Structure and mechanism of a mitochondrial AAA+ disaggregase CLBP

Damu Wu[1], Yan Liu[2], Yuhao Dai[1], Guopeng Wang[1], Guoliang Lu[2], Yan Chen[1], Ningning Li[1], Jinzhong Lin[2], Ning Gao[1]

[1]State Key Laboratory of Membrane Biology, Peking-Tsinghua Joint Center for Life Sciences, School of Life Sciences, Peking University, Beijing 100871, China.

[2]State Key Laboratory of Genetic Engineering, School of Life Sciences, Zhongshan Hospital, Fudan University, Shanghai 200438, China

*Correspondence to: gaon@pku.edu.cn

Abstract

The human AAA+ ATPase CLPB (aka, HSP78 and Skd3) is a protein disaggregase and functions to promote the solubilization of proteins in the mitochondrial intermembrane space. Unlike other AAA+ protein unfoldases or disaggregases, CLPB contains an ankyrin repeat containing domain (ANK) at its N-terminus. Mutations of CLPB are closely associated with a few human diseases, such as 3-methylglutaconic aciduria (3-MGA) and severe congenital neutropenia (SCN). The mechanism of CLPB functions as a disaggregase and the role of its unique ANK domain are currently unclear. Herein, we report a comprehensive structural characterization of human CLPB in both the apo- and substrate-bound states. Our data show that CLPB assembles into homo-tetradecamers in apo-state and dodecamers in the presence of a substrate, mediated by ANK domains via a “head-to-head” organization. Conserved pore-loops (PLs) on the ATPase domains form a spiral staircase to grasp and translocate the substrate in a step-size of two amino acid residues. We also show that the ANK domain is functionally essential, and more important, responsible for direct binding to various mitocondiral proteins in the inner membrane or intermembrane space. These results suggest that CLPB functions as a general disaggregase in mitochondria and highlight its potential as a target for the treatment of various mitochondria-related diseases.

Introduction

Protein misfolding and aberrant aggregation are devastating to many fundamental processes in the cell and failures to remediate them are closed related to many human diseases (Eisele et al., 2015; Horwich, 2002). To maintain a healthy proteome, cells have evolved multiple dedicated systems, one of which is the Hsp100/Clp chaperone family. Hsp100/Clp chaperones are AAA+ ATPases that are widely found in bacteria, protozoa, fungi and plants (Kedzierska-Mieszkowska and Zolkiewski, 2021). HSP100 proteins generally contain an N-terminal domain (NTD), one or two ATPase domains (or nucleotide-binding domain, NBD), and usually function in hexameric forms. Taking yeast Hsp104 as an example, the NTD is responsible for substrate binding and recognition (Kedzierska-Mieszkowska and Zolkiewski, 2021), while the NBD1 and NBD2 bind and hydrolyze ATP to power substrate unfolding and translocation (Erzberger and Berger, 2006; Lee et al., 2013). Similar to many other AAA+ ATPases, Hsp104 unfolds and transports substrates through its central pore by a ratchet-like motion of the highly conserved pore-loops (PLs) on it ATPase domain (Gates et al., 2017). The ATP-hydrolysis cycle dependent conformational change of each subunit results in both inter- and intra-subunit structural remodeling, which collectively lead to the threading and unidirectional movement of peptide within the central (Gates et al., 2017).

Recently, a new type of HSP100 family proteins, CLPB (aka, HSP78 and Skd3) was reported to act as a protein disaggregase in the intermembrane space of mitochondria (Cupo and Shorter, 2020; Mroz et al., 2020; Thevarajan et al., 2020; Wortmann et al., 2015). The N-terminus of CLPB coding sequence has a mitochondrial targeting signal (MTS), followed by an ankyrin repeat (ANK) domain, and ends with a C-terminal NBD (Fig. 1a). There is a short hydrophobic stretch between the MTS and the first ankyrin motif, as well as a long linker helix (LH) between ANK domain and NBD (Fig. 1a). The MTS is cleaved by mitochondrial processing peptidase (MPP), followed by a second cleavage by presenilin-associated rhomboid-like (PARL) protease by removing another 35 residues from the N-terminus (Mroz et al., 2020; Saita et al., 2017). The ANK domain is a unique feature of CLPB compared to other AAA+ ATPases, and it is also the first protein that was reported to contain both ANK domain and AAA+ activity (Li et al., 2006). Importantly, the removal of ANK domain disrupts the disaggregase activity of CLPB (Cupo and Shorter, 2020).

Dysfunction of CLPB was found to be closely related to several human diseases. A number of *CLPB* mutations were found in patients of 3-methylglutaconic aciduria (3-MGA) (Capo-Chichi et al., 2015; Kanabus et al., 2015; Kiykim et al., 2016; Saunders et al., 2015; Wortmann et al., 2015). A common feature of these patients is increased urinary 3-methylglutaric acid excretion, with varying degree of microcephaly, small birth weight, neutropenia, severe encephalopathy, intellectual disability, movement disorder and cataracts (Capo-Chichi et al., 2015; Kanabus et al., 2015; Kiykim et al., 2016; Saunders et al., 2015; Wortmann et al., 2015). Moreover, heterozygous variants of CLPB were also identified in severe congenital neutropenia (SCN) that disrupts granulocyte differentiation of human

hematopoietic progenitors and increase apoptosis (Warren et al., 2022). Many of these mutations have been shown to impair the CLPB disaggregase activity, such as T268M, R408G, R475Q, A591V, R650P in 3-MGA and N499K, E557K, R561G, R620C in SCN (Cupo and Shorter, 2020; Warren et al., 2022). At cellular level, CLPB is important to mitochondria-mediated apoptosis and maintenance of the mitochondrial cristae structure (Chen et al., 2019b). CLPB interacts with HAX1, the anti-apoptotic factor of BCL-2 family, to promote cell survival (Chen et al., 2019b; Wortmann et al., 2015). Upregulated cellular level of CLPB in acute myeloid leukemia (AML) patients was found to mediate the resistance to BCL2 inhibitor venetoclax (Chen et al., 2019b). Recently, CLPB was also found to have negative correlation with the progression-free survival in castration-resistant prostate cancer (Pudova et al., 2020).

It is currently not clear why CLBP has such a broad role in many aspects of mitochondrial function, and how the ANK and NBD domains work together to catalyze its essential disaggregase function. Here, we present a comprehensive structural characterization of human CLPB in both the apo and substrate-bound state. Unexpectedly, we found that CLPB assembles into homo-tetradecamers in apo-state and dodecamers in the presence of a substrate, mediated by ANK domains via a “head-to-head” organization. While the CLPB works in a highly conserved manner to grasp and translocate substrate using the aromatic PL on its NBD, the ANK domain is responsible for recruiting a variety of substrates, including those proteins en route to the matrix. Together, our results suggest that CLPB functions as a general disaggregase in mitochondria and highlight the possibility of targeting CLPB for therapeutic purposes.

Results

CLPB forms higher-order double-heptamers or double-hexamers

The coding sequence of human *CLPB* was amplified from a home-made cDNA library of HEK-293T cells, and was revealed to be the isoform 2 (Uniprot identifier: Q9H078-2). First, CLPB without the mitochondrial targeting sequence (MTS) (Wortmann et al., 2015), named CLPB-N92 (Figure 1a), was expressed and purified from *E. coli* cells. Consistent with a previous study (Mroz et al., 2020), CLPB-N92 formed large aggregates and appeared to be highly heterogenous in size as shown by the negative-staining electron microscopy (nsEM) (Supplementary Fig. 1a-b). CLPB was reported to undergo another proteolytic step of maturation by PARL, and the predicted cleavage site is between C126 and Y 127(Saita et al., 2017). Notably, a hydrophobic stretch (S) within the region removed by PARL (Figure 1a) was reported to inhibit the disaggregase activity of CLPB (Cupo and Shorter, 2020). Therefore, as a validation experiment, we expressed the full-length of CLPB exogenously in HEK-293T cells, and examined the size of protein products. Indeed, two bands are clearly visible, and the precursor form gradually decreases over time (Supplementary Fig. 1c). Next, another construct harboring the sequence of CLPB starting from PARL cleave site (named

CLPB-N127) was made and used for preparation of CLPB complexes from *E. coli* cells (Supplementary Fig. 1D).

Interestingly, both the gel filtration and nsEM showed that the purified CLPB complex is in a higher oligomeric state rather than a hexamer as expected from other typical AAA+ unfoldases and disaggregases (Supplementary Fig. 2a-b). And this organization is irrelevant of exogenously supplemented AMPPNP. Despite of this unusual observation, these CLPB complexes do possess strong ATPase and disaggregase activities (Supplementary Fig. 1e-f), indicating that they are functional forms. CLPB complexes were then subjected to cryo-EM analysis, which revealed that most CLPB complexes are tetradecamers in a double-heptamer form. Only 1.3% of particles show a hexameric top view (Figure 1b). The side view of CLPB complex displays a four-layered architecture, with two tightly associated inner layers and two slightly curved outer layers. 3D classification indicated that these double-heptamers are extremely diverse in conformation, involving both inter- and intra-heptamer structural variances. With an exhausted manner of classification strategies (See method), we could only push the resolution to a range of 6-7 Å from a relatively homogenous class of particles (Supplementary Fig. 3). In this map, secondary structures were fairly resolved in certain regions, which allowed an accurate rigid-body fitting of the CLPB model predicted by alpha-fold (Jumper et al., 2021). CLPB subunits display a pseudo-D7 symmetry with an open channel in the center (24 Å in diameter) (Figure 1c). The overall size of the structure is 175 Å and 160 Å in height and width, respectively. The "head-to-head" organization of the two heptamers is mediated by their N-terminal ANK domains, and the two AAA+ ATPase rings locate at the ends of the complex (Fig. 1c). Although a heptamer form of ATPase modules is very rare among AAA+ ATPases, another member of mitochondrial AAA+ ATPases Bcs1 was recently reported to function also in homo-heptamer forms (Kater et al., 2020; Tang et al., 2020).

Since double-heptamer contains no substrate and is wide open in the central pore, we hope to capture a CLPB complex with co-purified substrates in the central channel by introducing a Walker B mutation (E425Q) to the NBD domain of CLPB. This mutation would abolish ATP hydrolysis but not ATP binding. Notably, this strategy has been shown to be successful in obtaining substrate-bound forms for many types of AAA+ ATPases (Fei et al., 2020; Lo et al., 2019; Twomey et al., 2019). This mutant version of $CLPB^{E425Q}$ was similarly purified from *E. coli* cells (Supplementary Fig. 2d-f) and analyzed by cryo-EM (Supplementary Fig. 4). Similar to the wild type CLPB complex, the mutant form is also highly dynamic and could only be refined to a global resolution of 7.9 Å， and the resolution of the NBD-ring could be improved to 5.2 Å (Supplementary Fig. 5). However, a sharp difference is that the mutant form of CLPB complexes is a dodecamer (double-hexamer). Top views of 2D classification showed a hexameric ring exclusively (Fig. 1b). Compared with the WT complex, while the width is reduced to 135 Å, the overall height has increased to about 200 Å (Fig. 1d). This result indicates that upon binding to the substrate in the central channel (Supplementary Fig.

6), CLPB subunits take a more compact form to organize its higher-order structure. In particular, the NBDs adopt an upward spiral positions around the substrate polypeptide, in a counter-clockwise rotation pattern when viewed from the ANK domain direction, resulting in an asymmetric helical structure (Supplementary Fig. 6). Not surprisingly, this arrangement is highly similar to that of other substrate-bound AAA+ ATPases (Cooney et al., 2019; Gates et al., 2017; Twomey et al., 2019).

In short, unlike most AAA+ ATPases, CLPB assemble into homo-tetradecamers in apo-state and dodecamers in the presence of substrate. The formation of the double heptamer or double hexamer is mediated by the CLPB-specific ANK domain upstream the AAA+ domain.

ANK domain is essential for the assembly and activity of CLPB complexes

To confirm the role in ANK domain in the assembly of CLPB complexes, we created an NBD-only CLPB variant ($CLPB^{NBD}$) by removing the N-terminal ANK domain sequence. Both the gel filtration and nsEM confirmed that the removal of ANK domain led to the formation of much smaller sized CLPB particles (Supplementary Fig. 7). The ANK domain of CLPB was also proposed to be responsible for substrate recognition (Cupo and Shorter, 2020). A published work showed that removal of the ANK domain from CLPB abolished both the ATPase and disaggregase activities (Cupo and Shorter, 2020). Since ANK doman is far from the ATPase center, this result appeared to be unusual. Therefore, we tested a few truncations of CLPB in these two activities (Figure 2 a-b). Our results show that CLPB variants retain its ATPase activity as long as an intact NBD is included, and the NBD alone ($CLPB^{NBD}$) displays an even higher ATPase activity. In contrast, the disaggregase activity critically depends on the integrity of both domains, the E425Q mutation ($CLPB^{E425Q}$, $CLPB^{NBD/E425Q}$) or the removal of ANK domain ($CLPB^{NBD}$) resulted in a complete loss of the activity (Fig. 2c-d). As a reference for comparison, a point mutation Y400A in the conserved PL of the ATPase domain also abolished the disaggregase activity, and significantly reduced the ATPase activity as well. In general, our results of these functional domains or motifs are consistent with previous studies (Cupo and Shorter, 2020), except that we showed that the ATPase activity is not perturbed by ANK domain removal.

Therefore, there data demonstrate the essential role of CLPB-specific ANK domain in both the higher-order organization and chaperone function of CLPB.

