

Cryo-EM structure of Nma111p, a unique HtrA protease composed of two protease domains and four PDZ domains

Cell Research (2017) 27:582-585. doi:10.1038/cr.2017.5; published online 13 January 2017

Dear Editor,

Apoptosis, a physiological form of programmed cell death, is essential for the maintenance of normal cellular homeostasis [1]. In the unicellular eukaryote *Saccharomyces cerevisiae*, a number of evolutionarily conserved apoptosis-regulatory proteins have been identified, one of which is nuclear mediator of apoptosis 111 kDa protein (Nma111p), a protease targeting Bir1p, the sole inhibitor-of-apoptosis protein (IAP) in yeast [2]. Nma111p is a serine protease of the HtrA family, of which a common structural feature is the presence of the trypsin-like protease domain and the post-synaptic density 95, *Drosophila* discs large, zona-occludens-1 (PDZ) domain [3]. Based on the domain organization, the HtrA family proteins can be divided into distinct groups. Group 1 proteins (e.g., DegS in *E. coli* and Htra2 in mammal) contain one protease domain and one PDZ domain [4, 5], and group 2 proteins (e.g., DegP and DegQ in *E. coli*) contain one protease domain and two PDZ domains [6, 7]. Nma111p-like proteases in group 3 are distinctively characterized by possessing two protease domains and four PDZ domains, of which the N-terminal protease domain exhibits protease activity, whereas the C-terminal protease domain is supposed to be degenerated in protease activity. All members in group 3 have the length twice that of group 2 HtrA proteins and harbor an internal duplication sequence. They are widely present in various organisms and possess diverse functions [8].

The structures of several members of groups 1 and 2 of the HtrA family have been determined by X-ray crystallography or cryo-electron microscopy (cryo-EM) [6, 7, 9, 10]. However, none belongs to group 3. We thus attempted to perform structural and functional investigations of Nma111p. We first purified Nma111p and its substrate Bir1p. The absorption peak of size exclusion chromatography (SEC) of Nma111p corresponds to a trimer (~300 kDa) (Supplementary information, Figure S1A and S1B). The protease activity assay showed that

Bir1p was degraded by Nma111p (Supplementary information, Figure S1C and S1D), demonstrating that our sample preparation is effective. The full-length Bir1p, however, is unstable and tends to undergo degradation in solution. We then used a new model substrate, L2 (named subL2 hereafter), which had been serendipitously discovered to be degraded by Nma111p, in the subsequent analysis of the protease activity of Nma111p (Supplementary information, Figure S1E-S1G). Next, we performed single-particle analysis of Nma111p^{S235A/S236A}, in which the two serine residues critical for the protease activity were mutated to avoid protein degradation, and determined five structures of various conformational states (Figure 1A and Supplementary information, Figure S1H-S1Q). Among these structures, the state 1 structure with C3 symmetry, occupying the largest proportion, was determined at an overall resolution of 3.4 Å (Supplementary information, Figure S1K-S1Q). Sequentially, we built an atomic model of state 1 structure based on the EM density (Supplementary information, Table S1). In this structure, three Nma111p protomers are arranged in parallel along the C3 axis, forming a cage-like trimeric structure with a prismatic shell (Figure 1B). Each protomer is composed of six domains. The N-terminal half of Nma111p comprises the protease1, PDZ1, and PDZ2 domains, followed by the C-terminal half, which has a similar domain arrangement (Figure 1B and Supplementary information, Figure S2A and S2B). The three domains of each half form a triangle, and these two halves show a pseudo-C2 symmetry. Three protease1 domains and three protease2 domains are located at the two poles of the prismatic shell, whereas the sidewall comprises 12 PDZ domains. The PDZ2 and PDZ4 domains are situated around the equator region (Figure 1B).

The correct positioning of the catalytic triad and the main chain interactions between loop L2 and the substrate are essential features of active trypsin-like serine proteases, and may be used to define whether a protease is active or inactive [11]. In proteolytically competent

