

New Structural Insights into the Function of the Catalytically Active Human Taspase1

Nirupa Nagaratnam^{1, †}, Silvia L. Delker^{2, †}, Rebecca Jernigan^{1, †}, Thomas E. Edwards², Janey Snider³, Darren Thifault¹, Dewight Williams⁴, Brent L. Nannenga^{1,5}, Mary Stofega³, Lidia Sambucetti³, James J. Hsieh⁶, Andrew J. Flint⁷, Petra Fromme^{1*}, and Jose M. Martin-Garcia^{1,8*}

¹Center for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe, AZ

²Beryllium Discovery Corporation, Bainbridge Island, WA

³Division of Biosciences, SRI International Menlo Park, CA, 94025

⁴Eyring Materials Center, Arizona State University, Tempe, AZ

⁵Chemical Engineering, School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, Arizona

⁶Molecular Oncology, Division of Oncology, Department of Medicine, Washington University, St. Louis, MO

⁷Frederick National Lab for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD

⁸ Department of Crystallography and Structural Biology, Institute of Physical-Chemistry "Rocasolano", Spanish National Research Council (CSIC), Madrid, Spain

[†]These authors contributed equally to this work

*To whom correspondence should be addressed: Petra Fromme, petra.fromme@asu.edu and Jose M. Martin-Garcia, jmmartin@iqfr.csic.es, Tel. +34 (91) 561 94 00 (Ext. 961173)

Keywords: Taspase1, cancer, proteases, Ntn-hydrolases, X-ray crystallography, anisotropy

Abstract

Proteases can play essential roles in severe human pathology, ranging from degenerative and inflammatory illnesses to infectious diseases, with some, such as Taspase1, involved in growth and progression of tumors at primary and metastatic sites. Taspase1 is a N-terminal nucleophile (Ntn)-hydrolase overexpressed in primary human cancers, coordinating cancer cell proliferation, invasion, and metastasis. Loss of Taspase1 activity disrupts proliferation of human cancer cells *in vitro* and in mouse xenograft models of glioblastoma, thus this protein has the potential to become a novel anticancer drug target. It belongs to the family of Ntn-hydrolases, a unique family of proteins synthesized as enzymatically inactive proenzymes that become activated upon cleavage of the peptide bond on the N-terminal side of a threonine residue, which then becomes the catalytic site nucleophile. The activation process simultaneously changes the conformation of a long domain at the C-terminus of the alpha-subunit for which no full-length structural information exists and its function is poorly understood. Here we present a novel cloning strategy to generate a fully active, circularly permuted form of Taspase1 to determine the crystallographic structure of catalytically active human Taspase1 to 3.04Å. We discovered that this region forms a long helical domain and is indispensable for the catalytic activity of Taspase1. Together, our study highlights the importance of this element for the enzymatic activity of Ntn-hydrolases and suggests that this long domain could be a novel target for the design of inhibitors with the potential to be developed into anticancer therapeutics.

Introduction

Proteases are key enzymes in the human genome that carry out important biological processes including, but not limited to, metabolism, tissue remodeling, apoptosis, cell proliferation, and migration. Further, this class of enzymes is implicated in the progression and growth of tumors when their function is deregulated (1, 2). Taspase1 is an endopeptidase that promotes cell proliferation and permits oncogenic initiation through cleavage of the human mixed-lineage leukemia (MLL) nuclear protein. The MLL activation by Taspase1 has been shown to be a key step in several cancer types, including development of childhood leukemia (3, 4). Taspase1 is also involved in proper regulation of expression of *HOX* genes and regulation of cell cycle genes (5, 6). MLL is a large nuclear protein of 3,969 residues (500 kDa) that is synthesized in an inactive form. Hsieh and co-workers discovered that Taspase1 catalyzes the processing of MLL from its inactive to active state of the protein (4). Subsequent studies identified additional Taspase1 substrates, including MLL2, TFIIA, ALF (TFIIA), and *Drosophila* HCF (6-8). Taspase1 mediates cleavage of its substrates by recognizing conserved peptide motifs ($Q^3 X^2 D^1/G^{-1} X^{-2} D^{-3} D^{-4}$) with Asp and Gly residues at positions P1 and P-1, respectively (4). In the case of MLL, Taspase1 cleaves the protein at two sites (after D2666 and D2718), creating an N-terminal fragment of 320 kDa and a C-terminal fragment of 180 kDa, which heterodimerize to form a stable complex that localizes to a subnuclear compartment (4). Despite intensive work on Taspase1, the catalytic mechanism by which Taspase1 cleaves and thereby activates its very large target proteins, and the regulation of that activity leading to severe changes in gene expression that promote cancer development are still not well understood (9).

Taspase1 is a unique threonine aspartase that contains an N-terminal threonine residue as a nucleophile and cleaves after an aspartate residue in the target substrate (3, 4). The human

taspase1 gene encodes a highly conserved protein of 420 residues with a molecular weight of 50 kDa. Taspase1 belongs to a superfamily of proteases called N-terminal nucleophile hydrolases (Ntn hydrolases) (10) that are synthesized as enzymatically inactive proenzymes that become activated upon cleavage of the peptide bond on the N-terminal side of the active site threonine, serine or cysteine. In Taspase1, the intramolecular reaction takes place between residues Asp233 and Thr234 to generate an N-terminal fragment of 28 kDa (α subunit) and a C-terminal fragment of 22 kDa (β subunit). The N-terminal threonine residue of the β -subunit becomes the active site nucleophile(3) essential for proteolytic cleavage of substrates. The two subunits remain associated and assemble into a higher order $\alpha\beta\alpha$ fold (10-13).

No structural information is available for Taspase1 in a complex with any of its target substrates. Only high-resolution crystal structures of the proenzyme and a mature, 2-chain truncated version of human Taspase1 have been determined so far (14). Due to the high flexibility of the region between residues Pro183 and Asp233, Khan *et al.* removed residues 207-233 to enable crystallization (14). Both the proenzyme and the truncated Taspase1 were demonstrated to be dimers, and the overall structure showed no major differences between the proenzyme dimer and the dimer of the truncated protein, with only some reorganization at the dimer interface and significant conformational differences of the side chains in or near the catalytic site, including the catalytic nucleophile Thr234, also were reported (14). Importantly, the cleavage to activate Taspase1 releases and increases the flexibility a 56 amino acid segment at the C-terminus of the alpha-subunit previously constrained by attachment to Thr234. But, in this conformation the sequence had only been partially resolved in prior crystal structures of Taspase1 (14). Interestingly, this loop is also absent in all crystal structures of type 2 asparaginases solved so far (10-13, 15-17).

In a recent study by van den Boom and co-workers, they used molecular modeling, CD, and NMR spectroscopy to predict that this long loop was mostly alpha-helical with a defined helix-turn-helix secondary structure (18). They also hypothesized that, in the context of the full length Taspase1, the long loop might be in a closed conformation, sitting on top of the catalytic site of the proenzyme, and then undergoing a large conformational change to adopt a more flexible status, or open state, in the active Taspase1 (18). The proximity of this unresolved region to the active site and its crucial role in the activation of the cancer-related protease Taspase1 suggest this region could be a novel target for Taspase1 inhibitors. One approach could be the design of inhibitors that bind the loop region acting as clamps to keep the loop in a closed conformation, preventing activation of the Taspase1 proenzyme.

Although published studies (14, 18) have provided important clues to the biological function of Taspase1, structural information on the active form of the enzyme will be critical to help us understand how Taspase1 recognizes and processes its natural targets. Here, we present a novel cloning strategy to generate a fully active, circularly permuted form of Taspase1 by swapping the α and β subunits of the protein so that the region with the long domain is placed at the C-terminus of a single-chain enzyme. This strategy has allowed us to determine the crystal structure of catalytically active human Taspase1 at 3.04Å resolution. The structure reveals the conformation of a clearly defined, long helical domain (Pro183-Asp233) in the open state of the active enzyme. In addition, we demonstrate that this long domain is essential for Taspase1 activity. Taken together, these results lead us to hypothesize that the current view of the proteolytic mechanism of Ntn-hydrolases must be revisited and the long domain has to be included to fully understand how this family of proteases recognize and process their targets.

Results

The circularly permuted constructs of human Taspase1

As Taspase1 undergoes post-translational autoproteolytic cleavage between residues Asp233 and Thr234 and forms a hetero-tetramer composed of alpha and beta subunits (3), isolation of recombinant Taspase1 in homogenous form is complicated by the presence of two species; the inactive proenzyme and the active cleaved enzyme. Because these two species have essentially the same molecular weight, separation by biophysical techniques is difficult, and obtaining pure protein for structural studies is challenging (14). In order to obtain homogenous Taspase1, Khan and co-workers created a mutant of the inactive proenzyme (D233A) and successfully purified and crystallized it. The crystal structure revealed two disordered regions in the alpha subunit, 1-40 and ~203-233. To obtain Taspase1 in its processed cleaved form, they co-expressed and purified the associated alpha and beta subunits. However, attempts to crystallize the hetero-tetramer were unsuccessful due to the presence of disordered regions. A similar co-expression strategy in which residues 207-233 of the alpha subunit were deleted provided the first crystal structure of a mature, 2-chain truncated version of Taspase1 in its hetero-tetrameric form (14). However, residues 183-206 remained unresolved in this structure.

To obtain the first structure of human Taspase1 in its catalytically active state, we designed, expressed and purified three circularly permuted (cp) Taspase1 variants (cp-Taspase1_{α41-183/β}, cp-Taspase1_{α41-206/β}, cp-Taspase1_{α41-233/β}). These constructs have α and β subunits swapped and linked through a GSGS tetrapeptide linker. A comparison of the sequences of our cp constructs with the sequence of WT Taspase1 is shown in Figure 1A. In our constructs, the nucleophile Thr234 becomes the first residue of the N-terminus as it would be found in the wild type (WT) Taspase1 after autoproteolytic cleavage. We deleted the first 40 residues of the alpha subunit, which are

disordered in previous structures (14), and linked this new N-terminus to the C-terminus of the β subunit using the GSGS linker described above. The region 183-233 is also shown to be disordered in the structure of Khan *et al.* However, this region lies adjacent to the catalytic site, the deletion of which might perturb the active site of the enzyme (14). Yet, minimizing disordered regions increases the probability of crystallization and structure determination. Thus, we performed serial truncations in the long domain of the alpha subunit (residues 183-233) in the cp constructs to study the effect of this domain on the catalytic function of the enzyme. Figure 1B shows a comparison of the cp constructs used in our study to the constructs previously reported (14) and to the WT 2-chain Taspase1.

Catalytic activity of the cp human Taspase1 proteins

To assess the catalytic activity of the cp Taspase1 proteins, we developed a quenched FRET-based assay that measures the proteolysis of a labeled Taspase1 peptide substrate. The internally quenched peptide substrate includes 10 residues (P₋₆ – P₄) from cleavage site 2 (after D2718) in MLL, which is more efficiently processed by Taspase1 than cleavage site 1 (after D2666)(3). For the fluorescence emitter, we used an MCA (7-Methoxycoumarin-4-yl)acetyl) fluorophore attached to the N-terminus of the peptide through the first Lys residue; the quencher used was 2,4-dinitrophenyl-lysine (Lys(DNP)) attached to the lysine residue at the C terminus, which is not part of the native sequence in MLL. The peptide was also amidated at the N-terminus. The sequence of the peptide substrate was MCA-Lys-Ile-Ser-Gln-Leu-Asp↓Gly-Val-Asp-Asp-Lys(DNP)-NH₂ where the cleavage site is indicated by the arrow (↓). The proteolytic cleavage of the substrate by Taspase1 proteins was monitored for ~1.5 h after the addition of substrate. Figure 2 shows the reaction progress curves for the three cp proteins compared to that of the WT Taspase1.

