学位論文

Evolution of the complement genes in the Arthropoda

（節足動物における補体遺伝子の進化）

平成 26 年 12 月博士（理学）申請

東京大学大学院理学系研究科
生物科学専攻
関口　玲生

学位論文

Evolution of the complement genes in the Arthropoda

（節足動物における補体遺伝子の進化）

平成 26 年 12 月博士（理学）申請

東京大学大学院理学系研究科
生物科学専攻
関口　玲生

Contents

Abstract

The mammalian complement system comprises more than 30 serum and cell surface proteins, and plays a pivotal role in innate immunity. The third complement component (C3) plays the central role in the mammalian complement system both in the activation and effecter pathways. C3 belongs to the thioester-containing protein (TEP) family; the other members of this family are the nonspecific protease inhibitor alpha-2 macroglobulin (A2M), the glycosylphosphatidylinositol (GPI)-anchored protein CD109, and insect TEP (iTEP), shown to be orthologous to CD109. Evolutionary studies thus far performed have indicated that the evolutionary origin of the complement system can be traced back to the common ancestor of the eumetazoa, and the gene duplications and the subsequent functional diversification of TEP family genes seems to have been completed by the emergence of this common ancestor, because the genes for C3, A2M and iTEP/CD109 were identified from sea anemones, Cnidaria. In addition to C3, factor B (FB), mannan-binding lectin-associated serine protease (MASP) have been identified from sea anemone, indicating that the primitive complement system composed of at least three components, C3, FB, and MASP, was present in the eumetazoa common ancestor. These complement and TEP genes were identified from all deuterostome species analyzed thus far. On the other hand, some protostomes have lost certain complement and TEP genes, suggesting that these genes were lost multiple times independently in some protostome lineages. In the arthropod, the C3 and FB gene have been reported from horseshoe crab and tick species, and C3 gene has been reported from a spider, although water flea and several insect genomes deciphered to date lacked all complement gene.

To elucidate the evolution of C3 and other TEP genes in the arthropod, TEP

cDNAs were isolated from a sea spider, *Ammothea* sp. (Chelicerata), a centipede, *Scolopendra subspinipes* (Myriapoda) and sea lice, *Pseudocaligus fugu*, *Caligus* sp. and Pandaridae gen. et sp. (Crustacea) by reverse transcription polymerase chain reaction (RT–PCR) amplification using universal degenerate primers specific for the thioester region. In addition, to elucidate the evolutionary history of the complement system in the arthropod, *de novo* transcriptome analysis was performed with *Ammothea* sp. and a spider, *Hasarius adansoni* (Chelicerata), *S. subspinipes* and two millipedes, *Niponia nodulosa* and *Epanerchodus* sp. (Myriapoda), a sea firefly, *Vargula* sp. (Crustacea), and *P. fugu* and *Caligus* sp. Complement genes were identified based on their characteristic domain structures. As a result of these two approaches, the complement C3 and FB genes, but not other complement genes were identified from *Ammothea* sp., *H. adansoni*, and *S. subsupinipes*. In contrast, no complement genes were identified from *N. nodulosa*, *Epanerchodus* sp. and *Vargula* sp. Interestingly, three sea lice species *P. fugu*, *Caligus* sp., and Pandaridae gen. et sp. possessed C3 showing a closer similarity to fish C3 than to arthropod C3. No additional complement components identifiable by their characteristic domain structures were found from any of these species by RNA-seq.

These results together with genome sequence information for the water flea and several species of the insect suggest that the common ancestor of the arthropod possessed a simple complement system comprising C3 and FB, resembling the alternative pathway of the mammalian complement system. It was lost at least twice independently during the evolution of the arthropod in the millipede lineage and in the common ancestor of crustacean and Hexapoda. In addition, sea lice species may have secondary obtained the C3 gene by horizontal gene transfer from their host fish species.

General Introduction

All metazoans seem to possess the immune system to detect and eliminate non-self cells and molecules potentially harmful to the host such as pathogenic bacteria or viruses. The mammalian immune system is divided into two parts, which are the innate immunity and adaptive immunity. Innate immunity is a non-specific immune system which responds immediately upon infection assumed by the complement system, phagocytes, antimicrobial peptides, and so on. The innate immune cells, such as macrophages, dendritic cells, and neutrophils, express pattern recognition receptors (PRRs) on their surface (Medzhitov 2007). Innate immune responses are induced by the binding of PRRs to the foreign molecules that have a common structural motif or pattern in microorganisms called as pathogen-associated molecular patterns (PAMPs) (Janeway Jr. and Medzhitov 2002). The innate immune system is found in almost all organisms and is main immune system in plants, fungi, invertebrates, and primitive multicellular organisms (Buchmann 2014). On the other hand, the adaptive immune system, found only in vertebrates (Kasahara et al. 2004; Herrin and Cooper 2010), is antigen-specific immune system. The T and B lymphocytes playing the central role in the jawed vertebrate adaptive immunity express a huge repertoire of binding specificity generated by somatic gene rearrangement (Tonegawa 1983; Davis 1990). Unlike innate immunity, the adaptive immunity provides long-lasting protection after first response to a specific antigen, and leads to an enhanced response to subsequent encounters with same antigen (Kalia et al. 2006).

The mammalian complement system comprises more than 30 serum and cell surface proteins, and plays a pivotal role in innate immunity (Volanakis 1998). The activation of central component C3 by the C3 convertase induces almost all physiologically functions of the complement system including opsonic activity. Upon proteolytic activation of C3, the thioester bond is exposed to the molecular surface and bind to the pathogen or other target molecules. Evolutionary studies thus far performed have indicated that the evolutionary origin of the complement system can be traced back to the common ancestor of the eumetazoa, because the genes for the central component C3, factor B (FB)—the serine protease responsible for C3 activation—and mannan-binding lectin-associated serine protease (MASP)—possibly involved in FB activation—were identified from sea anemones, cnidaria (Kimura et al. 2009; Putnam et al. 2007), whereas no complement gene is present in the genomes of the sponge, *Amphimedon queenslandica* (Srivastava et al. 2010), or the choanoflagellate, *Monosiga brevicollis* (King et al. 2008). Although the C3 gene has been identified from all deuterostome species analyzed thus far (Nonaka 2014), earlier genomic analyses showed its absence from the genomes of a nematoda, *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998), a water flea, *Daphnia pulex* (Colbourne et al. 2011) and a fruit fly, *Drosophila melanogaster* (Adams et al. 2000) indicating that this gene has been lost at least in some lineages of protostomes. On the other hand, the C3 gene has been reported from several other protostome species, such as the clam (Prado-Alvarez et al. 2009), the squid (Castillo et al. 2009), horseshoe crabs (Ariki et al. 2008; Zhu et al. 2005) and ticks (Buresova et al. 2011; Urbanová et al. 2015), and FB has also been reported from the clam (Prado-Alvarez et al. 2009) and horseshoe crab (Tagawa et al. 2012; Zhu et al. 2005) (Fig. 1).

C3 belongs to the thioester-containing protein (TEP) family; the other members of this family are the nonspecific protease inhibitor alpha-2 macroglobulin (A2M) (Dodds and Law 1998) and the glycosylphosphatidylinositol (GPI)-anchored protein

CD109 (Lin et al. 2002) whose function is still poorly understood. In addition, insect TEP (iTEP) (Blandin and Levashina 2004) has been reported from the fruit fly and mosquito. These species lack C3 and A2M, and certain iTEPs show opsonic activity similar to C3 (Levashina et al. 2001), suggesting that insects compensate for the loss of the complement system by expanding the functions of iTEPs. Later, iTEP was shown to be orthologous to CD109 by extensive phylogenetic analysis (Sekiguchi et al. 2012). Because C3, A2M and iTEP/CD109 genes have been identified in Cnidarian sea anemones (Fujito et al. 2010; Kimura et al. 2009; Putnam et al. 2007), the gene duplications that generated them, and the subsequent functional diversification seem to have been completed before divergence of the Cnidaria from the Bilateria (Fig.1). Whereas a set of C3, A2M and iTEP/CD109 genes have been identified from all deutrostome, the A2M gene has not been identified in some insect genomes sequenced thus far (Adams et al. 2000; Holt et al. 2002; International Aphid Genomics Consortium 2010), indicating that not only C3 but also A2M has been lost in some protostome lineages. To elucidate evolution of the complement and TEP genes in protostome, I focused on the biggest phylum Arthropoda. The phylum Arthropoda is estimated to contain 5–10 million extant species (Ødegaard 2000), which are classified into four subphyla, the Chelicerata, Myriapoda, Crustacea, and Hexapoda. Recent molecular phylogenetic studies strongly suggest that the Crustacea subphylum is actually paraphyletic, forming a clade, the Pancrustacea, together with the Hexapoda (Regier et al. 2010, Fig. 2).

In my Master thesis, I analyzed the C3 and TEP genes of a jumping spider, *Hasarius adansoni*, and a millipede, *Niponia nodulosa,* using RT-PCR and exhaustive cloning method, and reported that C3, A2M and iTEP/CD109 were present in *H.*

adansoni, whereas only A2M and iTEP/CD109 were present in *N. nodulosa* (Sekiguchi et al. 2012, Fig. 2). This result suggested that C3 gene has been lost not only in the crustacean and insect lineages but also in the millipede lineage. Here, to elucidate more detail of the evolution of complement and TEP genes in the arthropods, I selected species which belong to evolutionally important lineage in the arthropod, and performed exhaustive analysis of the complement and TEP genes using RT-PCR and RNA-seq.

Chapter 1

Evolution of the TEP genes in the arthropod revealed by molecular cloning

Abstract

The TEP gene family, composed of the C3, A2M and iTEP/CD109 subfamilies, has an ancient evolutionary origin, predating the divergence of Cnidaria and Bilateria. Whereas the iTEP/CD109 subfamily has been retained by all Protostomes and Deuterostomes analyzed thus far, the C3 and A2M subfamilies have been lost in some Protostome lineages. In the arthropod, the C3 subfamily members have been reported from horseshoe crab, spider and tick species, although several insect genomes deciphered to date lacked the C3 subfamily genes. To elucidate the evolution of these genes in the arthropod, TEP cDNAs were isolated from a sea spider, *Ammothea* sp. (Chelicerata), a centipede, *Scolopendra subspinipes* (Myriapoda) and sea lice, *Pseudocaligus fugu*, *Caligus* sp., and Pandaridae gen. et sp. (Crustacea) by reverse transcription polymerase chain reaction (RT–PCR) amplification using universal degenerate primers specific for the thioester region. Four different TEP genes were identified from *Ammothea* sp., *S. subspinipes*, and *P. fugu*, and two different TEP genes were from *Caligus* sp. and Pandaridae gen. et sp.. BLAST search and phylogenetic analysis using the deduced amino acid sequence of these TEPs indicated that all species possess C3 and A2M gene, and iTEP/CD109 gene was identified from *S. subspinipes*. Interestingly, C3s possessed by sea lice showed a closer similarity to teleost C3 than to other arthropod C3. These results indicated that the C3 gene was lost in the Diplopoda in Myriapoda and common ancestor of Pancrustacea or Altocrustacea , and sea lice species secondarily gained the C3 gene from teleost.

Introduction

The TEP members possess the unique intrachain thioester bond, which was originally found in the human serum protease inhibitor, A2M and the central component of the complement system, C3 (Dodds and Law 1998). The thioester bond is present in the inside of native TEP molecules. Upon proteolytic activation of TEP, the thioester bond is exposed to the molecular surface and forms a covalent bond with the molecules of pathogen or other targets. In addition to human A2M and C3, many genes of the eumetazoa have been shown to contain the thioester-encoding region, defining the TEP family. Seven members of this family are encoded in the human genome: C3, C4, C5, A2M (Dodds and Law 1998), pregnancy zone protein (PZP) (Sottrup-Jensen et al. 1984), the C3 and PZP-like A2M domain-containing 8 (CPAMD8) (Li et al. 2004) and CD109 (Lin et al. 2002). Although CD109 is a GPI-anchored membrane protein, all other TEPs are secreted proteins. In addition, insect TEP (iTEP), shown to be orthologous to CD109 by later phylogenetic analysis, have been reported from several insect species. Phylogenetic analysis of many TEP family genes from various eumetazoa indicated the presence of three subfamilies, C3, A2M and iTEP/CD109. The C3 subfamily comprises C3, C4, C5, the A2M subfamily comprises A2M, PZP and CPAMD8, and the iTEP/CD109 subfamily comprises iTEP/CD109 (Sekiguchi et al. 2012). Three-dimensional structural analysis of human C3 revealed the presence of an unpredicted macroglobulin (MG) domain, which repeats eight times and constitutes the core of the TEP family proteins (Janssen et al. 2005). In addition to these eight MG domains, a CUB domain (C1r, C1s, uEGF and bone morphogenetic protein domain) holding the TED (thioester domain) in the middle is inserted between the seventh and eighth MG domains (Janssen et al. 2005). Moreover, specific domains are present in each of the TEP members, such as ANA and C345C domains in C3 and the bait domain in A2M.

In the arthropod, molecular analyses of TEPs have been performed in several species; horseshoe crabs (Ariki et al. 2008; Iwaki et al. 1996; Zhu et al. 2005); (Chelicerata), a crayfish (Wu et al. 2012) (Crustacea), a water flea (Colbourne et al. 2011) (Crustacea), and some insects (Adams et al. 2000; Holt et al. 2002; International Aphid Genomics Consortium et al. 2010) (Fig. 2). On the other hand, no information on TEPs was available in Pycnogonida at the root of Chelicerata, Chilopoda (Myriapoda) and Maxillopoda (Crustacea) thus far.

In this chapter, to elucidate the evolution of the TEP genes in the arthropod, I performed exhaustive analysis of the TEP genes of a sea spider, *Ammothea* sp. (Pycnogonida), a centipede, *Scolopendra subspinipes* (Chilopoda) and sea lice, *Pseudocaligus fugu, Caligus* sp., and Pandaridae gen. et sp. (Maxillopoda) using RT-PCR and molecular cloning.

Materials and Methods

Animal collection and isolation of RNA

A sea spider, *Ammothea* sp. (abbreviated below to Amsp), and a sea lice, *Pseudocaligus fugu* (Psfu) were collected at the Misaki Marine Biological Station of the University of Tokyo in Kanagawa, Japan. The sea lice, *Caligus* sp. (Casp) and Pandaridae gen. et sp. (Pasp) were collected at Yokohama Kaihin Park in Kanagawa and the Shimoda Marine Research Center of the University of Tsukuba in Shizuoka, Japan, respectively. A centipede, *Scolopendra subspinipes* (Scsu) was collected in Tokyo, Japan. In the following, the abbreviations shown above in parentheses based on the species names are used as a prefix for each protein name. Total RNA was isolated from the whole body of each animal using ISOGEN (NIPPON GENE Co. Ltd., Tokyo, Japan), except for *S. subspinipes* with its large body, where approximately 10 μg of the gut from several segments was used for RNA extraction.

RT–PCR and cloning

cDNA synthesis from total RNA was performed at 42℃ for 1.5 h using ReverTra Ace (Toyobo Corp., Osaka, Japan). The sea spider, centipede, and sea liceTEP cDNAs were amplified with the degenerate primers designed based on the amino acid sequences of the thioester site conserved among the TEP proteins of various species. Two sense and four antisense degenerate primers were designed to cover all possible nucleotide sequences for the PSGCGEQ(I/T)NM and (T/V)(W/F)LTA(F/Y)VVKV amino acid sequences at and near the thioester site, respectively. The sense primers were 5’-CCCTCGGGNTGYGGNGARCARAAYATG-3’ and

5’-CCCTCGGGTTGYGGNGARCARANNATG-3’ (where N, R and Y represent a mix of G, A, T and C, a mix of G and A, or a mix of T and C, respectively). The antisense primers were 5’-CTTGACCACRAANGCNGTNAGCCANGT-3’, 5’-CTTGACCACRAANGCNGTNAGCCANAC-3’, 5’-GACCTTGACCACRTANGCNGTNARCCA-3’, and 5’-GACCTTGACCACRAANGCNGTNAGRAA-3’, (where W and represents a mix of A and T). All combination of sense and antisense primers were tested for each species. PCR amplification was performed as follows: 94℃ for 5 min and 40 cycles of 94℃ for 30 sec, 54℃ for 30 sec and 72℃ for 1 min, followed by final extension at 72℃ for 5 min. RT-PCR products were about 200 base pairs (bp) long, as expected. Amplified DNA was gel-purified and cloned into the pGEM-T easy vector (Promega, Madison, WI, USA).

Rapid amplification of cDNA ends (RACE)

5’ and 3’ RACE was performed using the SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA, USA) following the manufacture’s protocol with the following modifications. RACE primers were designed based on the nucleotide sequences obtained by the RT-PCR analyses of each animal. After treatment of cDNA at 95℃ for 5 min, Taq polymerase was added at 70℃. The reaction mixture was incubated at 60℃ for 1 min, then at 72℃ for 10 min, and was mixed with the gene specific primer. After 35 or 40 cycles of 94℃ for 40 sec, annealing temperature for 1 min and 72℃ for 2～3 min, final extension was performed at 72℃ for 5 min. Nucleotide sequences of all RACE products were determined by direct sequencing without cloning after treatment with exonuclease I and antarctic phosphatase (New

England Biolabs Japan Inc, Tokyo, Japan).

Genomic PCR for amplification of PsfuC3 gene

To analyze structure of the C3 gene of *P. fugu*, the sense and the antisense primers were designed based on cDNA sequence of PsfuC3 corresponding with exon 24 and 25 of human C3 gene, respectively. The sense primer was 5'-CTTGATATTTGGGCCATACAG-3' and the antisense primer was 5'-GATAGCGGCCACCTATTTGG-3'. Genomic PCR was performed with the following condition: 94℃ for 3 min and 35 cycles of 94℃ for 30 sec, 55℃ for 30 sec and 72℃ for 5 min, followed by final extension at 72℃ for 7 min.

DNA sequence analysis of the TEP genes

DNA sequences were determined using the BigDye Terminator Cycle Sequencing kit version3.1 (Applied Biosystems, Foster City, CA, USA) and an ABI 3100 DNA sequencer (Applied Biosystems).

Phylogenetic analysis of the TEP family genes

BLAST searches were performed on the blast network service at NCBI (http://www.ncbi.nlm.nih.gov). The amino acid sequences of the TEP genes deduced by this study were aligned with those of various vertebrate and invertebrate species, using ClustalX (Thompson et al. 1997), with some corrections made by eyes using MEGA6 (Tamura et al. 2013). Evolutionary analyses were conducted in MEGA6. A phylogenetic tree was inferred by the maximum likelihood (ML) method (Goldman et al. 2000) with complete deletion of gaps.

Results

Cloning, characterization and RACE-PCR of the *Ammothea* sp. TEP genes

As a result of cloning and sequencing analysis of TEP cDNA fragments, four different TEPs were identified from *Ammothea* sp.. Upon BLASTX search, one of them showed the closest similarity to horseshoe crab C3 (Zhu et al. 2005) and the other three to amphioxus A2M (Putnum et al. 2008). Therefore, I named these four TEPs as AmspC3, AmspA2M-1, AmspA2M-2, and AmspA2M-3 (Table 1). Out of 93 clones encoding the TEP genes obtained by the RT-PCR, 10 were AmspC3, 59 were AmspA2M-1, 14 were AmspA2M-2, and 10 were AmspA2M-3. To obtain the entire coding sequences for AmspC3, AmspA2M-1, AmspA2M-2, and AmspA2M-3, 5'- and 3'- RACE were performed. Only one primer was used for each 3' RACE of four *Ammothea* sp. TEP genes, since the entire coding sequence at the 3' side was obtained by the first cycle of 3' RACE. In contrast, 5' RACE was repeated multiple times for each spider TEP gene. In total, 4, 3, and 4 primers were used for 5' RACE for AmspC3, AmspA2M-1, and AmspA2M-2, respectively (Supplementary Fig. S1). Since 5' RACE did not work well, the only partial C-terminal side sequence was obtained for AmspA2M-3 (Supplementary Fig. S1). Sequence informations obtained by the RT–PCR and RACE analyses were connected to elucidate the entire coding sequences, and integrity of the connected sequences was confirmed by RT–PCR using primers designed at the 5' and 3' ends of the connected sequences. The deduced entire amino acid sequences of AmspC3, AmspA2M-1, and AmspA2M-2 consisted of 1725 residues, 1527 residues, and 1614 residues, respectively. The partial amino acid sequence of AmspA2M-3 consisted of 465 residues.

Cloning, characterization and RACE-PCR of the *S. subspinipes* TEP genes

As a result of cloning and sequencing analysis of TEP cDNA fragments, four different TEPs were identified from *S. subspinipes*. Upon BLASTX searching, one of them showed the closest similarity to horseshoe crab C3 (Zhu et al. 2005), another one to amphioxus A2M (Putnum et al. 2008) and the other two to mosquito and bee iTEP/CD109 (Werren et al. 2010), respectively. Therefore, I named these four TEPs as ScsuC3, ScsuA2M, ScsuiTEP/CD109-1, and ScsuiTEP/CD109-2, respectively (Table 1). Out of 267 clones encoding the TEP genes obtained by the RT-PCR, 249 were ScsuC3, 7 were ScsuA2M, 5 were ScsuiTEP/CD109-1, and 6 were ScsuiTEP/CD109-2. To obtain the entire coding sequences for ScsuC3, ScsuA2M, ScsuiTEP/CD109-1 and ScsuiTEP/CD109-2, 5’ and 3’ RACE analyses were performed. Only one primer was used for each 3’ RACE of four *S. subspinipes* TEP genes, since the entire coding sequence at the 3’ side was obtained by the first cycle of 3’ RACE. In contrast, 5’ RACE was repeated multiple times for each *S. subspinipes* TEP gene. In total, 3, 6, and 5 primers were used for 5’ RACE for ScsuC3, ScsuA2M, and ScsuiTEP/CD109-2, respectively (Supplementary Fig. S2). Since 5’ RACE did not work well with ScsuiTEP/CD109-1, only the partial C-terminal side sequence was obtained for this TEP (Supplementary Fig. S2). Nucleotide sequences obtained by the RT–PCR and RACE analyses were connected to elucidate the entire coding sequences, and integrity of the connected sequences was confirmed by RT–PCR using primers designed at the 5’ and 3’ ends of the connected sequences. The deduced entire amino acid sequences of ScsuC3, ScsuA2M-1, and ScsuiTEPCD109-2 consisted of 1749 residues, 1625 residues, and 1423 residues, respectively. The partial amino acid sequence of ScsuiTEP/CD109-2 consisted of 500 residues.

Cloning, characterization and RACE-PCR of the *P. fugu* TEP genes

As a result of cloning and sequencing analysis of TEP cDNA fragments, four different TEPs were identified from *P. fugu*. Upon BLASTX searching, one of them showed the closest similarity to killifish C3 and the other three to zebrafish, hydra and copepod A2M, respectively. Therefore, I named these four TEPs as PsfuC3, PsfuA2M-1, PsfuA2M-2 and PsfuA2M-3, respectively (Table 1). Out of 120 clones encoding the TEP genes obtained by the RT-PCR, 1 clone was PsfuC3, 5 were PsfuA2M-1, 4 were PsfuA2M-2, and 110 were PsfuA2M-3. To obtain the entire coding sequences for PsfuC3, PsfuA2M-1, PsfuA2M-2, and PsfuA2M-3, 5' and 3' RACE analyses were performed. Only one primer was used for each 3' RACE of PsfuC3, PsfuA2M-1, and PsfuA2M-3, since the entire coding sequence at the 3' side was obtained by the first cycle of 3' RACE, whereas two primers were used for 3' RACE of PsfuA2M-2. In contrast, 5' RACE was repeated multiple times for all *P. fugu* TEP genes. In total, 6, 5, 3, and 6 primers were used for 5' RACE for PsfuC3, PsfuA2M-1, PsfuA2M-2, and PsfuA2M-3, respectively (Supplementary Fig. S3). Since 5' RACE for PsfuC3 did not work after the 4th round, I designed the degenerate primer based on the amino acid sequence at the upstream, IQTDK(P/T)IYTP (Supplementary Fig. S3 and Fig. S9) which is highly conserved by various C3s, and performed RT-PCR. Following this RT-PCR, 5th and 6th round of 5' RACE were performed. Nucleotide sequences obtained by the RT–PCR and RACE analyses were connected to elucidate the entire coding sequences. The deduced entire amino acid sequences of PsfuC3, PsfuA2M-1, PsfuA2M-2, and PsfuA2M-3 consisted of 1655 residues, 1536 residues, 1341 residues, and 1717 residues, respectively. Interestingly, the entire amino acid sequence of PsfuC3 showed approximately 70% identity with several teleost C3 including puffer fish C3, although it showed only 30-40% identity with the other arthropod C3 sequences reported thus far. This result suggests that PsfuC3 was horizontally transferred from fish to *P. fugu*. The relative expression level of PsfuC3 to total TEP mRNA inferred from RT-PCR analysis was only 1/120. This level is much lower than those for AmspC3 (10/93) and ScsuC3 (249/267), and contamination of host RNA was suspected. However, partial amino acid sequences of two C3s of the host species, *Takifugu niphobles,* showed only 65-70% sequence identity with PsfuC3, excluding the possibility that PsfuC3 mRNA was contaminated from the host.

Genomic sequence of PsfuC3

I tried to amplify a part of the PsfuC3 gene, and obtained a partial genomic sequence of PsfuC3 including an intron (Fig. 3). This intron was inserted at the same position and in the same phase as the corresponding introns of fish and mammals. The presence of the intron indicated that the horizontal transfer of the C3 gene from fish to sea lice occurred not at the RNA level but at the DNA level.

Cloning, characterization and RACE-PCR of the *Caligus* sp. and the Pandaridae gen. et sp. TEP genes

To clarify whether other sea lice species also possess fish-like C3, two species of sea lice, *Caligus* sp. (parasitize rabbitfish) and Pandaridae gen. et sp. (parasitize banded houndshark), were analyzed. As a results of cloning and sequencing analysis of the TEP cDNA fragments, two TEPs were identified from both *Caligus* sp. and Pandaridae gen. et sp.. Upon the BLASTX search, one *Caligus* sp. TEP cDNA sequence showed the closest similarity to carp C3 (Nakao et al. 2000) and the other one to sea squirt A2M. For Pandaridae gen. et sp., one TEP cDNA sequence showed the closest similarity to croker C3 (Meng et al. 2012) and the other one showed the closest similarity to sea squirt A2M. Therefore, I named these TEPs as CaspC3, CaspA2M, PaspC3, and PaspA2M, respectively (Table 1). Out of 68 and 72 clones encoding the *Caligus* sp. and Pandaridae gen. et sp. TEP genes obtained by the RT-PCR, respectively, only one clone each was C3. To obtain the entire coding sequences for CaspC3, CaspA2M, PaspC3, and PaspA2M, 5' and 3' RACE was performed. Only one primer was used for each 3' RACE of CaspA2M, PaspC3, and PaspA2M, since the entire coding sequence at the 3' side was obtained by the first cycle of 3' RACE. In contrast, 5' RACE was repeated multiple times for CaspA2M, PaspC3, and PaspA2M. In total, 8 and 2 primers were used for 5' RACE for CaspA2M and PaspA2M, respectively (Supplementary Fig. S4 and S5). Since 5' RACE for PaspC3 did not work after 3rd round, I designed the degenerate primer based on the amino acid sequence at the upstream, FEVKEYVLP (Supplementary Fig. S5 and Fig. S9) which is highly conserved by various TEPs, and performed RT-PCR. Although 4th round of 5' RACE were performed following RT-PCR, the 5' end of PaspC3 mRNA was not reached. Since both 5' and 3' RACE did not work well, the available nucleotide sequence information of CaspC3 was only approximately 200bp obtained by the original RT-PCR with the degenerate primers (Supplementary Fig. S4). The deduced partial amino acid sequence of PaspC3 consisted of 1604 residues, and the deduced entire amino acid sequences of CaspA2M and PaspA2M consisted of 1735 residues and 1704 residues, respectively. The amino acid sequence of PaspC3 showed approximately 70% identity with various teleost C3 and approximately 42% identity with C3 of the host, banded houndshark, whereas it showed only 25-30% identity with the other arthropod C3. The amino acid sequence identity between PsfuC3 and PaspC3 was only 67%.

Primary structures of the arthropod TEPs

To clarify the conserved and diverged residues of TEPs, the deduced amino acid sequences of arthropod TEPs were compared with human TEPs using ClustalX program and GeneDoc (Supplementary Fig S9). All identified TEPs had the signal peptide region except for PsfuA2M-2, whose signal peptide was not found by SignalP (Supplementary Fig S1-S5). The TED domain containing the thioester site, GCGEQ, was present in all TEPs. AmspC3, ScsuC3, PsfuC3, and PaspC3 had the functionally-important domains and residues of C3 such as the β-α processing site, RXXR, the anaphylatoxin (ANA) domain with its conserved six cysteine residues and the C345C domain consisting of about 150 amino acids. A histidine residue catalyzing the cleavage of the thioester bond was found in all these C3s. The conservations of the characteristic domains and residues suggest that these C3s play basically the same functional role as human C3. As described above, the expression levels of PsfuC3 and PaspC3 seem to be very low, whereas human C3 is one of the most abundant serum protein. Therefore, it is likely that the physiological function of PsfuC3 and PaspC3, if any, is different from that of human C3.

A2M is characterized by the presence of the bait region highly susceptible to proteolytic cleavage by proteases of all catalytic classes, and the cleaved A2M shows conformational changes leading to trapping and inhibition of the attacking proteases. Although the amino acid sequences of the bait region show poor evolutionary conservation even among mammalian species (Sottrup-Jensen et al. 1989), the bait region is flanked by conserved residues at both sides: by a cysteine residue at the N-terminal side and by FPEXW at the C-terminal side. Except for PsfuA2M-1 and PsfuA2M-2, all arthropod A2Ms possess these conserved residues, indicating the presence of the bait region. The presence of the bait region suggested that these A2Ms play basically the same functional role as human A2M as a nonspecific protease inhibitor. Although PsfuA2M-1 and PsfuA2M-2 possessed FPESW, they lacked the cysteine residue conserved at the N-terminal side of the bait region of most A2M (Supplementary Fig. S3 and Fig S9). Therefore, PsfuA2M-1 and PsfuA2M-2 may lack the bait region, making it impossible to be a nonspecific protease inhibitor. In addition, the β-α processing site was present in AmspA2M-1 and AmspA2M-2, suggesting that these molecules have two subunits chain structure. AmspA2M-2 had 40 extra amino acid residues compared to other A2Ms at the upstream of the β-α processing site, although no functional motif was found in this region (Supplementary Fig S9). Interestingly, PsfuA2M-3, CaspA2M and PaspA2M possessed the CUB domain at the N-terminal side (Supplementary Fig. S3, S4, and S5). This domain structure was presented in a cyclops A2M (Refseq No.: AII16542), suggesting that unique domain shuffling of A2M occurred in copepod lineage, although the function of these CUB domains is unclear.

Arthropod iTEP/CD109s did not show a shared characteristic structure except for the typical thioester region. Because human iTEP/CD109 is a GPI-anchored protein, the possible presence of the GPI-anchor attachment signal sequence at the C-terminal of ScsuiTEP/CD109-1 and ScsuiTEP/CD109-2 was analyzed using GPI-SOM (http://gpi.unibe.ch/). The GPI-anchor attachment signal was not found in both ScsuiTEPCD109-1 and ScsuiTEP/CD109-2, suggesting that they are secreted proteins.

Phylogenetic analysis of the arthropod TEPs

To construct a phylogenetic tree of TEPs, the deduced amino acid sequences of arthropod TEPs were aligned with various eumetazoan TEP sequences by ClustalX using the MEGA6 platform. This alignment includes the TEP sequences of a sea anemone, arthropods shown in Fig. 2, and urochordate, cephalochordate, and vertebrate TEPs. Since only partial amino acid sequences were elucidated for AmspA2M-3, ScsuiTEP/CD109-1 and CaspC3, these TEPs were not included in this alignment. The phylogenetic tree was constructed based on this alignment using the ML method excluding any positions with gaps. As shown in Fig. 4, C3, A2M and iTEP/CD109 formed their respective clades supported by bootstrap percentages of 100%, 63%, and 99%, respectively. Although PsfuA2M-3 was located within the A2M clade clustering with sea lice A2Ms, CaspA2M and PaspA2M, PsfuA2M-1 and PsfuA2M-2 were located out of the A2M clade, suggesting that these PsfuA2Ms are highly diverged molecules. As mentioned above, PsfuA2M-1 and PsfuA2M-2 seem to lack the bait region essential for the function as a nonspecific protease inhibitor (Supplementary Fig. S3 and Fig. S9). Upon the BLASTP search, these molecules showed a closer similarity to A2M at the N-terminal side (70th ~300th aa residues) whereas they showed a closer similarity to iTEP/CD109 at the C-terminal side (600th ~1400th aa). In addition, the central part (300th ~600th aa) did not show any similarity to other TEPs. These results suggest that PsfuA2M-1 and PsfuA2M-2 derived directly from the common ancestor of A2M and iTEP/CD109. Since PsfuA2M-1 and PsfuA2M-2 could disturb the phylogenetic tree analysis, I constructed another phylogenetic tree excluding PsfuA2M-1 and PsfuA2M-2 by the ML method (Fig. 5). In this phylogenetic tree, C3, A2M, and iTEP/CD109 clades were supported by bootstrap percentages of 100%, 82%, and 100%, respectively.

Phylogenetic analysis of sea lice C3s

PsfuC3 and PaspC3 were located in the vertebrate C3 cluster, showing the closest relationship with medaka C3 (Fig. 4 and 5). To analyze the phylogenetic relationship between these C3 and other fish C3, a phylogenetic tree including various teleost and shark C3 was constructed using the NJ method (Saitou and Nei 1987) (Fig. 6). PsfuC3 and PaspC3 did not form the sea lice-specific cluster. Moreover, PsfuC3 did not show any close relationship to C3s of *Takifugu rubripes* (TaruC3-1, TaruC3-2, and TaruC3-3) closely related to the host species *T. niphobles*. Also, PaspC3 did not show any close relationship to C3s of its host, TrscC3A and TrscC3B. These results may indicate that the horizontal transfer of the C3 gene occurred long time ago from a common ancestor of teleost to a common ancestor of sea lice.

Discussion

Evolution of the TEP genes in the Arthropoda

This study identified the C3 and A2M genes from five arthropod species, *Ammothea* sp. (Chelicerata), *S. subspinipes* (Myriapoda), *P. fugu*, *Caligus* sp., and Pandaridae gen. et sp. (Crustacea), and the iTEP/CD109 gene only from *S. subspinipes*. The evolution of the TEP genes in the arthropod elucidated by this study as well as previous reports is summarized in Fig. 7. Since all three TEP genes, C3, A2M, and iTEP/CD109, are present in cnidaria, vertebrates and some arthropods (Sekiguchi et al. 2012), the common ancestor of the arthropod should have possessed these three TEP genes. However, all these three TEP genes seem to have experienced disappearance in various lineages of the arthropod. C3 was lost at least twice independently in the diplopod lineage and in the common ancestor of Pancrustacea or Altocrustacea. A2M was lost at least twice in Insecta, and iTEP/CD109 was lost twice in Pycnogonida and Maxillopoda. Although the physiological function of iTEP/CD109 is still to be clarified, it is established that C3 and A2M play totally independent physiological roles as the central component of the complement system and a non-specific protease inhibitor, respectively. At the present moment, it is puzzling why all these three genes were lost multiple times in the Arthropoda. Some iTEP/CD109 is reported to play a role as opsonin like C3 (Levashina et al. 2001), and further elucidation of the function of iTEP/CD109 is expected to help understand the curious evolution of the TEP genes in the Arthropoda.

In this study, C3s were identified from sea lice species, *P. fugu*, *Caligus* sp., and Pandaridae gen. et sp.. This is the first report on the crustacean C3. However, sea lice C3s showed a closer similarity to teleost C3 than to arthropod C3. In addition, genomic analysis of *P. fugu* C3 showed the presence of an intron. These results suggest that the C3 gene was transferred horizontally from fish to sea lice at the DNA level. A similar horizontal gene transfer from a host vertebrate to a parasite arthropod has been reported recently in a tick species which obtained vasodilatory hormone from reptiles or primitive mammals (Iwanaga et al. 2014). In the case of sea lice C3, three identified sequenced did not show a close similarity to each other, although all of them showed the closest similarity to teleost C3s. Further analyses are required to clarify whether horizontal gene transfer occurred multiple times or it occurred long time ago and the C3 sequence diverged after sea lice speciation.

Chapter 2:

Evolution of the complement gene in the Arthropoda revealed by *de novo* transcriptome analysis

Abstract

To elucidate the evolutionary history of the complement system in the Arthropoda, de novo transcriptome analysis was performed with eight species among the Chelicerata, Myriapoda, and Crustacea, and complement genes were identified based on their characteristic domain structures. Complement C3 and factor B (FB) were identified from a sea spider, a jumping spider, and a centipede, but not from a sea firefly, two sea lice species, and two millipede species. Whereas C3 was identified from sea lice using RT-PCR method, both C3 and FB were not from *Pseudocaligus fugu* and *Caligus* sp.. No additional complement components identifiable by their characteristic domain structures were found from any of these eight species. These results together with genome sequence information for several species of the Hexapoda suggest that the common ancestor of the arthropod possessed a simple complement system comprising C3 and FB, and thus resembled the alternative pathway of the mammalian complement system. It was lost at least twice independently during the evolution of the arthropod in the millipede lineage and in the common ancestor of crustacean and insect.

Introduction

The mammalian complement system comprises more than 30 serum and cell surface proteins, and plays a pivotal role in innate immunity (Volanakis et al., 1998). Evolutionary studies thus far have indicated that the evolutionary origin of the complement system can be traced back to the common ancestor of the Eumetazoa, because the genes for the central component C3, factor B (FB), the serine protease responsible for C3 activation, and mannan-binding lectin-associated serine protease (MASP)—possibly involved in FB activation—were identified from sea anemones, Cnidaria (Kimura et al. 2009; Putnam et al. 2007), whereas no complement gene is present in the genomes of the sponge, *Amphimedon queenslandica* (Srivastava et al. 2010), or the choanoflagellate, *Monosiga brevicollis* (King et al. 2008). Although the C3 gene has been identified from all deuterostome species analyzed thus far (Nonaka 2014), earlier genomic analyses showed its absence from the genomes of *Drosophila melanogaster* (Adams et al. 2000) and *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998), indicating that this gene has been lost in at least some lineages of protostomes. On the other hand, the C3 gene has been reported from several other protostome species, such as the clam (Prado-Alvarez et al. 2009), the squid (Castillo et al. 2009), horseshoe crabs (Ariki et al. 2008; Zhu et al. 2005), the spider (Sekiguchi et al. 2012) and ticks (Buresova et al. 2011; Urbanová et al. 2014), and FB has also been reported from the clam (Prado-Alvarez et al. 2009) and horseshoe crab (Tagawa et al. 2012; Zhu et al. 2005). However, to date, no comprehensive analysis of the complement genes has been performed in protostome species possessing the C3 gene and it is still an unsolved question as to whether the protostome complement system shares complement components other than C3 and FB with deuterostomes.

C3 belongs to the thioester-containing protein (TEP) family; the other members of this family are the nonspecific protease inhibitor alpha-2 macroglobulin (A2M) (Dodds and Law 1998) and the glycosylphosphatidylinositol (GPI)-anchored protein CD109 (Lin et al. 2002) whose function is still poorly understood. In addition, insect TEP (iTEP) (Blandin and Levashina 2004) has been reported from the fly and mosquito, which lack C3 and A2M, and certain iTEPs show opsonic activity similar to C3 (Levashina et al. 2001), suggesting that insects compensate for the loss of the complement system by expanding the functions of iTEPs. Later, iTEP was shown to be orthologous to CD109 by extensive phylogenetic analysis (Sekiguchi et al. 2012). Macroglobulin complement-related (Mcr) found from *D. melanogaster* shows overall structural similarity to these TEP family members, although it lacks the thioester site and possesses low density lipoprotein receptor class A (LDLa) domain not found in C3, A2M or iTEP/CD109 (Stroschein-Stevenson et al. 2006). Therefore, here I treat the C3, A2M, and iTEP/CD109 subfamilies as authentic members of the TEP family, and Mcr as the closest relative of the TEP family. C3 and pregnancy zone protein (PZP)-like A2M domain-containing 8 (CPAMD8) (Li et al. 2004) thus far found only from deuterostomes are included in the A2M subfamily (Fujito et al. 2010). Because all these three subfamily genes have been identified in Cnidarian sea anemones (Fujito et al. 2010; Kimura et al. 2009; Putnam et al. 2007), the gene duplications that generated them, and the subsequent functional diversification seem to have been completed before divergence of the Cnidaria from the Bilateria. The A2M gene has not been identified in some insect genomes sequenced thus far (Adams et al. 2000; Holt et al. 2002; International Aphid Genomics Consortium 2010), indicating that not only C3 but also A2M has been lost in some of these lineages. In contrast, all three TEP genes have been reported from a spider (Sekiguchi et al. 2012) and ticks (Buresova et al. 2011; Urbanová et al. 2014), so the evolutionary history of TEP genes in the Arthropoda is still unresolved.

The phylum Arthropoda is estimated to contain 5–10 million extant species (Ødegaard 2000), which are classified into four subphyla, the Chelicerata, Myriapoda, Crustacea, and Hexapoda. Recent molecular phylogenetic studies strongly suggest that the Crustacea subphylum is actually paraphyletic, forming a clade, the Pancrustacea, together with the Hexapoda (Regier et al. 2010). Here I performed RNA sequencing (RNA-seq) analysis of eight species belonging to the Chelicerata, Myriapoda, and Crustacea to elucidate the evolution of complement and TEP genes in the Arthropoda.

Materials and Methods

Animal collection and isolation of RNA

A sea spider, *Ammothea* sp. (abbreviated below to Amsp), a sea firefly, *Vargula* sp. (Vasp), and a sea lice *Pseudocaligus fugu* (Psfu) were collected at the Misaki Marine Biological Station of the University of Tokyo in Kanagawa, Japan. Another sea lice, *Caligus* sp. (Casp) was collected at Yokohama Kaihinn Park in Kanagawa. A jumping spider, *Hasarius adansoni* (Haad), a centipede, *Scolopendra subspinipes* (Scsu), and the millipedes *Niponia nodulosa* (Nino) and *Epanerchodus* sp. (Epsp) were collected in Tokyo, Japan. In the following material, the abbreviations shown above in parentheses based on the species names are used as a prefix for each protein name. Total RNA was isolated from the whole body of each animal using ISOGEN (NIPPON GENE Co. Ltd., Tokyo, Japan), except for *S. subspinipes* with its large body, where approximately 10 μg of the gut from several segments was used.

Sequencing, data processing, and de novo assembly

Construction of cDNA libraries and sequencing were performed by the Beijing Genomics Institute (BGI; Shenzhen, China). The cDNA libraries were constructed using Illumina TruSeq RNA library preparation kits (Illumina Inc., San Diego, CA, USA), and pair-end sequencing (2 × 90 bp) was performed on an Illumina HiSeq 2000 platform. The adaptor sequences were removed from raw reads, and low quality reads (quality value ≤ 10 for more than 20%) were removed using filter_fq (BGI internal software). High quality reads were obtained and assembled using the Trinity program (Grabherr et al. 2011).

Gene annotation and estimation of expression levels

To annotate the complement and TEP genes from such massive amounts of sequence data, I used the local Basic Local Alignment Search Tool (BLAST) program. Local BLAST software was downloaded from the NCBI website (http://blast.be-md.ncbi.nlm.nih.gov/Blast.cgi). A database was constructed from the assembled contigs of each animal, and the amino acid sequences of eumetazoan complement component or TEP were used as queries for tBLASTN. To estimate the expression level of the annotated genes, Bowtie (ver. 1.0.0) (Langmead et al. 2009) and eXpress (ver. 1.5.1) (Roberts and Pachter 2013) software were used to calculate the values of 'fragments per kilobase of exon per million mapped fragments' (FPKM).

Phylogenetic analysis of the complement and TEP genes

The amino acid sequences of the C3, FB, A2M, and iTEP/CD109 proteins of various animal species occupying critical phylogenetic positions for understanding evolution were obtained from the NCBI database (see above). These sequences were aligned with the amino acid sequences of these proteins of the eight arthropod species deduced in this present study using ClustalX (http://www.clustal.org/clustal2) (Thompson et al. 1997). Some manual corrections of the alignments and following evolutionary analyses were conducted in MEGA6 (http://www.megasoftware.net) (Tamura et al. 2013). Upon model testing, the lowest Bayesian Information Criterion (BIC) score model was used and a phylogenetic tree was constructed using the maximum likelihood (ML) method (Goldman et al. 2000).

Results

Generation of raw reads and assembly of reads

The raw reads generated on the Illumina Hiseq 2000 platform were trimmed for adaptor sequences, and low quality reads were excluded from subsequent analyses. High quality reads were assembled using Trinity software. As shown in Table 2, approximately 50 million high quality reads were obtained for *H. adansoni*, *S. subsupinipes*, *N. nodulosa* and *Caligus* sp., and approximately 90 million high quality reads were obtained for *Ammothea* sp., *Epanerchodus* sp., *Vargula* sp., and *P. fugu*, which were assembled into 37,757 to 155,223 contigs. The mean contig length and N50 values (defined as the summed lengths of scaffolds/contigs from the longest to the shortest, and reaching 50% of the total assembly size) are also shown in Table 2. The values of *Vargula* sp. and *H. adansoni* were low, most probably because of the difficulty in obtaining high quality RNA samples from these species.

TEP family

To annotate the C3, A2M, and iTEP/CD109 proteins of each species, the amino acid sequences of arthropods and human TEP were used as queries for tBLASTN against local BLAST databases, which consisted of the assembled contigs of each arthropod species. The TEPs identified from each arthropod species are shown in Table 3. A2M and iTEP/CD109 were identified from all analyzed species, whereas C3 was identified only from *Ammothea* sp., *H. adansoni*, and *S. subspinipes*. The deduced protein sequences are named below as four-character prefixes representing the species name plus protein name. When multiple isotypes were present, numbers are given as suffixes. Thus, the six TEPs identified from *Ammothea* sp. have been designated AmspC3, AmspA2M-1, AmspA2M-2, AmspA2M-3, AmspA2M-4, and

AmspiTEP/CD109. ScsuC3 and CaspA2M were renamaed ScsuC3-1 and CaspA2M-1, respectively. Nearly entire amino acid sequences were determined for all TEP proteins analyzed here, except for PsfuA2M-1, PsfuA2M-2, VaspA2M-1, VaspA2M-2, and VaspiTEP/CD109 (Supplementary Fig. S1-S8). Only partial amino acid sequences were obtained for these TEPs of *P. fugu* and *Vargula* sp., presumably because of poor total RNA quality. Although PsfuC3 and CaspC3 were identified by RT-PCR method, not detected by RNA-seq. This suggests that PsfuC3 and CaspC3 showed very low expression level. The deduced amino acid sequences of arthropod TEPs were aligned with eumetazoan TEP sequences by ClustalX using the MEGA 6 program to examine for conservation of the characteristic domains, motifs, and residues of those TEPs known to have functional importance in mammals (Supplementary Fig. S9). The results of predictions of these domains, motifs, and residues are summarized in Table 4. All TEPs have a typical thioester sequence, GCGEQ, except for ScsuC3-2, whose thioester site is ACGEQ. AmspC3, HaadC3-1, HaadC3-2, ScsuC3-1, and ScsuC3-2have the motif sequences and domains of typical C3, such as the β-α cleavage site, RXXR, the anaphylatoxin domain (ANA) with its conserved six cysteine residues and the C345C domain consisting of about 150 amino acids, whereas ScsuC3-3 lacks the C345C domain. A histidine residue catalyzing the cleavage of the thioester bond was found in AmspC3, HaadC3-1, and ScsuC3-1. On the other hand, this histidine residue has been substituted with glutamic acid, glycine, and serine in HaadC3-2, ScsuC3-2, and ScsuC3-3, respectively. As shown by the FPKM values in Table 4, histidine-type C3s showed significantly higher expression levels than the other C3s. A2M is characterized by the presence of the bait region highly susceptible to proteolytic cleavage by proteases of all catalytic classes, and the cleaved A2M shows conformational changes leading to trapping and inhibition of the attacking proteases. Although the amino acid sequences of the bait region show poor evolutionary conservation even among mammalian species (Sottrup-Jensen et al. 1989), the bait region is flanked by conserved residues at both sides: by a cysteine residue at the N-terminal side and by FPETW at the C-terminal side. All arthropod A2Ms where the entire coding sequences were deciphered have these conserved residues, indicating the presence of the bait region, except for PsfuA2M-1 and PsfuA2M-2. Interestingly, PsfuA2M-3, CaspA2M-1, and CaspA2M-2 possessed CUB domain at N-terminal side (Supplementary Fig. S9). The β-α cleavage site indicative of the two-subunit chain structure is present in all C3s, AmspA2M-1, AmspA2M-2, AmspA2M-4, AmspiTEP/CD109, HaadA2M-1, HaadiTEP/CD109-1, HaadiTEP/CD109-2, NinoiTEP/CD109-1, and EpspiTEP/CD109-1. Because human iTEP/CD109 is a GPI-anchored protein, I analyzed the presence of the GPI-anchor attachment signal sequence at the C-terminal of each iTEP/CD109 using GPI-SOM (http://gpi.unibe.ch/). The GPI-anchor attachment signal was found in AmspiTEP/CD109, HaadiTEP/CD109-1, HaadiTEP/CD109-2, ScsuiTEP/CD109-3, NinoiTEP/CD109-2, EpspiTEP/CD109-2, and VaspiTEP/CD109. In contrast, ScsuiTEP/CD109-1, ScsuiTEP/CD109-2, ScsuiTEP/CD109-4, NinoiTEP/CD109-1, NinoiTEP/CD109-3, and EpspiTEP/CD109-1 did not have this attachment signal, suggesting that they are secreted proteins. Interestingly, the possible secreted proteins—NinoiTEP/CD109-1, NinoiTEP/CD109-3, and EpspiTEP/CD109-1—showed high FPKM values indicating high levels of expression (Table 4).

Phylogenetic analysis of arthropod TEPs

To construct a phylogenetic tree of TEPs, the deduced amino acid sequences of arthropod TEPs were aligned with various eumetazoan TEP sequences by ClustalX using the MEGA6 program. Only complete amino acid sequences were used, and PsfuA2M-1, PsfuA2M-2, VaspA2M-1, VaspA2M-2, and VaspiTEP/CD109 with partial sequence information were not included. The phylogenetic tree was constructed by the ML method based on a Le and Gascuel (LG) model (Le et al. 2008) excluding any positions with gaps. As shown in Fig. 8, C3, A2M, and iTEP/CD109 formed their respective clades supported by bootstrap percentages of 100%, 78%, and 100%, respectively, indicating that identification based on Blast search and domain structure was correct. Within the C3 clade, arthropod C3s formed a clade supported by a bootstrap percentage of 99%. In contrast, the arthropod A2M and arthropod iTEP/CD109s clades were supported by a bootstrap percentage of <50%. A neighbor-joining (NJ) tree (Saitou and Nei, 1987) based on p-distances showed a closely resembling branching pattern, and bootstrap percentages supporting the C3, A2M, and iTEP/CD109 clades were 99%, 94%, and 92%, respectively (Fig. 9). Two, one, two, two, two, one, one, and one Mcrs were found from *Ammothea* sp., *H. adansoni, S. subspinipes, N. nodulosa, Epanerchodus* sp., *Vargula* sp., *P. fugu*, and *Caligus* sp., respectively. Mcr sequences of these species and those of some arthropod species were also included in this NJ tree. Mcrs formed a well-supported clade with a long branch indicating a remote evolutionary relationship to the TEP family. In most cases, multiple isotypes of each species formed clusters, suggesting that gene duplications leading to multiple isotypes occurred in each species lineage. However, at least two lineages for iTEP/CD109 are shared by *S. subspinipes*, *N. nodulosa*, and *Epanerchodus* sp. (Fig. 8 and 9).

Complement components

To annotate the complement components with characteristic domain structures of each arthropod species, amino acid sequences of mammalian complement components, C1q, MBL, C1r/C1s/MASP, factor I (FI), FB/C2, and C6/C7/C8/C9, were used as queries for the tBLASTN search of the assembled contigs of each arthropod species. Two, three, and two FBs were identified from *Ammothea* sp., *H. adansoni*, and *S. subspinipes*, respectively. Although individual domains composing mammalian complement components—such as the carbohydrate recognition domain (CRD), C1r/C1s, uEGF and bone morphogenetic protein (CUB), epidermal growth factor (EGF)-like, short consensus repeat (SCR), serine protease (SP), scavenger receptor (SR), low-density lipoprotein receptor domain class A (LDLA), von Willebrand factor type A (vWFA), and thrombospondin type 1 repeats (TSP1)—were found in many deduced arthropod proteins, characteristic combinations of domain structures found only in complement components, C1q, MBL, C1r/C1s, MASP, FI, and C6/C7/C8/C9, were not identified from these six arthropod species. Therefore, I conclude that *Ammothea* sp., *H. adansoni*, and *S. subspinipes* have C3 and FB, and that *N. nodulosa*, *Epanerchodus* sp., and *Vargula* sp. have no complement component. Although C3 gene was identified from sea lice, *P. fugu*, *Caligus* sp., by RT-PCR and comprehensive cloning approach, both C3 and FB were not from sea lice by RNA-seq.

Comparisons and phylogenetic analysis of arthropod FB sequences

Entire coding sequences of FB gene were obtained from *Ammothea* sp. FBs (Supplementary Fig. S10), *H. adansoni* FBs (Supplementary Fig. S11) and *S. subspinipes* FBs (Supplementary Fig. S12). The deduced amino acid sequences of arthropod FBs were aligned with the human FB amino acid sequence (Supplementary Fig. S13). AmspFB-1, HaadFB-1, and ScsuFB-1 had seven SCR domains whereas AmspFB-2, HaadFB-3, and ScsuFB-2 had only two. *H. adansoni* had one more FB sequence with four SCR domains, HaadFB-2. Interestingly, the three residues forming the catalytic triad of serine proteases, histidine, aspartic acid, and serine, were substituted with serine, asparagine, and valine, respectively in HaadFB-2, and two of them—histidine and serine—were substituted with glutamine and aspartic acid, respectively in ScsuFB-2. Therefore, it is highly unlikely that HaadFB-2 and ScsuFB-2 possess protease activity.

To elucidate the evolutionary relationships among arthropod FBs and eumetazoan FB and C2, an NJ tree was constructed based on p-distances (Fig. 10). The arthropod FBs formed two clades supported by bootstrap percentages of 50% and 100%, respectively. All FBs in clade 1 have seven SCR domains except for HaadFB-2 and TatrFB-1, which possess four and five SCR domains, respectively. All three FBs in clade 2 have two SCR domains and two of them, HaadFB-2 and ScsuFB-2, seem to have lost protease activity. This phylogenetic tree suggests that the common ancestor of the Arthropoda had two lineages of FB sequences, one with seven SCR domains and the other with two.

Discussion

The distribution of the complement and TEP genes in the Arthropoda revealed by the present study and preceding studies including genome analyses in several insect species is summarized in Fig. 11 (Adams et al. 2000; Ariki et al. 2008; Buresova et al. 2011; Colbourne et al. 2011; Honeybee Genome Sequencing Consortium 2006; International Aphid Genomics Consortium 2010; Iwaki et al. 1996; Tribolium Genome Sequencing Consortium et al. 2008; Wu et al. 2012; Zhu et al. 2005). Here, + indicates the presence of any kind of evidence obtained by targeted reverse transcription polymerase chain reaction (RT–PCR), RNA-seq, or genome analysis, whereas a – symbol indicates failure of detection by the comprehensive methods such as RNA-seq or genome analysis. The ? symbol indicates that no comprehensive analysis has been performed in that species thus far, and the presence or absence of that gene is still unclear. As shown in Fig. 11, the only complement genes present in the Arthropoda are C3 and FB. No other complement genes with characteristic domain structures—including MASP present in some Cnidaria (Kimura et al. 2009; Putnam et al. 2007)—were identified by the present study or by genome analyses in several arthropod species. These results suggest that the common ancestor of the Arthropoda had a simple complement system comprising C3, FB, and a few additional components, in which FB most probably activates C3. At least one additional protease that activates FB proteolytically seems to be essential. In the mammalian complement system, FB is activated by factor D (FD) (Volanakis and Narayana 1996), and the FD cleavage site is conserved in all the arthropod FB sequences (Supplementary Fig. S13). However, human FD has a simple structure composed of a single serine protease domain only, and I could not identify any FD ortholog in the Arthropoda based on domain architecture.

Both C3 and FB have been retained by the Chelicerata and Chilopoda, whereas they seem to have been lost twice simultaneously and independently in the diplopod lineage and the common ancestor of the Pancrustacea (Fig. 11). Although why the complement system is not needed in these lineages is not clear at present, it is interesting to note that some iTEP/CD109s of flies and mosquitoes, without the GPI-anchor signal and therefore considered to be secreted proteins, are reported to show opsonic activity similar to C3. In the present study, a high level of expression (>100 FPKM) of secretion-type iTEP/CD109 was also observed for *Niponia nodulosa* and *Epanerchodus* sp. lacking C3, but not for *Ammothea* sp., *Hasarius adansoni*, and *Scolopendra subspinipes* possessing C3. Therefore, it is possible that the secretary iTEP/CD109 of the arthropod executes opsonic activity as a proxy for C3, although experimental evidence is still missing for millipede iTEP/CD109.

Compared with genome analysis, the RNA-seq analysis adopted in this study leaves some ambiguity when the sought molecule is not detected. Moreover, because I had to extract RNA from the whole bodies of most species because of their small size, I cannot exclude the possibility that undetected transcripts were actually expressed at low levels and/or locally. However, I still consider that an RNA-seq analysis is valid to trace the evolutionary history of the complement system, because large amounts of C3 seem to be essential for an opsonic role in the body’s defense system. Indeed, human C3 is one of the most abundant proteins in serum. Therefore, even if a minute amount of C3 was present in millipede, sea firefly, and sea lice, eluding my RNA-seq analysis, it is highly unlikely that it would play a role as a complement component.

The activation mechanism of the protostome complement system has been extensively analyzed in the horseshoe crab (Ariki et al. 2008; Tagawa et al. 2012; Zhu et al. 2005). Although both C3 and FB are present in horseshoe crab, there are two C3 activation pathways, one FB independent and the other FB dependent. Gram-negative bacteria activate the FB-independent pathway in which Factor C (FC)—originally identified as a lipopolysaccharide (LPS)-sensitive trigger of the hemolymph coagulation pathway—directly activates C3 (Ariki et al. 2008). On the other hand, Gram-positive bacteria activate the FB-dependent pathway, which also requires Ca2+-dependent lectins (Tagawa et al. 2012). Interestingly, FC-like genes were identified from *Ammothea* sp., *H. adansoni*, and *S. subspinipes*, which all possess a C3 sequence (Fig. 12). All the domains of *Tachypleus tridentatus* (Tatr) FC were present in AmspFC-like, although these two molecules showed some difference in the position of the SCR domain. HaadFC-like and ScsuFC-like lacked the CLECT (C-type lectin domain), and LCCL (Limulus factor C, Coch-5b2, and Lgl1) plus CLECT domains, respectively. Moreover, FC-like was identified from two tick species *(Ixodes scapularis* and *Ixodes richius*) which possess C3 gene (Urbanová et al. 2014). These results indicate the possibility that the FC-like proteins of these animals are also involved in activation of the complement system. It is interesting to note that FC was lost in the Pancrustacea lineage simultaneously with FB and C3 (Fig. 11). However, FC-like molecules possessing a strikingly similar domain architecture to TatrFC were also identified from two millipede species, *N. nodulosa* and *Epanerchodus* sp. lacking the C3 and FB genes (Fig. 11). Therefore, FC-like proteins of these animals likely have the major physiological role outside of the complement system.

This study was the first comprehensive search for complement components among protostome species possessing C3. Only FB was identified, suggesting that FI and later components were innovations in the deuterostome lineage, and that

MASP—present in the common ancestor of the Eumetazoa—was lost by the time of appearance of the Arthropoda. This simple complement system resembling to the mammalian alternative pathway was lost at least twice independently during arthropod evolution.

General Discussion

Evolutionary scenario of the complement system

The evolutionary process of the complement system elucidated by the present study as well as by the previous works (Nonaka and Kimura 2006; Nonaka 2014) is schematically shown in Fig. 13. Since C3, FB and MASP have been identified from sea anemone, the common ancestor of eumetazoa should have possessed the primitive complement system comprising at least these three components. Conservation of the protein domains and amino acid residues critical for basic functions of these components between cnidaria and humans suggests that the basic activation mechanism and physiological functions of the primitive complement system were similar to those of the human complement system. Thus, MASP appears to be the first protease to be activated, which in turn activates FB. Then FB activates C3 into the C3a and C3b fragments. C3b covalently tags microbes and enhances phagocytosis, whereas C3a induces inflammation as an anaphylatoxin. I performed the first comprehensive search for the complement genes in protostome species which possess the complement system, and identified only C3 and FB. Thus, MASP seems to be lost in the protostome lineage before the emergence of the arthropod, leaving the activating protease of FB in arthropod still to be clarified. In addition, absence of FI and C6 in protostome strengthened the previous conclusion that these components were innovated in the vertebrate lineage (Nonaka 2014). FI and C6 most probably endowed the vertebrate complement system with the regulatory mechanism and the cytolytic activity, respectively, rendering it much more sophisticated compared to that of protostomes. In deuterostomes, the complement system is retained by all members analyzed thus far. In contrast, the central component of the complement system, C3, has been lost at least three times independently in the protostome lineage (Fig. 13). Since C3 plays so pivotal role in the complement system that loss of C3 is interpreted as loss of the complement system itself. In addition to the recurrent loss in the protostome lineage, the complement system was lost also in cnidarian hydra. The reason why the complement system was lost so many times is still not clear. As discussed in Chapter 2, in some insects, iTEP/CD109 is reported to be multiplied and play an opsonic role like C3 (Levashina et al. 2001). Thus, it is possible that some protostomes and cnidaria developed unique immune mechanisms, making it unnecessary to retain the complement system.

Evolution of the TLR and PGRP gene in the Arthropoda

The complement system is one of the most ancient innate immune system conserved by cnidaria and vertebrate. However, it was lost at least twice during evolution of arthropods. To clarify whether a similar loss is observed in other innate immune gene in the Arthropoda or not, I analyzed Toll-like receptor (TLR) and peptidoglycan-recognition protein (PGRP) in eight arthropod species using the RNA-seq data (Table 4 and 5). Toll was originally identified as a developmental gene involved in the dorso-ventral axis formation of the fruit fly, *Drosophila* (Anderson et al. 1985), and later analyses revealed its function in innate immune response (Lemaitre et al. 1996; Tauszig et al. 2000). PGRP was first identified in the moth, *Bombyx mori* as a protein to bind peptidoglycan, the major structural component of the bacterial cell wall (Yoshida et al. 1996). Both TLR and PGRP are present in various invertebrates and vertebrates, and many of them were shown to play innate immune functions (Medzhitov et al. 1997: Kang et al. 1998). PGRPs are classified into the long and short types (Werner et al. 2000). As a result of BLAST search, TLR genes, characterized by the combination of the leucine-rich repeat (LRR) and Toll/interleukin-1 receptor (TIR) domains, were identified from all eight species. When multiple isotypes were present, arbitrary numbers are given as suffixes. To elucidate the relationship of the TLR genes, the phylogenetic tree was constructed (Fig. 14). As shown in Fig. 14, the phylogenetic tree of TLR genes consists of three clades supported by the bootstrap percentages of 68%, 99%, and 95% (Fig. 14). Clade 3 contained sea anemone TLR, human TLRs, and some arthropod TLRs, suggesting that this is the evolutionary conserved clade. Clades 1 and 2 contained only Arthrpoda members, suggesting that they were innovated in the arthropod lineage. Interestingly, some arthropod such as *D. melanogaster* without the complement system also lack the clade 3 TLR. Unlike the TLR gene, PGRP genes detected from only limited species. *Hasarius adansoni*, *Niponia nodulosa*, and *Epanerchodus* sp. had only the short type PGRPs, and *Scolopendra subspinipes* possessed both the short and long types. *Ammothea* sp., *Vargula* sp., *Pseudocaligus fugu*, and *Caligus* sp. did not have any PGRP. In addition, PGRP was not detected by genome analysis of a water flea, *Daphnia pulex* (Colbourne et al. 2011), and extensive pull-down binding assay of a shrimp, *Penaeus monodon* (Udompetcharaporn et al. 2014). These results suggest that the PGRP genes have been lost in some lineage of the arthropod. Interestingly, all arthropods without the PGRP genes are marine animals. Udompetcharaporn et al. (2014) reported the possible substitution molecule of PGRPs from *P. monodon*, and it is possible that other marine arthropods without PGRP also possess similar substitution molecules. To analyze the phylogenetic relationship of arthropod PGRPs, the phylogenetic tree was constructed (Fig. 15). This phylogenetic tree indicated that almost of the arthropod PGRPs were lineage specific-evolved. In addition, long type *S. subspinipes* PGRP, ScsuPGRP-2, was clustered with the fly and the mosquito PGRP-LD, whereas the bootstrap value supported by only 33%. These analyses revealed that not only the complement genes, but also the TLR and PGRP genes have been lost in several arthropod lineages. Therefore, the innate immune system of protostome seems to have experienced much more drastic evolutionary changes than the deuterostome counterpart.

