HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2018 July 05.

Published in final edited form as:

Biochemistry. 2017 May 02; 56(17): 2304–2314. doi:10.1021/acs.biochem.6b01066.

Structure Determination of *Mycobacterium tuberculosis* Serine Protease Hip1 (Rv2224c)

Jacqueline L. Naffin-Olivos[†], **Andrew Daab**[†,‡], **Andre White**[†], **Nathan E. Goldfarb**[§], **Amy C. Milne**[†], **Dali Liu**[‖], **Jacqueline Baikovitz**[⊥], **Ben M. Dunn**[#], **Jyothi Rengarajan**[▽], **Gregory A. Petsko**[○], and **Dagmar Ringe**[*,†]

[†]Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02454, United States

[‡]Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, United States

[§]Department of Pharmaceutical and Biomedical Sciences, California Health Sciences University, Clovis, California 93612, United States

[‖]Department of Chemistry and Biochemistry, Loyola University Chicago, Chicago, Illinois 60660, United States

[⊥]Department of Biology, Brandeis University, Waltham, Massachusetts 02454, United States

[#]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida 32610, United States

[▽]Division of Infectious Diseases, Department of Medicine, Emory Vaccine Center, Emory University School of Medicine, Atlanta, Georgia 30329, United States

[○]Appel Alzheimer's Disease Research Institute, Weill Cornell Medical College, New York, New York 10021, United States

Abstract

The *Mycobacterium tuberculosis* (Mtb) serine protease Hip1 (hydrolase important for pathogenesis; Rv2224c) promotes tuberculosis (TB) pathogenesis by impairing host immune responses through proteolysis of a protein substrate, Mtb GroEL2. The cell surface localization of Hip1 and its immunomodulatory functions make Hip1 a good drug target for new adjunctive immune therapies for TB. Here, we report the crystal structure of Hip1 to a resolution of 2.6 Å and

[*]Corresponding Author: Phone: 781-736-4902. ringe@brandeis.edu.

ORCID

Jacqueline L. Naffin-Olivos: 0000-0002-3356-2419

Author Contributions

D.R., G.A.P., J.L.N.O., and A.D. designed the research; J.L.N.O., A.D., A.C.M., D.L., and J.B. performed the research; J.L.N.O., A.W., A.D., D.R., and N.G. analyzed data; D.R., G.A.P., and B.M.D. contributed reagents/materials/analysis tools; and J.L.N.O., D.R., and J.R. wrote the paper.

Notes

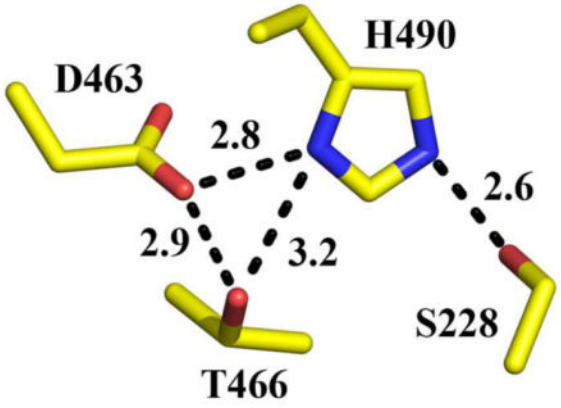
The authors declare no competing financial interest.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b01066. Crystallographic data and refinement statistics, sequence alignment, unmodeled electron density, structural alignment; as well as, anion exchange column chromatogram and CD spectra of Hip1wt and Hip1T466A proteins. (PDF)

the kinetic studies of the enzyme against model substrates and the protein GroEL2. The structure shows a two-domain protein, one of which contains the catalytic residues that are the signature of a serine protease. Surprisingly, a threonine is located within the active site close enough to hydrogen bond with the catalytic residues Asp463 and His490. Mutation of this residue, Thr466, to alanine established its importance for function. Our studies provide insights into the structure of a member of a novel family of proteases. Knowledge of the Hip1 structure will aid in designing inhibitors that could block Hip1 activity.

Graphical Abstract



Tuberculosis (TB) remains a serious global public health problem with about 1.5 million deaths reported in 2014 and an estimated 480,000 new cases of multidrug-resistant TB.[1] The bacterium that causes TB, *Mycobacterium tuberculosis* (Mtb), has evolved a wide array of strategies to evade host innate immunity and cause disease. Mtb can replicate within macrophages, inhibit phagosome maturation and acidification, interfere with responses to gamma interferon (IFN-γ), and resist the microbicidal functions of macrophages.[2,3] In addition, Mtb modulates macrophage responses and interferes with the function of dendritic cells (DCs), which enable this pathogen to escape early detection by innate immunity and prevent optimal adaptive immunity. Delineating the mycobacterial factors that modulate host immunity and promote pathogen virulence is therefore important for developing new therapeutics for TB. We have identified a cell wall-associated Mtb serine hydrolase, Hip1, (hydrolase important for pathogenesis; Rv2224c), as a key immunomodulatory virulence factor.[4] We have shown that Hip1 is a serine protease that dampens macrophage proinflammatory responses, inhibits maturation of DCs, and impairs antigen presentation and T cell responses.[5,6] Further, we have shown that Hip1-mediated proteolytic cleavage of a specific protein substrate, Mtb GroEL2, is an immune evasion mechanism that serves to dampen innate immunity.[7] Thus, while a cleaved form of GroEL2 predominates in wild-type Mtb, a *hip1* mutant strain of Mtb retains a full length, uncleaved form of GroEL2 that is more immunostimulatory. Indeed, mice infected with the *hip1*

mutant survive significantly longer than wild-type Mtb-infected mice and display only mild lung immunopathology despite high bacterial burdens.[7] These studies suggest that Hip1 functions to restrict beneficial immune responses and thereby promotes progression to TB

disease. The contribution of GroEL2 proteolysis to Hip1-dependent dampening of innate immune responses indicates that Hip1-mediated cleavage of GroEL2 represents a novel immune evasion mechanism that allows the pathogen to manipulate host immunity to its advantage and drive disease progression and pathology. Thus, Hip1 protease activity is a good drug target for developing adjunctive immunomodulatory therapeutics against Mtb.

Hip1 is predicted to be a lipoprotein based on an N-terminal type II signal peptide and a cysteine that is expected to be a lipoprotein anchor residue.[8] Sequence analysis shows Hip1 has limited similarity to the *Streptomyces lividans* tripeptidyl-peptidases; TPP A (Tap; 30% seq ID), TPP B (SlpD; 36% seq ID) and TPP C (SlpE; 36% seq ID), as well as proline iminopeptidases (PIPs; also known as prolyl aminopepidases, PAPs); such as XCPIP (24% seq ID) and SPAP (21%) (Figure S1). These enzymes are all serine peptidases and are categorized in clan SC, family S33.[9] Characteristics of the clan include a catalytic triad in the order of Ser, Asp, and His within the sequence with the active site Ser residue usually part of a GXSXG consensus sequence.[10] All families within the clan are structurally similar in that they share an α/β hydrolase fold, a fold common among other enzymes including hydrolases, esterases, lipases, and proteases.[11] The S33 family consists mainly of exopeptidases that act on the N-terminus of peptides, preferably cleaving at a Pro residue. However, there are other homologues within the family that are not peptidases such as epoxide hydrolases and haloalkane dehalogenase. At the level of sequence identity given above, it is not possible to predict the exact activity and substrate specificity of Hip1. Consequently, a three-dimensional structure is needed to clarify the biochemical role of the enzyme and provide an accurate model to aid in the design of specific inhibitors.

We have recently published a molecular and biochemical characterization of Hip1 enzymatic activity and demonstrated that Hip1 exhibits serine protease activity against protein and peptide substrates and against a physiological substrate, Mtb GroEL2. By screening standard inhibitors directed at serine proteases, we demonstrated that 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF) inactivated the enzyme. In an effort to identify substrate specificity, small peptide *p*-nitroanilide substrates were tested with Hip1. Substrates were chosen by analogy to the substrate specificity of the exopeptidases TPP A and TPP B that share some sequence similarity with Hip1. The enzyme can recognize and cleave at the C-terminus of small peptides such as Ala-Pro-Ala and Gly-Pro-Leu, albeit inefficiently. Unfortunately, these experiments did not address the substrate specificity of the enzyme. Surprisingly, we also observed that Hip1 cleaves human GroEL2 within the N-terminus of the protein (between Arg_{12} and Gly_{13}), releasing a peptide fragment of 12 amino acid residues.[12] Taken together, these experiments indicate that the specific substrate and cleavage site preferences of Hip1 remain unclear and that the protease may have both exo- and endopeptidase activity. In addition, it is likely that this protease has additional physiological substrates.

In order to better understand the mechanism of the Hip1 cleavage reaction and to gain insight into the determinants for substrate recognition, we have crystallized the protein in its apo form and determined its three-dimensional structure. The resulting model shows structural similarities to the S33 family of serine peptidases but also has several novel features that may be related to its specific recognition of a large protein substrate. The active

site region is accessed through a large opening that suggests that the enzyme has endopeptidase activity rather than exopeptidase activity. The P1 pocket (see Schechter and Berger for nomenclature[13]) can accommodate a side chain larger than the proline residue that is the signature for S33 family serine peptidases.