CLPB with ANK domain truncated assembles into helical structures

Due to the dynamic arrangement of the two heptamers or two hexamers, the ANK domain truncated form might be helpful towards high-resolution structural determination. Therefore, two mutant versions of CLBP (NBD and NBD^{E425Q}) were analyzed by cryo-EM (Supplementary Fig. 7a-b). 2D classification results confirmed that E425 mutation led to the binding of an endogenous peptide in the channel and the top views display hexameric features

exclusively (Supplementary Fig. 9). In contrast, top views of NBD^{WT} show both hexameric and heptameric arrangements (Supplementary Fig. 9). Consequently, we focused on the NBD^{E425Q} dataset for high-resolution analysis. After several rounds of 3D classification, three different oligomeric arrangements, hexameric, heptameric and nonameric were identified. The nonamer appeared to be more stable and could be resolved at an overall resolution of 3.7 Å (Supplementary Fig. 8), allowing the atomic modelling of the NBD of CLPB. The central channel of all three states contain the density of a substrate (Fig. 3a-c), and CLPB protomers clearly form a counter-clockwise helical structure, viewing from the ANK doamin (Fig. 3a-c). In the heptameric complex, an additional protomer (protomer 7, P7), is on the top of P1, whereas in the nonamer, 9 protomers form a 1.5-turn helix around the central substrate. Obviously, this is the effect of removing ANK domain, because the presence of the N-terminal ANK domain would block the association of the seventh CLPB subunit to the hexamer (Fig. 1d).

In addition to the variation in the number of protomers, these three oligomers also have some obvious structural differences. While the twist angle of subunits is roughly 60° in all three forms, their axial rises are markedly different (Fig. 3d). The axial rise of the conserved PL between P1 and P6 of the hexamer is ~24 Å, and is increased to 29 Å and 32 Å for the heptamer and nonamer, respectively (Fig. 3d). This is as expected, as more axial space is required to fit in the seventh and more subunits onto the hexamer.

In summary, these data show that the ANK domain is important for the formation of the correct oligomeric state of CLPB. Although this nonameric helical structure does not exist in a physiological context, it provides important structural insights into how the ATPase module of CLPB binds the substrate.

Conserved PLs interact with the substrate in the central channel

In the nucleotide-binding pockets of the nonameric structure, a total of eight ATP molecules could be modelled (Supplementary Fig. 10). The conserved functional motifs on the ATPase domain are well resolved, including the Walker A motif (351-GSSGIGKT-358), sensor-1 (464-TSN-466), and sensor-2 (588-GAR-590) (Fig. 4e). The conserved residues I317, I318 and F541 form a hydrophobic pocket to stabilize the adenine ring of ATP, while T358 stabilizes the β- and γ-phosphates through a Mg^{2+} ion (Fig. 4e). The Arginine finger (R531) from the adjacent protomer points to ATP and has a strong interaction with the γ-phosphate.

In addition to these conserved features, the substrate binding property of CLPB also exhibits a high conservation. The density of the polypeptide backbone is well resolved and modeled as a 17-residue poly-alanine fragment (Fig. 4a-c). The peptide spans about 53 Å length in the central pore (Fig. 4b). The PL-Is and PL-IIs form an upward spiral staircase around the substrate (Fig. 4c). For individual protomers, a canonical PL-I (399-GYVG-402) binds to the peptide through two hydrogen bonds formed between the main-chain atoms. The first is between the carbonyl oxygen of the substrate residue N and the main-chain nitrogen of

V401, and the other between the carbonyl oxygen of G399 and the main-chain nitrogen of the substrate residue N+2 (Fig. 4d). This pattern of interactions is nearly identical for all protomers, except that the hydrogen bond distances vary between 2.5-3.6 Å. Thus, the axial step size of PLs is exactly two residues. Since we and others showed that the conserved tyrosine in the PL-I is essential for the disaggregase activity (Fig. 2c-d) (Cupo and Shorter, 2020), the side-chain of Y400 could also play a role in interacting with the substrate sidechains or it is more required for substrate translocation (Fig. 4d) during an ATP-hydrolysis cycle.

In contrast, the PL-II, consisting of E386, R387, and H388, is further away from the substrate backbone, except the side-chain of H388 is within 4-Å distance with the β-carbon atom of the substrate. The polar residues of the PL-II likely functions to interact with different sidechains on the substrate during substrate processing.

Overall, these results indicate that the PLs of CLPB grasp and move the substrate to exert its disaggregase activity, in a conserved manner as classic AAA+ ATPases.

A **unique insertion of the ANK domain contributes to disaggregase activity of CLPB**

A typical ankyrin repeat motif is about 33 amino acid residues and exhibits a helix-turn-helix conformation (Li et al., 2006) (Fig. 5a). The ANK domain of CLPB contains 4 copies of ankyrin motif, and a unique insertion between ANK2 and ANK3 (Fig. 5a). The isoform 1 of CLPB (Uniprot: Q9H078-1) differs from the isoform 2 exactly in this region, with 30 residues more in the insertion (Fig. 5a). The AlphaFold prediction suggests that this insertion is highly flexible. Since the ANK domain is important for both the assembly and function of CLPB, we sought to test whether this insertion has isoform-specific functions. Based on the structures of the double hexamer, some extra density around the ankyrin motifs, extending towards the central channel and with a direct contact with the substrate, could be found (Supplementary Fig. 11a-b). This implies that this insertion may involve in substrate binding and transport. To explore this possibility, we created an additional CLPB mutant with the insertion deleted ($CLPB^{\Delta 201-232}$) (Fig. 5b). This mutant was then compared with both the isoform 1 and 2 in their ATPase and disaggregase activities. First, both the mutant $CLPB^{\Delta 201-232}$ and the isoform 1 showed the same oligomeric state as the isoform 2. Second, the two isoforms have comparable activities in these functions and the isoform 1 appears to be slightly more active (Fig. 5c). In contrast, while the deletion in $CLPB^{\Delta 201-232}$ has marginal impact on the ATPase activity, it significantly reduces disaggregase activity by more than 50% (Fig. 5c-d). These results indicate that this flexible insertion is of functional importance. Taking into consideration of axial movement of CLPB subunits during an ATP-hydrolysis cycle, it is likely that this insertion directly involves in substrate recognition and transport to the C-terminal AAA+ ring.

The ANK domain directly interacts with mitochondrial substrates of CLPB

Previous data showed that CLPB knockout cells exhibit decreased solubility of many proteins in the inner membrane and intermembrane space of mitochondria (Cupo and Shorter, 2020). Given the essentiality of ANK domain in the disaggregase activity of CLPB, we did a Mass Spectrometry (MS) based profiling of direct binders of ANK domain in HEK-293T cells. A plasmid harboring the MTS and ANK domain sequences were expressed in HEK-293T cells (Fig. 6a). Mitochondria were then isolated, lysed and dissolved in 1% detergent. The ANK domain and its binders were purified through a C-terminal Strep tag. Using DAVID Bioinformatics Resources (Huang et al., 2009; Huang et al., 2007) to analyze cellular components, 869 proteins localized to mitochondria and significantly enriched were selected for subsequent analysis (Supplementary File1). Among these captured mitochondrial proteins, most are located in the inner membrane and matrix (59.2%). Around 40 proteins from the intermembrane space were also detected (Fig. 6b). It was previously shown that CLPB could increase the solubility of HAX1, HTRA2, PHB1/2, SLC25 family and STOML2 (Cupo and Shorter, 2020). Indeed, these proteins were detected in our MS data (Fig. 6b). We also found that ANK domain interacts with OPA1, which is consistent with a previous report (Chen et al., 2019b) that in AML CLPB mediates venetoclax resistance by maintaining mitochondrial cristae structure through its interaction with OPA1 and HAX1.

The enriched proteins are not only from pathways of mitochondrial organization and apoptosis, but also mitochondrial protein translation (Fig. 5d), including many mito-ribosomal proteins. Since these proteins are located in the matrix, it is likely CLPB may contribute to the transport of these proteins from the cytosol, through the inter-membrane space, to the matrix. The KEGG pathway analysis showed that in addition to the apoptosis and amino acid metabolism pathways (valine, leucine and isoleucine), some proteins associated with neurodegenerative diseases are also enriched (Fig. 5e). Notably, CLPB is not only highly expressed in testis but also in brain and skeletal muscle (Saunders et al., 2015) (NCBI, Illumina bodyMap2 transcriptome, Bioproject: PRJEB2445). This may explain why intellectual disability/developmental delay and movement disorders are always associated with CLPB-mutation-related diseases such as 3-MGA (Wortmann et al., 2015).

In short, our data suggests that the ANK domain of CLPB is a direct binder of substrates, and CLBP has a very broad function in maintaining a healthy mitochondrial proteome, thus contributes to various functional aspects of mitochondria.

Discussion

In the present work, we characterized the structures of CLPB in both the apo and substrate-bound states. Unexpectedly, we found that CLPB assembles into a higher-order structure, in the form of double-heptamer or double-hexamer, through inter-molecular interactions between the N-terminal ANK domains (Fig. 1). We and others also showed that the ANK domain is essential for the disaggregase acitivty of CLPB (Cupo and Shorter, 2020) (Fig. 2),

and we further demonstrated that the deletion of a unique insertion in the ANK domain also could result in a more than 50% reduction on the disaggregase activity (Fig. 5), suggesting that the ANK domain may function to recognize and pass the substrate to the distal ATPase ring. In the double-hexamer, both hexamers are in an active conformation with the substrate being threaded through the central pores of their ATPase rings (Supplementary Fig. 6). Therefore, this "head to head" organization indicates that the two functional units of CLPB may work in a collaborative manner to perform the disaggregase activity. This reminds us of the recent findings on the double-hexamer form of p97/Cdc48 (Gao et al., 2022; Hoq et al., 2021), although they utilize a "back to back" interface through two C-terminal ATPase rings. Therefore, these two examples demonstrate that AAA+ proteins could adapt to different higher-order assemblies to serve different purposes.

Beside these unique features of CLPB, the mechanisms of substrate unfolding and translocation appear to be highly conserved, similar to many members of AAA+ ATPases. Our structures of $CLPB^{NBD}$ helical assemblies show that the aromatic PLs on the ATPase domains interact with substrate in a sequence-independent manner, forming a staircase around the substrate. The step-size between two PLs is exactly two residues. This arrangement is similar to what we see in a rope climbing exercise except that a functional unit of CLPB have more than two hands. Moreover, we have tried to model the substrate in both directions (N to C or C to N) in the substrate density, and both configurations could satisfy this repeated patter of interaction, indicating that CLPB could unfold substrate in both directions, which is similar to ClpXP and ClpAP that translocate the substrate in either the N to C or C to N direction (Olivares et al., 2017).

Importantly, the structures of CLPB allow us to map the disease related mutations (Capo-Chichi et al., 2015; Cupo and Shorter, 2020; Kanabus et al., 2015; Kiykim et al., 2016; Pronicka et al., 2017; Saunders et al., 2015; Wortmann et al., 2021; Wortmann et al., 2015) on the atomic model to understand their possible effect on the function of CLPB. Mutations associated with 3-MGA could be categorized into three types. The first type lies at the interface between two adjacent CLPB protomers, including R378G, R445Q, Y587C, R598C and E609K (Fig. 7a). These mutations likely perturb the inter-protomer communication to impair CLPB function. The second type consists of M381I, C456R, E471K, and Y537 in the large subdomain, and A561V, G616V, R620P, and I652N in the small subdomain (Fig. 7b). These mutations are away from the inter-protomer or inter-domain interfaces, and could potentially destabilize the domain structure (Fig. 7b). These mutations severely disrupt disaggregase activity, although some have less effect on ATPase activity (like R620P) (Cupo and Shorter, 2020). The third type of mutations are on the ANK domain, in the ankyrin motif 3. Besides two nonsense mutations (R250* and K321*), the equivalent positions of other three mutations (T238M, A239T and Y242C) are on the first helix of ANK3 (Fig. 7c). This region is highly conserved among ankyrin motifs, and a tetrapeptide (-TPLH-) motif (238-TALH-241 in CLPB) is a characteristic of ankyrin motif (Li et al., 2006). Therefore, these

mutations might destabilize the structure of ANK domain, or lead to misfolding of the ankyrin motifs. Indeed, T238M mutation does not affect the ATPase activity, but inhibits the disaggregase activity (Cupo and Shorter, 2020).

As for mutations related to SCN, they are exclusively located near the ATP-binding pocket of NBD (Fig. 7d). The mutations are distributed in sensor-1 (N466K), sensor-2 (R590C), AFs (R531Q/G), Walker A (T358K), and the conserved residue (E527K). These amino acids play important roles in the binding and hydrolysis of ATP (Fig. 4e) (Yedidi et al., 2017; Zhang et al., 2021; Zhang and Mao, 2020). Indeed, when these mutations were introduced, not only the disaggregase activity was abolished, but the ATPase activity was significantly impaired (Warren et al., 2022).

Furthermore, we analyzed the substrate candidates of CLPB through an affinity-purification based mass spectrometry. Only the ANK domain was used as bait, and therefore the hits may represent the direct binders of the ANK domain. Many of these candidates have been reported to interact with CLPB in previously large-scale proteomics studies (Botham et al., 2019; Huttlin et al., 2021), and several of them have been reported that CLPB can promote their solubility, such as HAX1, HTRA2, PHB1/2, SLC25A3/A4/A6, STOM L2, and OPA1 (Cupo and Shorter, 2020). The presence of many matrix proteins in the target list suggests these proteins likely rely on CLPB for stability and protection on the way to the matrix. Therefore, it suggests that CLPB is a chaperone of general purpose in mitochondria, highlighting its potential as a target to treat various mitochondrion-related disorders.

