Molecules and Cells

Crystal Structure of Histidine Triad Nucleotide-Binding Protein from the Pathogenic Fungus *Candida albicans*

Ahjin Jung[1,6], Ji-Sook Yun[1,6], Shinae Kim[1], Sang Ryong Kim[2], Minsang Shin[3], Dong Hyung Cho[2], Kwang Shik Choi[2,4,5,*], and Jeong Ho Chang[1,5,*]

[1]Department of Biology Education, [2]School of Life Sciences, BK21 Plus KNU Creative BioResearch Group, [3]Department of Microbiology, School of Medicine, [4]Research Institute for Dokdo and Ulleungdo Island, Kyungpook National University, [5]Research Institute for Phylogenomics and Evolution, Kyungpook National University, Daegu 41566, Korea, [6]These authors contributed equally to this work.
*Correspondence: ksc@knu.ac.kr (KSC); jhcbio@knu.ac.kr (JHC)
http://dx.doi.org/10.14348/molcells.2018.0377
www.molcells.org

Histidine triad nucleotide-binding protein (HINT) is a member of the histidine triad (HIT) superfamily, which has hydrolase activity owing to a histidine triad motif. The HIT superfamily can be divided to five classes with functions in galactose metabolism, DNA repair, and tumor suppression. HINTs are highly conserved from archaea to humans and function as tumor suppressors, translation regulators, and neuropathy inhibitors. Although the structures of HINT proteins from various species have been reported, limited structural information is available for fungal species. Here, to elucidate the structural features and functional diversity of HINTs, we determined the crystal structure of HINT from the pathogenic fungus *Candida albicans* (CaHINT) in complex with zinc ions at a resolution of 2.5 Å. Based on structural comparisons, the monomer of CaHINT overlaid best with HINT protein from the protozoal species *Leishmania major*. Additionally, structural comparisons with human HINT revealed an additional helix at the C-terminus of CaHINT. Interestingly, the extended C-terminal helix interacted with the N-terminal loop (α1-β1) and with the α3 helix, which appeared to stabilize the dimerization of CaHINT. In the C-terminal region, structural and sequence comparisons showed strong relationships among 19 diverse species from archea to humans, suggesting early separation in the course of evolution. Further studies are required to address the functional significance of variations in the C-terminal region. This structural analysis of CaHINT provided important insights into the molecular aspects of evolution within the HIT superfamily.

Keywords: *Candida albicans*, crystal structure, histidine triad, HIT family, HINT

INTRODUCTION

Histidine triad nucleotide-binding protein (HINT), which contains three highly conserved histidines as catalytic residues, is a member of the histidine triad (HIT) protein superfamily (Brenner et al., 1997). Members of the HIT superfamily, which contain a conserved sequence motif (H-X-H-X-H-X-X, where X represents a hydrophobic residue) in the active site, exhibit catalytic activities, including nucleotide hydrolase, phosphoramidate hydrolase, and nucleotidyltransferase activities (Krakowiak et al., 2004). Based on the catalytic specificities, sequence compositions, and structural similarities of its members, this superfamily has been historically grouped into five classes: HINT, galactose-1-phosphate uridylyltrans-

Received 14 September 2018; revised 5 November 2018; accepted 14 November 2018; published online 2 January, 2019

eISSN: 0219-1032

ferase (GalT), fragile HIT protein (Fhit), decapping scavenger (DcpS), and aprataxin (Martin et al., 2011).

Among these five classes, GalT is known to participate in galactose metabolism, where it catalyzes the transfer of UDP from glucose-UDP to galactose (McCorvie and Timson, 2011). Lack of GalT causes hereditary galactosemia, an inborn error of carbohydrate metabolism (Demirbas et al., 2018). Fhit, which is designated as a fragile HIT protein because of the high mutation rate of its gene locus in multiple cancers, has been characterized as a tumor suppressor (Huebner and Croce, 2001). Fhit has phosphoramidase activity and therefore participates in the hydrolysis of dinucleotide polyphosphates (Huang et al., 2004; Varnum et al., 2001). DcpS is known to involved in mRNA decay by specifically cleaving methylated mRNA transcripts (Milac et al., 2014). A distinguishing feature of this protein is its requirement for an additional N-terminal domain, which undergoes a large conformational change to switch between open and closed states during catalysis (Gu and Lima, 2005; Gu et al., 2004; Liu et al., 2002; 2004). Aprataxin participates in DNA repair systems by hydrolyzing both dinucleotide polyphosphates and phosphoramidates (Martin et al., 2011).

HINT belongs to the most ancient class within the HIT superfamily and is well conserved from archaea to eukaryotes. In humans, three isotypes of HINT (termed the hHint family), i.e., Hint1, Hint2, and Hint3, have been identified (Maize et al., 2013). Hint1 functions as a tumor suppressor by inducing apoptosis independent of its enzymatic activity (Weiske and Huber, 2006). Thus, Hint1 also hydrolyzes lysyl-adenylate, which is produced by lysyl-tRNA synthetase, and mutation in Hint1 causes heritable axonal neuropathy with neuromyotonia (Butland et al., 2005; Chou and Wagner, 2007; Zimon et al., 2012). Both hHint2 and hHint3 are upregulated in breast, pancreatic, and colon cancers, similar to hHint1 (Martin et al., 2011). However, the biological roles of HINT proteins from other species have not yet been fully characterized.

The second histidine of the histidine triad in the active site is strictly conserved and participates in catalysis with the third histidine, whereas the first histidine in this motif is not involved in the reaction (Maize et al., 2013). These conserved residues participate in catalysis by forming a covalent nucleotidyl-phosphohistidyl intermediate. HINT proteins also have a conserved zinc-binding motif 'C-X-X-C' (where C is a cysteine residue and X is a hydrophobic residue), and a zinc ion is coordinated by these cysteine residues, together with the first histidine residue (Klein et al., 1998; Lima et al., 1996).

The opportunistic pathogenic fungus *Candida albicans* causes candidiasis (Pfaller and Diekema, 2007). Candidiasis can be classified into two categories depending on the severity of the disease (Spampinato and Leonardi, 2013). The first category of candidiasis causes mucosal infections, including infections of the oral/vaginal cavities and respiratory/intestinal tracts of humans and animals (Meiller et al., 2009). The other type of candidiasis is a severe disease resulting in systemic infections and typically occurs in individuals with immunocompromised status, including human immunodeficiency virus-infected patients, transplant recipients, and chemotherapy patients, who show susceptibility to systemic *Candida* infections (Akter et al., 2018; Schulze and Sonnenborn, 2009). Although mucosal *C. albicans* infections are common and nonlethal (Li et al., 2002; Torosantucci et al., 2002), systemic candidiasis is associated with a high mortality rate (Pfaller and Diekema, 2007) and can cause neutropenia and gastrointestinal mucosal damage.

Multiple structures of HINT proteins have been deposited in the Protein Data Bank (PDB). However, few structural studies of HINT proteins from fungal/yeast species have been reported. In order to expand the molecular evolutionary pool, we determined the crystal structure of HINT from the pathogenic fungal species *C. albicans*. Our findings are expected to provide insights into the structural basis of phylogenetic relatedness of HINT proteins from to other closely related species.