MATERIALS AND METHODS

Cloning, Expression, and Purification of Proteins

Hip1wt, Hip1T466A, and Hip1SeMet—*Mtb* Hip1 protein was expressed from the plasmid, pET28Hip1Δ49, which was reported previously.[12] This construct lacks the first 49 amino acids of the protein, which removes the N-terminal signal sequence and contains a histidine affinity tag (6xHis-tag) at the N-terminus. To produce Hip1 protein with a threonine to alanine mutation at amino acid residue 466, plasmid pET28Hip1Δ49 T466A was obtained by site-directed mutagenesis using primers 5′-CCCATGACCCGGCGGCGCCGTACAAGGCC-3′ and 5′-GGCCTTGTACGGCGCCGCCGGGTCATGGG-3′. The plasmids were transformed separately into *E. coli* BL21 Star (DE3) cells (Invitrogen, Carlsbad, CA) and grown in Luria–Bertani (LB) broth for both the wild-type protein (Hip1wt) and Hip1 threonine to alanine mutation (Hip1T466A). In addition, cells containing the plasmid pET28Hip1Δ49 were also grown in minimal media to produce protein with the methionine residues replaced with selenomethionine (Hip1SeMet).

For expression of Hip1wt or Hip1T466A, 1 L of LB broth containing 50 μg/mL kanamycin was inoculated with 5 mL of overnight culture and incubated at 37 °C to an OD_{600} of 0.6 to 1.0. The cells were cooled to room temperature for 15–30 min, and then, protein production was induced with 1 mM IPTG (isopropyl β-D-thiogalactopyranoside, Gold Biotechnology, St. Louis, MO). After overnight incubation at 25 °C, the cells were harvested by centrifugation for 30 min to 1 h at 5,000 rpm. The pellet was resuspended in 25 mL of water to wash away any residual media. The cells were once again pelleted by centrifugation for 30 min at 3,500 rpm and frozen at −80 °C until further use.

For expression of Hip1SeMet, 3 mL of LB was inoculated and grown for 9 h at 37 °C. The cells were pelleted to remove LB media, resuspended in 20 mL of M9 minimal media, and grown overnight at 37 °C. The overnight culture (5 mL) was used to inoculate 1 L of M9 minimal media. Once an OD_{600} of 0.5 to 0.6 was reached, an amino acid mixture was added (100 mg K, F, and T; 50 mg I, L, and V; and 60 mg SeMet), and the cells were allowed to grow for an additional 15 min. Protein production was induced and cells harvested as given for Hip1wt. All media contained 50 μg/mL kanamycin.

Purification of Hip1wt and Hip1T466A was performed as previously reported.[12] The Hip1SeMet was purified the same way as Hip1wt. In summary, the cells were sonicated and the insoluble fraction containing inclusion bodies was washed and then resuspended in buffer containing 8 M urea. After the protein was solubilized, nickel resin (Qiagen) was added, and the protein was refolded on the beads within an anaerobic chamber by decreasing the concentration of urea in a stepwise gradient. The protein was eluted with 250 mM imidazole and dialyzed overnight against 50 mM Tris-HCl at pH 8.0 plus 10% glycerol

(Buffer A) using 10 kDa molecular cutoff dialysis tubing. The dialyzed protein was loaded onto a MonoQ column equilibrated with Buffer A and eluted using a gradient of Buffer B (50 mM Tris at pH 8.0, 1 M NaCl, and 10% glycerol). The protein eluted at approximately 100–160 mM NaCl. The protein fractions were aliquoted and stored at −80°C. Incorporation of selenium was validated by mass spectroscopy (data not shown).

GroEL2—GroEl2 protein was expressed and purified as reported previously.[12] Briefly, the plasmid pACYCDUET-1 GroEL2, which contains full-length GroEL2 with an added histidine affinity tag (6xHis-tag) at the N-terminus and a S-tag at the C-terminus, was transformed into *E. coli* BL21 Star (DE3) cells (Invitrogen). LB broth (1 L) containing 34 μg/mL chloramphenicol was inoculated with 5 mL of overnight culture and incubated at 37 °C to an OD_{600} of 0.6–0.8. After cooling cells for 15–30 min to room temperature, 1 mM IPTG was added, and the cells were incubated overnight at 28 °C followed by centrifugation. The cell pellet was lysed by sonication after resuspension in 50 mM $NaPO_4$ at pH 8.0, 300 mM NaCl, 10 mM imidazole, and protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN). The GroEl2 protein was purified via a series of three columns. First, the protein was bound and eluted from Ni-NTA beads, followed by a MonoQ anion exchange column, and a last size exclusion S200 column equilibrated with buffer containing 50 mM Tris at pH 8.0 and 150 mM NaCl. The final protein was concentrated, aliquoted, and stored at −80 °C.

All proteins were analyzed for purity by SDS–PAGE. The Bradford assay was used to determine the concentration of the purified proteins with bovine serum albumin (BSA) as a standard.

Circular Dichroism (CD)

CD spectra were acquired at room temperature using a Jasco J-810 Spectropolarimeter scanning from 200 to 280 nm at a rate of 20 nm/min and a bandwidth of 1.0 nm. The resulting spectra are an average of five scans. Hip1wt and Hip1T466A concentrations were 1.5 and 2.3 μM, respectively, in buffer containing 50 mM sodium phosphate at pH 7.0 plus 150 mM NaCl. A spectrum of the buffer only was obtained and subtracted from the protein spectra. The data were then converted to molar ellipticity $[\theta]$ using the equation $[\theta] = (100 \times \theta)/(C \times l)$, where θ is the degree of ellipticity or raw data, C is the concentration in molar, and l is the cell path length in cm. A plot for the molar ellipticity vs wavelength from 200 to 240 nm was generated using GraphPad Prism version 7.0a for MacOS X, GraphPad Software, La Jolla California USA (www.graphpad.com).

Thermofluor Assay

Screening for Crystal Growth Conditions.[14]—To determine the optimal solution conditions for stability, the protein was tested against a number of pH values, salts, and additives. Hip1wt protein was buffer exchanged into 50 mM Tris at pH 8.0 to remove the glycerol and concentrated to 125 μM. The protein was diluted to 5–10 μM in various buffers at 50 mM concentration including sodium phosphate at pH 4.25, 5.0, 6.0, 7.0; sodium acetate at pH 4.6; Hepes at pH 8.0; Tris at pH 8.0; CHES at pH 9.0. At pH 8, 40% PEG (range 200–6000) or 20% glycerol was added to test for their effects on protein stability.

Sypro orange dye (Invitrogen) was diluted to a final concentration of 10×. The assays were performed in a 96 well plate with a final volume of 25 μL. To determine thermo-stability, the temperature was increased from 25 to 95 °C in increments of 0.3 °C per second using a StepOnePLus real-time PCR system (Applied Biosystems). The raw data were normalized using Perl Script and then plotted using Excel (data not shown).

Stabilities of Hip1 Proteins—Hip1wt, Hip1SeMet, Hip1S228A, or Hip1T466A protein (2 μM) were added to a 96 well plate with buffer (50 mM sodium phosphate at pH 7.4, and 150 mM NaCl) and 10× Sypro orange dye to a final volume of 25 μL. The stability of the proteins was monitored by increasing the temperature from 25 to 95 °C in increments of 0.3 °C per second. The raw data were normalized using Perl Script and then plotted using GraphPad Prism 7.0a.

Enzyme Activity

Azocasein Assay—Hip1wt, Hip1SeMet, Hip1S228A, or Hip1T466A proteins (2 μg in 25 μL) were incubated for 30 min at 37 °C before the addition of the substrate. BSA (Thermo Scientific) and the protease Subtilisin Carlsberg (Sigma-Aldrich) were used as negative and positive controls, respectively, and treated the same as the Hip1 proteins. Azocasein (Sigma-Aldrich) was reconstituted in 1× phosphate buffered saline, pH 7.4 (Boston BioProducts), to produce a 5% solution and then passed through a 0.22 μm filter. Addition of the Hip1 proteins, BSA, or the protease Subtilisin Carlsberg resulted in a final concentration of 4% azocasein in 200 μL of total volume. After incubation for 1 h at 37 °C, the reactions were terminated by the addition of 200 μL of 10% trichloroacetic acid and incubated on ice for 30 min. The reactions were then centrifuged at 13,200 rpm for 10 min at 4 °C. The supernatant (200 μL) was transferred to a 96 well plate to which was added 50 μL of 1.8 N NaOH, and the absorbance was read at 440 nm. The enzyme activities are expressed as units of enzyme/mg protein (one enzyme unit is the quantity of enzyme required to increase the absorbance by 0.01 at 440 nm). The assay was performed in triplicate and graphed using GraphPad Prism 7.0a.

GroEL2 Substrate Assay—Hip1wt, Hip1SeMet, Hip1S228A, or Hip1T466A proteins (2 μM) and purified GroEl2 protein (0.65 μM) in 25 μL of 50 mM sodium phosphate at pH 7.4, and 150 mM NaCl were incubated for 16 h at 37 °C. SDS–PAGE loading dye (6×) was added to the samples, which were then boiled for 5 min and separated on a 10% SDS–PAGE gel. The protein bands were transferred onto a PDVF membrane (Millipore) and probed using an S-tag antibody (Novagen) followed by Goat Anti-Mouse IR Dye 800CW (LI-COR). Proteins were detected by IR fluorescence using an Odyssey Imager. Signal intensities for full-length GroELl2 and cleaved GroEL2 product were quantified with ImageJ software. The intensities for the two bands were converted to percentages of the two forms for each reaction. Since the GroEl2-only reaction does not have a band corresponding to the cleaved GroEl2 product, the percentage obtained for GroEl2 full-length was set to 100% and was used to normalize those reactions containing Hip1 proteins. The percent of GroEl2 cleaved product was graphed using GraphPad Prism 7.0. To compare Hip1wt and Hip1T466A cleavage of GroEl2 over time, the assay was repeated as previously described with the exception that Hip1wt or Hip1T466A proteins (1 μM) and GroEl2 protein (0.33

μM) were incubated in 1× Tris buffered saline, pH 7.4 (Boston BioProducts). Time points were taken at t = 0, 2, 4, 6, 14, 17, 21, and 25 h.

