listed in Table S1 of the Supporting Information.

NMR-Based Inhibitor Binding Assays. NMR samples used for the ^{15}N−^{1}H HSQC and ^{13}C−^{1}H HSQC experiments consisted of 0.025−0.050 mM uniformly or selectively [^{15}N]His-, [^{15}N]Cys-, or [^{13}C]Met-labeled cruzain in a 0.40−0.45 mL solution. Inhibitors were dissolved in d_6-DMSO (Cambridge Isotope Laboratories) resulting in 10 mM stock solutions. Titrations were performed by adding microliter aliquots of the 10 mM stock solution in a stepwise manner until a 10−20-fold stoichiometric excess with respect to cruzain was achieved. A total of five titration points (apo, 2.5×, 5×, 10×, and 20×) for each inhibitor were acquired. NMR solutions were supplemented with 1 mM DTT (final concentration) prior to acquiring the first (apo) titration point. Final DMSO concentrations did not exceed 5% (v/v). Further spectral parameters and data processing protocols are reported in the Supporting Information. Total chemical shift perturbation values were calculated as previously described.[21]

Circular Dichroism. All circular dichroism (CD) data were acquired on a JASCO J-715 spectropolarimeter at room temperature using a 0.2 cm path length QS cuvette. A single scan ranging from 260 to 200 nm was acquired for each sample. All other spectral parameters and the estimation of fractional helicity (f_H) values were as previously described.[21] Concentrated cruzain−K777 complex or procruzain stocks (500 μM) were kept in 20 mM potassium phosphate buffer containing no salt, at pH 5.0, 7.0, and 10.0. For each titration point, 10 μL of a concentrated enzyme stock was diluted with 490 μL of buffer containing various concentrations of guanidinium hydrochloride (GdnHCl), for a final cruzain concentration of 10 μM. Final GdnHCl concentrations in the assay solutions ranged from 0 to 8 M (cruzain−K777 complex) and from 0 to 5 M (procruzain). Assay solutions were incubated at room temperature for 30 min prior to data acquisition.

Structure Representations. All structural representations were generated using the X-ray crystal structure of the cruzain−K777 complex [Protein Data Bank (PDB) entry 2OZ2][6] with PyMol version 1.5.0 (Schrödinger, LLC).

■ RESULTS

High-Level Expression of Soluble Cruzain. Extensive modification of previously reported cruzain expression protocols[14,15] provided sufficient quantities of purified, active cruzain for use in NMR-based studies. The original protease construct was expressed as an insoluble fusion protein, requiring urea denaturation and multiple refolding and dialysis steps. Reported yields from the original protocol using rich media were 15−20 mg of the fusion protein per liter of bacterial culture, resulting in approximately 1−1.5 mg of active cruzain after the self-activation and purification steps.[14,15] In the current protocol, a His$_6$ tag sequence is placed at the N-terminus of the cruzain Δc construct[15] (Figure 1a) for ease of purification over a nickel Sepharose column. The construct was expressed as a soluble fraction zymogen using autoinduction methods described by Studier.[16] Compared to the original inclusion body preparation using rich medium,[14] autoinduction in both minimal and rich media is a significant improvement. Yields of the zymogen form of cruzain were >200 mg/L of rich autoinduction medium (A. Tochowicz and M. Merski, personal communication) and 150 mg/L of minimal autoinduction medium.

These higher protein yields allowed for faster autoactivation of the inactive zymogen into the catalytically active cruzain (Figure 1b): 20−40 min, compared to approximately 4−7 h in the previous protocol.[14] This result also verifies that self-activation of procruzain is concentration-dependent. Notably, incubation of procruzain with K777 impeded but did not completely abolish in vitro self-activation of the zymogen (Figure 1c and Supporting Information). Fluorescence-based activity assays using the peptide substrate Z-Phe-Arg-AMC (data not shown) verified the cleaved product to be the active form of cruzain. Final yields of purified active cruzain were 80−100 mg/L of autoinduction rich medium and 30−50 mg/L of autoinduction minimal medium. The corresponding ^{15}N−^{1}H HSQC spectra of the K777-inhibited and apo (Figure 1d and Figures S1 and S2 of the Supporting Information) cruzain display well-dispersed resonance peaks, verifying the presence of a folded protein.

NMR Assignments of the Cruzain−K777 Complex. Examination of the cruzain ^{15}N−^{1}H HSQC spectra indicated that the apo state was highly susceptible to continued self-proteolysis under the NMR conditions described here. Spectra of apo cruzain acquired at time intervals over a period of less than 1 day at 27 °C exhibited resonance peaks between 7.0 and 8.5 ppm in the ^{1}H dimension that were not present in the time "zero" spectrum (data not shown). The presence of these peaks, which grew in intensity over a period of time, is a hallmark of protein degradation and/or unfolding. Storing apo cruzain samples at 4 °C did not alleviate the problem, as the
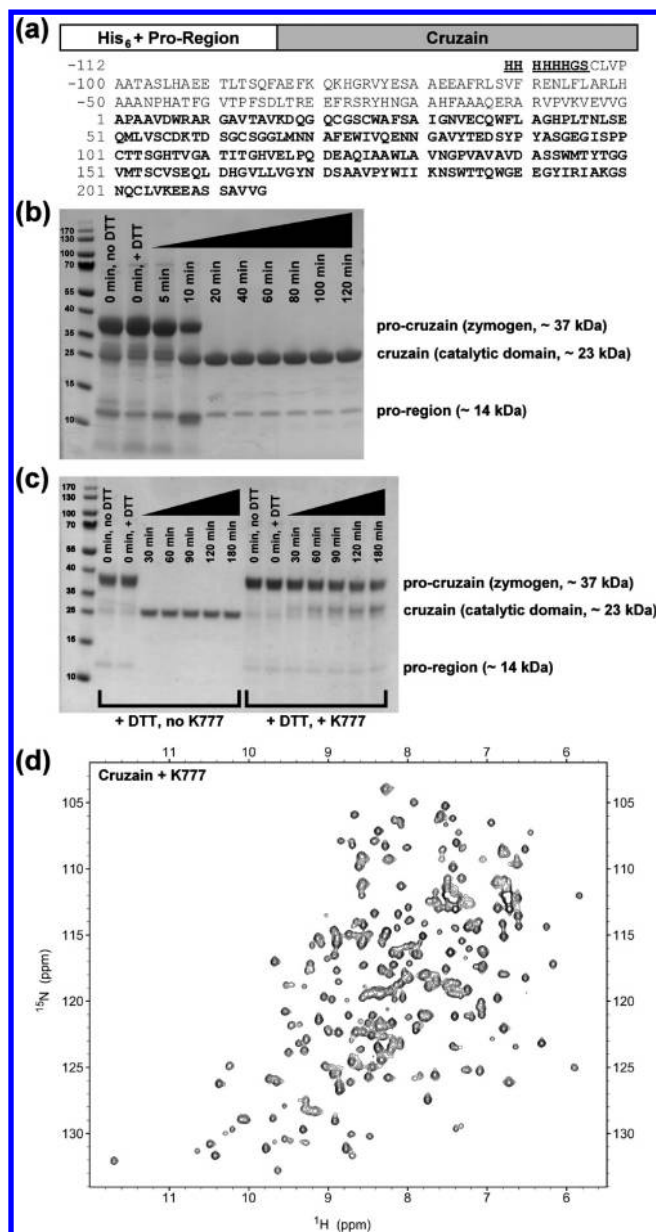
(a)