Methods

Protein expression and purification

The coding sequence (isoform 2) of human *CLPB*, amplified from a home-made cDNA library of HEK-293T cells, was cloned into pET28a vector with a N-terminal His6-SUMO tag followed by a TEV protease cleavage sequence, and expressed in *E. coli* Transetta (DE3) cells (Trans). CLPB protein variants were purified as reported previously (Mroz et al., 2020), with some modifications. Briefly, the culture was induced with 1mM IPTG at OD600 of 0.6-0.8 at 20°C overnight. Cells were harvested and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 1% Triton X-100 and 1mM PMSF) and lysed by sonication. Lysates were cleared by centrifugation in a JA-25.50 rotor (Beckman) for 30 min at 20,000 r.p.m. and the supernatants were precipitated with ammonium sulphate at 35% saturation, then incubated for 10 min at room temperature. The cloudy lysates were centrifuged for 30 min at 15,000 r.p.m. Dissolved the protein pellet with Buffer B (50 mM Tris-HCl, pH 8.5) supplemented with 5 mM DTT and centrifugation again. The solutions were loaded onto a Mono Q column (GE Healthcare). Proteins were eluted with a linear gradient to 60% Buffer C (50 mM Tris-HCl, pH 8.5, 1M NaCl). To remove His6-SUMO tags, 5 mM DTT and TEV protease were added into the eluates, and then incubated at RT for 2 hours. The tag and contaminations were removed by size-exclusion chromatography, which was pre-equilibrated

with 20 mM HEPES-KOH, pH 6.8, 300 mM KCl, 5 mM $MgCl_2$. Peak fractions were pooled and concentrated, then incubated with 5 mM nucleotide (AMPPNP for CLPB, ATP for $CLPB^{E425Q}$) for 2h at RT. 1 mM BS3 was added to improve the sample stability, and then loaded onto size-exclusion chromatography. Peak fractions were first examined by nsEM, then pooled and concentrated. 1mM nucleotide was added before vitrification. $CLPB^{Y400A}$ was prepared by site-specific mutagenesis, and purified as described above.

The NBD and NBD^{E425Q} sequences were cloned by site-specific mutagenesis. Harvested cells were resuspended in lysis buffer II (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 5% glycerol, 1% Triton X-100, 1mM PMSF, and 20 mM imidazole) and lysed by sonication. The supernatants were incubated with Ni–NTA agarose beads (GE Healthcare) for 2 h at 4°C. After washing 5 times with lysis buffer, protein was eluted with elution buffer (50 mM Tris-HCl, pH 7.4, 300 mM NaCl, 5 mM $MgCl_2$, and 500 mM imidazole). 5 mM DTT and TEV protease were add into the eluates to remove His6-SUMO tags, and then incubated at RT for 2 hours. Then the proteins were loaded onto size-exclusion chromatography, which was pre-equilibrated with 20 mM HEPES-KOH, pH 6.8, 300 mM KCl, 5 mM $MgCl_2$. Peak fractions were pooled and concentrated, then incubated with 5 mM nucleotide (AMPPNP for NBD, ATP for NBD^{E425Q}) for 2h at RT, and then loaded onto size-exclusion chromatography. Peak fractions were first examined by nsEM, then pooled and concentrated. 1mM nucleotide was added before vitrification.

For the ANK domain purification, cells were resuspended in lysis buffer III (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 1% Triton X-100, 1mM PMSF, and 20 mM imidazole). The supernatants were incubated with Ni–NTA agarose beads (GE Healthcare) for 2 h at 4°C. The eluates were loaded onto size-exclusion chromatography, and eluted with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl.

Cryo-EM specimen preparation

The sample homogeneity was first examined by nsEM. Proteins were applied onto carbon-coated copper grids, and examined with an FEI Tecnai T20 TEM at 120 kV. During specimen preparation for cryo-EM, we found that CLPB and $CLPB^{E425Q}$ had strong preferred orientation. 0.5 mM CHAPSO was added to reduce the preferred orientation (Chen et al., 2019a). All cryo-grids (R1.2/1.3, Au300, Quantifoil) were prepared with an FEI Vitrobot Mark IV with the chamber set at 6°C and 100% humidity with a blot force of -1, blot time 5 s, wait time 20 s for CLPB and $CLPB^{E425Q}$, and a blot force of -1, blot time 1s, wait time 4s for NBD and NBD^{E425Q}. Grids were screened with an FEI Talos Arctica (FEI Ceta camera) before data collection.

Cryo-EM data collection image processing for CLPB

For the sample of CLPB, micrographs were recorded with 300 kV FEI Titan Krios (Gatan K3 direct electron detector) with GIF Quantum energy filter (Gatan) at a nominal magnification of 81,000X and at a calibrated pixel size of 1.157 Å. 40 frames with exposure time of 3.2 s were collected with defocus ranged from -1 to -1.4 μm under low-

dose condition. A total of 9,341 micrographs from two datasets were collected with EPU. The micrographs were subjected to beam-induced motion correction using MotionCor2 (Zheng et al., 2017), and contrast transfer function (CTF) parameters for each micrograph were determined by Gctf (Zhang, 2016). RELION (versions 3.1) was used to perform data processing (Zivanov et al., 2020). After auto-picking, the particles were subjected to 2D classification and several round 3D classifications. 299,306 particles from two datasets were selected and applied for the next round of 3D classification (K=4), a final set of 231,212 particles were kept for 3D refinement. For double-heptamer, another round of 3D classification (K=6) without alignment was performed to improve the resolution. Finally, 153,005 particles were kept for the final 3D refinement (final resolution 6.8 Å). Due to the inter-heptamer flexibility, mask-based (only include a single heptamer) classification (K=6) and refinement was applied, which showed no significant improvement (Supplementary Fig. 3). In addition, multiple rounds of focused classification based on the NBD-ring or a selected combination of CLPB subunits using symmetry expansion derived half-particles were also tried, with different setting of parameters, however, none of them was able to reach better than 5 Å resolution.

For the sample of $CLPB^{E425Q}$, micrographs were recorded with 300 kV FEI Titan Krios (Gatan K2 camera) with GIF Quantum energy filter (Gatan) at a nominal magnification of 36,000X and at calibrated pixel sizes of 1.052 Å. 32 frames with exposure time of 8 s were collected with defocus ranged from -1.2 to -1.8 μm under low-dose condition. A total of 4574 micrographs were collected with SerialEM (Mastronarde, 2005). After 2D classification and several round 3D classifications, a 7.5-Å density map was achieved (Supplementary Fig. 4). In Addition, a density map of 5.2-Å could be achieved by applying a local mask of the NBD-ring during reconstruction (Supplementary Fig. 5).

For the sample of NBD^{E425Q}, micrographs were recorded with 300 kV FEI Titan Krios G3i (Gatan K3 direct electron detector) with GIF Quantum energy filter (Gatan) at a magnification of 64,000X and at pixel size of 1.08 Å. 32 frames with exposure time of 2.44 s were collected with defocus ranged from -1 to -1.4 μm. A total of 4574 micrographs were collected. 3,753,541 particles were split into 2 parts to facilitate processing. After several rounds of 2D classification and 3D classification, 3 major classes were found: hexamer, heptamer, and nonamer. A small group of hexamer particles were used for 3D refinement, resulting in a density map of 7.4 Å. The particles of heptamer and nonamer were subjected additional rounds of 3D classification, and after CTF refinement and Bayesian polishing, were reconstructed to the final resolutions of 4.1 Å and 3.7 Å, respectively.

Model building and refinement

Model building was based on the prediction structure of CLPB by alpha-fold (Jumper et al., 2021). The model was first docked into the cryo-EM density map using UCSF Chimera (Pettersen et al., 2004), and rebuilt manually in Coot (Emsley and Cowtan, 2004). Figure

preparation and structure analysis were performed with and UCSF ChimeraX (Pettersen et al., 2021) and Chimera (Pettersen et al., 2004).

ATPase activity measurement

ATPase activity was determined as reported previously (Cupo and Shorter, 2020), with some modifications. 0.1 μM CLPB and CLPB variants were incubated with 1 mM ATP at 30°C for 30 min in ATPase activity buffer (AAB: 20 mM HEPES-KOH, pH 6.8, 150 mM KAOc, 5 mM $MgCl_2$, 0.1% Tween-20). ATPase activity was measured by using the Malachite green phosphate assay (Lanzetta et al., 1979). For each protein at least two biological replicates were measured with three independent replicates.

Luciferase disaggregation assay

Disaggregase activity was measured as previously reported (Cupo and Shorter, 2020; Mroz et al., 2020). Briefly, 1 μM CLPB and CLPB variants were incubated with 50 nM firefly luciferase aggregates at 30°C for 90 min in luciferase reactivation buffer (LRB: 20 mM HEPES-KOH, pH 8.0, 150 mM KAOc, 10 mM KAOc, 10 mM DTT) supplemented with 5 mM ATP. Notably, the disaggregase activity of CLPB was detected using the luciferase reporter assay kit (Beyotime, CAT# RG005) alone. Others were detected with the luciferase reporter assay kit (Trans, CAT# FR101-01).

Western blots

HEK-293 T cells were transfected with pCAG vector containing CLPB variants with a C-terminal Strep. Cells were incubated with lysis buffer B (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.1 % SDS and 1mM PMSF) for 30 min on ice. α-Strep antibody (Huaxingbio, CAT# HX1816) was used to detect the target protein.

Mitochondrial isolation and mass spectrometry

Mitochondria were isolated as previously described (Cupo and Shorter, 2020; Frezza et al., 2007). Briefly, 1 liter cells were resuspended in 80 mL mitochondria isolation buffer (MIB: 20 mM HEPES-KOH, pH 7.6, 250 mM sucrose and 2 mM EDTA) and homogenized with a Dounce homogenizer at 4°C. Lysate was then centrifuged at 1,300g for 5 min. The supernatant was collected and then centrifuged at 13,000g for 15 min twice. And the pellet was resuspended with lysis buffer C (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Triton X-100, and 1mM PMSF) and incubated for 30 min at 4°C. Lysate was then centrifuged at 20,000g for 30 min. The supernatant was collected and incubated with Strep resin for 2 h at 4°C. After extensively washed, SDS-loading buffer was added to the sample. The samples were separated by MOS-PAGE on a gradient gel (4%-20%).

Acknowledgments

We thank the Core Facilities at the School of Life Sciences, Peking University for help with negative staining EM; the Electron Microscopy Laboratory and Cryo-EM Platform for help with data collection; the High-performance Computing Platform for help with computation; the National Centre for Protein Sciences at Peking University for assistance with

protein preparation and mass spectrum. The work was supported by the National Science Foundation of China (31725007 to N.G., 31922036 to N.L.), the National Key Research and Development Program of China (2019YFA0508904 to N.G.), and the Qidong-SLS Innovation Fund to N.G.

Author contributions

D.M.W. prepared the samples with the help of Y.L., Y.H.D. and G.L.L.. D.M.W. collected the cryo-EM data. D.M.W. performed EM analysis with the help of G.P.W. and Y.C. N.G., D.M.W. performed model building. N.G., D.M.W. wrote the manuscript.

Competing interests

Authors declare no competing interests.

Correspondence and requests for materials should be addressed to N.G.

Reference

Botham, A., Coyaud, E., Nirmalanandhan, V.S., Gronda, M., Hurren, R., Maclean, N., St-Germain, J., Mirali, S., Laurent, E., Raught, B., *et al.* (2019). Global Interactome Mapping of Mitochondrial Intermembrane Space Proteases Identifies a Novel Function for HTRA2. Proteomics *19*, e1900139.

Capo-Chichi, J.M., Boissel, S., Brustein, E., Pickles, S., Fallet-Bianco, C., Nassif, C., Patry, L., Dobrzeniecka, S., Liao, M., Labuda, D., *et al.* (2015). Disruption of CLPB is associated with congenital microcephaly, severe encephalopathy and 3-methylglutaconic aciduria. J Med Genet *52*, 303-311.

Chen, J., Noble, A.J., Kang, J.Y., and Darst, S.A. (2019a). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. J Struct Biol X *1*.

Chen, X., Glytsou, C., Zhou, H., Narang, S., Reyna, D.E., Lopez, A., Sakellaropoulos, T., Gong, Y., Kloetgen, A., Yap, Y.S., *et al.* (2019b). Targeting Mitochondrial Structure Sensitizes Acute Myeloid Leukemia to Venetoclax Treatment. Cancer Discov *9*, 890-909.

Cooney, I., Han, H., Stewart, M.G., Carson, R.H., Hansen, D.T., Iwasa, J.H., Price, J.C., Hill, C.P., and Shen, P.S. (2019). Structure of the Cdc48 segregase in the act of unfolding an authentic substrate. Science *365*, 502-505.

Cupo, R.R., and Shorter, J. (2020). Skd3 (human ClpB) is a potent mitochondrial protein disaggregase that is inactivated by 3-methylglutaconic aciduria-linked mutations. Elife *9*.

Eisele, Y.S., Monteiro, C., Fearns, C., Encalada, S.E., Wiseman, R.L., Powers, E.T., and Kelly, J.W. (2015). Targeting protein aggregation for the treatment of degenerative diseases. Nat Rev Drug Discov *14*, 759-780.

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

Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct *35*, 93-114.

Fei, X., Bell, T.A., Jenni, S., Stinson, B.M., Baker, T.A., Harrison, S.C., and Sauer, R.T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. Elife *9*.

Frezza, C., Cipolat, S., and Scorrano, L. (2007). Organelle isolation: functional mitochondria from mouse liver, muscle and cultured fibroblasts. Nat Protoc *2*, 287-295.

Gao, H., Li, F., Ji, Z., Shi, Z., Li, Y., and Yu, H. (2022). Cryo-EM structures of human p97 double hexamer capture potentiated ATPase-competent state. Cell Discov *8*, 19.

Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E., *et al.* (2017). Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. Science *357*, 273-279.

Hoq, M.R., Vago, F.S., Li, K., Kovaliov, M., Nicholas, R.J., Huryn, D.M., Wipf, P., Jiang, W., and Thompson, D.H. (2021). Affinity Capture of p97 with Small-Molecule Ligand Bait Reveals a 3.6 A Double-Hexamer Cryoelectron Microscopy Structure. ACS Nano *15*, 8376-8385.

Horwich, A. (2002). Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. J Clin Invest *110*, 1221-1232.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44-57.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., *et al.* (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res *35*, W169-175.

Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., *et al.* (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell *184*, 3022-3040 e3028.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

Kanabus, M., Shahni, R., Saldanha, J.W., Murphy, E., Plagnol, V., Hoff, W.V., Heales, S., and Rahman, S. (2015). Bi-allelic CLPB mutations cause cataract, renal cysts, nephrocalcinosis and 3-methylglutaconic aciduria, a novel disorder of mitochondrial protein disaggregation. J Inherit Metab Dis *38*, 211-219.

Kater, L., Wagener, N., Berninghausen, O., Becker, T., Neupert, W., and Beckmann, R. (2020). Structure of the Bcs1 AAA-ATPase suggests an airlock-like translocation mechanism for folded proteins. Nat Struct Mol Biol *27*, 142-149.

Kedzierska-Mieszkowska, S., and Zolkiewski, M. (2021). Hsp100 Molecular Chaperone ClpB and Its Role in Virulence of Bacterial Pathogens. Int J Mol Sci *22*.

Kiykim, A., Garncarz, W., Karakoc-Aydiner, E., Ozen, A., Kiykim, E., Yesil, G., Boztug, K., and Baris, S. (2016). Novel CLPB mutation in a patient with 3-methylglutaconic aciduria causing severe neurological involvement and congenital neutropenia. Clin Immunol *165*, 1-3.

Lanzetta, P.A., Alvarez, L.J., Reinach, P.S., and Candia, O.A. (1979). An improved assay for nanomole amounts of inorganic phosphate. Anal Biochem *100*, 95-97.

Lee, J., Kim, J.H., Biter, A.B., Sielaff, B., Lee, S., and Tsai, F.T. (2013). Heat shock protein (Hsp) 70 is an activator of the Hsp104 motor. Proc Natl Acad Sci U S A *110*, 8513-8518.

Li, J., Mahajan, A., and Tsai, M.D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry *45*, 15168-15178.

Lo, Y.H., Sobhany, M., Hsu, A.L., Ford, B.L., Krahn, J.M., Borgnia, M.J., and Stanley, R.E. (2019). Cryo-EM structure of the essential ribosome assembly AAA-ATPase Rix7. Nat Commun *10*, 513.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol *152*, 36-51.

Mroz, D., Wyszkowski, H., Szablewski, T., Zawieracz, K., Dutkiewicz, R., Bury, K., Wortmann, S.B., Wevers, R.A., and Zietkiewicz, S. (2020). CLPB (caseinolytic peptidase B homolog), the first mitochondrial protein refoldase associated with human disease. Biochim Biophys Acta Gen Subj *1864*, 129512.

Olivares, A.O., Kotamarthi, H.C., Stein, B.J., Sauer, R.T., and Baker, T.A. (2017). Effect of directional pulling on mechanical protein degradation by ATP-dependent proteolytic machines. Proc Natl Acad Sci U S A *114*, E6306-E6313.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci *30*, 70-82.

Pronicka, E., Ropacka-Lesiak, M., Trubicka, J., Pajdowska, M., Linke, M., Ostergaard, E., Saunders, C., Horsch, S., van Karnebeek, C., Yaplito-Lee, J., *et al.* (2017). A scoring system predicting the clinical course of CLPB defect based on the foetal and neonatal presentation of 31 patients. J Inherit Metab Dis

40, 853-860.

Pudova, E.A., Krasnov, G.S., Kobelyatskaya, A.A., Savvateeva, M.V., Fedorova, M.S., Pavlov, V.S., Nyushko, K.M., Kaprin, A.D., Alekseev, B.Y., Trofimov, D.Y., *et al.* (2020). Gene Expression Changes and Associated Pathways Involved in the Progression of Prostate Cancer Advanced Stages. Front Genet *11*, 613162.

Saita, S., Nolte, H., Fiedler, K.U., Kashkar, H., Venne, A.S., Zahedi, R.P., Kruger, M., and Langer, T. (2017). PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. Nat Cell Biol *19*, 318-328.

Saunders, C., Smith, L., Wibrand, F., Ravn, K., Bross, P., Thiffault, I., Christensen, M., Atherton, A., Farrow, E., Miller, N., *et al.* (2015). CLPB variants associated with autosomal-recessive mitochondrial disorder with cataract, neutropenia, epilepsy, and methylglutaconic aciduria. Am J Hum Genet *96*, 258-265.

Tang, W.K., Borgnia, M.J., Hsu, A.L., Esser, L., Fox, T., de Val, N., and Xia, D. (2020). Structures of AAA protein translocase Bcs1 suggest translocation mechanism of a folded protein. Nat Struct Mol Biol *27*, 202-209.

Thevarajan, I., Zolkiewski, M., and Zolkiewska, A. (2020). Human CLPB forms ATP-dependent complexes in the mitochondrial intermembrane space. Int J Biochem Cell Biol *127*, 105841.

Twomey, E.C., Ji, Z., Wales, T.E., Bodnar, N.O., Ficarro, S.B., Marto, J.A., Engen, J.R., and Rapoport, T.A. (2019). Substrate processing by the Cdc48 ATPase complex is initiated by ubiquitin unfolding. Science *365*.

Warren, J.T., Cupo, R.R., Wattanasirakul, P., Spencer, D.H., Locke, A.E., Makaryan, V., Bolyard, A.A., Kelley, M.L., Kingston, N.L., Shorter, J., *et al.* (2022). Heterozygous variants of CLPB are a cause of severe congenital neutropenia. Blood *139*, 779-791.

Wortmann, S.B., Zietkiewicz, S., Guerrero-Castillo, S., Feichtinger, R.G., Wagner, M., Russell, J., Ellaway, C., Mroz, D., Wyszkowski, H., Weis, D., *et al.* (2021). Neutropenia and intellectual disability are hallmarks of biallelic and de novo CLPB deficiency. Genet Med *23*, 1705-1714.

Wortmann, S.B., Zietkiewicz, S., Kousi, M., Szklarczyk, R., Haack, T.B., Gersting, S.W., Muntau, A.C., Rakovic, A., Renkema, G.H., Rodenburg, R.J., *et al.* (2015). CLPB mutations cause 3-methylglutaconic aciduria, progressive brain atrophy, intellectual disability, congenital neutropenia, cataracts, movement disorder. Am J Hum Genet *96*, 245-257.

Yedidi, R.S., Wendler, P., and Enenkel, C. (2017). AAA-ATPases in Protein Degradation. Front Mol Biosci *4*, 42.

Zhang, G., Li, S., Cheng, K.W., and Chou, T.F. (2021). AAA ATPases as therapeutic targets: Structure, functions, and small-molecule inhibitors. Eur J Med Chem *219*, 113446.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J Struct Biol *193*, 1-12.

Zhang, S., and Mao, Y. (2020). AAA+ ATPases in Protein Degradation: Structures, Functions and Mechanisms. Biomolecules *10*.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods *14*, 331-332.

Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253-267.

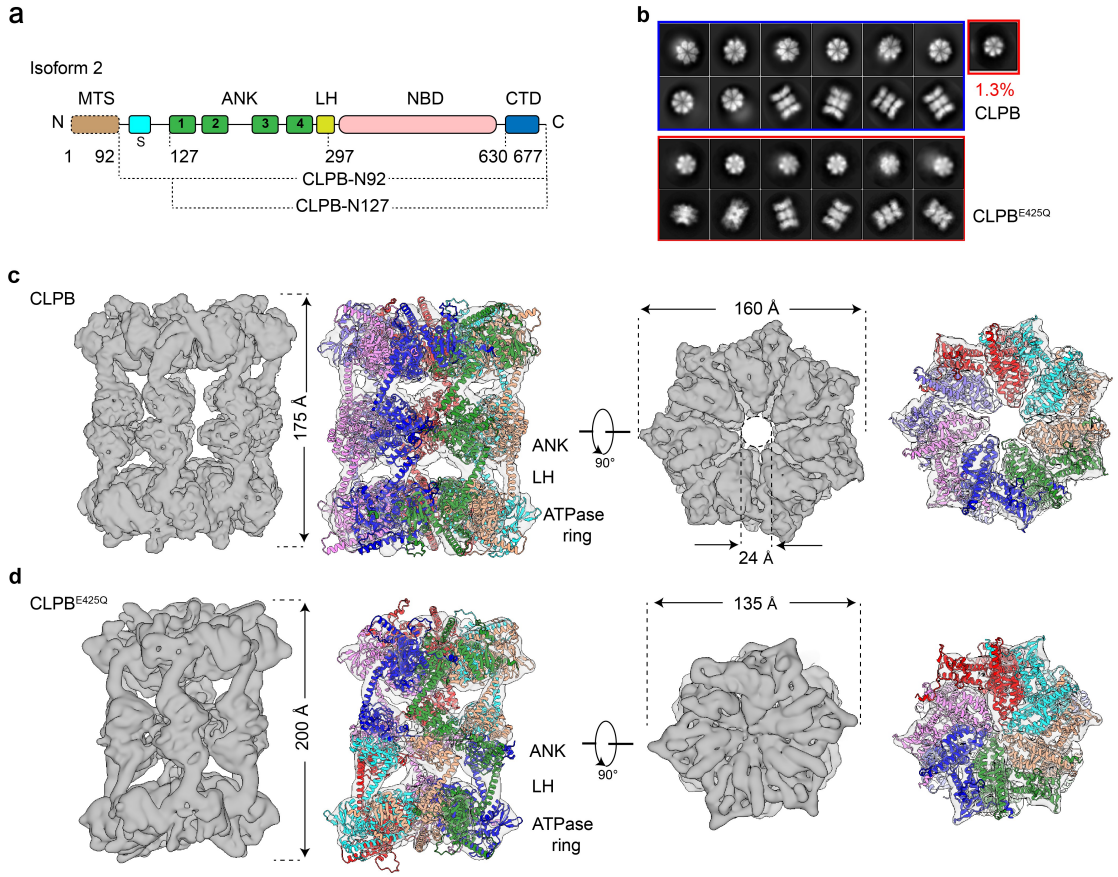


Fig. 1 Cryo-EM structures of CLPB double-heptamers in the apo-state and CLPB^{E425Q} double-hexamers in the substrate-bound state.

a, Domain organization of *H. sapiens* CLPB. CLPB is composed of an MTS, a short hydrophobic stretch (S), a linker helix (LH), four ankyrin-repeat (ANK) motifs, an NBD, and a CTD.

b, Representative 2D-classification averages of CLPB and CLPB^{E425Q} dataset.

c-d, The density maps of the double-heptamer and double-hexamer, superimposed with the predicted models of CLPB. The density maps are shown in the side and top (ATPase ring) views. The higher-order oligomer is mediated by ANK domains.

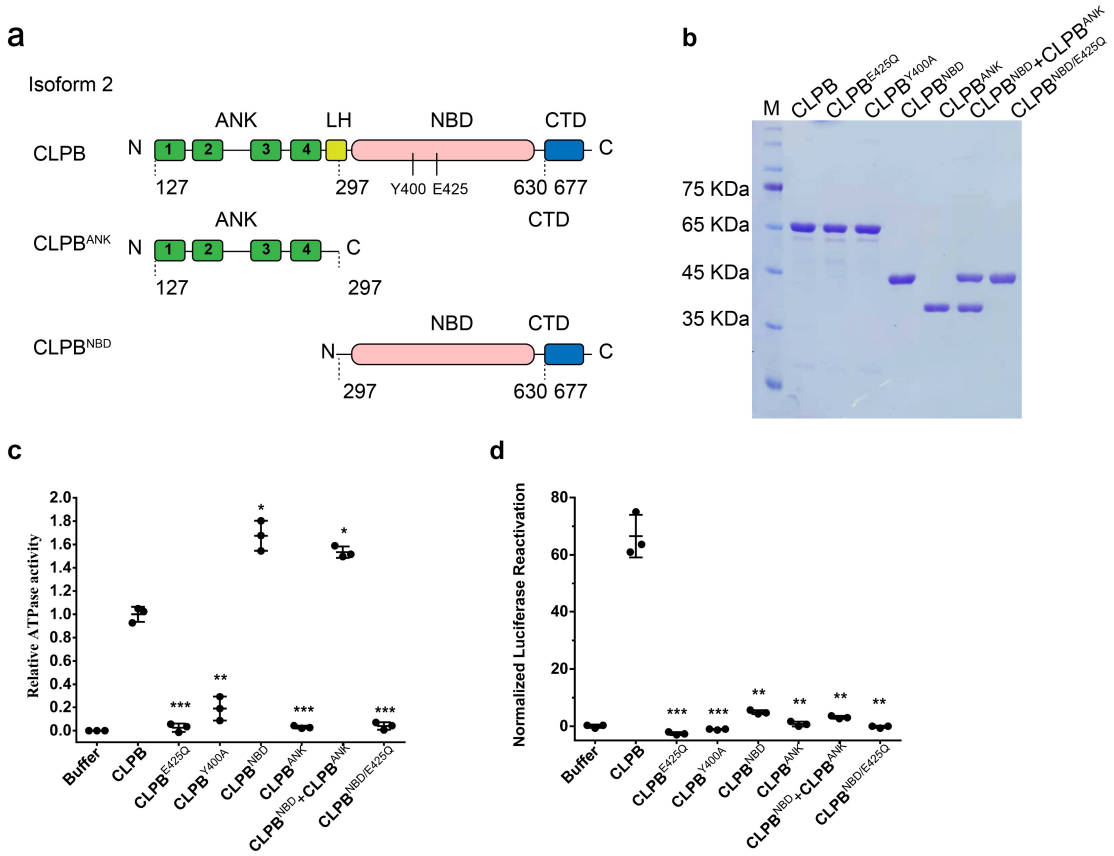


Fig. 2 ANK domain is essential for the disaggregase activity of CLPB.

a, Schematic overview of the CLPB, $CLPB^{ANK}$, $CLPB^{NBD}$, $CLPB^{Y400A}$ and $CLPB^{E425Q}$ constructs.

b, SDS-PAGE analysis of the purified proteins.

c, ATPase assays of CLPB, $CLPB^{E425Q}$, $CLPB^{Y400A}$, $CLPB^{NBD}$, $CLPB^{ANK}$, $CLPB^{NBD/E425Q}$ and $CLPB^{NBD}$ + $CLPB^{ANK}$. Results show that $CLPB^{E425Q}$, $CLPB^{ANK}$ and $CLPB^{NBD/E425Q}$ have nearly no ATPase activity. The ATPase activity of $CLPB^{Y400A}$ is significantly decreased. $CLPB^{NBD}$ and $CLPB^{NBD}$ + $CLPB^{ANK}$ have strong ATPase activity. ATPase activity was compared to CLPB (N=3, individual data points shown as dots, bars show mean ± SD, *p<0.05, **p<0.01, ***p<0.0001).

d, Disaggregase activity assay of CLPB, $CLPB^{E425Q}$, $CLPB^{Y400A}$, $CLPB^{NBD}$, $CLPB^{ANK}$, $CLPB^{NBD/E425Q}$ and $CLPB^{NBD}$ + $CLPB^{ANK}$ activity. The results show that these mutations (truncations) abolish the disaggregase activity of CLPB. Disaggregase activity was compared to CLPB (N=3, individual data points shown as dots, bars show mean ± SD, **p<0.01, ***p<0.0001).