MATERIALS AND METHODS

Construction of recombinant expression plasmids

The genes encoding CaHINT and HINT proteins from three other fungal species (*Debaryomyces hansenii*, *Kluyveromyces lactis*, and *Schizosaccharomyces pombe*), were amplified by polymerase chain reaction (PCR). The amplified fragments were digested with the restriction enzymes (Enzynomics, Korea) NdeI (cat. no. R0065) and XhoI (cat. no. R0075), for 3 h at 37℃ in a water bath. The fragments were ligated into pET28a and pET26b vectors using T4 ligase (cat. no. M0202S; Roche, Germany) overnight in a 20℃ incubator. These vectors harbored a hexahistidine (His_6)-tag at either the N- or C-terminus of the target protein. The ligated plasmids were transformed into *Escherichia coli* (*E. coli*), strain DH5α, and transformants were confirmed using colony PCR. The recombinant plasmids were verified by DNA sequencing.

Expression and purification of recombinant proteins

The resulting plasmids encoding HINT proteins were transformed into *E. coli* strain Rosetta. Cell cultures were grown at 37℃ to an OD_{600} of 0.6 in LB medium (cat. no. L4488; MBcell, Republic of Korea) containing 50 mg/L kanamycin (cat. no. A1493; Applichem, USA) and 50 mg/L chloramphenicol (cat. no. A1806; Applichem). Following induction with 0.3 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG; cat. no. 420322; Calbiochem, Germany) and addition of 100 mM zinc chloride (cat. no. 8B4079; Junsei Chemical, Tokyo), the cells were incubated for an additional 16 h at 20℃. The cultured cells were harvested by centrifugation at 4500 × *g* for 20 min at 4℃. The cell pellet was resuspended in a buffer containing 250 mM sodium chloride (NaCl; cat. no. A2942; Applichem), 5% glycerol (cat. no. 56515; Affymetrix, USA), 0.2% Triton X-100 (cat. no. 9002931; Sigma-Aldrich, USA), 10 mM β-mercaptoethanol (cat. no. 60242; BioBasic, Canada), 0.2 mM phenylmethylsulfonyl fluoride (cat. no. PMSF; P7326; Sigma-Aldrich), and 20 mM Tris (pH 8.0; cat. no. T1895; Sigma-Aldrich). Cells were disrupted by ultrasonication. The cell debris was removed by centrifugation at 13000 × *g* for 40 min. The lysate was bound to Ni-NTA agarose (cat. no. 30230; Qiagen, Germany) for 90 min at 4℃. After washing with buffer A (200 mM NaCl, 50 mM Tris, pH 8.0) containing 20 mM imidazole (cat. no. I5513;

Sigma-Aldrich), the bound proteins were eluted with buffer A containing 250 mM imidazole. Size-exclusion chromatography (SEC) was performed on the purified proteins using HiPrep 16/60 Sephacryl S-300 HR (cat. no. 17116701; GE Healthcare, Canada). The buffer used for SEC contained 150 mM NaCl, 2 mM dithiothreitol (DTT; cat. no. 233155; Calbiochem), and 20 mM Tris (pH 7.5). Following SEC, the proteins were stored at -80℃ until crystallization trials. The purified proteins were assessed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis using 18% acrylamide gels, and a single band corresponding to the calculated molecular weight of the target protein was observed. The purification of other constructs was performed using a method similar to that of full-length HINT.

Crystallization trials

All crystallization trials were performed at 4℃ using the sitting-drop vapor diffusion method in 96-well sitting-drop plates (cat. no. 102000100; Art Robbins Instruments, USA). Over 480 different conditions from sparse-matrix screening solution kits were used to identify crystallization conditions. The kits used included PEG/Ion (cat. nos. HR2-126 and HR2-098), Index (cat. no. HR2-144), SaltRx 1/2 (cat. nos. HR2-107 and HR2-109), Crystal Screen 1/2 (cat. nos. HR2-110 and HR2-112) from Hampton Research (USA) and Wizard (cat. no. CS-311) from Jena Bioscience (Germany). Crystals appeared for HINT proteins from all four species (CaHINT, DhHINT, KlHINT, and SpHINT). Crystals of KlHINT grew within 2 days in drops containing equal volumes (1 μl each) of mixed protein sample (50 mg/mL in 150 mM NaCl, 2 mM DTT, 20 mM Tris, pH 7.5) and reservoir solution (0.2 M ammonium sulfate, 0.1 M Tris [pH 8.5], 25% PEG 3350; 2% Tacsimate [pH 7.0], 0.1 M HEPES [pH 7.5], 20% PEG 3350). Crystals of SpHINT grew within 2 days in drops containing equal volumes (1 μl each) of mixed protein sample (50 mg/mL in 150 mM NaCl, 2 mM DTT, 20 mM Tris, pH 7.5) and reservoir solution (0.2 M ammonium acetate, 0.1 M Bis-Tris [pH 6.5], 25% PEG 3350; 0.1 M Bis-Tris [pH 6.5], 20% PEG-MME 5000). Crystals of DhHINT grew within 2 days in drops containing equal volumes (1 μl each) of mixed protein sample (50 mg/mL in 150 mM NaCl, 2 mM DTT, 20 mM Tris, pH 7.5) and reservoir solution (0.2 M ammonium acetate, 30% PEG 4000, 0.1 M sodium acetate [pH 4.6]; 1.8 M sodium acetate [pH 7.0], 0.1 M Bis-Tris propane [pH 7.0]).

Table 1. Data collection and refinement statistics for CaHINT

Data collection and refinement statistics	CaHINT
Data collection	
Space group	$P2_12_12$
Cell dimensions	
a, b, c (Å)	40.354, 101.907, 175.175
α, β, γ (°)	90, 90, 90
Resolution range (Å)	50.0-2.5 (2.59-2.5)[a]
R_{merge} (%)[b]	10.9 (41.3)
// σ/	35.2 (7.8)
Completeness (%)	100 (100)
Redundancy	7.5 (7.5)
Structure refinement	
Resolution	50.0-2.5
No. of reflections	48261
R_{work}[c]/ R_{free} (%)[d]	18.77 / 22.78
No. atoms	4719
Protein	4549
Ligand/ion	4
Water	166
r.m.s.d.	
Bond lengths (Å)	0.007
Bond angles (°)	0.928
Average *B*-factors (Å^2)	35.60
Ramachandran plot (%)	
Favored region	95.95
Allowed	4.05
Disallowed	0

[a]The numbers in parentheses are statistics from the highest resolution shell. [b]R_{merge}= Σ |I_{obs}- I_{avg}| / I_{obs}, where I_{obs} is the observed intensity of individual reflections, and I_{avg} is averaged over symmetry equivalents. [c]R_{work} = Σ ||F_o| - |F_c|| / Σ |F_o|, where |F_o| and |F_c| are the observed and calculated structure factor amplitudes, respectively. [d]R_{free} was calculated using 5% of the data.