Acknowledgments

I wish to express my deepest gratitude to Prof. Masaru Nonaka for his thoughtful suggestion and patient guidance throughout the course of this study. Thanks are also due to Dr. Mayumi Nonaka, Ms. Naoko Fujito, and the other members of laboratory of Evolutionary Immunology for their help and significant advice. I appreciate Dr. Akio Tanigawa (Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan) identifying species of the spider, *Hasarius adansoni*. I also thank Mr. Mamoru Sekifuji, Mr. Hisanori Kohtsuka, Ms. Natsuko Ito (The Misaki Marine Biological Station, The University of Tokyo, Kanagawa, Japan), and Mr. Ichiro Oshio for supplying me with a sea firefly, *Vargula* sp., and a millipede, *Niponia nodulosa*. I would like to thank Dr. Naoki Irie (The University of Tokyo) for providing the guidance on the technique of RNA-seq analysis.

References

Adams MD, Celniker SE, Holt RA et al., 2000. The genome sequence of *Drosophila melanogaster*. Science. 24, 2185-2195.

Anderson KV, Jurgens G, Nusslein-Volhard C, 1985. Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. Cell. 42, 779–789.

Ariki S, Takahara S, Shibata T, Fukuoka T, Ozaki A, Endo Y, Fujita T, Koshiba T, Kawabata S, 2008. Factor C acts as a lipopolysaccharide-responsive C3 convertase in horseshoe crab complement activation. J Immunol.181, 7994-8001.

Blandin S, Levashina EA, 2004. Thioester-containing proteins and insect immunity. Mol Immunol. 40, 903-908.

Buchmann K, 2014. Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals. Front Immunol. 5, 459.

Buresova V, Hajdusek O, Franta Z, Loosova G, Grunclova L, Levashina EA, Kopacek P, 2011. Functional genomics of tick thioester-containing proteins reveal the ancient origin of the complement system. J Innate Immun. 3, 623-630.

Castillo MG, Goodson MS, McFall-Ngai M, 2009. Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. Dev. Comp. Immunol. 33, 69-76.

C. elegans Sequencing Consortium, 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. Science. 282, 2012-2018

Colbourne JK, Pfrender ME, Gilbert D et al., 2011. The ecoresponsive genome of *Daphnia pulex*. Science. 331, 555-561

Davis MM, 1990. T cell receptor gene diversity and selection. Annu Rev Biochem. 59, 475-496.

Dodds AW, Law SK, 1998. The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and alpha 2-macroglobulin. Immunol Rev. 166, 15-26.

Fujito NT, Sugimoto S, Nonaka M, 2010. Evolution of thioester-containing proteins revealed by cloning and characterization of their genes from a cnidarian sea anemone, *Haliplanella lineate*. Dev Comp Immunol. 34, 775-784.

Grabherr MG, Haas BJ, Yassour M et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 29, 644-652.

Goldman N, Anderson JP, Rodrigo AG, 2000. Likelihood-based tests of topologies in phylogenetics. Syst Biol. 49, 652-670.

Herrin BR, Cooper MD, 2010. Alternative adaptive immunity in jawless vertebrates. J Immunol. 185, 1367-1374.

Holt RA, Subramanian GM, Halpern A et al., 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science. 298, 129-149.

Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature. 443, 931-949.

International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. PLoS Biol. 8, e1000313.

Iwaki D, Kawabata S, Miura Y, Kato A, Armstrong PB, Quigley JP, Nielsen KL, Dolmer K, Sottrup-Jensen L, Iwanaga S, 1996. Molecular cloning of *Limulus* alpha 2-macroglobulin. Eur J Biochem. 242, 822-831.

Iwanaga S, Isawa H, Yuda M., 2014. Horizontal gene transfer of a vertebrate vasodilatory hormone into ticks. Nat Commun. 5, 3373.

Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, Nilsson B, Gros P, 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature. 437, 505-511.

Janeway Jr. CA, Medzhitov R, 2002. Innate immune recognition. Annu Rev Immunol. 20, 197-216.

Kalia V, Sarkar S, Gourley TS, Rouse BT, Ahmed R, 2006. Differentiation of memory B and T cells. Curr Opin Immunol. 18, 255-264.

Kang D, Liu G, Lundström A, Gelius E, Steiner H, 1998. A peptidoglycan recognition protein in innate immunity conserved from insects to humans. Proc Natl Acad Sci U S A. 95, 10078-10082.

Kasahara M, Suzuki T, Pasquier LD, 2004. On the origins of the adaptive immune system: novel insights from invertebrates and cold-blooded vertebrates. Trends Immunol. 25, 105-111.

Kimura A, Sakaguchi E, Nonaka M, 2009. Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. Immunobiology. 214, 165-178.

King N, Westbrook MJ, Young SL et al., 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature. 451, 783-788.

Langmead B, Trapnell C, Pop M, Salzberg SL, 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009. 10, R25.

Le SQ, Lartillot N, Gascuel O, 2008. Phylogenetic mixture models for proteins. Philos Trans R Soc Lond B Biol Sci. 363, 3965-3976.

Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA, 1996. The dorsoventral regulatory gene cassette *spätzle*/*Toll*/*cactus* controls the potent antifungal response in *Drosophila* adults. Cell. 86, 973-983.

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC, 2001. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. Cell. 104, 709-718.

Li ZF, Wu XH, Engvall E, 2004. Identification and characterization of CPAMD8, a novel member of the complement 3/alpha2-macroglobulin family with a C-terminal Kazal domain. Genomics. 83, 1083-1093.

Lin M, Sutherland DR, Horsfall W, Totty N, Yeo E, Nayar R, W, XF, Schuh AC, 2002. Cell surface antigen CD109 is a novel member of the alpha(2) macroglobulin/C3, C4, C5 family of thioester-containing proteins. Blood. 99, 1683-1691.

Medzhitov R, 2007. Recognition of microorganisms and activation of the immune response. Nature. 449, 819-826.

Medzhitov R, Preston-Hurlburt P, Janeway CA Jr., 1997. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature. 388, 394-397.

Meng F, Sun Y, Liu X, Wang J, Xu T, Wang R, 2012. Analysis of C3 suggests three periods of positive selection events and different evolutionary patterns between fish and mammals. PLoS One. 5, e37489.

Nakao M, Mutsuro J, Obo R, Fujiki K, Nonaka M, Yano T, 2000. Molecular cloning and protein analysis of divergent forms of the complement component C3 from a bony fish, the common carp (*Cyprinus carpio*): presence of variants lacking the catalytic histidine. Eur J Immunol. 3, 858-866.

Nonaka M, 2014. Evolution of the complement system. Subcell Biochem. 80, 31-43.

Nonaka M, Kimura A, 2006. Genomic view of the evolution of the complement system. Immunogenetics. 58, 701-713.

Ødegaar F, 2000. How many species of arthropods? Erwin's estimate revised. Biological Journal of the Linnean Society. 71, 583-597.

Prado-Alvarez M, Rotllant J, Gestal C, Novoa B, Figueras A, 2009. Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. Fish Shellfish Immunol. 26, 305-315.

Putnam NH, Srivastava M, Hellsten U et al., 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 317, 86-94.

Putnam NH, Butts T, Ferrier DE et al., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature. 453, 1064-1071.

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW, 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature. 463, 1079-1083.

Roberts A, Pachter L, 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 10, 71-73.

Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4, 406-425.

Sekiguchi R, Fujito NT, Nonaka M, 2012. Evolution of the thioester-containing proteins (TEPs) of the arthropoda, revealed by molecular cloning of TEP genes from a spider, *Hasarius adansoni*. Dev. Comp. Immunol. 36, 483-489.

Sottrup-Jensen L, Folkersen J, Kristensen T, Tack BF, 1984. Partial primary structure of human pregnancy zone protein: extensive sequence homology with human alpha 2-macroglobulin. Proc Natl Acad Sci U S A .81, 7353-7357.

Sottrup-Jensen L, Sand O, Kristensen L, Fey GH, 1989. The alpha-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian alpha-macroglobulins. J Biol Chem. 264, 15781-15789.

Srivastava M, Simakov O, Chapman J et al., 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature. 466, 720-726.

Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD, 2006. Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. PLoS Biol.4, e4.

Tagawa K, Yoshihara T, Shibata T, Kitazaki K, Endo Y, Fujita T, Koshiba T, Kawabata S, 2012. Microbe-specific C3b deposition in the horseshoe crab complement system in a C2/factor B-dependent or -independent manner. PLoS ONE. 7, e36783.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 30, 2725-2729.

Tauszig S, Jouanguy E, Hoffmann JA, Imler JL., 2000. Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. Proc. Natl. Acad. Sci. USA. 97, 10520-10525.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876-4882.

Tonegawa S, 1983. Somatic generation of antibody diversity. Nature. 302, 575-581.

Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. Nature. 452, 949-955.

Udompetcharaporn A, Junkunlo K, Senapin S, Roytrakul S, Flegel TW, Sritunyalucksana K, 2014. Identification and characterization of a QM protein as a possible peptidoglycan recognition protein (PGRP) from the giant tiger shrimp *Penaeus monodon*. Dev Comp Immunol. 46, 146-154.

Urbanová V, Hartmann D, Grunclová L, Šíma R, Flemming T, Hajdušek O, Kopáček P, 2014. IrFC - An *Ixodes ricinus* injury-responsive molecule related to Limulus Factor C.Dev Comp Immunol. 46, 439-447.

Urbanová V, Síma R, Sauman I, Hajdušek O, Kopáček P, 2015. Thioester-containing proteins of the tick *Ixodes ricinus*: Gene expression, response to microbial challenge and their role in phagocytosis of the yeast *Candida albicans*. Dev Comp Immunol. 48, 55-64.

Volanakis JE, 1998. Overview of the complement system, in: Volanakis, J.E., Frank, M.M. (Eds.), The human complement system in health and disease. Marcel Dekker, New York, pp. 9–32.

Volanakis JE, Narayana SV, 1996. Protein Sci. 1996. Complement factor D, a novel serine protease. 5, 553-564. Review.

Werren JH, Richards S, Desjardins CA et al., 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science. 327, 343-348.

Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D., 2000. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. Proc Natl Acad Sci USA. 97, 13772-13777

Wu C, Noonin C, Jiravanichpaisal P, Söderhäll I, Söderhäll K, 2012. An insect TEP in a crustacean is specific for cuticular tissues and involved in intestinal defense. Insect Biochem Mol Biol. 42, 71-80.

Yoshida H, Kinoshita K, Ashida M, 1996. Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. J Biol Chem. 271, 13854-13860.

Zhu Y, Thangamani S, Ho B, Ding JL, 2005. The ancient origin of the complement system. Embo J. 24, 382-394.

Figure legends

Fig. 1. Phylogenetic relationship of metazoan animals and the distribution of the complement and TEP genes.

Phylogenetic relationship of metazoan animals is shown with choanoflagellate as an out group based on King et al. (2008). The distribution of the complement (MASP, FB, and C3) and TEP (C3, A2M, and iTEP/CD109) genes are shown in the right side. The + symbol indicates the presence of the gene, and the – symbol indicates failure of detection of the gene by the comprehensive methods. The ? symbol indicates that the presence or absence of the gene is still to be clarified. Species in which whole genomic analysis has been performed are shown by red underline.

Fig. 2. Phylogenetic relationship of the complement and TEP genes among arthropods

This phylum Arthoropoda is divided into three groups: the Chelicerata (sea spider, horseshoe crab, and spider); the Myriapoda (centipede and millipede); and the Pancrustacea comprising the crustacean (sea firefly, sea louse, water flea and crayfish) and the Hexapoda (aphid, beetle, honey bee, and fruit fly). The phylogenetic relationship of arthropods was based on Regier et al. (2010). The species analyzed in this study are shown in red bold letters. The + symbol indicates the presence of the gene, and the – symbol indicates failure of detection of the gene by the comprehensive methods. The ? symbol indicates that the presence or absence of the gene is still to be clarified. Species in which whole genomic analysis has been performed are shown by red underline.

Fig. 3. Genomic sequence of PsfuC3 aligned with vertebrate C3

The nucleotide sequence of the PsfuC3 gene was aligned together with that of *Oryzias latipes* (Orla) C3 (NC_019859), *Takifugu rubripes* (Taru) C3 (NC_018906), and *Homo sapiens* (Hosa) C3 (NC_000019). Exons were indicated by open boxes and sequences coding for the thioester site (GCGEQ) was indicated by dotted open box.

Fig. 4. Phylogenetic tree of TEP family

The phylogenetic tree was constructed using the amino acid sequences of the TEP family using the maximum likelihood method based on the LG model. Gaps in the alignment were excluded and there were 704 positions in the final dataset. Bootstrap percentages of more than 50 with 500 replicates are given. Bootstrap values of the C3, A2M, and iTEP/CD109 clades are in bold. Accession numbers of the amino acid sequences and the scientific names of animals used in this tree are; *Haliplanella lineata* (Hali) C3, A2M, and CD109 (BAJ05269, BAJ05271, and BAJ05272), *Euphaedusa tau* (Euta) CD109 (BAE44110), *Ammothea* sp. (Amsp) C3, A2M-1, and A2M-2, (LC009021, LC009022, and LC009023), *Tachypleus tridentatus* (Tatr) C3 and A2M (BAJ02276 and BAA19844), *Ixodes scapularis* (Ixsc) C3, A2M, and iTEP/CD109 (ISCW022535, ISCW023777, and XP_002409560), *Hasarius adansoni* (Haad) C3-1, C3-2, A2M, and iTEP/CD109 (BAK64109, BAK64110, BAK64111 and BAK64112), *Scolopendra subspinipes* (Scsu) C3, A2M, and iTEP/CD109-2 (LC009029, LC009032 and LC009034), *Niponia nodulosa* (Nino) A2M, iTEP/CD109-1, iTEP/CD109-2, and iTEP/CD109-3 (LC009014, LC009015, LC009016, and LC009017), *Paracyclopina nana* (Pana) A2M (AII16542), *Pseudocaligus fugu* (Psfu) C3, A2M-1, A2M-2, and A2M-3; *Caligus* sp. (Casp) A2M, Pandaridae gen. et sp. (Pasp) C3 and A2M, *Daphnia* *pulex* (Dapu) iTEP/CD109 (EFX86067), *Pacifastacus leniusculus* (Pale) A2M and iTEP/CD109 (AEC50080 and AEC50085), *Acyrthosiphon pisum* (Acpi) iTEP/CD109 (XP_001944348), *Tribolium castaneum* (Trca) A2M and iTEP/CD109 (EFA07508 and XP_972838), *Apis mellifera* (Apme) A2M and iTEP/CD109 (XP_392454 and XP_001122599), *Drosophila melanogaster* (Drme) iTEP/CD109 (NP_523578), *Anopheles gambiae* (Anga) iTEP/CD109 (XP_555086), *Strongylocentrotus purpuratus* (Stpu) C3, A2M, and CPAMD8 (NP_999868, XP_799248, and XP_785018), *Branchiostoma belcheri* (Brbe) C3 and CD109 (BAB47146 and AEG67300), *Branchiostoma floridae* (Brfl) CPAMD8 (XP_002586872), *Ciona intestinalis* (Ciin) C3, CPAMD8, and CD109 (NP_001027684, XP_002124325, and NP_001027688), *Lethenteron japonicum* (Leja) C3 and A2M (Q00685 and BAA02762), *Oryzias latipes* (Orla) C3 (NP_001098552), and *Homo sapiens* (Hosa) C3, C4, C5, A2M, PZP, CPAMD8, and CD109 (AAA85332, P0C0L4, AAA51925, P01023, CAA38255, NP_056507, and NP_598000). Names for the amino acid sequences of *Ammothea* sp., *S. subspinipes*, *P. fugu, Caligus* sp., and Pandaridae gen. et sp. TEPs are shown in bold. *H. adansoni* and *N. nodulosa* TEPs are indicated by underline.

Fig. 5. Phylogenetic tree of TEP family excluding PsfuA2M-1 and PsfuA2M-2

The phylogenetic tree was constructed using the amino acid sequences of the TEP family excluding PsfuA2M-1 and PsfuA2M-2 using the maximum likelihood method based on the LG model. Gaps in the alignment were excluded and there were 736 positions in the final dataset. Bootstrap percentages of more than 50 with 500 replicates are given. Bootstrap values of the C3, A2M, and iTEP/CD109 clades are in bold. Accession numbers of the amino acid sequences and the scientific names of animals used in the tree are described in the legend for Fig. 4.

Fig. 6. Phylogenetic tree using sea lice C3s and various fish C3s

The phylogenetic tree was constructed using the amino acid sequences of the TEP family using the neighbor joining method based on p-distance. Gaps in the alignment were excluded and there were 859 positions in the final dataset. Bootstrap percentages of more than 50 with 1000 replicates are given. Accession numbers of the amino acid sequences and the scientific names of animals used in the tree are; *Ammothea* sp. (Amsp) C3 (LC009021), *Hasarius adansoni* (Haad) C3 (BAK64109), *Scolopendra subspinipes* (Scsu) C3 (LC009029), *Pseudocaligus fugu* (Psfu) C3, Pandaridae gen. et sp. (Pasp) C3, *Triakis scyllium* (Trsc) C3A and C3B, *Miichthys miiuy* (Mimi) C3 (AFC89899), *Epinephelus coioides* (Epco) C3 (ADU33222), *Anarhichas minor* (Anmi) C3 (CAC29154), *Paralichthys olivaceus* (Paol) C3 (BAA88901), *Sparus aurata* (Spau) C3 (ADM13620), *Dicentrarchus labrax* (Dila) C3 (AEJ37034), *Xiphophorus hellerii* (Xihe) C3 (AEJ08067), *Tetraodon nigroviridis* (Teni) C3 (CAG06096), *Takifugu rubripes* (Taru) C3-1, C3-2, and C3-3 (XP_003978399, XP_003972136, and XP_003971979), *Oreochromis niloticus* (Orni) C3 (XP_005454885), *Oryzias latipes* (Orla) C3 (NP_001098552), *Cyprinus carpio* (Cyca) C3-H1, C3-H2, and C3-S (BAA36619, BAA36620, and BAA36621), *Danio rerio* (Dare) C3 (XP_002660623), and *Homo sapiens* (Hosa) C3 (AAA85332). Names for the amino acid sequences of *Pseudocaligus fugu* C3 and Pandaridae gen. et sp. C3 are indicated by red open box.

Fig. 7. Evolution of the TEP genes in the Arthropoda

The species analyzed in this study are shown in red bold letters. The + symbol indicates the presence of the gene, and the – symbol indicates failure of detection of the gene by the RT-PCR/cloning method. The ? symbol indicates that the presence or absence of the gene is still to be clarified. Triangles indicate evolutionary losses of the C3 gene. The * symbol upper the + of *P. fugu* C3 indicates that C3 gene was obtained by horizontal transfer independent from the evolution of the complement gene in the Arthropoda. Species in which whole genomic analysis has been performed are shown by red underline. The phylogenetic relationship of arthropods was based on Regier et al. (2010).

Fig. 8. Phylogenetic tree of TEP family including eight arthropod species

The phylogenetic tree was constructed using the amino acid sequences of the TEP family with the addition of the genes newly identified by the RNA-seq method, using the maximum likelihood method based on the LG model. Gaps in the alignment were excluded and there were 714 positions in the final dataset. Bootstrap percentages of more than 50 with 500 replicates are given. Bootstrap values of the C3, A2M, and iTEP/CD109 clades are shown in bold. Accession numbers of the amino acid sequences and the scientific names of animals used in this tree are; *Ammothea* sp. (Amsp) A2M-3, A2M-4, and iTEP/CD109 (LC009024, LC009025, and LC009026), *Hasarius adansoni* (Haad) A2M-2 and iTEP/CD109-2 (LC009009 and LC009010), *Scolopendra subspinipes* (Scsu) C3-2, C3-3, iTEP/CD109-1, iTEP/CD109-3, and iTEP/CD109-4 (LC009030 LC009031, LC009033, LC009035, and LC009036), *Epanerchodus* sp. (Epsp) A2M, iTEP/CD109-1, and iTEP/CD109-2 (LC009018, LC009019, and

LC009020), *Vargula* sp. (Vasp) A2M-3 (LC009041), *Pseudocaligus fugu* (Psfu) iTEP/CD109, and *Caligus* sp. (Casp) A2M-2, iTEP/CD109-1, and iTEP/CD109-2. Others are described in the legend for Fig. 5. *Ammothea* sp., *H. adansoni*, *S. subspinipes*, *N. nodulosa, Epanerchodus* sp., *Vargula* sp., *P. fugu, Caligus* sp., and Pandaridae gen. et sp. TEPs are shown in bold.

Fig. 9. Phylogenetic tree of TEP family with arthropod Mcrs

The phylogenetic tree was constructed based on the amino acid sequences of the TEP family with the addition of the Mcrs, using the Neighbor-joining method by p-distance. All sites were used in the final dataset. Bootstrap percentages of more than 50 with 1000 replicates are given. Bootstrap values of the C3, A2M, iTEP/CD109, and Mcr clades are in bold. Accession number of the amino acid sequences and scientific names of animals used in this tree are; *Ammothea* sp. (Amsp) Mcr-1, and Mcr-2, *Ixodes scapularis* (Ixsc) Mcr (XP_002410473), *Hasarius adansoni* (Haad) Mcr, *Scolopendra subspinipes* (Scsu) Mcr-1, and Mcr-2, *Niponia nodulosa* (Nino) Mcr-1, and Mcr-2, *Epanerchodus* sp. (Epsp) Mcr-1 and Mcr-2, *Vargula* sp. (Vasp) Mcr, *Pseudocaligus fugu* (Psfu) Mcr, and *Caligus* sp. (Casp) Mcr. Others are described in the legend for Fig. 5 and 8. *Ammothea* sp., *H. adansoni*, *S. subspinipes*, *N. nodulosa, Epanerchodus sp.*, *Vargula* sp., *P. fugu, Caligus* sp., and Pandaridae gen. et sp. TEPs are shown in bold.

Fig. 10. Phylogenetic tree of FB/C2

The phylogenetic tree was constructed based on the amino acid sequences of FB and C2 using the NJ method according to p-distances. Data with fewer than 5% alignment gaps were excluded and there were 510 positions in the final dataset.

Numbers indicate the percentage bootstrap statistical support of the corresponding node based on 1000 bootstrap replicates. Accession numbers of the amino acid sequences and the scientific names of animals used in this tree are; *Nematostella vectensis* (Neve) FB (BAH22728), *Ammothea* sp. (Amsp) FB-1 and FB-2 (LC009027 and LC009028), *Tachypleus tridentatus* (Tatr) FB-1 and FB-2 (BAM15262 and BAM15263), *Hasarius adansoni* (Haad) FB-1, FB-2, and FB-3 (LC009011, LC009012, and LC009013), *Scolopendra subspinipes* (Scsu) FB-1 and FB-2 (LC009037 and LC009038), *Ciona intestinalis* (Ciin) FB (NP_001027973), *Lethenteron japonicum* (Leja) FB (BAA02763), *Ginglymostoma cirratum* (Gici) fB (AAY55950), *Cyprinus carpio* (Cyca) FB (BAA34706), *Mus musculus* (Mumu) C2 (NP_038512), and *Homo sapiens* (Hosa) FB and C2 (AAH04143 and NP_000054). The *Ammothea* sp., *H. adansoni*, and *S. subspinipes* FBs that were newly identified in this study are shown in bold.

Fig. 11. Evolution of the complement, TEP and factor C (FC)/FC-like genes in the Arthropoda.

The species analyzed in this study are shown in red bold. The + symbol indicates the presence of the gene, and the – symbol indicates failure of detection of the gene by the comprehensive methods. The ? symbol indicates that the presence or absence of the gene is still to be clarified. Triangles indicate evolutionary losses of the C3 and FB genes. The * symbol upper the + of *P. fugu* C3 indicates that C3 gene was obtained by horizontal gene transfer independent from the evolution of the complement gene in Arthropoda. Species in which whole genomic analysis has been performed are shown by red underline. The phylogenetic relationship of arthropods was based on Regier et al. (2010).

Fig. 12. Domain structure of the FC-like proteins of the Arthropoda.

The FC-like proteins of *Ammothea* sp. (Amsp), *Hasarius adansoni* (Haad), *Scolopendra subspinipes* (Scsu), *Niponia nodulosa* (Nino) and *Epanerchodus* sp. (Epsp) are shown with *Tachypleus tridentatus* (Tatr) FC. Abbreviations of domain names are: EGF, epidermal growth factor domain; SCR, short consensus repeats; LCCL, Limulus factor C, Coch-5b2 and Lgl1; CLECT, c-lectin domain, and SP, serine protease domain.

Fig. 13. Evolution of the complement system in metazoan animals

The evolutionary process of the complement system is schematically shown. Vertical black and gray arrows indicate the appearance of the complement genes and innovation of additional components, respectively. Triangles indicate evolutionary losses of the complement genes.

Fig. 14. Phylogenetic tree of the TLR genes

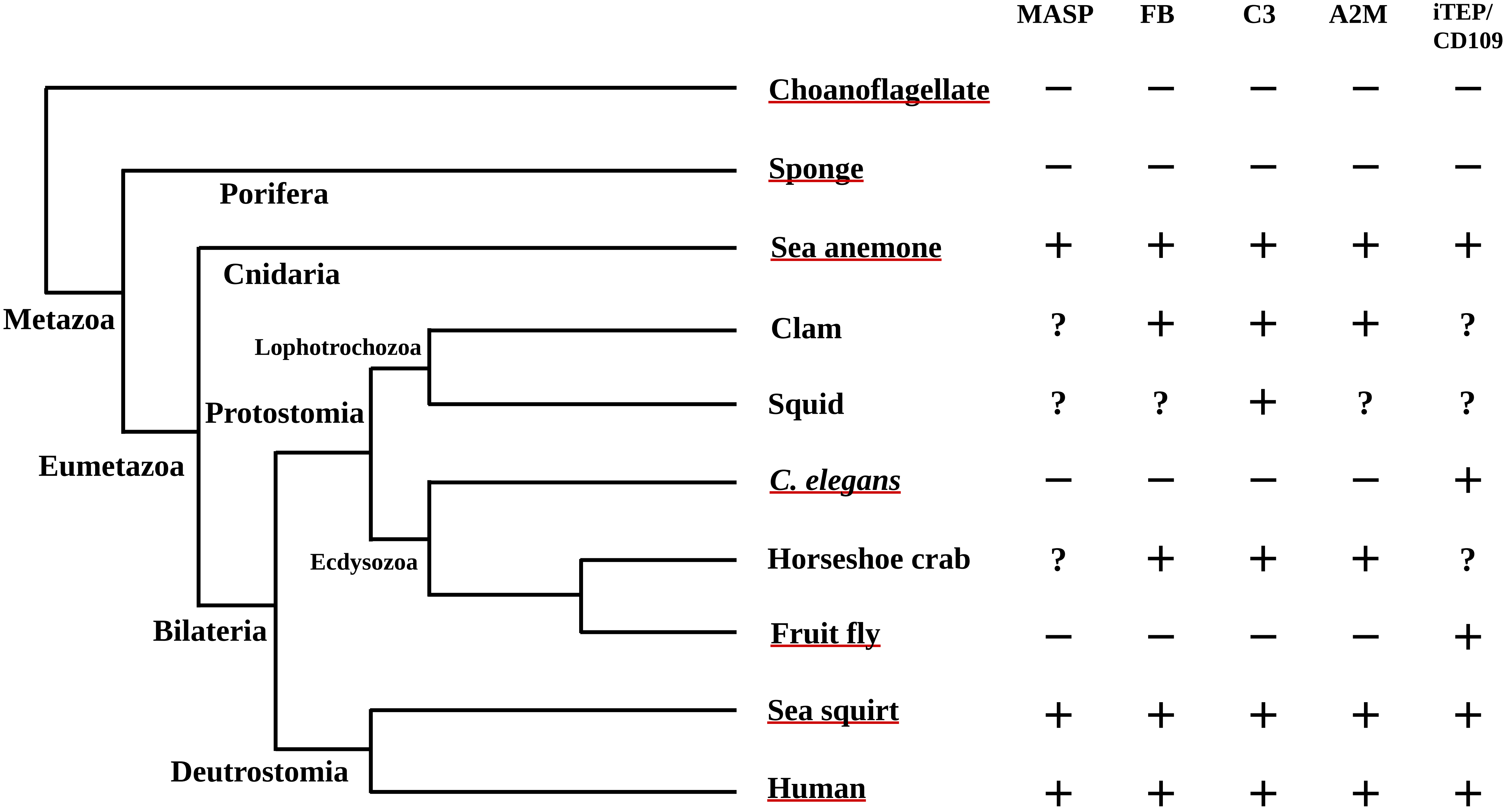
The phylogenetic tree of TLR was constructed based on the amino acid sequences of TLR using the NJ method according to p-distances. All sites were used in the final dataset. Bootstrap percentages of more than 50 with 1000 replicates are given. Accession numbers of the amino acid sequences and the scientific names of animals used in this tree are; *Nematostella vectensis* (Neve) TLR (XP_001641546), *Ammothea* sp. (Amsp) TLR-1, TLR-2, TLR-3, TLR-4, TLR-5, and TLR-6, *Hasarius adansoni* (Haad)TLR-1, TLR-2, TLR-3, TLR-4, and TLR-5, *Carcinoscorpius rotundicauda* (Caro) TLR (ABK88278), *Ixodes scapularis* (Ixsc) TLR (XP_002399580), *Scolopendra subspinipes* (Scsu) TLR, *Niponia nodulosa* (Nino) TLR-1, TLR-2, TLR-3, TLR-4, TLR-5, and TLR-6, *Epanerchodus* sp. (Epsp) TLR-1, TLR-2, TLR-3, and TLR-4, *Vargula* sp. (Vasp) TLR-1 and TLR-2, *Pseudocaligus fugu* (Psfu) TLR-1, TLR-2, and TLR-3, *Caligus* sp. (Casp) TLR-1, TLR-2, and TLR-3, *Daphnia pulex* (Dapu) TLR (EFX86424), *Fenneropenaeus chinensis* (Fech) TLR (ACC68670), *Carcinus maenas* (Cama) TLR (CDO91661), *Apis merifera* (Apme) TLR (NP_001013379), *Drosophila melanogaster* (Drme) Toll-1, Toll-2, Toll-3, Toll-4, Toll-5, Toll-6, Toll-7, and Toll-8 (NP_001262995, NP_476814, NP_649719, NP_523519, NP_477438, NP_524081, NP_523797, and NP_524757), Homo sapiens (Hosa) TLR-1, TLR-2, TLR-3, TLR-4, TLR-5, TLR-6, TLR-7, TLR-8, and TLR-9 (NP_003254, NP_003255, NP_003256, NP_003257, NP_003259, NP_006059, NP_057646, NP_619542, and NP_059138). *Ammothea* sp., *H. adansoni*, *S. subspinipes*, *N. nodulosa*, *Epanerchodus* sp., *Vargula* sp., *P. fugu*, and *Caligus* sp. TLRs were indicated by bold.

Fig. 15. Phylogenetic tree of the PGRP genes

The phylogenetic tree was constructed based on the amino acid sequences of PGRP using the NJ method according to p-distances. All sites were used in the final dataset. Bootstrap percentages of more than 50 with 1000 replicates are given. Accession numbers of the amino acid sequences and the scientific names of animals used in this tree are; *Hasarius adansoni* (Haad) PGRP-1, PGRP-2, PGRP-3, PGRP-4, PGRP-5, and PGRP-6, *Ixodes scapularis* (Ixsc) PGRP (XP_002433689), *Scolopendra subspinipes* (Scsu) PGRP-1 and PGRP-2, *Niponia nodulosa* (Nino) PGRP-1 and PGRP -2, *Epanerchodus* sp. (Epsp) PGRP-1 and PGRP-2, *Drosophila melanogaster* (Drme) PGRP-SA, PGRP-SB1, PGRP-SB2, PGRP-SC1a, PGRP-SC1b, PGRP-SC2, PGRP-SD, PGRP-LA PGRP-LB, PGRP-LC, PGRP-LD, PGRP-LE, and PGRP-LF (NP_572727, NP_648917, NP_648916, NP_610407, CAD89168, NP_610410, NP_648145,

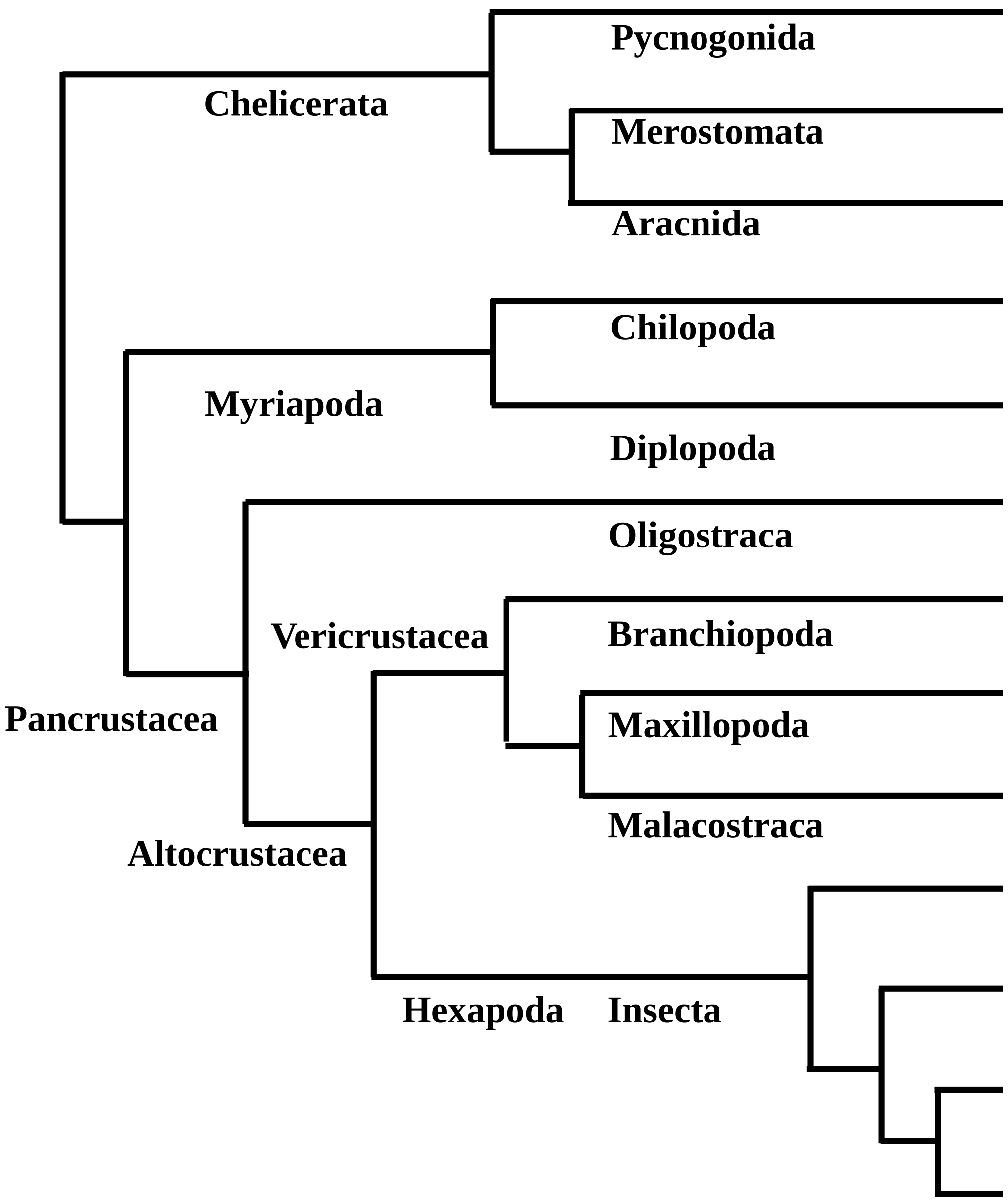
NP_996026, NP_650079, NP_729468, NP_001027111, NP_573078, and NP_648299), *Anopheles gambie* (Anga) PGRP-S1 PGRP-S2, PGRP-LA, PGRP-LB, PGRP-LC, and PGRP-LD (XP_310547, ADA54999, XP_001688528, XP_003435776, XP_558599, and XP_556195). *Hasarius adansoni*, *Scolopendra subspinipes*, *Niponia nodulosa*, *Epanerchodus* sp. PGRP were indicated in bold.

Figures and Tables



	MASP	FB	C3	A2M	iTEP/ CD109
Choanoflagellate	−	−	−	−	−
Sponge	−	−	−	−	−
Sea anemone	+	+	+	+	+
Clam	?	+	+	+	?
Squid	?	?	+	?	?
C. elegans	−	−	−	−	+
Horseshoe crab	?	+	+	+	?
Fruit fly	−	−	−	−	+
Sea squirt	+	+	+	+	+
Human	+	+	+	+	+

Fig. 1



	MASP	FB	C3	A2M	iTEP/ CD109
Sea spider	?	?	?	?	?
Horseshoe crab	?	+	+	+	?
Spider	?	?	+	+	+
Centipede	?	?	?	?	?
Millipede	?	?	−	+	+
Sea firefly	?	?	?	?	?
Sea lice	?	?	?	?	?
Water flea	−	−	−	+	+
Crayfish	?	?	−	+	+
Aphid	−	−	−	−	+
Beetle	−	−	−	+	+
Honey bee	−	−	−	+	+
Fruit fly	−	−	−	−	+

Fig. 2

```
                                                                                    Thioester site
                                                                                     G  C  G  E  Q
OrlaC3 : CGAGAACAAGTTTCTGGACTGGTGGAAAATGCCATCAGTGGAAAGTCAATGGGTACTCTGATTTATCAGCCCTCTGGTTGTGGAGAGCAGAACATGATCCACATGACCCTACCTGTCATTGCAACCATATATTTGGACAAAACCAACCAA :  150
TaruC3 : GTAGAACAAATATCTTCACTGGTGGAGAATGCCATTAGTGGGACATCCATGGGCAGTCTGATCTACCAGCCTTCTGGTTGTGGCGAGCAGAACATGATCCACACGACCTTGCCTGTCATCGCAACCACATACCTGGATAAAACGAACCAG :  150
PsfuC3 : -----------------------------------------------------------------------------------------------------------------------AGCGGCCACCTATTTGGACAAAACCAAGCAG :   31
HosaC3 : GGGACCCCAGTGGCCCAGATGACAGAGGATGCCGTCGACGCGGAACGGCTGAAGCACCTCATTGTGACCCCCTCGGGCTGCGGGGAACAGAACATGATCGGCATGACGCCCACGGTCATCGCTGTGCATTACCTGGATGAAACGGAGCAG :  150

OrlaC3 : TGGGAAGCCGTTGGCTTTCAGAAACGGGCTGAGGCCCTTCAGCACATAAAAACAGGTAAG----------ATGAAAG-----TTCGGGGTT-----GAAAAGCTGT-GGTTGTATG---CAAAGGGAAGTCATGGGTCATTCTCCAACAA :  276
TaruC3 : TGGGAGACTGTAGGGTTTCAGAAACGTGACGAAGCCCTCCAACACATAAAGACCGGTAAG----------CTCACCT-----TCAACGGGC-----CCAGACTGGTTAAACAAAGG---TTAGATCGAATCA--AATAAATGAAGGGTAA :  275
PsfuC3 : TGGGAGGCTGTCGGCTTCCAGAAGCGTAACGAAGCCATCAAACACATACAGACTGGTGGGTCCAAGAATCCCGACGC-----TCAGCAAGTTC--TCAAAAATCATGAGCAGGTCG---CAAAGAGAGGTCAAAAATCCGCAATTGATTC :  171
HosaC3 : TGGGAGAAGTTCGGCCTAGAGAAGCGGCAGGGGGCCTTGGAGCTCATCAAGAAGGGTGGGCTCCCTGCCCCTCTTGGAGACCCCAGGGACCCCTTTCCGAGCGCATCCCTCCCCTAAGATCCCACCTCATCTCAAGACCACGCCCTCCCC :  300

OrlaC3 : ACAACT--ACTTTTCTGTAC---------------------------------------------------------------------------------------------------------------------------------- :  294
TaruC3 : ATGTGT--TTTTTGTTGTCT---------------------------------------------------------------------------------------------------------------------------------- :  293
PsfuC3 : TCAAGTTGATTTTAATTTCC---------------------------------------------------------------------------------------------------------------------------------- :  191
HosaC3 : TGAGGCTCCACCTTCTCTCCTAGCCACTCCCCTCATTTGAGGCCCCACCTCTTCTCAAGGCTACGCCCTCTGAGGCCCTGACTCCTCCCAGGCCAGGCTTTTCATGAGACCCCGCCTCTCCTCAAGGCCATGCCCATCCCCTGAGGGCCC :  450

OrlaC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
TaruC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC3 : CCCACCTCTTCTCAAGGCCACGCCCTCTGAGGCCCTGACTCCTCCCAGGCCAGGCTCTTCATGAGACCCCGCCTCTCCTCAAGGCCATGCCCATCCCCTGAGGGCCCCCCACCTCTTCTCAAGGCCACGCCCTCTGAGGCCCTGACTCCT :  600

OrlaC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
TaruC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC3 : CCCAGGCCAGGCTCTTCATGAGACCCCGCCTCTCCTCAAGGCCATGCCCATCCCCTGAGGGCCTCCCACCTCTTCTCAAGGCCACGCCCTCTGAGGCCCTGACTCCTCCCAGGCCAGAATCTCGAGACCCTGCCTCTTTTCAAGGCCACG :  750

OrlaC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
TaruC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC3 : CCCATCCCCTGGGTCCCCACATCTTCTCAAGGCCACACCCTTCTGTGAGGCGCCACCTCCTGTCCCAGCCACTCTCATCTGAGGCCCCACGTCCTCTCCAGGCCATGCCTCTTCCCTGAGACTCCACCCCCTCTCTGAGAGCCCTCCCCT :  900

OrlaC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
TaruC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuC3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC3 : CCCTGAAAGCCCCCCACCCTCAATATCCTTCTCCTCTCTGAATCCCTTGTCCTCTTGAGAACTTTTCCACCTCCTCGTTCTGATCCCCCACCCTCTTTGAGTCCTTCCCTTTTTAAGGTCCCCTCCTCCCAGAACCCCTCCGCCACCCTG : 1050

OrlaC3 : ------------------------------------------------------------------------------------------------AGGCTACACCAACGAACTAGCCTACCTAAAAGGCGACGGCTCTTTTGCTGTATG :  348
TaruC3 : ------------------------------------------------------------------------------------------------AGGCTACAAGAACCAGCTTGCTTTCCGCAAAGCAGATGGATCTTTTAGTGTGTG :  347
PsfuC3 : ------------------------------------------------------------------------------------------------AGGTCTAAATAATCAGCAAGCCTACGCCAAAAAAGACGGATCATTTGCTGTATG :  245
HosaC3 : AGCCCCTGTCCCCTCTCTGCACCCCGCCCCTGCCCTTTCTGGCGTGCCCCCTCTGCTCAGCCCCGGCTCTTTTGGGGGTTCCTCTCTCTTCTCTGCAGGGTACACCCAGCAGCTGGCCTTCAGACAACCCAGCTCTGCCTTTGCGGCCTT : 1200

OrlaC3 : GGCTGATCATGGAAGCAGCTCCTGG :  373
TaruC3 : GGCCAATCGCAAAAGCAGCACCTGG :  372
PsfuC3 : GCCCAAATATCAAAG---------- :  260
HosaC3 : CGTGAAACGGGCACCCAGCACCTGG : 1225
```

Fig. 3

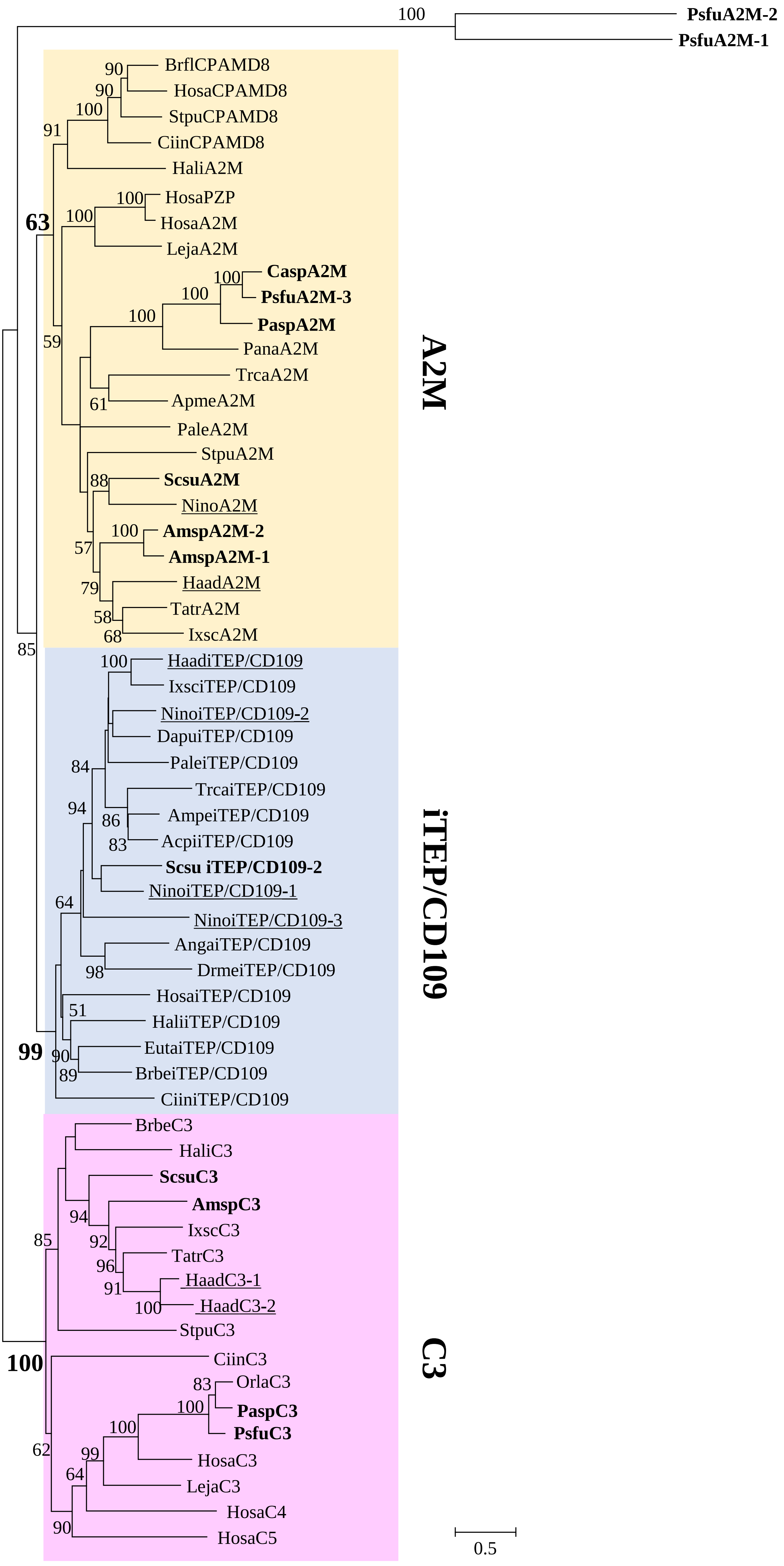

100
PsfuA2M-2
PsfuA2M-1
90
BrflCPAMD8
90
HosaCPAMD8
100
StpuCPAMD8
91
CiinCPAMD8
HaliA2M
100
HosaPZP
63
100
HosaA2M
LejaA2M
100
CaspA2M
100
PsfuA2M-3
100
PaspA2M
59
PanaA2M
TrcaA2M
61
ApmeA2M
PaleA2M
StpuA2M
88
ScsuA2M
NinoA2M
100
AmspA2M-2
57
AmspA2M-1
79
HaadA2M
58
TatrA2M
68
IxscA2M
A2M
85
100
HaadiTEP/CD109
IxsciTEP/CD109
NinoiTEP/CD109-2
DapuiTEP/CD109
84
PaleiTEP/CD109
TrcaiTEP/CD109
94
86
AmpeiTEP/CD109
83
AcpiiTEP/CD109
Scsu iTEP/CD109-2
NinoiTEP/CD109-1
64
NinoiTEP/CD109-3
AngaiTEP/CD109
98
DrmeiTEP/CD109
HosaiTEP/CD109
51
HaliiTEP/CD109
99
90
EutaiTEP/CD109
89
BrbeiTEP/CD109
CiiniTEP/CD109
iTEP/CD109
BrbeC3
HaliC3
ScsuC3
94
AmspC3
92
IxscC3
85
96
TatrC3
91
HaadC3-1
100
HaadC3-2
StpuC3
100
CiinC3
83
OrlaC3
100
PaspC3
100
PsfuC3
62
99
HosaC3
64
LejaC3
HosaC4
90
HosaC5
C3
0.5

Fig. 4

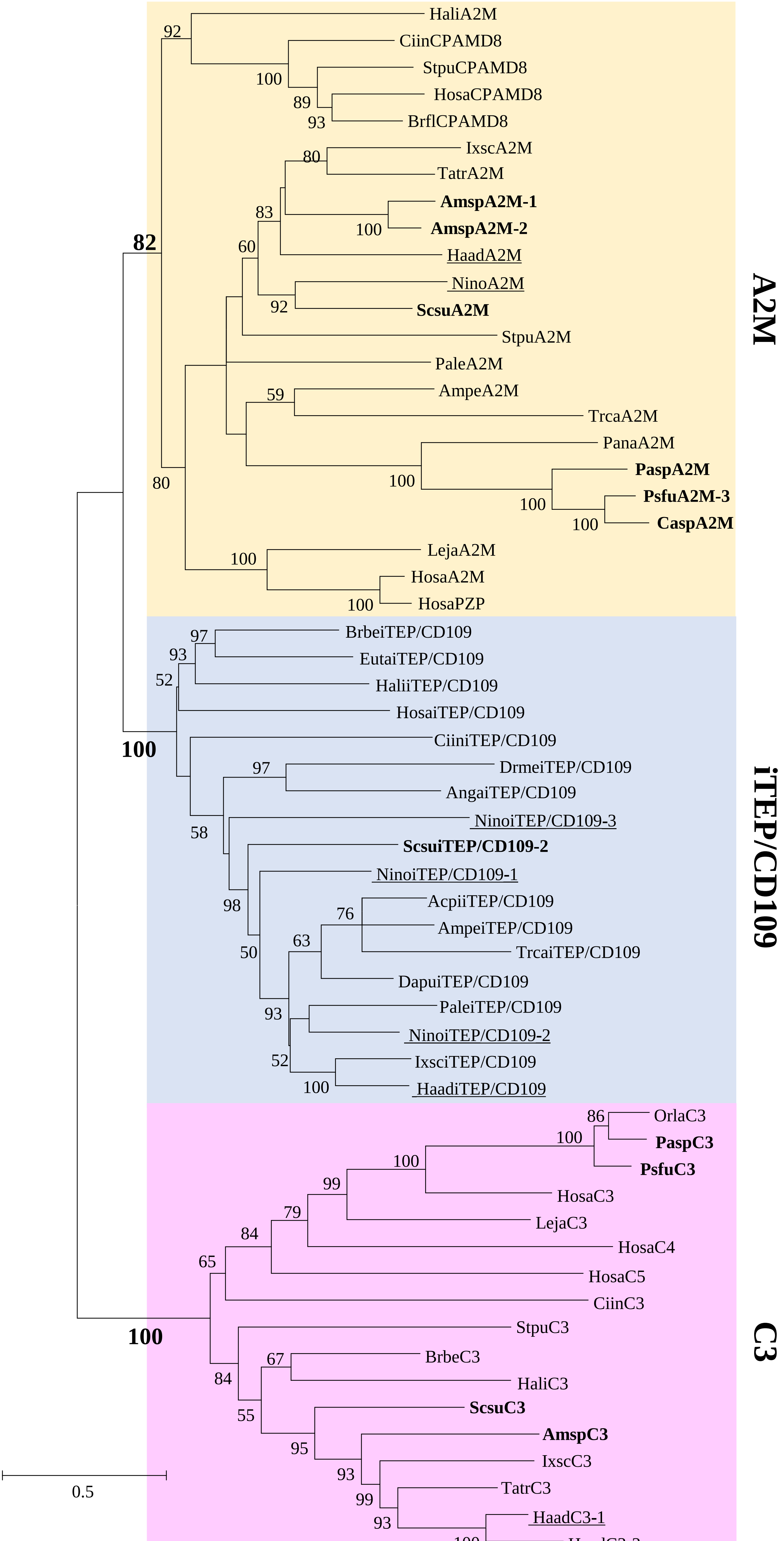

HaliA2M
92
CiinCPAMD8
StpuCPAMD8
100
HosaCPAMD8
89
BrflCPAMD8
93
IxscA2M
80
TatrA2M
AmspA2M-1
83
AmspA2M-2
100
82
HaadA2M
60
NinoA2M
ScsuA2M
92
StpuA2M
PaleA2M
AmpeA2M
59
TrcaA2M
PanaA2M
100
PaspA2M
80
PsfuA2M-3
100
CaspA2M
100
LejaA2M
100
HosaA2M
HosaPZP
100
A2M
BrbeiTEP/CD109
97
EutaiTEP/CD109
93
HaliiTEP/CD109
52
HosaiTEP/CD109
CiiniTEP/CD109
100
DrmeiTEP/CD109
97
AngaiTEP/CD109
NinoiTEP/CD109-3
58
ScsuiTEP/CD109-2
NinoiTEP/CD109-1
98
AcpiiTEP/CD109
76
AmpeiTEP/CD109
50
TrcaiTEP/CD109
63
DapuiTEP/CD109
PaleiTEP/CD109
93
NinoiTEP/CD109-2
52
IxsciTEP/CD109
100
HaadiTEP/CD109
iTEP/CD109
OrlaC3
86
PaspC3
100
PsfuC3
100
HosaC3
99
LejaC3
79
HosaC4
84
HosaC5
65
CiinC3
StpuC3
100
BrbeC3
67
84
HaliC3
ScsuC3
55
AmspC3
95
IxscC3
93
TatrC3
99
93
HaadC3-1
100
HaadC3-2
C3
0.5

Fig. 5

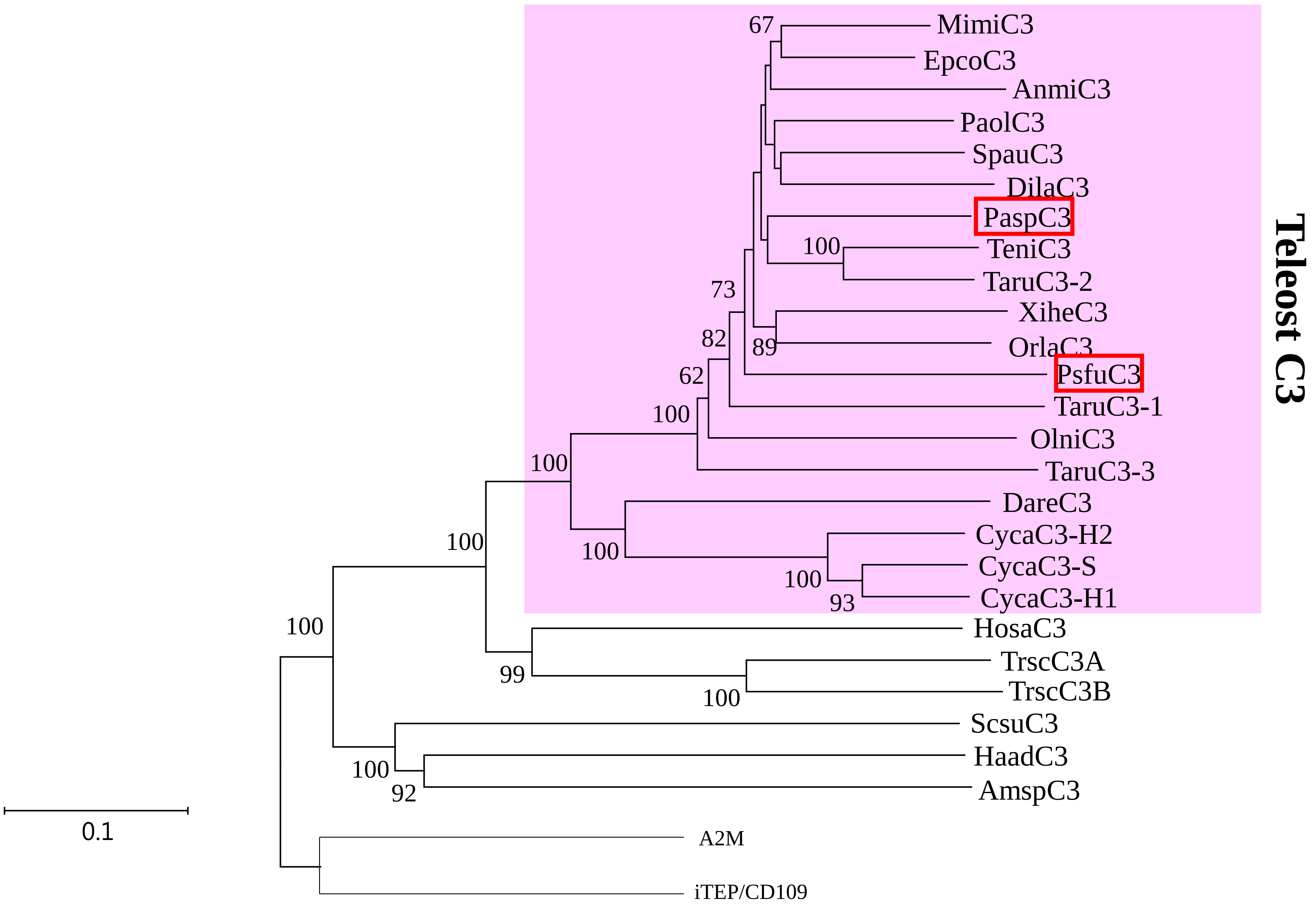
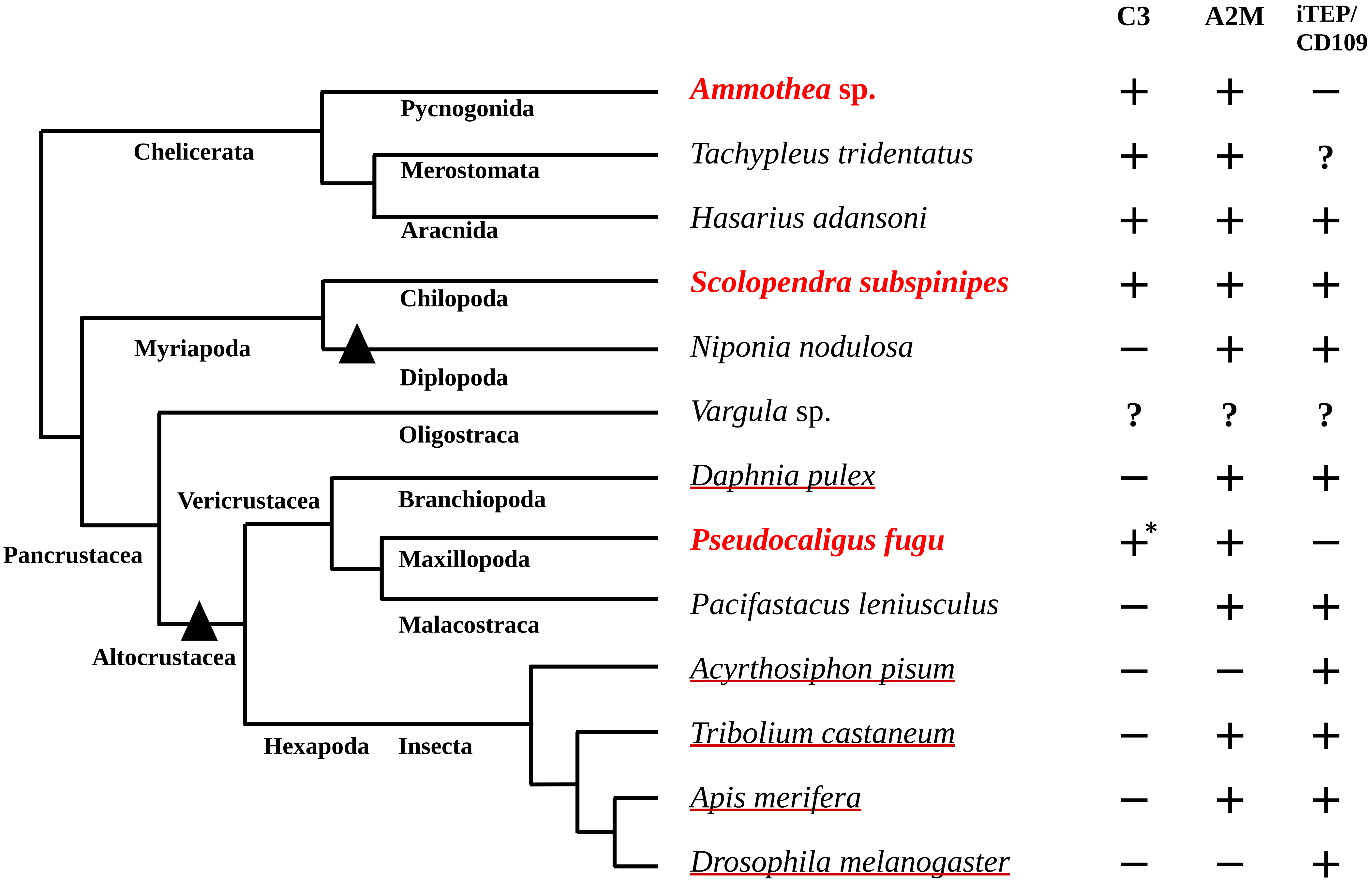

MimiC3
EpcoC3
AnmiC3
PaolC3
SpauC3
DilaC3
PaspC3
TeniC3
TaruC3-2
XiheC3
OrlaC3
PsfuC3
TaruC3-1
OlniC3
TaruC3-3
DareC3
CycaC3-H2
CycaC3-S
CycaC3-H1
HosaC3
TrscC3A
TrscC3B
ScsuC3
HaadC3
AmspC3
A2M
iTEP/CD109
Teleost C3
67
100
73
82
89
62
100
100
100
100
100
93
100
99
100
100
92
0.1

Fig. 6



	C3	A2M	iTEP/ CD109
***Ammothea* sp.**	+	+	−
Tachypleus tridentatus	+	+	?
Hasarius adansoni	+	+	+
Scolopendra subspinipes	+	+	+
Niponia nodulosa	−	+	+
Vargula sp.	?	?	?
Daphnia pulex	−	+	+
Pseudocaligus fugu	+*	+	−
Pacifastacus leniusculus	−	+	+
Acyrthosiphon pisum	−	−	+
Tribolium castaneum	−	+	+
Apis merifera	−	+	+
Drosophila melanogaster	−	−	+