The assay without enzyme (substrate alone) was included as a negative control. We observed high catalytic activity for the WT Taspase1 and comparable or slightly higher activity for the long circular permuted construct cp-Taspase1_{α41-233/β}. However, the catalytic activity of Taspase1 was completely abolished by deleting the fragment 183-233 in the shortest construct (cp-Taspase1_{α41-183/β} construct), giving a signal similar to the substrate alone. An intermediate catalytic activity was measured for the cp-Taspase1_{α41-206/β} protein. These results revealed that the region between 183 and 233 is essential for the catalytic activity of Taspase1, whereas previously it was assumed that the deletion of the fragment 183-233 would lead to a fully active enzyme (14). Thus, structural characterization of this functionally important fragment of the enzyme by crystallography became an even more important quest, as it might reveal features of the catalytic activity of Taspase1 that had previously been unappreciated.

Structure determination of the short cp human Taspase1

To determine if the circular permutations induced any structural alteration compared to the structure of the mature, 2-chain truncated Taspase1 from Khan *et al.* (14), we solved the crystal structure of the shortest circularly permuted (cp) constructs (cp-Taspase1_{α41-183/β}). The medium length circularly permuted cp-Taspase1_{α41-206/β} construct was also purified, but crystallization experiments were unsuccessful, likely due to the high flexibility of the extra residues 183-206.

The crystal structure of the cp-Taspase1_{α41-183/β} protein was determined at 2.15 Å resolution (Table 1) by molecular replacement using the structure of the mature, 2-chain truncated Taspase1 reported by Khan and co-workers (PDB 2A8J) (14). The crystals of cp-Taspase1_{α41-183/β} belonged to the space group P6₅. The crystals contain the characteristic hetero-tetramer in the asymmetric unit, as previously described for the mature, 2-chain truncated enzyme (14) (Figure S1A).

The structural alignment of cp-Taspase1_{α41-183/β} with the mature, 2-chain truncated Taspase1 protein from Khan *et al.*, showed that the overall structure of both inactive forms of Taspase1 are very similar to one another with average RMSDs of 0.44 Å and 0.60 Å for the C_α and all visible residues, respectively (Table S1). Thus, reassuringly, the circular permutation has not altered the general folding of the protein (Figure S1B). There are, however, minor differences in the loop regions between the two structures (Figure S1B). A comparison of the two chains in the asymmetric unit showed that the α and β subunits have essentially the same conformation in the dimeric structure with average RMSDs of 0.32 Å and 0.56 Å for the C_α atoms of the α and β subunit, respectively (Table S1). Chain B shows a slightly higher flexibility compared to that of chain A, which is reflected in a higher number of unmodeled residues (Table S2). A chloride ion was found in the catalytic site of the enzyme in the same position as it was in the truncated Taspase1 protein reported by Khan *et al.*(14) (Figure S2A), and an additional chloride ion was identified between the two protein monomers (Figure S2B). Also, a sodium ion was modeled just behind the active site (Figure S2C). The presence of Na⁺ ions in similar locations have been reported in the literature for type 2 asparaginases, not having any influence in the function of these enzymes.

Structure determination of the active cp human Taspase1

While the crystallization of the cp-Taspase1_{α41-183/β} construct was rather straightforward, the fully active, circularly permuted Taspase1 (cp-Taspase1_{α41-233/β}) form of the enzyme was very difficult to crystallize. Highly purified enzyme was subjected to large-scale, high throughput screening and optimization procedures to grow crystals of sufficient size and quality for X-ray structure analysis. After anisotropy correction (see Materials and Methods section for details), the

crystal structure of the catalytically active cp-Taspase1_{α41-233/β} protein was determined to 3.04 Å resolution by molecular replacement using the structure of the inactive circularly permuted cp-Taspase1_{α41-183/β} protein as initial search model for phasing. The final structure was refined to R_{work} and R_{free} values of 25.5% and 31.7%, respectively.

Crystals of cp-Taspase1_{α41-233/β} protein contained the dimer of the α and β subunits (or hetero-tetramer) in the asymmetric unit with the typical αββα folding observed in this hydrolase family, in which the β strands are located in an antiparallel sandwich, and the α helices are located on the faces of the sandwich (Figure 3). All residues of the α and β subunits of both monomers have been correctly modeled except for a small loop (Ser352-Gln362), the last four residues of the long helix domain (Gly230, Thr231, Leu232, and Asp233), and the His-tag at the N-terminus. The structure of Taspase1 presented here has very good refinement statistics and excellent geometry (bond lengths and bond angles) as shown in Table 1.

Despite the severe anisotropy featured by the crystals (Figure S3), the electron density maps of our structure show a high level of structural details. The 2mF_o-DF_c electron density maps of the long helix domain (residues 183-228) can be used to assess the quality of our structure (Figure S4). Model-induced structural bias is absent in our Taspase1 structure, as positive electron density of the residues comprising the long helical domain was clearly visible already in the very initial stages of refinement (Figure S5). Two more regions from the core of the protein (the helical fragment from Ala58 to Ser96 and the beta strand fragment from Thr234 to Ser252) have been randomly selected to prove the substantial improvement of the electron density after anisotropy correction (Figure S6).

The most relevant finding of this new structure was the identification of the long helix, extending between Pro183 and Asp228, that is required for the activity of the enzyme. This

segment was either absent in previous Taspase1 structures (14), or only partially resolved (Pro183-Lys202) in the proenzyme (14). Recently, van de Boom and co-workers proposed conformations for this long fragment before and after autoproteolytic cleavage based on molecular dynamic simulations, circular dichroism, and NMR (18) with the structure of the proenzyme as reference (PDB 2A8I (14)). While they predicted this domain to form a partially folded helix-turn-helix conformation (18), our structure instead reveals a novel conformation for this long fragment forming a long straight helix pointing outwards in opposite directions of the dimer (Figure 3). Similar to what was previously described by van de Boom and co-workers, the long helical fragment is interrupted by a turn-like element between residues Glu204 and Glu207 (Figure 3).

A structural comparison of cp-Taspase1_{α41-233/β} with previous Taspase1 structures shows that, while there is a high degree of similarity in the core of the protein, significant structural differences have been found in the loop areas and in the long domain that confers activity to the protein and was missing in all previous structures. Statistical results and RMSD values for the comparisons are in Table S3. A superposition of the active cp-Taspase1_{α41-233/β} structure to the published Taspase1 structures as well as to our inactive short cp-Taspase1_{α41-183/β} structure is shown in Figure S7. Interestingly, the long helix domain is in close proximity to the catalytic site, and slight but potentially important differences have been found in residues such as Tyr52 and Lys57 in the catalytic site (Figure 4). These differences become more pronounced when comparing cp-Taspase1_{α41-233/β} to the proenzyme. For instance, the side chain of Tyr52 is pointing toward the catalytic site in cp-Taspase1_{α41-233/β} and outward in the proenzyme. A comparison of the catalytic sites of chains A and B of the cp-Taspase1_{α41-233/β} with all Taspase1 structures is shown in Figure S8.

As mentioned above, the largest structural differences are found at loop areas. One of those is the loop formed between Gly153 and Arg159, which has not been resolved in the previous mature, 2-chain truncated Taspase1 structure (14) and the cp-Taspase1_{α41-183/β} (this study), due to its high flexibility. In contrast, this loop is ordered in the structure of the active Taspase1 and has been fully modeled in our cp-Taspase1_{α41-233/β} structure. The tip of this loop of one of the monomers is slightly bent backwards and points towards the middle of the long helix domain of the neighboring monomer that stabilizes it (Figure S9). The loop fragment (Gly153-Arg159) was also fully modeled in the proenzyme (14), but has a slightly different conformation (Figure S9), as the amino acids of the helical domain from Leu203 onward were flexible and thereby could not be resolved in the structure of the inactive proenzyme.

Taken together, the change of the conformation of Tyr52 in the catalytic site and the bent conformation of the loop from Gly153 and Arg159 along with the long helix domain itself make the catalytic site of the protein adopt a narrower and deeper conformation than that of the truncated Taspase1 structures and the proenzyme (Figure S10). The calculated accessible surface area (ASA) of the catalytic site for the active cp-Taspase1_{α41-233/β} is larger than those calculated for the truncated proteins (cp-Taspase1_{α41-183/β} and Taspase1_{α41-206/β} (PDB 2A8J (14)) and the proenzyme (Table S4).

Overall structure of the active cp human Taspase1

Crystals of the cp-Taspase1_{α41-233/β} protein belonged to the space group H32. An extended view of the asymmetric unit reveals that the αββα hetero-tetramer is part of a larger assembly, a double-ring like structure. An overview of the crystal packing of the cp-Taspase1_{α41-233/β} protein showing the rearrangement of the double-ring assembly is illustrated in Figure S11. This packing

is completely different from what has been so far observed for all the truncated forms of the enzyme or the proenzyme, in which each single-ring forms a trimer of $\alpha\beta\beta\alpha$ hetero-tetramers, with the long helix domain being disposed sequentially in an up-and-down configuration. A schematic overview of the arrangement of the $\alpha\beta\beta\alpha$ hetero-tetramers in the single-ring assembly is shown in Figure 5A-C along the crystallographic ternary axis. In this arrangement, the three $\alpha\beta\beta\alpha$ units are identical, where the beta subunits are located closer to the 3-fold axis and form tight interactions within the trimer, while the alpha subunits are located at the periphery. As shown in Figure S12A showing the side view of the double-ring structure, the single-ring structure is arranged as layers that are mostly connected through the long alpha helix. Analysis of the dimer interfaces that form the double-ring assembly carried out by PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) server (19) revealed that the double-ring structure is stabilized through seven highly symmetric interfaces. Information about all the interfaces is listed in Table 2. All interfaces are shown in Figures 5D and S12B. A careful analysis of the interfaces is described in the Supplementary Information material.

DLS confirms a concentration dependent oligomeric state

To study whether the single and/or double-ring structures observed in the crystals exist in solution or are a crystallization artifact, we performed DLS experiments with Taspase1 at various concentrations (2.5, 5, and 10 mg/ml) in the purification buffer in the absence of crystallization agents. As shown in Figure S13 and S14, a uniform and narrow size distribution was observed for all samples, indicating that only one species was present in solution at each concentration. However, the hydrodynamic radius of the observed species increased from ~5 nm to approximately ~5.5-5.6 nm as the protein concentration increases. Those hydrodynamic radii were estimated to correspond to protein species between 150 kDa and ~225 kDa, indicating that the oligomeric state

of Taspase1 is strongly concentration-dependent varying from tetrameric at 2.5 mg/ml to hexameric at 10 mg/ml. This behavior suggests that different oligomeric forms of Taspase1 may exist in solution, and possibly, also in cells.

EM suggests the dimer as predominant assembly in solution

To further examine the oligomeric state of the protein using an alternative method, we performed negative staining electron microscopy experiments. The electron micrographs show a highly uniform distribution of particles, with an estimated size of ~7-8 nm (Figure 6A). The sample was further analyzed by using 2D class averaging of negative staining EM micrographs. We identified a total of 250 particle classes, three of which are shown in Figure 6B. The sizes of all those particles was estimated to be ~7.5 nm. Thus, even at extremely low concentrations, Taspase1 forms large particles that resemble in both shape and size those of the $\alpha\beta\alpha$ hetero-tetramer (Figure 6C). Interestingly, the orientation of some of the particles allowed us to identify the long domain. However, no evidence of double-ring structures was found, which indicates that the higher assembly is an artifact of crystallization.