Kinetic Measurements

BApNA (Nα-Benzoyl-D,L-arginine 4-nitroanilide) Substrate—Hip1wt or Hip1T466A proteins at 40 μM was preincubated at 37 °C for 1 h in buffer A (50 mM sodium phosphate, pH 7.4, 150 mM NaCl). Trypsin and BSA were used as a positive and negative control, respectively, and treated the same as Hip1 proteins with the exception of the addition of 10 mM $CaCl_2$ to the trypsin buffer. BApNA (Sigma-Aldrich) was dissolved in 100% DMSO, heated at 37 °C for 10 min, and then diluted to concentrations ranging from 1 mM to 3 mM in buffer A + 4% DMSO. To a 96 well plate, 50 μL of each concentration of BApNA (1 mM to 3 mM) was added and incubated at 37 °C for 10 min. Hip1wt, Hip1T466A, BSA, or trypsin (50 μL) was added to BApNA for a final total volume of 100 μL, a DMSO concentration of 2%, a protein concentration of 20 μM, and BApNA concentrations ranging from 0.5 to 1.5 mM. The reaction was incubated at 37 °C and allowed to proceed for 2 h with a reading every 1 min. The enzyme activity was followed by monitoring the release of the *para*-nitroaniline in a continuous assay at 410 nm. Samples containing BApNA only at each concentration were used to subtract any absorbance due to the instability of the substrate itself. A standard curve was generated using *para*-nitroaniline (Sigma-Aldrich) to convert absorbance to μmol product released. GraphPad Prism 7.0a was used to plot μmol product released versus time.

GroEL2 Substrate—Hip1wt protein (74 μM in 50 mM Tris at pH 8.0, 150 mM NaCl, and 10% glycerol) was preincubated at 37 °C for 30 min. GroEL2 (at six concentrations between 10 μM to 200 μM in 50 mM sodium phosphate at pH 7.4, and 150 mM NaCl) was also preincubated for 10 min at 37 °C. The Hip1wt protein was diluted 1:18.5 with the GroEL2 solution for a final Hip1wt concentration of 4 μM. The final solution was kept at 37 °C, and aliquots were taken every 10 or 20 min up to 2 h and then every hour up to 5 h. SDS gel loading dye was added to stop the reaction and then stored at −80 °C. Once all time points were collected, the protein samples were separated on a 10% SDS–PAGE gel, transferred onto a PDVF membrane (Millipore), and probed using an S-tag antibody (Novagen) followed by Goat Anti-Mouse IR Dye 800CW (LI-COR). Proteins were detected by IR fluorescence using an Odyssey Imager. Signal intensities for full-length GroEL2 and cleaved GroEL2 product were quantified with ImageJ software, and the intensities for the two bands were converted to percentages of the two forms for each reaction. Using the value for the initial concentration of GroEL2, the percentages were converted to molarity, and the initial velocities of the reactions were obtained. The data were then plotted and fitted to the Michaelis–Menten equation by nonlinear regression using GraphPad Prism 7.0a software. The errors are reported for k_{cat} and K_m, and are determined from the standard deviation of three replicates.

Crystallization

Hip1wt, Hip1SeMet, or Hip1T466A, in storage buffer (50 mM Tris at pH 8.0, 10% glycerol, and approximately 130 mM NaCl) were concentrated to 15–30 mg/mL. To determine crystallization conditions, screening solutions were made based on the results of the

thermofluor assays. Hip1 was more stable at lower pH values, as well as with the additives PEG 3350, PEG6000, and glycerol. In addition, various salts were screened for their effects on crystallization. Optimal conditions for crystal formation and growth of Hip1wt were the following: crystals were grown at room temperature by the hanging drop method in 3 μL by mixing protein solution at 25 mg/mL (1:1) with well solution containing 30% PEG 3350, 0.1 M sodium acetate at pH 5.6, 0.2 M ammonium sulfate, and 10% glycerol. Crystals formed in 4–5 days as long thin rods, typically 0.6 mm × 0.06 mm. Seeding increased the rate of formation (crystals were observed in 2–4 days) but did not significantly improve the size of the crystals. For cryo protection, crystals were transferred to a solution consisting of 20% PEG400 plus native mother liquor and then flash-frozen with liquid nitrogen. Crystals of Hip1SeMet and Hip1T466A were obtained under the same conditions as for Hip1wt.

Data Collection

Data were collected at the National Synchrotron Light Source (NSLS-I), Brookhaven National Laboratory (Upton, NY) using beamline X29A equipped with an ADSC Q315 CCD X-ray area detector. Data for the SeMet derivative crystals were obtained by scanning for the optimal wavelength of the Se anomalous signal. The peak of the anomalous signal was found to be at 0.9790 Å. Data were collected from a single crystal at 100 K for 360° with a 1° oscillation angle and an exposure time of 0.6 s per image. Data were indexed, integrated, and scaled using the HKL 2000 software package.[15] The crystals possess symmetry corresponding to space group $P3_121$ and contain one molecule per asymmetric unit. See Table S1 for details.

For Hip1wt, data were collected at the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL) using the SBC-CAT beamline 19-ID equipped with a ADSC Q315r 315 mm × 315 mm mosaic CCD X-ray detector. Data were collected at a wavelength of 0.9793 Å from a single crystal at 100 K for 100° with a 1° oscillation angle and an exposure time of 2 s per image. Hip1T466A data were also collected at the Advanced Photon Source (APS), Argonne National Laboratory (Argonne, IL) using the GM/CA-CAT beamline 23-ID-B equipped with a MAR300 CCD X-ray detector. Data were collected at a wavelength of 1.033 Å from a single crystal at 100 K for 100° with a 0.5° oscillation angle and an exposure time of 1.5 s per image. The data were processed as previously described.

Phasing and Refinement

Se atoms were identified using the Autosol program in Phenix.[16] The eight expected sites were found; however, only six were used to determine phases as the other two Se atom sites had low occupancy. These two methionine residues are located at the N-terminus, a region with weak overall electron density. Initial phasing gave a figure of merit (FOM) of 0.34, followed by density modification to give a FOM of 0.66. Application of the sequence in the Autobuild program in Phenix gave an initial model with an R_{work} and R_{free} of 0.31 and 0.37, respectively. Manual rebuilding using Coot,[17] followed by several rounds of refinement in Phenix, resulted in a final model for Hip1SeMet with an R_{work} and R_{free} of 0.20 and 0.24. The Hip1SeMet model was then used for molecular replacement with diffraction data obtained for Hip1wt. After several rounds of refinement, a final model was obtained for Hip1wt with an R_{work} and R_{free} of 0.21 and 0.25. In turn, the Hip1wt model was then used

for molecular replacement with diffraction data obtained for Hip1T466A. After several rounds of refinement, a final model was obtained for Hip1T466A with an R_{work} and R_{free} of 0.22 and 0.26.

The final models for Hip1SetMet, Hip1wt, and Hip1T466A all contain 463 out of 492 residues. The first 21 amino acids of the N-terminus that include the His6 tag and linker region are not modeled due to poor electron density. The remaining 471 amino acids were renumbered to match the full-length native protein (residues 50–520). Residue numbers 57–64 (SSNPQVKI) were also not included due to poor electron density. Residues with electron density for the backbone but missing electron density for side chains (Glu51, Arg56, Lys128, and Lys423) were modeled and then refined with the occupancies of the side chain atoms being set to 0. The model was then validated using RAMPAGE: Ramachandran Plot Assessment.[18] There is one (0.2%) outlier, residue V492. Figures were generated with Pymol.[19] Atomic coordinates and experimental structure factors are available in the RCSB protein data bank and are accessible with the codes 5UGQ (Hip1SeMet), 5UNO (Hip1wt), and 5UOH (Hip1T466A).

RESULTS AND DISCUSSION

Overall Structure and Topology of Hip1

Mtb *Rv2224c* encodes a 520 residue protein, Hip1, with a molecular weight of 55 kDa.[20] We cloned and purified an N-terminal 6XHis epitope tagged fusion protein that excludes the first 49 amino acids, which constitute a type II signal sequence (His_6Hip1Δ49). The protein was expressed in the presence of selenomethionine in order to incorporate selenium with the intent to derive phase information from the anomalous scattering of the selenium atoms. It is important to note that the selenium had no effect on enzyme activity or protein stability as the Hip1SeMet protein behaved the same as wild-type Hip1 in a thermostability assay (Figure 1) and protease assays (Figure 2). The resulting SeMet structure was used as a model to solve the native (wild-type His_6Hip1Δ49) structure by molecular replacement. In both structures, the N-terminal 6× His epitope tag was disordered. Therefore, the final model of Hip1 contains the residues 50–520, numbered according to the full-length protein sequence. The models for SeMet and wild-type Hip1 protein are essentially identical (rmsd of Cα atoms = 0.2 Å); the following discussion is focused on the wild-type Hip1 structure.

Hip1 crystallized with one molecule in the asymmetric unit. The Hip1 protein is kidney-shaped, with the approximate dimensions 68 Å × 48 Å × 42 Å (Figure 3A). Full-length Hip1 contains 11 cysteine residues and therefore has the potential to form 5 disulfide bonds. We had previously encountered difficulties with refolding of Hip1 in our initial efforts to obtain the active Hip1 protease, and this difficulty has been attributed to the problem of forming these disulfide bridges correctly. The electron density shows clearly that 5 disulfide bonds are formed: between residues Cys55–Cys70, Cys153–Cys189, Cys282–Cys288, Cys393–Cys433, and Cys499–Cys520. (The disulfide bonds are shown in black in Figure 3A.) The cleft of the kidney-shaped Hip1 separates the protein into two domains that we distinguish by their secondary structure characteristics: one domain is a mixed alpha/beta (α/β) domain and the other is an alpha (α) only domain (Figure 3B). The α/β-domain that includes residues 50–140, 203–251, and 456–520 consists of a β-sheet comprising eight β-strands

that are all parallel with the exception of antiparallel β-strand 2. This core β-sheet is surrounded by seven α-helices, three on one side of the β-sheet and four on the other side. The sheet is twisted such that the first and last strands are oriented at right angles to each other. Structure based similarity searches (Dali Server[23]) of the α/β domain resulted in an extensive list of proteins all containing structural features reminiscent of α/β hydrolase fold enzymes[11,24] such as the proline iminopeptidases (PIPs)/propyl aminopeptidases (PAPs).