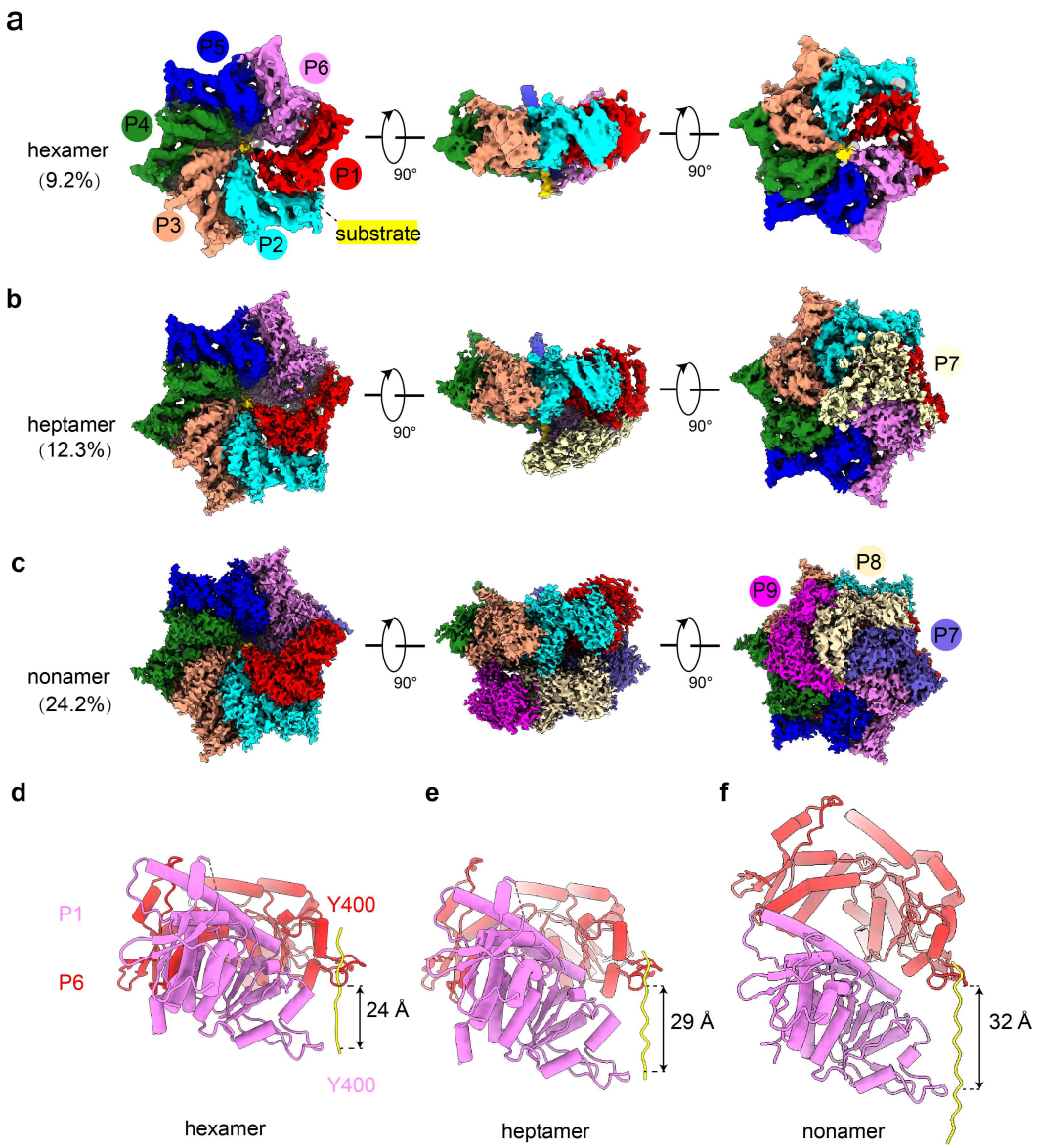


Fig. 3 Cryo-EM characterization of the NBD helical structures of CLPB.

a-c, Density maps of the NBD hexamer (a), heptamer (b) and nonamer (c), respectively. Each protomer is colored with different color.

d-f, Distance along the central substrate between Y400 of P1 and P6 in hexamer (d), heptamer (e) and nonamer (f) are 24 Å, 29 Å and 32 Å, respectively. During the transmition from hexamer to nonamer, the distance from P1 to P6 increases gradually for the next protomer binding.

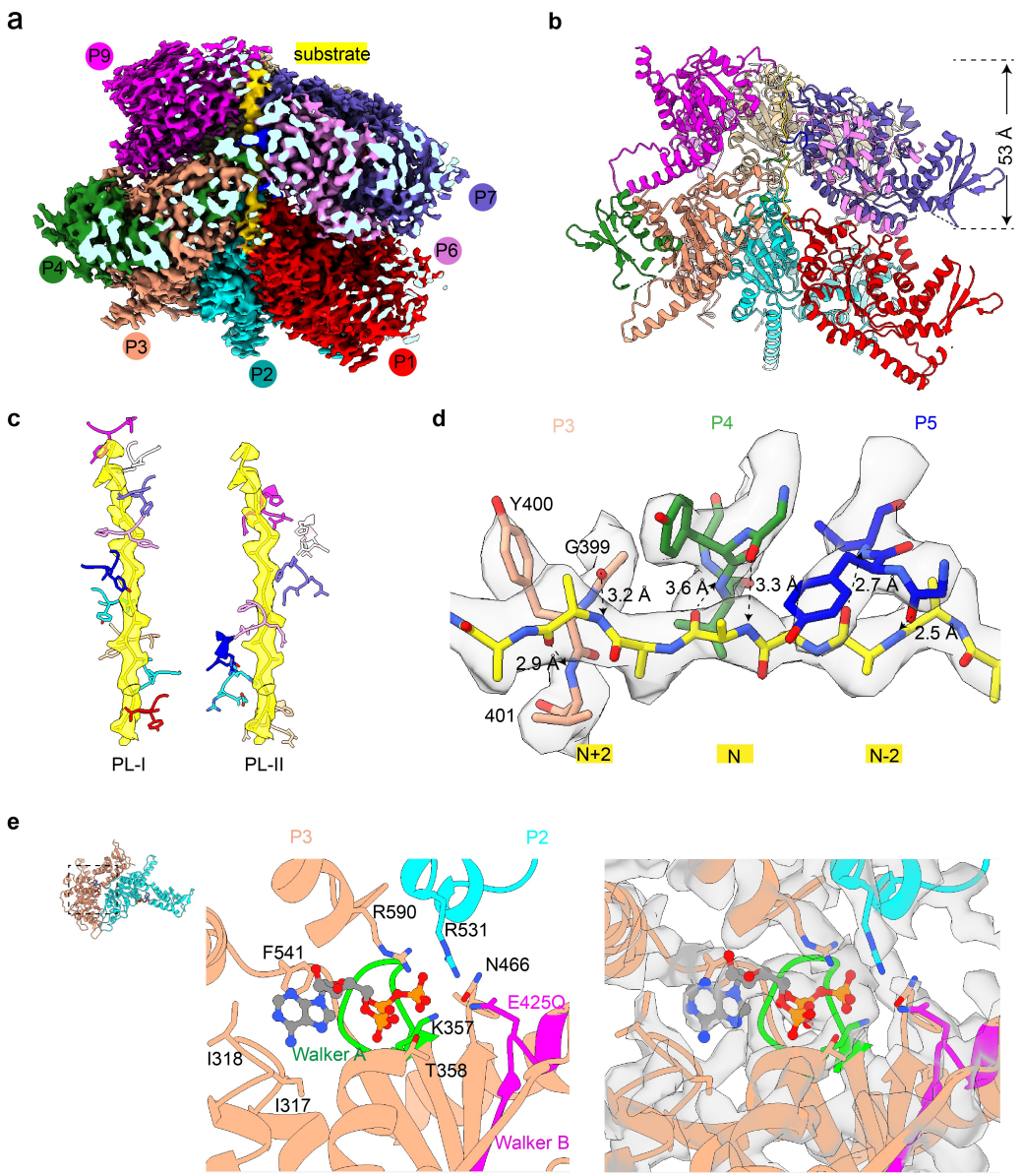


Fig. 4 NBD nonamer in processing a peptide substrate.

a-b, Density map (a) and atomic model (b) of the NBD nonamer in the substrate-processing state. Nine protomers are indicated as P1 to P9 and colored with different colors.

c, The spiral configuration of the PLs of CLPB protomers around the substrate in the central channel. PL-Is and PL-IIs are shown in the left and right panels, respectively. The conserved PL-I (G399-Y400-V401-G402) interacts with the substrate strongly, while the interactions between PL-II (E386-R387-H388) and substrate are less tightly.

d, Magnified view of the interaction between the PL-I of P3, P4 and P5 and the backbone of substrate.

e, Magnified view of the conserved ATP-binding pocket of CLPB. Functionally important residues of Walker A motif (K357, T358), Walker B motif (E425), sensor-1 (N466), sensor-2 (R590), arginine finger (R531) and the conserved residue I317, I318. Atomic model and density map are shown in the left and right panels, respectively.

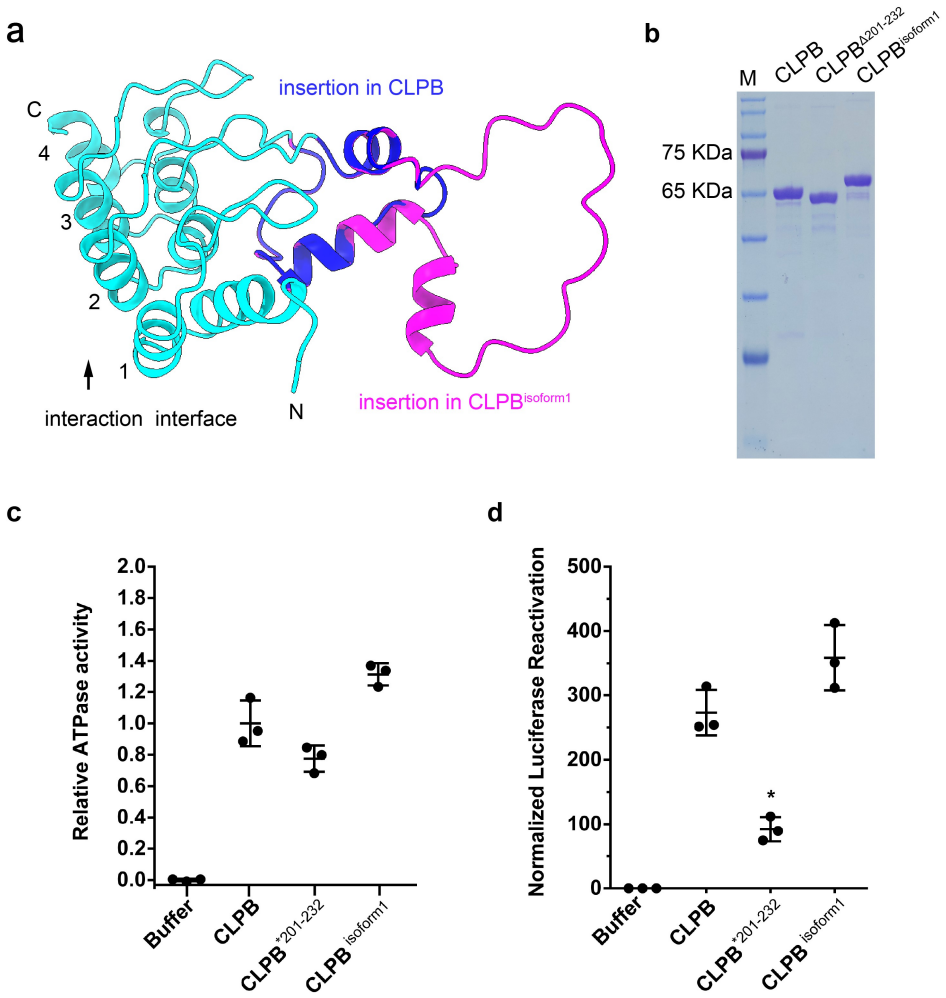


Fig. 5 The unique insertion in ANK domain is important for the disaggregase activity of CLPB.

a, Schematic and predicted models of ANK domain in CLPB and CLPB^{isoform1}. The unique insertion was colored in blue (CLPB) or magenta (CLPB^{isoform1}).

b, SDS-PAGE analysis of the purified proteins of CLPB, CLPB$^{\Delta 201\text{-}232}$ and CLPB^{isoform1}.

c, ATPase assay of CLPB, CLPB$^{\Delta 201\text{-}232}$ and CLPB^{isoform1}. Results show that CLPB$^{\Delta 201\text{-}232}$ and CLPB^{isoform1} have comparable ATPase activity. ATPase activity was compared to CLPB (N=3, individual data points shown as dots, bars show mean ± SD).

d, Disaggregase activity assay of CLPB, CLPB$^{\Delta 201\text{-}232}$ and CLPB^{isoform1} activity. Results show that no difference of the disaggregase activity between CLPB and CLPB^{isoform1}, while that of CLPB$^{\Delta 201\text{-}232}$ significantly reduced by more than 50%. Disaggregase activity was compared to CLPB (N=3, individual data points shown as dots, bars show mean ± SD, *$p<0.05$).