Crystals of CaHINT grew within 2 days in drops containing equal volumes (1 μl each) of mixed protein sample (50 mg/mL in 150 mM NaCl, 2 mM DTT, 20 mM Tris, pH 7.5) and reservoir solution (0.2 M magnesium acetate, 0.1 M sodium cacodylate [pH 6.5], 20% PEG 8000; 20% PEG 3350, 0.2 M magnesium acetate). To improve the plated-shaped crystals, additional screening was performed using Additive (cat. no. HR2-428) and Detergent (cat. no. HR2-406) screening kits (both from Hampton Research). The size of the optimized CaHINT crystals was approximately 200 × 100 ×15 μm. Prior to flash-cooling all crystals in liquid nitrogen, 15% glycerol was added to the reservoir solutions as a cryoprotectant.

Data collection and structure determination

All diffraction datasets were collected at -173℃ (100K) on Beamline 5C at the Pohang Accelerator Laboratory (PAL; Republic of Korea), using a Pilatus-3 6M detector (Switzerland)(Park et al., 2017). Data were processed using the HKL-2000 suite (Otwinowski and Minor, 1997). Crystals of CaHINT belonged to space group $P2_12_12$, diffracting to a resolution of 2.5 Å. The crystal structures were solved by molecular replacement methods using PHENIX software, version 1.9 (PHENIX, USA)(Adams et al., 2010). Then, structural models were built using the WinCoot program (Emsley and Cowtan, 2004). The structural models were refined using the PHENIX.refine program. Details of data collection and relevant statistics are provided in Table 1.

Phylogenetic analysis

The amino acid sequences of HINT proteins were obtained from the National Center for Biotechnology Information (NCBI) database (*C. albicans* [XP_713946.1], *Leishmania major* [XP_001687623.1], *Bacillus anthracis* [BAR78197.1], *Bacillus subtilis* [WP_041052754.1], *Methylobacillus flagellatus* [ABE50771.1], *Marinobacter hydrocarbonoclasticus* [WP_011785192.1], *Sulfurisphaera tokodaii* [WP_010980230.1], *Bartonella henselae* [WP_011180543.1], *Encephalitozoon cuniculi* [NP_597168.1], *Mycolicibacterium smegmatis* [WP_011730267.1], *Streptococcus mutans* [WP_002262609.1], *Clostridioides difficile* [WP_077711032.1], *Homo sapiens* HINT1 [NP_005331.1], *Homo sapiens* HINT2 [NP_115982.1], *Oryctolagus cuniculus* [NP_001076092.1], *Clostridium thermocellum* [WP_003519385.1], *Entamoeba histolytica* [XP_655618.1], *Escherichia coli* [WP_000807125.1], and *Helicobacter pylori* [EMH01398.1]). The 19 sequences were aligned in Bioedit v7.2.6.1 (Hall, 1999). There were a total of 211 positions in the final dataset. Maximum likelihood analysis based on JTT matrix-based model (Jones et al., 1992) was conducted using MEGA X (Kumar et al., 2018). The phylogenetic bootstrap consensus tree was inferred from 1000 replicates (Felsenstein, 1985).

RESULTS

Crystallization and X-ray diffraction of HINT proteins

To elucidate the structural features of fungal HINT proteins, we initially obtained His_6-tagged proteins from each of four fungal species (*Candida albicans*, *Debaryomyces hansenii*, *Kluyveromyces lactis*, and *Schizosaccharomyces pombe*; Fig. 1A). The initial crystals of the four fungal HINT proteins appeared with different morphologies mostly in 2 days after crystallizations, under various conditions at 4℃. After a series of in-depth improvements using additive screening, detergent screening, microseeding, and seed screenings, we obtained more than 14 different shapes of crystals, including rods, cubes, thin plates, and needles, from each of the four species (Supplementary Fig. S1). Interestingly, discrete forms of crystals between N- and C-terminal His_6-tagged HINT proteins were shown in all the four species.

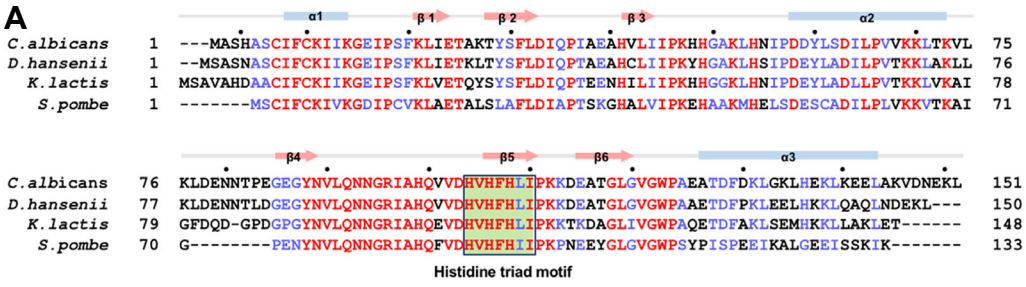
Subsequently, we performed diffraction data collection of the crystals. However, unexpectedly, most crystals, except CaHINT, exhibited poor diffraction (Supplementary Fig. S2). For example, diffraction of KlHINT and SpHINT showed split or overlapped peaks, causing the unit cell parameters to become abnormally large and blocking progression to the further stages. In contrast, DhHINT yielded clear spots with a high resolution of 2.2 Å; however, after data processing, the diffraction data were severely twinned, and we could not obtain the electron density map. In contrast to the crystals from these three species, crystals of both N- and C- terminal His_6-tag CaHINT diffracted well with fully separated single peaks. Therefore, we collected diffraction data from C-terminal His_6-tag CaHINT crystals at a resolution of 2.5 Å. Subsequently, the diffraction data were processed using scalepack in the HKL-2000 software package (Jung et al., 2018).

Determination of CaHINT structure

The crystals of CaHINT belonged to space group $P222_1$, with the following unit cell parameters: a = 40.354, b = 101.907, c = 175.175, $\alpha = \beta = \gamma = 90^\circ$ (Table 1). The values of VM and Matthews coefficients were 720381.812 and 2.49, respectively, which indicated that there were four molecules in the asymmetric unit, with 50.55% solvent content (Matthews, 1968). The initial phase was obtained from the structure of HINT from *Leishmania major* (PDB ID: 3KSV) by molecular replacement using AutoMR software in the PHENIX crystallographic software package (PHENIX)(Adams et al., 2010). Next, a crystal model was build using the Wincoot program (Emsley and Cowtan, 2004). After model building, the model was refined using PHENIX.refine software. The modeling buildings and refinements were performed several times until R and Rfree values of 18% and 24% were obtained.