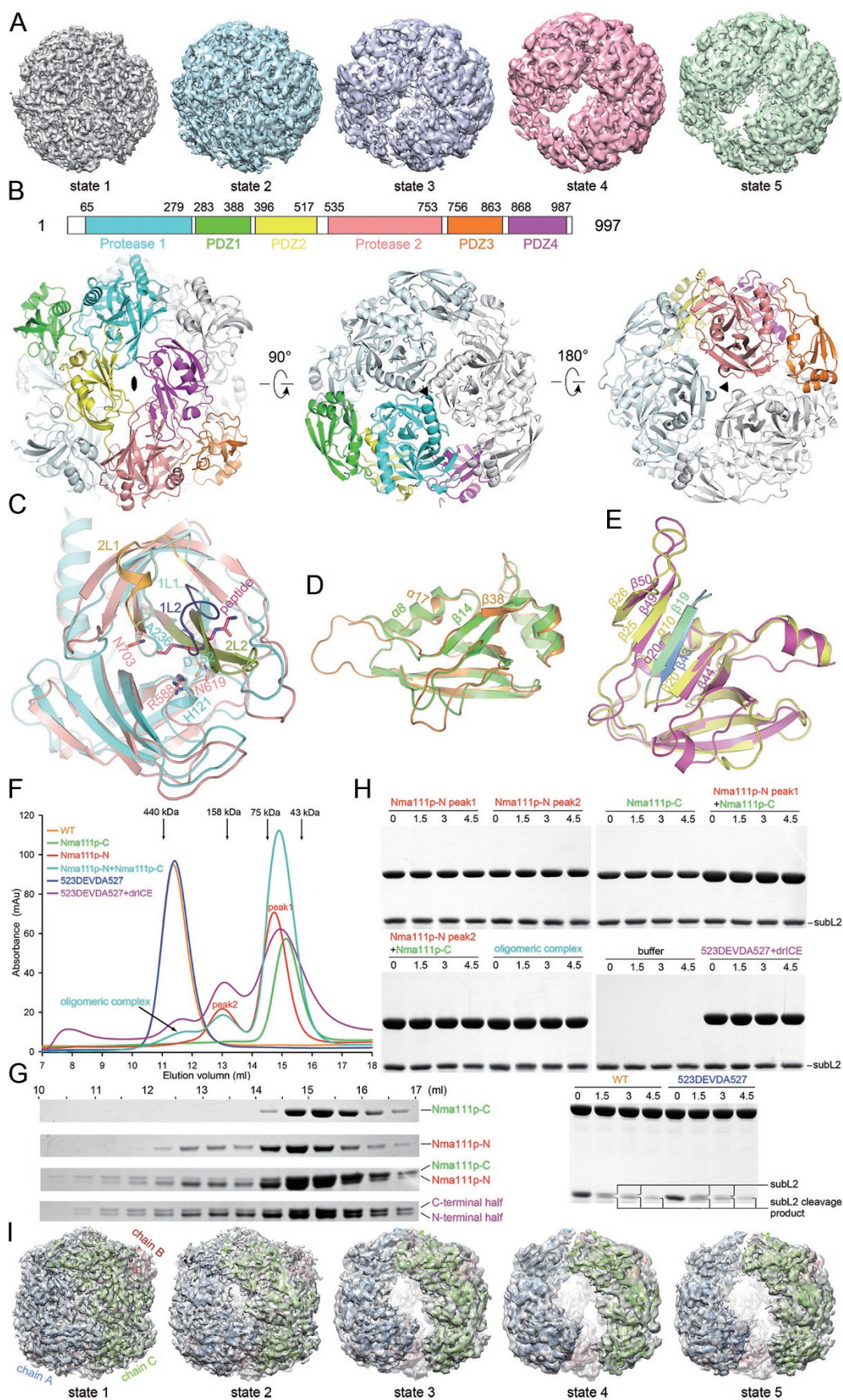


Figure 1 Structure of Nma111p. **(A)** Cryo-EM map of multi-conformations of trimeric Nma111p (3.1-3.5 δ). **(B)** State 1 of trimeric Nma111p is shown as ribbon diagram and visualized from three orientations. Protease1 (cyan), PDZ1 (green), and PDZ2 (yellow) are in the N-terminal half. Protease2 (salmon), PDZ3 (orange), and PDZ4 (magenta) are in the C-terminal half. The ellipse indicates the pseudo two-fold symmetry of intra-subunit. The black triangle indicates the three-fold symmetry of trimeric structure. **(C)** Comparison between protease1 (cyan) and protease2 (salmon). The substrate peptide (warm pink) of trypsin is modelled to protease1 and protease2 domains. The active site residues, H121, D152, and S236A in protease1, and R588, N619, and N703 in protease2, are shown in stick mode. The active loops are indicated by colors. **(D)** Alignment of PDZ1 (green) and PDZ3 (orange). **(E)** Alignment of PDZ2 (yellow) and PDZ4 (magenta). The internal PDZ2 and PDZ4 pseudo-ligands (β 19 and β 43) are shown in lime green and sky blue, respectively. **(F)** SEC Superdex-200 10/300 column elution profiles of Nma111p variants. **(G)** 4%-12% Nu-PAGE of purified Nma111p variants corresponding to 10-17 ml of products obtained from **F**. **(H)** SDS-PAGE analysis of the proteolytic activity of Nma111p variants against subL2. The nearly identical molecular weights of N- and C-halves result in bands of similar sizes visualized in 12% SDS-PAGE. **(I)** Cryo-EM maps of multi-conformational Nma111p viewed from the side view. Three individual subunits are color-coded.