His₆ + Pro-Region	Cruzain

```
-112                                              HH  HHHHGSCLVP
-100 AATASLHAEE  TLTSQFAEFK  QKHGRVYESA  AEEAFRLSVF  RENLFLARLH
 -50 AAANPHATFG  VTPFSDLTRE  EFRSRYHNGA  AHFAAAQERA  RVPVKVEVVG
   1 APAAVDWRAR  GAVTAVKDQG  QCGSCWAFSA  IGNVECQWFL  AGHPLTNLSE
  51 QMLVSCDKTD  SGCSGGLMNN  AFEWIVQENN  GAVYTEDSYP  YASGEGISPP
 101 CTTSGHTVGA  TITGHVELPQ  DEAQIAAWLA  VNGPVAVAVD  ASSWMTYTGG
 151 VMTSCVSEQL  DHGVLLVGYN  DSAAVPYWII  KNSWTTQWGE  EGYIRIAKGS
 201 NQCLVKEEAS  SAVVG
```

(b) (SDS-PAGE gel image)

Lanes: 0 min, no DTT; 0 min, + DTT; 5 min; 10 min; 20 min; 40 min; 60 min; 80 min; 100 min; 120 min

- pro-cruzain (zymogen, ~ 37 kDa)
- cruzain (catalytic domain, ~ 23 kDa)
- pro-region (~ 14 kDa)

(c) (SDS-PAGE gel image)

Lanes: 0 min, no DTT; 0 min, + DTT; 30 min; 60 min; 90 min; 120 min; 180 min (twice)

- pro-cruzain (zymogen, ~ 37 kDa)
- cruzain (catalytic domain, ~ 23 kDa)
- pro-region (~ 14 kDa)

+ DTT, no K777   + DTT, + K777

(d) Cruzain + K777 — $^{15}N-^{1}H$ HSQC spectrum; axes ^{1}H (ppm) vs ^{15}N (ppm)

Figure 1. Domain diagram of cruzain, activation and inhibition of its zymogen form, and $^{15}N-^{1}H$ HSQC spectrum. (a) Domain diagram and one-letter sequence of the zymogen (procruzain). The active catalytic domain (cruzain) used in this study is shown in bold. Residues of an engineered N-terminal His₆ tag (underlined) and the truncated pro-region are numbered from −112 to −1. Residue numbers of cruzain are numbered from 1 to 215. (b) Incubation at 37 °C and addition of DTT to PMSF- and MMTS-inhibited zymogen (~37 kDa) initiate autoproteolysis of the N-terminal pro-region (~14 kDa) from the catalytically active domain (~23 kDa), as monitored by sodium dodecyl sulfate−polyacrylamide gel electrophoresis. (c) Addition of the vinyl sulfone K777 to procruzain impedes but does not completely inhibit self-activation of the zymogen. (d) The $^{15}N-^{1}H$ HSQC spectrum of cruzain inhibited with K777 indicates a well-folded and stable enzyme structure.

degradation peaks were also present in their respective HSQC spectra.

To obtain backbone and side chain resonance assignments, we acquired standard triple-resonance data sets[20] using uniformly ^{13}C-, ^{15}N-, and ^{2}H-labeled cruzain inhibited with unlabeled K777. However, the $^{15}N-^{1}H$ HSQC spectra of the deuterated cruzain−K777 complex exhibited several missing peaks in comparison with that of the protonated analogue (Figure S1 of the Supporting Information), indicating incomplete ^{2}H to ^{1}H back-exchange of the amide protons during protein purification. Attempts to partially denature the cruzain−K777 complex at pH 5.0 with 2 M GdnHCl followed by rapid dilution and refolding yielded no appreciable amide proton back-exchange, as observed in the HSQC spectra (data not shown). Fortunately, the resulting triple-resonance spectra yielded backbone amide and side chain resonance assignments for 165 of the 206 (approximately 80%) non-proline residues in the cruzain−K777 complex (Table 1 of the Supporting Information). The majority of the absent backbone amide assignments correspond to residues in the central α-helix from Phe28 to Leu40, as well as portions of the adjacent β-sheets. Importantly, with the exception of Asn182 and Trp184, the backbone amide resonances of residues located within the active site and substrate binding pocket of the cruzain−K777 complex are assigned and provide sufficient coverage for mapping the binding modes of protease inhibitors.

The Cruzain−K777 Complex Is Exceptionally Stable. CD studies were performed to examine the structural stabilities of the cruzain−K777 complex and apo procruzain (Figure 2). Overall, the CD signatures acquired in the absence of denaturant at pH 5 exhibit significantly greater helical content than at pH 7 or 10 (Figure S3 of the Supporting Information). The CD spectrum of procruzain at pH 5 was not acquired because of significant protein aggregation. Under acidic (pH 5) and neutral conditions (pH 7), the CD signatures of inhibited cruzain remain relatively stable at increasing denaturant concentrations, with the θ_{222} band only losing half of its signal at 8 M GdnHCl. These results explain the lack of ^{2}H to ^{1}H back-exchange observed for the uniformly ^{13}C-, ^{15}N-, and ^{2}H-labeled cruzain−K777 NMR sample described above, as low denaturant concentrations are normally used to partially unfold deuterated proteins.[22] Conversely, at pH 10, the cruzain−K777 complex becomes structurally labile at >5 M GdnHCl. The apo form of procruzain at pH 10 displayed less structural stability at low GdnHCl concentrations than the cruzain−K777 complex under similar conditions. In this case, the θ_{222} bands of the apo procruzain CD spectra were more negative than that of the cruzain−K777 complex at 0 M GdnHCl, indicating greater helical content. However, by 2 M GdnHCl, the θ_{222} band of the procruzain CD curve has lost more than half of its intensity.

The structural stability of the cruzain−K777 complex at pH 5 is also reflected in the heteronuclear NOE data, which measure backbone mobility in terms of NOE ratios (Figure S4 of the Supporting Information). Including the N- and C-terminal residues, the majority of the backbone amides of the inhibited protease exhibit NOE ratios of >0.6, indicating a lack of overall backbone mobility. The regions that display lower NOE ratios, associated with higher levels of conformational mobility, are in loop regions, for example, residues 85−107 located between two β-strands. Taken together, the CD and heteronuclear NOE studies indicate that the cruzain−K777 complex has an extraordinarily stable structure, even in the presence of high denaturant concentrations, and is more structured under acidic conditions.