Fig. 7

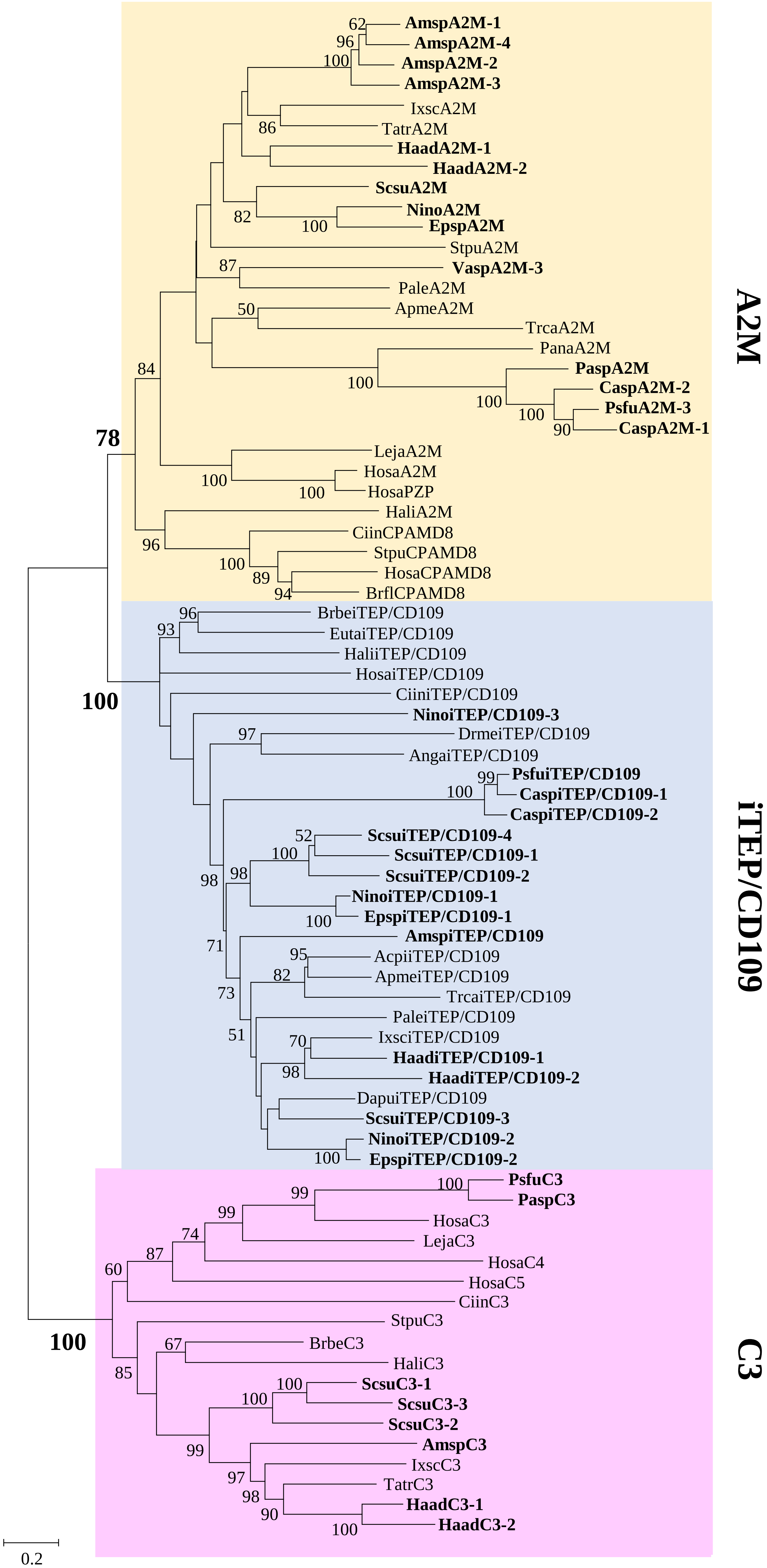

62
AmspA2M-1
96
AmspA2M-4
100
AmspA2M-2
AmspA2M-3
IxscA2M
86
TatrA2M
HaadA2M-1
HaadA2M-2
ScsuA2M
82
NinoA2M
100
EpspA2M
StpuA2M
87
VaspA2M-3
PaleA2M
50
ApmeA2M
TrcaA2M
84
PanaA2M
PaspA2M
100
CaspA2M-2
100
PsfuA2M-3
100
90
CaspA2M-1
78
LejaA2M
HosaA2M
100
100
HosaPZP
HaliA2M
CiinCPAMD8
96
StpuCPAMD8
100
HosaCPAMD8
89
94
BrflCPAMD8
A2M
96
BrbeiTEP/CD109
93
EutaiTEP/CD109
HaliiTEP/CD109
HosaiTEP/CD109
100
CiiniTEP/CD109
NinoiTEP/CD109-3
97
DrmeiTEP/CD109
AngaiTEP/CD109
99
PsfuiTEP/CD109
100
CaspiTEP/CD109-1
CaspiTEP/CD109-2
52
ScsuiTEP/CD109-4
100
ScsuiTEP/CD109-1
98
98
ScsuiTEP/CD109-2
NinoiTEP/CD109-1
100
EpspiTEP/CD109-1
AmspiTEP/CD109
71
95
AcpiiTEP/CD109
82
ApmeiTEP/CD109
73
TrcaiTEP/CD109
PaleiTEP/CD109
51
70
IxsciTEP/CD109
HaadiTEP/CD109-1
98
HaadiTEP/CD109-2
DapuiTEP/CD109
ScsuiTEP/CD109-3
NinoiTEP/CD109-2
100
EpspiTEP/CD109-2
iTEP/CD109
100
PsfuC3
99
PaspC3
99
HosaC3
74
LejaC3
87
HosaC4
60
HosaC5
CiinC3
StpuC3
100
67
BrbeC3
85
HaliC3
100
ScsuC3-1
100
ScsuC3-3
ScsuC3-2
99
AmspC3
IxscC3
97
TatrC3
98
HaadC3-1
90
100
HaadC3-2
C3
0.2

Fig. 8

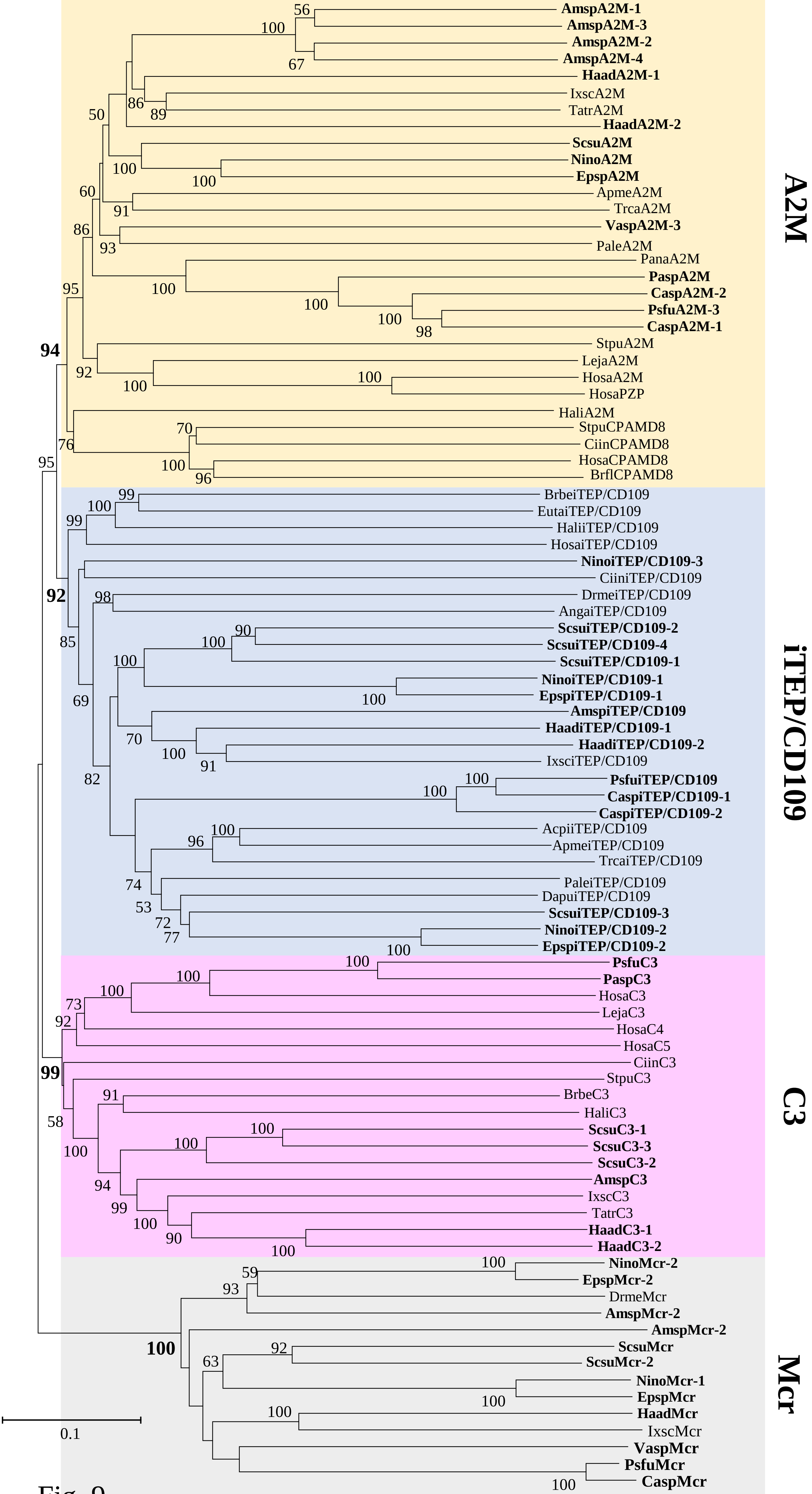

56
100
AmspA2M-1
AmspA2M-3
AmspA2M-2
67
AmspA2M-4
HaadA2M-1
IxscA2M
86
TatrA2M
89
50
HaadA2M-2
ScsuA2M
NinoA2M
100
EpspA2M
100
60
ApmeA2M
91
TrcaA2M
86
VaspA2M-3
93
PaleA2M
PanaA2M
95
PaspA2M
100
CaspA2M-2
100
PsfuA2M-3
100
CaspA2M-1
98
94
StpuA2M
LejaA2M
92
100
HosaA2M
100
HosaPZP
HaliA2M
70
StpuCPAMD8
76
CiinCPAMD8
95
HosaCPAMD8
100
BrflCPAMD8
96
A2M
99
BrbeiTEP/CD109
100
EutaiTEP/CD109
99
HaliiTEP/CD109
HosaiTEP/CD109
NinoiTEP/CD109-3
CiiniTEP/CD109
92
98
DrmeiTEP/CD109
AngaiTEP/CD109
90
ScsuiTEP/CD109-2
100
ScsuiTEP/CD109-4
85
100
ScsuiTEP/CD109-1
NinoiTEP/CD109-1
69
100
EpspiTEP/CD109-1
AmspiTEP/CD109
HaadiTEP/CD109-1
70
HaadiTEP/CD109-2
100
91
IxsciTEP/CD109
82
100
PsfuiTEP/CD109
100
CaspiTEP/CD109-1
CaspiTEP/CD109-2
100
AcpiiTEP/CD109
96
ApmeiTEP/CD109
TrcaiTEP/CD109
74
PaleiTEP/CD109
DapuiTEP/CD109
53
ScsuiTEP/CD109-3
72
NinoiTEP/CD109-2
77
EpspiTEP/CD109-2
100
iTEP/CD109
100
PsfuC3
100
PaspC3
100
HosaC3
73
LejaC3
92
HosaC4
HosaC5
99
CiinC3
StpuC3
91
BrbeC3
58
HaliC3
100
ScsuC3-1
100
ScsuC3-3
100
ScsuC3-2
94
AmspC3
99
IxscC3
100
TatrC3
90
HaadC3-1
HaadC3-2
100
C3
100
NinoMcr-2
59
EpspMcr-2
93
DrmeMcr
AmspMcr-2
AmspMcr-2
100
92
ScsuMcr
63
ScsuMcr-2
NinoMcr-1
100
EpspMcr
100
HaadMcr
IxscMcr
VaspMcr
PsfuMcr
100
CaspMcr
Mcr
0.1

Fig. 9

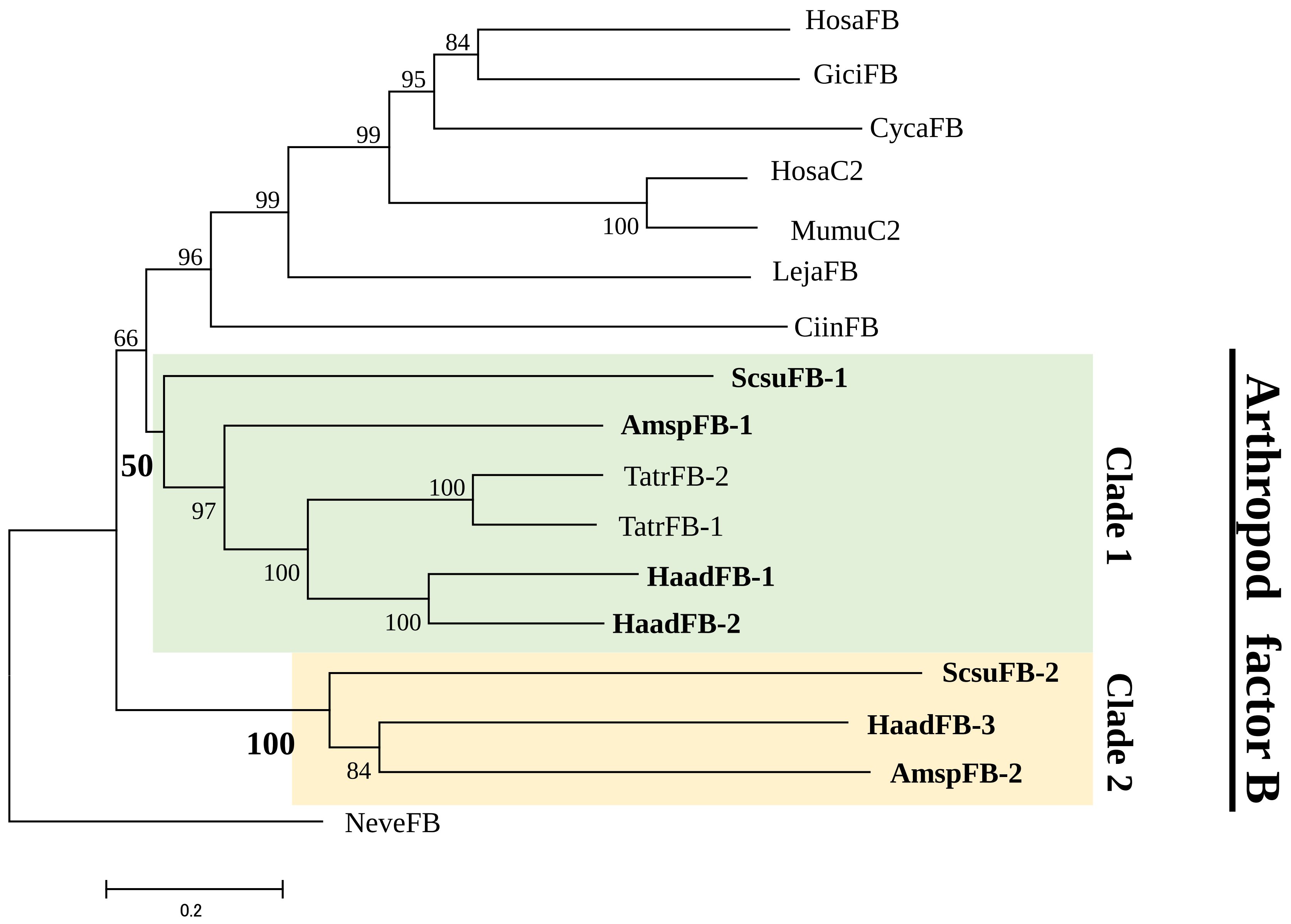
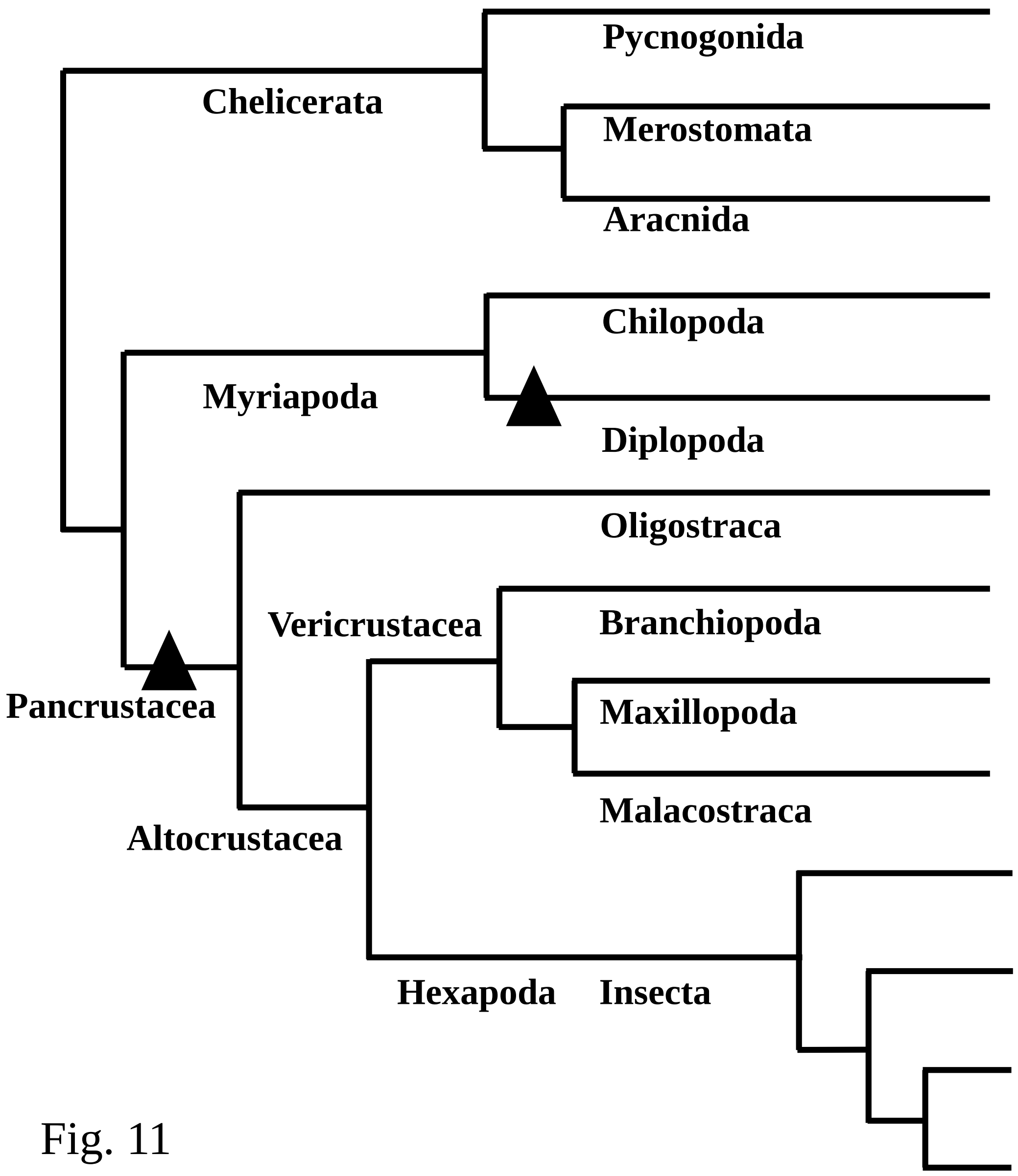

HosaFB
GiciFB
CycaFB
HosaC2
MumuC2
LejaFB
CiinFB
ScsuFB-1
AmspFB-1
TatrFB-2
TatrFB-1
HaadFB-1
HaadFB-2
ScsuFB-2
HaadFB-3
AmspFB-2
NeveFB
84
95
99
99
100
96
66
50
97
100
100
100
100
84
Clade 1
Clade 2
Arthropod factor B
0.2

Fig. 10



	MASP	FB	C3	A2M	iTEP/ CD109	FC/ FC-like
***Ammothea* sp.**	−	+	+	+	+	+
Tachypleus tridentatus	?	+	+	+	?	+
Hasarius adansoni	−	+	+	+	+	+
Scolopendra subspinipes	−	+	+	+	+	+
Niponia nodulosa	−	−	−	+	+	+
***Vargula* sp.**	−	−	−	+	+	−
Daphnia pulex	−	−	−	+	+	−
Pseudocaligus fugu	−	−	+*	+	+	−
Pacifastacus leniusculus	?	?	−	+	+	?
Acyrthosiphon pisum	−	−	−	−	+	−
Tribolium castaneum	−	−	−	+	+	−
Apis merifera	−	−	−	+	+	−
Drosophila melanogaster	−	−	−	−	+	−

Fig. 11

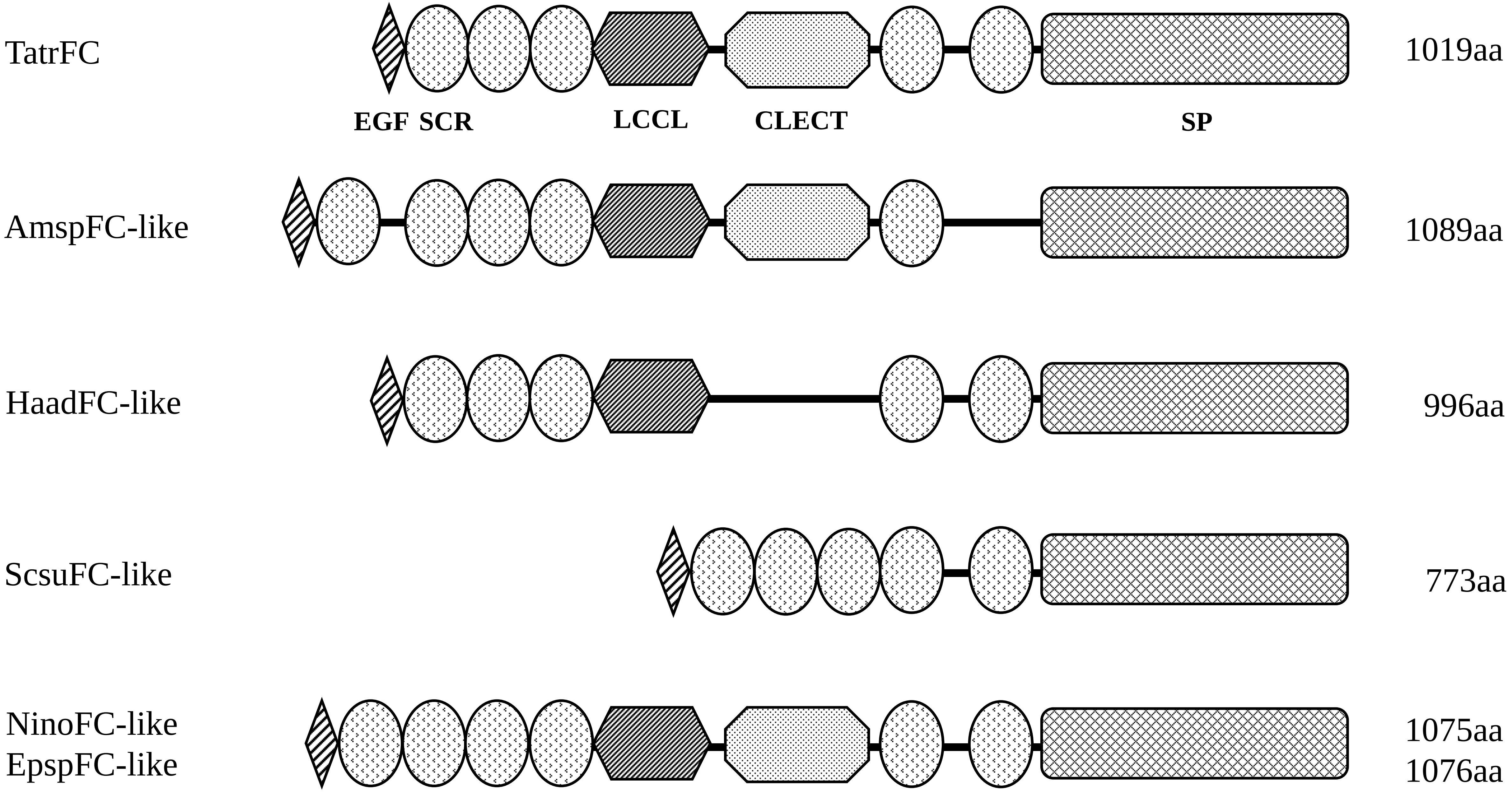

TatrFC
1019aa
EGF
SCR
LCCL
CLECT
SP
AmspFC-like
1089aa
HaadFC-like
996aa
ScsuFC-like
773aa
NinoFC-like
EpspFC-like
1075aa
1076aa

Fig. 12

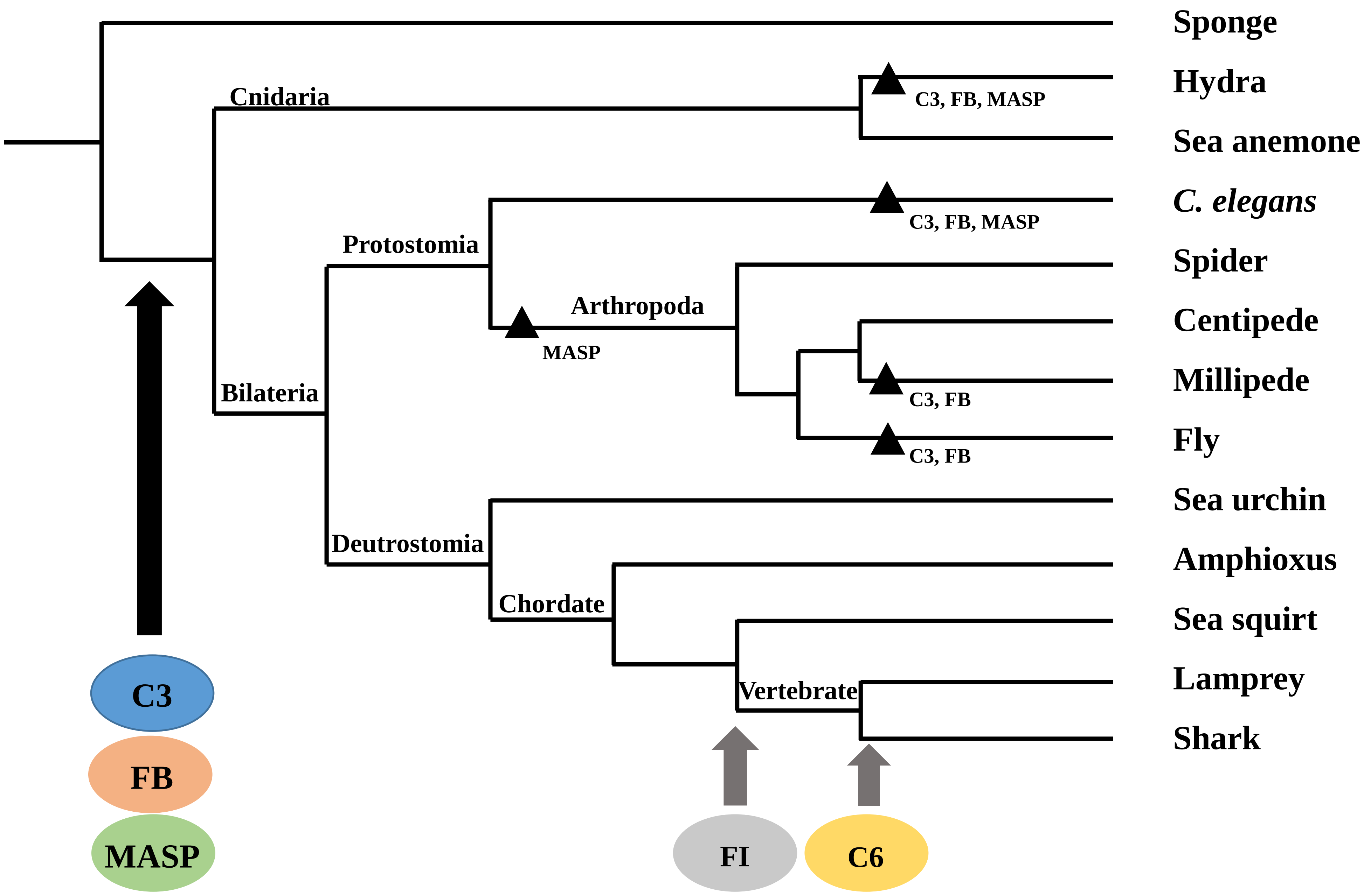

Sponge
Hydra
Sea anemone
C. elegans
Spider
Centipede
Millipede
Fly
Sea urchin
Amphioxus
Sea squirt
Lamprey
Shark
Cnidaria
Bilateria
Protostomia
Arthropoda
Deutrostomia
Chordate
Vertebrate
C3, FB, MASP
C3, FB, MASP
MASP
C3, FB
C3, FB
C3
FB
MASP
FI
C6

Fig. 13

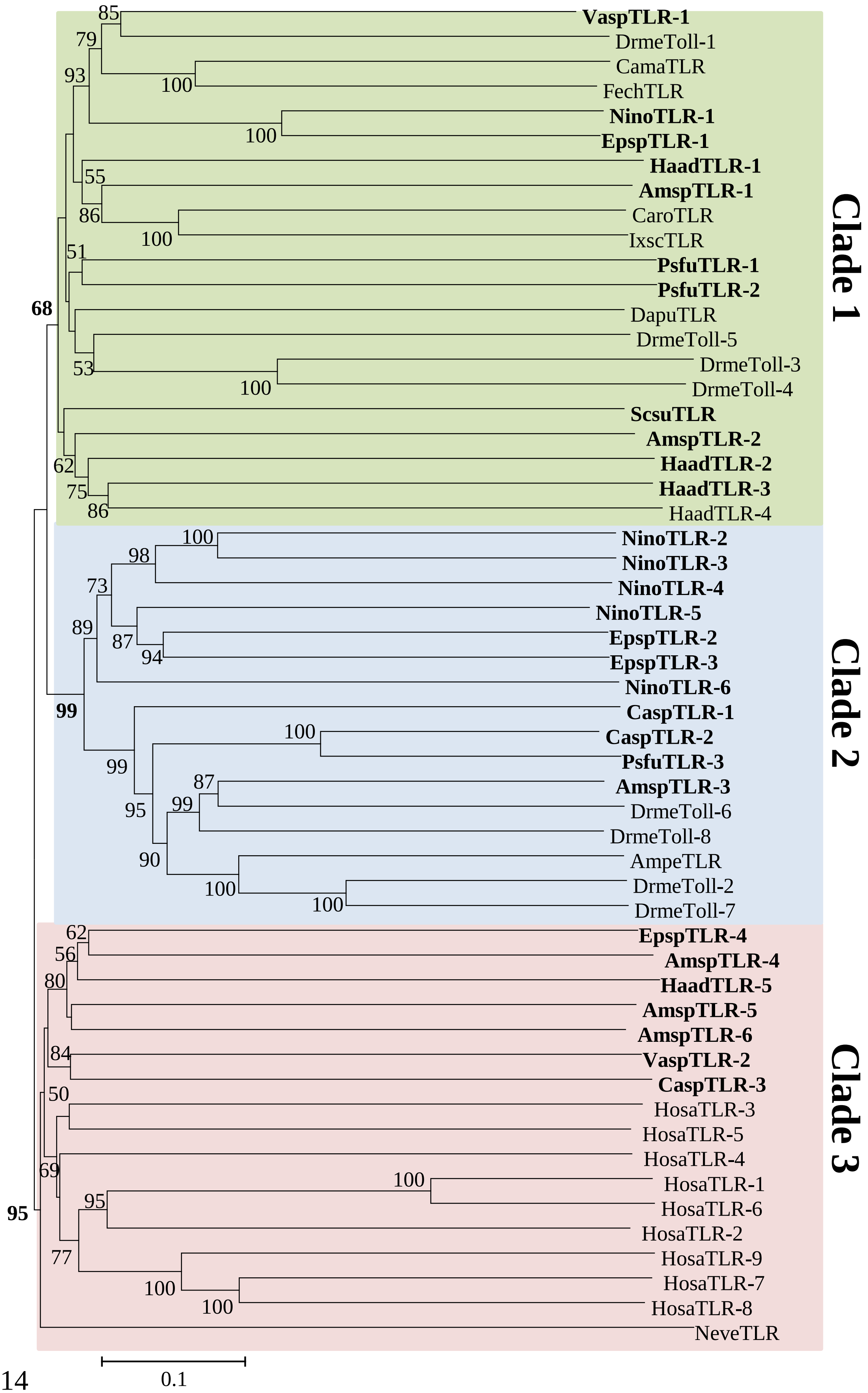
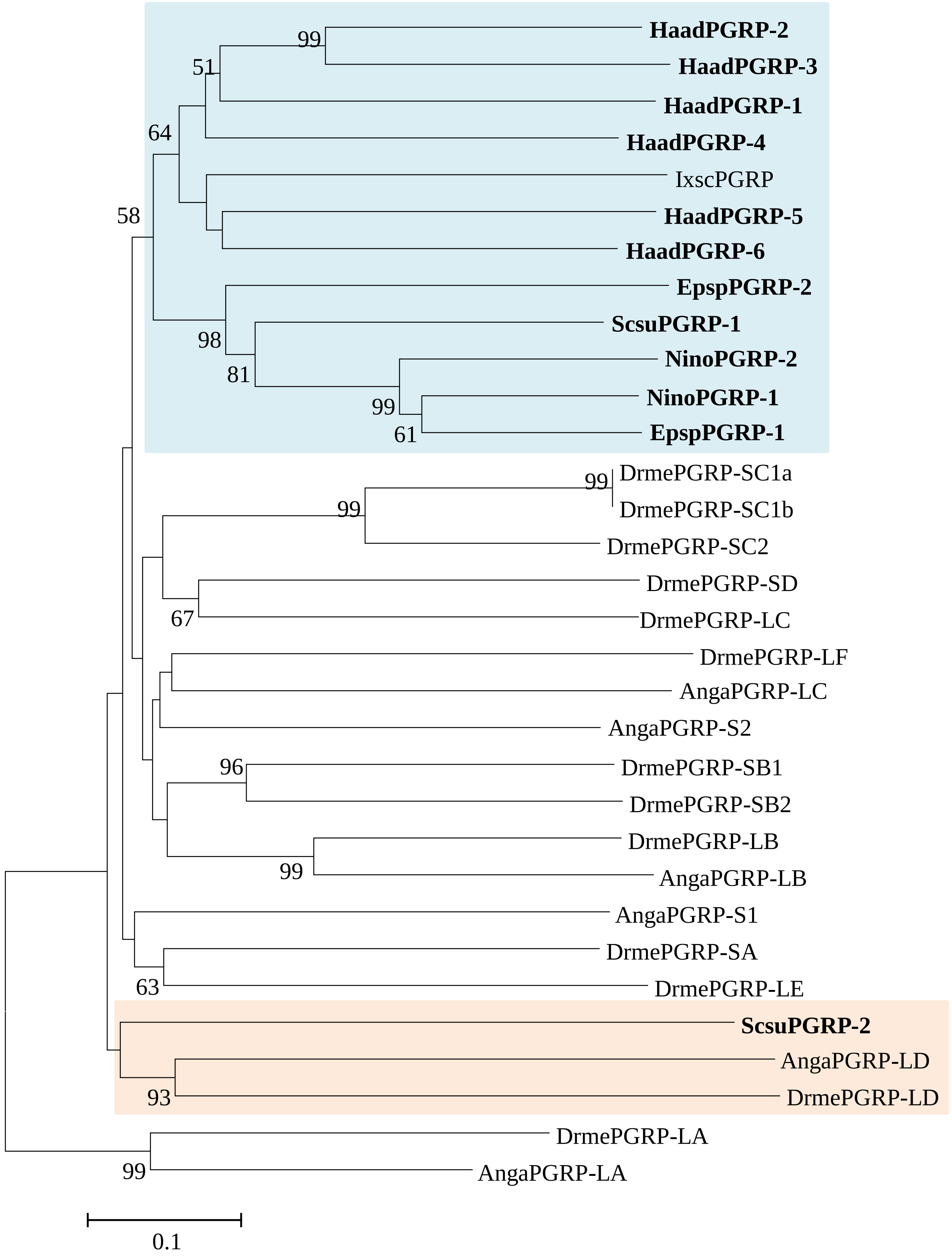

85
VaspTLR-1
79
DrmeToll-1
93
CamaTLR
100
FechTLR
NinoTLR-1
100
EpspTLR-1
HaadTLR-1
55
AmspTLR-1
86
CaroTLR
100
IxscTLR
51
PsfuTLR-1
PsfuTLR-2
68
DapuTLR
DrmeToll-5
53
DrmeToll-3
100
DrmeToll-4
ScsuTLR
AmspTLR-2
62
HaadTLR-2
75
HaadTLR-3
86
HaadTLR-4
Clade 1
100
NinoTLR-2
98
NinoTLR-3
73
NinoTLR-4
89
NinoTLR-5
87
EpspTLR-2
94
EpspTLR-3
NinoTLR-6
99
CaspTLR-1
100
CaspTLR-2
99
PsfuTLR-3
87
AmspTLR-3
95
99
DrmeToll-6
DrmeToll-8
90
AmpeTLR
100
DrmeToll-2
100
DrmeToll-7
Clade 2
62
EpspTLR-4
56
AmspTLR-4
80
HaadTLR-5
AmspTLR-5
AmspTLR-6
84
VaspTLR-2
CaspTLR-3
50
HosaTLR-3
HosaTLR-5
69
HosaTLR-4
100
HosaTLR-1
95
HosaTLR-6
95
HosaTLR-2
77
HosaTLR-9
100
HosaTLR-7
100
HosaTLR-8
NeveTLR
Clade 3
0.1

Fig. 14



HaadPGRP-2
HaadPGRP-3
HaadPGRP-1
HaadPGRP-4
IxscPGRP
HaadPGRP-5
HaadPGRP-6
EpspPGRP-2
ScsuPGRP-1
NinoPGRP-2
NinoPGRP-1
EpspPGRP-1
DrmePGRP-SC1a
DrmePGRP-SC1b
DrmePGRP-SC2
DrmePGRP-SD
DrmePGRP-LC
DrmePGRP-LF
AngaPGRP-LC
AngaPGRP-S2
DrmePGRP-SB1
DrmePGRP-SB2
DrmePGRP-LB
AngaPGRP-LB
AngaPGRP-S1
DrmePGRP-SA
DrmePGRP-LE
ScsuPGRP-2
AngaPGRP-LD
DrmePGRP-LD
DrmePGRP-LA
AngaPGRP-LA
99
51
64
58
98
81
99
61
99
99
67
96
99
63
93
99
0.1

Fig. 15

Table 1. Results of BLAST search against each arthropod TEP fragment

	TEP type	Subject[*]	Identity (%)	Number of clones
Ammothea sp.	C3	Horseshoe crab C3 [*Carcinoscorpius rotundicauda*]	55	10
	A2M-1	Amphioxus A2M [*Branchiostoma floridae*]	65	59
	A2M-2	Amphioxus A2M [*B. floridae*]	65	14
	A2M-3	Amphioxus A2M [*B. floridae*]	62	10
Scolopendra subspinipes	C3	Horseshoe crab C3 [*C. rotundicauda*]	50	249
	A2M	Amphioxus A2M [*B. floridae*]	58	7
	iTEP/CD109-1	Mosquito iTEP/CD109 [*Anopheles stephensi*]	70	5
	iTEP/CD109-2	Bee iTEP/CD109 [*Nasonia vitripennis*]	67	6
Pseudocaligus fugu	C3	Fish C3 [*Fundulus heteroclitus*]	73	1
	A2M-1	Fish A2M [*Danio rerio*]	42	5
	A2M-2	Hydra A2M [*Hydra vulgaris*]	41	4
	A2M-3	Copepod A2M [*Paracyclopina nana*]	51	110
Caligus sp.	C3	Fish C3 [*Cyprinus carpio*]	57	1
	A2M	Sea squirt A2M [*Botyllus schlosseri*]	53	67
Pandaridae gen. et sp.	C3	Fish C3 [*Miichthys miiuy*]	86	1
	A2M	Sea squirt A2M [*B. schlosseri*]	57	71

[*] Subjects shows results of BLAST search which indicated the highest score.

Table 2. Summary statistics of *de novo* transcriptome analysis of several arthropods

	High quality reads	Assembled contigs	Mean contigs length (bp)	N50 value (bp)
Ammothea sp.	93649672	59615	1031	2130
Hasarius adansoni	53164698	102190	635	1266
Scolopendra subspinipes	53851246	155223	1434	3090
Niponia nodulosa	51298384	124129	1961	3754
Epanerchodus sp.	92190042	44365	1392	2886
Vargula sp.	93520202	112077	646	925
Pseudocaligus fugu	91200218	32553	1171	2231
Caligus sp.	54948182	64459	814	1536

Table 3. Results of BLAST search against each Arthropoda TEP obtained by RNA-seq

	TEP type	Length of deduced amino acids	Subject[*]	E-value	Identity (%)
Ammothea sp.	C3	1725	Horseshoe crab C3 [*Carcinoscorpius rotundicauda*]	0	38
	A2M-1	1527	Tick A2M [*Ixodes ricinus*]	0	40
	A2M-2	1614	Tick A2M [*Ornithodoros moubata*]	0	38
	A2M-3	1552	Tick A2M [*I. ricinus*]	0	38
	A2M-4	1496	Tick A2M [*I. ricinus*]	0	38
	iTEP/CD109	1465	Tick iTEP/CD109 [*Ixodes scapularis*]	0	43
Hasarius adansoni	C3-1	1738	Spider C3-1 [*H. adansoni*]	0	100
	C3-2	1695	Spider C3-2 [*H. adansoni*]	0	100
	A2M-1	1717	Spider A2M [*H. adansoni*]	0	100
	A2M-2	1520	Tick A2M [*I. ricinus*]	0	37
	iTEP/CD109-1	1440	Spider iTEP/CD109 [*H. adansoni*]	0	100
	iTEP/CD109-2	1456	Spider iTEP/CD109 [*Stegodyphus mimosarum*]	0	76
Scolopendra subspinipes	C3-1	1749	Horseshoe crab C3 [*Tachypleus tridentatus*]	0	37
	C3-2	1755	Horseshoe crab C3 [*T. tridentatus*]	0	34
	C3-3	1631	Amphioxus C3 [*Branchiostoma belcheri*]	0	34
	A2M	1625	Ant A2M [*Harpegnathos saltator*]	0	36
	iTEP/CD109-1	1447	Spider iTEP/CD109 [*H. adansoni*]	0	39
	iTEP/CD109-2	1423	Spider iTEP/CD109 [*H. adansoni*]	0	39
	iTEP/CD109-3	1437	Water flea iTEP/CD109 [*Daphnia pulex*]	0	49
	iTEP/CD109-4	1430	Spider iTEP/CD109 [*S. mimosarum*]	0	43
Niponia nodulosa	A2M	1452	Tick A2M [*I. ricinus*]	0	34
	iTEP/CD109-1	1437	Spider iTEP/CD109 [*S. mimosarum*]	0	45
	iTEP/CD109-2	1457	Water flea iTEP/CD109 [*D. pulex*]	0	47
	iTEP/CD109-3	1427	Spider iTEP/CD109 [*H. adansoni*]	0	35
Epanerchodus sp.	A2M	1473	Tick A2M [*I. ricinus*]	0	34
	iTEP/CD109-1	1430	Spider iTEP/CD109 [*S. mimosarum*]	0	45
	iTEP/CD109-2	1465	Water flea iTEP/CD109 [*D. pulex*]	0	48
Vargula sp.	A2M-1	853	Shrimp A2M [*Fenneropenaeus chinensis*]	3e-133	34
	A2M-2	638	Horseshoe crab A2M [*Limulus* sp.]	1e-125	37
	A2M-3	1542	Ant A2M [*Acromyrmex echinatior*]	0	32
	iTEP/CD109	323	Termite iTEP/CD109 [*Zootermopsis nevadensis*]	9e-87	45

[*] Subject are results of BLAST search indicated the highest score.

Table 3 (continued)

	TEP type	Length of deduced amino acids	Subject[*]	E-value	Identity (%)
Pseudocaligus fugu	A2M-1	903	Flatworm A2M [*Opisthorchis viverrini*]	4e-51	25
	A2M-2	71	Flatworm A2M [*Echinococcus multilocularis*]	0.65	35
	A2M-3	1715	Copepod A2M [*Paracyclopina nana*]	0	38
	iTEP/CD109	1535	Wasp iTEP/CD109 [*Microplitis demolitor*]	0	36
Caligus sp.	A2M-1	1735	Copepod A2M [*Paracyclopina nana*]	0	37
	A2M-2	1701	Copepod A2M [*Paracyclopina nana*]	0	36
	iTEP/CD109-1	1539	Wasp iTEP/CD109 [*Microplitis demolitor*]	0	37
	iTEP/CD109-2	1429	Termite iTEP/CD109 [*Zootermopsis nevadensis*]	0	39

[*] Subject shows results of BLAST search which indicated the highest score.

Table 4. Characteristic domains, motiefs, and residues of TEPs of arthropods analyzed in this study

		N-terminal CUB domain	*B-α* cleavage site	ANA domain	Bait region	Thioester site	Catalytic histidine	C345C domain	FPKM[*]
Ammothea sp.	C3	−	RPKR	+	−	GCGEQ	H	+	115.33
Hasarius adansoni	C3-1	−	RRKR	+	−	GCGEQ	H	+	120.80
	C3-2	−	RWKR	+	−	GCGEQ	E	+	4.64
Scolopendra subspinipes	C3	−	RRKR	+	−	GCGEQ	H	+	391.14
	C3-2	−	RKKR	+	−	ACGEQ	G	+	7.37
	C3-3	−	RMKR	+	−	GCGEQ	S	−	19.28
Pseudocaligus fugu	C3	−	RKKR	+	−	GCGEQ	H	+	−
Caligus sp.	C3	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	−
Pandaridae gen. et sp	C3	−	RKKR	+	−	GCGEQ	H	+	−
Ammothea sp.	A2M-1	−	RSKR	−	+	GCGEQ	H	−	19.09
	A2M-2	−	RKKR	−	+	GCGEQ	H	−	89.56
	A2M-3	−	−	−	+	GCGEQ	H	−	7.61
	A2M-4	−	RPSR	−	+	GCGEQ	H	−	3.57
H. adansoni	A2M-1	−	RMPR	−	+	GCGEQ	N	−	5.79
	A2M-2	−	−	−	+	GCGEQ	D	−	57.42
S. subspinipes	A2M	−	−	−	+	GCGEQ	H	−	11.96
Niponia nodulosa	A2M	−	−	−	+	GCGEQ	H	−	90.88
Epanerchodus sp.	A2M	−	−	−	+	GCGEQ	H	−	14.34
Vargula sp.	A2M-1	n.d.	−	−	+	GCGEQ	S	−	5.22
	A2M-2	n.d.	n.d.	n.d	n.d.	GCGEQ	N	n.d.	1.99
	A2M-3	−	−	−	+	GCGEQ	H	−	75.56

[*] Fragments per kilobase of exon per million mapped fragments

Table 4 (continued)

		N-terminal CUB domain	*B-α* cleavage site	Bait region	Thioester site	Catalytic histidine	GPI anchor signal sequence	FPKM[*]
Pseudocaligus fugu	A2M-1	−	−	−	GCGEQ	H	−	1.49
	A2M-2	−	−	−	GCGEQ	H	−	1.62
	A2M-3	+	−	+	GCGEQ	S	−	31.34
Caligus sp.	A2M-1	+	−	+	GCGEQ	S	−	36.82
	A2M-2	+	−	+	GCGEQ	S	−	15.09
Pandaridae gen. et sp	A2M	+	−	+	GCGEQ	S	−	−
Ammothea sp.	iTEP/CD109	−	RRGR	−	GCGEQ	N	+	14.37
H. adansoni	iTEP/CD109-1	−	RRKR	−	GCGEQ	H	+	25.91
	iTEP/CD109-2	−	RRRR	−	GCGEQ	H	+	5.18
S. subspinipes	iTEP/CD109-1	−	−	−	GCGEQ	H	−	10.21
	iTEP/CD109-2	−	−	−	GCGEQ	Y	−	68.67
	iTEP/CD109-3	−	−	−	GCGEQ	N	+	4.69
	iTEP/CD109-4	−	−	−	GCGEQ	H	−	45.26
N. nodulosa	iTEP/CD109-1	−	RKKR	−	GCGEQ	S	−	383.84
	iTEP/CD109-2	−	−	−	GCGEQ	H	+	99.58
	iTEP/CD109-3	−	−	−	GCGEQ	D	−	309.18
Epanerchodus sp.	iTEP/CD109-1	−	RKKR	−	GCGEQ	N	−	156.65
	iTEP/CD109-2	−	−	−	GCGEQ	H	+	10.54
Vargula sp.	iTEP/CD109	n.d.	n.d.	−	n.d.	n.d.	+	1.21
P. fugu	iTEP/CD109	−	−	−	GCGEQ	F	+	8.13
Caligus sp.	iTEP/CD109-1	−	−	−	GCGEQ	F	+	3.51
	iTEP/CD109-2	−	−	−	GCGEQ	F	+	12.85

[*] Fragments per kilobase of exon per million mapped fragments

Table 5. Results of BLAST search for TLR gene

		Length of deduced amino acids	Subject*	Identity (%)
Ammothea sp.	TLR1	1098	Horseshoe crab TLR [*Tachypleus tridentatus*]	35
	TLR2	990	Spider TLR [*Stegodyphus mimosarum*]	28
	TLR3	1214	Tick TLR [*Ixodes scapularis*]	51
	TLR4	827	Spider TLR13 [S. *mimosarum*]	33
	TLR5	824	Butterfly TLR13 [*Danaus plexippus*]	29
	TLR6	772	Termite TLR4 [*Zootermopsis nevadensis*]	29
Hasarius adansoni	TLR1	1007	Spider TLR [*S. mimosarum*]	48
	TLR2	693	Spider TLR [*S. mimosarum*]	30
	TLR3	961	Spider TLR [*S.mimosarum*]	53
	TLR4	942	Spider TLR [*S. mimosarum*]	41
	TLR5	1046	Spider TLR4 [*S. mimosarum*]	53
Scolopendra Subspinipes	TLR	1604	Honey bee TLR [*Apis dorsata*]	28
Niponia nodulosa	TLR1	1006	Termite TLR [*Z. nevadensis*]	35
	TLR2	1070	Tick TLR [*I. scapularis*]	28
	TLR3	1047	Louse TLR [*Pediculus humanus corporis*]	28
	TLR4	1062	Louse TLR [*P. humanus corporis*]	30
	TLR5	661	Louse TLR [*P. humanus corporis*]	40
	TLR6	1115	Louse TLR [*P. humanus corporis*]	30
Epanerchodus sp.	TLR1	1003	Termite TLR [*Z. nevadensis*]	35
	TLR2	1113	Louse TLR [*P. humanus corporis*]	32
	TLR3	1120	Tick TLR [*I. scapularis*]	32
	TLR4	814	Termite TLR4 [*Z. nevadensis*]	31
Vargula sp.	TLR1	869	Termite TLR [*Z. nevadensis*]	34
	TLR2	863	Beetle TLR2 [*Tribolium castaneum*]	27
Pseudocaligus fugu	TLR1	678	Mosquito TLR [*Anopheles sinensis*]	39
	TLR2	1041	Acron worm TLR3 [*Saccoglossus kowalevskii*]	27
	TLR3	1229	Planthopper TLR6 [*Nilaparvata lugens*]	41
Caligus sp.	TLR1	833	Fruit fly Toll [*Drosophila persimilis*]	29
	TLR2	941	Louse TLR [*P. humanus corporis*]	39
	TLR3	873	Fruit fly Toll [*Drosophila willistoni*]	28

[*] Subject shows results of BLAST search which indicated the highest score.

Table 6. Results of BLAST search for PGRP gene

		Length of deduced amino acids	Subject*	Identity (%)
Ammothea sp.	—	—	—	—
Hasarius adansoni	PGRP1	221	Spider PGRP [*Stegodyphus mimosarum*]	59
	PGRP2	209	Spider PGRP [*Nephila pilipes*]	62
	PGRP3	212	Spider PGRP [*S. mimosarum*]	64
	PGRP4	198	Spider PGRP [*S. mimosarum*]	48
	PGRP5	203	Spider PGRP [*S. mimosarum*]	60
	PGRP6	194	Spider PGRP [*S. mimosarum*]	54
Scolopendra subspinipes	PGRP1	165	Spider PGRP [*S. mimosarum*]	51
	PGRP2	296	Frog PGRP [Xenopus tropicalis]	36
Niponia nodulosa	PGRP1	208	Clam PGRP S1 [Hyriopsis cumingii]	53
	PGRP2	208	Tick PGRP [*Ixodes scapularis*]	49
Epanerchodus sp.	PGRP1	207	Tick PGRP [*I. scapularis*]	51
	PGRP2	216	Clam PGRP S1 [*H. cumingii*]	49
Vargula sp.	—	—	—	—
Pseudocaligus fugu	—	—	—	—
Caligus sp.	—	—	—	—

* Subject shows results of BLAST search which indicated the highest score.

Supplementary Fig. S1. Primary structure of the *Ammothea* sp. TEP genes

The complete nucleotide and deduced amino acid sequences of AmspC3, AmspA2M-1, AmspA2M-2, AmspA2M-3, AmspA2M-4, and AmspiTEP/CD109 are shown. Signal peptide region is underlined. The ANA and C345C domains, the A2M bait region, and the thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), six conserved cysteine residues in the ANA domain (CC, C, C, CC), the conserved flanking residues of the A2M bait region (C and FPEXW), the catalytic histidine residue, and the conserved cysteine residues in the C345C domain are shown in bold. Putative signal peptide region for GPI-anchor attachment is double underlined. The sequences amplified by RT-PCR using degenerate primers are indicated by red open box. The positions of the primers used for RACE are indicated by the arrows above the nucleotide sequence. The sequence of AmspA2M-3 obtained by 3' RACE is

Supplementary Fig. S2. Primary structure of the *Scolopendra subspinesis* TEP genes

The complete nucleotide and deduced amino acid sequences of SuscC3, SuscC3-2, SuscC3-3, SuscA2M, SusciTEP/CD109-1, SusciTEP/CD109-2, ScsuiTEP/CD109-3, and SusciTEP/CD109-4 are shown. Signal peptide region is underlined. The ANA and C345C domains, the A2M bait region, and the thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), six conserved cysteine residues in the ANA domain (CC, C, C, CC), the conserved flanking residues of the A2M bait region (C and FPEXW), the catalytic histidine residue, and the conserved cysteine residues in the C345C domain are shown in bold. Putative signal peptide region for GPI-anchor attachment is double underlined. The sequence amplified by RT-PCR using degenerate primers is indicated by red open box. The positions of the primers used for RACE are indicated by the arrows above the nucleotide sequence.

Supplementary Fig. S3. Primary structure of the *Pseudocaligus fugu* TEP genes

The complete nucleotide and deduced amino acid sequences of PsfuC3, PsfuA2M-1, PsfuA2M-2, PsfuA2M-3, and PsfuiTEP/CD109 are shown. The sequence amplified by RT-PCR using degenerate primers is indicated by red open box. Signal peptide region is underlined. The CUB, ANA and C345C domains, the A2M bait region, and the thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), six conserved cysteine residues in the ANA domain (CC, C, C, CC), the conserved flanking residues of the A2M bait region (C and FPEXW), the catalytic histidine residue, and the conserved cysteine residues in C345C domain are shown in bold. Putative signal peptide region for GPI-anchor attachment is double underlined. The positions of the primers used for RACE are indicated by solid arrows above the nucleotide sequence. The positions of the degenerate primer and gene specific primer for RT-PCR are indicated by the dotted arrows above the nucleotide sequence in PsfuC3.

Supplementary Fig. S4. Primary structure of the *Caligus* sp. TEP genes

The complete nucleotide and deduced amino acid sequences of CaspA2M and CaspiTEP/CD109-1, and the partial nucleotide and deduced amino acid sequences of CaspC3, CaspA2M-2, and CaspiTEP/CD109-2 are shown. Signal peptide region is underlined. The CUB domain, A2M bait region, and thioester site (GCGEQ) are indicated by open boxes. The conserved flanking residues of the A2M bait region (C and FPEXW) are shown in bold. The sequence amplified by RT-PCR using degenerate primers is indicated by red open box. Putative signal peptide region for GPI-anchor attachment is double underlined. The positions of the primers used for RACE are indicated by the arrows above the nucleotide sequence.

Supplementary Fig. S5. Primary structure of the Pandaridae gen. et sp. TEP genes

The partial nucleotide and deduced amino acid sequences of PaspC3 and the complete nucleotide and deduced amino acid sequences of PaspA2M are shown. The sequence amplified by RT-PCR using degenerate primers is indicated by red open box. Signal peptide region is underlined. The CUB domain, ANA domain, C345C domain, A2M bait region and thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), six conserved cysteine residues in the ANA domain (CC, C, C, CC), the conserved flanking residues of the A2M bait region (C and FPEXW), the catalytic histidine residue and the conserved cysteine residues in the C345C domain are shown in bold. The positions of the primers used for RACE are indicated by the solid arrows above the nucleotide sequence. The positions of the degenerate and gene specific primers for RT-PCR are indicated by the dotted arrows above the nucleotide sequence in PaspC3.

Supplementary Fig. S6. Primary structure of the *Hasarius adnsoni* TEP genes

The complete nucleotide and deduced amino acid sequences of HaadA2M-2 and HaadiTEP/CD109-2 are shown. Signal peptide region is underlined. The A2M bait region and the thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), the conserved flanking residues of the A2M bait region (C and FPEXW) and the catalytic histidine residue are shown in bold. Putative signal peptide region for GPI-anchor attachment is double underlined.

Supplementary Fig. S7. Primary structure of the *Epanerchodus* sp. TEP genes

The complete nucleotide and deduced amino acid sequences of EpspA2M, EpspiTEP/CD109-1, and EpspiTEP/CD109-2 are shown. Signal peptide region is underlined. The A2M bait region and the thioester site (GCGEQ) are indicated by open boxes. The β-α cleavage site (RXXR), the conserved flanking residues of the A2M bait region (C and FPEXW), and the catalytic histidine residue are shown in bold. Putative signal peptie region for GPI-anchor attachment is double underlined.

Supplementary Fig. S8. Primary structure of the *Vargula* sp. TEP genes

The complete nucleotide and deduced amino acid sequences of VaspA2M-3 and the partial nucleotide and deduced amino acid sequences of VaspA2M-1, VaspA2M-2, and VaspiTEP/CD109 are shown. Signal peptide region is underlined. The A2M bait region and the thioester site (GCGEQ) are indicated by open boxes. The conserved flanking residues of the A2M bait region (C and FPEXW) are shown in bold. Putative signal peptide region for GPI-anchor attachment is double underlined.

Supplementary Fig. S9. Multiple alignment of the TEP amino acid sequences

The deduced amino acid sequences of *Ammothea* sp. (Amsp) C3, A2M-1, A2M-2, A2M-3, A2M-4 and iTEP/CD109 (LC009021, LC009022, LC009023, LC009024, LC009025 and LC009026), *Hasarius adansoni* (Haad) C3-1, C3-2, A2M-1, A2M-2, iTEP/CD109-1 and iTEP/CD109-2 (BAK64109, BAK64110, BAK64111, LC009009, BAK64112 and LC009010), *Scolopendra subspinipes* (Scsu) C3-1, C3-2, C3-3 A2M-1, iTEP/CD109-1, iTEP/CD109-2, iTEP/CD109-3 and iTEP/CD109-4 (LC009029, LC009030, LC009031, LC009032, LC009033, LC009034, LC009035 and LC009036), *Niponia nodulosa* (Nino) A2M, iTEP/CD109-1, iTEP/CD109-2 and iTEP/CD109-3 (LC009014, LC009015, LC009016 and LC009017), *Epanerchodus* sp. (Epsp) A2M, iTEP/CD109-1 and iTEP/CD109-2 (LC009018, LC009019 and LC009020), and *Vargula* sp. (Vasp) A2M-1, A2M-2, A2M-3 and iTEP/CD109 (LC009039, LC009040, LC009041 and LC009042), *Pseudocaligus fugu* (Psfu) C3, A2M-1, A2M-2, A2M-3 and iTEP/CD109, *Caligus* sp. (Casp) C3, A2M-1, A2M-2, iTEP/CD109-1 and iTEP/CD109-2, Pandaridae gen. et sp. (Pasp) C3 and A2M were aligned together with that of *Homo sapiens* (Hosa) C3, C4, C5, A2M, and iTEP/CD109 (AAA85332, P0C0L4, AAA51925, P01023, and NP_598000). Amino acid sequences were aligned using ClustalX and GeneDoc. Characteristic residues, domains and boundaries of domain are indicated above or bottom the aligned sequences. The β-α processing site (RXXR), the conserved cysteine residues in the C3 anaphylatoxin region (ANA), the thioester site (GCGEQ), the catalytic histidine, the conserved flanking residues of the A2M bait region, C and FPEXW, the conserved cysteine residues of C345C domain.

Supplementary Fig. S10. Primary structure of the *Ammothea* sp. FB genes

The complete nucleotide and deduced amino acid sequences of AmspFB-1 and AmspFB-2 are shown. Signal peptide region is underlined. The SCR (short consensus repeat) domain, the von Willebrand factor type A (vWFA) domain and the serine protease (SP) domain are indicated by the dotted, red, and green underlines, respectively. The factor D cleavage site, the Mg2+ binding site and the protease activate site are shown in bold.

Supplementary Fig. S11. Primary structure of the *Hasarius adansoni* FB genes

The complete nucleotide and deduced amino acid sequence of HaadFB-1, HaadFB-2 and HaadFB-3 are shown. Signal peptide region is underlined. The SCR, vWFA and SP domains are indicated by the dotted, red and green underlines, respectively. The factor D cleavage site, the Mg2+ binding site and the protease activate site are shown in bold.

Supplementary Fig. S12. Primary structure of the *Scolopendra subspinipes* FB genes

The complete nucleotide and deduced amino acid sequences of ScsuFB-1 and ScsuFB-2 are shown. Signal peptide region is underlined. The SCR, vWFA,and SP domains are indicated by the dotted, red and green underlines, respectively. The factor D cleavage site, the Mg2+ binding site and the protease activate site are shown in bold.

Supplementary Fig. S13. Multiple alignments of FB amino acid sequences

The deduced amino acid sequences of *Ammothea* sp. (Amsp) FB-1 and FB-2, *Tachypleus tridentatus* (Tatr) FB-1 and FB-2, *Hasarius adansoni* (Haad) FB-1, FB-2, and FB-3, *Scolopendra subspinipes* (Scsu) FB-1 and FB-2 were aligned together with that of *Homo sapiens* (Hosa) FB. The names for the characteristic domains and boundaries of domain are indicated above the aligned sequences. Functionally important amino acid residues reported in human FB, such as the factor D cleavage site, the Mg2+ binding site, and the protease active site, are indicated in bold.