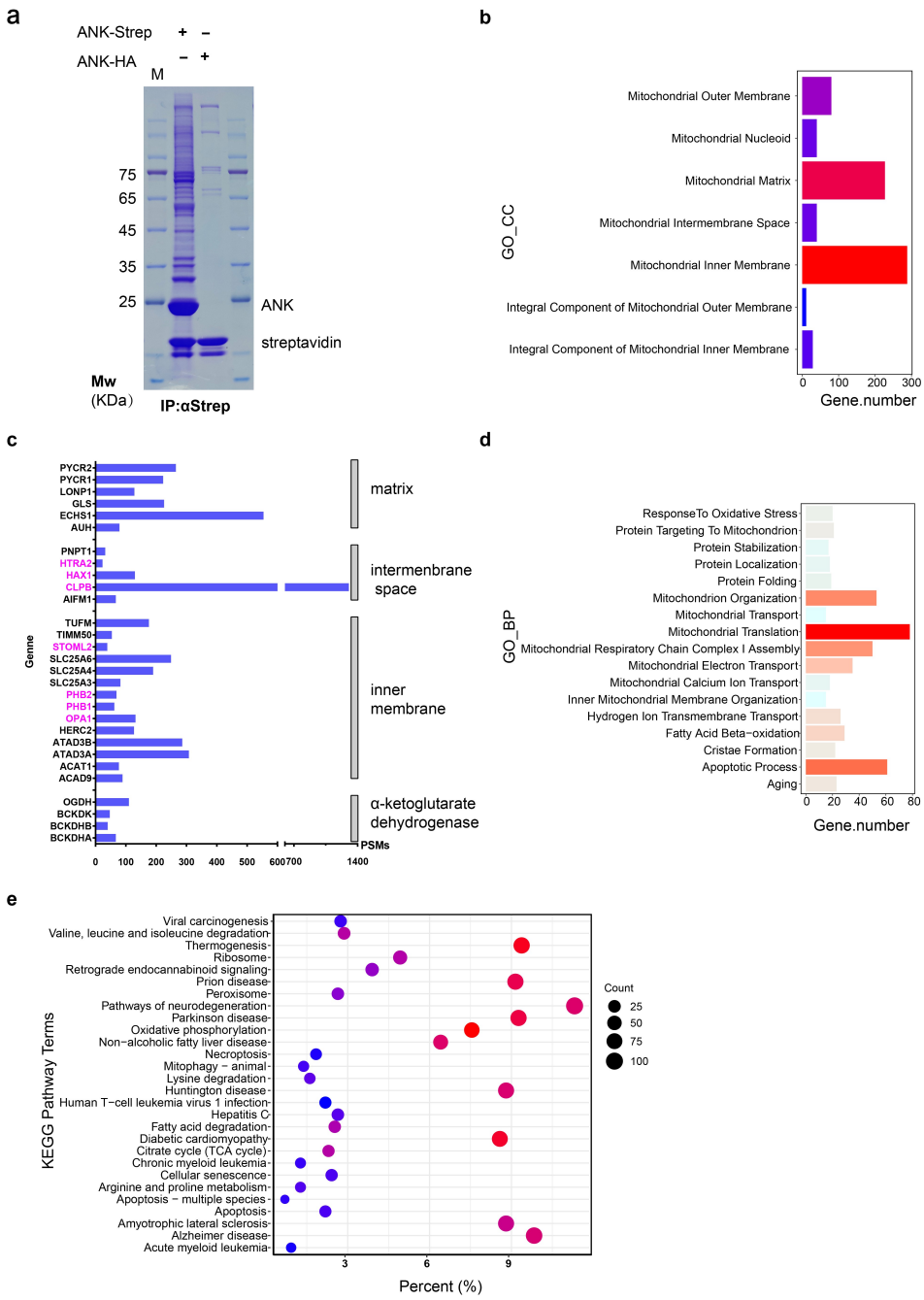


Fig. 6 The interactom of CLPB in mitochondrial.

a, HEK-293F cells expressing ANK-HA or ANK-Strep were subjected to anti-Strep immunoprecipitation. Precipitates were analyzed by SDS-PAGE, coomassie blue staining and mass-spectrometry (MS).

b, GO_CC analysis of the enrichment proteins involved different subcellular localizations in mitochondrial.

c, Representative enrichment proteins in the matrix, intermembrane space, inner membrane of mitochondrial, and some α-ketoglutarate dehydrogenases.

d, GO_BP analysis of the enrichment proteins involved in different mitochondrial bio-processes.

e, KEGG pathway enrichment analysis of the enrichment proteins.

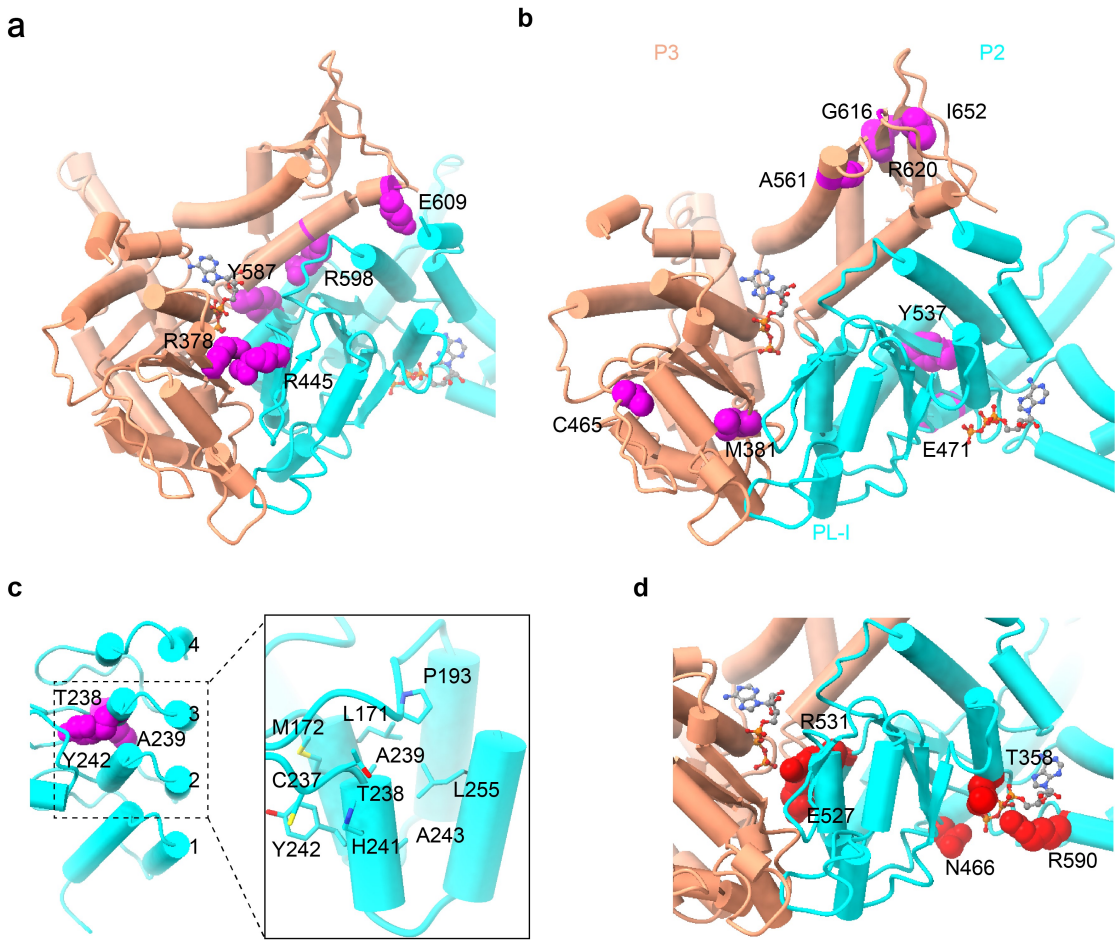


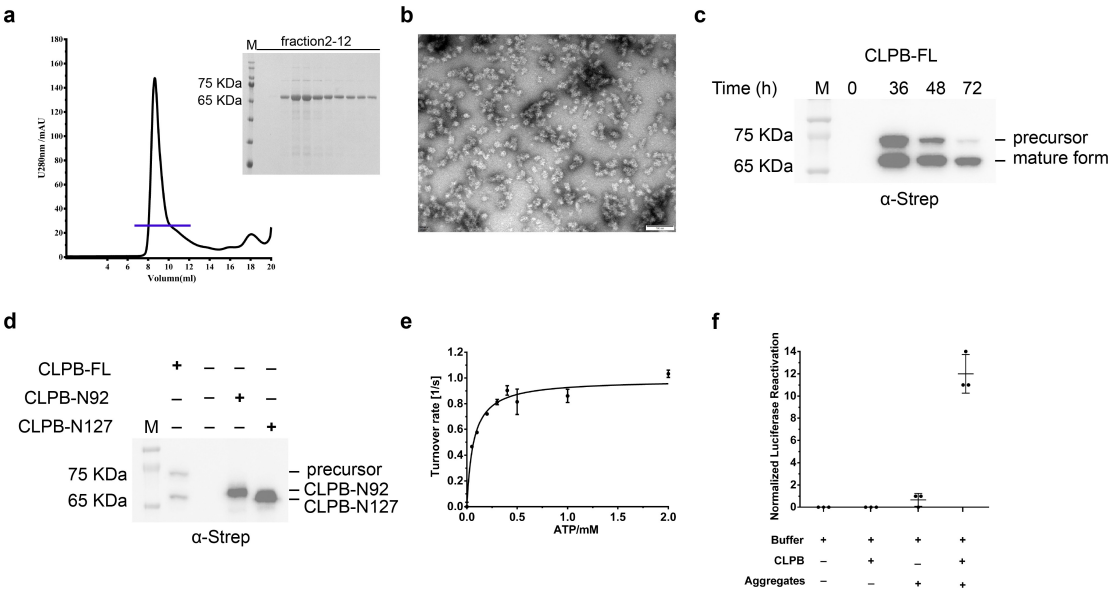
Fig. 7 The disease-related mutations of CLPB in 3-MGA and SCN

a, 3-MGA-related mutations in the interface of adjacent protomers.

b, 3-MGA-related mutations in the large and small subdomain of the ATPase domain.

c, 3-MGA-related mutations in the ANK domain.

c, SCN-related mutations in the ATP-binding pocket. Residues on the positions of these mutation are highlighted in magenta or red sphere models. ATP molecule is highlighted in stick model.



Supplementary Fig. 1 Determination the maturation cleavage sites of CLPB with biological functions.

a, The purification of CLPB-N92, analyzing by size-exclusion chromatography (left) and SDS-PAGE (right).

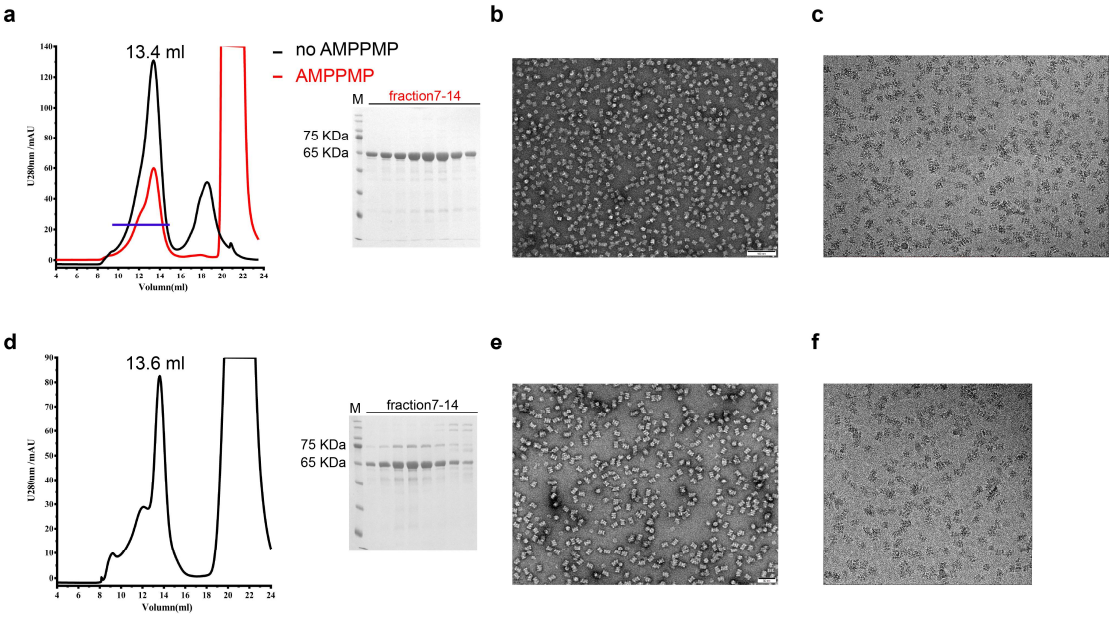
b, Negative staining electron microscopy (nsEM) of the peak fraction in (a). Result shows that CLPB-N92 forms aggregates in solution.

c, Time course of CLPB-FL-Strep expression in transiently transfected cells.

d, CLPB-FL, CLPB-N92 and CLPB-N127 expression in transiently transfected cells. Data show that the CLPB-N127 has the same molecular weight as the mature form of CLPB.

e, ATPase activity of CLPB-N127.

f, Disaggregase activity of CLPB-N127.