Overall structure of CaHINT

Although CaHINT appeared to form dimers in solution, there were four molecules of CaHINT in the asymmetric unit, which exist as two dimers shaped like a butterfly (Fig. 1B). Using the *PISA* server (Krissinel and Henrick, 2007), the total accessible surface area that became buried by the interaction between the two monomers was calculated to be ~1100 $Å^2$. These interactions were mostly hydrophobic. CaHINT monomers belonged to the class of α/β folds, consisting of a central six-stranded β-sheet (β1-β6) surrounded by three α-helices (denoted α1-α3). Thus, each monomer had a zinc ion near the α1 helix. Interestingly, the C-terminal helix (α3) protruded and was connected to β6 by an extended loop from the globular structure (Fig. 1C). Furthermore, the Ca-



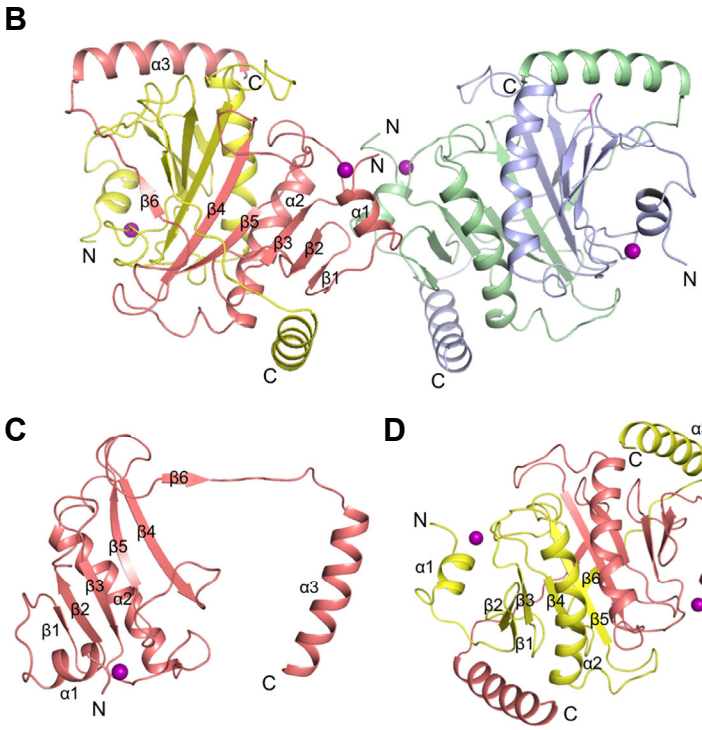


Fig. 1. Overall structures of CaHINT. (A) Sequence alignment of the four fungal HINT proteins. Amino acid sequences of the four fungal HINT proteins were aligned, and 100% identical residue are colored in red. Blue colored residues were 75% identical among the four fungal species. The green-colored box indicates the conserved histidine triad (H-X-H-X-H-X-X motif). (B) The tetrameric structure of CaHINT in the asymmetric unit. The purple sphere indicates a zinc ion. (C) The monomeric structure of CaHINT. (D) The dimeric structure of CaHINT.

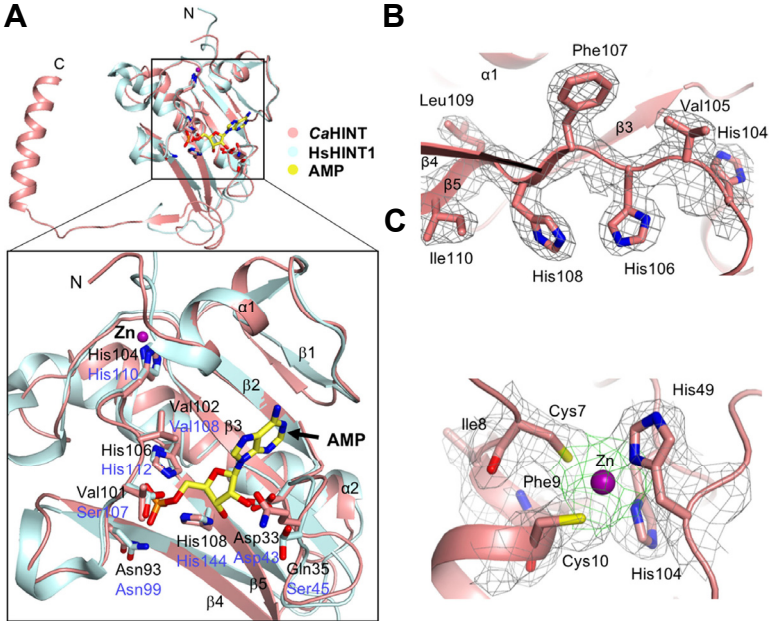


Fig. 2. Active site of CaHINT. (A) Overlaid structures of CaHINT and HsHINT1 in complex with AMP (PDB ID: 3TW2) are shown in the upper panel. The detailed interactions are indicated in the lower panel with a black box. Structures and residue-labels of CaHINT and HsHINT1 are colored in salmon and light cyan, respectively. The AMP molecule and zinc ion are shown in yellow and purple, respectively (HsHINT1: human HINT1). (B) The histidine triad motif of CaHINT. The 2Fo-Fc electron density of the histidine triad motif, including the residues His104, Val105, His106, Phe107, His108, Leu109, and Ile110 of CaHINT, is shown at the 1.0σ contoured level. (C) Zinc binding motif of CaHINT. The 2Fo-Fc electron density for the zinc binding motif (Cys7-Ile8-Phe9-Cys10) and two histidines from the histidine triad contoured at the 1.0σ level. The density of zinc (purple) interacting with the side chains of Cys7, Cys10, His49, and His104 and the main chains of Ile8 and Phe9 is highlighted in green.

HINT dimer was mainly formed by crosslinking of the C-terminal α3 helices, similar to a clamp, making extensive hydrophobic contacts with the N-terminal loop (α1-β1) and polar interactions between the each of the β4 strands from both monomers (Fig. 1D), resulting in the formation of a 10-stranded antiparallel sheet. Therefore, the dimer was found to be a stable structural unit, which may affect enzymatic function. In the current crystal structure, no significant differences were observed in the sizes or structures of the active sites of the two dimers; structural superimposition returned a root mean square deviation (r.m.s.d.) of 0.6 Å.

The active site of CaHINT

To elucidate the catalytic mechanism of CaHINT, the structure of CaHINT was superimposed on that of human HINT1 in complex with the ligand AMP (PDB ID: 3TW2). The position of the histidine triad was similar in both structures (Fig. 2A). The active site of CaHINT was composed of a histidine triad and a zinc binding motif. The first two His residues (His104 and His106) were located on the β4-β5 loop, whereas the third His (His106) was on the β5 strand (Fig. 2B). In addition, the nucleotide binding pocket of human HINT1 was formed by the α1 helix, α3 helix, and β-sheet (β1-β5); CaHINT likewise contained an α1 helix and β-sheet (β1-β5) in the active site. The zinc ions were coordinated by a C-X-X-C motif (where X is a hydrophobic residue), which was composed of an α1 helix and the first histidine (His104) of the histidine triad in CaHINT (Fig. 2C)(Klein et al., 1998; Lima et al., 1996).

Comparisons with other homologous structures

To determine the structural/functional relationships of CaHINT, we attempted to explore the other homologous structures using the *DALI* server (Holm and Rosenstrom, 2010). Eighteen homologous structures were identified based on their *Z*-scores and r.m.s.d. values, as shown in Table 2 and Fig. 3. The most similar structure was HINT from the protozoal species *Leishmania major* (LmHINT); the *Z*-score was 18.8, and the r.m.s.d. value was 2.0 Å. The next similar structures were HINT proteins from the bacteria *Bacillus anthracis* (*Z*-score: 18.4; r.m.s.d.: 1.8 Å) and *Bacillus subtilis* (*Z*-score: 17.9; r.m.s.d.: 1.6 Å).