AmspC3

```
   1  TGCTAAAAACCATGTATCGATCGAAATTGCATTTTTTCTAGAATGGCGAAAAATTTGTTCTTTATTATTTGTGTTTGCATGTTGCCATGCTCCATCTTCT  100
                                                M  A  K  N  L  F  F  I  I  C  V  C  M  L  P  C  S  I  F

 101  CTCAGGCAGACTACTTTGTCGTGGCACCGAATATGTTCAGATTTGGTGTTCCGGAAACCGTAAGTATCAACATACACGGAATTCAAGGTTCAGTCGATGT  200
      S  Q  A  D  Y  F  V  V  A  P  N  M  F  R  F  G  V  P  E  T  V  S  I  N  I  H  G  I  Q  G  S  V  D  V

 201  TGAAGTTTTTCTTCAAGATTCACCCGGTCGTGGCTTTACGTTCTCCCAAAAAGTTGTCAAAGTTGAAAATGAAAAATCAAGTCAAGTTTCGGTGCTGGTG  300
        E  V  F  L  Q  D  S  P  G  R  G  F  T  F  S  Q  K  V  V  K  V  E  N  E  K  S  S  Q  V  S  V  L  V

 301  AAGCCCTCCGATATACCTGACGTTGTTGTGACGGATCCATCGGCCCGAATATATGTCTACCTGGTTGTGAAATGTACAAGTCCCAAGCTCACGTTTAAGA  400
       K  P  S  D  I  P  D  V  V  V  T  D  P  S  A  R  I  Y  V  Y  L  V  V  K  C  T  S  P  K  L  T  F  K

 401  AAGAGGCTCTCGTGCTCGTTAGTCCGCACAACGGTTTTGTCTTTGTGCAAACAGACAAACCAATCTACACGCCAAGACAAGAAGTCAGGATGAGGATTAT  500
      K  E  A  L  V  L  V  S  P  H  N  G  F  V  F  V  Q  T  D  K  P  I  Y  T  P  R  Q  E  V  R  M  R  I  I

 501  TCCCTTGGATCAAGATCTAACCTACATGAAATCTCCCTTTCGGCTTCGGATTTTGAACCCACAAAATATCACCGTGAGGCGTATAGACTACGATGAAGGT  600
        P  L  D  Q  D  L  T  Y  M  K  S  P  F  R  L  R  I  L  N  P  Q  N  I  T  V  R  R  I  D  Y  D  E  G

 601  AACAAGCAAGGAATCATCACCGAAATCTACAAGTTTCCGAGACAAGCCATAGTGGGCATTTGGGGCATTTCCGTCGAGCATGGTCATGAGCTTTCTACAA  700
       N  K  Q  G  I  I  T  E  I  Y  K  F  P  R  Q  A  I  V  G  I  W  G  I  S  V  E  H  G  H  E  L  S  T

 701  CGACGTATGTTTCATTCGAAGTAAAGAAATATGTGCTCCCAACTTTCGATGTCAAAATTGTTGTTCCCGCAACGATAATGGAGGATACCGAAGAGATAAT  800
      T  T  Y  V  S  F  E  V  K  K  Y  V  L  P  T  F  D  V  K  I  V  V  P  A  T  I  M  E  D  T  E  E  I  M

 801  GGGGAAAGTTATTGCTAGGTATGTGTATGATAAACCAGTTTTTGGAGCTGTTCGTCTGCGATATGGTGTCAAGGCAGTAAACGGAGATATTTCATACCTT  900
        G  K  V  I  A  R  Y  V  Y  D  K  P  V  F  G  A  V  R  L  R  Y  G  V  K  A  V  N  G  D  I  S  Y  L

 901  CCCGATGACACAAAAGTCGTAATTAAAAAGATAAGGGGCGAAGCCACGTTCTCCATACCTATGAAAATGGTGACCAACAAGTTCAAGAGCTTCCCCGAAG  1000
       P  D  D  T  K  V  V  I  K  K  I  R  G  E  A  T  F  S  I  P  M  K  M  V  T  N  K  F  K  S  F  P  E

1001  GTCACAAATTCCATGTCGAGGCGACCGTTGTCGAAGGTGTTTCTGGAAAAGAGTTCATGGAAATCAGCGAAGATGCTTATTTTCAAAAAAGTCCGTACAT  1100
      G  H  K  F  H  V  E  A  T  V  V  E  G  V  S  G  K  E  F  M  E  I  S  E  D  A  Y  F  Q  K  S  P  Y  I

1101  TATCAGCTTCAAAAAGTCTGTCAACACTTTCAAGATCGGACTCATTGTTTACATCCAGGTACTGGTTCAGTTCGCCTCAAAGGCACCGGCACCTGGAGTT  1200
        I  S  F  K  K  S  V  N  T  F  K  I  G  L  I  V  Y  I  Q  V  L  V  Q  F  A  S  K  A  P  A  P  G  V

1201  GATGTCGTGTTCACTGTGAAGACAGATGATGGAGTCGCTCTTGCTATAACAAATAATAAACGACGCTCCGACAATTCGGGTTACGTTGAAGTGAAGATTA  1300
       D  V  V  F  T  V  K  T  D  D  G  V  A  L  A  I  T  N  N  K  R  R  S  D  N  S  G  Y  V  E  V  K  I

1301  GTGTTCCCAATACGAAAAAACTAACTGTCGAAGTCGTCACGGATGATGTCAAGCTGAAGGCAAATCAGCAAGGAAAGGGGACATTTGTCATGCTTCCATA  1400
      S  V  P  N  T  K  K  L  T  V  E  V  V  T  D  D  V  K  L  K  A  N  Q  Q  G  K  G  T  F  V  M  L  P  Y

1401  TGCATCACCAAATGAAGCTTATCTTGGAATAAAACCATTGGACGCAAGTCGAAAGCTACAGAAAGGGGTTCAGTTTGACGGTACGGTCATAGTCAATCCT  1500
        A  S  P  N  E  A  Y  L  G  I  K  P  L  D  A  S  R  K  L  Q  K  G  V  Q  F  D  G  T  V  I  V  N  P

1501  TGGCAACATGTCAATCGAGTCATGTTTTCTGTCATCAGCAGAGGTAAAGTTTTGGTCACTGGAAAGACTGCTACGAAAGATGGCAAGAATACTCTTGAAC  1600
       W  Q  H  V  N  R  V  M  F  S  V  I  S  R  G  K  V  L  V  T  G  K  T  A  T  K  D  G  K  N  T  L  E

1601  GAACATTCACTTTTAAAGTTACCGAGCAGATGGTTCCAAGCATTCGGGTCATCGTGTTCGCAAATTACAAAGGTCATTTGATTGCCGATTCAATTTGGAT  1700
      R  T  F  T  F  K  V  T  E  Q  M  V  P  S  I  R  V  I  V  F  A  N  Y  K  G  H  L  I  A  D  S  I  W  M

1701  GGACGTCGAAGACACATGTCAACATGAAATCGAAACGAGCATCACTCCACAATTTGACCTGAAACGTCCGGGTGACGAAGGAAGCATCGTTATCAAGGCG  1800
        D  V  E  D  T  C  Q  H  E  I  E  T  S  I  T  P  Q  F  D  L  K  R  P  G  D  E  G  S  I  V  I  K  A

1801  GCAAAGAAAACAACCGTTGGATTACTTTCAGTCGATAAAGCTGTTTACGTATTGCGTAACAAGGGTCTTCTGTCTGCATCTAAGGTATATAAAACAATGG  1900
       A  K  K  T  T  V  G  L  L  S  V  D  K  A  V  Y  V  L  R  N  K  G  L  L  S  A  S  K  V  Y  K  T  M

1901  AAGGTTATGACATTGGATGCGGAGCTGGAGGCGGGCAAAATTCAGCGAACGTCTTGGAAAAAAGTGGCGCCATCATCTTGACCAGTTCAACCATTACGAA  2000
      E  G  Y  D  I  G  C  G  A  G  G  G  Q  N  S  A  N  V  L  E  K  S  G  A  I  I  L  T  S  S  T  I  T  N

2001  TGAGAAACGTGATCATTTCAGCTGTAATGATCATTCAAGGCCAAAACGAGCTCTCGAAGAGCATTCAGCCAAATACACCGATGCTTTTACAAAAGCCTGC  2100
        E  K  R  D  H  F  S  C  N  D  H  S  R  P  K  R  A  L  E  E  H  S  A  K  Y  T  D  A  F  T  K  A  C
                                        β-α cleabage site
2101  TGCATCATGGGTCAAAGACCAACGAAATTTAGCGTTTCTTGCCAAGTTAGAGCGCAGATCCTAAAAGTTCACCTGACAGAAAAGTATCCGGACAAGTTGA  2200
       C  I  M  G  Q  R  P  T  K  F  S  V  S  C  Q  V  R  A  Q  I  L  K  V  H  L  T  E  K  Y  P  D  K  L
                                               ANA domain
2201  ATTGTGCCGATGCGTTCCTCAGCTGCTGCAGCTCCAAAAGCGTCATAATGGCCAGATCAGGTGCCAACGAAGCTCTTCAATTTGTCGATGATATAGATGC  2300
      N  C  A  D  A  F  L  S  C  C  S  S  K  S  V  I  M  A  R  S  G  A  N  E  A  L  Q  F  V  D  D  I  D  A

2301  CAAGGAATCTCAATCAATTGTTCGGTCCTCCTTCCCTGAAAGCTGGCTGTATGTGGCAGAACATATGGGAGAATCAACGGAGAAAAAAATCGCCACCCAT  2400
        K  E  S  Q  S  I  V  R  S  S  F  P  E  S  W  L  Y  V  A  E  H  M  G  E  S  T  E  K  K  I  A  T  H

2401  TTGCCGCATAGCATAACAACATGGGTATTTCAAGCGTTAAGCGTCTCACCAAACCGAGGTCTGTGTATAGCCAAACCGCAGGAAGTTATCGTTCGGCAAA  2500
       L  P  H  S  I  T  T  W  V  F  Q  A  L  S  V  S  P  N  R  G  L  C  I  A  K  P  Q  E  V  I  V  R  Q

2501  CGGTATTTTTGGACGTCAATATGCCATACAGCGTTGTTCGAAACGAAGAAATTGAAATCAAAGTTAACGTTTTCAACTACCTTTCGACTACCGTTCCCGT  2600
      T  V  F  L  D  V  N  M  P  Y  S  V  V  R  N  E  E  I  E  I  K  V  N  V  F  N  Y  L  S  T  T  V  P  V

2601  TACTGTTTACATGTATGGAGTCGATGGACTTTGCACCGGAGCCGAAGCCGGTAAACGAACTGATAGAAAGCTGATACGTGTGCCGAAGAATGATGCCGTC  2700
        T  V  Y  M  Y  G  V  D  G  L  C  T  G  A  E  A  G  K  R  T  D  R  K  L  I  R  V  P  K  N  D  A  V

2701  TCAACCGGATTTTTCATCATCCCGTTGCGAATTGGAACTTACAAGTTAAGGTTTGTGGCATTGTCGGAAGCGAAGACGGACGTCGTCGAAAAAACTCTCC  2800
       S  T  G  F  F  I  I  P  L  R  I  G  T  Y  K  L  R  F  V  A  L  S  E  A  K  T  D  V  V  E  K  T  L

2801  ATGTCCGCGCAGAAGGTTCCCAACAGTTCAAGGACAGCTCGACCGTCCTTGATCCCAGCAATTACAATGGACGCCCGGGATCAGCAATTTGGATTCAAAA  2900
      H  V  R  A  E  G  S  Q  Q  F  K  D  S  S  T  V  L  D  P  S  N  Y  N  G  R  P  G  S  A  I  W  I  Q  K
```

Supplementary Fig. S1

```
2901  ACCGGATGGTTATTCGTACTTTTTGGACAGCAAAAACCGCAAGCTGGTCGTATCTCTACCGATCATACCACCGGTCAATTCCATCGCAGAAACCGAAATA  3000
        P  D  G  Y  S  Y  F  L  D  S  K  N  R  K  L  V  V  S  L  P  I  I  P  P  V  N  S  I  A  E  T  E  I

3001  GCAACAATCAATATTATCGGTGACAGCTTTGGGCCAACGGTTAAGACGGCTTTCACAAAGAATGAGCTCGGCAAGTTCATCAACAAGCCAAGAGGATGTG  3100
       A  T  I  N  I  I  G  D  S  F  G  P  T  V  K  T  A  F  T  K  N  E  L  G  K  F  I  N  K  P  R  G  C
                                                                                                 Thioester site
3101  GCGAGCAGAATATGATGCTTATGGCTCCGTGTCTCTACACGATGAAGTACCTTGCCGCAACTGGAAAAATTGGAGTCAATGAAGAGAAAACAGGATATGG  3200
      G  E  Q  N  M  M  L  M  A  P  C  L  Y  T  M  K  Y  L  A  A  T  G  K  I  G  V  N  E  E  K  T  G  Y  G

3201  ATGGATCAGACTAGGTTACGAGCGCCAGTTGAACTTCAGAAGAAAAGATGGTGGCTATTCTGCCGGAGGGACACGAAAATCCAGTACCTGGCTGACAGCA  3300
        W  I  R  L  G  Y  E  R  Q  L  N  F  R  R  K  D  G  G  Y  S  A  G  G  T  R  K  S  S  T  W  L  T  A

3301  TTTGTCATGAAATGCTTCTGTCAGGCTGATGCCTTCATAAACGTTGACGAAACCGTCGGTTGCACGGGACTTCAATGGCTGGTCAACAGCCAACAACAGG  3400
       F  V  M  K  C  F  C  Q  A  D  A  F  I  N  V  D  E  T  V  G  C  T  G  L  Q  W  L  V  N  S  Q  Q  Q

3401  ATGGGTCATTCGTCGACAAGCATCCAGTTTACCATCGAGAAATGACGGGCGGAATAAAAGGAAGTTTCGCCATGACAGCATTCGTTCTGATGTCGCTGAT  3500
      D  G  S  F  V  D  K  H  P  V  Y  H  R  E  M  T  G  G  I  K  G  S  F  A  M  T  A  F  V  L  M  S  L  M
                              Catalytic histidine
3501  GGAATGTACGACATGCAAGATTGAAGGACTGGAAATCACTCGCCAACAGGCAATTGCCTTCTTGGAGAAAGGCATCCAGACAAAACGGCGACCTTACGAG  3600
        E  C  T  T  C  K  I  E  G  L  E  I  T  R  Q  Q  A  I  A  F  L  E  K  G  I  Q  T  K  R  R  P  Y  E

3601  CTTGCTGTTGCCACTTATGCCCTGACTCTTGCCAAGAGTCCCAAGCATCAAGATGCAATTAAACTTCTCAAAGATTCATCCATCTACAATAAAAACGAAG  3700
       L  A  V  A  T  Y  A  L  T  L  A  K  S  P  K  H  Q  D  A  I  K  L  L  K  D  S  S  I  Y  N  K  N  E

3701  ACCAGCGTCAGTGGGAGAAAGGCAATGGGGCATTGAGCGTGGAAGCAACTGCTTATGGTCTTCTCACCTTGGTCTTGACCAATGACCTGCCTTACGCCAA  3800
      D  Q  R  Q  W  E  K  G  N  G  A  L  S  V  E  A  T  A  Y  G  L  L  T  L  V  L  T  N  D  L  P  Y  A  K

3801  AGCGGTTGTCGGTTGGCTGAACGCCAACAGAGGCTTCGGCGGTGTATGGAAGTCCACCCAGGATACCGTGGTAGCTCTGCAAGCAATGTCTGAGTATGCC  3900
        A  V  V  G  W  L  N  A  N  R  G  F  G  G  V  W  K  S  T  Q  D  T  V  V  A  L  Q  A  M  S  E  Y  A

3901  ATAAAGTCGCGACAGCCAAAGATCGATATGGCCATCAACATCACATCAGAAACAGATCGAAACTTCCTCAAAAGCTTGAGGATCAACAACGAAAATTTTC  4000
       I  K  S  R  Q  P  K  I  D  M  A  I  N  I  T  S  E  T  D  R  N  F  L  K  S  L  R  I  N  N  E  N  F

4001  AGGATTTGCAAACTATCGATATTAACAAGCCCGGTGGGTTGATATTCGTTGACGCTTATGGTGCTGGCATCGCCCATCTTAATCTCAGACTTCAGTACAA  4100
      Q  D  L  Q  T  I  D  I  N  K  P  G  G  L  I  F  V  D  A  Y  G  A  G  I  A  H  L  N  L  R  L  Q  Y  N

4101  TATTCCGGTTCCTCCTAACGAAACATGTGCCTTCGAACTGACCATTAAAAACGAAGAAGTTGAGGAAGATATCAAGCCTAAAGCACTGGCTGGTCTGGAG  4200
        I  P  V  P  P  N  E  T  C  A  F  E  L  T  I  K  N  E  E  V  E  E  D  I  K  P  K  A  L  A  G  L  E

4201  AATGATGCCTTGGATCAGCTGCCAGAAGATTTGCAAAGAACTGTGGGTGTTCAGCAGAGAGCCGATGTTTTTGAAAATTTGGACAATTGGGAAGAGGATG  4300
       N  D  A  L  D  Q  L  P  E  D  L  Q  R  T  V  G  V  Q  Q  R  A  D  V  F  E  N  L  D  N  W  E  E  D

4301  GGCGTGATAGAGTTGGACGCCAGAAAGAAGCATCCAAAAAAGTTCTCAAGATTGAAGTCTGTGCCAAGTTTAAAGGTCCCGAAGAAGAGACATCAATGGC  4400
      G  R  D  R  V  G  R  Q  K  E  A  S  K  K  V  L  K  I  E  V  C  A  K  F  K  G  P  E  E  E  T  S  M  A

4401  CATGATAGATGTTGGAATCTTCACCGGCTACGCAGTCATCAAGGACGATCTTGAAAACCTATTAAATGAAGCCGCCTCTCCTATTCAACAATACGAGCTT  4500
        M  I  D  V  G  I  F  T  G  Y  A  V  I  K  D  D  L  E  N  L  L  N  E  A  A  S  P  I  Q  Q  Y  E  L

4501  TCCTCACGCGGCGTAGTATTCTATCTGGATACGATTTCTAGTAAACAAAAAACGTGCATTAAATTCCGAGCCAAGCAAACGTTTAGGGTTGGAGTTGCTC  4600
       S  S  R  G  V  V  F  Y  L  D  T  I  S  S  K  Q  K  T  C  I  K  F  R  A  K  Q  T  F  R  V  G  V  A

4601  AACGAGCATCCGTCGTTGTTTATGATTATTACAAACCAGAACTGACGTGCACCGCTTTTTATGCGCCAGGTGAAACAAGTGTCATTAAAAAAAGTTTATG  4700
      Q  R  A  S  V  V  V  Y  D  Y  Y  K  P  E  L  T  C  T  A  F  Y  A  P  G  E  T  S  V  I  K  K  S  L  C

4701  CTCTGGCAATAAGGCATGCATATGTCCAAGAGGTGGTTGTTCAGCTTGCAAGGCATTTGACGATGTTTTGAAGTTGTCCGGTAATATTTTCCAGAAAATG  4800
        S  G  N  K  A  C  I  C  P  R  G  G  C  S  A  C  K  A  F  D  D  V  L  K  L  S  G  N  I  F  Q  K  M
                                                                       C345C domain
4801  ACTCGTGAATTGCTACACCAGACCTGCAACATGCACGATTACGTTTGGGTGGCATCCGTCGTAAATATTCGTTTGGAAAACGAACTGAAATTCATCGATG  4900
       T  R  E  L  L  H  Q  T  C  N  M  H  D  Y  V  W  V  A  S  V  V  N  I  R  L  E  N  E  L  K  F  I  D

4901  TCGAAATAGACGAAGTGATCAAGCCTGGTGTTGAAGGCGAAGCTCAGCAATTGAGGCATGGAAGCGTTCGGACGTTGATCTCATACAATCACTGCACTTG  5000
      V  E  I  D  E  V  I  K  P  G  V  E  G  E  A  Q  Q  L  R  H  G  S  V  R  T  L  I  S  Y  N  H  C  T  C

5001  TCCAGATATGTTTATTGGCAAAAAGTACTTCATAATGGGTCAAGATGGTACAAAATGGGAGTTCGAAGGAGTTGCACAATACAGATATCTGTTAAACACG  5100
        P  D  M  F  I  G  K  K  Y  F  I  M  G  Q  D  G  T  K  W  E  F  E  G  V  A  Q  Y  R  Y  L  L  N  T

5101  AATGCACGGATCTTCGAATCAAAGACAATTAGAAGTACAAGTCCAAAGTACAAATTGATGATGAAAGCCTTCAACTGGCTGCAGAAGCGTTTCATCCGAT  5200
       N  A  R  I  F  E  S  K  T  I  R  S  T  S  P  K  Y  K  L  M  M  K  A  F  N  W  L  Q  K  R  F  I  R

5201  ACAAAAGAGGATGTGAATAGTAGAAATTCATTTCACAGGCGTATATAGTTTTTACTCGAGATGTTCAGCAATTAATTATTCAACTGATATATGATGTTTT  5300
      Y  K  R  G  C  E  *

5301  CTTCAATAAAAGTTTTCTTGCACAAAAAAAAAA  5333
```

AmspA2M-1

```
   1  TCAGTAGAGTGCATGGTTGCAATTATGAAAATTATTTTTCTGGCCGCGACTTGTCTCGCATTTACTTGTCTTGTAAATGCCGAAAGGGGCTACTTATTAA  100
                   M  V  A  I  M  K  I  I  F  L  A  A  T  C  L  A  F  T  C  L  V  N  A  E  R  G  Y  L  L

 101  CAGTTCCCAAGGTCCTTGACTCAGGGACATCAGAACGCCTGTGCTTGACACTTACAGATGTTAAAGGACCTGGAACAGTCACTATACGCTTACTGCAAGA  200
      T  V  P  K  V  L  D  S  G  T  S  E  R  L  C  L  T  L  T  D  V  K  G  P  G  T  V  T  I  R  L  L  Q  D

 201  CAGGATAGACACCGCAATCGCCGAAAAGTCCATCACTTACCCTCGTGACGATTCATGTTTCTTTCTGGAAATTCCCAATGTCAAGACGCGCAATGCACGA  300
        R  I  D  T  A  I  A  E  K  S  I  T  Y  P  R  D  D  S  C  F  F  L  E  I  P  N  V  K  T  R  N  A  R

 301  ATTGAAATCAAAGGAGACTTTCAATCAGCTCGGTACAGCTTCGAATCTGAAACTACCGTTTCAATTCAATCTCGATCAACGTTAGCATTCATTCAAACCG  400
       I  E  I  K  G  D  F  Q  S  A  R  Y  S  F  E  S  E  T  T  V  S  I  Q  S  R  S  T  L  A  F  I  Q  T
```

Supplementary Fig. S1 (continued)

```
 401   ACAAAGCTATTTACAAACCTGGCCAAAAAGTTCAATTCAGAATTCTCACTGTCAATCATTTATTGGAACCAATCTCGACCGAAATTTCGAAGGTCTATAT  500
       D  K  A  I  Y  K  P  G  Q  K  V  Q  F  R  I  L  T  V  N  H  L  L  E  P  I  S  T  E  I  S  K  V  Y  I

 501   CAACAATCCCACTGGAGTTCGCATTGCTCAATGGTTGGGCGTCAAGAACGAGAATGGTCTGATACAATTGGATATGCAGTTGTCGGATGAACCAACTTTG  600
         N  N  P  T  G  V  R  I  A  Q  W  L  G  V  K  N  E  N  G  L  I  Q  L  D  M  Q  L  S  D  E  P  T  L

 601   GGCAAATGGAAAATATACGTCCAACTTCAGGGAGCGTCTAAGGTTCAAGAATTTGAAGTAGATGAATATGTTTTACCAAAGTTTGAAGTAAGCGTTACTC  700
        G  K  W  K  I  Y  V  Q  L  Q  G  A  S  K  V  Q  E  F  E  V  D  E  Y  V  L  P  K  F  E  V  S  V  T

 701   CTCCAAAATACCTGCTTGCCAATTCAGAAACTGCTACTTGGAAGATCTGTGCAAGGTACACTTATGGCAAAGATGTTCAAGGCACACTGACTGCTGAAAT  800
       P  P  K  Y  L  L  A  N  S  E  T  A  T  W  K  I  C  A  R  Y  T  Y  G  K  D  V  Q  G  T  L  T  A  E  M

 801   GAATTACAGCGGCTATGCTTGGGAGAGACGCGGCCCACCATTATTATCATTAACAACTAAAAAGATTAATGGTTGCTATGATTGGATTGTCCATGTACGC  900
         N  Y  S  G  Y  A  W  E  R  R  G  P  P  L  L  S  L  T  T  K  K  I  N  G  C  Y  D  W  I  V  H  V  R
                                                          ◄──────────────────────
 901   GACACGAAATGGAACATGAGCTCAGTGAGTCATAAAAGCATTATCGTCAAGGCCGATGTCAAAGAATCTGGAACAGACATTGAATTCTCCAGTTCATCAA  1000
        D  T  K  W  N  M  S  S  V  S  H  K  S  I  I  V  K  A  D  V  K  E  S  G  T  D  I  E  F  S  S  S  S

1001   CTACTTCTATAGGACACGATGTTATTAAAATTAATATGACGCCGAAGAATGGCGAAATATACTTCAGGCCAGGACTACCATACTTTGGAGAACTCAATGC  1100
       T  T  S  I  G  H  D  V  I  K  I  N  M  T  P  K  N  G  E  I  Y  F  R  P  G  L  P  Y  F  G  E  L  N  A

1101   TAAGAAGCCAGATGGTCAACCGGCATCTGATGTCCTCATCCTCATCTGCTATAGAGTATCTAACACCAAAGAATGCAGAAATTTTACCAGTGATACCAAT  1200
         K  K  P  D  G  Q  P  A  S  D  V  L  I  L  I  C  Y  R  V  S  N  T  K  E  C  R  N  F  T  S  D  T  N

1201   GGCATTGTAAAATTCTCCATTCCGCCACAAAAACCAGAAGTTTTATCTATTTCCGTTGATGCAACAGCTCCAAATTACAAACCAGAGTACTACGACGGAA  1300
        G  I  V  K  F  S  I  P  P  Q  K  P  E  V  L  S  I  S  V  D  A  T  A  P  N  Y  K  P  E  Y  Y  D  G

1301   ATAAATACAGATCAAAAGTTCGCCAGCCACACACTCGGAAAACTTTGCGACCGTGGTTCTCGCCAAGTGGTAGCTTCATGGACGTGAAACCTGTGCATAA  1400
       N  K  Y  R  S  K  V  R  Q  P  H  T  R  K  T  L  R  P  W  F  S  P  S  G  S  F  M  D  V  K  P  V  H  K

1401   GGTCATTTCTTGTGATGAAGAAGTACCCTTGGATGTGATTTACACTACAAATGGAGACGATATTACTTTGAACTATCAGGTCATGGCACGTGGTAGAATC  1500
         V  I  S  C  D  E  E  V  P  L  D  V  I  Y  T  T  N  G  D  D  I  T  L  N  Y  Q  V  M  A  R  G  R  I

1501   GTTGATTATGGGATGAAATCTTACAAATTCAGTGTGAATGATTACAAAGAAGATTCTGATGTCATACGGAATAAAACCGAAGGGTCGAGCAGAACAAAGC  1600
        V  D  Y  G  M  K  S  Y  K  F  S  V  N  D  Y  K  E  D  S  D  V  I  R  N  K  T  E  G  S  S  R  T  K

1601   GAACTATCATTGCACCAGACATCGATAAAACTCCTAAATACCGACTTCCACGTCACATTGGCAAGTTCAGTCTCCCAATAAAAATCAAAGCTGAAATGGC  1700
       R  T  I  I  A  P  D  I  D  K  T  P  K  Y  R  L  P  R  H  I  G  K  F  S  L  P  I  K  I  K  A  E  M  A

1701   GCCGATCGCTCGAGTACTTGTCTATTACACCCGGTCCGATGGCGAGGTCATCGCTGCATTTGCTACTTTGGACGTCATGCCGTGCTTCCTCAACAAAGCT  1800
         P  I  A  R  V  L  V  Y  Y  T  R  S  D  G  E  V  I  A  A  F  A  T  L  D  V  M  P  C  F  L  N  K  A

1801   GCTTTGACATTCGAGAAAAAATCAATCAAACCTGGAATGTCTGCAAAGTATAAAATCGCAGCAACTGCAAAATCTTTGTGTGCCGTTGGTGTAGTCGACA  1900
        A  L  T  F  E  K  K  S  I  K  P  G  M  S  A  K  Y  K  I  A  A  T  A  K  S  L  C  A  V  G  V  V  D

1901   AAAGTACCCATCTTTTGAAAACCAGTAATCAAATAACAGGAGATAAGGTTTTCAAGATTTTGAAGGGATTTGATTCAAACAGAAACACACGGCCAAAATT  2000
       K  S  T  H  L  L  K  T  S  N  Q  I  T  G  D  K  V  F  K  I  L  K  G  F  D  S  N  R  N  T  R  P  K  F

2001   TATCAACACGGATGAGTACTGTCGAAGCAAATCAAAAGAACCAAGACCCACATTCAATCCCTTACCTCCCTTCCCCATACTGCAGCGTCGTTCTAAACGC  2100
         I  N  T  D  E  Y  C  R  S  K  S  K  E  P  R  P  T  F  N  P  L  P  P  F  P  I  L  Q  R  R  S  K  R
                                                                                            β-α cleabage site
2101   CGATCACAGCCATTCTGGAGACCTCGACCTACCACAGAATATGTTGACGCATCCATGGCATTTGAGGCTTCTGGATTGACACTGTTGACAACCAATGACG  2200
        R  S  Q  P  F  W  R  P  R  P  T  T  E  Y  V  D  A  S  M  A  F  E  A  S  G  L  T  L  L  T  T  N  D

2201   TAAACACGAGACCCTGCAGGCACACGAGACACTATCCCGAAATATTTTTTGAGAGTGTTTCAATAAGACGTAAAAATGTTGCATTTAGTGCCACTAATGT  2300
       V  N  T  R  P  C  R  H  T  R  H  Y  P  E  I  F  F  E  S  V  S  I  R  R  K  N  V  A  F  S  A  T  N  V
                                                    Bait region
2301   ACTTCGAGGGCAATCTGGCCCGAGACCAACGCGAAGACCAACACCGAAGCCAGCACCGCCACAATCTGAAAGTTCGACGACTTTCAGTGATGACGATCTT  2400
         L  R  G  Q  S  G  P  R  P  T  R  R  P  T  P  K  P  A  P  P  Q  S  E  S  S  T  T  F  S  D  D  D  L

2401   ACAGATGATGCCGTAAACGAGGCAGTCGAAGTACGTACATACTTCCCAGAAACATGGCTATGGGATCTTGAAGTCGTTGGTGATGATGGATACTTAAATA  2500
        T  D  D  A  V  N  E  A  V  E  V  R  T  Y  F  P  E  T  W  L  W  D  L  E  V  V  G  D  D  G  Y  L  N
                                                        ◄──────────────────────
2501   AAGAAGCAGAAATTCCACACACGATTACAGAATGGGTAGGAAGTATGTTCTGCACATCGAAGTCAAAAGGTCTTGGTATTTCAGCACCGTCTGCCATCAG  2600
       K  E  A  E  I  P  H  T  I  T  E  W  V  G  S  M  F  C  T  S  K  S  K  G  L  G  I  S  A  P  S  A  I  R

2601   AGCTTTTCAACCGTTCTTCGTGTCCTATTCACTGCCATATTCCGTAATTCGAAATGAAAAAATTCCAATCACTGTTTCCGTCTTCAACTACCTATCAGAA  2700
         A  F  Q  P  F  F  V  S  Y  S  L  P  Y  S  V  I  R  N  E  K  I  P  I  T  V  S  V  F  N  Y  L  S  E

2701   TGTCTTCCGATCAAACTCAAATTGGAGAAGAATGATGGTTTCACTTTATTGAGCAACAGTTACACCCACTCCTTGTGCGTGTGCGGAGGCAAGCCTGCCA  2800
        C  L  P  I  K  L  K  L  E  K  N  D  G  F  T  L  L  S  N  S  Y  T  H  S  L  C  V  C  G  G  K  P  A

2801   TTCACAACTTCAGAATTCTGCCAACTAGTCTCGGAGAAGTCAACTTAACAGTCCACAGTTTTTCATTCGACGATAAAAATTATGAAGTTTGCTCACAAGA  2900
       I  H  N  F  R  I  L  P  T  S  L  G  E  V  N  L  T  V  H  S  F  S  F  D  D  K  N  Y  E  V  C  S  Q  D

2901   CAAAGTGTCTTCCAAGGTCGTTGCTCGTGATGCCATAACCAAGCCATTATTGGTTGAACCTGAAGGTTTTCCACAAGTAATCACAGAGAGTCTCTTGTTC  3000
         K  V  S  S  K  V  V  A  R  D  A  I  T  K  P  L  L  V  E  P  E  G  F  P  Q  V  I  T  E  S  L  L  F

3001   TGCCCATCAGAACATGGAAATAGATTCAAGAAAGATTTTGAACTGTCACTTCCTGATGATTTAGTAGAAGGATCAGCTCGGGCATTCTTGTATGTTTCAG  3100
        C  P  S  E  H  G  N  R  F  K  K  D  F  E  L  S  L  P  D  D  L  V  E  G  S  A  R  A  F  L  Y  V  S

3101   GCGATATCATGGGTCCATCTTTAAGTGGACTGGAGAATCTTGTGACGAGGCCAACAGGCTGTGGCGAACAAAATATGATCAAGTTTGCTCCAATTATTTT  3200
       G  D  I  M  G  P  S  L  S  G  L  E  N  L  V  T  R  P  T  G  C  G  F  Q  N  M  I  K  F  A  P  I  I  F
            ────────────────────►                               Thioester site               ◄────────────
3201   CGTTACGCAATATCTGGAAGGCGTGGGCAGCCTCACACCAGAAACGAAAAAGAAATCCCTTGAATTCATGAAAGTTGGATATCAAAGAGAGTTGACGTAC  3300
         V  T  Q  Y  L  E  G  V  G  S  L  T  P  E  T  K  K  K  S  L  E  F  M  K  V  G  Y  Q  R  E  L  T  Y
       ──────
3301   AGACACAATGATGGTTCATACAGTGCTTTTGGTCAAAGTGATGCTGAAGGAAGCTCGTGGTTGACTGCTTTTGTGGTAAAATCATTTGCACAGGCACGCG  3400
        R  H  N  D  G  S  Y  S  A  F  G  Q  S  D  A  E  G  S  S  W  L  T  A  F  V  V  K  S  F  A  Q  A  R
```

Supplementary Fig. S1 (continued)

```
3401  ATTTAATCGACATTGACGCCATTGACCTAAAGAAGAGCACTGATTGGCTACTGAGTAAGCAACAATCCGATGGATGCTTCCCGTTCATAGGAATGGTGCA  3500
      D  L  I  D  I  D  A  I  D  L  K  K  S  T  D  W  L  L  S  K  Q  Q  S  D  G  C  F  P  F  I  G  M  V  H

3501  CCATTCAGACATGAAGGGTGGAGTGGGCAAAAGTGTGCCAACGGCTTTGACAGCATACACTGTGATTTCCATTCTTGAATCAGAGACTCCAATTAGTCAA  3600
        H  S  D  M  K  G  G  V  G  K  S  V  P  T  A  L  T  A  Y  T  V  I  S  I  L  E  S  E  T  P  I  S  Q
Catalytic histidine
3601  CAGAAACTGAACAAAGCCTTTGAATGTATTTCTCAGCAAACGGATCCAAATTCATACACTCTTGCTTTGTCTGCATACGCCTACGCCCTTGCTGGGCGTT  3700
       Q  K  L  N  K  A  F  E  C  I  S  Q  Q  T  D  P  N  S  Y  T  L  A  L  S  A  Y  A  Y  A  L  A  G  R

3701  CTGATATGTCCAATAAACTCATTGACAAGCTATTCTCAAGAGCAATCGTAGAAGGCACGGATGTTCACTGGGAAGCTGCTTCCAAGTCTATTAGCGTAGA  3800
      S  D  M  S  N  K  L  I  D  K  L  F  S  R  A  I  V  E  G  T  D  V  H  W  E  A  A  S  K  S  I  S  V  E

3801  GCTCGGTTCATACGTGATTTTAACACTGATGAAAATGGGAGGAACAGCGAACCAAGGCAAGGCTCTGGGCGTTGTCAGGTGGATTGCCCGCCAGAGAAAT  3900
        L  G  S  Y  V  I  L  T  L  M  K  M  G  G  T  A  N  Q  G  K  A  L  G  V  V  R  W  I  A  R  Q  R  N

3901  TCAAATGGCGGATTTGTTTCTACTCAGGATACCGTGATAGCTCTGCAAGCTTTTGCTAAGGTTGCAGCCATTCTGAACAGGAACAAACAAGATTTAAAAG  4000
       S  N  G  G  F  V  S  T  Q  D  T  V  I  A  L  Q  A  F  A  K  V  A  A  I  L  N  R  N  K  Q  D  L  K

4001  TGACTGCCGAAGGTAATGGATTCTATAAAGAATATGCAGTCAACTCCACGAATCGCCTACTGATGCAAATGCACAAAATTGAAGAATTGCCGAACATTGT  4100
      V  T  A  E  G  N  G  F  Y  K  E  Y  A  V  N  S  T  N  R  L  L  M  Q  M  H  K  I  E  E  L  P  N  I  V

4101  TGATGTATTTGCTACTGGTGACGGATGCGGTTTAATCCAGACGACACTGAAGTACAATAAAAAGAACGTGAATGCAAGTGATGCCTTTGATCTTACCGTT  4200
        D  V  F  A  T  G  D  G  C  G  L  I  Q  T  T  L  K  Y  N  K  K  N  V  N  A  S  D  A  F  D  L  T  V

4201  GTCGGAAACAATGATAGAAGTAATTGCAAAAAACGTGAGCTCAGGATATGTACCAGGTACAAGCTTTTAAATGAACGATCAAACATGGCCATAGTTACTG  4300
       V  G  N  N  D  R  S  N  C  K  K  R  E  L  R  I  C  T  R  Y  K  L  L  N  E  R  S  N  M  A  I  V  T

4301  TAAAGATGGTATCCGGCTACATACCTATCAAGGATAATTTGAAACAATTGAAAAACGACAGGGCGTTGAATCTTAAACGTTATGAAGTTGATGCCAATTA  4400
      V  K  M  V  S  G  Y  I  P  I  K  D  N  L  K  Q  L  K  N  D  R  A  L  N  L  K  R  Y  E  V  D  A  N  Y

4401  CGTGAACTTATATTTCGATTACTTGAGCAACGATCAGACCTGTTTCAAGATGAATGTCGAGAAAGAAATTGACGTCGAAGATGCCAAACCAGCAACCATT  4500
        V  N  L  Y  F  D  Y  L  S  N  D  Q  T  C  F  K  M  N  V  E  K  E  I  D  V  E  D  A  K  P  A  T  I

4501  ACAGTCTCCGATTACTACACTCCAGAAAAGATACTTGAGAAATCATACACATTGCCAGCTGATCAGTGCTAGAAACATCTGCCACGCCATACTGAGCCAT  4600
       T  V  S  D  Y  Y  T  P  E  K  I  L  E  K  S  Y  T  L  P  A  D  Q  C  *

4601  AGAGCTTGAGATTTTACTTCTATTCTTCATTTTATCGGATTTTTAAATTCTACATTTTGTTTTATAAATCTTTTTTTGCTATTAAAAATGTTATAACAAA  4700
4701  AAAAAAAAAAAAAAAAAAAAAAAAAAAA  4728
```

AmspA2M-2

```
   1  ACAGAGGCTTTGATTTCAATTCGGAGTCAAACGTCTTGAGTTTTGTTTATATTTTTTTTTAAGTTCTTTTTTAGGCTTTTATCTCAGGTGGTTGTTTATA  100
 101  TCGTATTAACCATGAAGCTTATTCTTTTGGCTGTGGCATGCCTTGCCTTTTATTCCAAAGCAAACGCCGAGCGAGGTTACCTATTTACGGCACCAAAGGT  200
                  M  K  L  I  L  L  A  V  A  C  L  A  F  Y  S  K  A  N  A  E  R  G  Y  L  F  T  A  P  K  V

 201  ACTTGATGCAGGAACATCAGAACGTCTATGTTTGACACTCACAGATGTGAAGGGACCAGGTACAGTCACTGTAAGTTTACTGAAAGAAAAGATAGACGAA  300
        L  D  A  G  T  S  E  R  L  C  L  T  L  T  D  V  K  G  P  G  T  V  T  V  S  L  L  K  E  K  I  D  E

 301  GTCATCGCTGAGACATCAGTCAACTTCCCTCGTGATGATATGTGTTTCTTCATTGAAATTCGAAATACCAAGTTGACCCGCGGTAACCTTAAAATTCAAG  400
       V  I  A  E  T  S  V  N  F  P  R  D  D  M  C  F  F  I  E  I  R  N  T  K  L  T  R  G  N  L  K  I  Q

 401  GAAAGTTTGATTCAGATGATTACAGCTTTGAATCGGAATCACCTGTTTCAATAGCCAGCCGCTCCACATTGACATATGTACAAACAGACAAAGCCGTTTA  500
      G  K  F  D  S  D  D  Y  S  F  E  S  E  S  P  V  S  I  A  S  R  S  T  L  T  Y  V  Q  T  D  K  A  V  Y

 501  CAAGCCCGGACAAAAAGTTCAATTCCGGATTCTGACTGTGAATCATTTGTTGGAACCACTTCCGACTGAAATTTCTAAAGTTTACGTTCTAAACCCCAAC  600
        K  P  G  Q  K  V  Q  F  R  I  L  T  V  N  H  L  L  E  P  L  P  T  E  I  S  K  V  Y  V  L  N  P  N

 601  GGCATCCGTGTTGCTCAGTGGCTTGGAATAAAAAATGAAAATGGGCTTATTCAATTGGACTTGCAGTTGTCTGATGAATCTGCCAGGGGCCAGTGGAAAA  700
       G  I  R  V  A  Q  W  L  G  I  K  N  E  N  G  L  I  Q  L  D  L  Q  L  S  D  E  S  A  R  G  Q  W  K

 701  TAAATGCCCAGGTTAAAGGGTCGACAATATCCCAAGATTTTGAAGTTGATGAATATGTTTTGCCAAAGTTTGAAGTTACAATTACCCCACCAACGTTTCT  800
      I  N  A  Q  V  K  G  S  T  I  S  Q  D  F  E  V  D  E  Y  V  L  P  K  F  E  V  T  I  T  P  P  T  F  L

 801  CAGCTCCAATATGGATATCGCTACTTGGAAAATATGTGCTCGCTACACTTACGGTAAAGGTGTCCAAGGCACATTGAAGGCTAAATTGGAGTATGTCACC  900
        S  S  N  M  D  I  A  T  W  K  I  C  A  R  Y  T  Y  G  K  G  V  Q  G  T  L  K  A  K  L  E  Y  V  T

 901  TACAGCTATGAAAGAGACAGAGATTCGTTCCCTGAAACTAACTTGGAAGTAAAGATTAATGGGTGCCATGACCTCACGGTTTATGCAAAGACTATGAAAT  1000
       Y  S  Y  E  R  D  R  D  S  F  P  E  T  N  L  E  V  K  I  N  G  C  H  D  L  T  V  Y  A  K  T  M  K

1001  GGAACACAAGGAAAATGCATTATAAAAGTTTGAAATTGAACGCTGAAGTTGAGGAAGAAGGCACTGATATTAAATTTTCGAGTTCGTCATCAATGTCAAT  1100
      W  N  T  R  K  M  H  Y  K  S  L  K  L  N  A  E  V  E  E  E  G  T  D  I  K  F  S  S  S  S  S  M  S  I

1101  TAGTCATGAGCCCTTATTTTTGAATTTCAAAAATTGGGACAGACAAAAATATTTCCATCCTGGACTTCCATACCATGGACAACTCCACGTCACGTTACCC  1200
        S  H  E  P  L  F  L  N  F  K  N  W  D  R  Q  K  Y  F  H  P  G  L  P  Y  H  G  Q  L  H  V  T  L  P

1201  AATGGACAAAATGCTCCTGATGAACTCATCCAGGTGTGCTATGAAGCTATTGTAAGCTCATGTAGGAACTTCACAAGCGATAGCCATGGAATCATCAATT  1300
       N  G  Q  N  A  P  D  E  L  I  Q  V  C  Y  E  A  I  V  S  S  C  R  N  F  T  S  D  S  H  G  I  I  N

1301  TCACTGTGCCTCCACAAGGAGCCGAAGCTACATATGTGAAAATCAAAGCACTTGCTCCTAATCATCCCTCTGAGTTTTATCCGAATGATAGATATGCGCA  1400
      F  T  V  P  P  Q  G  A  E  A  T  Y  V  K  I  K  A  L  A  P  N  H  P  S  E  F  Y  P  N  D  R  Y  A  Q

1401  AATAAAAATACGTCAACCATCTGCTGAAAATTCATTAACTCCATGGTATTCCCCTAGTGGAAGCTTCTTGAAGTTAAAACCTGTACTTGGCACTTCATCG  1500
        I  K  I  R  Q  P  S  A  E  N  S  L  T  P  W  Y  S  P  S  G  S  F  L  K  L  K  P  V  L  G  T  S  S

1501  TGCAATGAAGAAGTACCTTTGGATGTGTTTTATACTACCAGTGGGGAAGACATTCAACTGCACTATCAGGTTATGTCAAGAGGCAGAATCGTGACCCACG  1600
       C  N  E  E  V  P  L  D  V  F  Y  T  T  S  G  E  D  I  Q  L  H  Y  Q  V  M  S  R  G  R  I  V  T  H

1601  GCATGAAATCCTACAAATTCAATGAAAACGATTACCAAGATGACTCTTATACAATTCAAAATGAAAATGTCAATGAACGACAGAAACGTTTCTACAGTGA  1700
      G  M  K  S  Y  K  F  N  E  N  D  Y  Q  D  D  S  Y  T  I  Q  N  E  N  V  N  E  R  Q  K  R  F  Y  S  E
```

Supplementary Fig. S1 (continued)

```
1701  AAAAGTCAACTACACCCTTCCCAAGCACATTGGAAAGTTCAGTCTTCCAATCAAAATAAGCGCACACATGTCTCCAGTAGCCAGAGTTCTCATTTATTAC  1800
        K  V  N  Y  T  L  P  K  H  I  G  K  F  S  L  P  I  K  I  S  A  H  M  S  P  V  A  R  V  L  I  Y  Y

1801  ATCAGAACCGACGGTGAAGTTGTTGCCGCATCTACAAGTTTGGATGTCATGCCATGCTTTGCCAACAAGGCCTCTTTCACATTTGAGAAAGATTCTGTCA  1900
       I  R  T  D  G  E  V  V  A  A  S  T  S  L  D  V  M  P  C  F  A  N  K  A  S  F  T  F  E  K  D  S  V

1901  AACCTGGAGAGGCGGCTAAATACAAAATTTCTGCTGCACCAAAATCCTTGTGTGCCGTTGGTGTCGTCGACAAGAGTGTGCATCTTTTGAAGTCCGATAA  2000
      K  P  G  E  A  A  K  Y  K  I  S  A  A  P  K  S  L  C  A  V  G  V  V  D  K  S  V  H  L  L  K  S  D  N

2001  TCAAATTACAAGCGAAAAAATTTTCAAGGTTTTGAAAAGCTTTGATACTGGCAGATACACTTACCCAACATTGATCGATGATAGCAAGTACTGTAAAGAA  2100
        Q  I  T  S  E  K  I  F  K  V  L  K  S  F  D  T  G  R  Y  T  Y  P  T  L  I  D  D  S  K  Y  C  K  E

2101  AGACTTGAAGGAAAAATTCCAACTGAATCTCCATTTGACTCAACAACCCAGCGTCCAGAGCCGATAAATCCATTTGGTTTCCAAGGCCTTAGACTCGATG  2200
       R  L  E  G  K  I  P  T  E  S  P  F  D  S  T  T  Q  R  P  E  P  I  N  P  F  G  F  Q  G  L  R  L  D

2201  AGCTCGTTTCATCAACCTCATCTGAGGCTCCAACAACTGAAGCTGAAGAACCTGGTTCAGAAACTACATCTGTGGCCCCAACAACTGAAGAGCCAACTTC  2300
      E  L  V  S  S  T  S  S  E  A  P  T  T  E  A  E  E  P  G  S  E  T  T  S  V  A  P  T  T  E  E  P  T  S

2301  AACAGAATTTGTAGACAAGCAGCCTGAAGGAGCACCACCTTCTGTATTTAGAAAGAAACGATCAAGCCCACCACGATTTTACAGTGACAGTTACTCATCA  2400
        T  E  F  V  D  K  Q  P  E  G  A  P  P  S  V  F  R  K  K  R  S  S  P  P  R  F  Y  S  D  S  Y  S  S
                                                    β-α cleabage site
2401  AGTTATGTTGATGCATTGTTGGCTTTTGAGAGTTCTGGAGTTTTGGCGATGTCTGACCGCAGTTTGGAATCTCGTCTATGCAGCTTTTTCCAACGACGAC  2500
       S  Y  V  D  A  L  L  A  F  E  S  S  G  V  L  A  M  S  D  R  S  L  E  S  R  L  C  S  F  F  Q  R  R

2501  ATTTGGTATATGCATCTCCTGGTTTTGGAGGAAACGGGGTTCAAAAATTTGGAATCCATGATGATGAAAAAGGCGCAGGCGTTCCTTCTGCAGCTAGCGG  2600
      H  L  V  Y  A  S  P  G  F  G  G  N  G  V  Q  K  F  G  I  H  D  D  E  K  G  A  G  V  P  S  A  A  S  G
                                                   Bait region
2601  CTTTGGTTCAGGAGGAGGAGGAGGTGTAAATGAGGCTGTTGAAGTGCGTACCTACTTCCCAGAAACATGGTTGTGGGATCTTGAAGTTGTAGGCGATGCC  2700
        F  G  S  G  G  G  G  G  V  N  E  A  V  E  V  R  T  Y  F  P  E  T  W  L  W  D  L  E  V  V  G  D  A

2701  GGATACACTGACAAAGAAGCAGAAATACCTCACACAATTACCGAATGGGTTGGAAGCATGTTCTGTACATCCAAAACTAATGGTCTCGGAATTTCATCAC  2800
       G  Y  T  D  K  E  A  E  I  P  H  T  I  T  E  W  V  G  S  M  F  C  T  S  K  T  N  G  L  G  I  S  S

2801  CAACAGCAATTAAAGCTTTCCAACCATTCTTTGTGTCCTATGCTTTGCCCTATTCCGTCGTACGTAAAGAAAAAGTTCCAATCATCGTTTCTGTCTTCAA  2900
      P  T  A  I  K  A  F  Q  P  F  F  V  S  Y  A  L  P  Y  S  V  V  R  K  E  K  V  P  I  I  V  S  V  F  N

2901  CTATCTTTCAGAATGTCTTCCAATTCAGCTCAAATTGGAAAAAAGCGATGAATTCACTCTCCTGAGCGACAGTTACACTCACAGAATGTGTGTCTGTGGA  3000
        Y  L  S  E  C  L  P  I  Q  L  K  L  E  K  S  D  E  F  T  L  L  S  D  S  Y  T  H  R  M  C  V  C  G

3001  GGTCAACCTGCAACACATCGCTTCAGAATTCTGCCAACAGGCCTAGGAGAAGTCAACCTGACTGTTTACAGCCATTCATTCAACGATGCAAATAATGAAG  3100
       G  Q  P  A  T  H  R  F  R  I  L  P  T  G  L  G  E  V  N  L  T  V  Y  S  H  S  F  N  D  A  N  N  E

3101  TTTGCTCGAAAGACAAAAAAGCATCTACACTGGAGGCTCGTGATGCTATCACCAAACCATTGTTAGTCGAGCCTGAAGGTTTCCCACAGGAGTCCACCGA  3200
      V  C  S  K  D  K  K  A  S  T  L  E  A  R  D  A  I  T  K  P  L  L  V  E  P  E  G  F  P  Q  E  S  T  E

3201  AAGCACCTTGTTTTGTCCATCCGAATATCAAAATGGATTCAAGAAATCTTTTGAGTTGATGCTTCCTGACGATTTGGTCGAAGGCTCGGCAAGGGCATTC  3300
        S  T  L  F  C  P  S  E  Y  Q  N  G  F  K  K  S  F  E  L  M  L  P  D  D  L  V  E  G  S  A  R  A  F

3301  CTGTCCGTTAGCGGTGATATTATGGGTCCGTCACTGAGTGGATTAGAGAAGCTTGTGGCAAGGCCAACTGGTTGTGGTGAACAAAATATGATCAGGTTTG  3400
       L  S  V  S  G  D  I  M  G  P  S  L  S  G  L  E  K  L  V  A  R  P  T  G  C  G  E  Q  N  M  I  R  F
                        -------------------->                               Thioester site
3401  CTCCAAACATCTTCGTCATGCAATACCTTCAAGGTACGAGCAGTCTCACCCCAGAAATTGAAAAGAAAGCTCTTGACTTCATGAGAATTGGATATCAAAG  3500
      A  P  N  I  F  V  M  Q  Y  L  Q  G  T  S  S  L  T  P  E  I  E  K  K  A  L  D  F  M  R  I  G  Y  Q  R
                      <-----------------------
3501  AGAACTGAACTACAGGCATGACGATGGCTCTTACAGCGCCTTTGGTAAAAGTGATGCTGAGGGAAGTTCGTGGCTGACAGCATTTGTTGTGAAATCTTTT  3600
        E  L  N  Y  R  H  D  D  G  S  Y  S  A  F  G  K  S  D  A  E  G  S  S  W  L  T  A  F  V  V  K  S  F

3601  GCCCAAGCACGCCAGTTCATCGACATCGATCCCGTTGATTTGAAGAAGAGCACCGATTGGTTGCTGAGTAAGCAACAAGCCGATGGCTGCTTCCCATTCA  3700
       A  Q  A  R  Q  F  I  D  I  D  P  V  D  L  K  K  S  T  D  W  L  L  S  K  Q  Q  A  D  G  C  F  P  F

3701  TTGGCATGGTTCATCATCAAGACATGAAGGGTGGAGTAGGAAAAAGTGTCCCAACTGCATTGACAGCATATACTGTTATATCTCTTCTGGAAGCTGAAAC  3800
      I  G  M  V  H  H  Q  D  M  K  G  G  V  G  K  S  V  P  T  A  L  T  A  Y  T  V  I  S  L  L  E  A  E  T
            Catalytic histidine
3801  TCCAATCAGTCAAGACAAATTGGACAAAGCCTTTGAATGCATCTCAAAACAAACTGATCCCAATTCTTATACCATGGCCTTAACCGCCTATGCCTACGCC  3900
        P  I  S  Q  D  K  L  D  K  A  F  E  C  I  S  K  Q  T  D  P  N  S  Y  T  M  A  L  T  A  Y  A  Y  A

3901  CTTGCCGGGCGCTACGAGCTCGCTAACAAAATAATTGACTCCTTGTTCTCACGAGCCACCATTCAAGGCACAGATGTTTACTGGTCAACCTCTTCAAAGT  4000
       L  A  G  R  Y  E  L  A  N  K  I  I  D  S  L  F  S  R  A  T  I  Q  G  T  D  V  Y  W  S  T  S  S  K

4001  CTATCAGCGTTGAGCTTGGTTCTTACGTCATTCTATCACTCATGAAAATGGGAGGAGCTGCTAACCAAGCTAAAGCTCTGAGCATCGTCAGGTGGATTGC  4100
      S  I  S  V  E  L  G  S  Y  V  I  L  S  L  M  K  M  G  G  A  A  N  Q  A  K  A  L  S  I  V  R  W  I  A

4101  TCGCCAGAGAAACGCCAATGGAGGATTTGTTTCAACGCAGGATACTGTGATAGCTCTTCAGGCTTTTGCGAAGTTTGCAGTCATCCACTCAAAGAATAAA  4200
        R  Q  R  N  A  N  G  G  F  V  S  T  Q  D  T  V  I  A  L  Q  A  F  A  K  F  A  V  I  H  S  K  N  K

4201  CAAGATTTAGAAGTAATTGCCGAGGGCAATAATTTTAACCAGAAGTATGCAATTAACTCAACAAATCGTTTGCTCATGCAAAAGGACAAGGTCGTAGAAT  4300
       Q  D  L  E  V  I  A  E  G  N  N  F  N  Q  K  Y  A  I  N  S  T  N  R  L  L  M  Q  K  D  K  V  V  E

4301  TGCCAAATATAATTGATGTTTCTGCTGTTGGTGATGGATGTGGCTTAATTCAGACCACTTTAAAATACAACAAGGATAACGTGAATGCGAGCGATGCTTT  4400
      L  P  N  I  I  D  V  S  A  V  G  D  G  C  G  L  I  Q  T  T  L  K  Y  N  K  D  N  V  N  A  S  D  A  L

4401  GGAACTTATTATTGTTGGCAAAGCGGACAGATGGAATTGCAAACGACCTCAACTTGACATCTGTGCAAGATACAAGATTCTCGGAGAAAAGTCCAATATG  4500
        E  L  I  I  V  G  K  A  D  R  W  N  C  K  R  P  Q  L  D  I  C  A  R  Y  K  I  L  G  E  K  S  N  M

4501  GCTGTGATTAGTGTAAAGATGATCTCCGGATATATTCCAGTCAAGTCCTTGTTGGCAGATTTGAAAGATGTCCCGGAATTGAATTTGAAACGTTATGAAG  4600
       A  V  I  S  V  K  M  I  S  G  Y  I  P  V  K  S  L  L  A  D  L  K  D  V  P  E  L  N  L  K  R  Y  E

4601  TTGATGCAAACTATGTTAACTTCTACTTTGACTACCTAAGTAATAAACAGACCTGTTTCGCACTGCATGTTGAAAAGGAAATCGATGTGGAAGATGCTAA  4700
      V  D  A  N  Y  V  N  F  Y  F  D  Y  L  S  N  K  Q  T  C  F  A  L  H  V  E  K  E  I  D  V  E  D  A  K
```

Supplementary Fig. S1 (continued)

```
4701  GCCAGCGATTGCTTCTGTCTACGATTACTACGTTACAGAGCTGAAGTTGGAGAAATCTTATTCATTGCCTGTTGTTGACTGCAAGAAGACACCTCATGTT  4800
        P  A  I  A  S  V  Y  D  Y  Y  V  T  E  L  K  L  E  K  S  Y  S  L  P  V  V  D  C  K  K  T  P  H  V

4801  GAACCTTTGGATGATGTACCCGGACTGGTTGAATTAACAACTGCTGCTGCTGTTGAATCAACAACAGTTGATGATGAACAATCAACAACGGTTGATGATG  4900
       E  P  L  D  D  V  P  G  L  V  E  L  T  T  A  A  A  V  E  S  T  T  V  D  D  E  Q  S  T  T  V  D  D

4901  AACAATCAACAACGGTTGATGCCGAACAATCAACAACAGCCGGTGCTCAGGAGTAATGAGTTGTTGTAAAATTAAGTTTTCATAATTTTTTTTTAATGCC  5000
      E  Q  S  T  T  V  D  A  E  Q  S  T  T  A  G  A  Q  E  *

5001  ACTTTTGCCTTTTTTTATTTTTATGGCAACATAAGCTTTTTGAACATGCAGCACGTTTCAAATGCATTCTTCGGTTATCAAAGAACTTATCGTTATTGTT  5100
5101  ATCATTTTTTTTTTTTATCAATTAAAACCAAAAAAAAATTTTATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  5175
```

AmspA2M-3

```
   1  GCTCTTCCGATCTGTTTGGAGAATTCTTATGACACTGGAGTACAGGAGCCTGGTCATATGATGAAGACGATGCATACTGTTTACCTGGCAGCTATATTCC  100
                                                                  M  M  K  T  M  H  T  V  Y  L  A  A  I  F

 101  TGGTATCCGTATGTCAAGTTGGCGCTGAACGGAGTTACTTGCTGACGGCACCTAAAGTACTTGATGCTGGATCTGAACGTCTGTGTTTAACACTTACAGA  200
      L  V  S  V  C  Q  V  G  A  E  R  S  Y  L  L  T  A  P  K  V  L  D  A  G  S  E  R  L  C  L  T  L  T  D

 201  CATTGAAGGACCTGGTACAGTCAATGTACGTTTGTTGCATGAAAACGACACCACACTCTCCGAGACGTCGATCACTTACCCTCGCCATGAACCTTGCTTT  300
        I  E  G  P  G  T  V  N  V  R  L  L  H  E  N  D  T  T  L  S  E  T  S  I  T  Y  P  R  H  E  P  C  F

 301  TTTCTGAAAATTCCCAACATCAAGTCCACAAATGCTTATATCGAGGTCAATGGGAGTTTTGAATCCGACCAGTACCACTTCGGATCGAAAACTGGCATTG  400
       F  L  K  I  P  N  I  K  S  T  N  A  Y  I  E  V  N  G  S  F  E  S  D  Q  Y  H  F  G  S  K  T  G  I

 401  TCATCGCTCCTCTGGCGCCGGTGACTGTCGTACAAACCGACAGAGCCGTTTACAAACCTGGACAAAAAGTTCAATTCCGGATTCTGAACGTCAACCACTT  500
      V  I  A  P  L  A  P  V  T  V  V  Q  T  D  R  A  V  Y  K  P  G  Q  K  V  Q  F  R  I  L  N  V  N  H  L

 501  GTTGGAGCCAGTTGCAAACGAAATAACAAAAGTCTACATCCATAATCCAAATGAAGTTCGCGTCGCCCAATGGTTGGGAATCCAGAATGAAAATGGTCTG  600
        L  E  P  V  A  N  E  I  T  K  V  Y  I  H  N  P  N  E  V  R  V  A  Q  W  L  G  I  Q  N  E  N  G  L

 601  GTGCAGCTAGATATGCAGTTGTCTGATGAGCCAACTTTGGGCCTCTGGAAGATATATGTTGAGATTCAAGGGCACAAACGGGAACAAGGTTTTGAAGTTT  700
       V  Q  L  D  M  Q  L  S  D  E  P  T  L  G  L  W  K  I  Y  V  E  I  Q  G  H  K  R  E  Q  G  F  E  V

 701  CCGAATACGTTTTGCCGAAGTTTGAAGTAACAGTGACTCCTCCAACTACTCTGTTTCCTGATTCAGATTCTGCAACTTGGAAGATCTGTGCAAAATACAC  800
      S  E  Y  V  L  P  K  F  E  V  T  V  T  P  P  T  T  L  F  P  D  S  D  S  A  T  W  K  I  C  A  K  Y  T

 801  ATATGGCAAAGGCGTTGAAGGTGTATTGACAGCTGAAATCGAAAATACTGCCACCAAGTATTGGCGAATACCGATGAAAAGGAACTTGGAAACTAAGATT  900
        Y  G  K  G  V  E  G  V  L  T  A  E  I  E  N  T  A  T  K  Y  W  R  I  P  M  K  R  N  L  E  T  K  I

 901  AATGAATGCTATAACTGGACTTTAAGAAAAGAAGATTCGTTCTGGAACGGAATAGACTTGGATTTTGGAGTTATTAACGTCTCTGCTAAGGTCAAAGAAT  1000
       N  E  C  Y  N  W  T  L  R  K  E  D  S  F  W  N  G  I  D  L  D  F  G  V  I  N  V  S  A  K  V  K  E

1001  CTGGAACTGATATCGAATTCTCCAGTTCGTCGTCCACTAAAGTCGAACGCTGGCCTTATACAATTGAGGGCAGGTCAAATGAAGATAGGTATTACTTTAG  1100
      S  G  T  D  I  E  F  S  S  S  S  S  T  K  V  E  R  W  P  Y  T  I  E  G  R  S  N  E  D  R  Y  Y  F  R

1101  ACCTGGACTTCCATACTTTGGAGAGTTTGTTGTGAAGAAACCAGACAAAGAGCCGGCAGCTGATGTACTCATACTAGTCTGCTACGAAGTAAACAATACC  1200
        P  G  L  P  Y  F  G  E  F  V  V  K  K  P  D  K  E  P  A  A  D  V  L  I  L  V  C  Y  E  V  N  N  T

1201  AAAGAATGTAGAAACTTTACAAGCGATGACAAAGGAATTATAAAATTCACAATTCCGCCACAAAAACCGGAAGTTGTAGCCATTTCCGTCGAAGCAACAC  1300
       K  E  C  R  N  F  T  S  D  D  K  G  I  I  K  F  T  I  P  P  Q  K  P  E  V  V  A  I  S  V  E  A  T

1301  TTGCACTCTTCGAATCAGAGCACTATAACAATCAATCGTATCAAGAAAAACTTTATCAACCGCAACAACAAATTCCTTTGGCACCATGGTACTCGCCAAG  1400
      L  A  L  F  E  S  E  H  Y  N  N  Q  S  Y  Q  E  K  L  Y  Q  P  Q  Q  Q  I  P  L  A  P  W  Y  S  P  S

1401  TGGTAGTTTCTTGGATGTGAAGCCTGTACTTGACATTCTTTCATGCGATACGGAAGTACCTTTGAACATCATGTATACTACTAATGGTGAAGACATCATT  1500
        G  S  F  L  D  V  K  P  V  L  D  I  L  S  C  D  T  E  V  P  L  N  I  M  Y  T  T  N  G  E  D  I  I

1501  TTGAATTATCAGGTCTTGTCACGTGGAAGAATCATTGATTATGGTAAGAAATCCTACAAGTTCAACATAGACGATTACAATGAAGATCATTCAGTCATAC  1600
       L  N  Y  Q  V  L  S  R  G  R  I  I  D  Y  G  K  K  S  Y  K  F  N  I  D  D  Y  N  E  D  H  S  V  I

1601  GAAACGAAACTGAAGATTCTGGCAAAGTTAAAAGATCCATCGATCCATCCATCAATGTGAATTACAGTCTTCCTCAGCACATCGGAAAATTCAGTCTTCC  1700
      R  N  E  T  E  D  S  G  K  V  K  R  S  I  D  P  S  I  N  V  N  Y  S  L  P  Q  H  I  G  K  F  S  L  P

1701  AATACAAATCAAAGCCGAAATGGCCCCAATAACCAGAGTGTTTGTCTACTACATTCGGCCTGATGGCGAAGTCGTTGCTTCATATACCACTTTGAAAGTG  1800
        I  Q  I  K  A  E  M  A  P  I  T  R  V  F  V  Y  Y  I  R  P  D  G  E  V  V  A  S  Y  T  T  L  K  V

1801  ATGCCATGTTTCGTAAACAAAGCTTCGTTTACATTTGAAAAAAAGTCCATCAAACCAGGAGTATCTGCCAAGTACAAAATCAGAGCTTCTCCAAAGTCAT  1900
       M  P  C  F  V  N  K  A  S  F  T  F  E  K  K  S  I  K  P  G  V  S  A  K  Y  K  I  R  A  S  P  K  S

1901  TATGTGCCATTGGCGTGGTCGATAAAAGCTCCAATCTTTTGAAAACCGGTCATCAAATCACAGCGGAAAGACTTTTCGAGATCATGAAAGCGTTTGATGT  2000
      L  C  A  I  G  V  V  D  K  S  S  N  L  L  K  T  G  H  Q  I  T  A  E  R  L  F  E  I  M  K  A  F  D  V

2001  GAACGTGTACAACCTGCCAGTAGTGGCCAACAATCAAGCGTACTGTCAAGAGAAGTATAAAGATTCCAATAGTACTCGTCCCCCGCCAATTCGTCCTCGT  2100
        N  V  Y  N  L  P  V  V  A  N  N  Q  A  Y  C  Q  E  K  Y  K  D  S  N  S  T  R  P  P  P  I  R  P  R

2101  ATAATTCGAACAATTATTCCTGCACCTCGACCAATTCCAATACCTCCACGACTTGTTCCACGACCAGTACCACCAAGTCAAACAGTTTTTCGTCCACAAT  2200
       I  I  R  T  I  I  P  A  P  R  P  I  P  I  P  P  R  L  V  P  R  P  V  P  P  S  Q  T  V  F  R  P  Q

2201  ATTCCTATTTTTGGCTGAATAGATATGATCAATCAAAGTTAAAACTTGTTGATGCATCTATGGCATTTGAGACTTCTGGACTGACATTCTTGACATCCAA  2300
      Y  S  Y  F  W  L  N  R  Y  D  Q  S  K  L  K  L  V  D  A  S  M  A  F  E  T  S  G  L  T  F  L  T  S  N

2301  TGATGTGAACACCAGGCCATGCAGCCCTGAAGATATTTATGATGGAAGGTACACATTTCCACCTCCACCTATTCGAAGACCTCAACCTGTTCGACGACCT  2400
        D  V  N  T  R  P  C  S  P  E  D  I  Y  D  G  R  Y  T  F  P  P  P  P  I  R  R  P  Q  P  V  R  R  P
                                                  Bait region
2401  AGACCTTCCCCACGGCCTATACCTGTTGCACCAGCTCGGCCTTTACCAAGACCTTTACCTGCTCCACAAGTTCGGCCTATTCCACGACCTCCACCTGTTC  2500
       R  P  S  P  R  P  I  P  V  A  P  A  R  P  L  P  R  P  L  P  A  P  Q  V  R  P  I  P  R  P  P  P  V
```