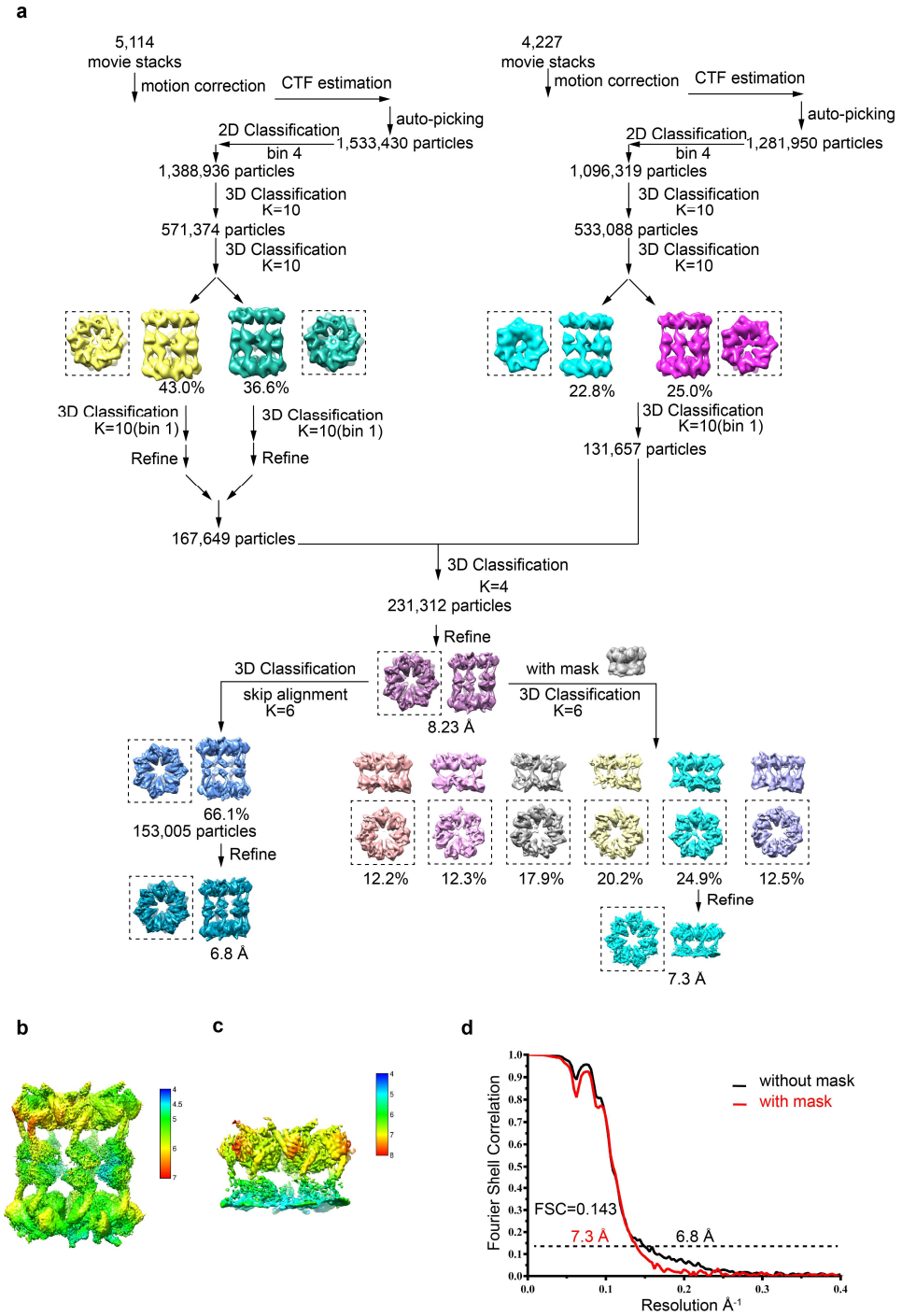
Supplementary Fig. 2 Sample preparation of CLPB and CLPB^{E425Q}.

a, Purification of CLPB using size-exclusion chromatography with AMPPNP (red line) or without AMPPNP (black line). Corresponding fractions were analyzed by SDS-PAGE (right panel).

b-c, Representative nsEM (b) and cryo-EM micrograph (c) of the peak fraction in (a).

d, Purification of CLPB^{E425Q} using size-exclusion chromatography in the presence of ATP. Corresponding fractions were analyzed by SDS-PAGE (right panel).

e-f, Representative nsEM (e) and cryo-EM micrograph (f) of the peak fraction in (d).

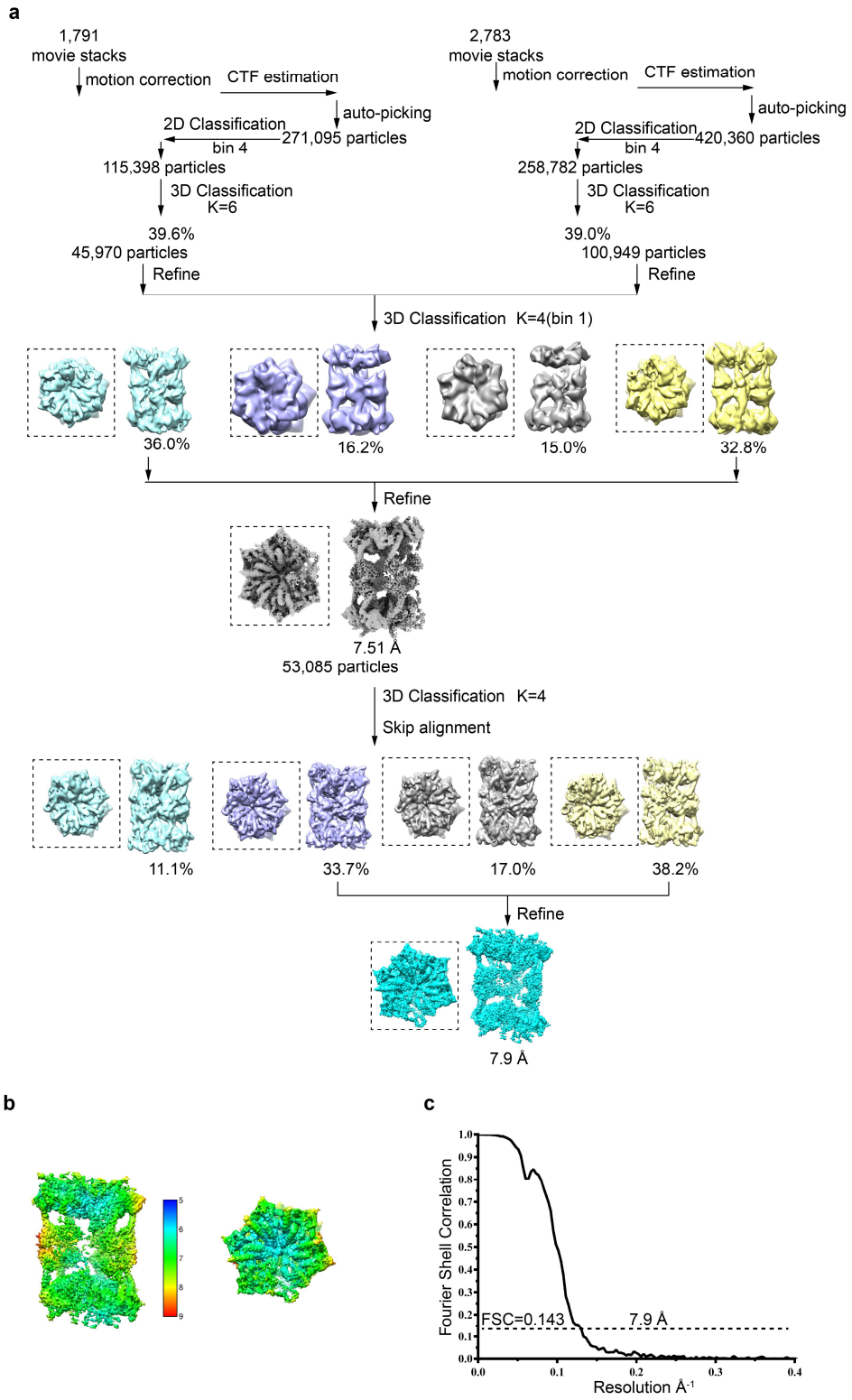


Supplementary Fig. 3 Image processing workflow of the CLPB dataset.

a, Image processing workflow of the CLPB dataset. The micrographs were subjected to motion correction and CTF estimation. The auto-picked particles were subjected to 2D classifications and 3D multiple rounds of classifications. The particles of double-heptamer were subjected to 3D classification without alignment to improve the homogeneity. A half mask was added to improve the resolution.

b-c, Local resolution estimation of the map of double-heptamer (b) and heptamer (c) of CLPB.

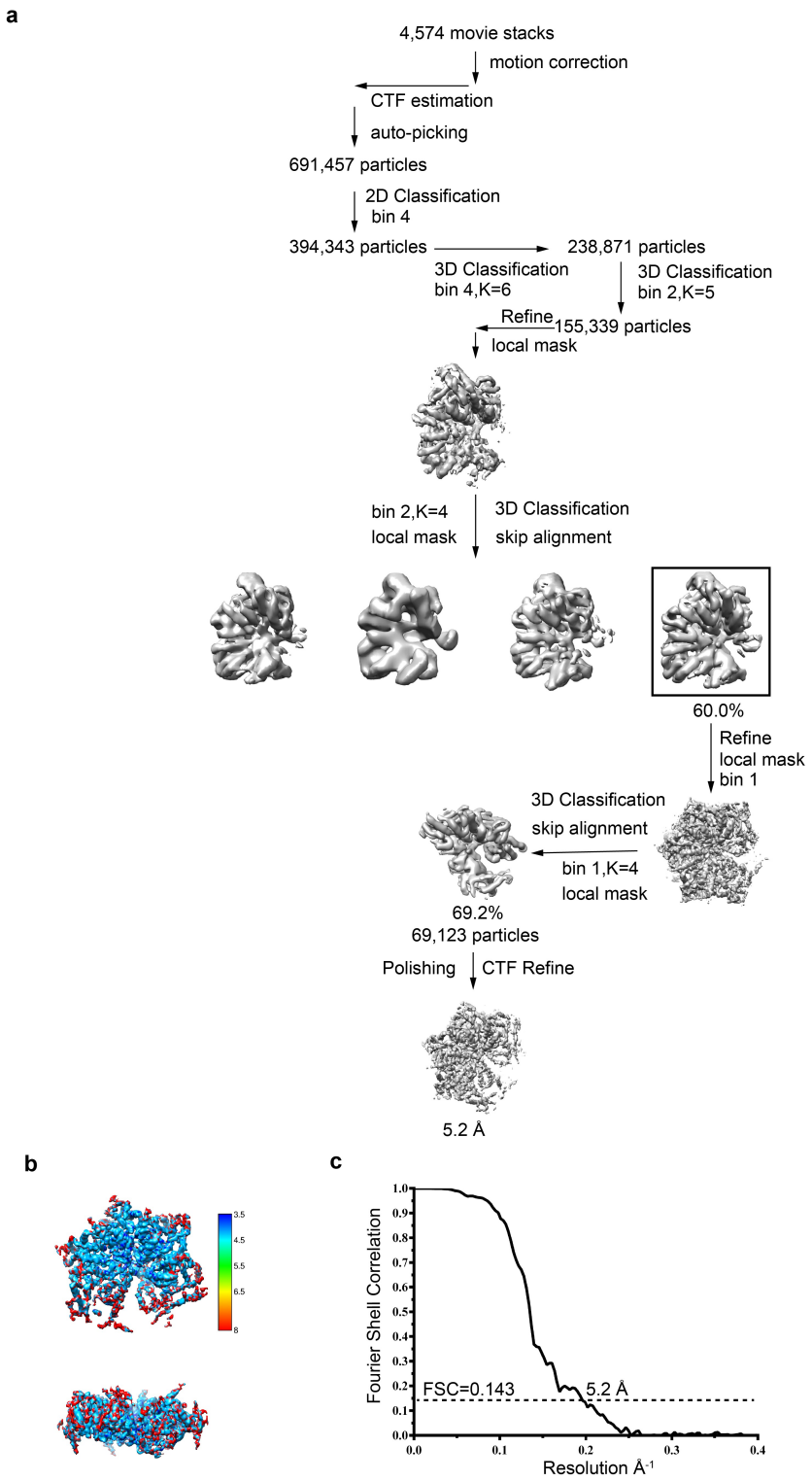
d, Fourier Shell Correlation (FSC) curves for the final cryo-EM map using the gold standard FSC 0.143 criteria.



Supplementary Fig. 4 Image processing workflow of the CLPB^{E425Q} dataset.

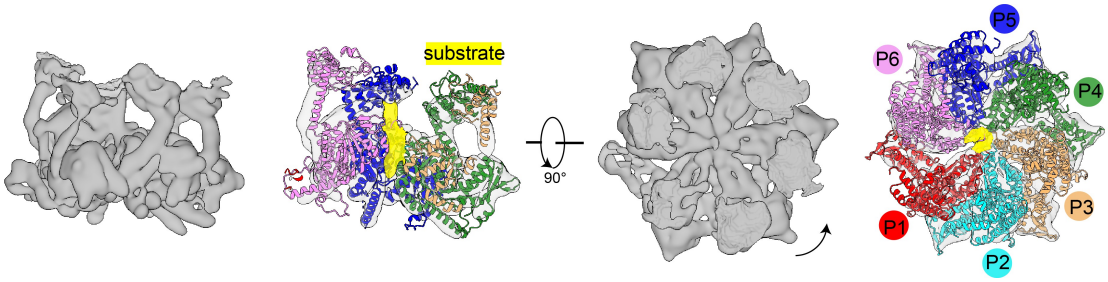
a, Image processing workflow of the CLPB^{E425Q} dataset.

b-c, Local resolution estimation (b) and fourier shell correlation (FSC) curve (c) of the final map of CLPB^{E425Q} double-hexamer.