Disorder prediction and coiled coil structure analyses

The propensity of Taspase1 to harbor intrinsically disordered regions (IDRs) was predicted using four different web servers: PONDR-FIT (20), MetaDisorder (21), MobiDB (22), and SPOT-Disorder2 (23). All four tools predicted the long domain (Pro183- Asp233) of Taspase1 to be highly disordered (Figure 7A). According to PONDR-FIT and MetaDisorder, this could be a structural disorder, as it encompasses at least 25 or more consecutive residues of the helix (Figure 7B and Table S6). Further analysis with the Protparam-Expasy tool (24), showed that the long

domain is enriched in charged and flexible residues (disorder-promoting residues) such as R, S, E, D and K, as well as lack/low abundance of hydrophobic residues (order-promoting residues) such as G, C, P, V, W and Y, which explains the high tendency of this fragment to be disordered (Figure 7C). The long domain was also inspected by several coiled coil structure predictors including COILS (25), DEEPCOIL (26), and MARCOIL (27). All three predicted the long fragment to have a high tendency to form a coiled coil structure (Figure S15).

Discussion

In this study, we have determined the 3D structure of the fully active form of human Taspase1 to 3.04 Å resolution, capturing the protein with a previously unobserved, important structural feature, a long C-terminal domain presented as a nearly continuous helix projecting outward from the dimer, which is essential for enzymatic activity. Previous molecular dynamic simulations and NMR methods suggested that this piece of the protein (Pro183-Asp233) forms a helix-turn-helix conformation and adopted two different states, a closed state in the proenzyme and an open state that is acquired upon autoproteolytic cleavage (18). Yet, little is known about this newly discovered long helix domain of Taspase1 or the structurally related type 2 asparaginases. Despite its proximity to the catalytic site, this portion of the protein had not been thought to be important for catalytic activity, perhaps because its mobility had prevented it from being observed in the crystal structures previously reported (10-14, 28-36). While it has been demonstrated that mutations within the catalytic site of type 2 asparaginases, including Taspase1 (14), hASRGL1 (37), glycosylasparaginase (38), and hASNase3 (39), and others, abolish or significantly diminish the catalytic activity of the proteins, an essential role for a segment like the Pro183-Asp233 helix of Taspase1 had not previously been considered. In addition to cp-

Taspase1_{α41-233/β}, we have generated two shorter protein variants, the cp-Taspase1_{α41-206/β} containing half the helix (Pro183-Asp206), and the cp-Taspase1_{α41-183/β}, which lacks the entire fragment. While cp-Taspase1_{α41-233/β} was fully active the catalytic activity was diminished in cp-Taspase1_{α41-206/β} and completely abolished in cp-Taspase1_{α41-183/β}. These results suggest that a similar behavior could be expected for the long fragment of structurally-related type 2 asparaginases in which the full length of the helical domain may be needed for the enzymes to develop their function.

Mutational studies of residues in the catalytic site (14) led to the currently accepted model for the proteolytic mechanism of Taspase1 in which a small set of amino acid residues (e.g., the nucleophile Thr234) are considered important to the catalytic reaction. However, we have demonstrated that the loss of the long domain (partially or fully) dramatically affects the activity of Taspase1 and thus raises questions as to the critical role of this region in the function of Taspase1. We have also observed that the presence of the long domain leads to conformational changes, specially Try52 and the loop Gly153-Arg159, which could be somewhat implicated in the catalytic mechanism of Taspase1. Those conformational modifications create a narrower and deeper catalytic site, and possibly could help stabilize and facilitate substrate binding. In the case of the Tyr52, even though is far from the active site, it could contribute to stabilize and redirect the substrates into the catalytic site by interacting with them. Thus, the most likely role for the long domain in the proteolytic cleavage of substrates by Taspase1 would be to facilitate substrate binding. However, we currently have no experimental evidence, such as structures with bound substrate or product, that might support this proposal. Having crystal structures of cp-Taspase1_{α41-233/β} protein in complex with substrates is absolutely required to fully understand how Taspase1 recognizes and processes its targets. We therefore conclude that the current proteolytic mechanism

of Taspase1 and, overall, the type 2 asparaginases family must be revised, and the long domain, which has been largely ignored so far, must now also be considered as a key part of the mechanism.

The observed straight helical conformation that points outwards and interacts with another long helix of a neighboring molecule resembles that of a typical coiled coil-type structure. This novel feature has been further supported by sequence analyses where this fragment was predicted to have a high tendency to form coiled-coil type structures. In the recent study by van de Boom and co-workers (18), the long fragment was predicted to form two helices separated by a short turn (a helix-turn-helix conformation), in which the two helices interact with each other to adopt a folded conformation typical of coiled-coil structures (18). Additionally, we have shown that the long domain of Taspase1 has also been predicted to be intrinsically disordered using various disorder predictors algorithms, which is consistent with the fact that the long domain has been unresolved in all Taspase1 structures reported so far. We also have used MoRFPred (40) to predict that molecular recognition features (MoRFs) within the long-disordered region. MoRFs are short functional IDRs located in long disordered regions and undergo disorder-to-order transition upon interacting with their partners (41). MoRFPred predicted several MoRF residues within both middle- and C-terminal- disordered regions (Figure 8A, B; and Table S6) of Taspase1. Disordered regions in proteins have often been associated with function (42-45). A common characteristic of disordered regions is their ability to become ordered, typically leading to the formation of α helices, upon binding of substrates from natural targets (44).

Taken together, the long domain of Taspase1 might adopt a coiled coil-type structure promoted by its natural substrates. Taspase1 has been shown to contain a highly evolutionary conserved bipartite NLS at the long domain between residues 197 and 220 (¹⁹⁷KRNKRKLELAERVDTDFMQLKKRR²²⁰), which could be utilized to translocate the protein

into the nucleus by direct binding to importin- α . Therefore, a plausible hypothesis is that Taspase1, initially expressed as an inactive precursor, could use its long domain to translocate from the cytoplasm to the nucleus, where ultimately undergoes intramolecular cleavage to become the active protease, through the interaction of its nuclear localization sequence (NLS) with importins. Importins α possess a highly structured helical domain comprised of ten tandem armadillo (ARM) repeats (46-48), whose interior serves as the NLS-binding groove by wrapping it in a fashion that resembles a twisted slug (49-51). The structure of the cp-Taspase1 _{α 41-233/ β} protein reported in this study provides the first structural visualization of the NLS within the context of the Taspase1 protein. We can, therefore, propose that the long domain of Taspase1 binds to importin α in a similar fashion to that found for the NLS of the Fused in Sarcoma (FUS) protein (49-51). The NLS of FUS protein, which is initially intrinsically disordered, becomes ordered by adopting a helical conformation upon binding to importin α . Taken together, we hypothesize that, as observed for FUS protein, the binding of Taspase1 to importin α may trigger large conformational changes in the initially disordered long domain inducing it to become the long straight helix observed in our crystals.

Controversy concerning the functional multimeric state of Taspase1 exists. All structural studies reported so far support the formation of the hetero-tetramer, in which the $\alpha\beta\beta\alpha$ subunits remain together upon autoproteolytic cleavage (14), which was postulated by others based on the high homology of Taspase1 with the Type 2 asparaginase and the Ntn hydrolase families (3, 4, 52). However, studies carried out in living cancer cells suggested Taspase1 functions as an $\alpha\beta$ monomer (53, 54). The approach pursued by Bier *et al.* examined whether interference of Taspase1 dimer formation via expression of dominant-negative mutants could impede Taspase1 from processing its natural targets. However, this strategy, that has been successfully applied to other

systems (55-59), was shown to be not applicable to Taspase1 as the activity was not inhibited when WT Taspase1 was in contact with inactive mutants (53). Bier, *et al.* argued that this outcome is due to the failure of Taspase1 to form the $\alpha\beta\beta\alpha$ hetero-tetramer *in vivo*. More recently, Sabiani and co-workers proposed a model for the proteolytic mechanism in which the hetero-tetramer formation process was suggested to be the key event that triggers the activation (or maturation) of Taspase1 (60). The results of Bier and co-workers argue against the hetero-tetramer formation. However, in their studies, Taspase1 was expressed as a fusion protein (i.e., with GFP or mCherry) to monitor fluorescence and both fusion tags are over half the size (~27 kDa) of Taspase1 (50 kDa). Thus, it is possible that hetero-tetramer formation might be impeded. In addition to this, Bier and co-workers argue that the Khan *et al.*'s crystallographic results show that the $\alpha\beta\beta\alpha$ structure is a crystallization artifact (53). However, Khan *et al.* co-expressed and purified the individual α and β subunits, leading to the $\alpha\beta\beta\alpha$ hetero-tetramer later seen in their crystallization studies. Also, as shown in our results, the hetero-tetramer interface buries a large surface area that is stabilized by numerous contacts (see Table 2 and S5). Large and stable interfaces such as these are unlikely to result only from crystal contacts but are more likely to be interfaces forming functional oligomeric states. Our DLS and EM studies also demonstrate that the smallest assembly of Taspase1 in solution is the $\alpha\beta\beta\alpha$ hetero-tetramer, confirming what has been proposed by published Taspase1 structures (14). DLS also demonstrated that the oligomeric state of Taspase1 is concentration dependent, forming higher oligomeric states -plausibly tetramers and hexamers- in solution at higher protein concentrations. We therefore hypothesize that the biological multimeric state of Taspase1 is the $\alpha\beta\beta\alpha$ hetero-tetramer that eventually self-oligomerizes to form a higher oligomeric assembly in the form of a trimer of $\alpha\beta\beta\alpha$ hetero-tetramers, leading to the single-ring structure. Further *in vitro* and *in vivo* studies, as well as structural analysis, will be needed to

understand how and why Taspase1 oligomerizes and what are the biological functions or consequences of this self-oligomerization.

Taspase1 is classified as a “non-oncogene addiction” protease overexpressed in primary human cancers and its deficiency disrupts proliferation of human cancer cells *in vitro* and in mouse tumor xenograft models of glioblastoma. Taspase1 is, therefore, considered a novel anticancer drug target; however, to date the only reported inhibitors are an arsenic acid-based small molecule (NSC48300) and succinimidy- containing peptides with modest potency (61)–(62). In order to advance the design of Taspase1 inhibitors for cancer therapy, more structural information in combination with biochemical and cellular assays to specifically assess inhibition of Taspase1 will be needed. In this regard, the structure of Taspase1 reported in this study, along with the activity assay, offers new insights into the requirements for Taspase1 activity. Our results demonstrate that, even though residues essential for catalytic activity lie within the presumed catalytic site of Taspase1, another region of the protein strongly influences the catalytic activity of Taspase1, which was previously unappreciated. Further investigation of the role of the long domain in the catalytic activity of Taspase1 may offer opportunities for drug targeting and suggest potential alternative routes to inhibition.

Materials and Methods

Cloning, expression, and purification of the circularly permuted human Taspase1 proteins

When examining the structure of the proenzyme Taspase1 (PDB 2A8I) (14), the proximity of the C-terminus of the C-terminal domain (CTD) to the first AA observed in the N-terminal domain (NTD), G41, was noted. This juxtaposition suggested that insertion of a short linker (GSGS) between these ends might enable the creation of circularly permuted (cp), single chain

constructs, which might simplify expression and purification of different forms of Taspase1. Three circularly permuted constructs (cpTaspase1_{α41-183/β}, cpTaspase1_{α41-206/β} and cp-Taspase1_{α41-233/β}) were designed. Expression and purification of each of these proteins has enabled us to characterize their enzymatic activity and determine novel crystal structures.