Choosing [^{15}N]Cys, [^{15}N]His, and [^{13}C]Met Labels as Cruzain Inhibitor Binding Probes. Selective labeling of cruzain with [^{15}N]Cys, [^{15}N]His, and [^{13}C]Met was performed to more quickly assess potential inhibitor−protease interactions via NMR spectroscopy. This labeling scheme was chosen
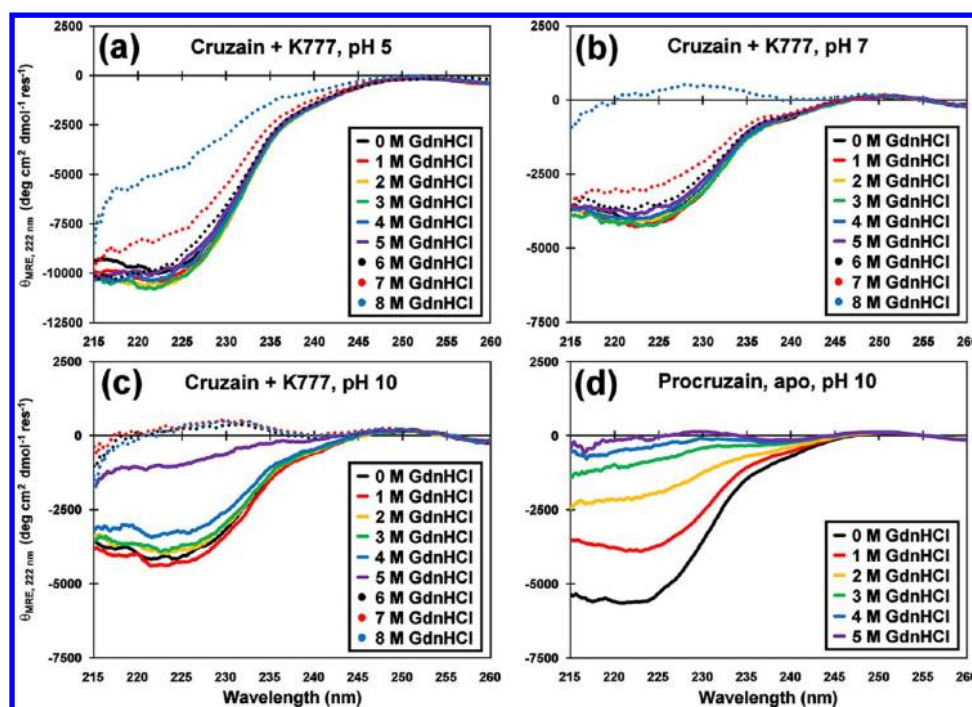
Figure 2. Circular dichroism denaturation study of the cruzain−K777 complex and procruzain. Minimal perturbations of the CD spectra of the cruzain−K777 complex with guanidinium hydrochloride (final concentrations from 0 to 8 M, colored as indicated) measured at (a) pH 5, (b) 7, and (c) 10 reflect the structural stability of the inhibited protease against chemical denaturation. (d) CD spectra of procruzain in the absence of inhibitor measured at pH 10 indicate the weaker overall stability vs GdnHCl denaturation (final concentrations from 0 to 5 M, colored as indicated). Comparisons of the θ_{222} band and estimated fractional helicity data are displayed in Figure S3 of the Supporting Information.

because the two catalytic residues are Cys25 and His162. In addition, there are relatively few cysteine (eight), histidine (four), and methionine (four) residues within the cruzain sequence, which are all well-dispersed within the cruzain structure (Figure 3) and in their respective HSQC spectra (Figure S2 of the Supporting Information). Six of the eight cysteine residues are sequestered in disulfide bonds (Cys22−Cys63, Cys56−Cys101, and Cys155−Cys203). Of the free cysteine residues, one (Cys36) is buried in the central helical core, while the catalytic Cys25 is surface-exposed. In the case of the crystal structure of the cruzain−K777 complex,[6] only the Cys22−Cys63 pair is located within 8 Å of the catalytic Cys25. Three of the four histidine residues (His43, His106, and His115) are located on the distal side of the molecule with respect to the catalytic His162. Although all the methionine residues are positioned more than 8 Å from the active site Cys25, Met68 and Met145 are present and surface-exposed in the substrate binding pocket, while Met152 is located in a proximal loop. Importantly, Met68 forms a portion of the S2 pocket of cruzain. This binding pocket subsite, which interacts with the P2 position (two residues N-terminal to the scissile bond) of the peptide substrate, is known to be a determinant of cysteine cathepsin enzyme specificity.[23] The loop containing Met145 and Met152 is separated from the catalytic His162 by a cluster of tryptophan residues (Trp144, Trp184, and Trp188). Therefore, any perturbations in the conformation of His162 upon inhibitor binding may also propagate through the tryptophan cluster toward the two methionine residues.

Classification of Cruzain−Inhibitor Interactions Using Selectively Labeled Probes. The selectively labeled cruzain samples were titrated with several inhibitors with known binding modes as a proof of principle. All the listed compounds (Table 1) had been previously identified through either

chemical modification of known peptide substrates or via computational docking experiments and verified as having some inhibitory effect on protease activity.[6,12,13,24] The cysteine, histidine, and methionine residues were segregated into three categories according to their respective distance from the active site: (1) "active site", (2) "proximal", and (3) "remote". The "active site" residues (Cys25, His162, Met68, and Met145) are surface-exposed and are located in the substrate binding pocket within 10 Å of the catalytic Cys25 thiol group. The "proximal" residues (Cys22, Cys63, and Met152) border the active site region and are located within 10−15 Å of the Cys25 thiol. The remaining Cys, His, and Met residues, which are generally located more than 15 Å from the Cys25 thiol, are categorized as "remote".

Covalent, Irreversible Inhibitor. HSQC titrations with the vinyl sulfone K777, a known covalent, irreversible inhibitor, yielded dramatic chemical shift perturbations of peaks corresponding to the catalytic residues (Cys25 and His162) and residues located within, or proximal to, the active site and substrate binding pocket (Cys22, Cys63, Met68, Met145, and Met152) (Figure 4a−c). Residues located remotely with respect to this region displayed negligible shift perturbations. The apparent maximal shift perturbations were reached at low stoichiometric inhibitor ratios (<2.5-fold molar excess with respect to cruzain) and were not significantly affected by increasing K777 concentrations (Figure 4d−f).