The α-domain comprised a cluster of α-helices arising from the α/β-domain in two sections (Figure 3B). The first insert (annotated as Chain B in Figure 3B) is composed of residues 260–435 that connect β-strands 6 and 7, forming a series of nine α-helices (α8- α16). Other members of the α/β hydrolase family have this insert, often referred to as the cap or lid, although it varies greatly in size and make up.[11,24] The second insert (Chain A), comprising residues 155–201 that connects β-strand 3 and α-helix 5, contains two α-helices (α3–α4). This feature is unique to Hip1 (Figure 4).

Structure-based similarity searches (Dali Server[23]) of the α-domain only resulted in one hit, dipeptidyl peptidase 2 (PDB 3n0t), with poor alignment (Z score = 1.8; rmsd = 5.6 Å). Like Hip1, dipeptidyl peptidase 2 (DPP2) also contains two domains: an α/β domain with an α/β hydrolase fold and an α-domain comprising nine α-helices connecting β strands 6 and 7 (Chain B). The nine α-helices are arranged differently and are of different lengths than those in Hip1. In another notable difference, DPP2 has a loop (residues 106–129) with just a small, five-residue, α-helix in place of Chain A of Hip1 (Figure 4).

Residues from the helices of the α-domain-Chain A form a pocket that contains unmodeled electron density as seen in both Hip1 SeMet and Hip1 wt models. Mainly hydrophobic residues, including Leu162, Tyr170, Ile178, Phe185, and Val389, as well as residues Gln167, Gly385, Pro429, and Thr432, line this pocket and therefore would most likely bind a hydrophobic molecule. Since Hip1 is a cell wall-associated lipoprotein, it is possible that a hydrocarbon chain could be modeled into the density. Hip1 was unfolded, then washed and refolded bound to Ni-NTA resin, so it is unlikely that this unknown bound molecule was carried over from the purification. Therefore, we looked to the crystallization conditions and modeled a small molecular weight polyethylene glycol (PEG) molecule. Tetra-ethylene glycol fit the density nicely, although possibly a longer PEG chain could also be modeled (Figure S2). What binds in the pocket *in vivo*, as well as the function of this domain, remains to be determined.

Predicted Active Site

Catalytic residues of the protein were previously predicted by sequence alignment of the *hip1* gene product to its then available homologues.[25] Ser228 is located within a GXSXG motif, a conserved region among α/β hydrolase family members. Residues Asp463 and His490 are also conserved and suspected to be part of the catalytic triad. Single point mutations of these residues to alanine abolished esterase activity. Further studies in our laboratory confirmed that Ser228 is part of the catalytic triad and responsible for protease activity.[12]

In the crystal structure, residues Ser228-Asp463-His490 are in close enough proximity to form a hydrogen bond network similar to those described for catalytic triads of serine proteases.[26] These residues are located in the α/β-domain within the cleft of the kidney-shaped protein and are at the center of the cavity (Figure 5). The cavity is formed by parts of both the α/β-and α-domains and extends approximately 20 Å deep. The entrance to the cavity is lined with hydrophilic residues (Glu113, Glu117, Gln124, Thr125, Ser343, Asn345, Arg374, Asn382, and Gln495) with a hydrophobic pocket located at the upper part of the cavity (Met339, Tyr342, Leu346, Met 371, and Try 372). Close to the catalytic triad are three loops forming an active site pocket. One loop contains the catalytic residue Asp463, as well as residues Ala465 and Thr466. The other two loops are made up mainly of aliphatic residues; the first loop with residues Gly252 and Val254 lies at the bottom of the pocket. The second loop, on the opposite side of the pocket to the first loop, contains residues Gly109 and Gly110.

As previously mentioned, the eponymous serine of this serine protease, Ser228, is contained within the conserved sequence, GXSXG. More specifically, sequence alignment of Hip1 with tripetidyl peptidases (TPPs) and proline iminopeptidases (PIPs) shows a consensus sequence of GXSBG where X is any residue, and B is a bulky aromatic residue (Tyr229 for Hip1) (Figure S1). Structurally, Ser228 is located in a tight turn between the β-strand 5 and α-helix 7, referred to as the nucleophilic elbow. Another common feature of the α/β hydrolase family is that the Ser228 is in a strained position, placing the side chain in a "generously allowed" region according to Ramachandran angles ($\varphi = 61.71$, $\psi = -111.74$). We have observed previously in many other enzymes that residues involved in catalysis frequently have conformation torsion angles indicative of strain.[27] Furthermore, Ser228 is exposed to the bulk solvent and is spatially flanked by Gly109–Gly110 and Gly252.

Hip1 residues Gly109 and Gly110 are also in a conserved region, GGPG (Figure S1). In PIPs, the backbone amides from the second Gly in the GGPG sequence and the bulky aromatic residue B in the GXSBG sequence have been shown to form the oxyanion hole, a small pocket that stabilizes the tetrahedral intermediate in the catalytic mechanism by interacting with the carbonyl oxygen of the scissile residue. The Pro111 residue within this conserved sequence (GGPG) is in a cis conformation to allow the NH group of Gly110 to be positioned in a manner necessary to form this oxyanion hole, as seen also in PIPs. Hip1 residues Tyr229 and Gly110 are in the same spatial position and overlay nicely with the oxyanion binding site residues of PIPs and PAPs. Therefore, we conclude that Gly110 and Tyr229 are involved in forming the oxyanion hole in Hip1.

Like Ser228, residue His490 is also exposed to the bulk solvent. It is located in a loop between β-strand 8 and α-helix 18 and is spatially flanked by the two threonine residues, Thr466 and Thr491. Importantly, the imidazole ring is perpendicular to the hydroxyl group of Ser228 and is therefore in a position to abstract a proton to assist the nucleophilicity of the catalytic serine. In contrast, the Asp463 side chain is approximately 60° to the His490 side chain; however, both carboxylate oxygen atoms of Asp463 are within hydrogen bonding distance to Nδ1 of His490, 3.0 and 2.8 Å, respectively, to allow for proton transfer. Although Thr466 is not part of the traditional catalytic triad, it is located on the same loop as Asp463, in between β strand 7 and α helix 17 and is close enough to form hydrogen bonds with

either (or both) His490 or Asp463 at 3.2 and 2.9 Å, respectively (Figure 5). The effect of Thr466 on the enzymatic activity of Hip1 will be discussed later.

Structural Comparison to PIPs/PAPs

Although proline iminopeptidase (PIPs) and propyl aminopeptidases (PAPs) show low sequence identity to Hip1 (Seq ID approximately 20%), a Dali search showed high structural similarity in the α/β-domain (rmsd approximately 2.0 Å). The highest statistics (Z score = 15.4 ; rmsd = 2.0 Å) were for Tricorn-interacting aminopeptidase F1. The eight β-strands of the α/β-domains align and are nearly identical in length. The β-sheet is sandwiched between six α-helices, four on one side and two on the other. However, the α-domains are strikingly different. Although both proteins have a series of α-helices forming the α-domains, that of F1 is much smaller in size with large segments missing relative to Hip1. In particular, two α-helices (α3–α4) from Chain A of Hip1 do not appear in the F1 structure. (Figure S3)

The active sites of these proteases and Hip1 also show some significant differences. Fl has a small opening to the active site and a small active site cavity, with a network of tunnels allowing for the access of small peptides and an exit tunnel for the cleaved amino acids.[28] Other PIPs and PAPs, such as XCIP and SPAP, have larger openings to the active site but a rather small cavity that allows for small dipeptides to enter and exit.[29,30] In contrast, Hip1 has a rather large opening to access the active site as well as a deep cavity to accommodate larger peptides. These structural features are consistent with the fact that PIPs, PAPs, and F1 are exopeptidases, whereas our data indicate that Hip1 may be an endopeptidase.[31] The large active site region of Hip1 can accommodate not only a single strand of a longer peptide but could also interact with the complex structure of a protein surface.

The catalytic residues of F1 were shown to adopt two forms, an “inactive” form and an “active” form upon covalently binding the irreversible inhibitor phenylalanyl chloromethylketone (PCK). In the inactive form, the oxygen (O-γ) from the catalytic Ser105 resides in the oxyanion pocket preventing interaction with His271. Binding of the inhibitor, PCK, displaced the oxygen of Ser105 allowing the former carbonyl oxygen of PCK to occupy the oxyanion pocket. The oxygen of Ser105 was moved toward the nitrogen (N-ε) of His271, which also is reoriented, thereby reducing the distance of the N-ε-O-γ hydrogen bond from 3.9 to 2.8 Å. Hence binding of the inhibitor causes a conformational change of the catalytic residues of F1 to distances more optimal for protease activity.[28] The Hip1 catalytic residues, Ser228 and His460, overlay nicely with the catalytic residues of the active bound form of F1 (Figure 6A). However, one difference is the slight rotation of the catalytic Asp463 residue of Hip1 in reference to the catalytic Asp244 residue of F1. Still, the Hip1 Asp463 residue does align with the catalytic Asp in other PIPs: XCPIP and SPAP (Figure 6B). Therefore, this slight rotation of the Asp residue should allow for protease activity to occur and is not unique to Hip1.