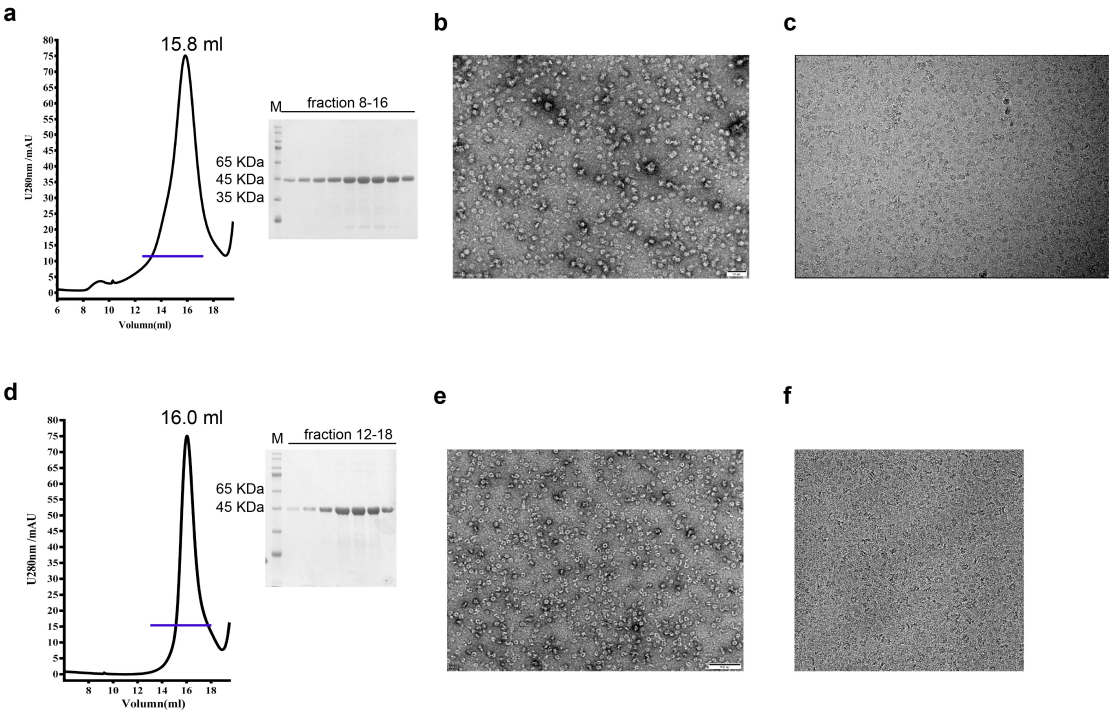
a, Image processing workflow of the NBD of CLPB^{E425Q} dataset. The micrographs were subjected to motion correction and CTF estimation. The auto-picked particles were subjected to 2D classification and multiple rounds of 3D classifications. The 3D classifications were done with local masks.

b-c, Local resolution estimation (b) and fourier shell correlation (FSC) curve (c) of the final map of CLPB^{E425Q}-NBD-ring.



Supplementary Fig. 6 CLPB binds to the substrate in the central channel, and forms a helical structure.

The density map of the double-hexamer was shown in half. Six protomers adopt an upward spiral staircase around the substrate polypeptide, in a counter-clockwise rotation pattern when viewed from the ANK domain direction. The density of substrate is shown in yellow and CLPB in gray.



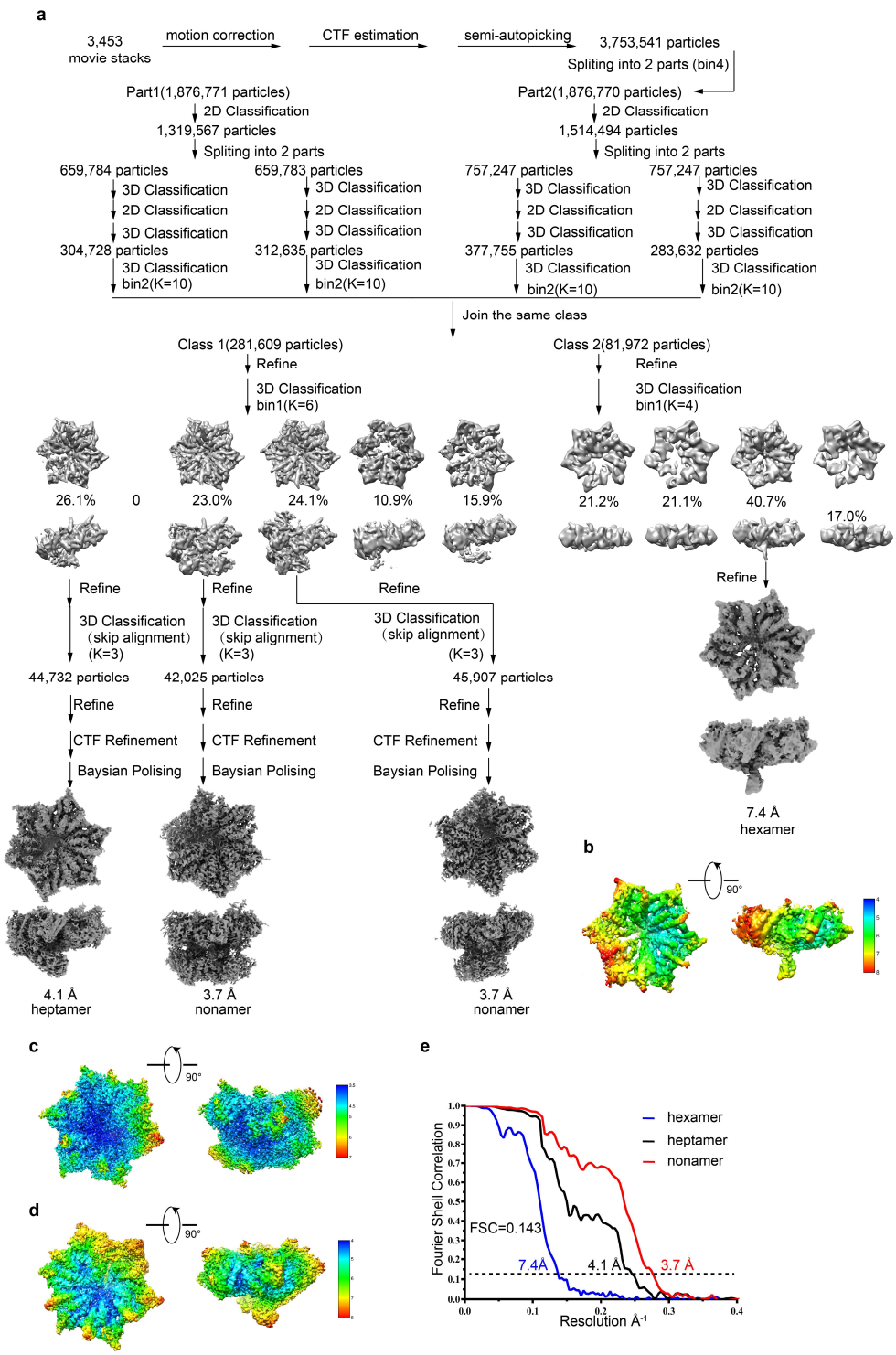
Supplementary Fig. 7 Sample preparation of the NBD^{E425Q} and NBD.

a, Purification of NBD^{E425Q} using size-exclusion chromatography in the presence of ATP. Corresponding fractions were analyzed by SDS-PAGE (right panel).

b-c, Representative nsEM (b) and cryo-EM micrograph (c) of the peak fraction in (a).

d, Purification of NBD using size-exclusion chromatography in the presence of AMPPNP. Corresponding fractions were analyzed by SDS-PAGE (right panel).

e-f, Representative nsEM (e) and cryo-EM micrograph (f) of the peak fraction in (d).

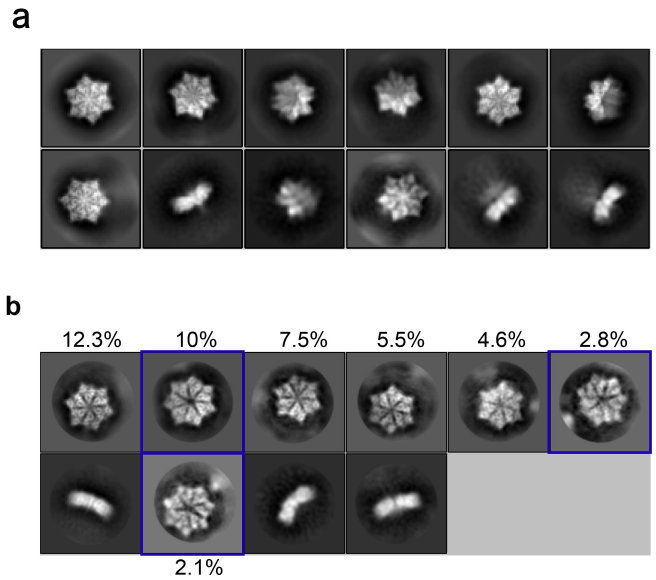


Supplementary Fig. 8 Image processing workflow of the NBD^{E425Q} dataset.

a, Image processing workflow of the CLPB dataset. The micrographs were subjected to motion correction and CTF estimation. The auto-picked particles were subjected to multiple rounds of 2D and 3D classifications. Three different oligomeric arrangements, hexameric, heptameric and nonameric were identified.

b-d, Local resolution estimation of hexamer (b), heptamer (c) and nonamer (d) in (a).

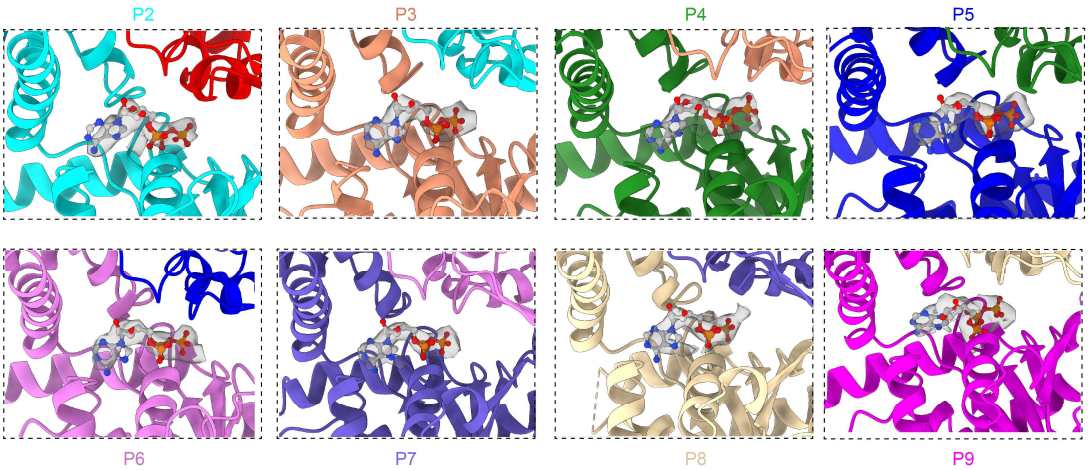
e, Fourier shell correlation (FSC) curves of the final cryo-EM map of hexamer (blue line), heptamer (black line) and nonamer (red line), using the gold standard FSC 0.143 criteria.



Supplementary Fig. 9 2D-classification averages cof NBD^{E425Q} and NBD.

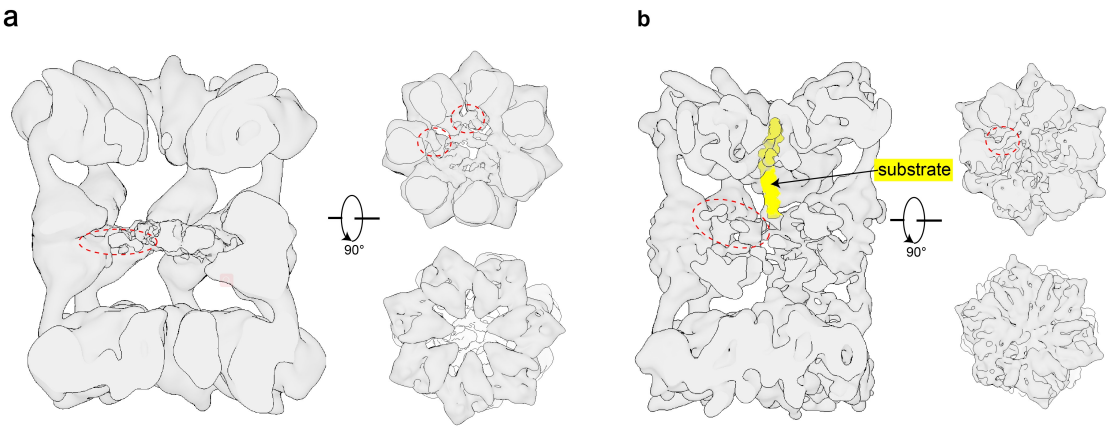
a, 2D-classification average of NBD^{E425Q} dataset. Only top views of hexameric ring were found, and no hepatameric ring was observed.

b, 2D-classification average of NBD dataset. Top views of both hexameric ring and hepatameric (circled by blue squares) ring were observed. The hexameric ring is more compact than hepatameric ring, and the diameter of central pore of hexamer is much smaller than that of heptamer.



Supplementary Fig. 10 Nucleotide binding states of the nine ATPase sites in the nonamer.

ATP occupation of the all the ATPase sites in the nonamer. The atomic models are color-coded for different protomers. The density map of ATP was superimposed with the atomic model.



Supplementary Fig. 11 Extra densities around the ANK domains, extending towards the central channel.

a-b, Extra densities were found around the ANK domains in the structures of both CLPB apo-state (a) and CLPB^{E425Q} substrate-bond state (b), extending towards the central channel. The extra densities are highlighted by dotted ellipse. The density of the substrate is shown in yellow.