The expression vector pEMB7013 harboring the sequence of the cp-Taspase1_{α41-233/β} (a.a. 234-416-GSGS-41-233), the partially truncated cpTaspase1_{α41-206/β} (a.a. 234-416-GSGS-41-206) or the fully truncated cpTaspase1_{α41-183/β} (a.a. 234-416-GSGS-41-183) human Taspase1 constructs with a hexa-histidine tag at the C-terminus was transformed in the *E. coli* BL21 (DE3) strain. Transformed cells were grown at 37°C in 2xYT medium supplemented with 50 µg/ml kanamycin, and expression of Taspase1 was induced with 0.5 mM IPTG when the OD₆₀₀ reached values between 0.5 and 0.7. Following growth overnight at 20°C, cells were harvested by centrifugation at 7,500 x g for 20 min at 4°C and stored at -80°C for purification later. Purification of the three proteins was carried out using the same procedure. Cell pellets were re-suspended in 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol, 0.25 mM TCEP, 0.5% CHAPS, 1 mg/ml lysozyme and 30 U/ml Benzonase, supplemented with EDTA-free protease inhibitor cocktail (Sigma-Aldrich). Cells were lysed by sonication using 5 cycles in pulsed mode with 1 min rest on ice between cycles. Each cycle consisted of thirty 2 second (s)-pulses at 70% amplitude and cooling for 2s between pulses. The lysate was subsequently centrifuged at 45,000 x g for 1 h at 4°C to remove unbroken cells and cell debris. The supernatant was then filtered through 0.22 µm filters and loaded onto a Ni-NTA HisTrapTM HP column previously equilibrated with equilibration buffer (EB) 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol, 0.25 mM TCEP, 10 mM imidazole. After collection of the flow through, the column was washed with 20 column volumes of EB followed by another wash with 15 column volumes of EB containing 25 mM imidazole. Taspase1 proteins were finally

eluted with EB using a linear gradient of imidazole between 25 and 250 mM. Fractions containing Taspase1 protein were combined and concentrated to 12-15 mg/ml using 30 kDa cut-off concentrators. The protein was further purified by size exclusion chromatography (SEC) using a HiPrep™ 26/60 Sephacryl™ S-200 HR (GE healthcare) column previously equilibrated in 20 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol. Fractions containing Taspase1 protein were pooled, concentrated as described above, flash-frozen using liquid N₂, and stored at -80°C for later use. All purification steps were carried out at 4°C using the ÄKTA Pure fast protein liquid chromatography (FPLC) system (GE healthcare). Protein purity was assessed by SDS-PAGE. Purification results of the cp-Taspase1_{α41-233/β} are illustrated in Figure S16A, B.

Activity assays of the circularly permuted human Taspase1 proteins

To assess the catalytic activity of the designed cp Taspase1 constructs, a quenched fluorescence resonance energy transfer (FRET) based enzyme assay was performed (63) that measures proteolysis of a labeled Taspase1 substrate by the purified cp Taspase1 constructs. The substrate consisted of a 10-residues peptide derived from the second cleavage site of the endogenous Taspase1 substrate MLL, flanked by a MCA (7-Methoxycoumarin-4-yl)acetyl) fluorophore and 2,4-dinitrophenyl-lysine (Lys(DNP)) as the quencher: MCA-Lys-Ile-Ser-Gln-Leu-Asp↓Gly-Val-Asp-Asp-Lys(DNP)-NH₂. The peptide was purchased from CPC Scientific Inc (San Jose, CA). In the absence of Taspase1, the fluorophore does not emit due to close proximity of the quencher. The proteolysis of the peptide substrate by Taspase1 releases the quencher from the fluorophore and generates a fluorescence signal. A schematic of the proteolytic reaction is illustrated in Figure S17. Fluorescence intensities were measured with excitation 340/emission 455 filters using a Perkin Elmer Envision multi-mode plate reader. The reaction buffer contained

100 mM ammonium acetate pH 8.3, 625 μ M β -mercapto-ethanol (BME), 10 μ M substrate, and 100 nM enzyme.

Taspase1 enzymes were tested for activity by a kinetic read method in triplicate alongside a “substrate only” negative control in a 384 well plate. Buffer (5 μ l) and enzyme (10 μ l) were added to each well at final concentrations of 100 mM NH₄Ac pH 8.3, 625 μ M BME plus 100 nM Taspase1 enzymes, except for the “substrate only” wells that received 10 μ l of additional assay buffer instead of enzyme. The reactions were started by adding 10 μ l of 10 μ M MCA-DNP substrate to all the wells. Plates were read on the EnVision multi-mode plate reader, which was immediately initialized to generate reaction progress curves. Readings were taken every minute for 90 min at room temperature.

Crystallization of the circularly permuted human Taspase1 proteins

Automated, high-throughput crystallization experiments were performed using various commercial kits. Initial crystallization conditions of truncated cp Taspase1 (cpTaspase1 _{α 41-183/ β}) at ~9-10 mg/ml were identified from PACT screen (B11) condition (0.2 M calcium chloride, 0.1 M MES pH 6, 20% (w/v) PEG 6000). For cpTaspase1 _{α 41-233/ β} , initial hits were obtained and further optimized with the best crystals of cpTaspase1 _{α 41-233/ β} being obtained at 20°C using hanging drop vapor diffusion method. Protein at 15 mg/ml was mixed in a 2:1 ratio with precipitant solution composed of 0.1 M sodium citrate pH 4.0, 1.0 M ammonium sulfate, 10% 2-methyl-2,4-pentanediol (MPD), and 0.4 M NaCl. Crystals in these conditions grew in 3 days and reached sizes between 200 and 500 μ m in their longest dimension (Figure S16C).

Data collection, structure determination, and refinement of the cpTaspase1 _{α 41-183/ β} protein

Crystals of cpTaspase1_{α41-183/β} grown in the above conditions were cryo-protected using 20% ethylene glycol in the crystallization buffer. Data sets were collected at beamline 21-ID-F at the Advanced Photon Source (APS), with a RAYONIX MX-300 detector at a wavelength of 0.97872 Å under a stream of nitrogen (100K). Data were indexed and integrated with XDS/XSCALE (64). The structure was solved by molecular replacement with MOLREP (65) using the structure of the mature, 2-chain truncated Taspase1 (PDB 2A8J(14)) as search model. The final structure was solved at a resolution of 2.15 Å after numerous iterative rounds of manual rebuilding in COOT (66) and refinement in PHENIX (67). The structure was examined, validated, and improved using MolProbity. The crystallographic data collection and structure refinement statistics are summarized in Table 1.

Data collection, structure determination, and refinement of the cpTaspase1_{α41-233/β} protein

For data collection, crystals grown in the conditions above indicated were transferred into a cryoprotectant solution of 30% glycerol in the crystallization buffer for less than 5 seconds before being flash cryo-cooled in liquid nitrogen. X-ray data sets were collected for the cpTaspase1_{α41-233/β} crystals at 100 K at the APS GM/CA 23-ID-D beamline using a Pilatus 16M detector. The diffraction images were indexed and integrated with XDS (64) in H32 space group with unit cell dimensions of a=196Å, b=196Å, c=196.91Å. The dataset was scaled, and the structure factors generated using AIMLESS (68) and TRUNCATE (69), respectively. Using XTRIAGE program (70) in PHENIX suite (67), the crystals were confirmed not to be merohedrally twinned.

Phasing was successfully performed by molecular replacement with PHASER (71) using the structure of the circularly permuted truncated Taspase1 (cpTaspase1_{α41-183/β}) as the search model upon removal of all solvent molecules. Phaser found a unique solution (TFZ 9.5 and LLG

434) with two monomers in the asymmetric unit consisting of the hetero-tetramer as previously reported for the crystallographic structures of the proenzyme and the truncated Taspase1(14). The structure was built and refined using REFMAC5 (72) and COOT (66). Extra positive electron density was clearly visible for the long fragment (Pro183-Asp233), that was missing in all previous structures, which was built and refined in later steps. As our crystals showed anisotropic diffraction, the initial refinement generated moderate refinement statistics, with R_{work} and R_{free} values of 32% and 36%, respectively, as well as poor electron density maps.

As shown in Figure S3A, the diffraction was highly anisotropic with the useful data extending only to a modest-low resolution (Table 1). Thus, structure determination of the cpTaspase1_{α41-233/β} was challenging. The TRUNCATE program (69) revealed two strong-diffracting directions (a^* and b^*) and one very weak-diffracting direction (c^*) (Figure S3B). Anisotropy analysis was addressed with the STARANISO server (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>) using the DEBYE and STARANISO programs to perform an anisotropic cut-off of merged intensity data, to perform Bayesian estimation of structure amplitudes, and to apply an anisotropic correction to the data. The diffraction images were re-processed with XSCALE/XDS (64) with no resolution cut-off applied; an *ahkl* output file containing unmerged and unscaled intensities was generated and was provided as input for the STARANISO server using the unmerged STARANISO protocol, which prevents the noisy measurements in the weakly diffracting directions from adversely affecting the scaling and error-model estimation at the final scaling/merging step, using the default cut-off level of 1.2 for the local $I/\sigma\langle I \rangle$. STARANISO indicated that the data were strongly anisotropic (the anisotropic ratio and S/N ratio were 1.474 and 121.60, respectively), which yielded an ellipsoidal resolution boundary with resolution limits of 2.9 Å, 2.9 Å, and 5.3 Å along three principal component a^* , b^* ,

and c^* axes, respectively. After the anisotropic correction, the final dataset showed a best-resolution limit of 3.04 Å along the a^* , b^* axes, and a lowest-resolution limit of 5.8 Å along the c^* axis. The axes-dependent truncated and corrected structure factors generated by STARANISO were then used to repeat the molecular replacement phasing using the initially refined model that resulted in significantly better solution (TFZ 37.1 and LLG 1070). This new model was refined with REFMAC5 with non-crystallographic symmetry (NCS) using local restraints. Density modification with PARROT program (73) using NCS averaging and solvent flattening was also used. The final model was refined to 3.04 Å resolution with R_{work} and R_{free} values of 25.5% and 31.7%, respectively. Moreover, the final model yielded electron density maps that were much more interpretable. The crystallographic data collection and structure refinement statistics are summarized in Table 1.

Structure alignments

The alignments between crystallographic structures were carried out by LSQKAB using the method described by Kabsch (74). The root-mean square deviations (RMSDs) excluding the long domain residues (Pro183-Asp233) were calculated for the $C\alpha$ and all atoms.

Bioinformatics analyses

The sequence of the full length human Taspase1 (UniProtKB accession number Q9H6P5) was further analyzed. Intrinsic Disorder analysis was performed using five web-based disorder prediction tools: PONDR-FIT (20), MetaDisorder (21), MoRFpred (40), SPOT-Disorder2 (23), and MobiDB (22). ExPASy-Protparam was used to calculate the frequency of occurrence of different amino acids (<https://web.expasy.org/protparam/>). Coiled-coil structural predictions were

performed using three available predictors from EXPASY server: COILS (25), DEEPCOIL (26), and MARCOIL (27).