The catalytic Asp residue for Hip1 and F1, as well as XCPIP and SPAP, is located in a loop 5–7 amino acids in length. For Hip1, this loop also contains Thr466, which places the residue in close proximity to interact with the catalytic residues histidine and aspartic acid, 3.2 and 2.9 Å, respectively. Since Thr can also act as an active site nucleophile, we have looked at this residue more carefully in comparison to other serine proteases in this family.

In F1, this residue is Val246 which cannot form hydrogen bonds and is in the same position spatially as Thr466 of Hip1 with the side chain turned away from the His and Asp. In XCPIP and SPAP, a Cys residue located in the same position of each structure was thought to be the active site nucleophile. Specifically, a decrease in protease activity of XCPIP and SPAP is observed with cysteine protease inhibitors that exclusively interact with the thiol group of cysteines. These cysteines have since been shown to be important but not necessary for protease activity. Although the Cys residue is located in the active site, it is approximately 4.5–5 Å away from the residues of the catalytic triad. More importantly, mutagenesis studies showed that the Ser residue and not the Cys residue is necessary for activity.[32] Likewise, we have previously shown by mutagenesis that Ser228 is critical for cleavage of GroEL2.

Characterization of Hip1 Mutant T466A

We were interested in testing the effect residue Thr466 has on the activity and stability of Hip1. To accomplish this task, we made a single mutation of Hip1 residue 466 from threonine to alanine to produce Hip1T466A. We purified Hip1wt alongside of Hip1T466A to have a direct comparison of the two proteins. The chromatograms of the MonoQ (anion exchange) column, which is the final step in purification, are practically identical (Figure S4A). Also, the circular dichroism (CD) spectra of the two proteins are similar indicating that the two proteins have the same secondary structure (Figure S4B). However, according to thermostability assays, the melting temperature for Hip1T446A is approximately 2 °C lower than that for Hip1wt (46.6 °C vs 44.8 °C) (Figure 1). Other mutations and modifications of Hip1, such as Hip1S228A (inactive mutant) and Hip1SeMet, have little (46.9 °C for Hip1SeMet) or even a positive effect (47.8 °C for Hip1S228A) on protein stability.

Next, we wanted to determine the activity of Hip1T446A mutant as a protease. First, we tested azocasein, a general protease substrate, and detected a small decrease in activity (approximately 15–25%) (Figure 2A). Then, we tested Hip1T446A against GroEl2, Hip1's cognate substrate, and found that cleavage of GroEL2 was almost completely abolished (Figure 2B). It is important to note that in both of these assays, Hip1S228A and Hip1SeMet, were also tested and behaved as expected with Hip1S228A completely inactive and Hip1SeMet comparable to Hip1wt. Also, we tested Hip1T466A in the same pH range in which Hip1wt is active (pH 5.4–7.4), and the result was still that Hip1T466A is practically inactive (data not shown). These data suggest Hip1's ability to cleave GroEl2 is very specific and that residue Thr466 is an important player in its ability to act upon GroEL2.

To gain insight as to how residue T466 may help in the cleavage of GroEl2, we crystallized the mutant under the same conditions as Hip1wt. The Hip1wt structure was used as a model to solve the Hip1T466A structure by molecular replacement. The model for Hip1wt and Hip1T466A protein are nearly identical (rmsd of Cα atoms = 0.12 Å). Closer examination of the active site confirms that there are no differences in the positioning of the catalytic triad. At least under crystallographic conditions, there are no differences that can account for the role of residue T466.

One possibility may be the stability of the protein. As mentioned previously, Hip1T466A is less stable than Hip1wt. In addition, the mutation of the threonine to alanine does not alter the conformations of the active site residues, and therefore, we expect that the mutant

enzyme will still hydrolyze a nonspecific peptide such as azocasein. These results suggest that Thr466 may be needed to maintain a competent conformation of the catalytic triad after binding to GroEL2.

Kinetic Studies of Hip1wt and GroEL2

On the basis of the structure and the positioning of the catalytic residues, one would suspect Hip1 to be a robust serine protease. To determine the efficiency of Hip1 at cleaving GroEL2, we studied the enzyme reaction over time. GroEL2 was incubated at 37 °C with Hip1wt or Hip1T466A and cleavage monitored by stopping the reaction at several time points over a course of 25 h (Figure 7A). Cleavage of GroEL2 by Hip1wt was observed even at 2 h and was 50% complete by 6 h, whereas no cleavage occurred with the Hip1T466A mutant.

To obtain k_{cat} and K_m for the proteolysis of GroEL2 by Hip1, the reaction was again monitored over time, and time points were taken every 10 or 20 min to observe the initial velocity. Greater concentrations of GroEL2 protein were used (10 to 200 μM) to saturate the Hip1 enzyme. Also, in order to obtain reproducible kinetic parameters, it is necessary to preincubate the enzyme at 37 °C as has been observed by others.[33] For Hip1wt protein, k_{cat}, K_m, and k_{cat}/K_m were calculated to be 0.45 ± 0.27 min^{-1}, 1.9 × 10^{-4} ± 1.6 × 10^{-4} M, and 2.4 × 10^3 $M^{-1}min^{-1}$, respectively. However, under these conditions Hip1T466A was still inactive. Although in the 3D structure the alignment and distances of the catalytic residues are in accordance with other serine proteases, the activity against its native protein substrate indicates that Hip1 is a relatively inefficient enzyme compared to the standard serine proteases. Other proteases have been studied that are also inefficient except with their specific cognate substrates.[34,35]

Specificity of Active Site

Previous work from our groups showed that Hip1 cleaves GroEL2 within the N-terminus of the protein (between Arg_{12} and Gly_{13}) releasing a peptide fragment of 12 amino acid residues.[12] Therefore, we wanted to determine if Hip1 could cleave a small peptide or peptidomimetic containing an arginine residue in the P1 position. We chose the simplest substrate BApNA (Nα-benzoyl-D,L-arginine 4-nitroanilide) to compare Hip1wt to the Hip1T466A mutant. BApNA was incubated with Hip1wt or Hip1T466A at 37 °C and monitored continuously for 2 h. Hip1T466A was not active in this assay. Hip1wt, however, was active and exhibited hydrolysis at 1 mM BApNA concentration (Figure 7B) reproducibly. The substrate was insoluble at higher concentrations, and therefore, we are unable to calculate k_{cat} or K_m for the reaction.

We have seen that Hip1 is able to process substrates with arginine in the P1 position. However, results from a recent publication differ from our findings. Screening of combinatorial libraries of small peptides was used to profile substrate specificity of Hip1 and resulted in the observation of a positively charged residue being favored in the P2 position of the peptide substrate.[33] These differences may lie in the binding of a large protein substrate to the enzyme versus small peptides.

Putative S1 Binding Pocket from Peptide Model

Substrates for Hip1 range from a large protein, GroEL2, to a small molecule BApNA, both containing an arginine residue at the cleavage site. Therefore, the Hip1 S1 pocket should be able to accommodate a large positively charged residue in the P1 position. In comparison, F1, PIPs, and PAPs can specifically cleave after a Pro residue or in some cases, to a lesser extent, Ala, but are unable to cleave after an Arg or Asp residue. The S1 binding pocket of XCPIP and SPAP, formed by large hydrophobic residues, is small and can accommodate Ala or Pro residues. The S1 pocket of F1 is also small (8 × 10 × 12 Å) and comprises the residues from both domains.

To predict the S1 pocket of Hip1, we modeled a dipeptide into the active site of Hip1 based on an overlay of the Hip1 model and the model of F1 with a bound peptide (Figure 8). Using the positions of the oxyanion hole, the S1 site, and the Phe-Leu peptide that is observed in the structure of F1,[36] we modeled Arg-Gly into the active site of Hip1 by substituting the Leu residue with Gly and the Phe residue with an Arg and selecting the most common rotamers of the Arg side chain without change to main chain configuration. As a result, the S1 pocket of Hip1 is clearly defined and comprises residues from the α-domain, including Pro260, Ala263, Glu264, and Gln267. The large pocket of Hip1 can clearly accommodate a large side chain such as Arg in the P1 site. In addition, residue Glu264 is in a prime position to interact with the Arg side chain, possibly forming a salt-bridge.

Structural Comparison to Other Mtb Proteases

The structure reported here contributes new insights into Mtb proteases. Structures for a number of other proteases from Mtb that are important for Mtb function in the host cell have been reported previously. For instance, Rv3671c, a putative serine protease, is crucial for the persistence of Mtb in the hostile environment of the phagosome.[37] Another is Mycosin-1, one of five proteases essential for secretion and shown to be involved in processing of type VII-exported proteins. The secretion systems in which these proteases function are particularly important to disease progression.[38] The structure of a ClpP protease from Mtb has recently been reported. These proteases are important for protein degradation, highlighted by the fact that genes for *clP1* and *clpP2* are essential and that mechanism-based inhibitors have been shown to suppress growth.[39] Although all of these proteases are related to the function of the bacterium in the host environment, none of these proteases are similar to each other structurally, nor is Hip1 related to these others. Therefore, Hip1 is a unique protease with novel function.

Summary

Mycobacterium tuberculosis (Mtb) protein Hip1 is a serine protease with limited similarity to the *Streptomyces lividans* tripeptidyl-peptidases, TPPs (approximately 30% seq ID), and proline iminopeptidases (PIPs; also known as prolyl aminopepidases, PAPs); such as XCPIP (24% seq ID) and SPAP (21%). Although no structures are available for the *Streptomyces lividans* TPPs for comparison, structures for PAPs and PIPs are known and have many features similar to those of Hip1. Nevertheless, Hip1 has features that are distinct from these other proteases and serve to specify its function.

There are four main structural features common to Hip1, F1, and PIPs/PAPs: (1) an α/β hydrolase fold as seen in α/β hydrolase enzymes, (2) two distinct domains, (3) active site residues in the same spatial position, and (4) residues Tyr229 and Gly110 that overlay nicely with residues that are known to form the oxyanion hole.