Covalent, Reversible Inhibitor. Spectral data of nitrile compound **2**,[24] a covalent, reversible inhibitor, displayed chemical shift perturbation profiles similar to those observed for the K777 titration (Figures S5−S7 of the Supporting Information). However, formation of a precipitate within the NMR tube occurred during the titration of compound **2**. Because none of the HSQC resonance peaks were concurrently
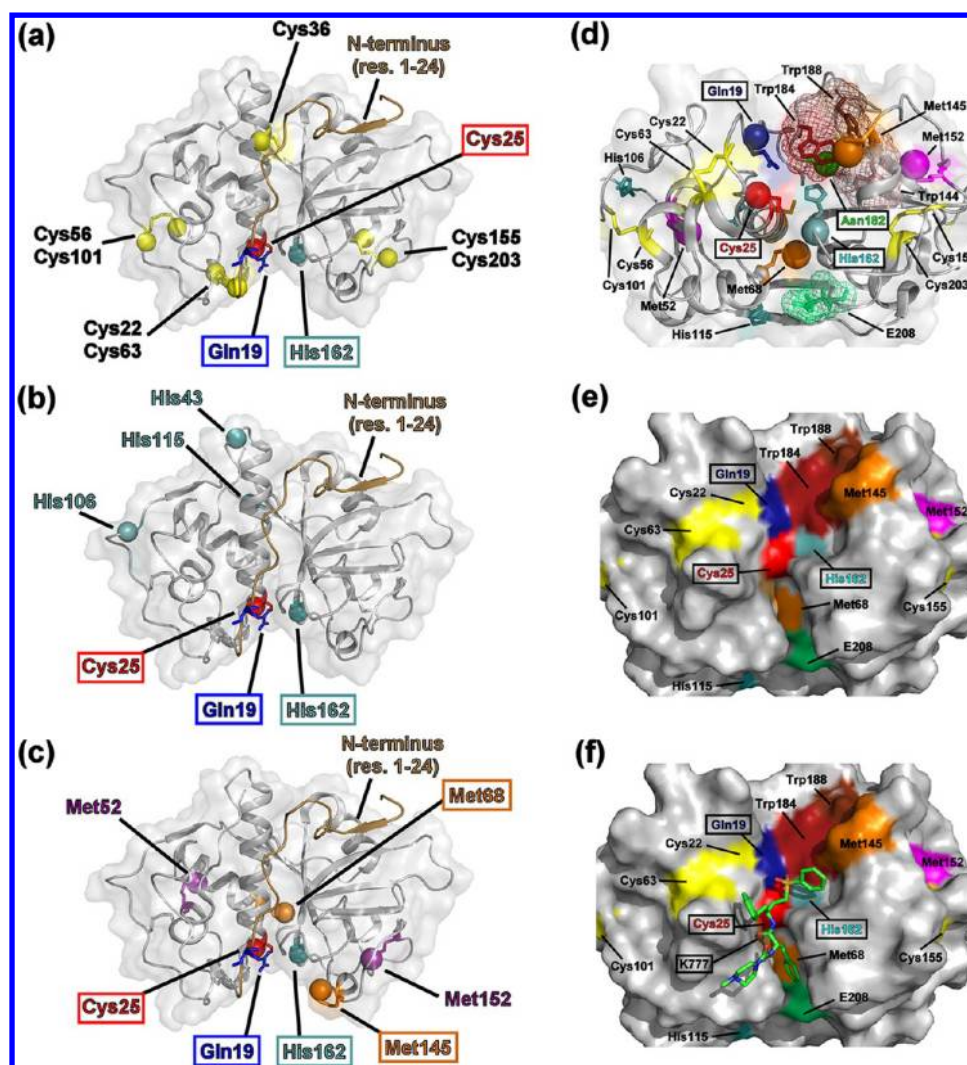
Figure 3. Choosing cysteine, histidine, and methionine residues as cruzain inhibitor binding probes. (a−c) Distributions of (a) cysteine (yellow spheres), (b) histidine (cyan spheres), and (c) methionine (orange and magenta spheres) residues mapped on the crystal structure of the cruzain−K777 complex (PDB entry 2OZ2).[6] The catalytic Gln19 (dark blue), Cys25 (red), and His162 (cyan) residues are denoted with boxes. The putative catalytic residue Asn182 has been omitted for the sake of clarity. Met68 and Met145 (orange), both located on the surface of the substrate binding pocket, and the relatively unstructured N-terminal region (residues 1−24, brown cartoon) are also indicated. (d) Cruzain active site, represented as a 90° rotation about the horizontal axis with respect to panels a−c. The positions of the catalytic Gln19 (blue), Cys25 (red), His162 (cyan), and Asn182 (green) amide groups are indicated with spheres. Orange spheres denote positions of the Met68 and Met145 ε-methyl groups; purple spheres denote positions of the Met52 and Met152 ε-methyl groups. A cluster of tryptophan residues (Trp184, dark red; Trp188, brown; and Trp144, pink) connects Met145 and Met152 with His162. Glu208 (light green), which forms a portion of the S2 pocket, is also displayed. The van der Waals surfaces of the tryptophan cluster and Glu208 are denoted with mesh. (e and f) Surface representation of the cruzain active site with (e) and without (f) the K777 structure. Residue colors are the same as in panel d.

broadened as the inhibitor concentration increased, compound **2**, not cruzain, is likely the precipitating and/or aggregating species in solution. Major differences in the overall perturbation patterns between the two data sets, especially noticeable for the methionines, may be attributed to the different chemical moieties present in K777 and compound **2**, as well as their relative positioning on the enzyme. For example, in the crystal structure of the cruzain−compound **2** complex (PDB entry 3I06),[24] the inhibitor is in contact with Met68 but is positioned more than 10 Å from Met145. Conversely, in the cruzain−K777 complex, both methionine methyl groups are in contact with the inhibitor. Moreover, the terminal sulfyl phenyl ring of K777 is in contact with Trp184, allowing conformational perturbations of the tryptophan indole ring to affect the observed Met145 and Met152 chemical shifts.

Noncovalent Inhibitors. In comparison to that of the covalently bound K777 and compound **2**, titration with a large number of molar equivalents of the known noncovalent cruzain inhibitors **4**−**7**[13] displays significantly reduced, sometimes negligible, [^{15}N]Cys chemical shift perturbations (Figure S5 of the Supporting Information). In the case of the [^{15}N]His and [^{13}C]Met spectra, only the catalytic His162, as well as Met68 and Met145, displayed any strong chemical shift perturbations (Figures S6 and S7 of the Supporting Information). In general, the shift perturbations of the [^{15}N]His and [^{13}C]Met resonances do not reach a maximal value at >20-fold molar excess inhibitor concentrations. Additionally, in the case of the titration with compound **4**, the His162 amide resonance exhibited extensive peak broadening at low inhibitor concentrations while the intensities of the other three histidine peaks
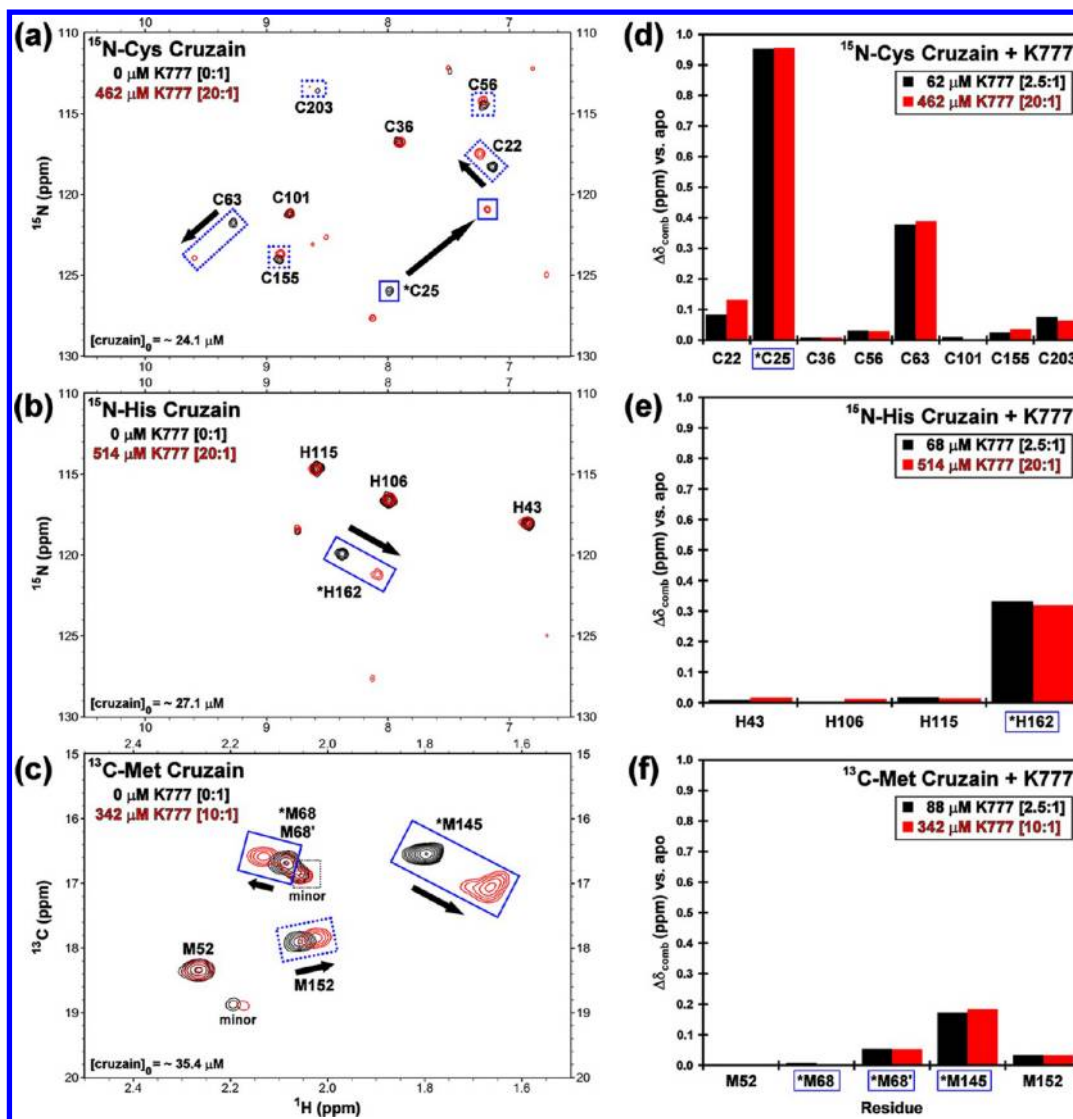
Figure 4. Chemical shift perturbation data of cruzain reporter groups with K777. The HSQC titration spectra of selectively (a) [15N]Cys-, (b) [15N]His-, and (c) [13C]Met-labeled cruzain with 0 (black) and 10 or 20 (red) molar equiv of K777. Solid boxes denote residues located in the substrate binding pocket. Dotted lines denote other perturbed residues. The chemical shift perturbations relative to the apo state at 2.5 (black) and 10 or 20 (red) molar equiv of K777 observed for the selectively (d) [15N]Cys-, (e) [15N]His-, and (f) [13C]Met-labeled cruzain resonances. Residues located in (asterisks and boxed) or proximal to the substrate binding pocket of cruzain exhibit the largest shift perturbations.