Negative staining electron microscopy

Freshly purified cpTaspase1_{α41-233/β} protein was negative stained and examined by transmission electron microscopy (TEM) to assess cpTaspase1_{α41-233/β} protein homogeneity. Taspase1 at 0.52 μg/ml in the purification buffer (20 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol) was applied as 5 μl drops for 2 min at room temperature to carbon-coated 400 mesh copper grids, which were previously glow-discharged for 60 seconds. The excess liquid was removed by blotting then replaced with a drop (5 μl) of stain 0.75% (w/v) pH neutralized uranyl formate solution. Excess staining solution was removed promptly with filter paper and the staining step was repeated once more. Grids were air-dried at room temperature prior to data acquisition. Taspase1 particles were initially examined with a Philips CM12 transmission electron microscope operated at 120 kV. A second round of negative staining experiments was performed to obtain 2D-class averages. For these experiments Taspase1 was used at 1.6 μg/ml and examined with FEI Titan/ETEM microscope on a Gatan US1000XP CCD camera. A total of 69 micrographs were acquired with a 2.5 x 2.5 Å pixel and processed within the cisTEM software package (75). A total of 40,365 particles were used for statistical analysis and a total of 250 classes were generated and averaged in the cisTEM software suite.

Dynamic light scattering (DLS)

DLS measurements were performed on two different instruments for comparison, a SpectroSize 300 spectrometer instrument (Molecular Dimensions), and a DynaPro Nanostar instrument (Wyatt

Technology). Samples of cpTaspase1_{α41-233/β} were prepared at three different concentrations (2.5 mg/ml, 5 mg/ml, and 10 mg/ml) in the purification buffer (20 mM HEPES pH 8.0, 500 mM NaCl, 5% glycerol). Immediately before measurements protein solutions were centrifuged for 30 min at 14000 rpm, 4°C, in order to remove any aggregates and dust. Instruments software were used in data collection and processing. All measurements were carried out by averaging 10 accumulations of 20 sec each at 20 °C.

Data Availability

The structure models and structure factors for the cp-Taspase1_{α41-233/β} and cp-Taspase1_{α41-183/β} have been deposited at the wwPDB under accession codes PDB 6VIN, and PDB UGK, respectively.

Supplementary information

The supplementary information includes a detailed description of all the interfaces found in the crystallographic structure of cp-Taspase1_{α41-233/β}, Figures S1 to S23, and Tables S1 to S6.

Acknowledgements

This project has been funded in whole with Federal funds from the National Cancer Institute, National Institutes of Health, under Chemical Biology Consortium Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. We would like to recognize the protein expression, purification, and crystallization teams at Beryllium for

their assistance. We thank Dr. Barbara Mroczkowski, Dr. Joel Schneider, and Dr. Michele Zacks, for their careful scientific editing and proofreading.

Author contributions

J.M. Martin-Garcia, S. L. Delker, T. E. Edwards, P. Fromme, A. Flint, and J. Hsieh designed the experiments and provided with the most relevant scientific discussions. S. L. Delker and T. E. Edwards designed and cloned all proteins presented in this study. S. L. Delker and T. E. Edwards crystallized and determined the structure of the cp-Taspase1_{α41-183/β} protein variant. L. Sambucetti, M. Stofega, and J. Snider performed all catalytic activity assays of all proteins presented in this study. J.M. Martin-Garcia and P. Fromme designed all experiments of the cp-Taspase1_{α41-233/β} protein. J.M. Martin-Garcia, N. Nagaratnam, and R. Jernigan did the bioinformatics analyses. J.M. Martin-Garcia, N. Nagaratnam, and R. Jernigan crystallized and characterized crystals as well as collected data for the cp-Taspase1_{α41-233/β} protein. J.M. Martin-Garcia and D. Thifault processed anisotropic data of cp-Taspase1_{α41-233/β}. J.M. Martin-Garcia and P. Fromme determined and evaluated the structure of cp-Taspase1_{α41-233/β} protein. D. Williams, B. Nannenga and N. Nagaratnam prepared and imaged samples for EM experiments. D. Williams analyzed negative staining EM experiments of cp-Taspase1_{α41-233/β}. J.M. Martin-Garcia and N. Nagaratnam prepared all figures for the manuscript. J.M. Martin-Garcia, P. Fromme, N. Nagaratnam, and R. Jernigan wrote the manuscript with contributions from all co-authors.

Declaration of interests

The authors declare that they have no conflicts of interest.

References

1. S. Rakash, Role of proteases in cancer: A review. *Biotechnology and Molecular Biology Reviews* **7**, 90-101 (2012).
2. Y. Yang, H. Hong, Y. Zhang, W. Cai, Molecular Imaging of Proteases in Cancer. *Cancer Growth Metastasis* **2**, 13-27 (2009).
3. J. J. Hsieh, E. H. Cheng, S. J. Korsmeyer, Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**, 293-303 (2003).
4. J. J. Hsieh, P. Ernst, H. Erdjument-Bromage, P. Tempst, S. J. Korsmeyer, Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. *Mol Cell Biol* **23**, 186-194 (2003).
5. H. Liu, S. Takeda, E. H. Cheng, J. J. Hsieh, Biphasic MLL takes helm at cell cycle control: implications in human mixed lineage leukemia. *Cell Cycle* **7**, 428-435 (2008).
6. S. Takeda *et al.*, Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes Dev* **20**, 2397-2409 (2006).
7. F. Capotosti, J. J. Hsieh, W. Herr, Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways. *Mol Cell Biol* **27**, 7063-7072 (2007).
8. H. Zhou *et al.*, Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. *Mol Cell Biol* **26**, 2728-2735 (2006).
9. D. Wunsch *et al.*, Taspase1: a 'misunderstood' protease with translational cancer relevance. *Oncogene* **35**, 3351-3364 (2016).
10. C. Oinonen, J. Rouvinen, Structural comparison of Ntn-hydrolases. *Protein Sci* **9**, 2329-2337 (2000).
11. H. C. Guo, Q. Xu, D. Buckley, C. Guan, Crystal structures of Flavobacterium glycosylasparaginase. An N-terminal nucleophile hydrolase activated by intramolecular proteolysis. *J Biol Chem* **273**, 20205-20212 (1998).
12. C. Oinonen, R. Tikkanen, J. Rouvinen, L. Peltonen, Three-dimensional structure of human lysosomal aspartylglucosaminidase. *Nat Struct Biol* **2**, 1102-1108 (1995).
13. Q. Xu, D. Buckley, C. Guan, H. C. Guo, Structural insights into the mechanism of intramolecular proteolysis. *Cell* **98**, 651-661 (1999).
14. J. A. Khan, B. M. Dunn, L. Tong, Crystal structure of human Taspase1, a crucial protease regulating the function of MLL. *Structure* **13**, 1443-1452 (2005).
15. M. Beijer *et al.*, Na(+)/K(+) exchange switches the catalytic apparatus of potassium-dependent plant L-asparaginase. *Acta Crystallogr D Biol Crystallogr* **70**, 1854-1872 (2014).
16. J. Lubkowski, W. Chan, A. Wlodawer, Opportunistic complexes of E. coli L-asparaginases with citrate anions. *Sci Rep* **9**, 11070 (2019).
17. K. Michalska, D. Borek, A. Hernandez-Santoyo, M. Jaskolski, Crystal packing of plant-type L-asparaginase from Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **64**, 309-320 (2008).
18. J. van den Boom, F. Trusch, L. Hoppstock, C. Beuck, P. Bayer, Structural Characterization of the Loop at the Alpha-Subunit C-Terminus of the Mixed Lineage Leukemia Protein Activating Protease Taspase1. *PLoS One* **11**, e0151431 (2016).
19. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007).

20. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**, 996-1010 (2010).
21. L. P. Kozlowski, J. M. Bujnicki, MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111 (2012).
22. D. Piovesan *et al.*, MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res* **46**, D471-D476 (2018).
23. J. Hanson, K. Paliwal, Y. Zhou, Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. *J Chem Inf Model* **58**, 2369-2376 (2018).
24. E. Gasteiger *et al.*, "Protein Identification and Analysis Tools on the ExPASy Server" in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, Totowa, NJ, 2005), 10.1385/1-59259-890-0:571, pp. 571-607.
25. A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164 (1991).
26. J. Ludwiczak, A. Winski, K. Szczepaniak, V. Alva, S. Dunin-Horkawicz, DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics* **35**, 2790-2795 (2019).
27. M. Delorenzi, T. Speed, An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617-625 (2002).
28. H. L. Ammon *et al.*, Preliminary crystal structure of *Acinetobacter glutaminasificans* glutaminase-asparaginase. *J Biol Chem* **263**, 150-156 (1988).
29. H. P. Aung, M. Bocola, S. Schleper, K. H. Rohm, Dynamics of a mobile loop at the active site of *Escherichia coli* asparaginase. *Biochim Biophys Acta* **1481**, 349-359 (2000).
30. D. Borek, M. Jaskolski, Crystallization and preliminary crystallographic studies of a new L-asparaginase encoded by the *Escherichia coli* genome. *Acta Crystallogr D Biol Crystallogr* **56**, 1505-1507 (2000).
31. P. P. Ho *et al.*, Crystalline L-asparaginase from *Escherichia coli* B. I. Purification and chemical characterization. *J Biol Chem* **245**, 3708-3715 (1970).
32. K. Michalska, K. Brzezinski, M. Jaskolski, Crystal structure of isoaspartyl aminopeptidase in complex with L-aspartate. *J Biol Chem* **280**, 28484-28491 (2005).
33. K. Michalska, M. Jaskolski, Structural aspects of L-asparaginases, their friends and relations. *Acta Biochim Pol* **53**, 627-640 (2006).
34. A. Prahl, M. Pazgier, M. Hejazi, W. Lockau, J. Lubkowski, Structure of the isoaspartyl peptidase with L-asparaginase activity from *Escherichia coli*. *Acta Crystallogr D Biol Crystallogr* **60**, 1173-1176 (2004).
35. A. L. Swain, M. Jaskolski, D. Housset, J. K. Rao, A. Wlodawer, Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. *Proc Natl Acad Sci U S A* **90**, 1474-1478 (1993).
36. J. Xuan *et al.*, Crystal structure of glycosylasparaginase from *Flavobacterium meningosepticum*. *Protein Sci* **7**, 774-781 (1998).
37. W. Li *et al.*, Intramolecular Cleavage of the hASRGL1 Homodimer Occurs in Two Stages. *Biochemistry* **55**, 960-969 (2016).
38. Y. Wang, H. C. Guo, Two-step dimerization for autoproteolysis to activate glycosylasparaginase. *J Biol Chem* **278**, 3210-3219 (2003).