Despite the similarities mentioned above, the structure of Hip1 can account for the differences in specificity within these proteins. PIPs and PAPs are specific for cleavage after Pro or Ala residues from small two residue peptide amide substrates. F1, also specific for cleavage after Pro or Ala, with limited access to the active site due to the small opening, also cleaves small peptides. As exopeptidases, PIPs, PAPs, and F1 all possess two Glu residues within the active site that are shown to be involved in binding the imino group of a substrate's N-terminal Pro. Not surprisingly, there are no Glu residues in the same spatial position for Hip1 consistent with its role as an endopeptidase. Instead, a large opening leading to the active site, as well as a deep cavity, can account for Hip1's ability to accommodate larger peptide chains and to interact with a protein surface. The deep cavity, as well as a Glu residue (E264) in the putative S1 pocket, can potentially recognize a positively charged side chain in the P1 position. This is consistent with our previous findings that Hip1 exhibits endopeptidase activity, cleaving the peptide bond at a positively charged residue.

Another unique feature of Hip1 is the placement of residue Thr466 in close proximity to the Asp463 and His490 of the catalytic triad. This residue seems to be important for function as mutation of the threonine to alanine nearly abolished cleavage of GroEL2. Although the structure of the protein remained unchanged according to CD and a model obtained from crystallography, thermostability assays indicate the HipT466A mutant is somewhat less stable than the wild type.

Overall, the model of Hip1 will be very useful for further studies. First, modeling of the Hip1 protein may help elucidate the structures of proteins within S33 serine protease family not yet known, including the *Streptomyces lividans* TPPs. Second, we will use the Hip1 protein model to perform studies to gain a greater understanding as to the specificity of Hip1 and more importantly its selectivity toward the Mtb protein GroEL2 including the role of residue Thr466. Third, of particular interest is the use of the model for drug design, as Hip1 has been shown to be an important immunomodulatory protein. Blocking Hip1 activity is attractive for designing adjunctive immune therapeutics that could synergize with anti-TB antibiotics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by funds from National Institutes of Health grants R00TW008043 and 5R01AI083366 (to J.R.), GM 32415 (to G.A.P. and D.R.), and R37AI28571 (to B.M.D.).

We thank Jared Auclair for performing mass spectroscopy to validate the incorporation of selenium in the protein. We also thank Cheryl Kreinbring, Quyen Hoang, and CeFeng Liu for their help with figure preparation as well as

members of the Rengarajan, Dunn, Petsko, and Ringe laboratories for helpful discussions and insights. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. This research also used resources of the National Synchrotron Light Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-AC02-98CH10886.

References

1. Global Tuberculosis Report 2015. World Health Organization (WHO); Geneva, Switzerland: 2015.
2. Ehrt S, Schnappinger D. Mycobacterial survival strategies in the phagosome: defence against host stresses. Cell Microbiol. 2009; 11:1170–1178. [PubMed: 19438516]
3. Russell DG. Mycobacterium tuberculosis: here today, and here tomorrow. Nat Rev Mol Cell Biol. 2001; 2:569–586. [PubMed: 11483990]
4. Rengarajan J, Bloom BR, Rubin EJ. Genome-wide requirements for Mycobacterium tuberculosis adaptation and survival in macrophages. Proc Natl Acad Sci U S A. 2005; 102:8327–8332. [PubMed: 15928073]
5. Madan-Lala R, Peixoto KV, Re F, Rengarajan J. Mycobacterium tuberculosis Hip1 Dampens Macrophage Proinflammatory Responses by Limiting Toll-Like Receptor 2 Activation. Infect Immun. 2011; 79:4828–4838. [PubMed: 21947769]
6. Madan-Lala R, Sia JK, King R, Adekambi T, Monin L, Khader SA, Pulendran B, Rengarajan J. Mycobacterium tuberculosis Impairs Dendritic Cell Functions through the Serine Hydrolase Hip1. J Immunol. 2014; 192:4263–4272. [PubMed: 24659689]
7. Rengarajan J, Murphy E, Park A, Krone CL, Hett EC, Bloom BR, Glimcher LH, Rubin EJ. Mycobacterium tuberculosis Rv2224c modulates innate immune responses. Proc Natl Acad Sci U S A. 2008; 105:264–269. [PubMed: 18172199]
8. Sutcliffe IC, Harrington DJ. Lipoproteins of *Mycobacterium tuberculosis*: an abundant and functionally diverse class of cell envelope components. FEMS Microbiol Rev. 2004; 28:645–659. [PubMed: 15539077]
9. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2014; 42:D503–D509. [PubMed: 24157837]
10. Laskar A, Chatterjee A, Chatterjee S, Rodger EJ. Three-Dimensional Molecular Modeling of a Diverse Range of SC Clan Serine Proteases. Mol Biol Int. 2012; 2012:1–9.
11. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, et al. The α/β hydrolase fold. Protein Eng, Des Sel. 1992; 5:197–211.
12. Naffin-Olivos JL, Georgieva M, Goldfarb N, Madan-Lala R, Dong L, Bizzell E, Valinetz E, Brandt GS, Yu S, Shabashvili DE, Ringe D, Dunn BM, Petsko GA, Rengarajan J. *Mycobacterium tuberculosis* Hip1 modulates macrophage responses through proteolysis of GroEL2. PLoS Pathog. 2014; 10:e1004132. [PubMed: 24830429]
13. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun. 1967; 27:157–162. [PubMed: 6035483]
14. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem. 2006; 357:289–298. [PubMed: 16962548]
15. Otwinowski, Z., MInor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In: Carter, CW., Jr, Sweet, RM., editors. Methods in Enzymology. Academic Press; New York: 1997. p. 307-326.
16. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr, Sect D: Biol Crystallogr. 2002; 58:1948–1954. [PubMed: 12393927]
17. Emsley P, Cowtan K. *Coot*: model-building tools for molecular graphics. Acta Crystallogr, Sect D: Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
18. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins: Struct, Funct, Genet. 2003; 50:437–450. [PubMed: 12557186]

19. The Pymol Molecular Graphics System, version 1.3. Schrödinger, LLC; New York:
20. Lew JM, Kapopoulou A, Jones LM, Cole ST. TubercuList – 10 years after. Tuberculosis. 2011; 91:1–7. [PubMed: 20980199]
21. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM. PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends Biochem Sci. 1997; 22:488–490. [PubMed: 9433130]
22. de Beer TAP, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic Acids Res. 2014; 42:D292–D296. [PubMed: 24153109]
23. Holm L, Rosenström P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38:W545–W549. [PubMed: 20457744]
24. Holmquist M. Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. Curr Protein Pept Sci. 2000; 1:209–235. [PubMed: 12369917]
25. Lun S, Bishai WR. Characterization of a Novel Cell Wall-anchored Protein with Carboxylesterase Activity Required for Virulence in Mycobacterium tuberculosis. J Biol Chem. 2007; 282:18348–18356. [PubMed: 17428787]
26. Torrance JW, Holliday GL, Mitchell JBO, Thornton JM. The Geometry of Interactions between Catalytic Residues and their Substrates. J Mol Biol. 2007; 369:1140–1152. [PubMed: 17466330]
27. Stojanoski V, Adamski CJ, Hu L, Mehta SC, Sankaran B, Zwart P, Prasad BVV, Palzkill T. Removal of the Side Chain at the Active-Site Serine by a Glycine Substitution Increases the Stability of a Wide Range of Serine β-Lactamases by Relieving Steric Strain. Biochemistry. 2016; 55:2479–2490. [PubMed: 27073009]
28. Goettig P, Groll M, Kim J-S, Huber R, Brandstetter H. Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. EMBO J. 2002; 21:5343–5352. [PubMed: 12374735]
29. Medrano FJ, Alonso J, García JL, Romero A, Bode W, Gomis-Rüth FX. Structure of proline iminopeptidase from Xanthomonas campestris pv. citri: a prototype for the prolyl oligopeptidase family. EMBO J. 1998; 17:1–9. [PubMed: 9427736]
30. Yoshimoto T, Kabashima T, Uchikawa K, Inoue T, Tanaka N, Nakamura RT, Tsuru M, Ito E. Crystal Structure of Prolyl Aminopeptidase from Serratia marcescens. J Biochem. 1999; 126:559–565. [PubMed: 10467172]
31. Barrett, AJ. [1] Classification of Peptidases. In: Barrett, AJ., editor. Methods in Enzymology. Vol. 244. Academic Press; San Diego, California: 1994. p. 1-15.
32. Kitazono A, Ito K, Yoshimoto T. Prolyl Aminopeptidase Is Not a Sulfhydryl Enzyme: Identification of the Active Serine Residue by Site-Directed Mutagenesis. J Biochem. 1994; 116:943–945. [PubMed: 7896753]
33. Lentz CS, Ordonez AA, Kasperkiewicz P, La Greca F, O'Donoghue AJ, Schulze CJ, Powers JC, Craik CS, Drag M, Jain SK, Bogyo M. Design of Selective Substrates and Activity-Based Probes for Hydrolase Important for Pathogenesis 1 (HIP1) from *Mycobacterium tuberculosis*. ACS Infect Dis. 2016; 2:807–815. [PubMed: 27739665]
34. Taylor FR, Bixler SA, Budman JI, Wen D, Karpusas M, Ryan ST, Jaworski GJ, Safari-Fard A, Pollard S, Whitty A. Induced Fit Activation Mechanism of the Exceptionally Specific Serine Protease, Complement †. Biochemistry. 1999; 38:2849–2859. [PubMed: 10052957]
35. Das C, Hoang QQ, Kreinbring CA, Luchansky SJ, Meray RK, Ray SS, Lansbury PT, Ringe D, Petsko GA. Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. Proc Natl Acad Sci U S A. 2006; 103:4675–4680. [PubMed: 16537382]
36. Goettig P, Brandstetter H, Groll M, Göhring W, Konarev PV, Svergun DI, Huber R, Kim J-S. X-ray Snapshots of Peptide Processing in Mutants of Tricorn-interacting Factor F1 from Thermoplasma acidophilum. J Biol Chem. 2005; 280:33387–33396. [PubMed: 15994304]
37. Biswas T, Small J, Vandal O, Odaira T, Deng H, Ehrt S, Tsodikov OV. Structural Insight into Serine Protease Rv3671c that Protects M. tuberculosis from Oxidative and Acidic Stress. Structure. 2010; 18:1353–1363. [PubMed: 20947023]