remained relatively unchanged. This broadening of a specific amino acid resonance peak is indicative of an "intermediate" NMR time scale conformational exchange, rather than protein aggregation. Similar peak broadening was observed for the catalytic Cys25 at low concentrations of compound 5. Multiple peaks, likely corresponding to degradation products, were also observed in both the [15N]Cys and [15N]His HSQC spectra of the compound 5 titration. These results confirm that the noncovalent inhibitors have overall weaker binding affinity for cruzain than their covalent counterparts, as reflected in their micromolar K_i and IC_{50} values (Table 1).

Noninteracting Compounds. Oxadiazole 8[12] represents a scaffold subsequently considered to be a noninteracting compound and/or aggregator (J. Ellman, personal communication). Titration of the cruzain samples with compound 8 resulted in aggregates forming in the NMR tubes. Overall chemical shift perturbations for the three reporter groups were minor in comparison to those observed for the other compounds (Figures S5−S7 of the Supporting Information).

Interestingly, peaks corresponding to the catalytic Cys25 and His162 appear to exchange broaden more dramatically than the other cysteine and histidine residues. As with compound 2, the remaining resonance peaks in the HSQC spectra did not significantly broaden over the course of the titration. In addition, the Met145 ε-methyl group is the only resonance among the other methionine residues that is perturbed upon addition of compound 8. Together, these results suggest that, although the compound is aggregating in solution, it may have a limited binding affinity for cruzain.

The only scaffold listed in Table 1 with a previously unknown binding mode is compound 3, a chlorinated intermediate of compound 2. In this case, compound 3 was added to the cruzain NMR samples to test the hypothesis that the nitrile group of compound 2 was necessary for covalent binding to Cys25. Like that of compounds 2 and 8, titration of compound 3 resulted in the formation of a precipitate in solution. In this case, however, negligible effects on resonance shift perturbations or peak broadening were observed in the
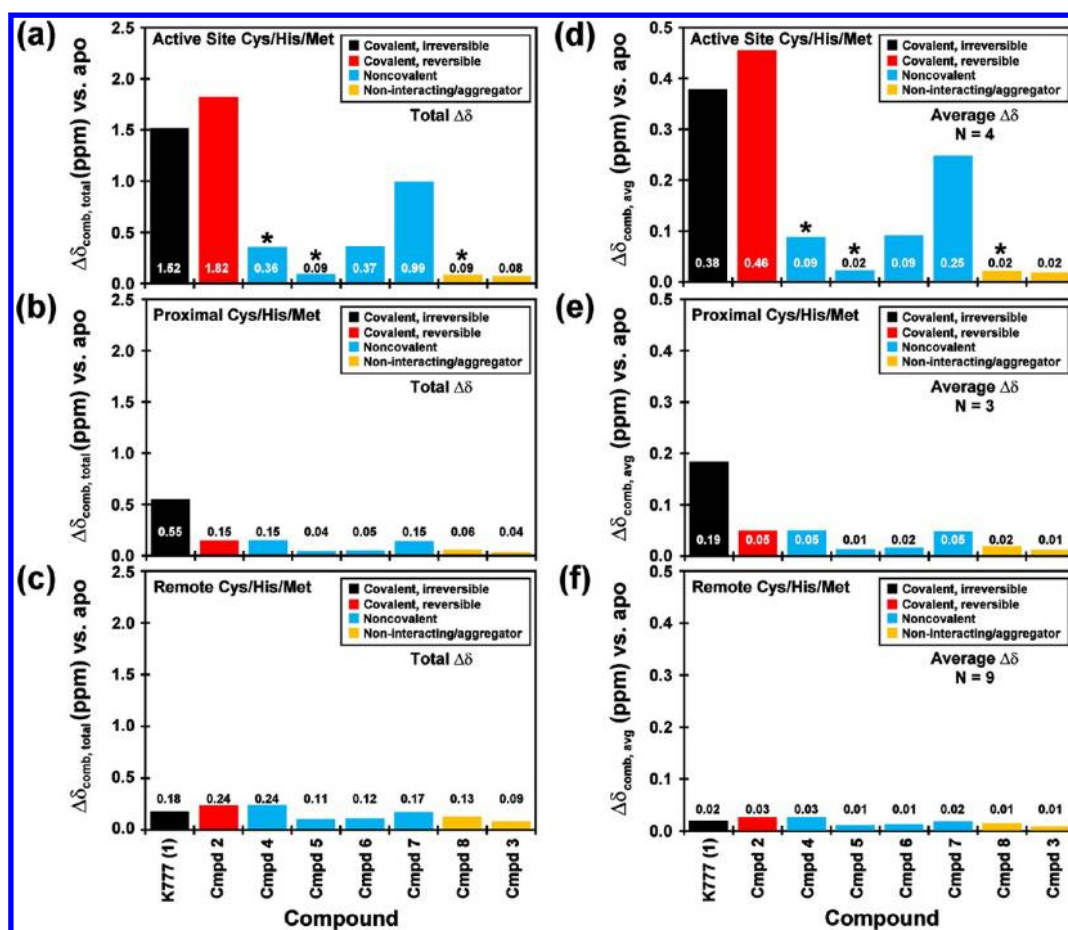
Figure 5. Summary of the cruzain−inhibitor NMR titration data. Chemical shift perturbations of the selectively labeled [15N]Cys, [15N]His, and [13C]Met cruzain resonances upon addition of the inhibitors listed in Table 1. The bar charts represent the summation of the chemical shift perturbations and are separated as Cys, His, and Met residues located (a) in the substrate binding pocket, (b) proximally to the substrate binding pocket, and (c) outside the substrate binding pocket. Average shift perturbations corresponding to the same residue classifications (d−f) are also plotted. The inhibitors are categorized as covalent and irreversible (black), covalent and reversible (red), noncovalent (blue), and noninteracting or aggregators (yellow). Asterisks indicate exchange broadening of the resonance peaks upon addition of inhibitor. Shift perturbations of the Cys, His, and Met residues plotted individually are displayed in Figures S6−S8 of the Supporting Information.