A comparison between the structures of monomeric LmHINT and CaHINT by superimposition revealed that the two structures overlaid well, except for the C-terminal helix (Fig. 4A). The location of the zinc ion and the conformations of coordinating residues were also quite similar. However, dimers from both species did not superimpose well (Fig. 4B).

Table 2. Analyses for the structural similarities[a]

Species[b]	C-terminal region	Z-score	RMSD (Å)	Identity (%)	Cα	PDB code	NCBI ID
L. major	I	18.8	2.0	43	136	3KSV	XP_001687623.1
B. anthracis	I	18.3	1.9	35	133	3IMI	BAR78197.1
B. subtilis	I	17.7	2.0	36	134	1Y23	WP_041052754.1
E. cuniculi	I	17.4	1.4	33	129	3R6F	NP_597168.1
C. thermocellum	III	16.5	2.0	39	109	5UVM	WP_003519385.1
H. pylori	III	16.5	1.4	36	101	4ZGL	EMH01398.1
C. difficile	III	16.1	1.9	37	110	4EGU	WP_077711032.1
E. coli	III	16.0	1.8	30	109	3N1S	WP_000807125.1
B. henselae	II	15.8	1.9	35	107	3LB5	WP_011180543.1
O. cuniculus	III	15.8	2.0	33	108	3O1Z	NP_001076092.1
H. sapiens (HINT1)	III	15.7	1.6	34	106	3TW2	NP_005331.1
S. mutans	I	15.3	1.7	32	132	3L7X	WP_002262609.1
M. smegmatis	II	14.9	3.6	29	110	3O0M	WP_011730267.1
E. histolytica	III	14.9	2.1	41	106	3OJ7	XP_655618.1
S. tokodaii	II	14.1	2.1	35	104	2EO4	WP_010980230.1
H. sapiens (HINT2)	III	12.9	2.0	29	95	5KM5	NP_115982.1
M. flagellates	II	9.8	2.8	16	99	2OIK	ABE50771.1
M. hydrocarbonoclasticus	II	8.0	3.3	19	95	3OHE	WP_011785192.1
C. albicans	I	-	-	-	-	6IQ1	XP_713946.1

[a]This server computes optimal and suboptimal structural alignments between two protein structures using the DaliLite-pairwise option (http://ekhidna.biocenter.helsinki.fi/dali_lite/start)

[b]The full names of the species are as follows: *C. albicans, Candida albicans; L. major, Leishmania major; B. anthracis, Bacillus anthracis; B. subtilis, Bacillus subtilis; M. flagellates, Methylobacillus flagellates; M. hydrocarbonoclasticus, Marinobacter hydrocarbonoclasticus; S. tokodii, Sulfolobus tokodii; B. henselae, Bartonella henselae; E. cuniculi, Encephalitozoon cuniculi; M. smegmatis, Mycobacterium semgmatis; S. mutans, Streptococcus mutans; C. difficile, Clostridioides difficile; H. sapiens, Homo sapiens; O. cuniculus, Oryctolagus cuniculus; C. thermocellum, Closridium thermocellum; E. histolytica, Entamoeba histolytica; E. coli, Escherichia coli; H. pylori, Helicobacter pylori.*

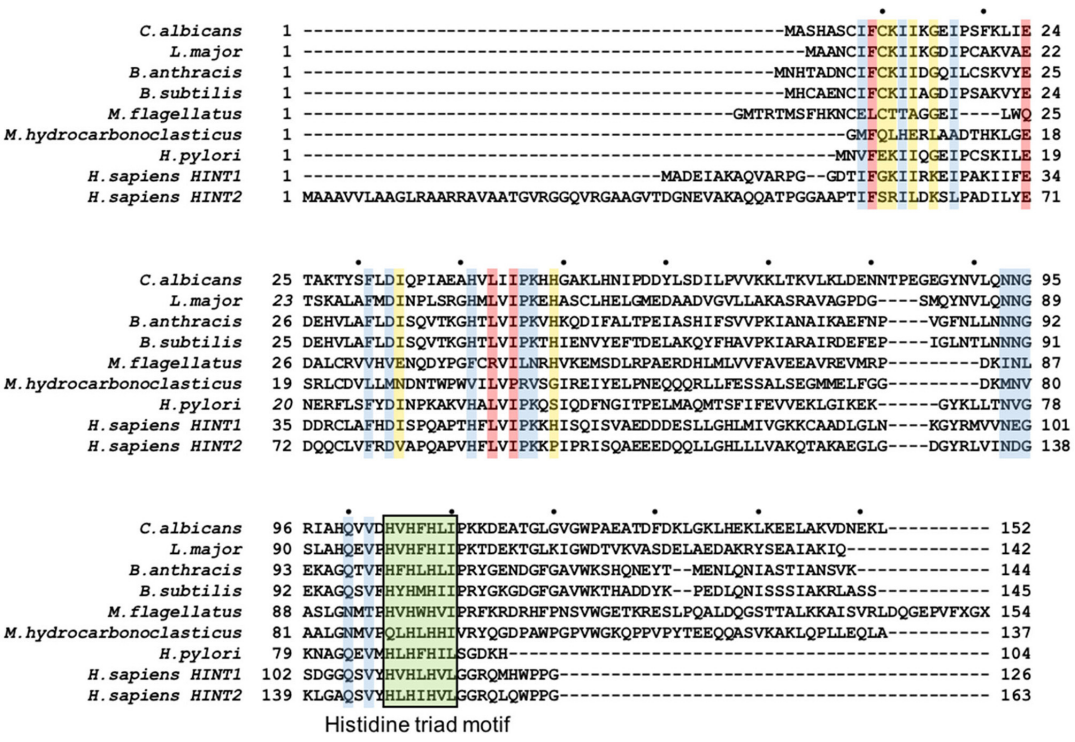


Fig. 3. Alignment of the amino acid sequences of homologous HINT proteins. Residues with 80% identity are colored in red. Blue and yellow highlighted residues indicate 70% and 60% identities, respectively. The green box indicates the conserved histidine triad (H-X-H-X-H-X-X motif). Every 10th residue is marked with a black circle.