38. Solomonson M, Huesgen PF, Wasney GA, Watanabe N, Gruninger RJ, Prehna G, Overall CM, Strynadka NCJ. Structure of the Mycosin-1 Protease from the Mycobacterial ESX-1 Protein Type VII Secretion System. J Biol Chem. 2013; 288:17782–17790. [PubMed: 23620593]

39. Schmitz KR, Carney DW, Sello JK, Sauer RT. Crystal structure of Mycobacterium tuberculosis ClpP1P2 suggests a model for peptidase activation by AAA+ partner binding and substrate delivery. Proc Natl Acad Sci U S A. 2014; 111:E4587–E4595. [PubMed: 25267638]

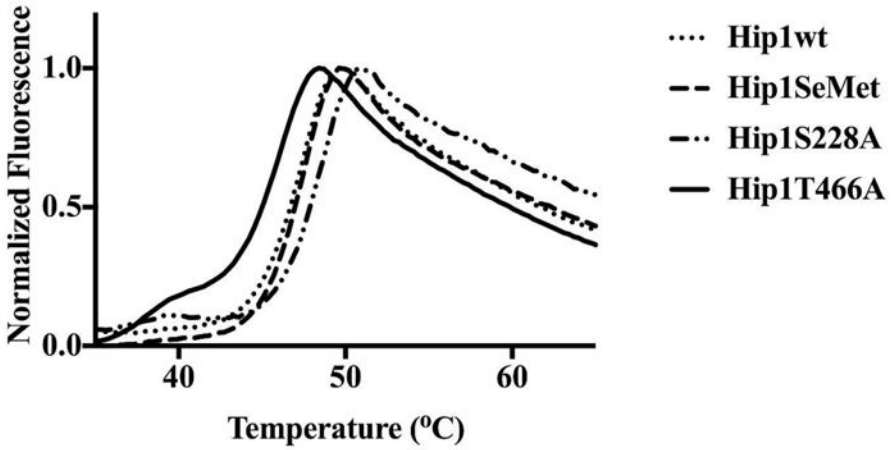


Figure 1.

Thermo-stability assay of Hip1 proteins. See Materials and Methods for assay conditions.

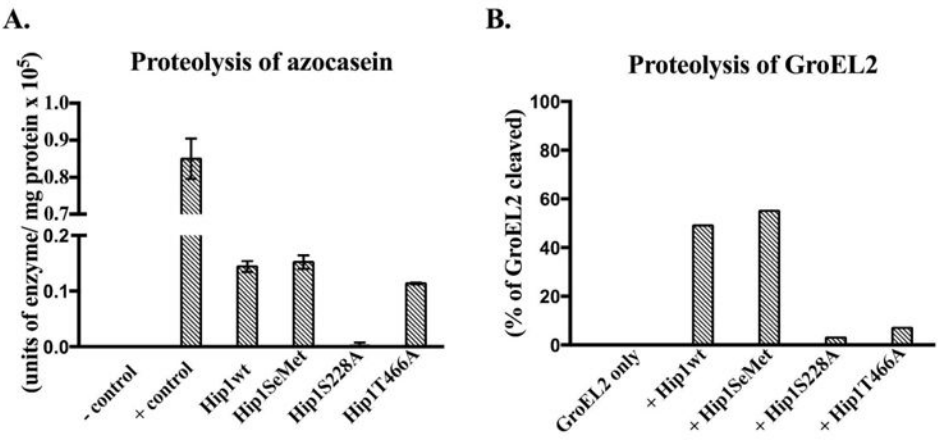


Figure 2.

Protease activity of Hip1wt, Hip1SeMet, Hip1S228A, and Hip1T466A proteins observed using two different protein substrates, azocasein and GroEL2. (A) Bovine serum albumin (BSA) and protease Subtilisin Carlsberg were used as a negative and positive control, respectively. (B) The GroEL2-only sample contained no cleaved product; therefore, this band was set to 100% and used to normalize the samples containing Hip1 proteins. The graph represents the precentage of cleaved GroEl2 product formed.

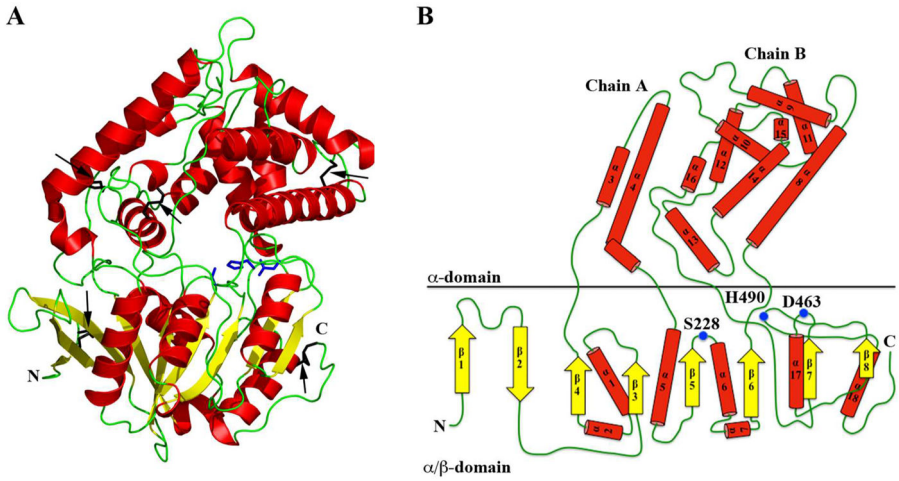


Figure 3.

Overall structure of the Hip1 monomer. (A) Ribbon diagram based on Cα's of the final model. Helices are colored red, β-strands are yellow, and the loop regions are green. The catalytic residues, as well as the cysteine residues involved in the formation of disulfide bridges, are represented as sticks and are colored blue and black, respectively. Black arrows are used to indicate the five disulfide bridges. N and C termini are indicated. (B) Topology model based on structural assignments made according to DSSP. The diagram was first generated using PDBSUM[21,22] to determine the length of helices and then redrawn to depict a more accurate representation of the two domains. A solid black line is used to indicate a distinction between the two domains referred to as the α-domain and the α/β-domain. The same color scheme is used as that for the ribbon diagram in part A. The catalytic residues are denoted by the blue circles and are annotated. The N-terminus is on the left side, which is the same orientation relative to the ribbon diagram in part A.

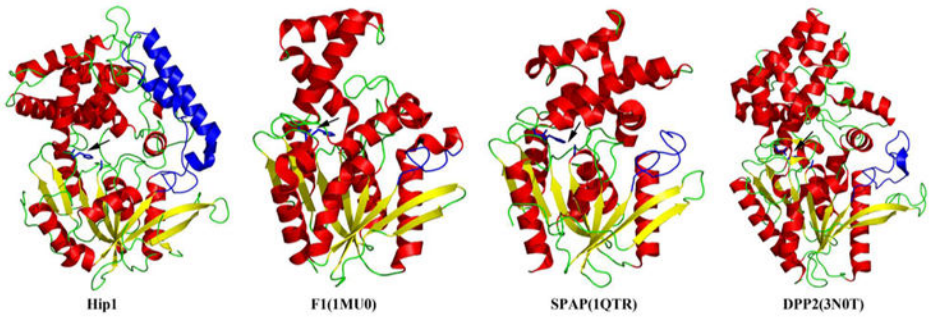


Figure 4.

Comparison of Hip1, F1, SPAP, and DPP2 structures. Ribbon diagrams are shown of Hip1, F1, SPAP, and DPP2 structures to compare the two domains of each protein with emphasis on the different α-domains, in particular, the insert in chain A of Hip1 shown in blue (residues inserted between β-strand 4 and α-helix 5 that form part of the α-domain). Also shown are the corresponding regions of F1 (residues Asp62–Pro72), SPAP (residues Asp70–Asn86), and DPP2 (residues Glu106-L129) colored in dark blue. For all proteins, helices are colored red, β-strands are yellow, the loop regions are green, and the catalytic residues, shown in stick representation, are dark blue. Arrows indicate the locations of the active site in each protein.