Supplementary Fig. S1 (continued)

```
2501  CAAAAGTTCAGCCTATTCCACGACCTCCACTTCGTCGCCCAGCTTCACCAGATGACGGCGTGAATGAAGCTGTCGAAGTACGTAATTACTTCCCAGAAAC  2600
      P  K  V  Q  P  I  P  R  P  P  L  R  R  P  A  S  P  D  D  G  V  N  E  A  V  E  V  R  N  Y  F  P  E  T

2601  GTGGCTGTGGGACCTTGAAGTTGTTGGCGATGATGGAATCACTAACAAAGAAGCAGAAATTCCACACACAATCACCGAATGGACTGGAAGCATGTTCTGC  2700
       W  L  W  D  L  E  V  V  G  D  D  G  I  T  N  K  E  A  E  I  P  H  T  I  T  E  W  T  G  S  M  F  C

2701  ACATCCAAAACCGACGGCCTTGGAATTTCACCATCTGCAGCAATCAAATCTTTCCAACCCTTTTTTGTGTCTTACACATTGCCTTATTCGGTGATACGCA  2800
       T  S  K  T  D  G  L  G  I  S  P  S  A  A  I  K  S  F  Q  P  F  F  V  S  Y  T  L  P  Y  S  V  I  R

2801  ATGAAAAAGTTCCGGTCATTGTCACTGTCTTCAACTATCTTCCAGAATGCCTTCCAATTGAGCTCAGATTGGAAGAAAGTGAAGACTTCGAATTATTGAG  2900
      N  E  K  V  P  V  I  V  T  V  F  N  Y  L  P  E  C  L  P  I  E  L  R  L  E  E  S  E  D  F  E  L  L  S

2901  CAACAATACCCACAGAATGTGCGTCTGCAGCGGTCCTGCCACTCATCGTTTTAGGATCCAACCAACAGACCTCGGAAAAATAAATCTAACAGTCCACAGC  3000
        N  N  T  H  R  M  C  V  C  S  G  P  A  T  H  R  F  R  I  Q  P  T  D  L  G  K  I  N  L  T  V  H  S

3001  GATTCATTTGTTGATGCAACCCATGAAGTGTGCCCAGAAGATGGAGGTGCGTCAACACTCGTGGCTCGTGATGCCATCACCAAACCGCTATTGGTCGAAG  3100
       D  S  F  V  D  A  T  H  E  V  C  P  E  D  G  G  A  S  T  L  V  A  R  D  A  I  T  K  P  L  L  V  E

3101  CTGAAGGCTTTCCACAAGAGTCAATTCAAAGCGTTCTCTTTTGTCCATCTGAACATCAAAATGGATTCAAGAAAGCGTTTGAACTGATGCTTCCGGACGA  3200
      A  E  G  F  P  Q  E  S  I  Q  S  V  L  F  C  P  S  E  H  Q  N  G  F  K  K  A  F  E  L  M  L  P  D  D

3201  TTTGGTACAAGGTTCTGCTCGGGCATTCCTGCATGTCACCGGCGATATTATGGGACCATCTTTGAGTGGTTTGGAGCGTCTTGTACGATTCCCAACGGGG  3300
        L  V  Q  G  S  A  R  A  F  L  H  V  T  G  D  I  M  G  P  S  L  S  G  L  E  R  L  V  R  F  P  T  G

3301  TGTGGCGAACAAAATATGGTCTTGTTCGCCCCAAATATCTTTGTCATACAATACTTACAAGGCATCGATTCTCTTACACCTGAAGTTGAAAACAAAGCCC  3400
       C  G  E  Q  N  M  V  L  F  A  P  N  I  F  V  I  Q  Y  L  Q  G  I  D  S  L  T  P  E  V  E  N  K  A
      Thioester site
3401  TTGGTTTCATGAGAGCCGGATATCAAAGAGAACTGAACTACAGGCACGGCGATGGTTCTTACAGTGCTTTTGGTGAAAGCGATCCAGAAGGAAGTTCATG  3500
      L  G  F  M  R  A  G  Y  Q  R  E  L  N  Y  R  H  G  D  G  S  Y  S  A  F  G  E  S  D  P  E  G  S  S  W

3501  GTTGACTGCTTTTGTGGTGAAATCGTTTGCCCAAGCACGGCATCTAATTGACATTGATCCAGTCGATCTGAAGAAGAGCGCTGATTGGCTACTGAGTAAG  3600
        L  T  A  F  V  V  K  S  F  A  Q  A  R  H  L  I  D  I  D  P  V  D  L  K  K  S  A  D  W  L  L  S  K

3601  CAGCAAGACGATGGTTGTTTCCCATTCATTGGAATGGTTCATCATCAAGCAATGAGGGGTGGAGTCGGAAGGAATGAACCAACAGCTTTGACTGCCTACA  3700
       Q  Q  D  D  G  C  F  P  F  I  G  M  V  H  H  Q  A  M  R  G  G  V  G  R  N  E  P  T  A  L  T  A  Y
                                        Catalytic histidine
3701  CTGTGATTGCCATTCTTGAATCAGAAACACCAATAAGTCAAGACAAACTCGACAAAGCCTTTGGATGTATTTCACGACGAACAAGCCCAGATTCGTATGC  3800
      T  V  I  A  I  L  E  S  E  T  P  I  S  Q  D  K  L  D  K  A  F  G  C  I  S  R  R  T  S  P  D  S  Y  A

3801  CCTTGCCTTGTCTGTTTATGCCTACACCCTTGCCGGACGTTACGAGATCGCCAACAAACTTCTCGATCAACTATATTCACATGCTACAACTGAAGGCGCA  3900
        L  A  L  S  V  Y  A  Y  T  L  A  G  R  Y  E  I  A  N  K  L  L  D  Q  L  Y  S  H  A  T  T  E  G  A

3901  GATGTCTACTGGGCAGCAGATTCAAAATCAATTAGTGTTGAGATCGGTTCTTACGTCATTTTGTCACTTATGAAATTGGGAGGAACTGCTAACCAGGCCA  4000
       D  V  Y  W  A  A  D  S  K  S  I  S  V  E  I  G  S  Y  V  I  L  S  L  M  K  L  G  G  T  A  N  Q  A

4001  AAGCAATGCAAATCGTCAAGTGGATCACCCGCCAGAGAAACGCCAATGGTGGATTTGTTTCAACACAGGATACCGTTATAGCTCTTCAGGCATTTGCAAA  4100
      K  A  M  Q  I  V  K  W  I  T  R  Q  R  N  A  N  G  G  F  V  S  T  Q  D  T  V  I  A  L  Q  A  F  A  K

4101  GTTTGCGGTCCATCTGAACAGAAACAAGCAAGACTTAGAAGTGGCCGTTGAAGGAAATGGATTAAATGGCAAATATGCAATCAATTCTACCAATCATCTT  4200
        F  A  V  H  L  N  R  N  K  Q  D  L  E  V  A  V  E  G  N  G  L  N  G  K  Y  A  I  N  S  T  N  H  L

4201  CTCATGCAAACGAACAAGATTGAGGAATTGCCAAATATCATTGATGTTGAGGCTGTTGGTGAAGGATGTGGTTTAATCCAGACTACTTTAAAATTCAACA  4300
       L  M  Q  T  N  K  I  E  E  L  P  N  I  I  D  V  E  A  V  G  E  G  C  G  L  I  Q  T  T  L  K  F  N

4301  AAAACATCGCCAATGCAAGTGAAGTGTTTGACCTGACCGTTAAAGGCAGATTCCACAGACGGGATTGCGACAAGGACAAGCACACGATCGACATTTGTGC  4400
      K  N  I  A  N  A  S  E  V  F  D  L  T  V  K  G  R  F  H  R  R  D  C  D  K  D  K  H  T  I  D  I  C  A

4401  AAAGTACAAGATTACAAATGAAAAATCCAACATGGCAGTCATCTCCTTGAAAATGATATCTGGTTACATTCCTGTCAAGAGAGTTTTGAACAATTTGAAA  4500
        K  Y  K  I  T  N  E  K  S  N  M  A  V  I  S  L  K  M  I  S  G  Y  I  P  V  K  R  V  L  N  N  L  K

4501  TTTGACGAGGAATTGAATTTGAAACGCTACGAAATTAGCAACAACTTCGTGAACTTCTATTTCGATCACTTGAACAACGATCAGATTTGTTTCTCCATTG  4600
       F  D  E  E  L  N  L  K  R  Y  E  I  S  N  N  F  V  N  F  Y  F  D  H  L  N  N  D  Q  I  C  F  S  I

4601  ATGTTGAAAAGGAAATTGAAGTTGAAGAAACCAAGCCAGCGACAGTTTCGGTCTACGATTACTACAATTCAGAACTGAAACTGGATAAATCATACGAATT  4700
      D  V  E  K  E  I  E  V  E  E  T  K  P  A  T  V  S  V  Y  D  Y  Y  N  S  E  L  K  L  D  K  S  Y  E  L

4701  GCCTTCTACGTGCTGAGATAATATTACTGCATTGAATTTGCCGTTTGTCAATTGCAATAGTCAGCCAATCGTATAACTTTCAAATAAAAAAAAAATGCTA  4800
        P  S  T  C  *

4801  AATGATAGTAAATTTATGGTACGTAAGCGCAAAGAGAGGTTCAAAATGGTGATTAATGCAAGACATGTTCGAGACATTCAACTTTGCTGTATTTTGTTGT  4900
4901  ACCAGACACCCAGAAATTACAACGCCATTTTCCTCTTATTTTGAACTGAGATACTGGAGTATTCTTCATACAATATTCGCCACACTTTGGTTAAAAAATT  5000
5001  AACAAATAAACATTTAAAAAAATTTTTAATAGTAATTTTTAATGCTTTAGCATAATTATCGCGCCTGCTTTAGCAATT  5078
```

AmspA2M-4

```
   1  CATATGAGGAGGACTGACAAATTTCAGTTTCAAGTCAAGTAAGCCAGGGTGAATACAGATATTTGGTTCACCATGGTTGCCACCATGAAGCTACTTTTTG  100
                                                                                               M  K  L  L  F

 101  TGACTTCGATTTTCATCGCCTTCTCTTTTCCTATAAATGCAGAGCGAGGTTATTTGCTAACTGCACCAAAGGAATTTCATGCAGGATCATCGGAGCGCCT  200
      V  T  S  I  F  I  A  F  S  F  P  I  N  A  E  R  G  Y  L  L  T  A  P  K  E  F  H  A  G  S  S  E  R  L

 201  GTGTTTGACACTCACGGACGTCAAAGAGCGTGGCCAAGTCAGGGTACGTTTATTACAAGAAAGGAAAGATATTGCCATAGCCGAAAAATTGATCAATTAT  300
        C  L  T  L  T  D  V  K  E  R  G  Q  V  R  V  R  L  L  Q  E  R  K  D  I  A  I  A  E  K  L  I  N  Y

 301  CCCTCTGAAGACCCTTGCTTTCTTCTGGAAATTCCCAATACCAACTTCATCAAAGGTCGATTTGAAATCAAAGGACGGTTCGGTTCAGTTCAATACGATT  400
       P  S  E  D  P  C  F  L  L  E  I  P  N  T  N  F  I  K  G  R  F  E  I  K  G  R  F  G  S  V  Q  Y  D

 401  TTGATTCAGATGATGACTACCTGGACTACAGTTTTGATTCAGATCAATACACTTTCGATTCTGAAACTAAAGTTTCCATCAAATCCAGATCAACTTTAAC  500
      F  D  S  D  D  D  Y  L  D  Y  S  F  D  S  D  Q  Y  T  F  D  S  E  T  K  V  S  I  K  S  R  S  T  L  T
```

Supplementary Fig. S1 (continued)

```
 501  ATTCGTTCAAACCGACAAAGCAGTTTACAAGCCTGGACAAAAAGTCCAATTTAGGATTCTTACAGTTAATCATGTATTGGAGCCTGACTCAACGGAAATA  600
        F  V  Q  T  D  K  A  V  Y  K  P  G  Q  K  V  Q  F  R  I  L  T  V  N  H  V  L  E  P  D  S  T  E  I

 601  TCGAAGATTTTCATCGAAAACCCCAAAGGAATTCGTATTGCTCAATGGCTTGCAGTCAAAAATGAGAATGGCCTCATCCAATTGGATATGCAGTTATCTG  700
       S  K  I  F  I  E  N  P  K  G  I  R  I  A  Q  W  L  A  V  K  N  E  N  G  L  I  Q  L  D  M  Q  L  S

 701  ATGAAGCTCCAAAGGGTACGTGGAAAATAAGTGCTCAAGTTCAGGGCTTAACTAAGGAGCAAAGCTTTGAAGTGGATGAATATGTTTTGCCAAAGTTTGA  800
      D  E  A  P  K  G  T  W  K  I  S  A  Q  V  Q  G  L  T  K  E  Q  S  F  E  V  D  E  Y  V  L  P  K  F  E

 801  AGTGAAAGTCGTCCCTCCAAGATTTCTACTTCCAAACATGGACTCCGCAACGTGGAAGATCTGTGCTAAGTATACATATGGCAAAAATGTTCAAGGCACG  900
        V  K  V  V  P  P  R  F  L  L  P  N  M  D  S  A  T  W  K  I  C  A  K  Y  T  Y  G  K  N  V  Q  G  T

 901  TCGACCATAGTTATTGGATACCGAAATACTCCATTGAATGAAGGCAACATAGCACATTATCCACATGAAAAGTATGAAGTCAAGGTTGACGGATGCTACG  1000
       S  T  I  V  I  G  Y  R  N  T  P  L  N  E  G  N  I  A  H  Y  P  H  E  K  Y  E  V  K  V  D  G  C  Y

1001  AGTTGACTGTCAACAAGACTGTGATGAGATGGTATTCTGCTTTGGTGGATTCAAAGTATATTGGCGTCACTTCTAGTGTCAAAGAAGATGGAACTGGTGT  1100
      E  L  T  V  N  K  T  V  M  R  W  Y  S  A  L  V  D  S  K  Y  I  G  V  T  S  S  V  K  E  D  G  T  G  V

1101  TGAATTCACCGATTCATCAGTCACCCTTCTTCATGAAACGCCACTTAAAATTGACTTCGAAGCAAGGAATAACGAGAAATACTTCCGGCCAGGACTTCCA  1200
        E  F  T  D  S  S  V  T  L  L  H  E  T  P  L  K  I  D  F  E  A  R  N  N  E  K  Y  F  R  P  G  L  P

1201  TACTTTGGAAAACTTGTTGTGAAGCATCCAGATAAGCAAGCAGCATTTAATGCTCTCATCTTAATCTGCCACAAGACGAAGCATCTGACGGAATGCCGAA  1300
       Y  F  G  K  L  V  V  K  H  P  D  K  Q  A  A  F  N  A  L  I  L  I  C  H  K  T  K  H  L  T  E  C  R

1301  ACTTCACAAGTGATATCAACGGAATGATAAAATTTACCATACCACCACAGAAACCGGAAGTTCAGTCTCTTCAAATCTCTGCATCAGCTGCGCTTTATAA  1400
      N  F  T  S  D  I  N  G  M  I  K  F  T  I  P  P  Q  K  P  E  V  Q  S  L  Q  I  S  A  S  A  A  L  Y  K

1401  ATCTGAGAAAGTTGGCGGCGAGCTAGGCCGACTAATCCAACAGCCGAGCGGTGAAATGACTTTGAAACCCTGGTACTCACCAACTGGTAGTTTCATGGAT  1500
        S  E  K  V  G  G  E  L  G  R  L  I  Q  Q  P  S  G  E  M  T  L  K  P  W  Y  S  P  T  G  S  F  M  D

1501  GTGAAGCCCGTACATGATGTCATCTCATGTGGTGAAGAAGTACCTTTGGAAGTGCTTTATACTACCAGTGGCGAAGACATTACTTTGCATTATGAGGTCA  1600
       V  K  P  V  H  D  V  I  S  C  G  E  E  V  P  L  E  V  L  Y  T  T  S  G  E  D  I  T  L  H  Y  E  V

1601  TGTCAAGAGGTAGAATTGTTGATTATGGAAAGAAAATGTACAAGTACAACGCAAACGATTACAAAGAAGATGCTTTCGTCGTACGAAACACCATGAACAT  1700
      M  S  R  G  R  I  V  D  Y  G  K  K  M  Y  K  Y  N  A  N  D  Y  K  E  D  A  F  V  V  R  N  T  M  N  I

1701  ATCGAACGAAAAGAAAGAATTAACCAATGAAACTGTTAACTACATTCTTCCAGGAAGTATCGGAAAGTTCCACCTGCCAATAAAAATGAAAGCCGAAATG  1800
        S  N  E  K  K  E  L  T  N  E  T  V  N  Y  I  L  P  G  S  I  G  K  F  H  L  P  I  K  M  K  A  E  M

1801  TCGCCAATCGCCCGAGTACTCGTCTACTACATCCGCTCCAATAGCGAAGTTATCTCTGCATTCGCCACTTTGGATGTCATGCCATGCTTTCTAAACAAAG  1900
       S  P  I  A  R  V  L  V  Y  Y  I  R  S  N  S  E  V  I  S  A  F  A  T  L  D  V  M  P  C  F  L  N  K

1901  TTTCGCTCAATTTTGAAAAAAATTCCATCAAACCCGGAACGCTTGCCAGGTACAAAATTTCCGCTTCTGCAAAATCTCTGTGCGCAGTTGGCGTGGTCGA  2000
      V  S  L  N  F  E  K  N  S  I  K  P  G  T  L  A  R  Y  K  I  S  A  S  A  K  S  L  C  A  V  G  V  V  D

2001  CAAAAGTGTCCATCTTTTGAAATCCAGTAACCAAATTACGGTCGAAAAGATTTTCAACATTTTGAAAAGCTTCGATGCAACTAACGATAATGTACACTGC  2100
        K  S  V  H  L  L  K  S  S  N  Q  I  T  V  E  K  I  F  N  I  L  K  S  F  D  A  T  N  D  N  V  H  C

2101  AGAAGGTCCCGACGCCGAAGAACTCGATCTGATGCACGTATTGTACGCTCGGTTCCATTTGCTCCTCGACCATCAAGAACGGAATTTGCAGACGCACCGA  2200
       R  R  S  R  R  R  R  T  R  S  D  A  R  I  V  R  S  V  P  F  A  P  R  P  S  R  T  E  F  A  D  A  P
                                                                     β-α cleavage site
2201  TGTCGTTTGAGTTTTCCGGACTGACCTACTTGACCGACAATGCGGTGGGCACATTAGCGCCATGCAGAAGAATAATTCGATATAGGAAACCTGTAAACCG  2300
      M  S  F  E  F  S  G  L  T  Y  L  T  D  N  A  V  G  T  L  A  P  C  R  R  I  I  R  Y  R  K  P  V  N  R
                                                                                   Bait region
2301  TATACAAGAAGATTCAAGAGGACGAAACCAAATTTCAGCACGACGAGGCGCCAATCGACCGTCAGAAATTGCCGAAGTCTTAAATGCAGATGATGACGAC  2400
        I  Q  E  D  S  R  G  R  N  Q  I  S  A  R  R  G  A  N  R  P  S  E  I  A  E  V  L  N  A  D  D  D  D

2401  AATGCCGGCGTCAATGAAGCTGCCGAAGTTCGTGCATACTTCCCAGAAACCTGGCTGTGGGACCTTGAAATTGTAGGCGATGGAGGATACGCAGATAAAG  2500
       N  A  G  V  N  E  A  A  E  V  R  A  Y  F  P  E  T  W  L  W  D  L  E  I  V  G  D  G  G  Y  A  D  K

2501  AAGCAGAAATTCCGCACACGATCACAGAATGGGTAGGAAGCATGTTCTGCACATCAAAAGCAAAAGGACTCGGTGTTTCACCACCAGCCACGATCAAAGC  2600
      E  A  E  I  P  H  T  I  T  E  W  V  G  S  M  F  C  T  S  K  A  K  G  L  G  V  S  P  P  A  T  I  K  A

2601  TTTCCAACCATTCTTCGTGTCGTATACACTGCCTTATTCCGTAATACGCAAAGAAAAAGTACCGATCGTCGTGTCCGTCTTCAACTACCTTCCAGAATGT  2700
        F  Q  P  F  F  V  S  Y  T  L  P  Y  S  V  I  R  K  E  K  V  P  I  V  V  S  V  F  N  Y  L  P  E  C

2701  CTTCCGATCCAAATTAAATTGCAGAAAAACGAAGGCTTCACGTTATTGAGCAACAACTATACTCACAGTATGTGCGTTTGCGGAGGGAAACCCGCAACCC  2800
       L  P  I  Q  I  K  L  Q  K  N  E  G  F  T  L  L  S  N  N  Y  T  H  S  M  C  V  C  G  G  K  P  A  T

2801  ATCAGTTCAGAATCCTCCCGACAAGCCTCGGCGAAGTCAACCTCACAGTTTACAGCCATTCATTCAACGACTCGAACAAAAAGATTTGCCCAGAAGACAA  2900
      H  Q  F  R  I  L  P  T  S  L  G  E  V  N  L  T  V  Y  S  H  S  F  N  D  S  N  K  K  I  C  P  E  D  K

2901  AAATGCGTCAGCGCTCATTGCACGTGATGCCATCACCAAACCATTATTGGTCGAACCCGAAGGTTTTCCACAAGAGCTCACCGAAAGCGTCTTGTTTTGT  3000
        N  A  S  A  L  I  A  R  D  A  I  T  K  P  L  L  V  E  P  E  G  F  P  Q  E  L  T  E  S  V  L  F  C

3001  CCATCTGAGCATCAAAACGGATTCGAACAGGATTTTGAATTCATGCTTCCGCACGATTTGGTTGAAGGATCCGCTCGGGCGTTCTTTTCAATCACCGGTG  3100
       P  S  E  H  Q  N  G  F  E  Q  D  F  E  F  M  L  P  H  D  L  V  E  G  S  A  R  A  F  F  S  I  T  G

3101  ATATCATGGGACCATCATTAAGCGGCCTGGAGAGACTTGTGGCTTTGCCAATCGGATGCGGTGAGCAGAATATGATAAGATTTGCTCCCAATATCTTCGT  3200
      D  I  M  G  P  S  L  S  G  L  E  R  L  V  A  L  P  I  G  C  G  E  Q  N  M  I  R  F  A  P  N  I  F  V
                                                           Thioester site
3201  TATGCAATATCTGCAAGGAACGAACAATGTGACGAAAGAAATTGAAAAGAAAGCTTTGAAGTTCATGAAAACTGGATATCAAAGACAATTGAACTACAGG  3300
        M  Q  Y  L  Q  G  T  N  N  V  T  K  E  I  E  K  K  A  L  K  F  M  K  T  G  Y  Q  R  Q  L  N  Y  R

3301  CACAGTGATGGCTCTTACAGCGCTTTCGGTGAAAATGATCCAGAAGGAAGTTCGTGGTTGACTGCCTTTGTCGTGAAATCTTTTGCTCAAGCACGTCAAT  3400
       H  S  D  G  S  Y  S  A  F  G  E  N  D  P  E  G  S  S  W  L  T  A  F  V  V  K  S  F  A  Q  A  R  Q

3401  TCATCGACATCGATCCTATTGATTTGAAGAAGAGTACCGAATGGTTGTTGAGTAAGCAGCAGGCCGACGGCTGCTTCCCATTCATTGGAATGGTTCATCA  3500
      F  I  D  I  D  P  I  D  L  K  K  S  T  E  W  L  L  S  K  Q  Q  A  D  G  C  F  P  F  I  G  M  V  H  H
                                                                                               Catalytic histidine
```

Supplementary Fig. S1 (continued)

```
 3501  TCAGGACATGAAGGGTGGAATAGGAAAAAATGTGCCGACTGCATTGACAGCATATGTTGTAATCTCCCTTCTGGAATCTGAAACGCCCATCGGCCAAGAC  3600
         Q  D  M  K  G  G  I  G  K  N  V  P  T  A  L  T  A  Y  V  V  I  S  L  L  E  S  E  T  P  I  G  Q  D

 3601  CAGCTTGACAAAGCCTTCGAGTGCATCACAAAAGAGACCGCTCCGAGTCAATACACCGTGGCTTTAACCGCATATGCTTACACTCTTGCCGGACGTTACA  3700
        Q  L  D  K  A  F  E  C  I  T  K  E  T  A  P  S  Q  Y  T  V  A  L  T  A  Y  A  Y  T  L  A  G  R  Y

 3701  ACCTCACCGAAAAACTCCTTGACGATCTGTTTTCAAAATCCAGCACTGAAGGCCGAGGCATGTACTGGCCAACGTCTTCAAAATCCGTCAGCATTGAGCT  3800
       N  L  T  E  K  L  L  D  D  L  F  S  K  S  S  T  E  G  R  G  M  Y  W  P  T  S  S  K  S  V  S  I  E  L

 3801  TGGCTCTTACGTCATTCTGTCCCTGATGAAATTGGGAGGAGCTGCCAACAAAGTCAAAGCAATGGACATTGTCAGGTGGATAGCTCGTCAGAGAAACTCC  3900
         G  S  Y  V  I  L  S  L  M  K  L  G  G  A  A  N  K  V  K  A  M  D  I  V  R  W  I  A  R  Q  R  N  S

 3901  AATGGAGGATTCGTCTCGACACAGGATACTGTTTTAGCTCTTCAGGCCTTTGCGAAGTACTCAGTTACTCTGACGCAGAATGCACAAGATGTAGCAGTGA  4000
        N  G  G  F  V  S  T  Q  D  T  V  L  A  L  Q  A  F  A  K  Y  S  V  T  L  T  Q  N  A  Q  D  V  A  V

 4001  CCGCCAAAGCTAACGGCTTTGATCATACATATGCAGTAAAATCGACCAACCGTTTACTCATGCAAACGGACAAAATTGAAGAATTGCCAAATATCGTGGA  4100
       T  A  K  A  N  G  F  D  H  T  Y  A  V  K  S  T  N  R  L  L  M  Q  T  D  K  I  E  E  L  P  N  I  V  D

 4101  TGTTCAAGCTACTGGTTCTGGATGCGGTTTAATACAGACGACATTGAAATACAATAGAAACAATGTGAATGCAAGCGAGGCATTTGAACTTTCCGCCGTC  4200
         V  Q  A  T  G  S  G  C  G  L  I  Q  T  T  L  K  Y  N  R  N  N  V  N  A  S  E  A  F  E  L  S  A  V

 4201  GGAAATACTCACAAGCAGGGCTGCAAATTAAGAACGCTTGACATTTGTGCTACGTACAAGATTCCAAAGGAAAATTCAAACATGGCAGTCATCACCGTTA  4300
        G  N  T  H  K  Q  G  C  K  L  R  T  L  D  I  C  A  T  Y  K  I  P  K  E  N  S  N  M  A  V  I  T  V

 4301  AAATGATATCAGGTTACATTCCAGTCAAGGATAGCTTAGCACGCTTGAAGAAAGACAAGAAATTGAATTTGAAACGGTTCGAAGTTGATGCCAACTATGT  4400
       K  M  I  S  G  Y  I  P  V  K  D  S  L  A  R  L  K  K  D  K  K  L  N  L  K  R  F  E  V  D  A  N  Y  V

 4401  CAACTTTTATTTCGATTTCTTGGGCAACGAGAAAACCTGCTTCACGATACATCTTGAAAAAGAAATCGATGTCGAAGATGCTAAACCAGCTACGATTTCG  4500
         N  F  Y  F  D  F  L  G  N  E  K  T  C  F  T  I  H  L  E  K  E  I  D  V  E  D  A  K  P  A  T  I  S

 4501  GTCTTCGATTACTATATTTCAGAATTGAAGCTGGAAAAATCATACTCATTGCCTTCTGTTGCTAACTGCTTTTAATTTTCTTGAAAATGTATTTGTTATA  4600
        V  F  D  Y  Y  I  S  E  L  K  L  E  K  S  Y  S  L  P  S  V  A  N  C  F  *

 4601  ATTAAAGCGTATCTGTGCTTCTGAATTTTATTTTTGTTATAAATACGAGTATAAATGTATAAATAAATAAATCGTACTCCAAG  4683
```

AmspiTEP/CD109

```
    1  ACATTACGCACACACCTGTTTGTGTACTTGTGTAGGCCCGGCGTACTACTCTCATTCCATCATTACCTCTCGACTAATCTGTGTCTCCAGTGTTAAGAAC  100
  101  GCTCATCGAGACGACTACAATTCTCGACCTGTCAGGAAAACTCATAAGAAGGAAAGCGTTCGATGAAGAAGAGAATCAACGATGTGAAAATGACATTGGC  200
                                                                      M  K  K  R  I  N  D  V  K  M  T  L  A

  201  AATTATCATAATATTCATCTTTTTATACCCAAGCGTCGAATCTCAAGGGTATTACACTGTTGTAGCTCCAAAGGTCTTACGGCCTGACACAAGGTACCAT  300
         I  I  I  I  F  I  F  L  Y  P  S  V  E  S  Q  G  Y  Y  T  V  V  A  P  K  V  L  R  P  D  T  R  Y  H

  301  ATTGGAGTATCCATATACAACACAACCTCGACCGTTCACGTCGCAGTACAACTAACGGGAAACCTCAGAGTTTCCAGCGAACTCGATGTCAGAGGTGGTG  400
        I  G  V  S  I  Y  N  T  T  S  T  V  H  V  A  V  Q  L  T  G  N  L  R  V  S  S  E  L  D  V  R  G  G

  401  ATACAGGACTAGTCACATTCCAAATTGGAAATTGGTCAGCAGGTGTATATAAGCTCGAAGTTGTTGGCTCCCGTGGCTTGGACTTCCGCAATTCAACGGA  500
       D  T  G  L  V  T  F  Q  I  G  N  W  S  A  G  V  Y  K  L  E  V  V  G  S  R  G  L  D  F  R  N  S  T  E

  501  AATAAAATTCGTCGCTCGAAGCCTCAACGTGTTTATCCAAACAGATAAGTCGGTGTATCGATCTTCGCAAAAGGTTCAATTCAGGGCCATAATTCTCGAC  600
         I  K  F  V  A  R  S  L  N  V  F  I  Q  T  D  K  S  V  Y  R  S  S  Q  K  V  Q  F  R  A  I  I  L  D

  601  AAAAGCCTGTGGCCAAGAAGAACAGCGGTTGAAATTTACATTACTGATGCTGATGGTAATAGAGTGAAGCACTACAGAGGACTGAACGCCAGGCTAGGTT  700
        K  S  L  W  P  R  R  T  A  V  E  I  Y  I  T  D  A  D  G  N  R  V  K  H  Y  R  G  L  N  A  R  L  G

  701  TAGTGAGTGAAGAACTGCAACTGCCCGACCAACCAGTTCTCGGTGTGTGGATAATCCACGTCGTTGCTTCCGGCCAAGAATACAAGAAATCCTTTTCGGT  800
       L  V  S  E  E  L  Q  L  P  D  Q  P  V  L  G  V  W  I  I  H  V  V  A  S  G  Q  E  Y  K  K  S  F  S  V

  801  TGCAGAGTATGTGTTGCCTGGATTCTATGTCAAGGTCAAGTTGTCGCCTTCGTTCGTCACTTACGACAACCCTCTAGTTAGGGCGACGGTTTCGGCCACA  900
         A  E  Y  V  L  P  G  F  Y  V  K  V  K  L  S  P  S  F  V  T  Y  D  N  P  L  V  R  A  T  V  S  A  T

  901  TACAATTATGGAAAGCCTGTAAAGGGAACTGTGACTTTGACAGTAATTCCAAAGCATCGAACGACAAGCATTTCAGTGCGACCTTTGGATTCTTACCAAA  1000
        Y  N  Y  G  K  P  V  K  G  T  V  T  L  T  V  I  P  K  H  R  T  T  S  I  S  V  R  P  L  D  S  Y  Q

 1001  CCATTTTACCGTTGAACAAAGAAGTTACGCACGAAATCAACGTCCGAAGAGTTCTGAACTTGATCACGGATAATTTGAAACGTGAGATTGAATTCATGGC  1100
       T  I  L  P  L  N  K  E  V  T  H  E  I  N  V  R  R  V  L  N  L  I  T  D  N  L  K  R  E  I  E  F  M  A

 1101  AATTGTCGAAGAGGAACTTACCGGTCGCCGATACAATGGGTCAAATAGTATTTTCATCTATAATGACCCGGTCAAATTGGAACTCATCAAGACATCACAA  1200
         I  V  E  E  E  L  T  G  R  R  Y  N  G  S  N  S  I  F  I  Y  N  D  P  V  K  L  E  L  I  K  T  S  Q

 1201  TCGTTCAAACCGGGTCTCGTCTACAAAGCTTTCTTGAAAGTCTCTCGTCAGGACAACACTCCTCTAAACCTTCCAAATGGAGCGCTGACGTTGAAATACG  1300
        S  F  K  P  G  L  V  Y  K  A  F  L  K  V  S  R  Q  D  N  T  P  L  N  L  P  N  G  A  L  T  L  K  Y

 1301  CCTACAACTACAAACCGGGATCCACAAGATCTGATCGGTACAGAATTCCAACCAATGGACTAATTGAACTGAATTTCTTCCCACCACTTTCGAAAGACAC  1400
       A  Y  N  Y  K  P  G  S  T  R  S  D  R  Y  R  I  P  T  N  G  L  I  E  L  N  F  F  P  P  L  S  K  D  T

 1401  TGTTACGATTTTCACTAAGGCTGACTTCAATGGAAAGGAATACGATCTGGCGTACGTCGACAAAGCGTACTCGCCGAGTAATACGTACATGCAGATAACG  1500
         V  T  I  F  T  K  A  D  F  N  G  K  E  Y  D  L  A  Y  V  D  K  A  Y  S  P  S  N  T  Y  M  Q  I  T

 1501  CTCAACACTCCTTTTCCACAAGTTGATCAGGAAGTGGAAGTTTTGGTCAACTGTACGTCACAACTTCCCCAATACGTCTACCAGGTCATTGCCAGAGGCA  1600
        L  N  T  P  F  P  Q  V  D  Q  E  V  E  V  L  V  N  C  T  S  Q  L  P  Q  Y  V  Y  Q  V  I  A  R  G

 1601  ACATATTGCGAACCAGATCTGTGAGACCACCAGGAGGCAACTCCCATTCGTTTAAAGTTCAAATGACCGATAACATGGCACCGTTGGTCCGCATCGTCGT  1700
       N  I  L  R  T  R  S  V  R  P  P  G  G  N  S  H  S  F  K  V  Q  M  T  D  N  M  A  P  L  V  R  I  V  V

 1701  CTATTTCACCCGTGATGATGGCGAAATTGTCGCTGATGGTCTCAGTCTGGACTTGGAAAAAATATTTGAGAATCAAATTTCTTTCACTGCTGGCCCTGGC  1800
         Y  F  T  R  D  D  G  E  I  V  A  D  G  L  S  L  D  L  E  K  I  F  E  N  Q  I  S  F  T  A  G  P  G
```

Supplementary Fig. S1 (continued)

```
1801  GTTCTTCGACCAAGAGAGAAAGTGAGAATATCACTGAATACAGATCCCAACTCCATGGTGGGTTTGATGGGCATTGACCAAAGAAACTTGGTACTGGATC  1900
       V  L  R  P  R  E  K  V  R  I  S  L  N  T  D  P  N  S  M  V  G  L  M  G  I  D  Q  R  N  L  V  L  D

1901  CGGGCAATGACATCACACAAAATGATGTCATCCGCAGCCTTGAAGGATTCGATTCCGGCAAGAAAGATTCGGATCAGCAAATGATTCTTCAGTTACCAGT  2000
      P  G  N  D  I  T  Q  N  D  V  I  R  S  L  E  G  F  D  S  G  K  K  D  S  D  Q  Q  M  I  L  Q  L  P  V

2001  TAGACGAGGAAGGGCTTTGTTTTATCCCGGCTCATTGTCTGCTGCTTCTGTATTTGATGACGCTGGAGTGACTGTTATGTCCAATGGCATCGTCAACTCT  2100
        R  R  G  R  A  L  F  Y  P  G  S  L  S  A  A  S  V  F  D  D  A  G  V  T  V  M  S  N  G  I  V  N  S
     β-α cleavage site
2101  TTCGATCGACGAAAAAGCAAAAGTGCACCTCCTGGCGAAGGCGTCTATGACTTCAAGATCGTGCCGTTTGGTGCCGCACGGCCTCCGGACACTCCACAGC  2200
       F  D  R  R  K  S  K  S  A  P  P  G  E  G  V  Y  D  F  K  I  V  P  F  G  A  A  R  P  P  D  T  P  Q

2201  GGCCAAGGACGCCCTTCATCAACATCGACCTTCCGCCAACTTGGCTCTGGTGGAACAGAACAGTCGGATCAGATGGATCTGCTTCGCTCGAAAGTCACGT  2300
      R  P  R  T  P  F  I  N  I  D  L  P  P  T  W  L  W  W  N  R  T  V  G  S  D  G  S  A  S  L  E  S  H  V

2301  TCCAGAAAATATGACGTCTTGGATCATCAGTGCCTTTTCGATTAGTCCAACAAATGGTTTGGCCCTTGCTCAAAATTCAGCCAAGGTGACAGTCTTTGAA  2400
        P  E  N  M  T  S  W  I  I  S  A  F  S  I  S  P  T  N  G  L  A  L  A  Q  N  S  A  K  V  T  V  F  E

2401  AGATTCTTTGTCAAGTTGATACTTCCACACTCTGTGATCTTAGGCGAAACACTGTCCGTTCAGGTCGTCGTCTTCAACTACAATGATCGTCCAGCACAGG  2500
       R  F  F  V  K  L  I  L  P  H  S  V  I  L  G  E  T  L  S  V  Q  V  V  V  F  N  Y  N  D  R  P  A  Q

2501  TTGAAGTGACAATGGAAAACAAAGGAGGATTTGAATTCACCACAGTAGAAGATGACCCTTCAATTCGCCGGGCGACAAGAATGAGAAAAGCAACGACGGT  2600
      V  E  V  T  M  E  N  K  G  G  F  E  F  T  T  V  E  D  D  P  S  I  R  R  A  T  R  M  R  K  A  T  T  V

2601  ACCTGCTCAAGAAGGAAAGGCAACGTATTTTATGATCAAGCCAAACAGATTGGGATACATTGACATCAAAGTCTTCGCTCGTTCGTCGTTCGCTGGTGAT  2700
        P  A  Q  E  G  K  A  T  Y  F  M  I  K  P  N  R  L  G  Y  I  D  I  K  V  F  A  R  S  S  F  A  G  D

2701  GGTGCGCAAGACAAACTTCTTGTCAAACCGACAGGAGGACCTCAATATTTCAACAGACCAATTTTGATAGATCGCAGATCGGCTGGTGGCGAACCTCTAT  2800
       G  A  Q  D  K  L  L  V  K  P  T  G  G  P  Q  Y  F  N  R  P  I  L  I  D  R  R  S  A  G  G  E  P  L

2801  CCGTTGATGTTGAATTGAACATTCCTCGTACTGTCATCAGAGGTTCAGAAAAAATTGAAGTGACGGCCATCGCTGATGTGATGGGACCGGTCATTGAAAA  2900
      S  V  D  V  E  L  N  I  P  R  T  V  I  R  G  S  E  K  I  E  V  T  A  I  A  D  V  M  G  P  V  I  E  N

2901  CTTGGGCGATCTACTTCGTATTCCACGAGGTTGCGGTGAACAGAACATGGTCAACTTTGTTCCGAACATACTCGTCATCAATTACTTGAAAGGAAATAAC  3000
        L  G  D  L  L  R  I  P  R  G  C  G  E  Q  N  M  V  N  F  V  P  N  I  L  V  I  N  Y  L  K  G  N  N
                                   Thioester site
3001  CGCCTTGCCAACGACATCTACTCCAAAGCCATCGGAAACATTGGAACCGGATATCTAAGAGAACTTACGTACAAGCACGATGATGGTTCTTTCAGTTCTT  3100
       R  L  A  N  D  I  Y  S  K  A  I  G  N  I  G  T  G  Y  L  R  E  L  T  Y  K  H  D  D  G  S  F  S  S

3101  TTGGCAGGACTGATAATAGTGGCAGCACGTGGTTGACTGCTTTCGTTTTGAAGACTTTCAAACAGGCCAAGGAACATATTCCAGACATGATCGACGATGA  3200
      F  G  R  T  D  N  S  G  S  T  W  L  T  A  F  V  L  K  T  F  K  Q  A  K  E  H  I  P  D  M  I  D  D  E

3201  AGTTGTGAAAGCTGCTATGTCGTGGCTCAGCAAGCAGCTGCAATTCAATGGCACGTTTGCAGAGCCCGGAAGTGTGCTCAACAAGGAACTTCAGGGTGGA  3300
        V  V  K  A  A  M  S  W  L  S  K  Q  L  Q  F  N  G  T  F  A  E  P  G  S  V  L  N  K  E  L  Q  G  G

3301  GCTAAGCAAGGCTTGCCTTTGACAGCCTACGTTCTCATAGCCCTCCACGAAAATCGGGAATATTACGATTCTGTTGCCGGAGCACTGGACAGTGCCAGAC  3400
       A  K  Q  G  L  P  L  T  A  Y  V  L  I  A  L  H  E  N  R  E  Y  Y  D  S  V  A  G  A  L  D  S  A  R

3401  AAGCCCAATCGAGTTTGGAACAAAACTACGCAAGCATTGACGACACTTACACGCTGGCTATCATCTCATACGCTCTACAACTAGTCAACAGCCCAAGAAG  3500
      Q  A  Q  S  S  L  E  Q  N  Y  A  S  I  D  D  T  Y  T  L  A  I  I  S  Y  A  L  Q  L  V  N  S  P  R  R

3501  AGACGCAGCCTTCCAGCAACTTGCCAGCCGTGCAAGATTCGGAGAAGAGACAATGTATTGGTCAGCCAATGTACGAAAGCCAAACAACCCGGAAGAAGTA  3600
        D  A  A  F  Q  Q  L  A  S  R  A  R  F  G  E  E  T  M  Y  W  S  A  N  V  R  K  P  N  N  P  E  E  V

3601  TTCCACTTGCCTAGCAGCTCGGACATCGAAATGACTTCGTATGCCCTGATGACGTACACGTTACGAGGCGACATCTCAACGTCGTTGAGGATCATGAAGT  3700
       F  H  L  P  S  S  S  D  I  E  M  T  S  Y  A  L  M  T  Y  T  L  R  G  D  I  S  T  S  L  R  I  M  K

3701  GGTTGGTCGAGAGGCGGAACAGTCTTGGTGGATTCACGTCAACTCAGGATACGGTTATCGGCATTCAAGCTTTGACTATGCTCACGAACAACTTGAATAT  3800
      W  L  V  E  R  R  N  S  L  G  G  F  T  S  T  Q  D  T  V  I  G  I  Q  A  L  T  M  L  T  N  N  L  N  I

3801  CCGAGGCAGCAACTTGGAAATAACTTACAGTTATAACGAAAACGCGGATCATAACAACACGTTTCCTATCCGCAAAAAGATCAACGTCAACGACCAGAAT  3900
        R  G  S  N  L  E  I  T  Y  S  Y  N  E  N  A  D  H  N  N  T  F  P  I  R  K  K  I  N  V  N  D  Q  N

3901  TCTCTCAACATGCAGAGCAGAACTTTGCCAGTTACAGTTAGAAAAGTGAGGATTTCGGCAGAAGGAAGAGGAATTGCCATGGTCCAGGTTTCATGGTCAT  4000
       S  L  N  M  Q  S  R  T  L  P  V  T  V  R  K  V  R  I  S  A  E  G  R  G  I  A  M  V  Q  V  S  W  S

4001  TCAATCTGAAAGTCTCTGCTCCAAACCCGTCTTTCGGTCTCAATCCATTGGTCGATAAAGTCTCCACAAAGGGATATCTACAAGTCAGCTCTTGTATAAA  4100
      F  N  L  K  V  S  A  P  N  P  S  F  G  L  N  P  L  V  D  K  V  S  T  K  G  Y  L  Q  V  S  S  C  I  N

4101  TTACATACCGGAAGGTGAAAGCGGAATGGCTGTTATGGAGTTCTATGCGCCATCTGGATACGTTGTTGACAGAAGCTCTTTGTCGAGTATCCGACAAGAA  4200
        Y  I  P  E  G  E  S  G  M  A  V  M  E  F  Y  A  P  S  G  Y  V  V  D  R  S  S  L  S  S  I  R  Q  E

4201  TCCATCATCAAACGTGTTGAGACCTATGACGACGAAACAATGGTTGCCATCTATTTCGACAAGATTGGTAAGGAACCAGTGTGCCCGACTGTGTCCGCGT  4300
       S  I  I  K  R  V  E  T  Y  D  D  E  T  M  V  A  I  Y  F  D  K  I  G  K  E  P  V  C  P  T  V  S  A

4301  ACAGAGTGCATAGAGTGGCGAATCAGCAGCAGAAACCAGTCGTCGTGTACGATTACTATAACAGAGCTCAAATTGCAAGAGTTTTCTATCGGTTGGCGCA  4400
      Y  R  V  H  R  V  A  N  Q  Q  Q  K  P  V  V  V  Y  D  Y  Y  N  R  A  Q  I  A  R  V  F  Y  R  L  A  H

4401  TGTAACAAAGAAGGAAGAGATCTGCGACGGTTACGAATGCGACAAGAGCAACACCAGACGAAACGATGCGAGTGTCGACCAACGAGACAAGAATTCTTCA  4500
        V  T  K  K  E  E  I  C  D  G  Y  E  C  D  K  S  N  T  R  R  N  D  A  S  V  D  Q  R  D  K  N  S  S

4501  GTTTCCTTAAACAGCTTTTCCAATATCCTAATCCTAACCCAAGTTTTAACTTTTTTAACATTTCAAGTACTGAAATCAATCGTTTGATTTGCAAATGTAC  4600
       V  S  L  N  S  F  S  N  I  L  I  L  T  Q  V  L  T  F  L  T  F  Q  V  L  K  S  I  V  *

4601  AAATTCGCTTTGTAAATACCTTGTTTCATATTTTATCCAAGACGTAACTTATAAGTCTGTTTGGCCTATCTGGAATCCCCTGATAAAGTACTAGCAGTCT  4700
4701  AATGAGCGATTTCATTATTTGACCTTGACGCCTCCATTATGACCCTTTAATGACATCATCCTTTTTTTAAAAACAATTTGGGGATTGATGTCCCTACTTG  4800
4801  ACATGGATATTCTCTATAAAATTATAGCCGGATTGCCAAAATATTGGCGAAT  4852
```

Supplementary Fig. S1 (continued)

ScsuC3

```
   1  TAGCTAGAGTTGATCAAAATGGGTTTCTTGGTCTCTCTTTCGGTGGTGGTTCTGGCCTTCAGTCAGTGTATAAATGGTTTACAATATTCGGTAACAGCCC  100
                         M  G  F  L  V  S  L  S  V  V  V  L  A  F  S  Q  C  I  N  G  L  Q  Y  S  V  T  A

 101  CAAATACATTCAGGATTGGTGTTCAAGAGACGGTAGGAGTCGCCATTACCAATTCTCCAACACCAGTTCAAGTGGAAATTTTCATCCATGATAGAACACC  200
      P  N  T  F  R  I  G  V  Q  E  T  V  G  V  A  I  T  N  S  P  T  P  V  Q  V  E  I  F  I  H  D  R  T  P

 201  CCAAAAGAAAGTCATCGATAGCCAGAAGTTAACACTGCAAAATGATAAACCTCAGATTACTACTCTTCTACTTCGTGCTCAAGATGTTCCAGCAGATCAA  300
        Q  K  K  V  I  D  S  Q  K  L  T  L  Q  N  D  K  P  Q  I  T  T  L  L  L  R  A  Q  D  V  P  A  D  Q

 301  AAAGACCAACCTCATATCTTCGTTGATCTAGCTGTTAAAGAATCAACAAACAAATTCTACAAGGAAATGACGATCCCTGTAACTAAATACTCGGGTTACG  400
       K  D  Q  P  H  I  F  V  D  L  A  V  K  E  S  T  N  K  F  Y  K  E  M  T  I  P  V  T  K  Y  S  G  Y

 401  TCTTCGTTCAGACAGACAAACCAATCTATTTACCTAATCAAAGAGTGCATATACGCCTATTCTATCTGGATGAAAACCTTCTTCCACTAACAGGAGACTT  500
      V  F  V  Q  T  D  K  P  I  Y  L  P  N  Q  R  V  H  I  R  L  F  Y  L  D  E  N  L  L  P  L  T  G  D  L

 501  AACACTGGAAGTTAAGAATCCCAATGGATCAAGAGTATTATACAAAGAAAATCTGCCAGCCACTCCTAGTGGCATTACAGAGGCTTCATTCAAATTTCCA  600
        T  L  E  V  K  N  P  N  G  S  R  V  L  Y  K  E  N  L  P  A  T  P  S  G  I  T  E  A  S  F  K  F  P

 601  TCATCTCCAGTGTTTGGGAATTGGACGGTTACAGCTTTTTATGGTTATAAAAAAGCAGCCAGAACTACAGTTCGATTTGAAGTAAAGGATTACGTTTTAC  700
       S  S  P  V  F  G  N  W  T  V  T  A  F  Y  G  Y  K  K  A  A  R  T  T  V  R  F  E  V  K  D  Y  V  L

 701  CTACATTCTCAGTAAAAATCAAATCACAAAAAGTTGTTCTAAAAACCGATAACTTGGTCAAGGTTGATATGATAGCAGAATATGTGTATGGAAAACCTGT  800
      P  T  F  S  V  K  I  K  S  Q  K  V  V  L  K  T  D  N  L  V  K  V  D  M  I  A  E  Y  V  Y  G  K  P  V

 801  TGAAGGATTTGTTAATTATAAATTTGCAATCCGAAAACCATCTGGAAGCATACATTCAATTGGAGGACATTCAAATTTAAAACTGCGAGATGGTAAATCT  900
        E  G  F  V  N  Y  K  F  A  I  R  K  P  S  G  S  I  H  S  I  G  G  H  S  N  L  K  L  R  D  G  K  S

 901  ACAATCACTATCAGAAAGTCAGACATCGTGAAAAAACTTCAATGGTTCCCAGCTATCGATAAATCAGTTCTCATCGTTGAAGCAGAAGTAATAGAACAAG  1000
       T  I  T  I  R  K  S  D  I  V  K  K  L  Q  W  F  P  A  I  D  K  S  V  L  I  V  E  A  E  V  I  E  Q

1001  CAACAGGAAAAAGAGAATCAGAATATGATGACAGCACCATATTTACAACATCACCATATGTGATAGATCTTTCCAGGTCACTAAATGAATTTAAACCAGG  1100
      A  T  G  K  R  E  S  E  Y  D  D  S  T  I  F  T  T  S  P  Y  V  I  D  L  S  R  S  L  N  E  F  K  P  G

1101  TGTTCCATATCAAGTGCAGGTTGATGTTCGACTTGTTAATAATCAACCTGTGAATGATAGAATACCAGTGACTGTGAATGCCCGAGCAAAGAAAGGAGGA  1200
        V  P  Y  Q  V  Q  V  D  V  R  L  V  N  N  Q  P  V  N  D  R  I  P  V  T  V  N  A  R  A  K  K  G  G

1201  AACTTCCAAGCTTTAGGAAAGAAACCAGATCTTAGGACTGATGTCCAAGGACGTGTCATGTTTCAATTTGACACTGATGAGAATTTTGAAGAACTAGTAA  1300
       N  F  Q  A  L  G  K  K  P  D  L  R  T  D  V  Q  G  R  V  M  F  Q  F  D  T  D  E  N  F  E  E  L  V

1301  TAGAAGTTCAAACAAAAGATGAAGCTGTGGGGGATAATCAAGCAAAAACAACTCTCAGTGTTATACGTTACAATACTCCTGTTTCAAAGACTTATGTTTG  1400
      I  E  V  Q  T  K  D  E  A  V  G  D  N  Q  A  K  T  T  L  S  V  I  R  Y  N  T  P  V  S  K  T  Y  V  W

1401  GATTGCTGCACCACACGAAGGAACATTATTTCAAGTTGGGAAGACATTCCAAACTCAAGTGACTGTATATCCTCAAGAAAGGCAAATGAAGCTTATGTAT  1500
        I  A  A  P  H  E  G  T  L  F  Q  V  G  K  T  F  Q  T  Q  V  T  V  Y  P  Q  E  R  Q  M  K  L  M  Y

1501  ATGGTTGTTAGCCGAGGAAAGATTTTGATGATGAATGAAACAGAAACTAGAGGAGAAGCTGTAGTGAGAACAATCCAATTTCCTGTAACTGTTGACATGT  1600
       M  V  V  S  R  G  K  I  L  M  M  N  E  T  E  T  R  G  E  A  V  V  R  T  I  Q  F  P  V  T  V  D  M

1601  CACCCAGTTTTCGATTAATAGTTTATTTCATTAAAAATGGGAAAATTATAGCTGATTCATTGAACATTGATGTAGAAAGAACTTGCAAGTACAATAATGG  1700
      S  P  S  F  R  L  I  V  Y  F  I  K  N  G  K  I  I  A  D  S  L  N  I  D  V  E  R  T  C  K  Y  N  N  G

1701  AAAATTCTCCGTCACTGCAGACTCGTCATCAGGATTCTTTAGCCCAAACCAAGAAGTAACTTTTAAAATAACAGGAGAGCCAGATTCTGTGGTTGGAATT  1800
        K  F  S  V  T  A  D  S  S  S  G  F  F  S  P  N  Q  E  V  T  F  K  I  T  G  E  P  D  S  V  V  G  I

1801  GGAGCTGTAGATGAAGCGGTGTATCTCCTCAACGATAGGGATGTTCTCACTAGAGACAAAATGTTTAAAGAACTTAGCAAACATGATCTTGGTACTGGAC  1900
       G  A  V  D  E  A  V  Y  L  L  N  D  R  D  V  L  T  R  D  K  M  F  K  E  L  S  K  H  D  L  G  T  G

1901  CTGGAGGAGGTATTAATCCAGCAGTTGTTTTCAAGAACGCTGGCATATTGATGCTTTCAAATAATGAAATTGGAGAACATGGAAGAAAAGAAGGCATAAC  2000
      P  G  G  G  I  N  P  A  V  V  F  K  N  A  G  I  L  M  L  S  N  N  E  I  G  E  H  G  R  K  E  G  I  T

2001  TCAACCAAAAGAACGAAGAAAACGTAGCCTTTTGGAAAAAGTTGAAGAGTATAGTGGACAGGCTGCGATATGTTGTAAATTCGGCCAATTTGAAGGCCCA  2100
        Q  P  K  E  R  R  K  R  S  L  L  E  K  V  E  E  Y  S  G  Q  A  A  I  C  C  K  F  G  Q  F  E  G  P
                β-α cleabage site                                           ANA domain
2101  GTTGATATGGATTGTGCTGATAGAGCAGCAATGATTTACGAGAAAATTGGAGAGAAATTCAACTGCAGCACCGCATTTCTTGATTGCTGTGAAAATAAAT  2200
       V  D  M  D  C  A  D  R  A  A  M  I  Y  E  K  I  G  E  K  F  N  C  S  T  A  F  L  D  C  C  E  N  K

2201  TGAAATATATGGCTGCTAATCCAGGGCGTGCTAATCAAGAGGTAGAAGATGAAAAGCCTATAAATGAATTAATAGAAATGGTTGAAGCAGATACTTTGAA  2300
      L  K  Y  M  A  A  N  P  G  R  A  N  Q  E  V  E  D  E  K  P  I  N  E  L  I  E  M  V  E  A  D  T  L  K

2301  ACATATTCGACACATCTTTCCTGAAACATGGTTTTTCAATACGTTACAAATTGGTGATGGTAATGAATGCAAAGGAGAGCAAGGCCAATGTATAACGAAG  2400
        H  I  R  H  I  F  P  E  T  W  F  F  N  T  L  Q  I  G  D  G  N  E  C  K  G  E  Q  G  Q  C  I  T  K

2401  TTTAATGCTCCACACAGTATAACAACCTGGGTAATCCAAGGCATAGCTGTTTCGAAGACAACAGGAATGTGTGTTGCTGAGCCATTGAAAATCACTGTTT  2500
       F  N  A  P  H  S  I  T  T  W  V  I  Q  G  I  A  V  S  K  T  T  G  M  C  V  A  E  P  L  K  I  T  V

2501  TCAAAAAGACATTTGTTCAGTTGAGCTTACCACCAGTAGCAATAAGAGGAGAACAGATTGAAGTGCTTGCAACTGTATTTAATTATGAACCTGAAGATTT  2600
      F  K  K  T  F  V  Q  L  S  L  P  P  V  A  I  R  G  E  Q  I  E  V  L  A  T  V  F  N  Y  E  P  E  D  L

2601  GGATGTTAGTGTTTACATGTTTGGGGTTGAAGGTGTCTGCATGGGAGCTGGTCCTGGTGAAAGAACGGAGATTAGAAAGCTCAAAGTGCCAGCCAATGGC  2700
        D  V  S  V  Y  M  F  G  V  E  G  V  C  M  G  A  G  P  G  E  R  T  E  I  R  K  L  K  V  P  A  N  G

2701  GCATCATCAGCCACGTTCTCTGTGATGCCTCTTGAAGTGAGCGAATACCAACTTCGTGTGGCAGCACTAAGTTATACTTCCAGTGATGCTGTCCAAAAAG  2800
       A  S  S  A  T  F  S  V  M  P  L  E  V  S  E  Y  Q  L  R  V  A  A  L  S  Y  T  S  S  D  A  V  Q  K

2801  TTCTCCGAGTTGTGCCAGAGGGTGCCCGTGTTGAAAAATCTATCTCCTTTACATTAGACCCAGGTGGCATTTACAGCAAGCGACCAAGAAGACAAGCAGA  2900
      V  L  R  V  V  P  E  G  A  R  V  E  K  S  I  S  F  T  L  D  P  G  G  I  Y  S  K  R  P  R  R  Q  A  D
```