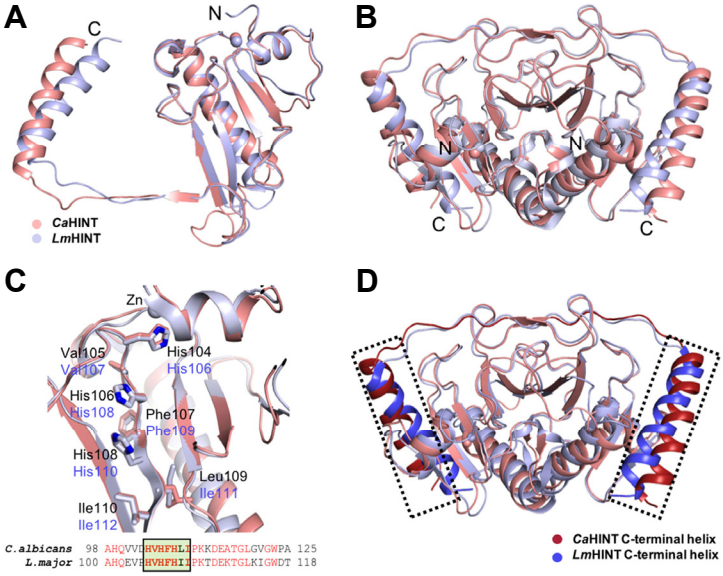


Fig. 4. Structural superposition of CaHINT and LmHINT. (A) Superposed structures of monomeric CaHINT and LmHINT colored in salmon and light blue. (B) Superposed dimeric structures of CaHINT and LmHINT are shown in the lower right portion. (C) Detailed view of the superposed histidine triads from CaHINT and LmHINT. Compared sequences including the histidine triad motif are indicated in the lower panel. Identical residues are show in red. (D) Superposed dimeric structures of CaHINT and LmHINT. The extended C-terminal helices of both structures are highlighted in red and blue, with dotted boxes, as indicated in the lower right of the figure.

Because a single molecule was located in the asymmetric unit of the LmHINT crystal structure, the crystallographic packing of LmHINT resulted in a dimeric conformation different from that of CaHINT. The residues of the histidine triad motif adopted similar conformations and superimposed well, except for Leu109 in CaHINT, which corresponded to Ile111 in LmHINT (Fig. 4C). Nevertheless, the C-terminal helix of LmHINT extended from the central portion to symmetry-related molecules, resulting in stable dimerization; this was similar to that in CaHINT (Fig. 4D). In addition, an extended C-terminal α3 helix was involved in dimerization in all of these structures.

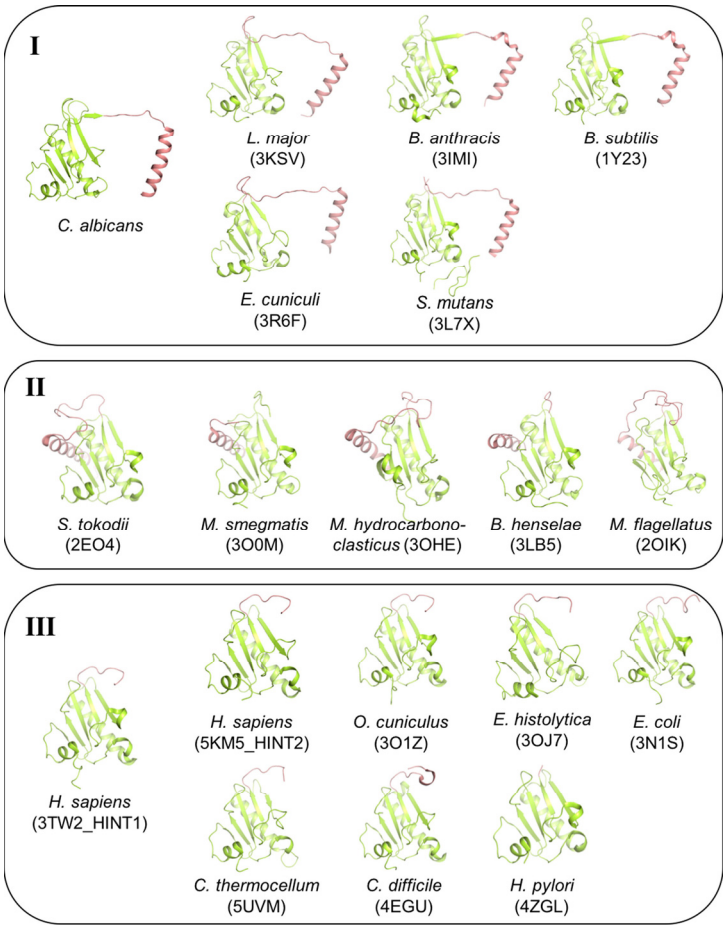


Fig. 5. Homologous structures of CaHINT. Comparison of the overall structures of monomeric HINT homologous proteins. The C-terminal regions of HINT are shown in brown. The PDB codes are indicated in parentheses. The full names of the species are as follows: *C. albicans*, *Candida albicans*; *L. major*, *Leishmania major*; *B. anthracis*, *Bacillus anthracis*; *B. subtilis*, *Bacillus subtilis*; *M. flagellates*, *Methylobacillus flagellates*; *M. hydrocarbonoclasticus*, *Marinobacter hydrocarbonoclasticus*; *S. tokodii*, *Sulfolobus tokodii*; *B. henselae*, *Bartonella henselae*; *E. cuniculi*, *Encephalitozoon cuniculi*; *M. smegmatis*, *Mycobacterium semgmatis*; *S. mutans*, *Streptococcus mutans*; *C. difficile*, *Clostridioides difficile*; *H. sapiens*, *Homo sapiens*; *O. cuniculus*, *Oryctolagus cuniculus*; *C. thermocellum*, *Closridium thermocellum*; *E. histolytica*, *Entamoeba histolytica*; *E. coli*, *Escherichia coli*; *H. pylori*, *Helicobacter pylori*.

However, not all HINT structures had an extended C-terminal α3 helix; for example, human HINT1 and HINT2 were unstructured in this region, and some bacterial HINTs had a C-terminal helix, but were located in different positions without protrusions (Fig. 5). Although the structures of the core regions in HINT were similar, the C-terminal regions were somewhat varied and could be divided into three conformations, i.e., type I (extended), type II (relocated), and type III (unstructured), based on the 19 reported structures (Fig. 5). Interestingly, the C-terminal regions of the type I conformation were associated with the corresponded region in the dimer, unlike in the type II and III structures.

Phylogenetic analysis of HINT

In the C-terminal region of HINT, the correlation between structure and sequence was not clear. Accordingly, phylogenetic analysis was performed with the 19 reported HINT protein sequences and structures (Table 2). The data in the phylogenetic tree were analyzed using Maximum-Likelihood analysis (Fig. 6). Interestingly, the taxa were mostly arranged according to the structure of HINT depending on the C-terminal region type, except in *Bartonella henselae*. Briefly, there were three main clades in the phylogenetic tree. The first clade included *Candida albicans*, *Leishmania major*, *Encephalitozoon cuniculi*, *Bacillus anthracis*, *Bacillus subtilis*, and *Streptococcus mutans* for the type I region (extended C-terminal helix) and *Methylobacillus flagellatus*, *Marinobacter hydrocarbonoclasticus*, and *Mycolicibacterium smegmatis* for the type II region (relocated C-terminal helix). The second clade contained *Sulfurisphaera tokodaii* for the type II region and *Clostridium difficile*, *Entamoeba histolytica*, *Homo sapiens* HINT1, *Oryctolagus cuniculus*, *Homo sapiens* HINT2, *Escherichia coli*, *Clostridium thermocellum*, and *Helicobacter pylori* for the type III region (unstructured C-terminus). *Bartonella henselae* was identified for the type II region in the third clade.