39. J. Nomme, Y. Su, A. Lavie, Elucidation of the specific function of the conserved threonine triad responsible for human L-asparaginase autocleavage and substrate hydrolysis. *J Mol Biol* **426**, 2471-2485 (2014).
40. F. M. Disfani *et al.*, MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75-83 (2012).
41. A. Mohan *et al.*, Analysis of molecular recognition features (MoRFs). *J Mol Biol* **362**, 1043-1059 (2006).
42. L. Almagor, I. S. Ufimtsev, A. Ayer, J. Li, W. I. Weis, Structural insights into the aPKC regulatory switch mechanism of the human cell polarity protein lethal giant larvae 2. *Proc Natl Acad Sci U S A* **116**, 10804-10812 (2019).
43. M. Strong *et al.*, Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, 8060-8065 (2006).
44. O. V. Tsodikov, C. Hou, C. T. Walsh, S. Garneau-Tsodikova, Crystal structure of O-methyltransferase CalO6 from the calicheamicin biosynthetic pathway: a case of challenging structure determination at low resolution. *BMC Struct Biol* **15**, 13 (2015).
45. K. Vidilaseris *et al.*, Asymmetry in catalysis by Thermotoga maritima membrane-bound pyrophosphatase demonstrated by a nonphosphorus allosteric inhibitor. *Sci Adv* **5**, eaav7574 (2019).
46. E. Conti, M. Uy, L. Leighton, G. Blobel, J. Kuriyan, Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. *Cell* **94**, 193-204 (1998).
47. M. R. Fontes, T. Teh, B. Kobe, Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. *J Mol Biol* **297**, 1183-1194 (2000).
48. A. Herold, R. Truant, H. Wiegand, B. R. Cullen, Determination of the functional domain organization of the importin alpha nuclear import factor. *J Cell Biol* **143**, 309-318 (1998).
49. C. Niu *et al.*, FUS-NLS/Transportin 1 complex structure provides insights into the nuclear targeting mechanism of FUS and the implications in ALS. *PLoS One* **7**, e47056 (2012).
50. T. Yoshizawa *et al.*, Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell* **173**, 693-705 e622 (2018).
51. Z. C. Zhang, Y. M. Chook, Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc Natl Acad Sci U S A* **109**, 12017-12021 (2012).
52. D. Y. Chen *et al.*, Taspase1 functions as a non-oncogene addiction protease that coordinates cancer cell proliferation and apoptosis. *Cancer Res* **70**, 5358-5367 (2010).
53. C. Bier *et al.*, Overexpression of the catalytically impaired Taspase1 T234V or Taspase1 D233A variants does not have a dominant negative effect in T(4;11) leukemia cells. *PLoS One* **7**, e34142 (2012).
54. C. Bier *et al.*, Allosteric inhibition of Taspase1's pathobiological activity by enforced dimerization in vivo. *FASEB J* **26**, 3421-3429 (2012).
55. M. R. Arkin, A. Whitty, The road less traveled: modulating signal transduction enzymes by inhibiting their protein-protein interactions. *Curr Opin Chem Biol* **13**, 284-290 (2009).
56. R. S. Srinivasan *et al.*, The synthetic peptide PFWT disrupts AF4-AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. *Leukemia* **18**, 1364-1372 (2004).

57. T. Tanaka, T. H. Rabbitts, Interfering with protein-protein interactions: potential for cancer therapy. *Cell Cycle* **7**, 1569-1574 (2008).
58. L. Weber, Patented inhibitors of p53-Mdm2 interaction (2006 - 2008). *Expert Opin Ther Pat* **20**, 179-191 (2010).
59. C. Wichmann *et al.*, Targeting the oligomerization domain of ETO interferes with RUNX1/ETO oncogenic activity in t(8;21)-positive leukemic cells. *Cancer Res* **67**, 2280-2289 (2007).
60. S. Sabiani *et al.*, Unraveling the Activation Mechanism of Taspase1 which Controls the Oncogenic AF4-MLL Fusion Protein. *EBioMedicine* **2**, 386-395 (2015).
61. J. van den Boom *et al.*, Peptidyl succinimidyl peptides as taspase 1 inhibitors. *Chembiochem* **15**, 2233-2237 (2014).
62. D. Y. Chen *et al.*, A pharmacologic inhibitor of the protease Taspase1 effectively inhibits breast and brain tumor growth. *Cancer Res* **72**, 736-746 (2012).
63. B. Kim, S. S. Tarchevskaya, A. Eggel, M. Vogel, T. S. Jardetzky, A time-resolved fluorescence resonance energy transfer assay suitable for high-throughput screening for inhibitors of immunoglobulin E-receptor interactions. *Anal Biochem* **431**, 84-89 (2012).
64. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
65. A. Vagin, A. Teplyakov, MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography* **30**, 1022-1025 (1997).
66. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallographica Section D* **66**, 486-501 (2010).
67. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
68. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).
69. S. French, K. Wilson, On the treatment of negative intensity observations. *Acta Crystallographica Section A* **34**, 517-525 (1978).
70. P. H. Zwart, Anomalous signal indicators in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **61**, 1437-1448 (2005).
71. A. J. McCoy, Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41 (2007).
72. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255 (1997).
73. K. Cowtan, Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470-478 (2010).
74. W. Kabsch, A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A* **32**, 922-923 (1976).
75. T. Grant, A. Rohou, N. Grigorieff, cisTEM, user-friendly software for single-particle image processing. *Elife* **7** (2018).

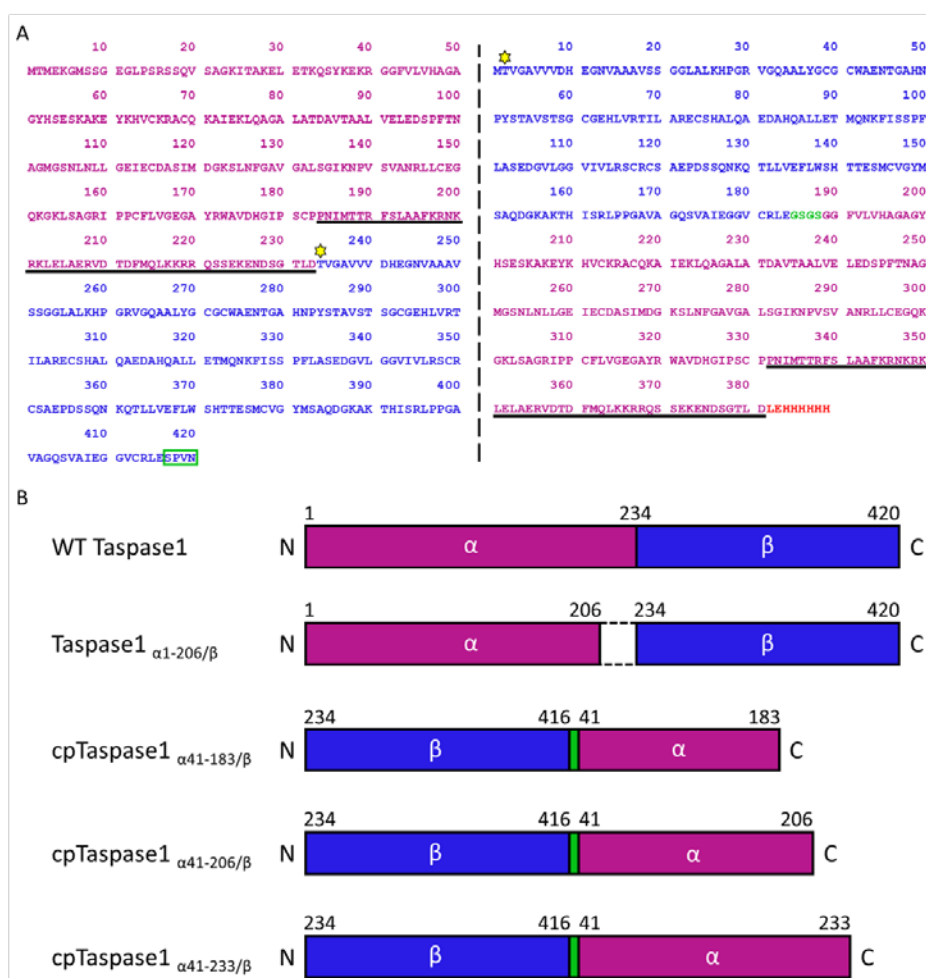


Figure 1. Sequence and domain organization of circularly permuted constructs. A) Comparison of the sequence of the circularly permuted construct, cpTaspase1_{α41-233/β} (right), with the sequence of the WT Taspase1 (left). The catalytic site nucleophile Thr234 is highlighted by the yellow star. The α and β subunits are colored in magenta and blue, respectively. In the circularly permuted construct, α and β subunits are swapped and linked through a GSGS linker (green). The GSGS residues replace the SPVN residues in the WT Taspase1 (green, boxed). The residues comprising the long C-terminal helix fragment are underlined. The first 40 residues of the alpha subunit are deleted except for the N-terminal initiating methionine. The two extra LE residues and the His (H) tag residues at the C-terminus (highlighted in red) are for cloning and for affinity purification purposes, respectively. B) Schematic representation of Taspase1 constructs. Amino acid positions in all constructs are numbered according to the wild type Taspase1 (WT Taspase1). The α and β subunits are highlighted in magenta and blue, respectively. Taspase1_{α1-206/β} construct (PDB 2A8J(14)) shows the two subunits are co-expressed with the missing fragment highlighted by the dashed lines. The three circularly permuted constructs (cp-Taspase1_{α41-183/β}, cp-Taspase1_{α41-206/β}, cp-Taspase1_{α41-233/β}) used in our study are shown: both subunits are swapped and linked through a GSGS tetrapeptide (green). These circularly permuted constructs differ from each other only in the length of the alpha subunit sequence.

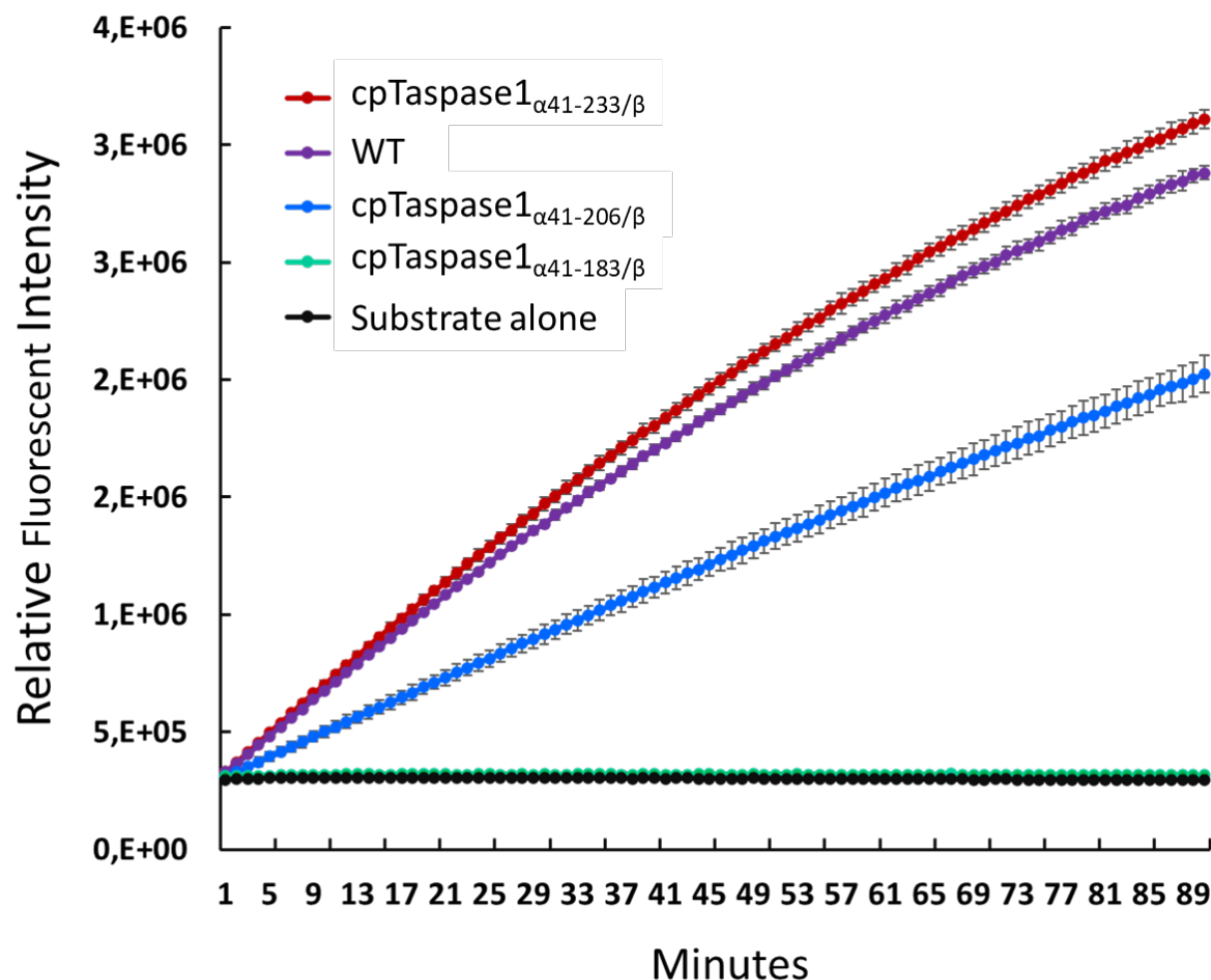


Figure 2. Activity assays of the circularly permuted constructs (cpTaspase1 α_{41-233}/β (red), cpTaspase1 α_{41-206}/β (blue), and cpTaspase1 α_{41-183}/β (green)), and the WT Taspase1 (purple). Reaction progress curves of the proteolytic cleavage of the labeled peptidic substrate MCA-Lys-Ile-Ser-Gln-Leu-Asp↓Gly-Val-Asp-Asp-Lys-(DNP)-NH₂ by various Taspase1 proteins. The experiment with the substrate alone (control) is represented in black. Error bars have been included.