Supplementary Fig. S2

```
2901  TCATGAAGGCACAATAAAAGACGTGTATGATGAGGTGCTTCGCAAGCAGTTAATAACAGTTGATTTAGATTTACCAAATAATCATATACCTGGAACAGAA  3000
        H  E  G  T  I  K  D  V  Y  D  E  V  L  R  K  Q  L  I  T  V  D  L  D  L  P  N  N  H  I  P  G  T  E

3001  AAATGTTTCATCAGCGTTGTAGGTGATCCTGTTGGGCAAGCAGTGAATACTACACTTTCTGGATTAGGAGGATTTCTGAAAATGCCAACAGGGTGTGGCG  3100
       K  C  F  I  S  V  V  G  D  P  V  G  Q  A  V  N  T  T  L  S  G  L  G  G  F  L  K  M  P  T  G  C  G
                                                                                              Thioester site
3101  AACAAACTATGATTGCTTTAGGACCACTAGTTTATACAATGTCCTATTTAAAAAAAACAAAACAAATGACAGCAAATATTGAAGATACTGGCTACAAATA  3200
      E  Q  T  M  I  A  L  G  P  L  V  Y  T  M  S  Y  L  K  K  T  K  Q  M  T  A  N  I  E  D  T  G  Y  K  Y

3201  TCTGTGGGGAGGATATTCCCTTCAGCAAAAGTATAGAAAAGCTGATGGATCCTATGCTGTGTGGGCAACTTATGCATCAAGTACTTGGTTGACGGCATTT  3300
        L  W  G  G  Y  S  L  Q  Q  K  Y  R  K  A  D  G  S  Y  A  V  W  A  T  Y  A  S  S  T  W  L  T  A  F

3301  GTGGCAAAAGTTTTTTGTCAGGCTTCAGCTTTCATTCCTATTCCTCCTGAAAATATTAATACAGCTCTGGAATGGTTGGTGAAAAGACAGCATGCTAATG  3400
       V  A  K  V  F  C  Q  A  S  A  F  I  P  I  P  P  E  N  I  N  T  A  L  E  W  L  V  K  R  Q  H  A  N

3401  GAATTTTTGCTGAAGCATATAAAGTACATCATAGAGAAATGACAGGTGGCATTCAAGGAGATGTTACACTGACGGCTTATGTTTTAATTTCACTATTGGA  3500
      G  I  F  A  E  A  Y  K  V  H  H  R  E  M  T  G  G  I  Q  G  D  V  T  L  T  A  Y  V  L  I  S  L  L  E
                        Catalytic histidine
3501  ATGCAATAAATATGACTCTGTAACCAAAAAAACAGCAGTAAACAGAGCCATTTTATTTTTGGAGCAAAATATTCCACATCTTCAGAGGCCCTATACCATT  3600
        C  N  K  Y  D  S  V  T  K  K  T  A  V  N  R  A  I  L  F  L  E  Q  N  I  P  H  L  Q  R  P  Y  T  I

3601  GCTATTGTTTCTTATGCTCTCGCTTTAACGAACAGTACAAAGCGTCAAGAAGCGAATCAAAAATTAAAAAATATTGCCAAATTCCATCAAGAAAGTTATA  3700
       A  I  V  S  Y  A  L  A  L  T  N  S  T  K  R  Q  E  A  N  Q  K  L  K  N  I  A  K  F  H  Q  E  S  Y

3701  CCCGTTACTGGAACTGGGATGCTACAGAGTTTGGAGCAGGACCAAAACCTTGGGTTTATCAACACAAACCAGCTGCTGTTGCAGTGGAAACCACCAGCTA  3800
      T  R  Y  W  N  W  D  A  T  E  F  G  A  G  P  K  P  W  V  Y  Q  H  K  P  A  A  V  A  V  E  T  T  S  Y

3801  TGCTTTACTTGCACAGTTAGCCTATGATGACTTAGATTATGCTCATCCCATTGTCAACTGGTTGAATCGTCAAAGATCAGCTAGTGGATCATTCGTGTCA  3900
        A  L  L  A  Q  L  A  Y  D  D  L  D  Y  A  H  P  I  V  N  W  L  N  R  Q  R  S  A  S  G  S  F  V  S

3901  ACTCAGGATACAGTGATGGCACTTCAAGCTTTGACTGAATACAATATTAAAGCTAATGTCCCAGCACTTGATATTATGTGCAATATTTCTAGCAGTGCAT  4000
       T  Q  D  T  V  M  A  L  Q  A  L  T  E  Y  N  I  K  A  N  V  P  A  L  D  I  M  C  N  I  S  S  S  A

4001  CTGCCCGTGTCAAAAGAACCATATTGTTGAAAAAAGATAGACCTCAAGAAATCCAAGAGATTGAGGTTCCTCCAAAAGGAAGGTTATATTTTGATGTCAC  4100
      S  A  R  V  K  R  T  I  L  L  K  K  D  R  P  Q  E  I  Q  E  I  E  V  P  P  K  G  R  L  Y  F  D  V  T

4101  TGGAAAAGGAATGGGAACATTGTCACTATCTATGCGTTTTAATGTAGAAAAGAATCCTGAAGACAGTTGTCATTATGATCTGACAATTACAAGCGAAGAG  4200
        G  K  G  M  G  T  L  S  L  S  M  R  F  N  V  E  K  N  P  E  D  S  C  H  Y  D  L  T  I  T  S  E  E

4201  GCAGATGAAATTATTAAACCTGTGAATCTTAAGCCAGAATTTGCTGGAAATGATATTCTTCCTGCTGCTGTAGTTCGATCAGTATTTGATGAGAGAGTTC  4300
       A  D  E  I  I  K  P  V  N  L  K  P  E  F  A  G  N  D  I  L  P  A  A  V  V  R  S  V  F  D  E  R  V

4301  AAAAAGAGAAATTTGGTTATGATGTTGAAGGCAAAGATGAAAATAATCCTGCAGTAGATCGTGCTGGAAAAGAAATTTACAATGTTGGAAGTCAACATGT  4400
      Q  K  E  K  F  G  Y  D  V  E  G  K  D  E  N  N  P  A  V  D  R  A  G  K  E  I  Y  N  V  G  S  Q  H  V

4401  TATAAAAATAAACATCTGTGTTAAATATAAGGAAAAACAAAATGCTGGCATGTCAATCCTAGATGTTGGCTTGTTTACAGGTTACAAACCAATAAAAGAA  4500
        I  K  I  N  I  C  V  K  Y  K  E  K  Q  N  A  G  M  S  I  L  D  V  G  L  F  T  G  Y  K  P  I  K  E

4501  GATTTAATTACTTTAACTATGAAAAAAGAATTAAAAGTAGGTCAGTTTGAAATTACTGACAGATCTGTCATCTTATATCTCGATGAAGTCCCTATTAATA  4600
       D  L  I  T  L  T  M  K  K  E  L  K  V  G  Q  F  E  I  T  D  R  S  V  I  L  Y  L  D  E  V  P  I  N

4601  AACCAATTTGTCTGAACTTCCGAGCATCCAAAGAAATTCATGTTGGAAAAGTTCAACCAACAGCTGTTAAAATTTATAACTATTATGATCCAGACAAATC  4700
      K  P  I  C  L  N  F  R  A  S  K  E  I  H  V  G  K  V  Q  P  T  A  V  K  I  Y  N  Y  Y  D  P  D  K  S

4701  GTGTACACAATTTTATGGTCCAGACAAAGGCAGCGTGATGTTGAAAAAAATTTGTGAAGGCAAACAATGTGTTTGTGTTGAAGGAATCTGCCCATCCTGC  4800
        C  T  Q  F  Y  G  P  D  K  G  S  V  M  L  K  K  I  C  E  G  K  Q  C  V  C  V  E  G  I  C  P  S  C

4801  TTTCCTTTCCAAGACATTCGAAATATTGCAAATGACATAGATCGTCGAATGGAGCTACTCAATGCTGTTTGTGATCGAAAGACTGATTACTTCTGGAATG  4900
       F  P  F  Q  D  I  R  N  I  A  N  D  I  D  R  R  M  E  L  L  N  A  V  C  D  R  K  T  D  Y  F  W  N
                                              C345C domain
4901  GAACTATCAAAAATATACGTGAAGATGGAAGTTTTAAATATTTTGAGTTTGAAGTTACAGATGTTCTGAAAGAAGGAGTGCAACAAGAACGTGAAATACA  5000
      G  T  I  K  N  I  R  E  D  G  S  F  K  Y  F  E  F  E  V  T  D  V  L  K  E  G  V  Q  Q  E  R  E  I  Q

5001  AGATGAGACTGTAACATTTGTTGGAAAAGCTGCATGTAATTGTCCCAAATTGACCATCGGAGAAGTTTATATGATAATGGGTGAAGATGGATTTGAATAC  5100
        D  E  T  V  T  F  V  G  K  A  A  C  N  C  P  K  L  T  I  G  E  V  Y  M  I  M  G  E  D  G  F  E  Y

5101  AAAACAGACACCGGTGAAAAGAAATTCAAATATCTTTTTAGCAAGTATACTCGTATCTATCACTCCCGCTCACTAAGAATGATTCAAGACCAAAGAGGAC  5200
       K  T  D  T  G  E  K  K  F  K  Y  L  F  S  K  Y  T  R  I  Y  H  S  R  S  L  R  M  I  Q  D  Q  R  G

5201  AAAAACTTCAGAAAACTTTCAACACGATTTATATGCGATTTAAGAGAGGAGAACGTTGCATTCATTAACATGCTGTAATTGTATTACAAACAAAGTATGG  5300
      Q  K  L  Q  K  T  F  N  T  I  Y  M  R  F  K  R  G  E  R  C  I  H  *

5301  GTTAATATTTTTTGTTAAATAAAGAACTTTTTTATGGAAAAAAAAAAAAAAAAAAAA  5357
```

ScsuC3-2

```
   1  TATGATTCCAGAGTTAACTCTTTCTGAAATTTAAGCATGAAAGTTCCCTCTGACAACCTACTTTCTATATTTTTCATTGCGTTAATAGAAACTTATGCTG  100
                                                 M  K  V  P  S  D  N  L  L  S  I  F  F  I  A  L  I  E  T  Y  A

 101  TAGAATATGTCTCATGCATAAATATTGTTGTTTCCGCTCCCAATGTCATCCGATTTGGTGTTGAAGAAACGGTTGCTGTCAGCATATTAAACCATCCATC  200
      V  E  Y  V  S  C  I  N  I  V  V  S  A  P  N  V  I  R  F  G  V  E  E  T  V  A  V  S  I  L  N  H  P  S

 201  CCAAGTTCAAGTTGATGTATGTTTGCATGATCATCCGTTAAAAATATCAACAATATCTTGCCAAAGGATAAGTTTATATGAAGGAAAACCAAAGATTGTC  300
        Q  V  Q  V  D  V  C  L  H  D  H  P  L  K  I  S  T  I  S  C  Q  R  I  S  L  Y  E  G  K  P  K  I  V

 301  AACGTAGTTTTACATCCAGAAAATATTCCTGAACAACAAAGATCGGATGCTAACTTTTTTGTCCTTCTAGTTGTTAGAGCAGTTTCTGGCGAATTTACAC  400
       N  V  V  L  H  P  E  N  I  P  E  Q  Q  R  S  D  A  N  F  F  V  L  L  V  V  R  A  V  S  G  E  F  T
```

Supplementary Fig. S2 (continued)

```
 401  AAAAACAGGCCGTGATTCCGGTGATGAAATATTCCGGTTATGTCTTTGTTCAAACTGATAAACCTCTCTATTTACCCAATCAAAAAGTACACATTCGAGT  500
      Q  K  Q  A  V  I  P  V  M  K  Y  S  G  Y  V  F  V  Q  T  D  K  P  L  Y  L  P  N  Q  K  V  H  I  R  V

 501  ATTGCGTCTAGACGAAAAATTACTTCCTAAAAATGACGTCGTTCGATTGGAAATAAAGAATCCAAAAGATATGATCGTAAGACAACAAGAACTGATGCCG  600
        L  R  L  D  E  K  L  L  P  K  N  D  V  V  R  L  E  I  K  N  P  K  D  M  I  V  R  Q  Q  E  L  M  P

 601  AATACGAAAGGCATCAGAGAAGCTGTCTTTACAATTTCACAAGTGCCCATTTTAGGGAATTGGTCAGTCAGCGTTCATTATGGTTTTAAATTTGTGGCAA  700
       N  T  K  G  I  R  E  A  V  F  T  I  S  Q  V  P  I  L  G  N  W  S  V  S  V  H  Y  G  F  K  F  V  A

 701  AAACAACTGTTCGCTTCGAAATAAAGGAATATGTATTGCCAACATATTCTGTGACGATAGATCTTCCAAAAGTTATTTTAGAAACACAAAAAGATCTTGA  800
      K  T  T  V  R  F  E  I  K  E  Y  V  L  P  T  Y  S  V  T  I  D  L  P  K  V  I  L  E  T  Q  K  D  L  E

 801  AGGAAAAGTGATAGCAAAATACGTATATGGTAAACCTGTGCAAGGTTTTGTCAATTTTAAATATTCTATAAAAAATTCTGTTGGGGTTCAAATTTTAATA  900
        G  K  V  I  A  K  Y  V  Y  G  K  P  V  Q  G  F  V  N  F  K  Y  S  I  K  N  S  V  G  V  Q  I  L  I

 901  GGCATGCGTAGAAATGTTAAGCTTTCCTCAGGAGAAATACAGTTTAAGATTCCACTTGAAAAAGAAATCAAAGCTAAAAATCTTCCTTGGTTTCCAAATA  1000
       G  M  R  R  N  V  K  L  S  S  G  E  I  Q  F  K  I  P  L  E  K  E  I  K  A  K  N  L  P  W  F  P  N

1001  TTGAGAAAAGTAGGCTGATCATGGAAGCCGAAGCAATTGAACAAGCATCGGGAAAAAGAGAAACAACAATTGTCGACAACACTGTATTCACTACTTCTCC  1100
      I  E  K  S  R  L  I  M  E  A  E  A  I  E  Q  A  S  G  K  R  E  T  T  I  V  D  N  T  V  F  T  T  S  P

1101  TTATAAAATCAGCTTCAAAAATTCATTAAAAAGTTTCAAACCTGGATTTCCTTATCAAGTTCAAGTTGAAGTCCTTTTATTTGACAATCAACAAATACCA  1200
        Y  K  I  S  F  K  N  S  L  K  S  F  K  P  G  F  P  Y  Q  V  Q  V  E  V  L  L  F  D  N  Q  Q  I  P

1201  AAGCCTGTCACTGTTGAAATAAGTGCAACAGCCTTCAAGAAAGATTACTCTTCTAGACAAATTGGAAAAATTGCTCCTGAAGTAACTGATGAATTCGGTC  1300
       K  P  V  T  V  E  I  S  A  T  A  F  K  K  D  Y  S  S  R  Q  I  G  K  I  A  P  E  V  T  D  E  F  G

1301  GAGCTGTGTTTCAATTCGACACGGATGGAACAGACACACGAATAGTAATTACAGCGAAAACTAATGACAAAAATTATGAAGAAAAGAATCAAGCTAATCA  1400
      R  A  V  F  Q  F  D  T  D  G  T  D  T  R  I  V  I  T  A  K  T  N  D  K  N  Y  E  E  K  N  Q  A  N  Q

1401  GACTTTCACTGGCTATCAATTTTCCTCTCCGTCTAACAGTTTTATTTGGCTGCAATCTCCAAAAGAAGGACTTCGTTTTAAAGTAGGAAAAACTCTTCAT  1500
        T  F  T  G  Y  Q  F  S  S  P  S  N  S  F  I  W  L  Q  S  P  K  E  G  L  R  F  K  V  G  K  T  L  H

1501  ACTACAGCAACTCTTCGATCATCTGATGCGTCTCAGAGACTTTACTACATGGTTGTTAATCGAGGAAATATTTTGGCAATGAATCAAACAGAAGCCCAGT  1600
       T  T  A  T  L  R  S  S  D  A  S  Q  R  L  Y  Y  M  V  V  N  R  G  N  I  L  A  M  N  Q  T  E  A  Q

1601  CAGCTGTGATTCGAACTTTGCATGTGCCTATTACAACCGAAATGGTTCCTAGCTTTCGCCTTGTTGTTTTTGTTATTGAAAAATATTATTTAGTGGCCGA  1700
      S  A  V  I  R  T  L  H  V  P  I  T  T  E  M  V  P  S  F  R  L  V  V  F  V  I  E  K  Y  Y  L  V  A  D

1701  TTCAATGCAGTTCGAAGTTGAAAGGATTTGCAAATACAACGAAGGCAAAGGGCTAATCATTGAGGCAAGCACACCGCTGGCCAGTCCTGGGGAAAATATT  1800
        S  M  Q  F  E  V  E  R  I  C  K  Y  N  E  G  K  G  L  I  I  E  A  S  T  P  L  A  S  P  G  E  N  I

1801  AACTTCAAAATTAAAGGCGAAGAAGAATCTTATGTTGGTTTATTAGGAGTTGACGAAGCCGTTTATGTCTTGAACAAACAAGATTTGCTGACAAAAGAAA  1900
       N  F  K  I  K  G  E  E  E  S  Y  V  G  L  L  G  V  D  E  A  V  Y  V  L  N  K  Q  D  L  L  T  K  E

1901  AGATGTTCAGGGAACTAAGAAATCATGATCTTGGAAAGGGTCCTGGAGGTGGGATTTCTACTGAAGCTGTATTAAGAGATTCGGGTATCATAATATTGTC  2000
      K  M  F  R  E  L  R  N  H  D  L  G  K  G  P  G  G  G  I  S  T  E  A  V  L  R  D  S  G  I  I  I  L  S

2001  AAGTGTTTACATCGGAGAACATGGAAGGGAAGAGTCTTTAATACAAAGTCAATCGCGGAAGAAACGCAGTCTACCAGATAAAGTCAACGAATATTCTGGA  2100
        S  V  Y  I  G  E  H  G  R  E  E  S  L  I  Q  S  Q  S  R  K  K  R  S  L  P  D  K  V  N  E  Y  S  G
                                                          β-α cleabage site
2101  AAAGCTGCTATCTGTTGTCGAATGGGCCAATTCGAAGGACCACAGCATTTAAACTGCACTTCGAGGGCCACAATGATTGAAGATTCGATTGGTGAAAAAC  2200
       K  A  A  I  C  C  R  M  G  Q  F  E  G  P  Q  H  L  N  C  T  S  R  A  T  M  I  E  D  S  I  G  E  K
                                                ANA domain
2201  ACAACTGCTCAGTAGCTTTCTTAGACTGTTGTCAACACGCTGAGGAAATAAGGAAAACATTCGGTTCAGGAGTTGGTCGTTCTTTAGATGAAGAAGATGA  2300
      H  N  C  S  V  A  F  L  D  C  C  Q  H  A  E  E  I  R  K  T  F  G  S  G  V  G  R  S  L  D  E  E  D  E

2301  ATCAGATCCCAATTTTGCTGATATTATGCAGGTTATCGAAACGTTTGAACAAGAAACTTTGGATAACATCCGTCGTTACTTCCCTGAGACGTGGATTTTG  2400
        S  D  P  N  F  A  D  I  M  Q  V  I  E  T  F  E  Q  E  T  L  D  N  I  R  R  Y  F  P  E  T  W  I  L

2401  GACATTTTCCAAATTACAAATTCTGAATGTAAGAGAGAAGATTTAAGTGTATGTGAGAAAACATACACAGCTCCTCATAGCATCACTACTTGGATGGTTC  2500
       D  I  F  Q  I  T  N  S  E  C  K  R  E  D  L  S  V  C  E  K  T  Y  T  A  P  H  S  I  T  T  W  M  V

2501  AAGGTTTTGGATTATCTCGTACCACGGGACTGTGTATAGCAGATCCCATAAGAATTCCTGTTTTTAAACCGATGTTCGTTGAACTGAATCTTCCACCAGC  2600
      Q  G  F  G  L  S  R  T  T  G  L  C  I  A  D  P  I  R  I  P  V  F  K  P  M  F  V  E  L  N  L  P  P  A

2601  AGCTGTCTTGGGAGAACAGATAGAAGTAGTAGCTACTGTGTTCAATTATGGACAGGAATCACTGAAAGTAACAGTGTATATGTATGGTGTAGAAGGTATT  2700
        A  V  L  G  E  Q  I  E  V  V  A  T  V  F  N  Y  G  Q  E  S  L  K  V  T  V  Y  M  Y  G  V  E  G  I

2701  TGCATGGGAGCAGCAGCAGGAGAAAAGAGTCCCGTTAGACAGGTCGAAGTTTCAGCGAACAGTGCAACTTCTGTAAGTTTTCCTGTTATGCCTCTGGAAG  2800
       C  M  G  A  A  A  G  E  K  S  P  V  R  Q  V  E  V  S  A  N  S  A  T  S  V  S  F  P  V  M  P  L  E

2801  TTAGCGAATATCCATTACGAGTGGTAGCTTTAAGTTGGAGAGCTAACGACGCAATCGAAAAGAAATTAAGAATTGTGCCAGAAGGCGTTACCAAAGACAA  2900
      V  S  E  Y  P  L  R  V  V  A  L  S  W  R  A  N  D  A  I  E  K  K  L  R  I  V  P  E  G  V  T  K  D  K

2901  ATCACTTTCTTTCTTCCTGGATCCTTCAGGGCTCATCAGAAATAAACATCGGAAACGTGAAGTAACTCCCGCAGGTGTTATTGAATACTCCACAAATAAA  3000
        S  L  S  F  F  L  D  P  S  G  L  I  R  N  K  H  R  K  R  E  V  T  P  A  G  V  I  E  Y  S  T  N  K

3001  CAAAAGATGAAGATAAACCTGACGTTGCCGGAAAATTATGTTCCAGGAACAGAAGAATGTTTTGTCAGTGTTATAGGAGATTCAATGGGTTCAGTTGTGA  3100
       Q  K  M  K  I  N  L  T  L  P  E  N  Y  V  P  G  T  E  E  C  F  V  S  V  I  G  D  S  M  G  S  V  V

3101  GCACTTCACTTAAAGGATTGGATCAGTTTTTAGTAGCCGCAGGTCCTCATGCCTGTGGAGAACAAACATTAGTGAAGTTAGCACCATTAGTATACACTAT  3200
      S  T  S  L  K  G  L  D  Q  F  L  V  A  A  G  P  H  A  C  G  E  Q  T  L  V  K  L  A  P  L  V  Y  T  I
                                                    Thioester site
3201  AAATTACTTGAAAAAGACTAAACAACTGACGTCGTCGTTTGAAAGTAAAGGATATAGTTATATATCACAGAGTTATGACCAACAGATGAAATACAGGAAA  3300
        N  Y  L  K  K  T  K  Q  L  T  S  S  F  E  S  K  G  Y  S  Y  I  S  Q  S  Y  D  Q  Q  M  K  Y  R  K

3301  GCAGACGGATCATTTTCTCTATGGACATTTACCGAATCGGGAACATGGCTAACTGCTTATGTATTAAAAGTATTTTGTCAGGCATATTCCAGTAATATTT  3400
       A  D  G  S  F  S  L  W  T  F  T  E  S  G  T  W  L  T  A  Y  V  L  K  V  F  C  Q  A  Y  S  S  N  I
```

Supplementary Fig. S2 (continued)

```
3401 TTTCGATACCTAATTATATTAGATGTGATGGCATTAAGTGGCTACTAACAAAACAATATACCAATGGGGAATTTTTTCCTTTAAGTTATAGAATATATGG 3500
     F  S  I  P  N  Y  I  R  C  D  G  I  K  W  L  L  T  K  Q  Y  T  N  G  E  F  F  P  L  S  Y  R  I  Y  G

3501 AGATCAAAAAATTGGTTTGACTGCTTTCGTCCTGATTACAATGCTTGAATGTGATTCCTGCTATTCGGCGGAAAAAAATCTAGCAGTGTCAAGAGCTACT 3600
       D  Q  K  I  G  L  T  A  F  V  L  I  T  M  L  E  C  D  S  C  Y  S  A  E  K  N  L  A  V  S  R  A  T

3601 GCATATTTAAGAAGTAGACTTGATTCCATAACAAATTATTACACCATGGCTATAGTTGCATATGCCCTCGCTTTAAATAACGATGATAAAGCTAATGAAG 3700
      A  Y  L  R  S  R  L  D  S  I  T  N  Y  Y  T  M  A  I  V  A  Y  A  L  A  L  N  N  D  D  K  A  N  E

3701 CAAATGAAAAACTTAAAAATATGTCATATTACAGTGAAGAAAATAATATCCGATATTGGAGCTGGAAAAAATTACATGATTCTGACTTATATAGACCATG 3800
     A  N  E  K  L  K  N  M  S  Y  Y  S  E  E  N  N  I  R  Y  W  S  W  K  K  L  H  D  S  D  L  Y  R  P  W

3801 GCTGTATAGAAGCAAACCTTGGAGCGGAGATATCGAAGCCACTGCCTACGCGCTTCTTACTCAATTACAACTCAATAATATCAATTACAGTCATCCTATA 3900
       L  Y  R  S  K  P  W  S  G  D  I  E  A  T  A  Y  A  L  L  T  Q  L  Q  L  N  N  I  N  Y  S  H  P  I

3901 GTCAACTGGTTAAATCAACATCGTTCGTATTATGGATTTTTGTCATCAACGCAGGATTCAGTTGTTACTCTACAGGCTTTGACACAATATAGTGTGAAAG 4000
      V  N  W  L  N  Q  H  R  S  Y  Y  G  F  L  S  S  T  Q  D  S  V  V  T  L  Q  A  L  T  Q  Y  S  V  K

4001 CACGAAATCCAAAAATGGATATGCATTGTAACATTGCTAGTACTGCTTCTAGTACTGCAAGAGGAGCTTTCCACTTGACTTCGAACAATCCATTGGAACT 4100
     A  R  N  P  K  M  D  M  H  C  N  I  A  S  T  A  S  S  T  A  R  G  A  F  H  L  T  S  N  N  P  L  E  L

4101 TTACGATTTGAAGATCTCCCCACAAGCTGACCTATTCGTGGAAGCTGAAGGCACTGGCTTGGCTAGTATGTCGCTTCTTATGAGATACAATGTAGCCCAA 4200
       Y  D  L  K  I  S  P  Q  A  D  L  F  V  E  A  E  G  T  G  L  A  S  M  S  L  L  M  R  Y  N  V  A  Q

4201 GAACCCGAAAAAACCTGTAAGTTTCATTTAAACATTACTGTAGAAGAGTACGATGACATCATTCGCCCGGTTGCTCCTACTGGAGAGCTTGAAGGGATCG 4300
      E  P  E  K  T  C  K  F  H  L  N  I  T  V  E  E  Y  D  D  I  I  R  P  V  A  P  T  G  E  L  E  G  I

4301 ACATCATACCCGAAAACGTTACTCGATCTCTCTTTACACAAAATGAATTAAGAGACCGATTTGGAATAAGAGATGAAAATGAAGCTGCAGAAGATTCAGA 4400
     D  I  I  P  E  N  V  T  R  S  L  F  T  Q  N  E  L  R  D  R  F  G  I  R  D  E  N  E  A  A  E  D  S  D

4401 TGAGGAAAATAATGGAGAAAAAGTTCACGTCGTGGAACTGAATATTTGCATGCGTTATTTAGAAAAAGAAGGAAATAGTGGAATGTCTATATTAGATGTG 4500
       E  E  N  N  G  E  K  V  H  V  V  E  L  N  I  C  M  R  Y  L  E  K  E  G  N  S  G  M  S  I  L  D  V

4501 GGACTATTTACTGGATATTCAATGAAAATGGAAGAACTCAAAAATTTAATCCGCAGTATTGAAACATCTTTAACTCAATTTGAACAGAATGAACGATCGA 4600
      G  L  F  T  G  Y  S  M  K  M  E  E  L  K  N  L  I  R  S  I  E  T  S  L  T  Q  F  E  Q  N  E  R  S

4601 CTGTTTTATATTTCGACGAGGTGCCAAATAAAGAAAGAATGTGCATTAGCTTACGAACATATCAGGATTTCCATGTTGGAAAAGTACAGCCTGCATCTGT 4700
     T  V  L  Y  F  D  E  V  P  N  K  E  R  M  C  I  S  L  R  T  Y  Q  D  F  H  V  G  K  V  Q  P  A  S  V

4701 CAAGATTTATTCCTATTATGAACCAAGCAAATCGTGCACGAAGTTTTATGCACCAAGAGACAGAAGTCCCATGCTGACTAAGATATGCGAAGGAAAACAG 4800
       K  I  Y  S  Y  Y  E  P  S  K  S  C  T  K  F  Y  A  P  R  D  R  S  P  M  L  T  K  I  C  E  G  K  Q

4801 TGTTTTTGTGCAGAAGGCAAATGTCCGTCTTCTACGCCATTTAAGGAGATTAGAGCGAAAATTTCAGATACTGAAAGACGAAGAGCACTATTGGATATTG 4900
      C  F  C  A  E  G  K  C  P  S  S  T  P  F  K  E  I  R  A  K  I  S  D  T  E  R  R  R  A  L  L  D  I
                                                     C345C domain
4901 CTTGTCATAAGTCACAACATTTCATTTGGAATGTGACATTGGAGAAAATAACTTACGAAAACAGCTTTAAACTGTTTCATGTCAACGTTTTTCATGTTAT 5000
     A  C  H  K  S  Q  H  F  I  W  N  V  T  L  E  K  I  T  Y  E  N  S  F  K  L  F  H  V  N  V  F  H  V  M

5001 GAAAGCAGGTATCCAAAGCAACAAAGAAGTTGAAGGAGAGACTGTCATCTTTCATGGGAGAAATGAATGTCGTTACCCTGAAATGACAGAGGGAAATTCT 5100
       K  A  G  I  Q  S  N  K  E  V  E  G  E  T  V  I  F  H  G  R  N  E  C  R  Y  P  E  M  T  E  G  N  S

5101 TACATTGTAATGGGAACCGACGGTTATCCAGTGACAACTGAGGATGGGGAAATCAAATTCAAATATATGTTCGACAGACATTCCCGTATTTATCATTCGA 5200
      Y  I  V  M  G  T  D  G  Y  P  V  T  T  E  D  G  E  I  K  F  K  Y  M  F  D  R  H  S  R  I  Y  H  S

5201 GATCCCTTCGAGATATTGCGGACACCAAAGGCAGAAACTTACAGAAGACGTTCAACTCGCTTTATAATCGATTTGTTATTAGAAAAGAAGGATGTGATAA 5300
     R  S  L  R  D  I  A  D  T  K  G  R  N  L  Q  K  T  F  N  S  L  Y  N  R  F  V  I  R  K  E  G  C  D  N

5301 TTAACACGTTGATTTTAAATCCAACTGATTGTGTGTAATCTTCATATAATCAACTCCTTCAATAAGATTTACTTTTTATTAACTGATATTAATCATTGAG 5400
       *

5401 AAAACGACATTCGTTGAAAGAATCGAAAAAAATTATATAAATAAAAATAAAATAAACTTCTCGTAAAAAATAATAACTTGTATTGATGAAATGTAAACCT 5500
5501 AAAGCTTAAATTGTACG  5517
```

ScsuC3-3

```
   1 AAGTTAGACAGTGATGGGATTTCAAGAAGCTTCGATCGCATTCAAAGTTCCAGCTCCATAGATAATATGAAGGGACCGCTGTTGATTCTGACTTTCTTGT 100
                                                                            M  K  G  P  L  L  I  L  T  F  L

 101 TGCTGTCATCATTGGTCCAATGTAATAACAAAAATATTCAATATACTGTAGCAGCACCAAACACGTTTCGGGTTGGAGTTGCGGAGACAGTGAGTGTTGT 200
     L  L  S  S  L  V  Q  C  N  N  K  N  I  Q  Y  T  V  A  A  P  N  T  F  R  V  G  V  A  E  T  V  S  V  V

 201 TGTTACAGGTAATCGGGAACCAGTCAGAGTGGACATTTTTCTCAAAGACAATACGGCAAATAAGGGAATCATTGTCAGCCAATCTTTAACTATTTCTGAT 300
       V  T  G  N  R  E  P  V  R  V  D  I  F  L  K  D  N  T  A  N  K  G  I  I  V  S  Q  S  L  T  I  S  D

 301 GGGAAACCTCAAATAACAACTCTTTTACTTCGTCCAGAAGATATCCCTAAAAGCCAAAAAAATTACCGAAATATTTTTGTGTATTTAAATGTCACCGAAT 400
      G  K  P  Q  I  T  T  L  L  L  R  P  E  D  I  P  K  S  Q  K  N  Y  R  N  I  F  V  Y  L  N  V  T  E

 401 CAACAGGTAAATTCAATAAAGAAGAGCGCATTCCAGTAACTAAATATTCTGGTTACCTCTTTATCCAGACAGACAAGCCTCTGTATACATCATCTGACTC 500
     S  T  G  K  F  N  K  E  E  R  I  P  V  T  K  Y  S  G  Y  L  F  I  Q  T  D  K  P  L  Y  T  S  S  D  S

 501 AGTTCATATCCGTATAATGTATGTTGATGAAAAATTGATTCCTATAACTGAAGAGGTGAAGCTGGAAGTTAAAAATCCAAATGACACCATCGTGTATTAC 600
       V  H  I  R  I  M  Y  V  D  E  K  L  I  P  I  T  E  E  V  K  L  E  V  K  N  P  N  D  T  I  V  Y  Y

 601 AAAGAAAGTCTACCATGTAAACCAAATGGCTTTACAGAAGTTGTTTTTAAACTGTCTTCTTCCCCAGTATTTGGCAATTGGTCCGTTAGTGTGTCCTACG 700
      K  E  S  L  P  C  K  P  N  G  F  T  E  V  V  F  K  L  S  S  S  P  V  F  G  N  W  S  V  S  V  S  Y

 701 GATATGAAATGGCAGCAAAAACCACAGTGGGCTTTGAAGTGAAGCAATATGTATTACCAACATTTTCAGTGACAATAACACCAAGAAAATATTTTGTCTT 800
     G  Y  E  M  A  A  K  T  T  V  G  F  E  V  K  Q  Y  V  L  P  T  F  S  V  T  I  T  P  R  K  Y  F  V  L
```

Supplementary Fig. S2 (continued)

```
 801 AAGTACTGATGATGCAATTATAGGTGATATAAAAGCAGAATATGTTTATGGAAAAGCAGTTCAAGGAGCTGTTACATATAAATATTATATTCGACAACCA  900
       S  T  D  D  A  I  I  G  D  I  K  A  E  Y  V  Y  G  K  A  V  Q  G  A  V  T  Y  K  Y  Y  I  R  Q  P

 901 TCCGGCATCCAAAGTCCTATTGGTAAACTTGTACATTACCCATTAAATAAAGATGGTACATCAACATACAATATTCCTAAGAGGGTAATAGATGAAAATC  1000
      S  G  I  Q  S  P  I  G  K  L  V  H  Y  P  L  N  K  D  G  T  S  T  Y  N  I  P  K  R  V  I  D  E  N

1001 GAATTCCTTGGTTTCCAGCAATTGATAAATCCACAGTAATTGTTGAAGCGGAAGTAACAGAGAAAGCAACAGGGAAAAAAGAATCAGCCATTAATGATGA  1100
     R  I  P  W  F  P  A  I  D  K  S  T  V  I  V  E  A  E  V  T  E  K  A  T  G  K  K  E  S  A  I  N  D  D

1101 TACTATATTTACAACAACACCTTATATGATAAGGTTCCACAGATCTTTGAAAGAATTTAAACCTGGAGTCCCTTATCAATTACAGGTTGATGTACATCAT  1200
       T  I  F  T  T  T  P  Y  M  I  R  F  H  R  S  L  K  E  F  K  P  G  V  P  Y  Q  L  Q  V  D  V  H  H

1201 ATCAACAACCGACCCTTGAAATATAAAGTCCCTGTCATAATTTCTGGTAGCGCTAAAAAATCTGGAAATGCTGTTAAAACTATTTTCACAAAAAATCTTC  1300
      I  N  N  R  P  L  K  Y  K  V  P  V  I  I  S  G  S  A  K  K  S  G  N  A  V  K  T  I  F  T  K  N  L

1301 AGACAGATTTACATGGAAGAGTTATGTTTCAAGTGGATACAGAAGATGGTTTTCAAGAATTAAACATACAAGTTGAAACAGCTGATCAAGAAATAGGAAA  1400
     Q  T  D  L  H  G  R  V  M  F  Q  V  D  T  E  D  G  F  Q  E  L  N  I  Q  V  E  T  A  D  Q  E  I  G  N

1401 TAATCAGGCAAAAGAAAAAATTTGTTGTTGTACGTTACAAAACTCCATTTCATAATTCGTACATTTGGATCAAAGTACCTGAGCATGGAAGATATTTTCAG  1500
       N  Q  A  K  E  K  F  V  V  V  R  Y  K  T  P  F  H  N  S  Y  I  W  I  K  V  P  E  H  G  R  Y  F  Q

1501 GTTGACAAATCTTTTCAAACTGTAGTGACAGTATATCCAGCAGATGAGCAAACAAAACTCTTTTTCATGGTTGTGAGTCGTGGAATTATTTTGCTGATGA  1600
      V  D  K  S  F  Q  T  V  V  T  V  Y  P  A  D  E  Q  T  K  L  F  F  M  V  V  S  R  G  I  I  L  L  M

1601 ATGAAACTGAAGCAAAAGGAAAGTATCTAGTAAGAACGATTCAGTTTCCTGTTACTGAGGATATGTCACCTAGTTTTCGTCTCATTGCTTACATTATAAA  1700
     N  E  T  E  A  K  G  K  Y  L  V  R  T  I  Q  F  P  V  T  E  D  M  S  P  S  F  R  L  I  A  Y  I  I  K

1701 GGACAACAAAATCATAGCGGATTCAGTGCAAATTGAAGTCGAAAGAGTTTGCAAATACAACGGTGGAAAAGGGTTCTCTATCAAAACAGATCGAAAAGCT  1800
       D  N  K  I  I  A  D  S  V  Q  I  E  V  E  R  V  C  K  Y  N  G  G  K  G  F  S  I  K  T  D  R  K  A

1801 GGAATTGCCGTACCAGGAGCAGCCGTTAATTTTATAATAACTGGTGAACAAGATTCTTTTATTGGACTAAGTGCTATTGATGAAGCTCTGTACTTCCTTA  1900
      G  I  A  V  P  G  A  A  V  N  F  I  I  T  G  E  Q  D  S  F  I  G  L  S  A  I  D  E  A  L  Y  F  L

1901 ATAACAGAAGTGTATTTACAAAGGAAAAGATGTTTAGAGAAATTCAGAAATATGATTTAGGTTTTGGACCAGGAGGAGGTATTGATCCTGCTGCTGTTTT  2000
     N  N  R  S  V  F  T  K  E  K  M  F  R  E  I  Q  K  Y  D  L  G  F  G  P  G  G  G  I  D  P  A  A  V  F

2001 TAAAAATGCTGGGATATTAATTCTATCCAATTCTCACATAGGACGGCATGGAAGAACTGAGGGCATTGATCATTCTCACCATCGTATGAAGCGCAGTTTA  2100
       K  N  A  G  I  L  I  L  S  N  S  H  I  G  R  H  G  R  T  E  G  I  D  H  S  H  H  R  M  K  R  S  L
                                                                              β-α cleabage site
2101 CAGAGTAAAATTGATGAGTATTTTGGGAATGCTGCTATA~~TGTTGTCGTTATGGACAGTTTGAAGGTCCCAAAGGCATGAACTGTACCACTCGAGCAGCCA~~  2200
      Q  S  K  I  D  E  Y  F  G  N  A  A  I  C  C  R  Y  G  Q  F  E  G  P  K  G  M  N  C  T  T  R  A  A
                                                                 ANA domain
2201 AGATTAAAGAACGAATGGGTGAAAAATTTAATTGTAGTGAAGCATTTTTAGACTGCTGTGAACATACAGAAGACTCAATACTAACTTTTGGAAGAAATTT  2300
     K  I  K  E  R  M  G  E  K  F  N  C  S  E  A  F  L  D  C  C  E  H  T  E  D  S  I  L  T  F  G  R  N  F

2301 TCGCTATGAAAAAGTTGAAAAGCCAGTGGAAAACCTTGTGGATGAACTCAAAGAAGACGATATAAAAAATATTCGTCGTTTTTTCCCTGAAAAGTGGTTG  2400
       R  Y  E  K  V  E  K  P  V  E  N  L  V  D  E  L  K  E  D  D  I  K  N  I  R  R  F  F  P  E  K  W  L

2401 TTCGATACGTATCATATTGGCAACCAGAAAGATTGTAAAAATGAAAAGGATATTTGCATAGTTAGTGCAACAGCACCTGACAGCATAACAAAATGGGTGG  2500
      F  D  T  Y  H  I  G  N  Q  K  D  C  K  N  E  K  D  I  C  I  V  S  A  T  A  P  D  S  I  T  K  W  V

2501 TTCAAGCAGTCGGTATTTCAAAGAAAACAGGCATGTGCATTGCAGAACCATTAGAATTAATTGTTTTCAAATCAATGTTCGTTCAGTTGTCTTTACCAGC  2600
     V  Q  A  V  G  I  S  K  K  T  G  M  C  I  A  E  P  L  E  L  I  V  F  K  S  M  F  V  Q  L  S  L  P  A

2601 TGTTGCAATGAGGGGAGAACAAATCGAAATTCTTGCTACAGTTTTTAATTATGATTCTTCTGATTTAGATGTATTTGTATATTTGTATGGAGTAGAAGGA  2700
       V  A  M  R  G  E  Q  I  E  I  L  A  T  V  F  N  Y  D  S  S  D  L  D  V  F  V  Y  L  Y  G  V  E  G

2701 CTTTGCACTGGAGCAATTCCTGGAGAGAGAACAGAACCAAAAAGAATCAAAGTGCTGGGTAACAGTGCGTCCACTGTAACATTTCCTATAATGCCTTTAA  2800
      L  C  T  G  A  I  P  G  E  R  T  E  P  K  R  I  K  V  L  G  N  S  A  S  T  V  T  F  P  I  M  P  L

2801 CAGTGAGCGATTTCAATATACAAGTTTTAGCATCCGCTGGACATTTATATGATGCAGTCATAAAAGTTCTCAAAGTCGTACACGAAGGAATTCCAACCGA  2900
     T  V  S  D  F  N  I  Q  V  L  A  S  A  G  H  L  Y  D  A  V  I  K  V  L  K  V  V  H  E  G  I  P  T  E

2901 AAAGACCATATCATTTCCATTGGATCCAGAAGGAAAAATTAATAAAAGAAAAAAGAGAGCTATTGATCAGAGTATATCAGAAGTCTATAATGAAGTTCAG  3000
       K  T  I  S  F  P  L  D  P  E  G  K  I  N  K  R  K  K  R  A  I  D  Q  S  I  S  E  V  Y  N  E  V  Q

3001 AAGAGGCAAGAGATAACTGTTGACGTAACATTTCCACATGATCACATCAAAGGAACGGAAAAATGCTTTGTTAATTTCATTGGTGATCCTGTTGGTCAAG  3100
      K  R  Q  E  I  T  V  D  V  T  F  P  H  D  H  I  K  G  T  E  K  C  F  V  N  F  I  G  D  P  V  G  Q

3101 CCGTCAATGTTACACTCTCTGGAGTTGAAGAAGAATTCCTCAAGCTTCCACAAGGATGTGGAGAACAAACTATGATCAAATTAGCTCCACTTGTTTCTAC  3200
     A  V  N  V  T  L  S  G  V  E  E  E  F  L  K  L  P  Q  G  C  G  E  Q  T  M  I  K  L  A  P  L  V  S  T
                                                     Thioester site
3201 AATGCATTATCTGAAAAAAACAAATCAGTTTTCAGCAACAGCAGAAAAAAAGGGCTATGATCTTATATGGAAAGGATACGATAACATGCAAAAATTTAAA  3300
       M  H  Y  L  K  K  T  N  Q  F  S  A  T  A  E  K  K  G  Y  D  L  I  W  K  G  Y  D  N  M  Q  K  F  K

3301 AAGAACGATGGATCTTATTCCATATTTACAGAAAGTCCTTCTAGTACTTGGTTGACTGCGTTTGTTATAAAAGTTTATTGCCAAGCTTCTGAGTTCATAC  3400
      K  N  D  G  S  Y  S  I  F  T  E  S  P  S  S  T  W  L  T  A  F  V  I  K  V  Y  C  Q  A  S  E  F  I

3401 AGATACCTCTTGAGAATATCGGAACTGCAGTTGAATGGTTGATTAAGCACCAGTCATCATCTGGCAAATTTTTTGATGATTACAAAGTACATTCAAATAC  3500
     Q  I  P  L  E  N  I  G  T  A  V  E  W  L  I  K  H  Q  S  S  S  G  K  F  F  D  D  Y  K  V  H  S  N  T

3501 TATAGCAGGCGGTCTTAACGGAGATGTTACACTCACAGCTTATGTATTAATAGCAATGATGGAATGCAATCGGTACAACTCAGTGACTAAAAAAACAGCT  3600
       I  A  G  G  L  N  G  D  V  T  L  T  A  Y  V  L  I  A  M  M  E  C  N  R  Y  N  S  V  T  K  K  T  A

3601 GTTCAAAGAGCCATAACATTTCTCGAAAAAAATATTGATCATACAAGACAACCATATAATCTTGCTATTGTTGCTTATGCTCTTTCTTTGACTGAAAATA  3700
      V  Q  R  A  I  T  F  L  E  K  N  I  D  H  T  R  Q  P  Y  N  L  A  I  V  A  Y  A  L  S  L  T  E  N

3701 GACGACGCTATGATGTAAATGAAGATTTAAAAAATATAGCTAAATTCATATCAGGTAGTAACGTACGCTATTGGAATTGGGATTCGTTAGAATTTGGAAC  3800
     R  R  R  Y  D  V  N  E  D  L  K  N  I  A  K  F  I  S  G  S  N  V  R  Y  W  N  W  D  S  L  E  F  G  T
```

Supplementary Fig. S2 (continued)

```
3801  AGGAAATGTACCATGGATTTATCAAAAGAAGGCTGATGCAGCAGCAGTAGAAACGACAAGTTATGCTTTACTTGCACAACTACAGTTTGATGAGATTGAT  3900
        G  N  V  P  W  I  Y  Q  K  K  A  D  A  A  A  V  E  T  T  S  Y  A  L  L  A  Q  L  Q  F  D  E  I  D

3901  TATTCTCATCCCATTGTTAATTGGTTGACCAGACAACGATCGTCTGGTGGTGCTTTTGTTTCTACACAGGATACTGTGATCACTCTGCAAGCTCTGGCAG  4000
       Y  S  H  P  I  V  N  W  L  T  R  Q  R  S  S  G  G  A  F  V  S  T  Q  D  T  V  I  T  L  Q  A  L  A

4001  AATATAACACAAAAACCAAAATTCCTCTTGTGGATATGCAATGCAACATTACCAGTTCTGAGACCAGTCGGTTTAAAAGGAGTATCCAACTCACTAAAGA  4100
      E  Y  N  T  K  T  K  I  P  L  V  D  M  Q  C  N  I  T  S  S  E  T  S  R  F  K  R  S  I  Q  L  T  K  D

4101  TAAAGCTCAAAATATTGAAGAGATTGAGGTGCCACCAAAAGGTAAACTATATGTAGATGTTGGAGGGAAAGGTATTGGTTCAATGTCTCTTTCTTTACAA  4200
        K  A  Q  N  I  E  E  I  E  V  P  P  K  G  K  L  Y  V  D  V  G  G  K  G  I  G  S  M  S  L  S  L  Q

4201  TACAATTCAGAATATACTCCGGAAACAAAATGTGAATATGATTTGATAGTCAAAACTCATGAATATAGAGATTCGTTTCAACCTCCTAACCTTGCTGAAT  4300
       Y  N  S  E  Y  T  P  E  T  K  C  E  Y  D  L  I  V  K  T  H  E  Y  R  D  S  F  Q  P  P  N  L  A  E

4301  ATGCAAACGGACCAATTCCTGATAATTTAAAAGAGGAGATTATTAACAAGAATGTTCAAGAAGTGTTTAATGAAGCAGTAAATAGGAAAAGAGATGGTAG  4400
      Y  A  N  G  P  I  P  D  N  L  K  E  E  I  I  N  K  N  V  Q  E  V  F  N  E  A  V  N  R  K  R  D  G  S

4401  TCACGATGAAGAAAAGGACAATGCAAATAACCAGGAACATTTAATCAACATAAACATTTGTGTTAAATATCAAGATAAAGAAAAGGCAGGAATGTCAATT  4500
        H  D  E  E  K  D  N  A  N  N  Q  E  H  L  I  N  I  N  I  C  V  K  Y  Q  D  K  E  K  A  G  M  S  I

4501  TTGGAAATAGGATTTCTAACAGGTTACAGAATTGATAAAGAAGAATTGTCCAAATTAGAAAACAGACCAAAAGTAAAATGTGTTGAAACTTCAGACAGAG  4600
       L  E  I  G  F  L  T  G  Y  R  I  D  K  E  E  L  S  K  L  E  N  R  P  K  V  K  C  V  E  T  S  D  R

4601  CTTTGATTTTATATTTGGAAGAGGTTCCCAATGATAGAACAATATGTTTAGATGTGAAATTAAGAAAATTCATAACTGTTGGATTAGTTCAGCCCACGAC  4700
      A  L  I  L  Y  L  E  E  V  P  N  D  R  T  I  C  L  D  V  K  L  R  K  F  I  T  V  G  L  V  Q  P  T  T

4701  TGTAAAAATTTATAATTATTATAAACTAGATAAATCATGCACAACGTTCTACGGACCAGATGAGGATAGTGTTATGTTGCAAACAATTTGTGAAGGAAAG  4800
        V  K  I  Y  N  Y  Y  K  L  D  K  S  C  T  T  F  Y  G  P  D  E  D  S  V  M  L  Q  T  I  C  E  G  K

4801  CAATGCAGATGCATGGAAGGAGCATGTCCTCCTCTTAATCCGTTCAAACATGTGTGGGAAAAGAACGAGGAGAAAGAACGTAATAAAGAATTACTACGCA  4900
       Q  C  R  C  M  E  G  A  c  P  P  L  N  P  F  K  H  V  W  E  K  N  E  E  K  E  R  N  K  E  L  L  R
                                                  C345C domain
4901  TCATCTGTGACGATAAAAAAACAAAAGATAATTTTGTATGGCTTGGAATACTGAAGAGCG  4960
      I  I  c  D  D  K  K  T  K  D  N  F  V  W  L  G  I  L  K  S
```

ScsuA2M

```
   1  ATGACGACTGAAAAGATGGCAAGAGGAACGCGGACCGACCCGAAAAAGTTTTTCCTGTTCCTCGTTTGCTTTGTATTTTGTACAAGAGCAGCTGAAGAAC  100
                      M  A  R  G  T  R  T  D  P  K  K  F  F  L  F  L  V  C  F  V  F  C  T  R  A  A  E  E

 101  AGGACAAAAGAGGTTTTGTATTAACAGCACCAAAACATTTACTGGCTGAAACAGTGGAACATATATGTTTATTTTTTCACAATATTAATTATGACGGAGA  200
      Q  D  K  R  G  F  V  L  T  A  P  K  H  L  L  A  E  T  V  E  H  I  C  L  F  F  H  N  I  N  Y  D  G  E

 201  AATTTGGCTTGAACTTCTGTCTGAAAATTCAACTGTCATTAGTACTTCATCACAAAAAATTTTTAAGGGAAAAGGAGAATGCATTGAAATGTTTATTCCG  300
        I  W  L  E  L  L  S  E  N  S  T  V  I  S  T  S  S  Q  K  I  F  K  G  K  G  E  C  I  E  M  F  I  P

 301  AATCTATTCATCGGTAATGCAAAATTATCTGTGCGTGGGATATTCCCAACAGAAGATGGCAGTGATCCATACAACATCCATCAAACGAAATCTGTTATTA  400
       N  L  F  I  G  N  A  K  L  S  V  R  G  I  F  P  T  E  D  G  S  D  P  Y  N  I  H  Q  T  K  S  V  I

 401  TTAAACATTATAATTCTTTAGTTTTCATTCAAACTGATAAACCAGTTTACAAACCTGGAGAAAAAATAAGATTTCGAATTTTGCATGTGACAATGGATTT  500
      I  K  H  Y  N  S  L  V  F  I  Q  T  D  K  P  V  Y  K  P  G  E  K  I  R  F  R  I  L  H  V  T  M  D  L

 501  AAAACCAATTTCAGATGAGATTCCATCAATTTGGATAGAAGATCCATCAGGAGTACGCATTTCACAGTGGCTGAATGAAAAACCTGAACTTGGTTTAATT  600
        K  P  I  S  D  E  I  P  S  I  W  I  E  D  P  S  G  V  R  I  S  Q  W  L  N  E  K  P  E  L  G  L  I

 601  GATTTACAAATGTCTTTATCCACAGAACCAATTTTAGGAAAATGGAATATTAAAGCAAGTATAGGAAAATTGCGAAAAACTCAAAGTTTTGATGTAGAAG  700
       D  L  Q  M  S  L  S  T  E  P  I  L  G  K  W  N  I  K  A  S  I  G  K  L  R  K  T  Q  S  F  D  V  E

 701  AATATGTTCTTCCCAAATTTGAAGTCAAAATTTCACCACCACCATTTATACTAGCCAATCAATTAAATGCAGTATGGAATGTCTGTGCTCATTACTCATA  800
      E  Y  V  L  P  K  F  E  V  K  I  S  P  P  P  F  I  L  A  N  Q  L  N  A  V  W  N  V  C  A  H  Y  S  Y

 801  TGGTAAACCCGTTCAGGGATATGCTGTTATCAAAGCTGTACTTGGAAGCATTGAATTTCCTACAAAAACCGTTGCCCAATACGAAGGCAAGATCACAGGT  900
        G  K  P  V  Q  G  Y  A  V  I  K  A  V  L  G  S  I  E  F  P  T  K  T  V  A  Q  Y  E  G  K  I  T  G

 901  TGTCATTCAATCATAGGAAAAATTCAACAGAAATTACGTAGCTATGGCCCATTTCAAATAGAAATCCATGCTGAAGTTACTGAACTGGGAACAAATGCAA  1000
       C  H  S  I  I  G  K  I  Q  Q  K  L  R  S  Y  G  P  F  Q  I  E  I  H  A  E  V  T  E  L  G  T  N  A

1001  CAATGACTGCCTTAGCTCAATCGAAAATTCATCTTGAGGCACTGATACTTGATATGTCATCGTATATGCCGTTCTATTTTAAACCTGGATTACCATTTCA  1100
      T  M  T  A  L  A  Q  S  K  I  H  L  E  A  L  I  L  D  M  S  S  Y  M  P  F  Y  F  K  P  G  L  P  F  H

1101  TGGCAAAATTAAAGTGACAACTCCTGATCATAATTCTGCATCGAACATAAGTGTTGAAATAAGTCTTGAAACTCGTACTAAGCGAGGAATTTCACAAATA  1200
        G  K  I  K  V  T  T  P  D  H  N  S  A  S  N  I  S  V  E  I  S  L  E  T  R  T  K  R  G  I  S  Q  I

1201  GTTGCTTCAGAGTCTTTCAAGTCAGATGATAATGGGATTATTTCATTTATCACAATACCACCTGTTGTCAGTGATACCGAAAACGTGATCATTAAGGCCA  1300
       V  A  S  E  S  F  K  S  D  D  N  G  I  I  S  F  I  T  I  P  P  V  V  S  D  T  E  N  V  I  I  K  A

1301  AAATATTACCAGTATCATCTGATTCGGGGCCCTCTGATCATGATCATTCTTACTTTTTACCTAATCATAATTTTGGCCCTACTTCCACAATATCTGCTCC  1400
      K  I  L  P  V  S  S  D  S  G  P  S  D  H  D  H  S  Y  F  L  P  N  H  N  F  G  P  T  S  T  I  S  A  P

1401  AATTTGGTACTCACCGAGTGGTAGTTTTCTTCAGATTCATAGACCTGCAGATAATCTTCCTTGTAATGGGGAATATTCATTTACTGTTTTCTACAGTGAA  1500
        I  W  Y  S  P  S  G  S  F  L  Q  I  H  R  P  A  D  N  L  P  C  N  G  E  Y  S  F  T  V  F  Y  S  E

1501  CTTTCTCAAAATATTCACTTATCATATCAGGTAATTTCTCGAGGCAACATTATCTATCAATCATCAAAGCTTCAGATACTCGAAGAAAATTCTCTCAATT  1600
       L  S  Q  N  I  H  L  S  Y  Q  V  I  S  R  G  N  I  I  Y  Q  S  S  K  L  Q  I  L  E  E  N  S  L  N

1601  TCAGTAATTCATGGCATCTTCTTCCTAGCAGGAATAAATTATCATTTAATTTTTTCACCATCCCAATTCCTCTCACCCCAAAAATGACTCCCGTAAGCAG  1700
      F  S  N  S  W  H  L  L  P  S  R  N  K  L  S  F  N  F  F  T  I  P  I  P  L  T  P  K  M  T  P  V  S  R
```

Supplementary Fig. S2 (continued)

```
1701  ATTGCTTGTATTTTATGTTCGAGATGATGGAGAAGTTGTAGCTGACAGTCTTAAATTTGAAATAGAAAAATGTACTGAAAATGATGTCAGTCTTAATTTT  1800
        L  L  V  F  Y  V  R  D  D  G  E  V  V  A  D  S  L  K  F  E  I  E  K  C  T  E  N  D  V  S  L  N  F

1801  GACACCCAAAGGGTAATTCCAGCTACTCATACAGAAATATCTATCAAAGCTTCTCCTTATTCATTATGTGCTGTGGGGATTGTTGATAGAGCTGTTCATT  1900
       D  T  Q  R  V  I  P  A  T  H  T  E  I  S  I  K  A  S  P  Y  S  L  C  A  V  G  I  V  D  R  A  V  H

1901  TTTTACGAGCTAATAATCAACTGACATTAACAAAGATTTTTAATGGTTTGTCTGCCTTTGATATTACCAAGGATAGTTTACCAGAACAGAGCAAAGTGAA  2000
      F  L  R  A  N  N  Q  L  T  L  T  K  I  F  N  G  L  S  A  F  D  I  T  K  D  S  L  P  E  Q  S  K  V  K

2001  ATATTGCCATCAACATTTTGATAGTTATCCTCTCCATCCATTTGTGGACAACAGTTTGATCAGGGTGATGGAAGTAGAATCAGGCTATGCAGATGCAGCT  2100
        Y  C  H  Q  H  F  D  S  Y  P  L  H  P  F  V  D  N  S  L  I  R  V  M  E  V  E  S  G  Y  A  D  A  A

2101  ACAGCATTTGATGATACAGGAGTTATTATAATGTCTGATCTTACATTTGATACTCGACCTTGTGTAGATGTTACTGGAATTATGGCTTTGGCTCGGACTT  2200
       T  A  F  D  D  T  G  V  I  I  M  S  D  L  T  F  D  T  R  P  C  V  D  V  T  G  I  M  A  L  A  R  T
                                                                                  Bait region
2201  TTGCCATGCCAATATCACTAGAATTCCAACAACCAGGGCCACCAGGTCCACAACGGCAGAGCAGAGAGTTTTCACCACCACTTTTTGAACTTCCTGTGAT  2300
      F  A  M  P  I  S  L  E  F  Q  Q  P  G  P  P  G  P  Q  R  Q  S  R  E  F  S  P  P  L  F  E  L  P  V  I

2301  AACATTTAGGAATAAAGAAAAACAAATTTTCGAAAAGAAGCAGAAAAAATCTGCTGTGGAAATTAGAAATTTTTTCCTGAGACTTGGTTATGGGAATTA  2400
        T  F  R  N  K  E  K  Q  I  F  E  K  K  Q  K  K  S  A  V  E  I  R  N  F  F  P  E  T  W  L  W  E  L

2401  CACAAAATTGGAAATTCAGGTCAACAGAAAGTTCCATTAAAAGTTCCTCACAGTATAACACACTGGGTAGGCAATGCTTTTTGTATATCATCTTATGCAG  2500
       H  K  I  G  N  S  G  Q  Q  K  V  P  L  K  V  P  H  S  I  T  H  W  V  G  N  A  F  C  I  S  S  Y  A

2501  GCATTGGTGTAGCATTGCCTGCACATTTGAAAGCATTTCAACCTTTTTTTATGACATACACATTGCCTTATTCTGTGAAAAGAGGAGAAACCATGAAACT  2600
      G  I  G  V  A  L  P  A  H  L  K  A  F  Q  P  F  F  M  T  Y  T  L  P  Y  S  V  K  R  G  E  T  M  K  L

2601  TCTTGTTTCATTATCTAACTATTTGCAAGAATGTTTACCAATAAAAGCTACATTACAGGAGAGCAAAGTTTTCAGCATTGCTAGTGAAACTACTTTTCAA  2700
        L  V  S  L  S  N  Y  L  Q  E  C  L  P  I  K  A  T  L  Q  E  S  K  V  F  S  I  A  S  E  T  T  F  Q

2701  AAAATTTGTTTATGTAAAGAGGAAACAAAAACAATCCAATTTTTGATAAGGCCAAAAAGTATTGGTAAGATGAATATTACTGTATATGCTTACACTTCAA  2800
       K  I  C  L  C  K  E  E  T  K  T  I  Q  F  L  I  R  P  K  S  I  G  K  M  N  I  T  V  Y  A  Y  T  S

2801  CAAAAAACATTTGTGACAGTAATACTCCTGATCCTGTGAGTAATGAACAAGCCCGGGATGCTGTAACTAAATCACTTTTAGTGGAGCCTGAAGGTTTTCC  2900
      T  K  N  I  C  D  S  N  T  P  D  P  V  S  N  E  Q  A  R  D  A  V  T  K  S  L  L  V  E  P  E  G  F  P

2901  AAAAGAAGATACTTGGACTTCATTAATATGTCTTAATGATTCTGAAAGTAATAATGATACTGAATTTGAGGATGTTATAAAGGAATCAGTTGTTTTGTCA  3000
        K  E  D  T  W  T  S  L  I  C  L  N  D  S  E  S  N  N  D  T  E  F  E  D  V  I  K  E  S  V  V  L  S

3001  ATATCAGACACACAAGTAGTACCAGGATCTGTAAGGGCTTACATAACAGTAATTGGTGATACAATGGGTCCTAGTTTACAGGGATTAGATCATTTGGTTC  3100
       I  S  D  T  Q  V  V  P  G  S  V  R  A  Y  I  T  V  I  G  D  T  M  G  P  S  L  Q  G  L  D  H  L  V

3101  GTTTACCAGTTGGTTGTGGTGAACAGAATATGGTTCTGTTTGCACCTAACATATATGTATTACATTATCTTAAGAATACAAACCAATTGACAACAGCTAT  3200
      R  L  P  V  G  C  G  E  Q  N  M  V  L  F  A  P  N  I  Y  V  L  H  Y  L  K  N  T  N  Q  L  T  T  A  M
                  Thioester site
3201  GGAAAATAAGATAATATCTCATTTGAAGACAGGGTACCAAAGAGAATTAACATACAGGCGTGAAGATGGATCTTTCAGTGCATTTGGTAAAAGTGACAGA  3300
        E  N  K  I  I  S  H  L  K  T  G  Y  Q  R  E  L  T  Y  R  R  E  D  G  S  F  S  A  F  G  K  S  D  R

3301  GAAGGAAGCATTTGGTTAACAGCATTTGTTGTTAAATCTTTTGCACAAGCAAGAGAATTTATTTTCATTGATGATACAATCATGGATGAAAGCATATCTT  3400
       E  G  S  I  W  L  T  A  F  V  V  K  S  F  A  Q  A  R  E  F  I  F  I  D  D  T  I  M  D  E  S  I  S

3401  GGATTACAAATAAGCAAATGGAAAATGGATGTTACTCTTTTGTAGGGAAAGTGCTGCATAAAGACATGAAGGGTGGAGTAAGTGATTCTGAAAGCTCTTT  3500
      W  I  T  N  K  Q  M  E  N  G  C  Y  S  F  V  G  K  V  L  H  K  D  M  K  G  G  V  S  D  S  E  S  S  F
                                                      Catalytic histidine
3501  TGCTGCTCTAACAGCCTATATCACCATAGCATTGTTGGAAGCGGGAATAGCAAATGACAGTAAACCCATTGTTAATGCATTTTTTTGTTTGAAAGCAGAA  3600
        A  A  L  T  A  Y  I  T  I  A  L  L  E  A  G  I  A  N  D  S  K  P  I  V  N  A  F  F  C  L  K  A  E

3601  AAAGAACCCGATATATATACTCTTGTGTTATCTACATATGCTTCTATTCTTGCTAAAGATGAAAACTATACTTCCTTACTCATGAAACGACTGCTTGGAC  3700
       K  E  P  D  I  Y  T  L  V  L  S  T  Y  A  S  I  L  A  K  D  E  N  Y  T  S  L  L  M  K  R  L  L  G

3701  TCGGTATTTCTAAAGATAATTTATTGTTTTGGGAAAAGCAGTCTAAAAAATCTCTTGCACTTAATGTTGAAATGAGTGCATATGCATTACTTTCATTAGT  3800
      L  G  I  S  K  D  N  L  L  F  W  E  K  Q  S  K  K  S  L  A  L  N  V  E  M  S  A  Y  A  L  L  S  L  V

3801  TTCTTTGGGTGATCAAGAAAGTATATTGAAAGCACAGAAAGTATTTCGTTGGATTACACAACAAAGAAATTCGCATGGTGGTTTTATTTCAACACAGGAC  3900
        S  L  G  D  Q  E  S  I  L  K  A  Q  K  V  F  R  W  I  T  Q  Q  R  N  S  H  G  G  F  I  S  T  Q  D

3901  ACTGTTCTTGCATTACAAGCTCTGTCTGAGTTTGCTGGCAAATTTCAGTCTAATGAGCTTGAAATAGAAATTTCAGTCGAAGCTGGAAAACTTAATCATG  4000
       T  V  L  A  L  Q  A  L  S  E  F  A  G  K  F  Q  S  N  E  L  E  I  E  I  S  V  E  A  G  K  L  N  H

4001  TTTTTGAAGTAAACAATGAAAATAAGCTCGTACAGCAGATTATTAAGATACCAGAAGTACCAACTGTAGTAGACTTTATTGCTCTGGGAAAAGGTTGCTC  4100
      V  F  E  V  N  N  E  N  K  L  V  Q  Q  I  I  K  I  P  E  V  P  T  V  V  D  F  I  A  L  G  K  G  C  S

4101  TATTTTGCAGACTGTGCTTAAATATAATGTTGAACATACAGAAGGCAGTGATGCATTTAACTTAGATATTCGTTCAGAAAATATTGGTACAACTAGTGCT  4200
        I  L  Q  T  V  L  K  Y  N  V  E  H  T  E  G  S  D  A  F  N  L  D  I  R  S  E  N  I  G  T  T  S  A

4201  GCTTGTAAAAGACATCGTTTAGAAATTTGTGCAAGGTATTTGCTAGAAGATGAATTTTCTAATATGATAGTAATCGAGATAAAAATGGTTTCTGGTTTTG  4300
       A  C  K  R  H  R  L  E  I  C  A  R  Y  L  L  E  D  E  F  S  N  M  I  V  I  E  I  K  M  V  S  G  F

4301  AACCTGATAAAAAATCTTTAGCTGAGTTATTAGAGAAGAAAGATATAAAACTGAAACGATGGGACACAGAAGGGGATCAACTCAACTTATATTTTGACCA  4400
      E  P  D  K  K  S  L  A  E  L  L  E  K  K  D  I  K  L  K  R  W  D  T  E  G  D  Q  L  N  L  Y  F  D  Q

4401  ATTGAATGCTCAAGAAAAATGTTTCTCAATCAGCATTACTGAAAAAGTTGAAGTTAAAGATACCAAACCAGCAATTGTTACAATTTACGATTACTATCAA  4500
        L  N  A  Q  E  K  C  F  S  I  S  I  T  E  K  V  E  V  K  D  T  K  P  A  I  V  T  I  Y  D  Y  Y  Q

4501  CCAGAATTGTTTGTCAGGAAGAATTACAGTATCGAAGGATGTAACAAGGAAACTTTAGCTCCATTTACTGATGAAGAATTGACAGAATTAGAATCATTAA  4600
       P  E  L  F  V  R  K  N  Y  S  I  E  G  C  N  K  E  T  L  A  P  F  T  D  E  E  L  T  E  L  E  S  L

4601  TCGTCCAAGGATTAGATGATTTTGAAAAACCATCCAGTAAAAGCAGTCAAACATCCACTGTCTCAATTCCTGGAGAAAATGGAAAACATCCAGTTGAGGA  4700
      I  V  Q  G  L  D  D  F  E  K  P  S  S  K  S  S  Q  T  S  T  V  S  I  P  G  E  N  G  K  H  P  V  E  E
```