The type I and III conformations belonged to the first and second clades, respectively (Fig. 6). However, the structures of the type II region, including that in *Bartonella henselae*, were distributed in all three clades. Among the five species in the type II region, four were bacteria and one was archaea (*S. tokodaii*). A detailed view of the phylogenetic tree indicated that *B. henselae* was initially separated from the two clades due to low sequence identity of 14-29% among the type II species. Moreover, *S. tokodaii* was branched out at the earliest stage of the second clade because of its discrete domain, grouped as archaea. Based on these evolutionary data, the type II region may be the ancient conformation of the C-terminal region of HINT proteins. The type III region

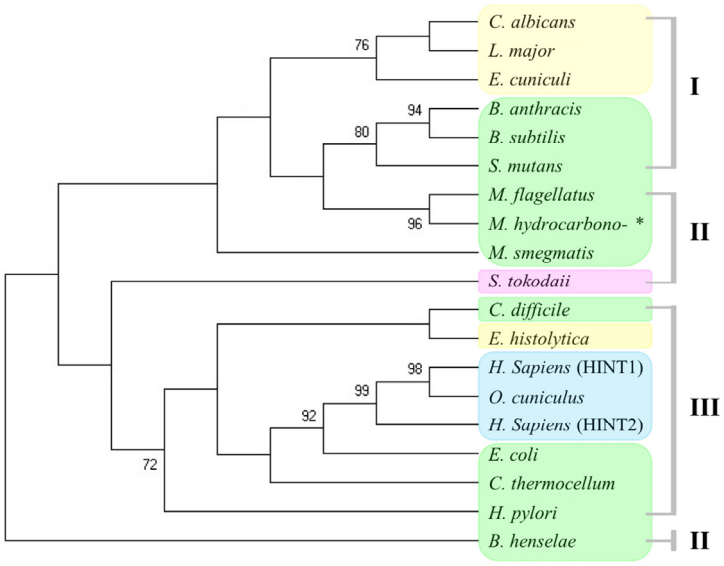


Fig. 6. Phylogenetic tree of HINT. The phylogenetic tree inferred from the *HINT* gene with bootstrap percentages for 1,000 replicates. No outgroup was included due to the large divergence in this study. Bootstrap values under 70% are not shown. The full names of the genera with species are shown in Fig. 5 with abbreviations. The yellow and green highlighted species indicate eukaryotic cells (including protozoa and fungi) and bacteria, respectively. The pink and blue highlighted species represent archaea and animals, respectively. Three groups based the types of C-terminal regions are shown on the right side of the figure. **Marinobacter hydrocarbonoclasticus.*

was observed in human HINT1 and HINT2, rabbits, protozoans (amoeba), and bacteria, which were all found in the second clade of the tree. Although the type I region of *L. major* was classified as the first clade, the presence of *E. histolytica* in the second clade indicated the effects of structural correlations grouped as the type III region.

DISCUSSION

HINT family proteins are classified as hydrolases based on their signature HIT motif, and the catalytic reaction mechanism of these proteins has been well established. Because HINTs also cleave various substrates, including dinucleotide, aminoacyl-adenlyate, and phosphoramidate, their substrate specificity was suggested by crystal structures of several HINT family proteins (Maize et al., 2017; Shah et al., 2017). Despite the abundance of structural information on HINT proteins, limited structural information from fungal species is currently available.

We obtained quite a few good morphological crystals from the four fungal species *C. albicans*, *K. lactis*, *D. hansenii*, and *S. pombe*; however, most crystals, except those of CaHINT, diffracted poorly, producing split peaks or twinned patterns. There may be several reasons for the poor diffractions. First, some of the local connecting loops disrupted the perfect packing of the crystals, thereby negatively affecting diffraction. Although the overall folds of the fungal HINT proteins were expected to be similar, local structural differences may have also affected X-ray diffraction. Moreover, all of the proteins crystallized under different conditions. Second, the extended C-terminal region may have affected crystal diffraction. Sequence alignments revealed several differences among the amino acid sequences. It is possible that these differences adversely affected molecular packing and crystal quality. Third, although the morphologies of the crystals were good, the quality of the crystals was insufficient to result in proper diffraction. This suggested that inadequate cryoprotectants or strong X-ray exposure may have disrupted the crystal quality by slight irregular packing of protein molecules.

Although the active sites and zinc binding motifs of HINT proteins are structurally conserved, the C-terminal regions varied and could be categorized into three different types. In CaHINT, which belonged to type I, the conformation of the C-terminal region was an extended helix, which was strongly involved in dimerization by clamping. In the other two types, the C-terminal regions were not necessary for dimerization. However, interestingly, while both human and *E. coli* HINTs harbored type III C-terminal regions, there was a discrepancy in substrate specificity between human and *E. coli* HINTs due to the involvement of the unstructured C-terminal region in mediating catalysis (Chou et al., 2007). In addition, these structural variations at the C-terminal region have been shown to be related to variations in the substrate specificity among members of the HIT superfamily (Lima et al., 1997). Based on our phylogenetic analysis, we found a strong relationship between the C-terminal region type and clade. Whereas bacterial, fungal, and protozoan HINT proteins harbored C-terminal regions of all types, the reported animal HINT structures could be classified as having type III C-terminal regions. This result may be related to the different speciation processes between animals and other organisms. Although further investigations are required, these findings indicate that the C-terminal region may be involved in catalysis rather than dimerization, particularly in higher organisms. To date, studies on the catalytic mechanism and substrate specificity have focused on human HINT family proteins, which exhibit several complex structures and various substrates and products. However, only a few other complex structures of HINTs have been elucidated with products of AMP or GMP. Therefore, further studies are needed to determine the ligand complex structures with a wide range of substrate specificities and to assess the conformations of the C-terminal regions.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org). The atomic coordinate has been deposited at the Protein Data Bank, with an accession code 6IQ1.

ACKNOWLEDGEMENTS

We would like to thank beamline staffs Yeon-Gil Kim and Sung Chul Ha at beamlines 5C and 7A of the Pohang Accelerator Laboratory (Pohang, Korea) for data collection. This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (grand No. 2016R1C 1B2009691 to JHC and 2017R1D1A2B03029971 to KSC).

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. D. Biol. Crystallogr. *66*, 213-221.

Akter, H., Yoon, J.H., Yoo, Y.S., and Kang, M.J. (2018). Validation of neurotensin receptor 1 as a therapeutic target for gastric cancer. Mol. Cells *41*, 591-602.

Brenner, C., Garrison, P., Gilmour, J., Peisach, D., Ringe, D., Petsko, G.A., and Lowenstein, J.M. (1997). Crystal structures of HINT demonstrate that histidine triad proteins are GalT-related nucleotide-binding proteins. Nat. Struct. Biol. *4*, 231-238.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature *433*, 531-537.

Chou, T.F., Sham, Y.Y., and Wagner, C.R. (2007). Impact of the C-terminal loop of histidine triad nucleotide binding protein1 (Hint1) on substrate specificity. Biochemistry *46*, 13074-13079.