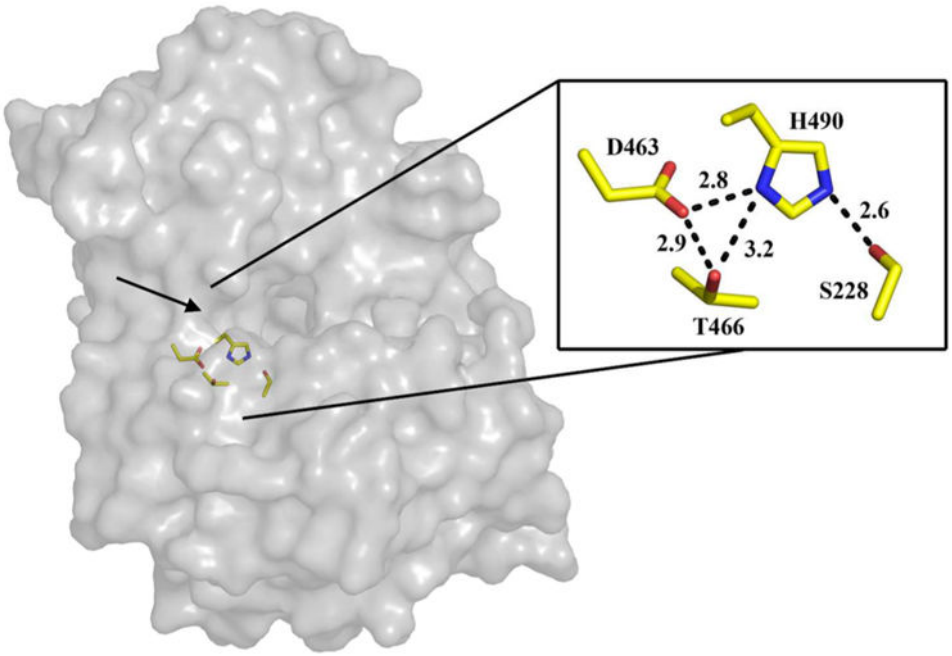


Figure 5.

Surface diagram of the Hip1 model. The active site opening is indicated by an arrow. Catalytic triad residues are colored by element: carbon in yellow, oxygen in red, and nitrogen in blue. Also included is residue Thr466, which is in close proximity spatially to Asp463 and His490. Measurements of the distances (in Å) between potential interactions of the residues are also depicted.

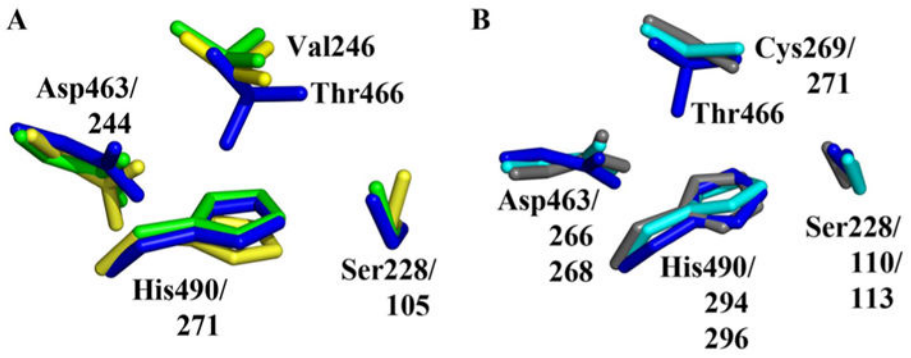


Figure 6.

Overlay of Hip1 with F1, XCPIP, and SPAP active sites. (A) Overlay of Hip1 and F1 catalytic residues. Hip1 residues are colored blue, F1 (active form) residues are green, and F1 (inactive form) residues are yellow. His490 and Ser228 of Hip1 align well with His271 and Ser105 of the active form of F1. Note the Ser105 of the inactive form is turned away from the His271 residue. Also, the His residue has moved slightly in the active form to align with the Ser. The position of the Asp244 residue of F1 does not change upon activation. However, Asp463 of Hip1 is slightly tilted out of line with His490; instead of being perpendicular (90° angle) such as with F1, it is approximately at a 60° angle. (B) Alignment of Hip1, XCPIP, and SPAP catalytic residues. Hip1 residues are colored blue, XCPIP residues are light blue, and SPAP residues are gray.

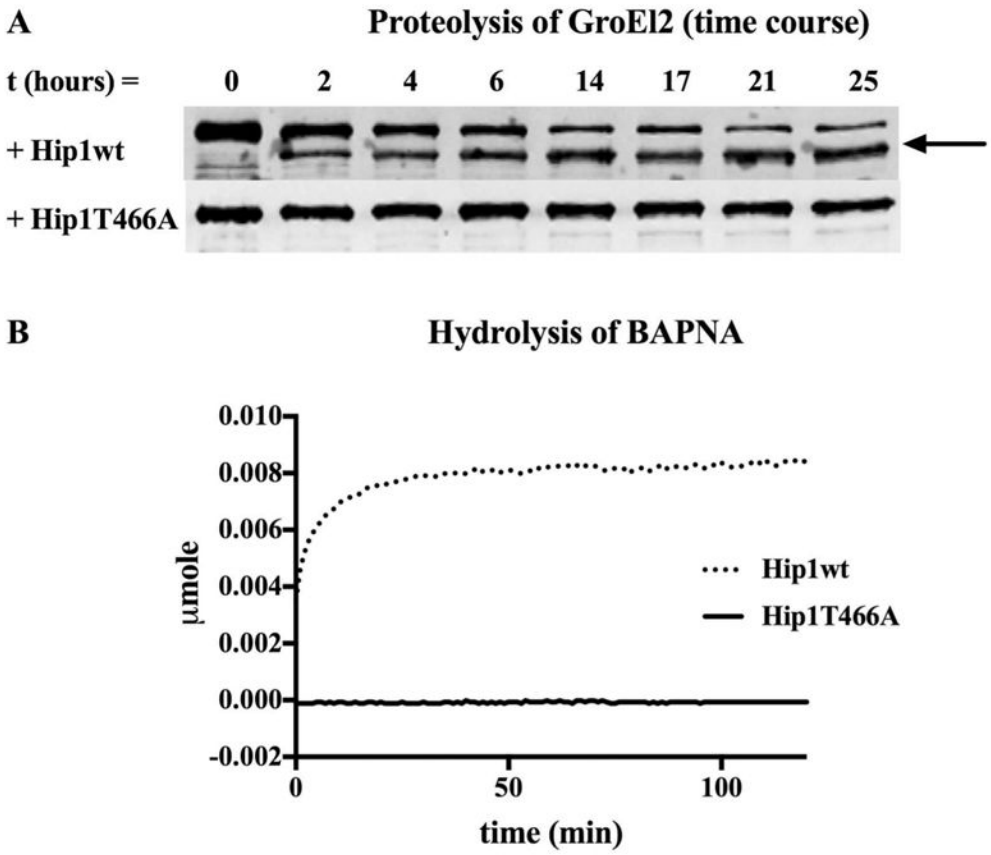


Figure 7.

Kinetic studies of Hip1wt and Hip1T466A. (A) Time course comparing the proteolysis of GroEL2 protein by Hip1wt and Hip1T466A using a Western blot probed with S-tag antibody to observe only the GroEL2 protein. An arrow indicates the cleaved product formed upon addition of Hip1wt. (B) Representative graph of the hydrolysis of the substrate BApNA at 1 mM.

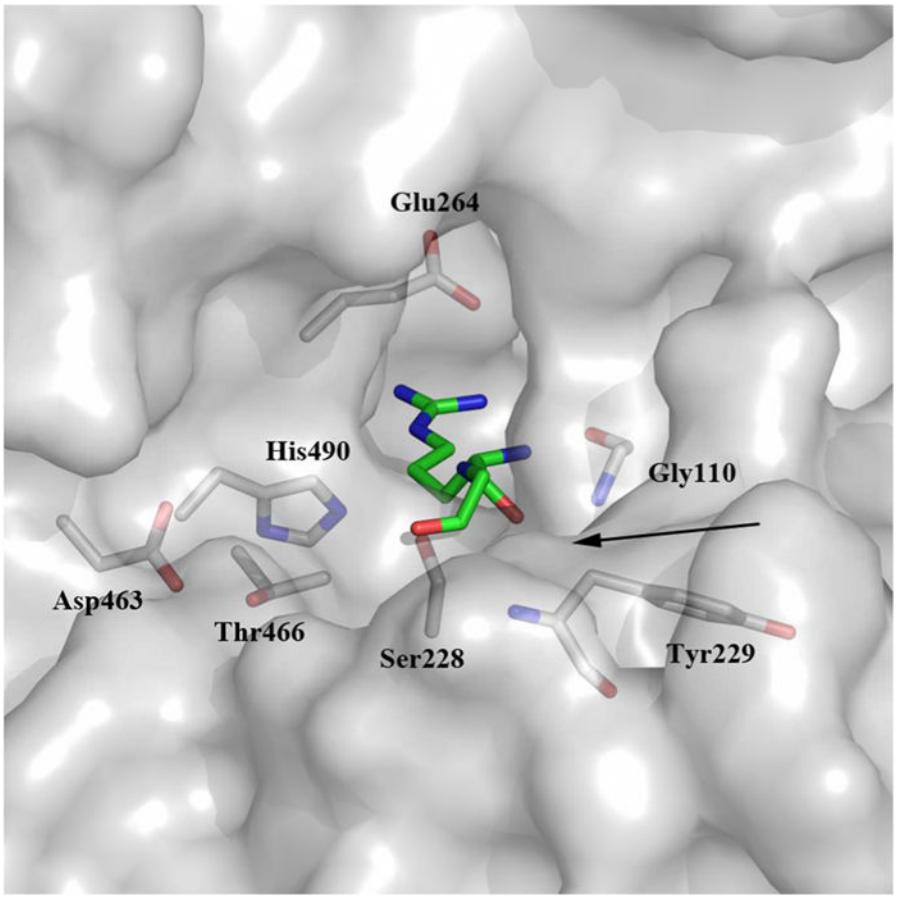


Figure 8.

S1 pocket of Hip1 with a model of the Arg-Gly dipeptide. Surface diagram of Hip1 showing a model of the RG dipeptide in the active site of Hip1 based on an overlay of the Hip1 model and the model of F1 with a bound peptide. The RG dipeptide is colored by element: carbon in green, oxygen in red, and nitrogen in blue. Residues within the active site are represented as sticks and are colored by element: carbon in light gray, oxygen in red, and nitrogen in blue. The putative oxyanion hole is indicated by an arrow and comprises the backbone nitrogen of Tyr229 and Gly110. Residue Glu264, which is suspected to be part of the S1 pocket, is also indicated.