Supplementary Fig. S2 (continued)

```
4701  AATTATTCCCGACCCTGTGTACATCCCTCCTTTAGGCACCCAGGAAAAGGAAGAATTCAAATCCTCAAAAGAACAATTCAAAAACTTTAAATCAGTTGAA  4800
        I  I  P  D  P  V  Y  I  P  P  L  G  T  Q  E  K  E  E  F  K  S  S  K  E  Q  F  K  N  F  K  S  V  E

4801  TGGGAATTAGATTTTCCTGATGGAATTGATGGACCACCACCTCATCATCATCCAACACCAAATGATTTAAATAACGGAGAAAATACTACTTAACAAATTT  4900
       W  E  L  D  F  P  D  G  I  D  G  P  P  P  H  H  H  P  T  P  N  D  L  N  N  G  E  N  T  T  *

4901  TAATAGCTTTTTGTAAATTAGTCTTTCTTTTATTGTTTAAAATTTTATCTTAGAAAACATCGTATATTTTTGCACATTTTTCATGCATCACTTAGTAAAG  5000
5001  AAATAAATTAAGTACTCTAGCGACATTAAGTTACAAAAAAAAAAAA  5046
```

ScsuiTEP/CD109-1

```
   1  TAGAATCTTTACTGAAATTTTCTTAAAAGCTAAGGGAGTGTAAAAGAGAAATTTCATGATAAAGGATAACAACTCAAAACCATGGTTAAACTTACCTGCT  100
                                                                                        M  V  K  L  T  C

 101  TTACCATAATTTTGCTATTTCTCACATTTTCTGGGGGAAACGGAACATCGTACTATGCAATAACAGCACCGGAAATTCTTCGACCGAATGCTCTTTATCG  200
      F  T  I  I  L  L  F  L  T  F  S  G  G  N  G  T  S  Y  Y  A  I  T  A  P  E  I  L  R  P  N  A  L  Y  R

 201  AGTTTTCGTAACTGTTTTCGAAGTGACTGATCCAGTTATTGTCGAAGCTACTTTAAAACCTGCTTCTCATTTGATAAATGAAAATTTGACAGCTATTGCT  300
        V  F  V  T  V  F  E  V  T  D  P  V  I  V  E  A  T  L  K  P  A  S  H  L  I  N  E  N  L  T  A  I  A

 301  TCTATTAACAGTGGAAAAACTGAAGAAATAAATTTTCAAGTGGGCAATTGGAATCATCAATTATATCGTTTAGAGATTAAAGGATCAGGAGGAATCACTT  400
       S  I  N  S  G  K  T  E  E  I  N  F  Q  V  G  N  W  N  H  Q  L  Y  R  L  E  I  K  G  S  G  G  I  T

 401  TTCGTGAAACAGCTACAGTAAAGCAGTCTTTTAAAATCTTTTCTATTTTCATTCAAACTGATAAAGGAATTTACACTCCAAGTCAAACAGTTTATTTTCG  500
      F  R  E  T  A  T  V  K  Q  S  F  K  I  F  S  I  F  I  Q  T  D  K  G  I  Y  T  P  S  Q  T  V  Y  F  R

 501  TATAATTGTTACCAAACCATCGCTTCTTCCTCACAATCCGAATGGACTCGTGGTATACATAAATGATCCTGAAGCCAATAGAATTAAACAATGGACAAAT  600
        I  I  V  T  K  P  S  L  L  P  H  N  P  N  G  L  V  V  Y  I  N  D  P  E  A  N  R  I  K  Q  W  T  N

 601  GTTACATTTACTCATGGAATTTATTCTGGAAAATTTCAACTAGCCGATCAAGTTAACTTTGGCAAATGGAAAATTTCTGCTGATTTTCAGGGTCAAACGG  700
       V  T  F  T  H  G  I  Y  S  G  K  F  Q  L  A  D  Q  V  N  F  G  K  W  K  I  S  A  D  F  Q  G  Q  T

 701  AATCTGTCCACTTCACCGTAGAAGAATATGTTTTACCAAAATTTGAAGTAATTATTCAATTACCACCATTTGTCACATGGGATGATACAGATGTAACCGC  800
      E  S  V  H  F  T  V  E  E  Y  V  L  P  K  F  E  V  I  I  Q  L  P  P  F  V  T  W  D  D  T  D  V  T  A

 801  ATTTATAGAAGCAAGATATACGTATGGAAAACCTGTCAAAGGTCAGTTAACTCTTAAAGTGACAGTACCATTCTTTTGGTGGACCAGAGACAGTGAAAAG  900
        F  I  E  A  R  Y  T  Y  G  K  P  V  K  G  Q  L  T  L  K  V  T  V  P  F  F  W  W  T  R  D  S  E  K

 901  AGGATAGCTGTATTTAACACTTCTATCGATGGAAAAGCTGAAGTGAAACTTAATATAGTCAGAGATTTACTTCTCAATCAAACGTACTCATTTACGCGAG  1000
       R  I  A  V  F  N  T  S  I  D  G  K  A  E  V  K  L  N  I  V  R  D  L  L  L  N  Q  T  Y  S  F  T  R

1001  AAATAAAATTTCACGCAAAGGTAACCGAGGCTCTTACTGGAAAGCAAATGAATGCGACAAATGAATTGCAAATTCATCTTTCCAAATACAAATTGAATTT  1100
      E  I  K  F  H  A  K  V  T  E  A  L  T  G  K  Q  M  N  A  T  N  E  L  Q  I  H  L  S  K  Y  K  L  N  F

1101  TTCTAATTTAAATAAATTCAAACCTGGTCTACCTTATACAGTATATCTTAATCTTGTTTTGCAAGATGGAACTCCATTCAGTGATAACGTTCACTTAATT  1200
        S  N  L  N  K  F  K  P  G  L  P  Y  T  V  Y  L  N  L  V  L  Q  D  G  T  P  F  S  D  N  V  H  L  I

1201  ACTGTAAACTGTACTTTCCACAAAAAGGATATTTGGGATCGATTTTCAGTTTACGATCATTCAAAAAATGTAATGAATACAATGAATTTCTCAGTTCCAA  1300
       T  V  N  C  T  F  H  K  K  D  I  W  D  R  F  S  V  Y  D  H  S  K  N  V  M  N  T  M  N  F  S  V  P

1301  TAGATGGAAGAGTCAAAATTCAAATTCTTCCTCCAAAATCATCTAAGTCAATTACGTTTAAGGCAAAATTTTTGGATGTTTTTACTGTCAGCTTGGCAAG  1400
      I  D  G  R  V  K  I  Q  I  L  P  P  K  S  S  K  S  I  T  F  K  A  K  F  L  D  V  F  T  V  S  L  A  R

1401  AAGAGCTCCGATGCATCCCAATTCGGATAGATTTATTCAACTATTACCACCTTTAACAGTTCCTCAAATTGGTGAAAAAATAGAATTACTTCTCCAAAGC  1500
        R  A  P  M  H  P  N  S  D  R  F  I  Q  L  L  P  P  L  T  V  P  Q  I  G  E  K  I  E  L  L  L  Q  S

1501  ACACATACTCTCGATGATCCTTTAATATTGCAGGTCATGGGTAGAGGGAAGATTTTGCTCACTAATAACACGGCTGGAAACAAAAGCAAAATACAAATCA  1600
       T  H  T  L  D  D  P  L  I  L  Q  V  M  G  R  G  K  I  L  L  T  N  N  T  A  G  N  K  S  K  I  Q  I

1601  TCCGATTTACTGTAACGGAAGAAATGGCTCCTGCAGTTCATGTTATAATTTTCTCTCTGACTAACAATGGCTTAGTAATGGCTGATTCAATTCTGTTCGG  1700
      I  R  F  T  V  T  E  E  M  A  P  A  V  H  V  I  I  F  S  L  T  N  N  G  L  V  M  A  D  S  I  L  F  G

1701  AGTTAAAGGATTGTTCAAAACACCAGTTTCTGTAAATGTGGTTCCAAAATCTGCAAAACCTGGAACAACAATGGAAGTGTCAGTGAAAACAAATCCCCAC  1800
        V  K  G  L  F  K  T  P  V  S  V  N  V  V  P  K  S  A  K  P  G  T  T  M  E  V  S  V  K  T  N  P  H

1801  GCTTTTGTAGCTTTGTCTGCTGTTGATCAAAGTGTTCTTTTATTAAACAAAGGAAACGACCTAAAAAGTAGTAGAGTGCTATCAAAATTAAGGAATATTG  1900
       A  F  V  A  L  S  A  V  D  Q  S  V  L  L  L  N  K  G  N  D  L  K  S  S  R  V  L  S  K  L  R  N  I

1901  AATCAAGCAACGCTCTGAAATTTTATGACGACCAGGAAGGTAAATTTTTAGATCCAATAACCACTTTGGAATTATTTAAGGATGCTGGACTACTGTTATT  2000
      E  S  S  N  A  L  K  F  Y  D  D  Q  E  G  K  F  L  D  P  I  T  T  L  E  L  F  K  D  A  G  L  L  L  L

2001  GACTAATGGTTATCAAATGTATTCAACTCCGTTCTTTTGGAACCGTGAGAGGTTTCCAGAGAGAGAGATGCCAGTTATGCCAGGTAGGCCAGGTGTTCCA  2100
        T  N  G  Y  Q  M  Y  S  T  P  F  F  W  N  R  E  R  F  P  E  R  E  M  P  V  M  P  G  R  P  G  V  P

2101  GCAATGGATAGTTTAGAAATTACAAGCCGTTTAAGTCCAGAAAATATGCCTTATGAACCACCATCTGCTCTACCCATGCCAAAAGAGCCACTGCATGTCA  2200
       A  M  D  S  L  E  I  T  S  R  L  S  P  E  N  M  P  Y  E  P  P  S  A  L  P  M  P  K  E  P  L  H  V

2201  GAAGTTATTTTCCCGAAACATTTTTATGGACAAATGAAACCGCCAGTAATAATGGTACAGTTCATATCAAAACGACAGTTCCTGATACAATTACTTCCTA  2300
      R  S  Y  F  P  E  T  F  L  W  T  N  E  T  A  S  N  N  G  T  V  H  I  K  T  T  V  P  D  T  I  T  S  Y

2301  TTATATTAATGCATTTGCGATGGATGATGTAAATGGAATAGGCCTAGCTGAACAACCTGCCAAGCTACAAATATTCCGTCCCTTCTTTGTCACTCTAAAC  2400
        Y  I  N  A  F  A  M  D  D  V  N  G  I  G  L  A  E  Q  P  A  K  L  Q  I  F  R  P  F  F  V  T  L  N

2401  CTTCCATATTCTGTTGTTCGTGGAGAAACTTTGGCTCTACAGGCTCTCGTCTTTAACTACATGACGGAGAATCTAATGGTGTGGTTGACTCTAGAAAACG  2500
       L  P  Y  S  V  V  R  G  E  T  L  A  L  Q  A  L  V  F  N  Y  M  T  E  N  L  M  V  W  L  T  L  E  N

2501  AGAACAATGAATTCGTTCTTATTAATTTAGAAAATGAAATGAATGTTGATGAAAACTCAAAATATATAGAAATCCAGGTGAAATCAGGCGAAGGAGCTTC  2600
      E  N  N  E  F  V  L  I  N  L  E  N  E  M  N  V  D  E  N  S  K  Y  I  E  I  Q  V  K  S  G  E  G  A  S
```

Supplementary Fig. S2 (continued)

```
2601  TGTATTTTTTTACATAGTCCCTAAAAAAATAGGCTACATTGATATTAAAATCGCTGCCAGGTCACATATTGCTGCTGATGCAATTCACAAAAAACTTCTC  2700
        V  F  F  Y  I  V  P  K  K  I  G  Y  I  D  I  K  I  A  A  R  S  H  I  A  A  D  A  I  H  K  K  L  L

2701  GTAAAGGCGGAAGGAATACCAATGTACAATAGCAAAACTGTGTTAGCTGATTTAAGAAAACAGTCTCAGTTTCAAGATAAGATACAAATTCTAATGCCGA  2800
       V  K  A  E  G  I  P  M  Y  N  S  K  T  V  L  A  D  L  R  K  Q  S  Q  F  Q  D  K  I  Q  I  L  M  P

2801  AGAATTTCGTTGATGGTTCTGAAAGAATAGAGATTTCTGCAACAAGGGATATAATGGGAACCGCCATCAATAACATTGACCAACTTCTGAGAATGCCCTC  2900
      K  N  F  V  D  G  S  E  R  I  E  I  S  A  T  R  D  I  M  G  T  A  I  N  N  I  D  Q  L  L  R  M  P  S

2901  TGGATGTGGAGAGCAAAATATGTTAAATTTTGTGCCAAACATTGTCATCATGGACTATTTCACTGCTACAAATCGATTAACTCCTCAGATTGAAGACAAA  3000
        G  C  G  E  Q  N  M  L  N  F  V  P  N  I  V  I  M  D  Y  F  T  A  T  N  R  L  T  P  Q  I  E  D  K
       Thioester site
3001  GCCATTCGATTTATGGAATCGGGTTACCAGAGAGAATTGACTTACAAAAGAAAAGAAGGTTCCTTTAGTGCTTTTGGTGAAAACGATTCAAAAGGCAGTA  3100
       A  I  R  F  M  E  S  G  Y  Q  R  E  L  T  Y  K  R  K  E  G  S  F  S  A  F  G  E  N  D  S  K  G  S

3101  CTTGGTTGACGGCATTTGTTGTGAAATCCTTCATTCAAGCGAAAAAGTACATATCGATAGATGAAAACGTCATTCGAGAATCTTTGATCTGGTTATCGCG  3200
      T  W  L  T  A  F  V  V  K  S  F  I  Q  A  K  K  Y  I  S  I  D  E  N  V  I  R  E  S  L  I  W  L  S  R

3201  ACAAAAAAATGCAAACCATTTGTTTCCAGAAGTAGGAACGGTTTTGCATAAAGAAATGCAGGGTGGCTCTTCACACGGATTGGGTTTGACTGCTTACGTT  3300
        Q  K  N  A  N  H  L  F  P  E  V  G  T  V  L  H  K  E  M  Q  G  G  S  S  H  G  L  G  L  T  A  Y  V

3301  TTATCGGCATTTTTAGAGAGCAAACTCAATGATGTTGATTCGAGTGATAGATTTCATCCACTGATAGATAATTCTATAAACATATTGGAAAACTATTTAG  3400
       L  S  A  F  L  E  S  K  L  N  D  V  D  S  S  D  R  F  H  P  L  I  D  N  S  I  N  I  L  E  N  Y  L

3401  AAGAACTAAAAACAGATTATGATCTTGTTCTTGTGACTTACGTTCTTCATCTAGCCAACAGTAGTGTAAAGGATGTTGCATTCGAGAAAATGAATATGCG  3500
      E  E  L  K  T  D  Y  D  L  V  L  V  T  Y  V  L  H  L  A  N  S  S  V  K  D  V  A  F  E  K  M  N  M  R
                                             Catalytic histidine
3501  TTCGAAGAAAGACGGAGACAAAATGTTCTGGACTATGTCGTCAACAAGTTCATCAGATGATAAATTAGATCCATTTTCGTACAAAAGCAAACCAAAATCA  3600
        S  K  K  D  G  D  K  M  F  W  T  M  S  S  T  S  S  S  D  D  K  L  D  P  F  S  Y  K  S  K  P  K  S

3601  GTTGATATTGAAATGACATCGTACGCACTGATGACGTATTCATTAAGAAACATGATTGCTGAGGGTTTGCCTATCATGAGATGGCTTCTAAGCAAACAAA  3700
       V  D  I  E  M  T  S  Y  A  L  M  T  Y  S  L  R  N  M  I  A  E  G  L  P  I  M  R  W  L  L  S  K  Q

3701  ATGCTAATGGTGGATTTCAATCGACTCAGGACACAGTTGTTGGCATCCAAGCTTTAGCACAAATTGCCAAAGAAATTTCTTTTAGTGATGACGATTTTCA  3800
      N  A  N  G  G  F  Q  S  T  Q  D  T  V  V  G  I  Q  A  L  A  Q  I  A  K  E  I  S  F  S  D  D  D  F  H

3801  TTTAGATGTAAAATTTTTTTATGAAGGAGGAGAAAAAATTATGAGTCTCACAAAAGATAATGATTTAGTTCTGTATATAGAACAAATACCAGGGAATGTT  3900
        L  D  V  K  F  F  Y  E  G  G  E  K  I  M  S  L  T  K  D  N  D  L  V  L  Y  I  E  Q  I  P  G  N  V

3901  AGACAGATTGATATTCAAGCATCGGGTAGTGGATTTGGTATTTTTCAGGTGAGTTGGTCCTACAACGTTTTGACTCTTCAAGAAAATCCTCCCTTTGAAG  4000
       R  Q  I  D  I  Q  A  S  G  S  G  F  G  I  F  Q  V  S  W  S  Y  N  V  L  T  L  Q  E  N  P  P  F  E

4001  TTGGGATTGAAATCAATAACGAAAACAATGAGCTGGCAGTGGAAGCTTGTGTCAATGCTTCTAGATATCTCTACGAATCACATGGGGAAACAAACATGGC  4100
      V  G  I  E  I  N  N  E  N  N  E  L  A  V  E  A  C  V  N  A  S  R  Y  L  Y  E  S  H  G  E  T  N  M  A

4101  GGTCATGGAACTTGCTCTCCCATCTGGTTATGTGGCAGATAAAGAACATTTGCCACATGTTGATGCCAGAAAATTAATTAAACGAGTTGAAACTAAAGAC  4200
        V  M  E  L  A  L  P  S  G  Y  V  A  D  K  E  H  L  P  H  V  D  A  R  K  L  I  K  R  V  E  T  K  D

4201  GGAGACAGTGTTGTTGTGATTTATTTTGATAAGATTGGTGAACAGGTATGTGTAACTGCTATGGCTGAACGCAAAATCATGATTGCTGATGTGAAACCAG  4300
       G  D  S  V  V  V  I  Y  F  D  K  I  G  E  Q  V  C  V  T  A  M  A  E  R  K  I  M  I  A  D  V  K  P

4301  CTTTGGTTCAAGTCTACGATTATTATAAACCAGAAAAACGTGGAGAAGCATTCTACAATCCACCAGCTCTTTCTAAGTGTGAGATTTGTCAAAATGAAGA  4400
      A  L  V  Q  V  Y  D  Y  Y  K  P  E  K  R  G  E  A  F  Y  N  P  P  A  L  S  K  C  E  I  C  Q  N  E  E

4401  ATGCAAACAGACATGTGATAGATAAAGAAATCCAAATAGGCTAAGAACTTCTATTTAGAATAAAAGTTTAATTAATAGAAAATAATTTTCATATCTCATA  4500
        C  K  Q  T  C  D  R  *

4501  CCTAGGACGTTTAGACAAAGAAATGATTATCACACTGAAAATGTGAATTTATTAAATTAAATTTTAACAAACCATTAAATAACAGTACGT  4590
```

ScsuiTEP/CD109-2

```
   1  ATAACCAATGGTTAAATTGTAAAAATTTAAATCGGCAAGATGGATAAAGTGATTTATTTTTTTCTTCTTTTTGCGTTTGCATCATCTGTAGAAAAGAACC  100
                                              M  D  K  V  I  Y  F  F  L  L  F  A  F  A  S  S  V  E  K  N

 101  AACCATTTTACACCGTAACAGCTCCTAATACTATTCGAGCAAATACTCCCTATCAAGTGGTGGTGACTGTGGCCTATGTTAATAAACCTGTCAATGTTGT  200
      Q  P  F  Y  T  V  T  A  P  N  T  I  R  A  N  T  P  Y  Q  V  V  V  T  V  A  Y  V  N  K  P  V  N  V  V

 201  AGTATCCTTAAATCCTGTACATGTTAATGTCACTGAGGATGAGGAACTGAGAGTTGTCGGTACACTTAATAATGGTGAAACGAAAACTTTAGAATTAAAA  300
        V  S  L  N  P  V  H  V  N  V  T  E  D  E  E  L  R  V  V  G  T  L  N  N  G  E  T  K  T  L  E  L  K

 301  ATTGGAAATTTGAACGTTCGTGCATATGAACTAATAGTACAAGGCAGTGGAGGAATCGATTTTAATCAAACAGAAAGTGTAAAAATTAATGAAAATCCAT  400
       I  G  N  L  N  V  R  A  Y  E  L  I  V  Q  G  S  G  G  I  D  F  N  Q  T  E  S  V  K  I  N  E  N  P

 401  TTTCTGTTTTCATCCAAACCGATAAAGGAATTTACCAACCAGGTCAAATAGTACATTTCCGTGTGATTGCAACAAAACCTGTACTTCTTCCTTATAATCC  500
      F  S  V  F  I  Q  T  D  K  G  I  Y  Q  P  G  Q  I  V  H  F  R  V  I  A  T  K  P  V  L  L  P  Y  N  P

 501  CGAAGACATTGAGATTTATGTAAATGATCCCAAATATAACAGAATTAAACAATGGAGAAATGTTACTTTTACTCATGGGGTTTACTCTGGTGAGTTTCAA  600
        E  D  I  E  I  Y  V  N  D  P  K  Y  N  R  I  K  Q  W  R  N  V  T  F  T  H  G  V  Y  S  G  E  F  Q

 601  CTTTCCGATCAAGTCAATTTTGGAGGCTGGATGATTAATGTGAATCATAAGGCGAATATCCTAAGAAGTACGTCTATCACAGTGGAAGAATATGTTTTGC  700
       L  S  D  Q  V  N  F  G  G  W  M  I  N  V  N  H  K  A  N  I  L  R  S  T  S  I  T  V  E  E  Y  V  L

 701  CAAAATTTGAAGTGAATATTCATTTACCAAAGTTTATCATTTCGAATGATACTGATTTAACTGCCTTTGTGGATGCAAGATATACTTATGGGAAACCAGT  800
      P  K  F  E  V  N  I  H  L  P  K  F  I  I  S  N  D  T  D  L  T  A  F  V  D  A  R  Y  T  Y  G  K  P  V

 801  CAGAGGAAAACTAGTTTTAAATATAAGTGATTACTATTGTGATTGGCCGTGCTCTTACTGGAGTGTAAAACCTTTATCTCTCAACACATCGATCGACGGC  900
        R  G  K  L  V  L  N  I  S  D  Y  Y  C  D  W  P  C  S  Y  W  S  V  K  P  L  S  L  N  T  S  I  D  G
```

Supplementary Fig. S2 (continued)

```
 901 AAAGCTGAGTTGAATATAAATCTTAAGGACTTGGAGATACCCGAATGGTTCAGACACGGTAATCGAATAAAATTTTCAGCCACGGTCACAGATCAATTAA  1000
      K  A  E  L  N  I  N  L  K  D  L  E  I  P  E  W  F  R  H  G  N  R  I  K  F  S  A  T  V  T  D  Q  L

1001 CTGGGCGACAAATGATTGGAATGAACGAACTGCGGATATTTTCTTCGAAATATAAGCTGAATTTTGAAACCCCTCCACGATTTAAGCCTGGATTAACCTT  1100
     T  G  R  Q  M  I  G  M  N  E  L  R  I  F  S  S  K  Y  K  L  N  F  E  T  P  P  R  F  K  P  G  L  T  F

1101 TACAACCTACCTTATCGTTCGTTTACAAGATGACACTCCGGTGGTGGATGATGAAGTTAATAAAGTAAAAGTGACGTATTCATTTGCTCACCATGGCAAA  1200
      T  T  Y  L  I  V  R  L  Q  D  D  T  P  V  V  D  D  E  V  N  K  V  K  V  T  Y  S  F  A  H  H  G  K

1201 GTGCACGAAATGGATTTACCCATCTCTAAAAATGGCAAAATAAAAATAGAATTCGTTCCCCCTGAATCAGCTTCAATGATTTCATTGGAAGCGCATTATA  1300
      V  H  E  M  D  L  P  I  S  K  N  G  K  I  K  I  E  F  V  P  P  E  S  A  S  M  I  S  L  E  A  H  Y

1301 AGGATGTTAATTCATACTCAAGAATAGATAAATCCTATTCACTCAGCAATAAATACATACAAATTGTATCAAATGTAGAAAAACCCAAGATAGGTGATGA  1400
     K  D  V  N  S  Y  S  R  I  D  K  S  Y  S  L  S  N  K  Y  I  Q  I  V  S  N  V  E  K  P  K  I  G  D  E

1401 AGTAGAATTTTTCATCTCAAGCACAAGACCCCTTGAAGATTTCTTAACAGTCGAAGTTATCGGCAGAGGCAAAATTCTCTACACTCATAATATTCCAGCA  1500
      V  E  F  F  I  S  S  T  R  P  L  E  D  F  L  T  V  E  V  I  G  R  G  K  I  L  Y  T  H  N  I  P  A

1501 AACGGACAAAAATTAGAAACCTTCAAATTAAAATTAACAAAGGAAATGGCTCCAGAAATTAGAGTAATAGTTCATTATGTTACTTCTTGTGGTGAGGTTA  1600
      N  G  Q  K  L  E  T  F  K  L  K  L  T  K  E  M  A  P  E  I  R  V  I  V  H  Y  V  T  S  C  G  E  V

1601 TAGCAGAAGGTCTTAATTTGGGTGTAGAAGGCGTTTTTAAGACACTGGTGGAATTAAACGTGGATCCAAAATCTGCAAAACCAGGAGCACCGATGGAAGT  1700
     I  A  E  G  L  N  L  G  V  E  G  V  F  K  T  L  V  E  L  N  V  D  P  K  S  A  K  P  G  A  P  M  E  V

1701 ATCAGTCAAGACAAACCCCAATGCTTTTGTTGGTTTATCAGCTGTCGATCAAAGTGTTCTCCTACTAAGGAAAGGAAACGATTTAACAACGGGAGAATTG  1800
      S  V  K  T  N  P  N  A  F  V  G  L  S  A  V  D  Q  S  V  L  L  L  R  K  G  N  D  L  T  T  G  E  L

1801 TTATCTGATCTAAGAAAATATGAAATAGGAAGTCAGTATCAATTTTACTATGGACAACGCACTCCACTGGCTTACTGTCTACCTCCTCGACCTCTATGGG  1900
      L  S  D  L  R  K  Y  E  I  G  S  Q  Y  Q  F  Y  Y  G  Q  R  T  P  L  A  Y  C  L  P  P  R  P  L  W

1901 CTGCCACTACGTCCGAAATGTTTGAGACTTCTGGATTGTTATTATTCACAAATGGATTACTTATTCCACAAAGACAATCGTCTTACGGTTATTCTCCTCA  2000
     A  A  T  T  S  E  M  F  E  T  S  G  L  L  L  F  T  N  G  L  L  I  P  Q  R  Q  S  S  Y  G  Y  S  P  H

2001 TCAATCTTACGGAATGGAATACGATAGTTCCCTAGCTGCGCCAGAACGCGTGGGCGTTGCTCAAACTTCGTATGAACCTGCAATGCGAGTCCGGAGCTAT  2100
      Q  S  Y  G  M  E  Y  D  S  S  L  A  A  P  E  R  V  G  V  A  Q  T  S  Y  E  P  A  M  R  V  R  S  Y

2101 TTTCCAGAAACTTTCCTTTGGACAAACGCTAGTACCAATGATGATGGCTCCTTGATTATCAAGGCTACTGTGCCTGATACAATCACTTCTTATTTCATTA  2200
      F  P  E  T  F  L  W  T  N  A  S  T  N  D  D  G  S  L  I  I  K  A  T  V  P  D  T  I  T  S  Y  F  I

2201 ATGCTTTTGCAGTTGATGATAAAACAGGAATTGGCCTTTCCGATCAGCCTGCTAAGCTACAAATATTCCGCCCTTTTTTTGTCACACTGAATCTTCCGTA  2300
     N  A  F  A  V  D  D  K  T  G  I  G  L  S  D  Q  P  A  K  L  Q  I  F  R  P  F  F  V  T  L  N  L  P  Y

2301 CTCGGTCATCAGAGGAGAAACTTTGGGTCTGCAGGCTCTGGTCTTTAATTATATGACTGAAGATATAGAAGCAGTAGTTACTCTTGATAATGAACAGCAA  2400
      S  V  I  R  G  E  T  L  G  L  Q  A  L  V  F  N  Y  M  T  E  D  I  E  A  V  V  T  L  D  N  E  Q  Q

2401 GAATTCACTATTATCGATAAGGAAAACGAAATAACTGATGGCTCTGCAACAGTGAATGAAATATCAAAAACTGTAAAGGTAGCATCTGGTGATGGTGTTT  2500
      E  F  T  I  I  D  K  E  N  E  I  T  D  G  S  A  T  V  N  E  I  S  K  T  V  K  V  A  S  G  D  G  V

2501 CGGTCTTATTTTATGTTAAACCTAAAAAATTAGGATATTTGAATGTCAAAGTTACGGCCAGATCAAAAACAGCGGGTGATGCAATTCTAAAAAAGCTTCT  2600
     S  V  L  F  Y  V  K  P  K  K  L  G  Y  L  N  V  K  V  T  A  R  S  K  T  A  G  D  A  I  L  K  K  L  L

2601 CGTAAAAGCGGAAGGAAAGACAGTGTATGAAACTCGAGGTGTTGTCGCTGATCTGCGGAAATTCACACAAATTGGAGAGGAAATTCATCTCAATTTCCCT  2700
      V  K  A  E  G  K  T  V  Y  E  T  R  G  V  V  A  D  L  R  K  F  T  Q  I  G  E  E  I  H  L  N  F  P

2701 TCGGATACAGTTGAAGATTCTGAACGAATTGAAGTTTCTGCAATAAGTGACATTATGGGGACAACTATAAACAATCTCCACAATCTTTTGAGAATGCCAG  2800
      S  D  T  V  E  D  S  E  R  I  E  V  S  A  I  S  D  I  M  G  T  T  I  N  N  L  H  N  L  L  R  M  P

2801 GTGGATGTGGAGAACAAAATTTGTTACGGTTCGTTCCAAATATTGTCATCACGGATTATCTCACCAACACCAATCAATTAACTCCTGCCCTAAAGGAGAA  2900
     G  G  C  G  E  Q  N  L  L  R  F  V  P  N  I  V  I  T  D  Y  L  T  N  T  N  Q  L  T  P  A  L  K  E  K
        Thioester site
2901 GGCCATACGTTATATGGAAACCGGTTTTCAAGGACAACTTATATATAGGAGATATGACGATTCATTCAGTGGTTTCGGTAACCGAGATTCCAGTGGCAGT  3000
      A  I  R  Y  M  E  T  G  F  Q  G  Q  L  I  Y  R  R  Y  D  D  S  F  S  G  F  G  N  R  D  S  S  G  S

3001 ACATGGTTGACAGCATTTGTAGTAAAATCTTTCGTTCAGGCGGACAAATATATAACGATCGATCACACAATAATTCGTTCGGCTCTAAATTGGTTAGCCA  3100
      T  W  L  T  A  F  V  V  K  S  F  V  Q  A  D  K  Y  I  T  I  D  H  T  I  I  R  S  A  L  N  W  L  A

3101 AAACTCAGTCACCAAACGGTTCATTTAGTGAAATTGGAAACGTATTTTATAAATCAATGCAGGGAGGATCCTCTCAAGATCTAGGGTTGACAGCCTATGT  3200
     K  T  Q  S  P  N  G  S  F  S  E  I  G  N  V  F  Y  K  S  M  Q  G  G  S  S  Q  D  L  G  L  T  A  Y  V

3201 TTTATCAGCGTTTTTGGAAAGCAAGAGAAATGGTTTTAATGAAGACCAATATCTGATAGAAAACGTAATAAATACAACTGTGAGTGTGTTAGAAAGAGAT  3300
      L  S  A  F  L  E  S  K  R  N  G  F  N  E  D  Q  Y  L  I  E  N  V  I  N  T  T  V  S  V  L  E  R  D

3301 TTGGAGAGCATAGAATCAGACTATGATCTTGTATTCGTAACCTACGTTCTTCATTTGGCAGACAGTCCCAAGAAAGATCAGGCTTTTAACTTAATGAATG  3400
      L  E  S  I  E  S  D  Y  D  L  V  F  V  T  Y  V  L  H  L  A  D  S  P  K  K  D  Q  A  F  N  L  M  N

3401 GACGTTCGAAGACAGTTGGAGATACAAAGTATTGGACTGTACCATTACCAGAAGTTAACGAGACATATTCATATGCTTATTACTACAAACCCAGATCGGT  3500
     G  R  S  K  T  V  G  D  T  K  Y  W  T  V  P  L  P  E  V  N  E  T  Y  S  Y  A  Y  Y  Y  K  P  R  S  V

3501 TGATGTTGAAATAACAGCTTATGCTTTACTGACTTATTCTTTGAGAAATATGGTCGCCGAAGGATTACCAATCATGCGATGGCTTCTAAGTAAGCGAAAC  3600
      D  V  E  I  T  A  Y  A  L  L  T  Y  S  L  R  N  M  V  A  E  G  L  P  I  M  R  W  L  L  S  K  R  N

3601 GCTTATGGTGGATTCGAATCAACCCAGGACACAGTAGTTGGTATCCAAGCCTTAGCTGAATTTACAAAACATCTGTATTATAGTGATTCCAATGTACAAG  3700
      A  Y  G  G  F  E  S  T  Q  D  T  V  V  G  I  Q  A  L  A  E  F  T  K  H  L  Y  Y  S  D  S  N  V  Q

3701 CGGTATTTAGTTACGATGGAGGAGCCAATGCAATGATTCTTACAAATGAAAATGCTTTAGTTCTGCACAAAGAAAAAATTCCAAGTAAAGTTCGAGATAT  3800
     A  V  F  S  Y  D  G  G  A  N  A  M  I  L  T  N  E  N  A  L  V  L  H  K  E  K  I  P  S  K  V  R  D  I

3801 TGAAGTTTCGGCATCAGGAAAAGGCATCGCTGTTCTACAGGTTAGTTGGTCTTATAATGTTCTGCACACTGAGGAGCATCCTGCTTTTGAGATTACATTA  3900
      E  V  S  A  S  G  K  G  I  A  V  L  Q  V  S  W  S  Y  N  V  L  H  T  E  E  H  P  A  F  E  I  T  L
```

Supplementary Fig. S2 (continued)

```
3901  CAATTCGACCCTTTTACGTTTGGAATAGTAGTAGGCGCATGCACCAAATATATATACGAAGATGGAGGTCAAAGTAACATGGCTATCATGGAATTTGGTT  4000
       Q  F  D  P  F  T  F  G  I  V  V  G  A  C  T  K  Y  I  Y  E  D  G  G  Q  S  N  M  A  I  M  E  F  G

4001  TTCCTTCCGGTTACCTCGTAGACAAAGAACGTCTGCCTCAATTGGTCAATTCAATCAAACGTGTCGAGACTAAAAATGCAGACACTGTTCTTGTAATATA  4100
      F  P  S  G  Y  L  V  D  K  E  R  L  P  Q  L  V  N  S  I  K  R  V  E  T  K  N  A  D  T  V  L  V  I  Y

4101  TTTCGATAATATCGGAAATGAACAGGTTTGCGTTAAAATTGAAGGTTATCACAATATTGACGTTAAAGATTTAAAACCTGCTATGGTTCAAGTGTATGAT  4200
        F  D  N  I  G  N  E  Q  V  C  V  K  I  E  G  Y  H  N  I  D  V  K  D  L  K  P  A  M  V  Q  V  Y  D

4201  TATTACGAACCAGAAAAACGAGTAGAAATTTTTTACGATCTTCCGACTAAGGTTATATGTGAAACTTGCGTAACAGAGGAATGCATGCTTGCATGTTCCA  4300
       Y  Y  E  P  E  K  R  V  E  I  F  Y  D  L  P  T  K  V  I  C  E  T  C  V  T  E  E  C  M  L  A  C  S

4301  TGAATTCCTAAAATATAGTTTTCTAGCCTTTTTTAAAAGACAGTTTGGATATTGAAATGGTTTGTTTAGTGAAGAAACGTTAAATCTTAGATGAATAAAA  4400
      M  N  S  *

4401  AATGTTAGTATCATTCGATAAAGAAATAAAAAATATATGTTTACAAAAAAAAAAAAAAAAAAAAAAAAAAAA  4472
```

ScsuiTEP/CD109-3

```
   1  AGGAAGCCAACAGGTGAGTGAGCGCACAAATATGATCCTGGTACCGATGTTGAAGCTCTTTTCTAATTAAACCATATTTATACGGATGTAATTATTTATT  100
 101  TCCATAGTCCAGAAACAATGACTTTTAACTACAGAATATTTATTGGATTAATTTTTTTTGTATTTATTTTTCAAAATGTTCTTGCAATTGGACATTATAT  200
                        M  T  F  N  Y  R  I  F  I  G  L  I  F  F  V  F  I  F  Q  N  V  L  A  I  G  H  Y  M

 201  GATAGTGGCTCCTAAAATTTTACGCCCGGCTTTAAATTATCATGTTAGTGTTTCTGTCCACAATGTAACACAACCTGTACGTATAGACATCACAATAGGA  300
        I  V  A  P  K  I  L  R  P  A  L  N  Y  H  V  S  V  S  V  H  N  V  T  Q  P  V  R  I  D  I  T  I  G

 301  GGTATATCCGATTCTGGAGTTCTTGTATCTTTACCACAACAAGTGTTTCTAGAATCAGAAAAAAGTCAAGTTGTCAATTTTCAGATTGGCATGTGGGGAA  400
       G  I  S  D  S  G  V  L  V  S  L  P  Q  Q  V  F  L  E  S  E  K  S  Q  V  V  N  F  Q  I  G  M  W  G

 401  TGGGAAAATATTCTATGAAAGCTATAGGTTCTGGAGGATTACAGTTTGCCAATGAAACAGATTTGTCATACGAACTTAAAAGTTACTTAGTTTTCATTCA  500
      M  G  K  Y  S  M  K  A  I  G  S  G  G  L  Q  F  A  N  E  T  D  L  S  Y  E  L  K  S  Y  L  V  F  I  Q

 501  AACAGATAAAGCAATTTATAAACCTGGACAACCTGTTCGTATACGAGTTATTGTTGTCAGTCCATCATTAAGGCCTGCAGGTACTGAACCTCTTGACATG  600
        T  D  K  A  I  Y  K  P  G  Q  P  V  R  I  R  V  I  V  V  S  P  S  L  R  P  A  G  T  E  P  L  D  M

 601  TACATTACTGATGGACAAGGAAATAGAATTAAACAGTGGAAACGAGCCTTTACTTCAAGAGGAGTTTTCAATGCAGAAATGCAATTGTCAAATGAACCTG  700
       Y  I  T  D  G  Q  G  N  R  I  K  Q  W  K  R  A  F  T  S  R  G  V  F  N  A  E  M  Q  L  S  N  E  P

 701  TTTTGGGTGACTGGAATATTACAGTATTAATACAGGATCAAGTTTACAAAAAAAGCTTCACTGTTGCTGAATATGTTTTACCCTCTTTTGAAGTTCGAGT  800
      V  L  G  D  W  N  I  T  V  L  I  Q  D  Q  V  Y  K  K  S  F  T  V  A  E  Y  V  L  P  S  F  E  V  R  V

 801  TGATTTACCACCATTTGTAACATTTAACGCTTCTGATATGGTAGCTACTGTTCGTGCCAAATACACTTATGGTAAACCTGTCAAAGGAAAGGTAGTATTA  900
        D  L  P  P  F  V  T  F  N  A  S  D  M  V  A  T  V  R  A  K  Y  T  Y  G  K  P  V  K  G  K  V  V  L

 901  ATGGTTACACCTCTTGTTCGTTCTCCAAAAATCAGAACATATTACACAGATCCACTTAGAAGAACTGCAGAGATAGATGGAAAAGTGGATATTCATTTTA  1000
       M  V  T  P  L  V  R  S  P  K  I  R  T  Y  Y  T  D  P  L  R  R  T  A  E  I  D  G  K  V  D  I  H  F

1001  ATCTTTTTTCTGATCTTAACCTAAAAGACGATTATCATCGCATGATCAGATTTGAAGCTATCGTTACTGAAGCAGTAACTGAACGCCGGGAAAATGCAAC  1100
      N  L  F  S  D  L  N  L  K  D  D  Y  H  R  M  I  R  F  E  A  I  V  T  E  A  V  T  E  R  R  E  N  A  T

1101  AAATACTATGGGAATTTTTAAATATAAACATAAAGTTGAACTCGTTAAATTATCTGAAACTTTTAAACCAGGATTAAAATTTTCAGTGATATTCAAAGTT  1200
        N  T  M  G  I  F  K  Y  K  H  K  V  E  L  V  K  L  S  E  T  F  K  P  G  L  K  F  S  V  I  F  K  V

1201  GCAACCCAAGATGATATTCCTGTAATGGATGAAATTAATCCATTGATCATAAGATATGGTTATACATATGAAGAATCTATATATGATCGTCTTGAATTGA  1300
       A  T  Q  D  D  I  P  V  M  D  E  I  N  P  L  I  I  R  Y  G  Y  T  Y  E  E  S  I  Y  D  R  L  E  L

1301  AAATACCTAAAAATGGAACTGTTCCTTTGGATTTATATCCACCTTTTGCAGATAATGTTAATCAGATTATTATAATTGCAGAATATAAAGATGTAAAACA  1400
      K  I  P  K  N  G  T  V  P  L  D  L  Y  P  P  F  A  D  N  V  N  Q  I  I  I  I  A  E  Y  K  D  V  K  Q

1401  GCAATTTCCTCCAATACGTCGAGCTGAATCTCCAAGCAATACATATATTCAAGCTGTATTAACTACTCTTCAACCAAAAGCTAATGAAGAAGTAATTATT  1500
        Q  F  P  P  I  R  R  A  E  S  P  S  N  T  Y  I  Q  A  V  L  T  T  L  Q  P  K  A  N  E  E  V  I  I

1501  ACAGTAAATTCAACTGTACGTTTAAATACATTCAATGTTGTAGTAATGGGGAGAGGTGATATTGTTTTTGCTGAAACTGTTAATGCAGGAGGTGAGAGGT  1600
       T  V  N  S  T  V  R  L  N  T  F  N  V  V  V  M  G  R  G  D  I  V  F  A  E  T  V  N  A  G  G  E  R

1601  CAATTAAGGTGAAATTTTTTATTACAAGAGCAATGGCTCCCAATATTCGCTTTATTGTTTATTACACGGATTTATCTGGAGAAGTAGTAGCGGACGGTAT  1700
      S  I  K  V  K  F  F  I  T  R  A  M  A  P  N  I  R  F  I  V  Y  Y  T  D  L  S  G  E  V  V  A  D  G  I

1701  AAGTTTTGAAGTTGAAGGAGTCTTCCAGAATTATGTAAATATTGAATCTAATTCAAAAGATGTGAAACCAAAAGACACTGTTAATTTGCAAATAAGCACT  1800
        S  F  E  V  E  G  V  F  Q  N  Y  V  N  I  E  S  N  S  K  D  V  K  P  K  D  T  V  N  L  Q  I  S  T

1801  AATCCAAACTCATTTGTTGGTTTACTTGGAATTGATCAAAGTGTGTTGATTTTAAAAACTGGAAATGATATTTCACAGCAAGAGATTTTATATCAGTTGG  1900
       N  P  N  S  F  V  G  L  L  G  I  D  Q  S  V  L  I  L  K  T  G  N  D  I  S  Q  Q  E  I  L  Y  Q  L

1901  ATGAATTTGATCCTGGAAAACAACCATATAGCAAAGAAGATTTATATTATAACTCTGTTTGGTTTCCAGGTTCTGCTACTGCCAGTGAAGTTTTCAAGGA  2000
      D  E  F  D  P  G  K  Q  P  Y  S  K  E  D  L  Y  Y  N  S  V  W  F  P  G  S  A  T  A  S  E  V  F  K  E

2001  AGCTGGATTTATTACCCTTAGCAATGCCATGATTTATTCAATTTTTCCTTATCTGATGTATCGATCATTTGCTGAAACTGTTGATCAGACTTCTGATATA  2100
        A  G  F  I  T  L  S  N  A  M  I  Y  S  I  F  P  Y  L  M  Y  R  S  F  A  E  T  V  D  Q  T  S  D  I

2101  TCTCTATTAAATGTGGATGAAAGTATTAATGCAAACCTTAGAATGCATTTTCCTGAAACTTGGCTATGGGAAAATTTAACAGCAGGACCTGATGGCAAAG  2200
       S  L  L  N  V  D  E  S  I  N  A  N  L  R  M  H  F  P  E  T  W  L  W  E  N  L  T  A  G  P  D  G  K

2201  CTGTCATCACTAGAGAAGTTCCTGATACAATAACTTCTTGGATGATAACTGCTTTTTCAGTTGATATGATTACAGGCTTGGCTGTGACTGAATCGCCAAC  2300
      A  V  I  T  R  E  V  P  D  T  I  T  S  W  M  I  T  A  F  S  V  D  M  I  T  G  L  A  V  T  E  S  P  T

2301  CAAAATTACTGTTTTTCGTCCCTTTTTTGTGTCTATGAATCTTCCATATTCTGTAATAAGAGGAGAAGCTATTGCTGTTCAAGCCATTGTTTTCAATTAT  2400
        K  I  T  V  F  R  P  F  F  V  S  M  N  L  P  Y  S  V  I  R  G  E  A  I  A  V  Q  A  I  V  F  N  Y
```

Supplementary Fig. S2 (continued)

```
2401  ATGAAACAAACCGTTTCTGTTACTGTTACTTTAGAAAATACCCATCAGTTTGATTTTGTAACAGTTGAAGATGAAATCAATGAAGTTTTCAATGCAAATT  2500
       M  K  Q  T  V  S  V  T  V  T  L  E  N  T  H  Q  F  D  F  V  T  V  E  D  E  I  N  E  V  F  N  A  N

2501  TAAAATCAAAGAGAATTGCTGTAAAATCAGGTGAACCAGAATCGGTTTACTTTATGATCATTCCGAAAGAATTAGGATTTATTGATATCAAAATTACAGC  2600
      L  K  S  K  R  I  A  V  K  S  G  E  P  E  S  V  Y  F  M  I  I  P  K  E  L  G  F  I  D  I  K  I  T  A

2601  TGAAACAGAAAGAGCTGGAGATGCAATGCTTAAAAAGTTGCTCGTAAAGCCTGAAGGCATACAACAAAAATTTAATAAGGCCTTCTTAGTAGATCTGCAA  2700
        E  T  E  R  A  G  D  A  M  L  K  K  L  L  V  K  P  E  G  I  Q  Q  K  F  N  K  A  F  L  V  D  L  Q

2701  ACACCATCAGTGTTTAATGCATATGTCAATGTTGATATACCAAAGCATGTTGTTTCTGGTTCAGAAAAAATTGAAATTTCAGCTATAGCGGATGTAATGG  2800
       T  P  S  V  F  N  A  Y  V  N  V  D  I  P  K  H  V  V  S  G  S  E  K  I  E  I  S  A  I  A  D  V  M

2801  GACCAACCATAAATAATTTTGATGATTTATTACAAATGCCTTTTGGATGTGGCGAACAAAATATGATGCGATTTGTTCCAAATATAGTGGTTTTGGAGTA  2900
      G  P  T  I  N  N  F  D  D  L  L  Q  M  P  F  G  C  G  E  Q  N  M  M  R  F  V  P  N  I  V  V  L  E  Y
                                                   Thioester site
2901  CTTAAGTAATACAGAACAGTTAACAGATTCATTACGATCAAAGGCTATTCTTAATATGGAAACTGGTTATCAGCGAGAACTAGTTTATAAACGTGACGAT  3000
        L  S  N  T  E  Q  L  T  D  S  L  R  S  K  A  I  L  N  M  E  T  G  Y  Q  R  E  L  V  Y  K  R  D  D

3001  GGATCATTTAGTGCATTTGGAAGTCGTGATGAAAATGGAAGCACTTGGTTGACTGCATTTGTTGTTAAATCATTTGGTCAAGCTTTGCCATATATCACTA  3100
       G  S  F  S  A  F  G  S  R  D  E  N  G  S  T  W  L  T  A  F  V  V  K  S  F  G  Q  A  L  P  Y  I  T

3101  TTGACGAAAAAGTAATTGAAAATTCACTTAAATGGCTAGCATTACAACAAGCAGAAAATGGGTCATTCCCAGAAGTTGGCAGCATTAATAATAAAGAAAT  3200
      I  D  E  K  V  I  E  N  S  L  K  W  L  A  L  Q  Q  A  E  N  G  S  F  P  E  V  G  S  I  N  N  K  E  I

3201  ACAGGGTGGATCAAGTGATGGTATTGCTTTAACTGCTTATGTATTACTAGCTTTTTTCCAAAACAAGGGTGAAACTTCTCATGGACCTGTTGTGATGAAA  3300
        Q  G  G  S  S  D  G  I  A  L  T  A  Y  V  L  L  A  F  F  Q  N  K  G  E  T  S  H  G  P  V  V  M  K

3301  GCATTACAGCTACTAGAAAAGGAACTTGAAGCACTAACTGACTTATATTCTATTGTTATTACTACTTATGCATTACATGTAGCAAGTAGCAATTTTAAAG  3400
       A  L  Q  L  L  E  K  E  L  E  A  L  T  D  L  Y  S  I  V  I  T  T  Y  A  L  H  V  A  S  S  N  F  K

3401  AAGTAGCATATCAAAAATTGCAAGCAGCTGCTACTGTTAAAGGTGATTTAAGATACTGGCAGAAACCAGAACCAACTGCAGCTGGTGAAATTGAAGTGAA  3500
      E  V  A  Y  Q  K  L  Q  A  A  A  T  V  K  G  D  L  R  Y  W  Q  K  P  E  P  T  A  A  G  E  I  E  V  K

3501  ATCTGTGGATGTGGAAATGACTTCTTACGCTCTCATGACATACATTCTTCAAAATGATTTATCAGAAGCAATGCAAATTTTGAAGTGGTTGATAAGTGAA  3600
        S  V  D  V  E  M  T  S  Y  A  L  M  T  Y  I  L  Q  N  D  L  S  E  A  M  Q  I  L  K  W  L  I  S  E

3601  CGTAATTCAAATGGTGGCTTTAAATCAACACAGGACACAATGATTGCAATTCAGGCATTGGCAAAACTAGCTCAGCGTATAAGTGATCCTCAAGTGAAAA  3700
       R  N  S  N  G  G  F  K  S  T  Q  D  T  M  I  A  I  Q  A  L  A  K  L  A  Q  R  I  S  D  P  Q  V  K

3701  TCACAGTAACTTTTACTTATTCAGGACAACAGAAAACATTTTCTCTTAATAGGGAAAATGCCATGATTTTACAGACAGATGAGATTCCAGCTGTAGAAAA  3800
      I  T  V  T  F  T  Y  S  G  Q  Q  K  T  F  S  L  N  R  E  N  A  M  I  L  Q  T  D  E  I  P  A  V  E  K

3801  GAATGTTAATATTTCAGCCACAGGGTATGGATTTGGAATTGTACAGGTTTCTTATCAGTATAACATCATGGTTTCAAAAGAATTCCCAGCTTTTCAAGTA  3900
        N  V  N  I  S  A  T  G  Y  G  F  G  I  V  Q  V  S  Y  Q  Y  N  I  M  V  S  K  E  F  P  A  F  Q  V

3901  AACCCTTTGGTTGACCGCTCTTCAACAAAAAATCGTTTGCAACTGAATGTTTGTGCAGCCTATGGAGAAAAGAATGGTGTAAGTAACATGGCTGTTATGG  4000
       N  P  L  V  D  R  S  S  T  K  N  R  L  Q  L  N  V  C  A  A  Y  G  E  K  N  G  V  S  N  M  A  V  M

4001  AAGTTACACTGCCATCAGGTTACGTTATTGACAGAGATTCTCTCCCAGCTTTACATCGTGTAGATGAAGTAAAAAGAGTAGATGTCAAAGATATGGACAC  4100
      E  V  T  L  P  S  G  Y  V  I  D  R  D  S  L  P  A  L  H  R  V  D  E  V  K  R  V  D  V  K  D  M  D  T

4101  TGGCATTATTATTTACTTTGATAAGTTGGATAATAAACTAGTGTGCCCAACAATAAAAGCATATCGGACATTCAGAGTTGCAAAGCAACGCCAAACAGCT  4200
        G  I  I  I  Y  F  D  K  L  D  N  K  L  V  C  P  T  I  K  A  Y  R  T  F  R  V  A  K  Q  R  Q  T  A

4201  GTTTATTTATATGACTATTATGATCAAGGCAAAGCAGCTCGATATTTTTACCAAACACCACAAGCTAGTTTATGTGACATTTGTGAAGGTGAAGAATGCA  4300
       V  Y  L  Y  D  Y  Y  D  Q  G  K  A  A  R  Y  F  Y  Q  T  P  Q  A  S  L  C  D  I  C  E  G  E  E  C

4301  ACAGTTCAAAATGCAGGAAAGAAATAGAAATATCAAAAGAAAGTAACAGTAACAATGGTTGTTCAAAAGCAAGCATGCAAATCTGGACTTTATTAACAAA  4400
      N  S  S  K  C  R  K  E  I  E  I  S  K  E  S  N  S  N  N  G  C  S  K  A  S  M  Q  I  W  T  L  L  T  K

4401  AATGATATTAATTTATTATTATTGCCATTAAATTGTCTTCAAGTTTTATAACTGCACAATATTACTTAATCATGAAAAAAAAATTCAGTCGTGTATATGG  4500
        M  I  L  I  Y  Y  Y  C  H  *

4501  ATTTATTATTTAAACTCAGGTTTATTAAGGAAAACTCTGTGCTGCTCTTTCCATTTCTTTGAAACAATTTATTTAATATGCATCAGTAAGAAAATGGCTA  4600
4601  TTAAATACTTGATTATTTTTTATTTACCATTTTATTTATTTAATAATAAATTTACCATTTTGTTATTTAACAACTAAAGATCAAAATATTAGGAACTAAT  4700
4701  T  4701
```

ScsuiTEP/CD109-4

```
   1  GTTTTTTTTTTTTGATTAGAATTGACCCGCATTAACTGGCCATTCTTAACCGAGTGGTTACCTAGGCAACAGGATGTAGATGGTGTACAAATCAATTTCC  100
 101  CCCCTCGGCAAATTTTGTCGTCATTAGTTCTTGGATTCTGGTTACAAAAGTGACCTCAGAAAACCGAGTTGCAGTTGAAATTCATTGACGTCACGGCTAT  200
 201  TTAAATAGTCTTTTTCTGCCCTGTCGGTGCATTCGGGCTTGAGAAGTGACCCTCACTCGCTCATCTAACAACCGGCAGATGTGGCTAAGGATTCTCTCCA  300
 301  TCAACAACGACCGAAGGAAAACATAAACATATCATCGCTCACATTTCATTTATTGATTCTTTTCAAAATATACTTTTAAGAATGGATAAAATCATAATCT  400
                                                                                        M  D  K  I  I  I

 401  TTTCATTTATTTTGGGGAGCTTATATGCTGTGCAAGGTTTAGATCCATTATATACTGTTACTGCTCCGAAAACACTTCGTGCAAATTCACCTTACCATGT  500
      F  S  F  I  L  G  S  L  Y  A  V  Q  G  L  D  P  L  Y  T  V  T  A  P  K  T  L  R  A  N  S  P  Y  H  V

 501  GGTGATCACAGTTCATCAAGCTCCAGAACCAGTCAATGTAGAAATTTCTTTAAAACCAACAGAAAATGTGACAGAAGAATTAAAAGAGAAAGATGTTATT  600
        V  I  T  V  H  Q  A  P  E  P  V  N  V  E  I  S  L  K  P  T  E  N  V  T  E  E  L  K  E  K  D  V  I

 601  AGTAATGGAGAAACAAAAAAGTTACAACTAAATGTTGGTGATTGGAATTATCCATCGTACCGTTTGAATGTACAAGGAAGTGGAGGAATTGTATTTAATA  700
       S  N  G  E  T  K  K  L  Q  L  N  V  G  D  W  N  Y  P  S  Y  R  L  N  V  Q  G  S  G  G  I  V  F  N

 701  AAACGGAAACAGTGAATTTCAATTCAAAGTCTTATTCAGTTTTTATTCAAACTGATAAAGCAATATATCAACCAGGGCAAATAGTACATTTTCGTGTGAT  800
      K  T  E  T  V  N  F  N  S  K  S  Y  S  V  F  I  Q  T  D  K  A  I  Y  Q  P  G  Q  I  V  H  F  R  V  I

 801  AGTGACCAATTCAACATTGCAACCAATAAGTCCTGAAACTGTTATTTACGTATCAGATTCTCAAGGCAATAGAATTAAGCAATGGATGAATGTTACTTTT  900
        V  T  N  S  T  L  Q  P  I  S  P  E  T  V  I  Y  V  S  D  S  Q  G  N  R  I  K  Q  W  M  N  V  T  F
```