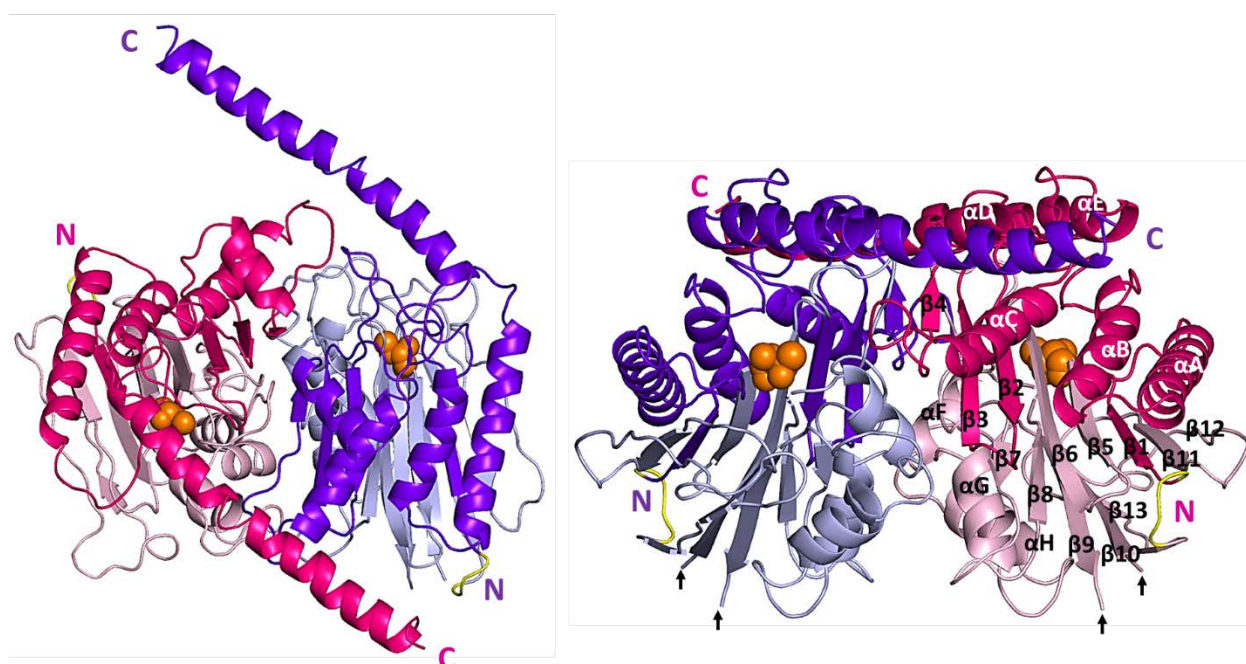


Figure 3. Crystal structure of cp-Taspase1_{α41-233/β}. Side (left) and top (right) views of the heterotetramer. Cartoon representation of the α subunit (pink) and β subunit (light pink) of chain A and α subunit (purple) and β subunit (light blue) chain B of the dimer. All secondary structure elements as well as the N- and C-termini are indicated. The newly identified long helical domain (helix αE), points away from the dimer. The nucleophile Thr234 is shown as orange spheres. The linker connecting the alpha and beta subunits is depicted in yellow. The missing fragment (Ser352-Gln362) is indicated by the arrows.

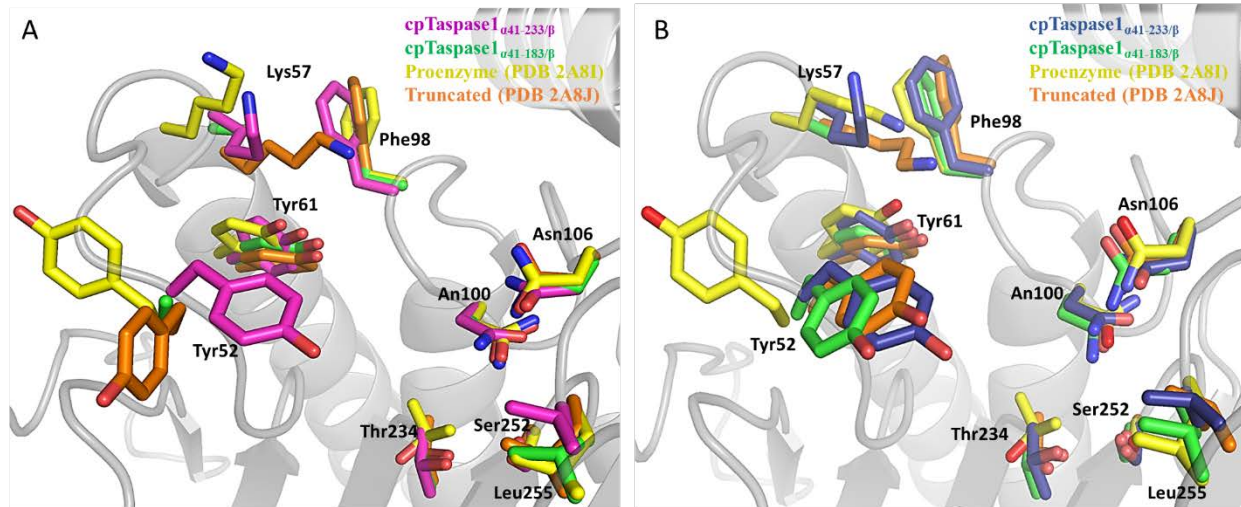


Figure 4. Comparison of the catalytic site of cpTaspase1 $_{\alpha 41-233/\beta}$ with other Taspase1 structures. Superimposition of chain A (A), and chain B (B) of cpTaspase1 $_{\alpha 41-233/\beta}$ to the proenzyme (PDB 2A8I (14)), the truncated Taspase1 (PDB 2A8J (14)), and the cpTaspase1 $_{\alpha 41-183/\beta}$. Residues important for the catalytic activity of Taspase1 are shown in stick representation.

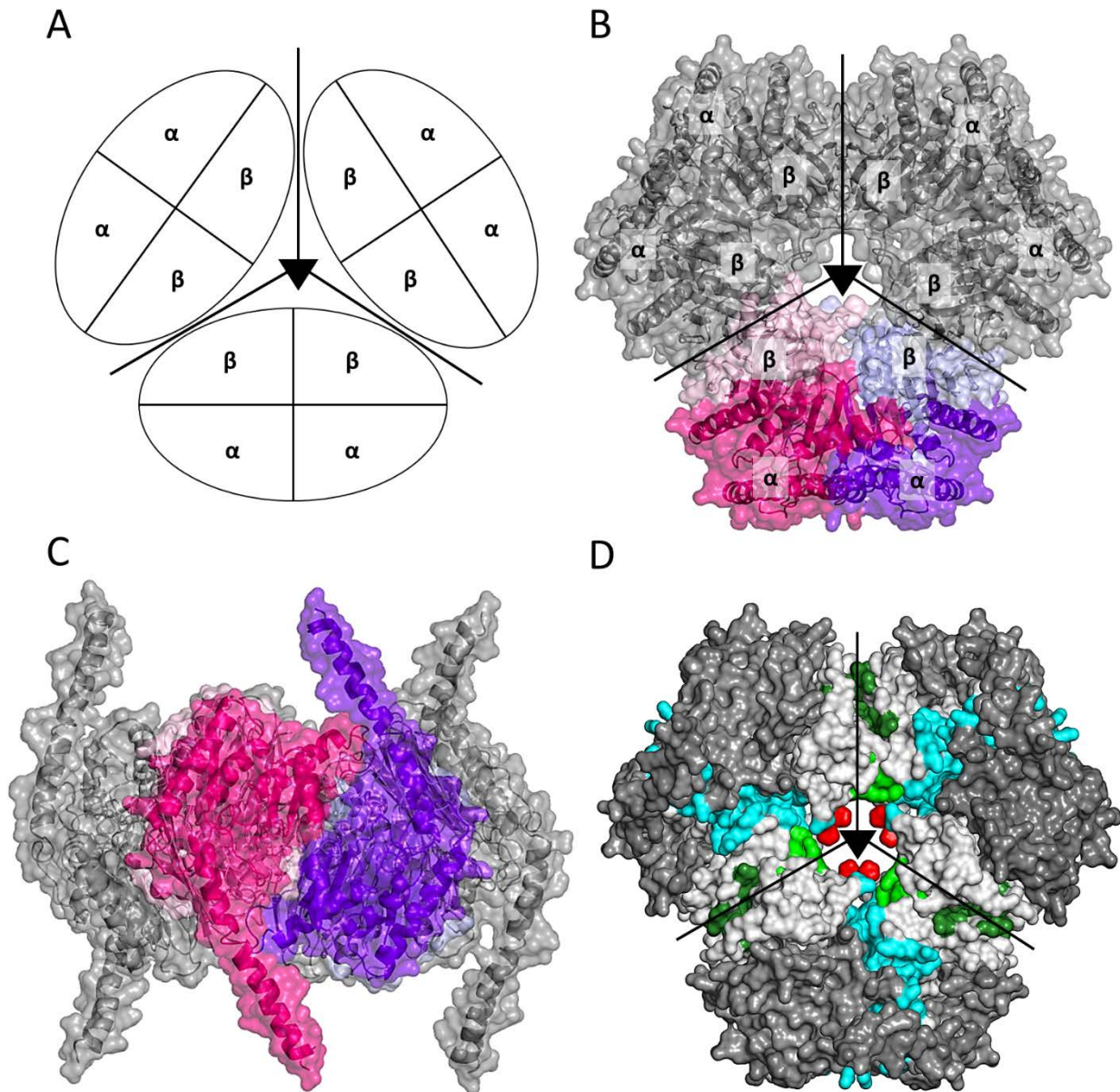


Figure 5. Subunit interfaces that maintain the single-ring assembly of cpTaspase1 _{α 41-233/ β} in the crystals. A) Schematic view of the single-ring structure. B) Top view of the surface representation of the cpTaspase1 _{α 41-233/ β} of the single-ring structure. C) Side view of the surface representation of the cpTaspase1 _{α 41-233/ β} of the double-ring structure D) Interfaces at the single-ring structure. The interfaces are colored as follows: cyan for the dimer interface, green for the intra-ring interface, and red for the histidine interface. The cpTaspase1 _{α 41-233/ β} dimers are represented with the same color as in Figure 3 and the dimers generated by symmetry are shown in a grey and black.