trypsin, the active loop L2 directly forms three hydrogen bonds with the substrate (PDB:1OX1 [11]; Supplementary information, Figure S2C). In the protease1 domain of Nma111p, the 1L2 (1 represents the protease1 domain) blocks the substrate-binding groove, thus preventing substrate binding and hydrogen bond formation (Figure 1C). This feature renders Nma111p in an inactive state. In addition, the loops of the protease1 domain (i.e., 1LD, 1L1, and 1L2) have weak density, suggesting a flexible conformation (Supplementary information, Figure S2D), reminiscent of many HtrA family proteins in inactive states, in which these loops are also too flexible to be traced [12, 13]. The catalytic triad of trypsin is composed of His, Asp, and Ser [11]. In the active state, interactions between His and Asp or Ser facilitate proteolytic activity. In the protease1 domain, the His121, Asp152, and Ser236 (mutated to Ala) catalytic residues can be clearly observed (Figure 1C). His121 has two conformations; however, in either conformation, His121 fails to form hydrogen bonds with Asp152 or Ser236 due to the distances between them (Supplementary information, Figure S2E), further suggesting that the protease1 domain is in an inactive state.

In Nma111p, the overall structures of the protease1 and protease2 domains both possess the folding of trypsin (Figure 1C); however, they show characteristic differences. First, structural and sequence alignments revealed that the catalytic triad in protease1 is formed by His121, Asp152, and Ser236, whereas Arg588, Asn619, and Asn703 constitute the corresponding triad in protease2 (Figure 1C and Supplementary information, Figure S2A). Thus, the protease2 domain lacks the basic characteristic of a serine protease. Second, the position of 2L1 (2 represents the protease2 domain), which contains Asn703 (corresponding to the position of Ser236), is distant from Arg588 (corresponding to the position of His121) (Figure 1C). Third, 2L2 extends to two rigid β -strands, and thus fully rejects substrate entry (Figure 1C). Based on these structural features, we conclude that the protease1 domain is responsible for the protease activity, albeit in-

active in this structure.

The overall fold of the PDZ1 and PDZ3 domains is similar to that of the PDZ domains of other HtrA proteins binding to the substrate C-terminal residues [14]. The difference of PDZ1 and PDZ3 is in the peptide-binding groove (Figure 1D). The binding groove in PDZ1 may accommodate the peptide, whereas the groove of PDZ3 seems to be occupied due to bulging (Supplementary information, Figure S2F). In addition, the α 8 helix and β 14 strand of PDZ1 are the structural elements to form the binding groove, consistent with other HtrA proteins [14]. The α 17 and β 38 of PDZ3, corresponding to α 8 and β 14 of PDZ1, however, lack the structural features for groove forming (Figure 1D). PDZ2 and PDZ4 have the higher similarity in folding. β 19 in PDZ2 and β 43 in PDZ4 occupy their corresponding peptide-binding grooves, resembling pseudo-peptides of the PDZ domain. PDZ2 and PDZ4 each contains two bonus β -strands (β 25/ β 26 in PDZ2 and β 49/50 in PDZ4), representing a unique structural feature in the HtrA family (Figure 1E).

The N-terminal half of Nma111p is 22% identical to the C-terminal half (Supplementary information, Figure S2A). We then purified N- and C-terminal halves to investigate their functions. Biochemical analysis revealed that the C-terminal half (Nma111p-C) only formed monomers, and the N-terminal half (Nma111p-N) mainly formed monomers with a small amount of trimers. When we incubated the two halves together and performed SEC, we observed a small peak corresponding to a trimer of full-length Nma111p, indicating a spontaneous formation of oligomers (Figure 1F and 1G). None of these peak elution products exhibited significant proteolytic activity towards subL2 (Figure 1H). Next we inserted a cleavage site specifically recognized by the protease drICE [15] into the linker region of Nma111p N- and C-terminal halves, generating a variant designated as 523DEVDA527. Both the oligomeric state and the proteolytic activity of 523DEVDA527 were similar to those of the wild type (Figure 1F-1H). Upon treatment with drICE, 523DEVDA527 lost its proteolytic activity, and

revealed a similar SEC profile to the mixture of the two halves (Figure 1F-1H). These observations suggest that first, the degenerated C-terminal half is indispensable for the formation of oligomers, and second, full-length Nma111p, with the N- and C-terminal halves on a single chain, is essential to maintain the oligomeric state and proteolytic activity.