Chou, T.F., and Wagner, C.R. (2007). Lysyl-tRNA synthetase-generated lysyl-adenylate is a substrate for histidine triad nucleotide binding proteins. J. Biol. Chem. *282*, 4719-4727.

Demirbas, D., Coelho, A.I., Rubio-Gozalbo, M.E., and Berry, G.T. (2018). Hereditary galactosemia. Metabolism *83*, 188-196.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta. Crystallogr. D. Biol. Crystallogr. *60*, 2126-2132.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution *39*, 783-791.

Gu, M., Fabrega, C., Liu, S.W., Liu, H., Kiledjian, M., and Lima, C.D. (2004). Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. Mol. Cell *14*, 67-80.

Gu, M., and Lima, C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Curr. Opin. Struct. Biol. *15*, 99-106.

Hall, T.A. (1999). Bioedit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series *41*, 95-98.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic. Acids Res. *38*, W545-549.

Huang, K., Arabshahi, A., Wei, Y., and Frey, P.A. (2004). The mechanism of action of the fragile histidine triad, Fhit: isolation of a covalent adenylyl enzyme and chemical rescue of H96G-Fhit. Biochemistry *43*, 7637-7642.

Huebner, K., and Croce, C.M. (2001). FRA3B and other common fragile sites: the weakest links. Nat. Rev. Cancer *1*, 214-221.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. *8*, 275-282.

Jung, A., Kim, S., and Chang, J.H. (2018). Purification, crystallization, and X-ray crystallographic analysis of histidine triad nucleotide-binding protein from *Candida albicans*. Biodesign *6*, 71-74.

Klein, M.G., Yao, Y., Slosberg, E.D., Lima, C.D., Doki, Y., and Weinstein, I.B. (1998). Characterization of PKCI and comparative studies with FHIT, related members of the HIT protein family. Exp. Cell Res. *244*, 26-32.

Krakowiak, A., Pace, H.C., Blackburn, G.M., Adams, M., Mekhalfia, A., Kaczmarek, R., Baraniak, J., Stec, W.J., and Brenner, C. (2004). Biochemical, crystallographic, and mutagenic characterization of hint, the AMP-lysine hydrolase, with novel substrates and inhibitors. J. Biol. Chem. *279*, 18711-18716.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774-797.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547-1549.

Li, D., Bernhardt, J., and Calderone, R. (2002). Temporal expression of the *Candida albicans* genes CHK1 and CSSK1, adherence, and morphogenesis in a model of reconstituted human esophageal epithelial candidiasis. Inf. Immun. *70*, 1558-1565.

Lima, C.D., Klein, M.G., and Hendrickson, W.A. (1997). Structure-based analysis of catalysis and substrate definition in the HIT protein family. Science *278*, 286-290.

Lima, C.D., Klein, M.G., Weinstein, I.B., and Hendrickson, W.A. (1996). Three-dimensional structure of human protein kinase C interacting protein 1, a member of the HIT family of proteins. Proc. Natl. Acad. Sci. USA *93*, 5357-5362.

Liu, H., Rodgers, N.D., Jiao, X., and Kiledjian, M. (2002). The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J. *21*, 4699-4708.

Liu, S.W., Jiao, X., Liu, H., Gu, M., Lima, C.D., and Kiledjian, M. (2004). Functional analysis of mRNA scavenger decapping enzymes. RNA *10*, 1412-1422.

Maize, K.M., Shah, R., Strom, A., Kumarapperuma, S., Zhou, A., Wagner, C.R., and Finzel, B.C. (2017). A crystal structure based guide to the design of human histidine triad nucleotide binding protein 1 (hHint1) activated ProTides. Mol. Pharm. *14*, 3987-3997.

Maize, K.M., Wagner, C.R., and Finzel, B.C. (2013). Structural characterization of human histidine triad nucleotide-binding protein 2, a member of the histidine triad superfamily. FEBS J. *280*, 3389-3398.

Martin, J., St-Pierre, M.V., and Dufour, J.F. (2011). Hit proteins, mitochondria and cancer. Biochim. Biophys. Acta. *1807*, 626-632.

Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

McCorvie, T.J., and Timson, D.J. (2011). Structural and molecular biology of type I galactosemia: disease-associated mutations. IUBMB Life *63*, 949-954.

Meiller, T.F., Hube, B., Schild, L., Shirtliff, M.E., Scheper, M.A., Winkler, R., Ton, A., and Jabra-Rizk, M.A. (2009). A novel immune evasion strategy of *Candida albicans*: proteolytic cleavage of a salivary antimicrobial peptide. PLoS One *4*, e5039.

Milac, A.L., Bojarska, E., and Wypijewska del Nogal, A. (2014). Decapping Scavenger (DcpS) enzyme: advances in its structure,

activity and roles in the cap-dependent mRNA metabolism. Biochim. Biophys. Acta. *1839*, 452-462.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method. Enzymol. *276*, 307-326.

Park, S.Y., Ha, S.C., and Kim, Y.G. (2017). The protein crystallography beamlines at the pohang light source II. Biodesign *5*, 30-34.

Pfaller, M.A., and Diekema, D.J. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. Clin. Microbiol. Rev. *20*, 133-163.

Schulze, J., and Sonnenborn, U. (2009). Yeasts in the gut: from commensals to infectious agents. Dtsch. Arztebl. Int. *106*, 837-842.

Shah, R., Maize, K.M., Zhou, X., Finzel, B.C., and Wagner, C.R. (2017). Caught before released: structural mapping of the reaction trajectory for the sofosbuvir activating enzyme, human histidine triad nucleotide binding protein 1 (hHint1). Biochemistry *56*, 3559-3570.

Spampinato, C., and Leonardi, D. (2013). Candida infections, causes, targets, and resistance mechanisms: traditional and alternative antifungal agents. Biomed. Res. Int. *2013*, 204237.

Torosantucci, A., Chiani, P., De Bernardis, F., Cassone, A., Calera, J.A., and Calderone, R. (2002). Deletion of the two-component histidine kinase gene (CHK1) of *Candida albicans* contributes to enhanced growth inhibition and killing by human neutrophils *in vitro*. Infect. Immun. *70*, 985-987.

Varnum, J.M., Baraniak, J., Kaczmarek, R., Stec, W.J., and Brenner, C. (2001). Di-, tri- and tetra-5'-O-phosphorothioadenosyl substituted polyols as inhibitors of Fhit: Importance of the alpha-beta bridging oxygen and beta phosphorus replacement. BMC Chem. Biol. *1*, 3.

Weiske, J., and Huber, O. (2006). The histidine triad protein Hint1 triggers apoptosis independent of its enzymatic activity. J. Biol. Chem. *281*, 27356-27366.

Zimon, M., Baets, J., Almeida-Souza, L., De Vriendt, E., Nikodinovic, J., Parman, Y., Battaloglu, E., Matur, Z., Guergueltcheva, V., Tournev, I., et al. (2012). Loss-of-function mutations in HINT1 cause axonal neuropathy with neuromyotonia. Nat. Genet. *44*, 1080-1083.