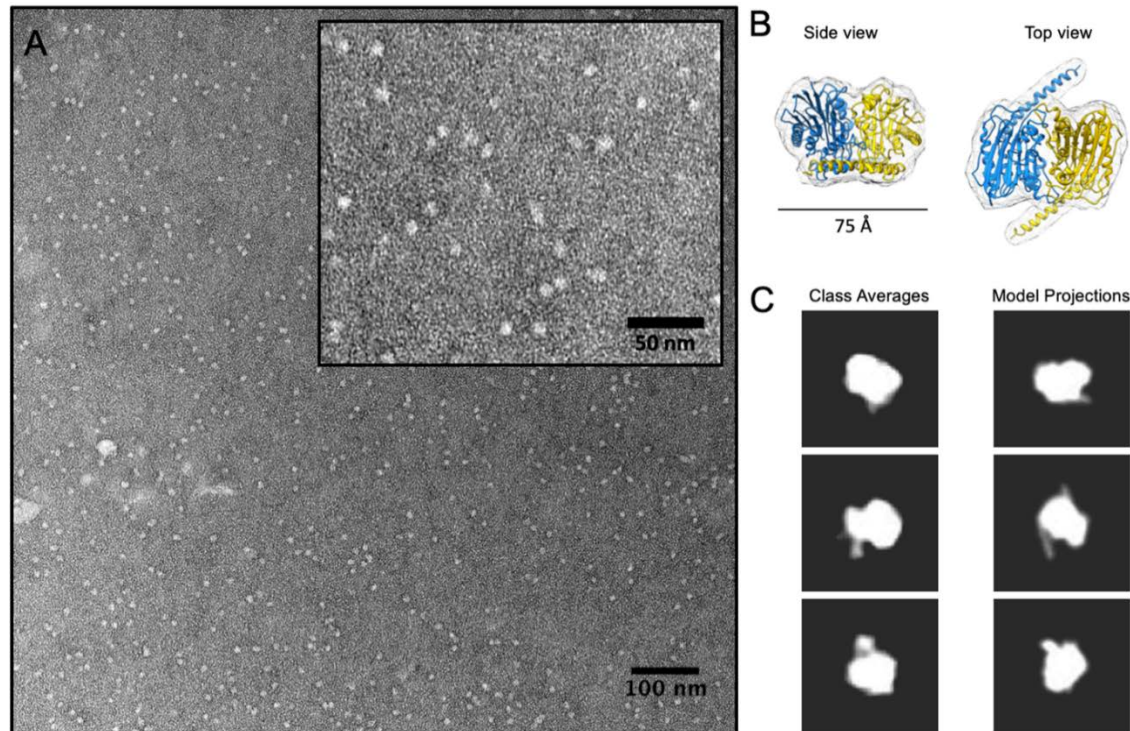


Figure 6. Negative stain electron microscopy (EM) of cpTaspase1 α 41-233/ β . A) EM micrograph showing a distribution of cpTaspase1 α 41-233/ β particles. Inset panel shows a closer view of the particles. Particles were estimated to be ~6-7 nm in size. B) Model of the cpTaspase1 α 41-233/ β dimer that was used to generate 15Å density map (shown as mesh around the model). C) Projection images were generated using the calculated map in B) and these were compared with 2D class averages calculated from the negative stain images collected from cpTaspase1 α 41-233/ β . Representative class averages and projection images of the model are shown, and it can be seen that the dimeric model matches well with the experimental class averages.

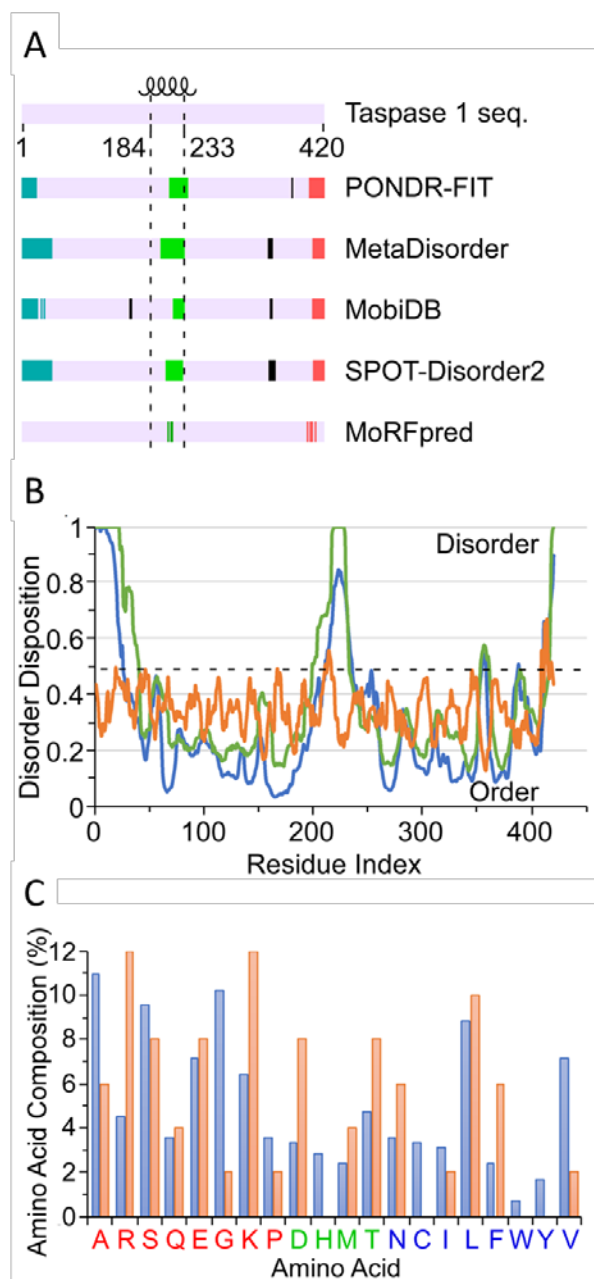


Figure 7: Prediction of intrinsically disordered regions (IDRs) of Taspase1. A) Disorder predictions in three discrete regions of Taspase1 (N-terminal (cyan), middle (green) and C-terminal (red)) by POND-R-FIT (20), MetaDisorder (21), MobiDB (22), MoRFpred (40), and SPOT-Disorder2 (23). Short disordered regions are represented in solid black. The location of the long helix fragment of Taspase1 is delimited by the black dotted line. B) Disorder disposition of each residue in Taspase1 sequence predicted by POND-R-FIT (blue), MetaDisorder (green) and MoRFpred (dark orange). Residues with disorder probability higher than 0.5 (black dotted line) are predicted to be disordered. C) Comparison of the composition of each amino acid in the long helix (coral orange) with full-length Taspase1 (blue). Disorder-promoting residues are red, order-promoting residues are blue and disorder-order neutral residues are green.

Table 1. Crystallographic data collection and refinement statistics

	cp-Taspase I _{α41-233/β}	cp-Taspase I _{α41-183/β}
Data collection statistics		
Radiation source	APS 23-ID-D	APS 21-ID-F
Wavelength (Å)	1.04	0.97
Detector	PILATUS3 6M	RAYONIX MX300HE
Temperature (K)	100	100
Space group	H32	P6 ₅
Unit cell dimensions		
a, b, c (Å)	196.00, 196.00, 196.91	60.30, 60.30, 319.01
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution range (Å)	3.04-49.05 (3.04-3.11)	2.15-40.41 (2.15- 2.20)
I/δ(I)	17.0 (1.9)	15.24 (3.74)
Multiplicity	19.6 (20.4)	5.8 (5.8)
Completeness (%)	99.8 (99.4)	99.8 (85.8)
CC _{1/2} (%)	100 (91.1)	99.8 (85.8)
R _{merge} (%)	6.8 (187.5)	7.8 (52.0)
Total number reflections	343,892	206,406
No. of unique reflections	17,964	35,513
Ellipsoidal truncation		
Ellipsoidal ^a resolution (Å) (direction) ^b	2.9 (0.894a* – 0.447b*) 2.9 (b*) 4.8 (c*)	N/A
Best/Worst diffraction limits after cut-off (Å)	3.04 / 5.3	N/A
I/δ(I) (ellipsoidal ^a) ^c	21.2 (1.5)	N/A
Completeness (%) (spherical ^a) ^{c,d}	54.4 (13.7)	N/A
Completeness (%) (ellipsoidal ^a) ^{c,d}	93.9 (65.5)	N/A
Total no. of reflections (ellipsoidal ^a) ^c	299,919	N/A
No. of unique reflections (ellipsoidal ^a) ^c	15,315	N/A
Data refinement statistics for truncated data		
Number of reflections used in refinement	13,749	35,417
Number of free reflections used in refinement	1,568	1,709
R _{work} / R _{free} (%)	23.7 / 30.7	15.9 / 19.7
Wilson B-factor (Å ²)	152.0	43.5
Solvent content (%) / Matthews coefficient (V _m)	70 / 4.3	48 / 2.4
No. of atoms		
Protein	10,510	4,368
Water and others (ligands or ions)	0	189
Overall B factor (Å ²)	165.0	43.48
R.M.S. deviations from ideal values		
Bonds lengths (Å)	0.007	0.007

Bond angles (°)	1.744	0.827
Ramachandran plot statistics (%)		
Favored	85	96.55
Allowed	12	2.80
Outliers	3	0
Molprobrity score / percentile	3.5 / 38 th	1.55 / 99 th
PDB	6VIN	6UGK

^aThese statistics are for data that were truncated by STARANISO to remove poorly measured reflections affected by anisotropy.

^bThe resolution limits for three directions in reciprocal space are indicated here. To accomplish this, STARANISO computed an ellipsoid postfitted by least squares to the cutoff surface, removing points where the fit was poor. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

^cValues in parentheses are for the highest-resolution shell. Note that data collection and refinement statistics have different highest-resolution shells. ^dThe anisotropic completeness was obtained by least squares fitting an ellipsoid to the reciprocal lattice points at the cutoff surface defined by a local mean $I/\sigma I$ threshold of 1.2, rejecting outliers in the fit due to spurious deviations (including any cusp), and calculating the fraction of observed data lying inside the ellipsoid so defined. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

Table 2. Interfaces in Cp-Taspase1_{α41-233/β} calculated by PDBePISA(19)

Interface #	Name of Interface	Interface residues component 1	Interface residues component 2	BSA* (Å ²)	Hydrogen bonds	Salt bridges	ΔG Kcal/mol	CSS [#]
1	Dimer	60	59	2039.1	21	7	-23.8	1.000
2	Inter-single ring	42	43	1355.1	13	2	-9.5	1.000
3	Helix	19	19	648.2	2	0	-6.6	0.069
4	Intra-double ring	16	16	535.7	0	0	-9.7	0.000
5	Linker	8	8	168.0	0	2	-1.7	0.026
6	Histidines	1	1	36.1	0	2	0.1	0.006
7		1	1	33.6	0	2	0.1	0.006

*Buried surface area (BSA).

[#]Complex formation significance score (CSS).