The cage-like trimeric state 1 of Nma111p represents a relatively close state with the catalytic triad embedded inside the cavity. It raises an inevitable question that how the substrates get access to the catalytic triad. As mentioned above, we reconstructed four other trimeric conformations and revealed a substantial conformational heterogeneity (Figure 1A and Supplementary information, Figure S1P). From state 1 with C3 symmetry to states 2-5 with C1 symmetry, chains A and C gradually extend outwards and thus form a gap, while chain B remains almost static. The diameter of state 5 reaches 115 Å, the largest among the conformations. The gap space between chains A and C in the middle region is significantly larger than that at top and bottom regions (Figure 1I and Supplementary information, Figure S2G and S2H). The conformations opening at different extents reflect a dynamic feature of Nma111p, suggesting that substrates can exploit this gap to get access to the catalytic triad and then be degraded, which has not been discovered among other HtrA family proteins. The full-length Nma111p, with the N- and C-terminal halves on a single chain, is required for the coordination of the open gap, further indicating that the degenerated C-terminal half is essential for Nma111p function. As a distinctive feature, the multi-conformations of Nma111p suggest that the Nma111p cages exist dynamically in equilibrium between closed and opened states. The former may be the state that the substrate could not approach, whereas the latter is likely the state ready for substrate binding.

Acknowledgments

We thank Shan Sun and Desheng Liu for helpful discussion and comments on the manuscript. We thank Jianlin Lei and Yanji Xu for the EM support. We acknowledge the Tsinghua University

Branch of China National Center for Protein Sciences Beijing and the 'Explorer 100' cluster system of the Tsinghua National Laboratory for Information Science and Technology for providing EM and computation resources. This work was funded by grants from the National Natural Science Foundation of China (31230016/31370717), the National Basic Research Program of China (2016YFA0501101), and by Tsinghua-Peking Joint Center for Life Sciences, Advanced Innovation Center for Structural Biology and One-Thousand Talent Program of China.

Li Zhang^{1,*}, Xiaojing Wang^{1,*}, Fenghui Fan¹,
Hong-Wei Wang², Jiawei Wang², Xueming Li²,
Sen-Fang Sui¹

¹State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China; ²Ministry of Education Key Laboratory of Protein Science, Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

*These two authors contributed equally to this work.

Correspondence: Sen-Fang Sui^a, Xueming Li^b, Jiawei Wang^c

^aE-mail: suisf@mail.tsinghua.edu.cn

^bE-mail: lixueming@tsinghua.edu.cn

^cE-mail: jwwang@tsinghua.edu.cn

References

- Hengartner MO. *Nature* 2000; **407**:770-776.
- Yang QH, Church-Hajduk R, Ren J, *et al.* *Genes Dev* 2003; **17**:1487-1496.
- Clausen T, Southan C, Ehrmann M. *Mol Cell* 2002; **10**:443-455.
- Sohn J, Grant RA, Sauer RT. *Cell* 2007; **131**:572-583.
- Suzuki Y, Imai Y, Nakayama H, *et al.* *Mol Cell* 2001; **8**:613-621.
- Jiang J, Zhang X, Chen Y, *et al.* *Proc Natl Acad Sci USA* 2008; **105**:11939-11944.
- Bai XC, Pan XJ, Wang XJ, *et al.* *Structure* 2011; **19**:1328-1337.
- Schuhmann H, Mogg U, Adamska I. *Biochem J* 2011; **435**:167-174.
- Singh N, Kuppili RR, Bose K. *Arch Biochem Biophys* 2011; **516**:85-96.
- Shen QT, Bai XC, Chang LF, *et al.* *Proc Natl Acad Sci USA* 2009; **106**:4858-4863.
- Perona JJ, Craik CS. *Protein Sci* 1995; **4**:337-360.
- Sawa J, Malet H, Krojer T, *et al.* *J Biol Chem* 2011; **286**:30680-30690.
- Wilken C, Kitzing K, Kurzbauer R, *et al.* *Cell* 2004; **117**:483-494.
- Clausen T, Kaiser M, Huber R, *et al.* *Nat Rev Mol Cell Biol* 2011; **12**:152-162.
- Yan N, Wu JW, Chai J, *et al.* *Nat Struct Mol Biol* 2004; **11**:420-428.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)