[15N]Cys and [15N]His HSQC spectra (Figures S5 and S6 of the Supporting Information). As with compound 8, minimal perturbations are observed for the Met145 peak (Figure S7 of the Supporting Information). Direct comparison of the overall and average reporter group chemical shift perturbations between compounds 2 and 3 (Figure 5) indicates that the nitrile derivative interacts more strongly with cruzain than the chlorinated derivative. If the Cys25 thiol were to covalently bind directly to the purine ring of compound 3, significant chemical shift perturbations on par with those observed for compound 2 would have been observed for the reporter groups. Collectively, the NMR titration data suggest a possible weak transient noncovalent interaction between cruzain and compound 3. The NMR data also verify the previously reported cruzain−compound 2 binding mode:[24] the Cys25 thiol sulfur is covalently bonded to the nitrile group, and not directly to the purine group of the inhibitor.

General Trends. For the set of compounds listed in Table 1, all of which exhibit binding or inhibitory constants in the low micromolar range, the total summed and average [15N]Cys, [15N]His, and [13C]Met shift perturbations are largest for the covalently bound inhibitors (K777 and compound 2) and smallest for the noninteractors and aggregators (Figure 5). As expected for these competitive inhibitors, which were designed

to target the active site, residues located within the substrate binding site (Cys25, His162, Met68, and Met145) display larger total and average chemical shift perturbations relative to the "proximal" (Cys22, Cys63, and Met152) and "remote" residues. The data for compounds 4 and 5 are the exceptions to these trends, in which extensive peak broadening of the catalytic residues would limit any contribution to the observed total chemical shift perturbation.

Maximal chemical shift perturbations of the "catalytic" cysteine, histidine, and methionine residues were observed at a <2-fold excess of the compound relative to cruzain in the cases of the covalently bound inhibitors. Conversely for the noncovalent inhibitors, a >20-fold excess of the compound was required before apparent maximal chemical shift perturbations were reached. A comparison of the [15N]Cys data (Figure S5 of the Supporting Information) reveals distinct patterns: the covalently bound inhibitors impart chemical shift perturbations on the catalytic Cys25 of >0.5 ppm, while the noncovalent compounds display values of <0.15 ppm for all the cysteine reporter groups. In the case of the [15N]His data (Figure S6 of the Supporting Information), the chemical shift perturbation values of the catalytic His162 observed for the K777 and compound 2 titrations appear to have reached a maximal value at a 2.5-fold excess of compounds, with no significant

differences with a 20-fold excess of the inhibitor. Conversely, although significant perturbations are observed for His162 upon addition of 2.5 equiv of the noncovalent inhibitors, such as compounds **4**, **6**, and **7**, further changes in the chemical shift resonances are observed at 20-fold excesses. Similar patterns are observed for Met68 (Figure S7 of the Supporting Information).

■ DISCUSSION

We have established an expression and purification protocol resulting in high yields of activated cruzain. The optimized autoinduction protocol has not only assisted in completing recent crystallography efforts[13,24] but also allowed for the cost-effective production of isotopically labeled cruzain for use in NMR studies. Because of its inherent instability, active cruzain has not yet been crystallized in its apo form. All previously published cruzain SAR studies had relied on obtaining high-resolution X-ray crystal structures, usually with an inhibited protease. With the exception of compound **6** (PDB entry 3KKU),[13] none of the noncovalent inhibitors listed in Table 1 have been successfully cocrystallized with cruzain. The existing crystallographic structures of cruzain will continue to be utilized for in silico docking and modeling studies, and other high-throughput screening (HTS) methods such as fluorescence activity assays will help triage compound libraries for identifying potential lead inhibitors. However, neither the docking nor HTS studies will verify and characterize inhibitor binding sites or allow for more thorough biophysical studies of cruzain.

As we have illustrated here, with the advantages of selective isotopic labeling, binding modes of these inhibitors can be examined in a relatively fast manner using NMR-based methods. In particular, the use of selectively [^{15}N]Cys-, [^{15}N]His-, and [^{13}C]Met-labeled cruzain was helpful in quickly verifying the binding sites of eight small-molecule inhibitors. With the exception of the noninteractors and aggregators, titration of each compound using these cruzain samples perturbed the chemical shifts of residues at or near the active site. Importantly, these NMR titrations verified previously reported in silico docking results of selected noncovalently binding compounds,[13] including those for which there are no existing crystallographic data sets. Using current NMR automation technology with samples contained in standard 5 mm NMR tubes, a binary binding experiment (i.e., apo vs >20-fold excess of compound) can be performed in 1−2 h per sample. The rate of data acquisition can be further enhanced using NMR flow cells and 96-well plate sampling attachments, which requires smaller sample volumes.[25] This has the added advantage of allowing for higher protein and inhibitor concentrations relative to those in the standard 5 mm tubes and, more importantly, faster data acquisition times.

Results from the NMR titrations can be used to verify whether these candidate compounds bind to cruzain and if they are competitive inhibitors. In addition, NMR-based titrations may verify recently proposed allosteric binding sites located more than 10 Å from the catalytic Cys25.[26] Insights gained from the NMR titrations of potential inhibitors, especially those that exhibit weak binding affinities for cruzain, may also help determine whether efforts to optimize the compounds for further SAR studies are warranted. Although the NMR titration assays are not able to detect subtle differences between various chemical scaffolds, such as relative orientation within the binding pocket, they may be useful in determining differences between chemical analogues. For example, clear differences in the HSQC spectra are observed between compounds **2** (a covalent binder) and **3** (a noninteractor and/or aggregator). Both share the same chemical scaffold but differ in the functional group attached to the purine ring: a nitrile group in compound **2** and a chloro group in compound **3**. In these cases, NMR spectroscopy and X-ray crystallography can be used in conjunction to identify and optimize potential therapeutic leads for cruzain.

There has been intense and long-standing interest in cysteine proteases, particularly because of their involvement in a wide range of human diseases.[8,9,27,28] In particular, cysteine cathepsins, which make up 11 of the 15 canonical cathepsin family members, are known to play roles in extracellular matrix remodeling in humans, leading to the development of various pathologies, including cancer, cardiovascular diseases, and inflammatory diseases.[9,28] However, there are relatively few NMR-based studies that focus on their catalytic domains. One early effort focused on pH studies of active papain and relied on one-dimensional proton spectra.[29] A second study presented two-dimensional NMR data of the zymogen procathepsin L but did not report any useful resonance assignments.[30] The majority of the previous NMR-based structure calculations have centered on protein inhibitors such as chagasin (cruzain)[31] or p41icf (cathepsin L).[32]

More recently, NMR-based structural studies focusing on cysteine proteases such as streptopain,[33,34] foot and mouth disease virus leader protease,[35] ubiquitin C-terminal hydrolases,[36,37] and SARS coronavirus main protease (SARS CoV M^{pro})[38] have appeared in the literature. Importantly, in these enzyme constructs, the catalytic cysteine had been substituted with either serine or alanine or, in the case of SARS CoV M^{pro}, truncated, rendering the proteases inactive. To date, there are few examples of active papain-like cysteine proteases that have been extensively studied via NMR, including the Josephin domain of ataxin-3,[39−41] the NlpC/P60 domain of lipoprotein Spr,[42] and Sortase A.[43] All of these proteases contain the catalytic cysteine and histidine residues positioned in approximately the same relative positions as in papain. However, all have low degrees of overall sequential (<15%) and structural (backbone rmsd of >3.5 Å) homology with respect to papain and other members of the cathepsin cysteine protease family. Thus, to the best of our knowledge, the data presented herein represent the first NMR-based study focusing on the mature, active form of a papain/cathepsin-like cysteine protease in its wild-type state.

On a more basic level, production of NMR-ready cruzain will allow for more extensive study of structure−function relationships. For example, backbone dynamics of both the zymogen and mature cruzain can now be examined to more fully elucidate how conformational mobility influences proteolytic activity. In addition, selectively labeled samples may also be utilized to determine the pK_a values of residues such as the catalytic Cys25 and His162 (Supporting Information and Figures S8−S11 of the Supporting Information). Such information may be useful in further dissecting the mechanism of the enzyme.

Further insight into the mechanism of in vitro procruzain self-activation is provided by our observation that K777 impedes but does not completely abolish cleavage of the pro-region from the catalytic domain. Examination of the structurally homologous Cys25 to Ser variant of procathepsin L (PDB entry 1CJL, backbone rmsd for the catalytic domain with PDB entry 2OZ2 of 0.582 Å)[44] shows a short helix positioned orthogonally at the active site occluding the catalytic

residues. Notably, the 44 residues immediately preceding the N-terminus of the catalytic domain lie in the substrate binding cleft and are relatively unstructured. As has been suggested for all members of the papain/cathepsin cysteine protease family,[44] the pro-region in cruzain may likewise adopt a similar conformation and mode of autoinhibition. The relatively unstructured state of the pro-region may allow for transient interactions with the catalytic domain, which in turn allows for the active site to be exposed in a catalytically competent state. Once an "active" enzyme is available, it would be ready to proteolyze other procruzain molecules, initiating a cascade of autoactivation events.

The expression and purification strategies described herein may also be adapted to papain and the other 11 known structurally homologous cysteine cathepsins, allowing the possibility of characterizing their solution states via NMR spectroscopy or other biophysical methods. For example, K777 was recently demonstrated to inhibit the activities of cathepsins B, L, and S isolated from pancreatic extracts.[45] Results from these studies could improve our understanding of mechanisms that govern enzymatic activity and assist in the discovery of new inhibitors for this large family of closely related cysteine proteases.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional methodology and experimental results; a table listing the available chemical shift assignments for the [U-^{13}C,^{15}N,^{2}H]-cruzain−K777 complex; a table listing the apparent fitted [^{15}N]Cys, [^{15}N]His, and [^{13}C]Met pK_a values; HSQC overlay of the protonated and deuterated cruzain−K777 complex; HSQC spectrum annotated with available resonance assignments; HSQC overlays of uniformly and selectively labeled cruzain with or without K777; θ_{222} band and estimated fractional helicity bar charts from the CD denaturation study; steady-state ^{1}H−{^{15}N} NOE ratios for the cruzain−K777 complex; chemical shift perturbation bar charts summarizing the inhibitor−cruzain titrations; and [^{15}N]His, [^{15}N]Cys, and [^{13}C]Met pH titration curves of MMTS- and K777-inhibited cruzain. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The available chemical shift resonance assignments of the cruzain−K777 complex have been submitted to the Biological Magnetic Resonance Data Bank (http://www.bmrb.wisc.edu/) as entry 18236.

AUTHOR INFORMATION

Corresponding Author

*E-mail: charles.craik@ucsf.edu. Telephone: (415) 476-8146.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

AMC, 7-amino-4-methylcoumarin; DTT, dithiothreitol; GdnHCl, guanidinium hydrochloride; HSQC, heteronuclear single-quantum coherence spectroscopy; MMTS, methyl methanethiosulfonate; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PMSF, phenylmethanesulfonyl fluoride; rmsd, root-mean-square deviation; SAR, structure−activity relationship; Z or N-CBZ, N-carbobenzyloxy.

REFERENCES

(1) Sajid, M., Robertson, S. A., Brinen, L. S., and McKerrow, J. H. (2011) Cruzain: The path from target validation to the clinic. *Adv. Exp. Med. Biol. 712*, 100−115.

(2) Le Loup, G., Pialoux, G., and Lescure, F. X. (2011) Update in treatment of Chagas disease. *Curr. Opin. Infect. Dis. 24*, 428−434.

(3) Munoz-Saravia, S. G., Haberland, A., Wallukat, G., and Schimke, I. (2012) Chronic Chagas' heart disease: A disease on its way to becoming a worldwide health problem: Epidemiology, etiopathology, treatment, pathogenesis and laboratory medicine. *Heart Failure Rev. 17*, 45−64.

(4) Bern, C., Kjos, S., Yabsley, M. J., and Montgomery, S. P. (2011) *Trypanosoma cruzi* and Chagas' Disease in the United States. *Clin. Microbiol. Rev. 24*, 655−681.

(5) Gutteridge, W. E. (1985) Existing chemotherapy and its limitations. *Br. Med. Bull. 41*, 162−168.

(6) Kerr, I. D., Lee, J. H., Farady, C. J., Marion, R., Rickert, M., Sajid, M., Pandey, K. C., Caffrey, C. R., Legac, J., Hansell, E., McKerrow, J. H., Craik, C. S., Rosenthal, P. J., and Brinen, L. S. (2009) Vinyl sulfones as antiparasitic agents and a structural basis for drug design. *J. Biol. Chem. 284*, 25697−25703.

(7) Kamphuis, I. G., Kalk, K. H., Swarte, M. B., and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution. *J. Mol. Biol. 179*, 233−256.

(8) McGrath, M. E. (1999) The lysosomal cysteine proteases. *Annu. Rev. Biophys. Biomol. Struct. 28*, 181−204.

(9) Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012) Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta 1824*, 68−88.

(10) Doyle, P. S., Zhou, Y. M., Engel, J. C., and McKerrow, J. H. (2007) A Cysteine Protease Inhibitor Cures Chagas' Disease in an Immunodeficient-Mouse Model of Infection. *Antimicrob. Agents Chemother. 51*, 3932−3939.

(11) McKerrow, J. H., Doyle, P. S., Engel, J. C., Podust, L. M., Robertson, S. A., Ferreira, R, Saxton, T., Arkin, M., Kerr, I. D., Brinen, L. S., and Craik, C. S. (2009) Two approaches to discovering and developing new drugs for Chagas disease. *Mem. Inst. Oswaldo Cruz 104* (Suppl. 1), 263−269.

(12) Ferreira, R. S., Bryant, C., Ang, K. K., McKerrow, J. H., Shoichet, B. K., and Renslo, A. R. (2009) Divergent modes of enzyme inhibition in a homologous structure-activity series. *J. Med. Chem. 52*, 5005−5008.

(13) Ferreira, R. S., Simeonov, A., Jadhav, A., Eidam, O., Mott, B. T., Keiser, M. J., McKerrow, J. H., Maloney, D. J., Irwin, J. J., and Shoichet, B. K. (2010) Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *J. Med. Chem. 53*, 4891−4905.

(14) Eakin, A. E., Mills, A. A., Harth, G., McKerrow, J. H., and Craik, C. S. (1992) The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. *J. Biol. Chem. 267*, 7411−7420.

(15) Eakin, A. E., McGrath, M. E., McKerrow, J. H., Fletterick, R. J., and Craik, C. S. (1993) Production of crystallizable cruzain, the major cysteine protease from *Trypanosoma cruzi*. *J. Biol. Chem. 268*, 6115–6118.

(16) Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif. 41*, 207–234.

(17) Shahian, T., Lee, G. M., Lazic, A., Arnold, L. A., Velusamy, P., Roels, C. M., Guy, R. K., and Craik, C. S. (2009) Inhibition of a viral enzyme by a small-molecule dimer disruptor. *Nat. Chem. Biol. 5*, 640–646.

(18) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(19) Goddard, T. D., and Kneller, D. G. (1999) *Sparky 3*, University of California, San Francisco.

(20) Cavanagh, J., Fairbrother, W. E., Palmer, A. G., III, and Skelton, N. J. (1996) *Protein NMR Spectroscopy Principles and Practice*, Academic Press, San Diego.

(21) Lee, G. M., Shahian, T., Baharuddin, A., Gable, J. E., and Craik, C. S. (2011) Enzyme Inhibition by Allosteric Capture of an Inactive Conformation. *J. Mol. Biol. 411*, 999–1016.

(22) Gardner, K. H., and Kay, L. E. (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct. 27*, 357–406.

(23) Lecaille, F., Authie, E., Moreau, T., Serveau, C., Gauthier, F., and Lalmanach, G. (2001) Subsite specificity of trypanosomal cathepsin L-like cysteine proteases. Probing the S2 pocket with phenylalanine-derived amino acids. *Eur. J. Biochem. 268*, 2733–2741.

(24) Mott, B. T., Ferreira, R. S., Simeonov, A., Jadhav, A., Ang, K. K., Leister, W., Shen, M., Silveira, J. T., Doyle, P. S., Arkin, M. R., McKerrow, J. H., Inglese, J., Austin, C. P., Thomas, C. J., Shoichet, B. K., and Maloney, D. J. (2010) Identification and optimization of inhibitors of trypanosomal cysteine proteases: Cruzain, rhodesain, and TbCatB. *J. Med. Chem. 53*, 52–60.

(25) Kautz, R. A., Goetzinger, W. K., and Karger, B. L. (2005) High-Throughput Microcoil NMR of Compound Libraries Using Zero-Dispersion Segmented Flow Analysis. *J. Comb. Chem. 7*, 14–20.

(26) Durrant, J. D., Keranen, H., Wilson, B. A., and McCammon, J. A. (2010) Computational identification of uncharacterized cruzain binding sites. *PLoS Neglected Trop. Dis. 4*, e676.

(27) Vasiljeva, O., Reinheckel, T., Peters, C., Turk, D., Turk, V., and Turk, B. (2007) Emerging roles of cysteine cathepsins in disease and their potential as drug targets. *Curr. Pharm. Des. 13*, 387–403.

(28) Reiser, J., Adair, B., and Reinheckel, T. (2010) Specialized roles for cysteine cathepsins in health and disease. *J. Clin. Invest. 120*, 3421–3431.

(29) Johnson, F. A., Lewis, S. D., and Shafer, J. A. (1981) Determination of a low pK for histidine-159 in the S-methylthio derivative of papain by proton nuclear magnetic resonance spectroscopy. *Biochemistry 20*, 44–48.

(30) Jerala, R., Zerovnik, E., Kidric, J., and Turk, V. (1998) pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation. *J. Biol. Chem. 273*, 11498–11504.

(31) Salmon, D., do Aido-Machado, R., Diehl, A., Leidert, M., Schmetzer, O., de A Lima, A. P., Scharfstein, J., Oschkinat, H., and Pires, J. R. (2006) Solution structure and backbone dynamics of the *Trypanosoma cruzi* cysteine protease inhibitor chagasin. *J. Mol. Biol. 357*, 1511–1521.

(32) Chiva, C., Barthe, P., Codina, A., Gairi, M., Molina, F., Granier, C., Pugniere, M., Inui, T., Nishio, H., Nishiuchi, Y., Kimura, T., Sakakibara, S., Albericio, F., and Giralt, E. (2003) Synthesis and NMR structure of p41icf, a potent inhibitor of human cathepsin L. *J. Am. Chem. Soc. 125*, 1508–1517.

(33) Luo, S.-C., Chen, C.-Y., Lin, Y.-S., Jeng, W.-Y., and Chuang, W.-J. (2003) Letter to the Editor: Backbone ^{1}H, ^{15}N and ^{13}C resonance assignments of the 28 kDa mature form of streptopain. *J. Biomol. NMR 25*, 165–166.

(34) Wang, C. C., Houng, H. C., Chen, C. L., Wang, P. J., Kuo, C. F., Lin, Y. S., Wu, J. J., Lin, M. T., Liu, C. C., Huang, W., and Chuang, W. J. (2009) Solution structure and backbone dynamics of streptopain: Insight into diverse substrate specificity. *J. Biol. Chem. 284*, 10957–10967.

(35) Cencic, R., Mayer, C., Juliano, M. A., Juliano, L., Konrat, R., Kontaxis, G., and Skern, T. (2007) Investigating the substrate specificity and oligomerisation of the leader protease of foot and mouth disease virus using NMR. *J. Mol. Biol. 373*, 1071–1087.

(36) Harris, R., Eidhoff, U., Vinzenz, D., Renatus, M., Gerhartz, B., Hommel, U., and Driscoll, P. C. (2007) Backbone ^{1}H, ^{13}C, and ^{15}N resonance assignments for the 26-kD human de-ubiquitinating enzyme UCH-L3. *Biomol. NMR Assignments 1*, 51–53.

(37) Andersson, F. I., Jackson, S. E., and Hsu, S. T. (2010) Backbone assignments of the 26 kDa neuron-specific ubiquitin carboxyl-terminal hydrolase L1 (UCH-L1). *Biomol. NMR Assignments 4*, 41–43.

(38) Zhang, S., Zhong, N., Ren, X., Jin, C., and Xia, B. (2011) ^{1}H, ^{13}C and ^{15}N resonance assignments of SARS-CoV main protease N-terminal domain. *Biomol. NMR Assignments 5*, 143–145.

(39) Nicastro, G., Masino, L., Frenkiel, T. A., Kelly, G., McCormick, J., Menon, R. P., and Pastore, A. (2004) Letter to the Editor: Assignment of the ^{1}H, ^{13}C, and ^{15}N resonances of the Josephin domain of human ataxin-3. *J. Biomol. NMR 30*, 457–458.

(40) Nicastro, G., Menon, R. P., Masino, L., Knowles, P. P., McDonald, N. Q., and Pastore, A. (2005) The solution structure of the Josephin domain of ataxin-3: Structural determinants for molecular recognition. *Proc. Natl. Acad. Sci. U.S.A. 102*, 10493–10498.

(41) Nicastro, G., Todi, S. V., Karaca, E., Bonvin, A. M., Paulson, H. L., and Pastore, A. (2010) Understanding the role of the Josephin domain in the polyUb binding and cleavage properties of ataxin-3. *PLoS One 5*, e12430.

(42) Aramini, J. M., Rossi, P., Huang, Y. J., Zhao, L., Jiang, M., Maglaqui, M., Xiao, R., Locke, J., Nair, R., Rost, B., Acton, T. B., Inouye, M., and Montelione, G. T. (2008) Solution NMR structure of the NlpC/P60 domain of lipoprotein Spr from *Escherichia coli*: Structural evidence for a novel cysteine peptidase catalytic triad. *Biochemistry 47*, 9715–9717.

(43) Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O., and Clubb, R. T. (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A. 98*, 6056–6061.

(44) Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J. 15*, 5492–5503.

(45) Lyo, V., Cattaruzza, F., Kim, T. N., Walker, A. W., Paulick, M., Cox, D., Cloyd, J., Buxbaum, J., Ostroff, J., Bogyo, M., Grady, E. F., Bunnett, N. W., and Kirkwood, K. S. (2012) Active Cathepsins B, L and S in Murine and Human Pancreatitis. *Am. J. Physiol. 303*, G894–G903.

(46) Irwin, J. J., and Shoichet, B. K. (2005) ZINC: A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model. 45*, 177–182.