Supplementary Fig. S2 (continued)

```
  901 GATAGAGGAGTATATTCTGGAGATCTTCAACTTTCTGATCAAGTTATACTCGGTGATTGGAGGATTTCAGTAGATACTCAAAAAGTACTTCAGACTAAAT 1000
       D  R  G  V  Y  S  G  D  L  Q  L  S  D  Q  V  I  L  G  D  W  R  I  S  V  D  T  Q  K  V  L  Q  T  K

 1001 ATTTTACAATCGCTGAATATGTTCTGCCGACATTTGAAGTGATCATTCATTTACCTACATTTGTCACATTTAATGAATCAGATGTAGTTGCTACTGTAGA 1100
      Y  F  T  I  A  E  Y  V  L  P  T  F  E  V  I  I  H  L  P  T  F  V  T  F  N  E  S  D  V  V  A  T  V  E

 1101 AGCCAAATATACATATGGTAAACCTGTTAAAGGAGAAGTAACACTTAATGTTACAGATTATTATTGCCAGTGGCCATGTATCAGCAGCGCAGTTAAACCA 1200
        A  K  Y  T  Y  G  K  P  V  K  G  E  V  T  L  N  V  T  D  Y  Y  C  Q  W  P  C  I  S  S  A  V  K  P

 1201 TTTGCTTTAAAATCATCTATTGATGGAAAGGCTGATATGAAACTTAATTTAGTCAAAGAGTTAAATCTTCCTGATTGGTATCGCTACGGATCTAAGCGTT 1300
       F  A  L  K  S  S  I  D  G  K  A  D  M  K  L  N  L  V  K  E  L  N  L  P  D  W  Y  R  Y  G  S  K  R

 1301 TTACATTCATTGCAACAGTAACAGAAGCGCTGACAAAACGTCAACAAAATGGCACTAATGATTTAAATCTTTATTCTGATAAATACAAATTAAATTTTGA 1400
      F  T  F  I  A  T  V  T  E  A  L  T  K  R  Q  Q  N  G  T  N  D  L  N  L  Y  S  D  K  Y  K  L  N  F  D

 1401 TACCCCAGATTCTTTTAAACCTGGTCTTCTGTTCACAACTTATCTGAATGTTTTGTTACAAGATGGCACACCTATTATTGATGATGTGAATAATGTTACT 1500
        T  P  D  S  F  K  P  G  L  L  F  T  T  Y  L  N  V  L  L  Q  D  G  T  P  I  I  D  D  V  N  N  V  T

 1501 ATAAATTATTTTTATTCTTGGAATGAGGCTTCTCAAACTTTGAAATTTCCTGTTCCAAAAGATGGGAAAATTAAGTTAGAATTAGTTCCTCCTGAATCTG 1600
       I  N  Y  F  Y  S  W  N  E  A  S  Q  T  L  K  F  P  V  P  K  D  G  K  I  K  L  E  L  V  P  P  E  S

 1601 CAGAAATTATTCGGTTATCGGCATCTTTTATTGAAGCTTCTTCATATTCCACTGTTAATAGAGCTCAGTCTCTTAGTGAAAGATTTCTCCAGTTAAGCTT 1700
      A  E  I  I  R  L  S  A  S  F  I  E  A  S  S  Y  S  T  V  N  R  A  Q  S  L  S  E  R  F  L  Q  L  S  L

 1701 AATAACAGAGAATCCAAAAATTGGAGATGAAGTCGAATTACTGGTGAATGCCACAAAAAATCTGGAAGATCCATTGATCTTGGAGATTATTGGTAGAGGA 1800
        I  T  E  N  P  K  I  G  D  E  V  E  L  L  V  N  A  T  K  N  L  E  D  P  L  I  L  E  I  I  G  R  G

 1801 AAAATTCTTCATACGGAAAATATTCCTGGCAGTAAAACAAATCATCAGAAGATCAGTTTTAAATTATTGCCTGAAATGGCTCCAAAAATAAGAGTAATTG 1900
       K  I  L  H  T  E  N  I  P  G  S  K  T  N  H  Q  K  I  S  F  K  L  L  P  E  M  A  P  K  I  R  V  I

 1901 TTTATTATACAACTCCGTGCGGTGAAGTTGTTGCTGATGCTATTGATTTTGGAGTAGAAGGAATTTTTAAAACTCCAGTTAAAGTGAACGTGAATCCTAA 2000
      V  Y  Y  T  T  P  C  G  E  V  V  A  D  A  I  D  F  G  V  E  G  I  F  K  T  P  V  K  V  N  V  N  P  N

 2001 TTCTACAAAGCCAGGTTCAGAAATAGATGTTTCAGTACAAACTAATCCCAATGCATTTATTGGCTTATCTGCAATTGACCAAAGTGTGCTTTTGCTGAAA 2100
        S  T  K  P  G  S  E  I  D  V  S  V  Q  T  N  P  N  A  F  I  G  L  S  A  I  D  Q  S  V  L  L  L  K

 2101 AAAGGAAATGATATAACTACTAAAGAGGTCTTAACAGATTTGCAGAACTATGAAATTGGTGACAGATCTCCTTTTCAATTTGATGATTATGTTAGTCCTA 2200
       K  G  N  D  I  T  T  K  E  V  L  T  D  L  Q  N  Y  E  I  G  D  R  S  P  F  Q  F  D  D  Y  V  S  P

 2201 GACATTATTCATTGCGCCCTTGGAGACCTCGTTCTTCTTCTACTTTGGAACTCTTCACGAATGTTGGATTGATATTCTTAACTAATGGACTTCTTGCAAG 2300
      R  H  Y  S  L  R  P  W  R  P  R  S  S  S  T  L  E  L  F  T  N  V  G  L  I  F  L  T  N  G  L  L  A  R

 2301 ATATCCTTATTCTGGTTACGGAGGCTTTGCTGGAGGAGCAGGTGGAATTTCATTACGACCTGCACAAGCCTTTGCCACTACAGCATTAGGATCACCACCG 2400
        Y  P  Y  S  G  Y  G  G  F  A  G  G  A  G  G  I  S  L  R  P  A  Q  A  F  A  T  T  A  L  G  S  P  P

 2401 CCACCAGCAAGGCGTCCTCCCAACTCAGGATTAGTAGAACCTTCTCGAGTCAGAACCTATTTCCCTGAAACATTCCTTTGGATAAACACCACATCCAATG 2500
       P  P  A  R  R  P  P  N  S  G  L  V  E  P  S  R  V  R  T  Y  F  P  E  T  F  L  W  I  N  T  T  S  N

 2501 AGGAAGGTATATTCAATATTAAAACTGCAGCACCTGATACTATTACCTCTTATTTCATTAATGCATTTGCAATTGATAATGAAAATGGTTTAGGCTTATC 2600
      E  E  G  I  F  N  I  K  T  A  A  P  D  T  I  T  S  Y  F  I  N  A  F  A  I  D  N  E  N  G  L  G  L  S

 2601 TGATCAACCAGCTAAGTTGCAAATATTTCGTCCATTTTTTGTAACCATGAATTTACCATATTCTGTTGTTCGAGGAGAAACATTATCTCTCCAAGCATTA 2700
        D  Q  P  A  K  L  Q  I  F  R  P  F  F  V  T  M  N  L  P  Y  S  V  V  R  G  E  T  L  S  L  Q  A  L

 2701 GTATTCAATTACATGAAAGAAGACTTAGAAGCAGAAGTAACTCTTTCTATTGAAGATGGAAGTTTTACTTTTGTTGATTTGGAGAATGAAATAAATGATG 2800
       V  F  N  Y  M  K  E  D  L  E  A  E  V  T  L  S  I  E  D  G  S  F  T  F  V  D  L  E  N  E  I  N  D

 2801 ATGCTAATGCTAAACAGAACTTTATAACGAAAACTGTTAAGGTAAAATCAGGAGATGGTACATCAGTGTATTTTTATATTGTGCCTAAAAAGTTGGGTCA 2900
      D  A  N  A  K  Q  N  F  I  T  K  T  V  K  V  K  S  G  D  G  T  S  V  Y  F  Y  I  V  P  K  K  L  G  H

 2901 CTTAGATCTCAAAATGACAGCCAAAACAAATGTTGCTGCTGATGCACTTATCCGAAAACTGCTTGTTAAGGCCGAAGGCATGCCAGTATATGTAAATAAA 3000
        L  D  L  K  M  T  A  K  T  N  V  A  A  D  A  L  I  R  K  L  L  V  K  A  E  G  M  P  V  Y  V  N  K

 3001 GCATCGGTAGCTGATTTGCGTGAACAATCACAGTTTAATGAGAAGGTGAAGATTGAATTTCCTGAAGATTATGTCAAGGACTCTGAAAATATTGAAATTT 3100
       A  S  V  A  D  L  R  E  Q  S  Q  F  N  E  K  V  K  I  E  F  P  E  D  Y  V  K  D  S  E  N  I  E  I

 3101 CTGCAATAAGTGACATAATGGGAACCACTGTCAGCAATATTGATAAGCTTTTAAAAATGCCGTATGGATGTGGAGAACAAAACATGTTAAATTTTGTACC 3200
      S  A  I  S  D  I  M  G  T  T  V  S  N  I  D  K  L  L  K  M  P  Y  G  C  G  E  Q  N  M  L  N  F  V  P
                                                                        Thioester site
 3201 AAATATTGCCATAACAGATTATCTCAATGAAACAAATAAACTCACCCCAGAAATTAAAGAAAAAACTATACGATTTATGGAATCTGGATATCAACGACAA 3300
        N  I  A  I  T  D  Y  L  N  E  T  N  K  L  T  P  E  I  K  E  K  T  I  R  F  M  E  S  G  Y  Q  R  Q

 3301 CTCACATACAAAAGAACTAATAACTCTTTTAGTGCGTTTGGAAATTCTGATAAAGTTGGAAGTACATGGTTGACATCATTTGTAGTTAAATCATTTATCC 3400
       L  T  Y  K  R  T  N  N  S  F  S  A  F  G  N  S  D  K  V  G  S  T  W  L  T  S  F  V  V  K  S  F  I

 3401 AAGCAAAAAAATATATTACAATTGATGAAAATGTGATCCACTCATCTCTAATGTGGTTATCTCATCAACAATTTCCAAATGGTTCATTTCCTGAAGTAGG 3500
      Q  A  K  K  Y  I  T  I  D  E  N  V  I  H  S  S  L  M  W  L  S  H  Q  Q  F  P  N  G  S  F  P  E  V  G

 3501 AACTGTCTTTCACAAAGCAATGCAGGGAGGATCTTCTCAAGGATTAGGTTTAACTGCTTATGTTTTATCAGCATTTCTGGAAAGCAAAATGAGTGGTCTT 3600
        T  V  F  H  K  A  M  Q  G  G  S  S  Q  G  L  G  L  T  A  Y  V  L  S  A  F  L  E  S  K  M  S  G  L
       Catalytic histidine
 3601 GATTCAAGTAAAGAACTGGTTGATTCTGTAATTAGTAGTAGCATAGAAATACTAGAGAAAGATTTGGATTCTATTGAGTCAGACTACGACCTTGTATTTG 3700
       D  S  S  K  E  L  V  D  S  V  I  S  S  S  I  E  I  L  E  K  D  L  D  S  I  E  S  D  Y  D  L  V  F

 3701 TGACATATGTTCTTCATTTAGCAAATAGCTCTAGCAAAGATGTTGCTTTTCAGAAAATGAATGAACGTTCAAAAACTGTTGGAGATACTAAATTTTGGAC 3800
      V  T  Y  V  L  H  L  A  N  S  S  S  K  D  V  A  F  Q  K  M  N  E  R  S  K  T  V  G  D  T  K  F  W  T

 3801 AATGCCATTGCCAGAAATCAATCAATCAGATCCATATGCTTATTACAACAGACCAAGATCAGTTGATGTTGAAATGACATCATATGCATTATTGACTTAC 3900
        M  P  L  P  E  I  N  Q  S  D  P  Y  A  Y  Y  N  R  P  R  S  V  D  V  E  M  T  S  Y  A  L  L  T  Y
```

Supplementary Fig. S2 (continued)

```
3901  TCATTAAGAAACATGATAGCTGAAGGGTTACCAATTATGAGATGGCTTTTAACCAAAAGAAATTCTAATGGAGGATTTGAATCAACCCAAGATACAGTAG  4000
       S  L  R  N  M  I  A  E  G  L  P  I  M  R  W  L  L  T  K  R  N  S  N  G  G  F  E  S  T  Q  D  T  V

4001  TTGGTATTCAAGCTTTAGCACATTATGCAAAAAAAATCTCTGCAGGTGACGGGTCAAATATGAAGGTTAAGTTTTCTTACAAAGATGGTGAAAAAGAATT  4100
      V  G  I  Q  A  L  A  H  Y  A  K  K  I  S  A  G  D  G  S  N  M  K  V  K  F  S  Y  K  D  G  E  K  E  L

4101  GGAACTTACTAAGGAAAATGCTTTAATTTTGCACAGAGAACAAATTCCTGGGAGCACTCGAGAAATTGATATCTCTGCAACAGGAAAAGGCTTAGGCATA  4200
        E  L  T  K  E  N  A  L  I  L  H  R  E  Q  I  P  G  S  T  R  E  I  D  I  S  A  T  G  K  G  L  G  I

4201  CTTCAGGTCAGTTGGTCATACAATATTTTAACTGTTCAAGAACGTCCAGCTTTTGAAATTTTAACAGATGTTACTAATGAAAATAATGAGATGACAGTGA  4300
       L  Q  V  S  W  S  Y  N  I  L  T  V  Q  E  R  P  A  F  E  I  L  T  D  V  T  N  E  N  N  E  M  T  V

4301  AAGCTTGTACAAAATATGTTTATGAAGATGAAAATGAAAGTAACATGGCAGTCATGGAAATAGGTTTACCTTCTGGCTATGTAGCGGACAAAGAACATTT  4400
      K  A  C  T  K  Y  V  Y  E  D  E  N  E  S  N  M  A  V  M  E  I  G  L  P  S  G  Y  V  A  D  K  E  H  L

4401  GCCATCAATTGACGAATCAAAATCAATTAAACGAGTTGAGACAAAGGATGGTGATAGTGTCATTGTAATATACTTTGATAAGATTGGAGAAAAAGTTTGT  4500
        P  S  I  D  E  S  K  S  I  K  R  V  E  T  K  D  G  D  S  V  I  V  I  Y  F  D  K  I  G  E  K  V  C

4501  GCAGATGCCAAGGCATATCGTAACAACAAAGTAGCTGATTTAAAACCAGCTTTAATTGAAGTTTATGATTATTATGACCTAAAAAAACGTGGAGAAAAAT  4600
       A  D  A  K  A  Y  R  N  N  K  V  A  D  L  K  P  A  L  I  E  V  Y  D  Y  Y  D  L  K  K  R  G  E  K

4601  TCTACACTCCTCCATTGGTTACAGTTTGTGATCTTTGTGAGACAGATGAATGCAAACAGAAATGCAAAAAATAATTGAATAATAGTTTTTATAAAAAATT  4700
      F  Y  T  P  P  L  V  T  V  C  D  L  C  E  T  D  E  C  K  Q  K  C  K  K  *

4701  TTAGACTTATTTTAAATGTTTTGTGAATATTATTATTGTAAAAATGAATTTATTGTATAA  4760
```

Supplementary Fig. S2 (continued)

PsfuC3

```
   1  GGCAGGATGAGGAGCTCTCCGCTGTGGCTGCTGGCCTCTCTGGCCTTCACTTCTCTCTCTTCTCCAGCCGATGGAGCTGCAATGAAAGTCATGATGGCCC  100
             M  R  S  S  P  L  W  L  L  A  S  L  A  F  T  S  L  S  S  P  A  D  G  A  A  M  K  V  M  M  A

 101  CCAACTTGTTGCGGGTCGGAACCCCGGAGAAAATCTTTGTGGAATGCCAAGACTGCACAGGCGCGGAAATCCCCGTCGAAATCAAAGTCATGAACCATCC  200
      P  N  L  L  R  V  G  T  P  E  K  I  F  V  E  C  Q  D  C  T  G  A  E  I  P  V  E  I  K  V  M  N  H  P

 201  GACCAAAACTGAAACGCTGAAGAGCACACGTGTGAGGCTGAACAGTGGGAATCATTTCCAGGCGCTGGGGGAGCTCACGGTTCCCACCGCCAGCTTCAGC  300
        T  K  T  E  T  L  K  S  T  R  V  R  L  N  S  G  N  H  F  Q  A  L  G  E  L  T  V  P  T  A  S  F  S

 301  AGGGACCCGAGTGCCAAGGAGTACGTTTACCTGGAGGCTCACTTTCCGGACCGTGTGCTGGAGAAGGTGGTGATGGTGTCCTTCCAGGCTGGCTACATCT  400
       R  D  P  S  A  K  E  Y  V  Y  L  E  A  H  F  P  D  R  V  L  E  K  V  V  M  V  S  F  Q  A  G  Y  I

 401  TTATTCAAACTGACAAGCCGCTATACACACCGGACAGCAAAGTTCAATACAGGCTTTTTGCAGTGACACCTGGCATGGAGCCCGTCGAGCGAGACAACTC  500
      F  I  Q  T  D  K  P  L  Y  T  P  D  S  K  V  Q  Y  R  L  F  A  V  T  P  G  M  E  P  V  E  R  D  N  S

 501  ATCCAAAGCCGGCGCCTCTGTTGCCATCGAGATCGTGACTCCTGAGAACATCGTTTTAACACAAGATCCAGTCTCGATGATCGGCATTCACTCTGGAGAC  600
        S  K  A  G  A  S  V  A  I  E  I  V  T  P  E  N  I  V  L  T  Q  D  P  V  S  M  I  G  I  H  S  G  D

 601  TACACACTCGGTGACGTTGTCAGTTTTGGCGTGTGGAAGATTGTGGCAAAGTTCCAGAGCAACCCACAGCTGAGCTTCTCAGCAGAGTTTGAGGTCAAAG  700
       Y  T  L  G  D  V  V  S  F  G  V  W  K  I  V  A  K  F  Q  S  N  P  Q  L  S  F  S  A  E  F  E  V  K

 701  AGTATGTGTTGCCCAGTTTTGAGGTGAAACTGACGCCTTCGAACCCCTTCTTCTACGTGGACAGTCCATCGCTCGCCGTTGACATCAGAGCAACGTACCT  800
      E  Y  V  L  P  S  F  E  V  K  L  T  P  S  N  P  F  F  Y  V  D  S  P  S  L  A  V  D  I  R  A  T  Y  L

 801  GTTTGGTGAAGACGTGAGCGGGGTGGCCTACGTTGTGTTTGGGGTCTCGAAGGAAGACCGGAAGTATGGCCTTCCTGCCTCCCTTCAGAGAGTATCCATT  900
        F  G  E  D  V  S  G  V  A  Y  V  V  F  G  V  S  K  E  D  R  K  Y  G  L  P  A  S  L  Q  R  V  S  I

 901  GGTGGAGGCACTGGACAAGCCACACTGACGAGAGAGCAGATTACACAGACCTTCAGAAATGTGGAGGAACTGGTGGGGGGTTCTATTTTTGTGACCGTCA  1000
       G  G  G  T  G  Q  A  T  L  T  R  E  Q  I  T  Q  T  F  R  N  V  E  E  L  V  G  G  S  I  F  V  T  V

1001  GCGTGCTTACTGACAACGGTGGAGAAATGGTCGAGGCCGAGTTCAGAAACATCCAGATAGTCAAGTCACCGTACGCAATCAACTTCAAGAAGACCCCCAA  1100
      S  V  L  T  D  N  G  G  E  M  V  E  A  E  F  R  N  I  Q  I  V  K  S  P  Y  A  I  N  F  K  K  T  P  K

1101  ATTTTTCAAACCTGGAATGTCTTTCGATATTGCGGTTGAGGTGGAGAATCCAGATAAATCTCCCGCACGCGGCGTCAAGGTGGTGGTCAATCCTGGCAAT  1200
        F  F  K  P  G  M  S  F  D  I  A  V  E  V  E  N  P  D  K  S  P  A  R  G  V  K  V  V  V  N  P  G  N

1201  GTGGAGGGTTACACCGCCGCCAATGGCCTGGCAATGCTCACCATCAACTCTGATGCCGGGTCCAGAGAACTCGCCATCAATGCAAGGACCGACGACTCTC  1300
       V  E  G  Y  T  A  A  N  G  L  A  M  L  T  I  N  S  D  A  G  S  R  E  L  A  I  N  A  R  T  D  D  S

1301  TTCTTACGCGAGAGAGGCAAGCTACAGAAAGAATGGTAGCTTATCCATATAAAACCACAGGCAACAATTACATCCATATAAGTGCGGATACAACAGAGCT  1400
      L  L  T  R  E  R  Q  A  T  E  R  M  V  A  Y  P  Y  K  T  T  G  N  N  Y  I  H  I  S  A  D  T  T  E  L

1401  GGAGAAAGGGGAAAACTTCAAAATCAACCTGAACATCCAGCGAGCGATGACAGGAAAGATGGACATCACCTACCTAATCATGAGCAGAGGTCAGCTGGTG  1500
        E  K  G  E  N  F  K  I  N  L  N  I  Q  R  A  M  T  G  K  M  D  I  T  Y  L  I  M  S  R  G  Q  L  V

1501  AAAAGTGGCCGTCACACTTTAACAAACCAAGTGCTGGTTTCCCTTATCATCCCCATTACCAAAGAGATGCTACCGTCATTCCGAGTTGTGGCCTTTTACC  1600
       K  S  G  R  H  T  L  T  N  Q  V  L  V  S  L  I  I  P  I  T  K  E  M  L  P  S  F  R  V  V  A  F  Y

1601  ACACAAACGACAACAAAGTGGTGTCAGACTCGATTTGGGTGGATGTCAAGGATTCCTGCATGGGCACGCTGAGGCTGGAACCAGTGAGACCCGCTCCGTC  1700
      H  T  N  D  N  K  V  V  S  D  S  I  W  V  D  V  K  D  S  C  M  G  T  L  R  L  E  P  V  R  P  A  P  S

1701  ATTTGAGCCCCGCAGGATGTTCGGTCTGAGGGTCACCGGAGACCCAGGAGCCACAGTAGGACTGGCAGCTGTTGACAAAGGCATCTTCGTCCTAAACAAC  1800
        F  E  P  R  R  M  F  G  L  R  V  T  G  D  P  G  A  T  V  G  L  A  A  V  D  K  G  I  F  V  L  N  N

1801  AAGCACCGCCTCACTCAGAAAAAGATTTGGGACATTGTTGAGAAGTACGACACGGGCTGCACACCAGGTGGAGGGAAAGACAGTATGAGTGTCTTCTTTG  1900
       K  H  R  L  T  Q  K  K  I  W  D  I  V  E  K  Y  D  T  G  C  T  P  G  G  G  K  D  S  M  S  V  F  F

1901  ACGCAGGGCTGCTGTTTGAATCCGACACAGCCTCTGGGACTGCCTACCGACAAGAATCCAAATGTCCTACTGTGGAAAGGAGGAAACGAGCCACAAGTAG  2000
      D  A  G  L  L  F  E  S  D  T  A  S  G  T  A  Y  R  Q  E  S  K  C  P  T  V  E  R  R  K  R  A  T  S  R
                                                                                 β-α cleabage site
2001  AATGGATGTCATCACCAGCTTAGCGAGCGAATACGATGGTACACCTGTGAAAGAGTGCTGTGTGGAAGGCATGAAGTTACTCCCTGTCTCTTACTCCTGT  2100
        M  D  V  I  T  S  L  A  S  E  Y  D  G  T  P  V  K  E  C  C  V  E  G  M  K  L  L  P  V  S  Y  S  C
                                                                                                ANA domain
2101  GAGGTCCGCAGCGAGTACATTGTGGACGACGCCCCTTGTGCCGCCGCCTTCCTGCACTGTTGCAAAGCAATGGAGACCGAGCGAGTGGAGAGACAGGAGG  2200
       E  V  R  S  E  Y  I  V  D  D  A  P  C  A  A  A  F  L  H  C  C  K  A  M  E  T  E  R  V  E  R  Q  E

2201  ACAACCTTCAACTGGCCCGGAGCGAGGAAGACGACAGTTACATTGACAGTGATGAGATTACCTCCCGATCCAATTTCCCCGAAAGTTGGCTGTGGCTGGA  2300
      D  N  L  Q  L  A  R  S  E  E  D  D  S  Y  I  D  S  D  E  I  T  S  R  S  N  F  P  E  S  W  L  W  L  D

2301  CCAAGTCTTGCCCACCTGCCCTCAAAACACACCCAACTGTCGTTCAACGTCCTTCGAGAAAACCGTTCCCTTGCAAGATTCTATCACAACCTGGCAGTTT  2400
        Q  V  L  P  T  C  P  Q  N  T  P  N  C  R  S  T  S  F  E  K  T  V  P  L  Q  D  S  I  T  T  W  Q  F

2401  ATCGGCATCAGTCTGTCCAAAACCTATGGAATCTGCGTAGCTGAGCCATTAGAAGTGATTGTCCGCAAGGAATTTTTCATAGATCTGAGGCTGCCCTACT  2500
       I  G  I  S  L  S  K  T  Y  G  I  C  V  A  E  P  L  E  V  I  V  R  K  E  F  F  I  D  L  R  L  P  Y

2501  CTGCTGTCAAAGGAGAACAGATAGAAGTAAAGGCCATCCTTCACAACTACAGCCCTGATATTATCACCGTGCGAGTGGAGCTGCTTGAGGAAGAAGATAT  2600
      S  A  V  K  G  E  Q  I  E  V  K  A  I  L  H  N  Y  S  P  D  I  I  T  V  R  V  E  L  L  E  E  E  D  I

2601  TTGCAGTTCAGCCTCAAAGCGGGGAAGGTACCGACAAGAGGTGGAAGTTGGAGAACATTCAACGCGAGCTGTACCCTTCGTTATTATTCCCATGAAGGAA  2700
        C  S  S  A  S  K  R  G  R  Y  R  Q  E  V  E  V  G  E  H  S  T  R  A  V  P  F  V  I  I  P  M  K  E

2701  GGACAGTTTAGCATCCACGTCAAAGCTGCAGTTCGAGACTCGTCGCTCAGGGATGGAATTGAGAAGAAGCTGCGAGTGGTGCCTGCAGGAGTGTTAGTCA  2800
       G  Q  F  S  I  H  V  K  A  A  V  R  D  S  S  L  R  D  G  I  E  K  K  L  R  V  V  P  A  G  V  L  V

2801  AGACACCAGTAGTTGTAACTCTGGACCCAATTAAAAAAGGCCAAGACGGAGTCCAAACTGAAACCATCAACAGTGCGATCCCCAAGACAGACGTCGCTCC  2900
      K  T  P  V  V  V  T  L  D  P  I  K  K  G  Q  D  G  V  Q  T  E  T  I  N  S  A  I  P  K  T  D  V  A  P
```

Supplementary Fig. S3

```
2901 AAACACACCAACAAGCACACAGATCTCTATAACAGGGAAAGAAAATCTGGCGGGATTAAGCAGTGGTAACAACGCAGAGTACGCGGGGGGAACACTCATT  3000
       N  T  P  T  S  T  Q  I  S  I  T  G  K  E  N  L  A  G  L  S  S  G  N  N  A  E  Y  A  G  G  T  L  I

3001 TACCAGCCATCTGGCTGTGGAGAGCAGAACATGATCCACATGACCTTGCCCGTGATAGCGGCCACCTATTTGGACAAAACCAAGCAGTGGGAGGCTGTCG  3100
      Y  Q  P  S  G  C  G  E  Q  N  M  I  H  M  T  L  P  V  I  A  A  T  Y  L  D  K  T  K  Q  W  E  A  V
               Thioester site
3101 GCTTCCAGAAGCGTAACGAAGCCATCAAACACATACAGACTGGTCTAAATAATCAGCAAGCCTACGCCAAAAAAGACGGATCATTTGCTGTATGGCCCAA  3200
     G  F  Q  K  R  N  E  A  I  K  H  I  Q  T  G  L  N  N  Q  Q  A  Y  A  K  K  D  G  S  F  A  V  W  P  K

3201 ATATCAAAGCAGCACATGGCTAACAGCGTACGTTGCCAAGGTTTTCTCGATGGCCAACAGTCTGGTCGCCGTGCCGGGAGGGCAGATTTGCAACGCCATC  3300
       Y  Q  S  S  T  W  L  T  A  Y  V  A  K  V  F  S  M  A  N  S  L  V  A  V  P  G  G  Q  I  C  N  A  I

3301 AAGTATCTGATTCTCACATCACAAAAACCTGACGGCAGGTTTGAAGAACGCGGGAGGATGTACCACACGGAGATGATTGGTGACGTGACTGGTATCGATG  3400
      K  Y  L  I  L  T  S  Q  K  P  D  G  R  F  E  E  R  G  R  M  Y  H  T  E  M  I  G  D  V  T  G  I  D
                                                           Catalytic histidine
3401 CAGATGTCTCCATGACAGCCTTCTGTGTCATCGCCATGCAGGAGTCAGGTCCACTATGCACAGCGTCTCTGACGAGTCTGCCAGGAAGTATTGCCAAAGC  3500
     A  D  V  S  M  T  A  F  C  V  I  A  M  Q  E  S  G  P  L  C  T  A  S  L  T  S  L  P  G  S  I  A  K  A

3501 GGTTGGTTTTCTTGAACAGCGATTGCCACAAGTTGGCAACCCTTACGCCGCCGCCATAGCGTCCTATGCCCTGGCCAATGAGAACAGATTCAACCGCAAC  3600
       V  G  F  L  E  Q  R  L  P  Q  V  G  N  P  Y  A  A  A  I  A  S  Y  A  L  A  N  E  N  R  F  N  R  N

3601 ATCCTCTACAAGCACGTCCACCAAGGGTTGGACCACTGGCCATCGCCTCGTGGACGGGTTATTACCCTGGAGACCACAGCTTACGCTCTTCTGGCGCTGG  3700
      I  L  Y  K  H  V  H  Q  G  L  D  H  W  P  S  P  R  G  R  V  I  T  L  E  T  T  A  Y  A  L  L  A  L

3701 TCAGAGCCCAGTTATTTGAAGACGCCAGACCTGTTGTCAAGTGGTTGGGCAGGCAGCAGAAAGTGGGCGGAGGCTACGGCTCGACTCAGGCCACCATCAT  3800
     V  R  A  Q  L  F  E  D  A  R  P  V  V  K  W  L  G  R  Q  Q  K  V  G  G  G  Y  G  S  T  Q  A  T  I  M

3801 GGTTTACCAGGCGGTGGCAGAGTACTGGACCCACGCCCAAAAACCAGAATATAATCTGGACGTGAACCTCGGGCTTCCCGGGAGGTCAAACCCAATCAGG  3900
       V  Y  Q  A  V  A  E  Y  W  T  H  A  Q  K  P  E  Y  N  L  D  V  N  L  G  L  P  G  R  S  N  P  I  R

3901 ATCAATTTCAACAGGAACAACCACTACACCACTAGAACATCCAAGTTTAATGACATCAACCAGGACATCCAAGTGACGGCGAGGGGAAGAGGAGAGGCCA  4000
      I  N  F  N  R  N  N  H  Y  T  T  R  T  S  K  F  N  D  I  N  Q  D  I  Q  V  T  A  R  G  R  G  E  A

4001 CACTGACTCTGGTGTCGCTGTATTACGCTCTGCCCAAAGAAAAGGAGAGTGACTGCGACAAGTTCAACGTGTCAGTGAAACTGACTGAAGAGCAGTCGCA  4100
     T  L  T  L  V  S  L  Y  Y  A  L  P  K  E  K  E  S  D  C  D  K  F  N  V  S  V  K  L  T  E  E  Q  S  Q

4101 AGGCGACGACTTGACATACAAGCTGACAATCAAAGTCTTGTTCAAGAGTCGCGATCGTGATGCAACAATGTCAATTCTGGACATCGGCTTGTTGACCGGC  4200
       G  D  D  L  T  Y  K  L  T  I  K  V  L  F  K  S  R  D  R  D  A  T  M  S  I  L  D  I  G  L  L  T  G

4201 TTTACGGTTGACACAAAAGACCTGGACCTGTTGTCCACAGGACGTGCCCGCCTTATTTCAAAGTACGAGATGAACAAAGCGCTGTCAGAAAAAGGCTCCC  4300
      F  T  V  D  T  K  D  L  D  L  L  S  T  G  R  A  R  L  I  S  K  Y  E  M  N  K  A  L  S  E  K  G  S

4301 TCATCATTTACCTGGACAAGGTCTCTCACACTCGTCCTGAGGAAATCTCATTCAGGATCCACCAGAAGTTGAAAGTCGGCGTCTTGCAACCAGCTGCCGT  4400
     L  I  I  Y  L  D  K  V  S  H  T  R  P  E  E  I  S  F  R  I  H  Q  K  L  K  V  G  V  L  Q  P  A  A  V

4401 GTCCGTCTATGAATATTACGAAGAAACACCCTGCGTGAAGTTCTACCACCCAGAGAGGAGAGATGGACAGTTCCTGCAGCTCTGCAGAAACAATGAGTGC  4500
       S  V  Y  E  Y  Y  E  E  T  P  C  V  K  F  Y  H  P  E  R  R  D  G  Q  F  L  Q  L  C  R  N  N  E  C

4501 ACATGTGCTGAAGAGAACTGCAGCATGCAAAAGAAGGGCAAGATTGACAATAGTCTGCGCACAGAGAAGTCTTGTGAGACCACGCCCACCAACAAGATCG  4600
      T  C  A  E  E  N  C  S  M  Q  K  K  G  K  I  D  N  S  L  R  T  E  K  S  C  E  T  T  P  T  N  K  I
                                                  C345C domain
4601 ATTACATATACAAAGTGAAAGTGGAGAGTGTCACAAGCGAATTGATCACGGACGTTTACTCGATGAAGATTGAGACCGTCATCAAGGAAGGGACCACTGA  4700
     D  Y  I  Y  K  V  K  V  E  S  V  T  S  E  L  I  T  D  V  Y  S  M  K  I  E  T  V  I  K  E  G  T  T  D

4701 CGTGGCTCCGCTTGGAAAGGTGCGCTCCTACCTGAGTTACAAACAATGCAAAGAGGCGTTGGGACTGCAGACTGGCAGATCTTACCTCCTCATGGGGACG  4800
       V  A  P  L  G  K  V  R  S  Y  L  S  Y  K  Q  C  K  E  A  L  G  L  Q  T  G  R  S  Y  L  L  M  G  T

4801 TCCAAAGACATACATAAAAATGGAGCAAAGTACGAGTACGTGATTGGCGAGAAGACCTGGGTGGAGTACTGGCCCACTCCAGAGGAGTGTCAGACTAGCG  4900
      S  K  D  I  H  K  N  G  A  K  Y  E  Y  V  I  G  E  K  T  W  V  E  Y  W  P  T  P  E  E  C  Q  T  S

4901 CTTTCAGGGAGGTCTGCGTCGGCATCAGCGAGATGGAGCAGCAGTACTTGTACTTTGGATGTCCGCAGAAGTGAACAGCCGCAGTGTTCTGAGGAAAGTT  5000
     A  F  R  E  V  C  V  G  I  S  E  M  E  Q  Q  Y  L  Y  F  G  C  P  Q  K     *

5001 TCACAGTAGAAAATGCACAATGTTCTCAGATGTTTGTATCCAGTTTTCCAAAAGATGTGCGTGTCTTATAAGTAAAAACAATGGTGCTAAAAAAAAAAAA  5100
5101 AAAAAAAAAAAAAAAAAA  5118
```

PsfuA2M-1

```
   1 TAGTTAATTATGCTAAATTCAAAGTATGTTAGCTTAACATTGCTACTCAACTTTCTGCTACTTATTAGTCCAAATGTTTTTCCACCTAGACGTTATCCTC  100
               M  L  N  S  K  Y  V  S  L  T  L  L  L  N  F  L  L  L  I  S  P  N  V  F  P  P  R  R  Y  P

 101 AAAGTTTGTTAAAAGTATACCGACCAATTAATGTTTATCCAGGCCAGAAAAATTTTTATTATGGATTTTCAAAAAATATTAATTGTGAAAGTGGAAGTGG  200
     Q  S  L  L  K  V  Y  R  P  I  N  V  Y  P  G  Q  K  N  F  Y  Y  G  F  S  K  N  I  N  C  E  S  G  S  G

 201 AATTAAAAGTGAAATTAAAATTACAAATGGAATAGATAATTTTCCACTTATTTTTCCAACTGATATTGAAGAAACACCAAAAAATAGAAAATTAATCGAT  300
       I  K  S  E  I  K  I  T  N  G  I  D  N  F  P  L  I  F  P  T  D  I  E  E  T  P  K  N  R  K  L  I  D

 301 AAATTTGAAGTGCAACCCATAAAAGTTGGTAATTTCTGTTTTATTGCATTTGAAAAAATTTTGCCTGATGGTTTTGATTCATATTCCATAAATTTGATTC  400
      K  F  E  V  Q  P  I  K  V  G  N  F  C  F  I  A  F  E  K  I  L  P  D  G  F  D  S  Y  S  I  N  L  I

 401 AAGATAAGAAAGAATTTAGTTTTGGAAATCGAATTTTTACTAAAAATATAAAACCTTTCACATTGTTACAACTTAACAAGCCAATTTTCAGGCCTGGAGA  500
     Q  D  K  K  E  F  S  F  G  N  R  I  F  T  K  N  I  K  P  F  T  L  L  Q  L  N  K  P  I  F  R  P  G  E

 501 AAATCTTAAAATTTGGCTCAAAGTATTTTATCCTAAAGAAATGGAATTTTTGATAAGGAACAAAAATACAAAAAAATCAAGCATGAATAAAGTGAAAGTG  600
       N  L  K  I  W  L  K  V  F  Y  P  K  E  M  E  F  L  I  R  N  K  N  T  K  K  S  S  M  N  K  V  K  V

 601 AATATTTTAGATTCTAAAGTGCGCAAAGTTGACTTTTTTGAATCTGACATGAATAATGACATGTTTGTTGAGCACAAAATCAATAGAGAAAGTTTCTTAG  700
      N  I  L  D  S  K  V  R  K  V  D  F  F  E  S  D  M  N  N  D  M  F  V  E  H  K  I  N  R  E  S  F  L
```

Supplementary Fig. S3 (continued)

```
 701  GACAATGGAAAATTGTAGTTTTTCTCAATGGTGTAGTGATCGAAACAAAACGTTTCAAAGTGGATCAATATCTCATGCCCACTTTGGAAGTATCTTTGAA  800
      G  Q  W  K  I  V  V  F  L  N  G  V  V  I  E  T  K  R  F  K  V  D  Q  Y  L  M  P  T  L  E  V  S  L  K

 801  AAAGATTGCAGAAAAAAATATGACTGTAACGTATGAAATGTGTGCAGAAGATAACGAAAATAGGCCAATCATTGGAATAGCAAAAACTAAATTTTGTATA  900
        K  I  A  E  K  N  M  T  V  T  Y  E  M  C  A  E  D  N  E  N  R  P  I  I  G  I  A  K  T  K  F  C  I

 901  AGAAGAAGATCGATGAGTTGCAATGATAAAATATTTAAATTTCAAAATGAAAAATGCATGAAAAAAGATTTTGAAATAAATTCGAACGATTTGGACATTT  1000
       R  R  R  S  M  S  C  N  D  K  I  F  K  F  Q  N  E  K  C  M  K  K  D  F  E  I  N  S  N  D  L  D  I

1001  ACAGTCACTTTTTACAAGTTGATTTCATAGAAAGTGTGACTGGAATAAAACAGAGAAGAAAAGCATATTCAAATAGTGGATTTTTTGGAGATTTTGTAGT  1100
      Y  S  H  F  L  Q  V  D  F  I  E  S  V  T  G  I  K  Q  R  R  K  A  Y  S  N  S  G  F  F  G  D  F  V  V

1101  GAAAAAATATCAATACACATCAAACGATAAAAGTGGTGAGAGATTAAAAATTGATTTAATAACTAAATTTGCCGAAGCACCAATTTTGAGGTTAGACTTG  1200
        K  K  Y  Q  Y  T  S  N  D  K  S  G  E  R  L  K  I  D  L  I  T  K  F  A  E  A  P  I  L  R  L  D  L

1201  AAAGAAGATTTAACTATCAAATTTAAGATAAAATCAGCAAATTACGAATTTGAAAAAAAAATCTTAAAAGGAAAAAGCGAATTGAAAATTCCTATTTCTG  1300
       K  E  D  L  T  I  K  F  K  I  K  S  A  N  Y  E  F  E  K  K  I  L  K  G  K  S  E  L  K  I  P  I  S

1301  AAAATTCTTTCAAAGTAATAGGAAAAAATAAATTTGATGAAATCGTATTTAATGAATATTTCAAATGCGATTGGGACAAGTTGAATTATAAATTAGAAAA  1400
      E  N  S  F  K  V  I  G  K  N  K  F  D  E  I  V  F  N  E  Y  F  K  C  D  W  D  K  L  N  Y  K  L  E  K

1401  AATTGGAGGCGAATGTTCAAATGTATTCAAATTCAAACAGTTTGGAAGAAAGAAAAATATCAAAGTAATGTTGAGTTACAAGAAATATGCAATACTCGAT  1500
        I  G  G  E  C  S  N  V  F  K  F  K  Q  F  G  R  K  K  N  I  K  V  M  L  S  Y  K  K  Y  A  I  L  D

1501  GAAGATGAAAGTAATTTCGAAGAAAAATGCATCAATGAAGATTGTTTCAAAATTTTCGAAGATGAAATTCAGCTGGAAAACACTTTCATATTCGACAACA  1600
       E  D  E  S  N  F  E  E  K  C  I  N  E  D  C  F  K  I  F  E  D  E  I  Q  L  E  N  T  F  I  F  D  N

1601  AATTTTATGGAATCAATAGTCAAGCTTATGATAAAAGTTTTGAAGTAATTTATGACGATAAATGTTTGACAAAATATCAATTGGATTCTTCTGAATCTAA  1700
      K  F  Y  G  I  N  S  Q  A  Y  D  K  S  F  E  V  I  Y  D  D  K  C  L  T  K  Y  Q  L  D  S  S  E  S  K

1701  AGGAATAACTTTTAGTGAAAATAATGTTAAGCCTGGCCAAAGTATTGATTTGAAAATGGAACTTTCGAAAAAGCCAGATTATTGCTTGATTGCTTCGTAT  1800
        G  I  T  F  S  E  N  N  V  K  P  G  Q  S  I  D  L  K  M  E  L  S  K  K  P  D  Y  C  L  I  A  S  Y

1801  GATAAGAATTTAGTGAATTTAGTCAAAAAAGAACAAATTGCAATTTATTTGAAAACAATCAAAGATTCATTGAAGAATTATCTGATTGGCAATTTAGGAA  1900
       D  K  N  L  V  N  L  V  K  K  E  Q  I  A  I  Y  L  K  T  I  K  D  S  L  K  N  Y  L  I  G  N  L  G

1901  TGAGTTTGATGAAAGATCTAGATTTCTTATATGAAACAGTATTAAGATCAAATCAAGCTATATATATGAAAAAAGATAATTCAGTTAAGGTTAGAAAACA  2000
      M  S  L  M  K  D  L  D  F  L  Y  E  T  V  L  R  S  N  Q  A  I  Y  M  K  K  D  N  S  V  K  V  R  K  H

2001  TTTTCCGGAAAGTTGGATATTTGAAAAGTATTCTGAGTTTCCAATTAAAATCAAAGCTCCTGATTCCATAACAAAATATCAAACAACCGGAATTTGCTTT  2100
        F  P  E  S  W  I  F  E  K  Y  S  E  F  P  I  K  I  K  A  P  D  S  I  T  K  Y  Q  T  T  G  I  C  F

2101  CAAAAAATGATTTGGGTTTCTCCGACAGACACTTCTTTAACAATTTCTCAAGATTTTTATTCATACTTGAATCTTCCTAAATATTTGATTGGTTCTGAAG  2200
       Q  K  M  I  W  V  S  P  T  D  T  S  L  T  I  S  Q  D  F  Y  S  Y  L  N  L  P  K  Y  L  I  G  S  E

2201  TAGCTTTAATCACAGCAACACACTACTTGAACAAAACAATGATTGGTAATGGAAAAAATTATGTGAAATTCGATTTAGAAGTGAATAAAGATGTTACAGA  2300
      V  A  L  I  T  A  T  H  Y  L  N  K  T  M  I  G  N  G  K  N  Y  V  K  F  D  L  E  V  N  K  D  V  T  E

2301  GGTACAAAACAATAAGGAATTCAGCTTTTGCACTAACAAAAACGATTTAAACATACACAGTGCTATTTTCTTAGAATACGCTAAGGAATTTGGAACATTA  2400
        V  Q  N  N  K  E  F  S  F  C  T  N  K  N  D  L  N  I  H  S  A  I  F  L  E  Y  A  K  E  F  G  T  L

2401  AAAATGAGACTTAATTCTCAAATCTTTATTGATTCTGTGAAATGTGGCGGCTCGAAATCGGATGAAAAAGCTAATATTTTATATGATATTGTTGAAAAGA  2500
       K  M  R  L  N  S  Q  I  F  I  D  S  V  K  C  G  G  S  K  S  D  E  K  A  N  I  L  Y  D  I  V  E  K

2501  AAATTGAAGTTTTTCCATCAGGCGAGAAGACAGAAATTGTTAAAAGTGGTCTTATTTGTACTAATCAATTAAATCAAGTTATTGATTTGCCCATTGATCA  2600
      K  I  E  V  F  P  S  G  E  K  T  E  I  V  K  S  G  L  I  C  T  N  Q  L  N  Q  V  I  D  L  P  I  D  Q

2601  AATTGTTCCAAAAAGTTTAAAAATGAGTGTCACTTTAGCCACTGATGGAACTCAAATGATTTCTGAAAACTTGAAAAACATGATAAGACAGCCTAGTGGA  2700
        I  V  P  K  S  L  K  M  S  V  T  L  A  T  D  G  T  Q  M  I  S  E  N  L  K  N  M  I  R  Q  P  S  G
                                                                                                Thioester site
2701  TGTGGAGAGCAAAACATTGCGCTTGTAATTCCTTCAATTTCTGCTTTATCTTATTTGCAAAATGCAAACATTGTGATGGAGGAGGAAAAAAAAGAAAAGT  2800
       C  G  E  Q  N  I  A  L  V  I  P  S  I  S  A  L  S  Y  L  Q  N  A  N  I  V  M  E  E  E  K  K  E  K

2801  TGAAACTTAAAGCAACAAAGTTCATTTTGCAAGGTTATGTGAGACAATTGAAGTATCAGCATAAAAACATGGGTTTTTCAGCTTTTGGAGAATCAGATGG  2900
      L  K  L  K  A  T  K  F  I  L  Q  G  Y  V  R  Q  L  K  Y  Q  H  K  N  M  G  F  S  A  F  G  E  S  D  G

2901  AAATGCTTCAACTTGGCTCACAGCTCACGTGCTTGAAACTTTCAATGAAGCATCGAAATACTTTAAAAAAGATGAAAACTTTGATTATGCAATTTTGAAA  3000
        N  A  S  T  W  L  T  A  H  V  L  E  T  F  N  E  A  S  K  Y  F  K  K  D  E  N  F  D  Y  A  I  L  K

3001  AGTAGTTTCAGATTTCTTAAAAAGGTTCTTGGTGGCAAACAATGTGTGAAAGAGAATGGACGAGTTTTGCACAGAGATATAATTTCGAATGACGAGTATG  3100
       S  S  F  R  F  L  K  K  V  L  G  G  K  Q  C  V  K  E  N  G  R  V  L  H  R  D  I  I  S  N  D  E  Y
                                                                   Catalytic histidine
3101  CTCGCTATGAACACTTAAATTTAATTGGCAAAATAATCACAGCTTTGAGCAGTTTGAATAAAAACAAAAATGCTCATCCTCATTTTGAGTATAACAAAAA  3200
      A  R  Y  E  H  L  N  L  I  G  K  I  I  T  A  L  S  S  L  N  K  N  K  N  A  H  P  H  F  E  Y  N  K  K

3201  AGCAATTAAGCAAACTTTAGAGAATGTCCTCACTTGTGTAACTGAAAATTACAAAAAAGAATTTTCGAAAGAAAATATTCAAATGATGCCAGCTGTTTTG  3300
        A  I  K  Q  T  L  E  N  V  L  T  C  V  T  E  N  Y  K  K  E  F  S  K  E  N  I  Q  M  M  P  A  V  L

3301  TTGGCTAGATGGAATTACGCTTTGAATTCACTTTCTGATCGAGACAAATCGATATTTAATCAAATTGAATTTGAAAATAAGTTAAAATCGATTGTGAAAG  3400
       L  A  R  W  N  Y  A  L  N  S  L  S  D  R  D  K  S  I  F  N  Q  I  E  F  E  N  K  L  K  S  I  V  K

3401  AAATAACTTCGAATATAGCTGATGAAAAAGGAATGCAAAATGCAGAAAAAGTTGAAGTTATTTCTTTGTATTTAATGACTATTGATGATAAAGAAGTTGA  3500
      E  I  T  S  N  I  A  D  E  K  G  M  Q  N  A  E  K  V  E  V  I  S  L  Y  L  M  T  I  D  D  K  E  V  E

3501  ATTGCAACTTATTTCTTGGTTGTTAAGTCAACAAAATAGCAGAGGAGGATATTATTCGTCGTGGGACACTTCGTTAGCTGTACGAGCTTTAGCTGCTAAA  3600
        L  Q  L  I  S  W  L  L  S  Q  Q  N  S  R  G  G  Y  Y  S  S  W  D  T  S  L  A  V  R  A  L  A  A  K

3601  ACAGTTAATCCAATCGAACAAAGCAAATCAGTGACATTCAGCATGGATGGGGTACAGAAAAAAATGCAAATCGAGCCGAATAGCAAGACAATATTGATTG  3700
       T  V  N  P  I  E  Q  S  K  S  V  T  F  S  M  D  G  V  Q  K  K  M  Q  I  E  P  N  S  K  T  I  L  I
```

Supplementary Fig. S3 (continued)

```
3701  AAAATAAATTTATGAAAGAGAAAAAATCATTCGTTATCAAATGGACAACTGATTTGAAAGAAAGTAAAAATGAATGTTTAGTTGGTCGAGTGGCGGCTAT  3800
      E  N  K  F  M  K  E  K  K  S  F  V  I  K  W  T  T  D  L  K  E  S  K  N  E  C  L  V  G  R  V  A  A  I

3801  TTTTTACTCAGAGAAAAGCAATAAAAAAGAAAAGGAATATTTCAAAATAAATCACAAAATTATTGAAAATGAAAAATTGAAAATGTTTAAATTGAATATC  3900
        F  Y  S  E  K  S  N  K  K  E  K  E  Y  F  K  I  N  H  K  I  I  E  N  E  K  L  K  M  F  K  L  N  I

3901  GATGTGGAACCAATCGATAGTAAAGCAAGATCTGGAATGTTATTGGTTGAAGTGAGCATTCCAAGTGGATTTAAAGTGGTTTATAATAATTTGGATAAAT  4000
       D  V  E  P  I  D  S  K  A  R  S  G  M  L  L  V  E  V  S  I  P  S  G  F  K  V  V  Y  N  N  L  D  K

4001  TGATTGGAAAAGGAGTTGAAAGGCCAACAAGATATGAAAAGTATTTCACTGCAACTCGAATTCAATTATACTTTTGGGGATTCGGTAGTGTGAACAGGAC  4100
      L  I  G  K  G  V  E  R  P  T  R  Y  E  K  Y  F  T  A  T  R  I  Q  L  Y  F  W  G  F  G  S  V  N  R  T

4101  AAGATCTTTCCAACTTGAATTTGTGCAACAAGTCAAAGTTAAAGATGCAGCATTGGGAGTGATCCACGTGGAAGATTATTATTCGTCAGAGAAAAATGCA  4200
        R  S  F  Q  L  E  F  V  Q  Q  V  K  V  K  D  A  A  L  G  V  I  H  V  E  D  Y  Y  S  S  E  K  N  A

4201  GCATCCTACTATGAAGTTCCTTTTGTGAATGAAAATAAGGAAGTTAATGCGCAAATAGTGGAAATATTGAATAATTTGAATGACAAAGCAATCGATAATT  4300
       A  S  Y  Y  E  V  P  F  V  N  E  N  K  E  V  N  A  Q  I  V  E  I  L  N  N  L  N  D  K  A  I  D  N

4301  TAGATGAACATTTGAACAAATTTTCGGAAGAAATTTGTGAACAAAAATTGAAGGAAGAAGTGAAAGTGGTGATTTTGAAAGATGAAAAGCAAATTGGATT  4400
      L  D  E  H  L  N  K  F  S  E  E  I  C  E  Q  K  L  K  E  E  V  K  V  V  I  L  K  D  E  K  Q  I  G  L

4401  AAAATTAAGTGAAAATGATAAAATTGAAGTTGTCAAGATGCCTGCAGTGAAAAAAATGTCTAAAAAAATCATTTTTGTTGCAAATAATGGCAAAAATCAA  4500
        K  L  S  E  N  D  K  I  E  V  V  K  M  P  A  V  K  K  M  S  K  K  I  I  F  V  A  N  N  G  K  N  Q

4501  TTTTATCCCATTGAAAATGAATTAGTAATGAAAAATTACTTAGAGACTTTCATGAAAAAGTCCAAAAGATGCATTGCTCTTTCGAAACTTTTTAAAAATG  4600
       F  Y  P  I  E  N  E  L  V  M  K  N  Y  L  E  T  F  M  K  K  S  K  R  C  I  A  L  S  K  L  F  K  N

4601  TCATGTACTTATTGATTTGATTCATATTTT  4630
      V  M  Y  L  L  I  *
```

PsfuA2M-2

```
   1  GAAATGAAAATTCTTTTATGTCTTCTGCTTGTTACAAGCATATGGGCAAAGCGTGCACCCATCATCATAGGTTAACAAGATACTGAAGTGTATCTTCCCC  100
 101  AAACATTTATTCAAGGAATCGAACAACAAACCTACATTACACATAATACGAAAAATGGAAAAATTCCAACTTCTGTGACAATGGAACTTTATTATCAAAA  200
                                                                                     M  E  L  Y  Y  Q  N

 201  TAAATTAGCTTGCAGAGTAGTACAAGAAATGAGACCATCGATGAGAAATAATCAAATTTATGAAGCTTTAATCAAACACGGCTTTTCAAGGCCATGGATT  300
        K  L  A  C  R  V  V  Q  E  M  R  P  S  M  R  N  N  Q  I  Y  E  A  L  I  K  H  G  F  S  R  P  W  I

 301  CAAGATAATCAATATCTCAATTTGAGAATCAAAGTATCTGGAAGAATGATATTTTATTCTAAAAAAGTAGTGCATGCTGAAAGAAGAATGATGAGCTATA  400
       Q  D  N  Q  Y  L  N  L  R  I  K  V  S  G  R  M  I  F  Y  S  K  K  V  V  H  A  E  R  R  M  M  S  Y

 401  TTGGACACCATAATAACATCATTTACACAAAATCGATGAATATTAAAATGAGAAAGTTGAGCAAATTAACCTTTTTCGAGTTGGACAAAAAATATTACAG  500
      I  G  H  H  N  N  I  I  Y  T  K  S  M  N  I  K  M  R  K  L  S  K  L  T  F  F  E  L  D  K  K  Y  Y  R

 501  ACCTGGAGAAAGCATTAAAGTTTGGATTAAACAAGTTGACCATTCTCAGTTGGTTGATTTGAAAAAAATGACTAAAAATAGTTGCTTTGGTGGAAAAGGT  600
        P  G  E  S  I  K  V  W  I  K  Q  V  D  H  S  Q  L  V  D  L  K  K  M  T  K  N  S  C  F  G  G  K  G

 601  GCTTCAAATATGAAATTGCCTGAATTAACATTGAACATTGAGTATCCTAATGGAATGAAATTTGAAAAGAAGAAAATTGGCAACACAAATCAAGATTTGT  700
       A  S  N  M  K  L  P  E  L  T  L  N  I  E  Y  P  N  G  M  K  F  E  K  K  K  I  G  N  T  N  Q  D  L

 701  TCAAATTTGAAATTCCTAGTGAAACTTTGATGGGAAAATGGAAGGCTTATTTGGAATATGGAAAAGAAAAAAGTTATAAATATTTCGAGGTAAAATATTT  800
      F  K  F  E  I  P  S  E  T  L  M  G  K  W  K  A  Y  L  E  Y  G  K  E  K  S  Y  K  Y  F  E  V  K  Y  F

 801  CACTGTTCCTCGTTACAGTGTTACACAAAAATTGATTCCTTTGTCAATGAATGAGGTGAGTGTGGAAGCTTGTGCCAAGATGAGTAATGCAAGACAATTC  900
        T  V  P  R  Y  S  V  T  Q  K  L  I  P  L  S  M  N  E  V  S  V  E  A  C  A  K  M  S  N  A  R  Q  F

 901  AAAGGAACAGTGGAAATCAAATCCTGCATAAATTCCCGAAGATGTGTTACAAAGAATATTTCTTTTTCCACAACTAAATGTGTCACAGCTAAAATACATT  1000
       K  G  T  V  E  I  K  S  C  I  N  S  R  R  C  V  T  K  N  I  S  F  S  T  T  K  C  V  T  A  K  I  H

1001  TTAAAAACACATATTTCAGCAAATGGGGAGTAAACTTTAGTGCTGAATTTAAATACTTTGAATACAAAGGAGCAGAACCAATTATTAAGAAGTTAAGCTG  1100
      F  K  N  T  Y  F  S  K  W  G  V  N  F  S  A  E  F  K  Y  F  E  Y  K  G  A  E  P  I  I  K  K  L  S  C

1101  TTATTATCGATCATCAAGAATAGTTGCAAGACCTTATCAAATTATAAAATTAACTGCCTATAAAAATTTCTATGGTTTGATACATTTCAAAACCTTGAAA  1200
        Y  Y  R  S  S  R  I  V  A  R  P  Y  Q  I  I  K  L  T  A  Y  K  N  F  Y  G  L  I  H  F  K  T  L  K

1201  ATGGGTGCATACAGATTGGAATTGATCTCATCGAAAAACAAAGATTTGATTTGGAAAAGTTCATTTACTTATTCCAAAGATTTAAAAACATTTTTTGTGA  1300
       M  G  A  Y  R  L  E  L  I  S  S  K  N  K  D  L  I  W  K  S  S  F  T  Y  S  K  D  L  K  T  F  F  V

1301  AACAATTGAATGAACAGGCATATTTACACATTCACTTCAACAACAAATTAATCGAAAAACGACACATTAGTGTGAGCTCTGTAAAGTTTATGATAAACTA  1400
      K  Q  L  N  E  Q  A  Y  L  H  I  H  F  N  N  K  L  I  E  K  R  H  I  S  V  S  S  V  K  F  M  I  N  Y

1401  TGACGAATCACAAATGGAAAAAATGTGTAGAAAATTTCAAAGTCAAAATTTACTCAATCCATAAGATTCATCAGTGTATGCAATTTCTAACATACAATGAA  1500
        D  E  S  Q  M  E  K  C  V  E  N  F  K  V  K  I  Y  S  I  H  K  I  H  Q  C  M  Q  F  L  T  Y  N  E

1501  TTGATTACTTTGAAGCCAATGTTGGTGGCCAGCAACAGTAATATGTATTTATACACTGTGAATATCAGCTTGAAACATTTCAGACATATCCAATTGAATG  1600
       L  I  T  L  K  P  M  L  V  A  S  N  S  N  M  Y  L  Y  T  V  N  I  S  L  K  H  F  R  H  I  Q  L  N

1601  GAACACTTCTTTTTGCTGACTTTACTTCGCAAAAAATAACCATCAAATCGCTGAACAATTGTTTGATTAGAAGTCAAACAAACTTTGGTAAAATATATTT  1700
      G  T  L  L  F  A  D  F  T  S  Q  K  I  T  I  K  S  L  N  N  C  L  I  R  S  Q  T  N  F  G  K  I  Y  F

1701  CAACAAAAGTAAAGTTTATCCTGGCGACGCCTTAAAGTTGAATGCCAACTTTATTGAAAAACCAAAATTATGCTTCATCAACAGTTTTGACCAGAATTTG  1800
        N  K  S  K  V  Y  P  G  D  A  L  K  L  N  A  N  F  I  E  K  P  K  L  C  F  I  N  S  F  D  Q  N  L

1801  GAAAATATGGATGTCAAACAAGTTGGCACTTTGTATTCAAAATATTTTTCGATATTAATCAATTATTCGAACAGAAGTTTGGGAATGAATTTATACAAAC  1900
       E  N  M  D  V  K  Q  V  G  T  L  Y  S  K  Y  F  S  I  L  I  N  Y  S  N  R  S  L  G  M  N  L  Y  K

1901  CAATAATTTATTATCCCAGCTACAAAAGACCTAGACCTATGCGTAATCAAAATAGGGAAAAATTAATGGGTGATGGATTAGACAATTTAAATTTGAATCT  2000
      P  I  I  Y  Y  P  S  Y  K  R  P  R  P  M  R  N  Q  N  R  E  K  L  M  G  D  G  L  D  N  L  N  L  N  L
```

Supplementary Fig. S3 (continued)

```
2001  GGTTCAAGGAGTTATTTTGAGGGAATATTTTCCTGAATCTTGGCTTTTTAAAAGCATCAAATCTTTTCCCCTTTCAATCACTGCTCCACACACAATTACT  2100
        V  Q  G  V  I  L  R  E  Y  F  P  E  S  W  L  F  K  S  I  K  S  F  P  L  S  I  T  A  P  H  T  I  T

2101  ACTTACAAAACATCAGCTGTTTGTTTCTTTGAAAAAAGCATTTGGATTTCCTCAACTTATGCTTTACTGAAGGTTCATAAAGATTATTATGTTACAGTCA  2200
       T  Y  K  T  S  A  V  C  F  F  E  K  S  I  W  I  S  S  T  Y  A  L  L  K  V  H  K  D  Y  Y  V  T  V

2201  AAGCACCAGCTTTCGTTTACATTCAAAATACCTTTGCACTTCGAGTGAATACATTCCTCAAGCAAAGTACTTTGCAAAACATTCGTGTCATATTGAAAGT  2300
      K  A  P  A  F  V  Y  I  Q  N  T  F  A  L  R  V  N  T  F  L  K  Q  S  T  L  Q  N  I  R  V  I  L  K  V

2301  ATCTGTCTCTTCTTCAGATTTCCAAATTCTCGAATCAAAACAGAAAGTTTTTGATCAAAAATCAAATGAAGTGAAAGTTAATACTTTTATGTTGAAAGCT  2400
        S  V  S  S  S  D  F  Q  I  L  E  S  K  Q  K  V  F  D  Q  K  S  N  E  V  K  V  N  T  F  M  L  K  A

2401  CTTAAATACAGCAAAGGATTGAAAGTGATTGTCTCTGCTAAGTTTTATTTAATAAATGGATTCGCTTTGAAATATGAAGATGCAGTGAAAGTGTCTATTC  2500
       L  K  Y  S  K  G  L  K  V  I  V  S  A  K  F  Y  L  I  N  G  F  A  L  K  Y  E  D  A  V  K  V  S  I

2501  CTTCGATTCCTAATGGAGATCGTCAAGAATTAGTTCAAACGGTGTTCTTATGCAAAGACAAACTGTATTCTGGTGTGTTAGATTTCTCGTCAATAAAATC  2600
      P  S  I  P  N  G  D  R  Q  E  L  V  Q  T  V  F  L  C  K  D  K  L  Y  S  G  V  L  D  F  S  S  I  K  S

2601  AATAATTCCTTCTACCATGTCAATCAAAGCTGAAGCAGCAGCTGATGATATGCAATTGACTATGTCTCATTTAAGCAGTTTGCTTCTTCAACCAATGGGA  2700
        I  I  P  S  T  M  S  I  K  A  E  A  A  A  D  D  M  Q  L  T  M  S  H  L  S  S  L  L  L  Q  P  M  G
                                                                                                      Thioester site
2701  TGCGGAGAACAGAATATTGGAAGAGTTGGACCTTCAATTGCAGTCTTGATTTACTTGATAAATCATCAAATGGTTCATATGAAAAGTATGAAAGCAAAAG  2800
       C  G  E  Q  N  I  G  R  V  G  P  S  I  A  V  L  I  Y  L  I  N  H  Q  M  V  H  M  K  S  M  K  A  K

2801  CCATCCGGTACATTCGTCTTGGATTTATTAGACAACAAAAGTACATTCACGGAAATTATTCCTACTCAGCATTTGGAGTGAGAGATAAATTTGGCAGCAC  2900
      A  I  R  Y  I  R  L  G  F  I  R  Q  Q  K  Y  I  H  G  N  Y  S  Y  S  A  F  G  V  R  D  K  F  G  S  T

2901  ATGGCTCACTGCTCATGTTCTTGAATACTTTTCACTCGCCATGCAATACACAGATGTGAACAATGACGACATGCTCAACCTTATTCCAAAAATGTTGCAT  3000
        W  L  T  A  H  V  L  E  Y  F  S  L  A  M  Q  Y  T  D  V  N  N  D  D  M  L  N  L  I  P  K  M  L  H

3001  TTCTTCAGTACTATGAAACAAACTATTGGATTTTTGGAAAAAGGTCGAGTGATTCATCGAGACTTTTATGAATTTTCTCAATCTAAAATTGATTTGTACA  3100
       F  F  S  T  M  K  Q  T  I  G  F  L  E  K  G  R  V  I  H  R  D  F  Y  E  F  S  Q  S  K  I  D  L  Y
                                                    Catalytic histidine
3101  AGAAAATCAATTTGAATGCTAAGATTTTAGATTCTTTAAATCAACTAGAGAAAGCTTTAAGCGTTTATAAGATTAAACAATTCAACAGGACCATTTCAAT  3200
      K  K  I  N  L  N  A  K  I  L  D  S  L  N  Q  L  E  K  A  L  S  V  Y  K  I  K  Q  F  N  R  T  I  S  M

3201  GCTGATAAATCATTCTTTGAAAAGTTTAGCTTCTTTGCTTTCAAGTGTTAAACTGGCTTCAATGTCAACAGTTTTAACAGCTCGTATTCTTTTCGCTTTG  3300
        L  I  N  H  S  L  K  S  L  A  S  L  L  S  S  V  K  L  A  S  M  S  T  V  L  T  A  R  I  L  F  A  L

3301  ACTCATTTGAAAACAAAAACTCTTTTCAGTAGTCAAATGACCACAAGTCTTTTTGCTGTATCGAAAAGTCAAGTTGAAATTATCGTTGGTTACAAGAAAG  3400
       T  H  L  K  T  K  T  L  F  S  S  Q  M  T  T  S  L  F  A  V  S  K  S  Q  V  E  I  I  V  G  Y  K  K

3401  GAAAAATGACTGAAAATTTGTCGAAGAGTAACTCTGTGGAAGTAGTCGCTTATTTCATGATGTCAAGGCACATAATGACTCCTTTATACCACAGTTGCTA  3500
      G  K  M  T  E  N  L  S  K  S  N  S  V  E  V  V  A  Y  F  M  M  S  R  H  I  M  T  P  L  Y  H  S  C  Y

3501  CCTTTACATGTTAAGGGCACAAAATTCTAAAGGTGGATTCATCAGTAGCACTGACACTGTTGTAGCTCTTCGGGCTTTTGCTCAATCTTTATTCAAAACC  3600
        L  Y  M  L  R  A  Q  N  S  K  G  G  F  I  S  S  T  D  T  V  V  A  L  R  A  F  A  Q  S  L  F  K  T

3601  ATCGATAAAGATGAGATAATAAGTTTGAAAATTTTGAACAAAACAAAGGCTAAATTGTTGAAACCATCGTCAAAAGTTAATTTCGATTTGAATATCAATT  3700
       I  D  K  D  E  I  I  S  L  K  I  L  N  K  T  K  A  K  L  L  K  P  S  S  K  V  N  F  D  L  N  I  N

3701  ATTACACTCCTCCTAGCAACAGAAAAGTTTCCTATGTTTGGGAAAAGAAGAGTTCAAGCAATAAATGCGTGATAGCTAAAGTGACAAGTGTTTTCTATGT  3800
      Y  Y  T  P  P  S  N  R  K  V  S  Y  V  W  E  K  K  S  S  S  N  K  C  V  I  A  K  V  T  S  V  F  Y  V

3801  CATGTCTCCCATTTTGATTATTGAAATTTTCAGCATGAAATCAAAGATAATCAAACAGGATGATTTCAGAAAGGTTTATTCAGTTAGTTTATCTTTGAAA  3900
        M  S  P  I  L  I  I  E  I  F  S  M  K  S  K  I  I  K  Q  D  D  F  R  K  V  Y  S  V  S  L  S  L  K

3901  AAAGTTATCAAAATGACTGGAATGTTGCTGATGACAATTAGAATTCCATCCGGATATACTTTCAACATGAATTATTTGACCAATCAAATGAAAAAAGGAT  4000
       K  V  I  K  M  T  G  M  L  L  M  T  I  R  I  P  S  G  Y  T  F  N  M  N  Y  L  T  N  Q  M  K  K  G

4001  ATTTTGTGAAGAAATGTGAAGAAAACATGAAAAAAGGTTACGTCAAAGTTTATTTGTCAGCTCTGGGTGGTAAAGTGCAAAAGATCAGTGTTAATTTTGA  4100
      Y  F  V  K  K  C  E  E  N  M  K  K  G  Y  V  K  V  Y  L  S  A  L  G  G  K  V  Q  K  I  S  V  N  F  E

4101  ACTTTTGCATGAATTCAAAGTGAATTCTTCTCCAAATGTTGTTGTTATTGTAGAAGATTATTACAATCCAGTCGTGAAAGGAGAAATTACAATTAAAAGT  4200
        L  L  H  E  F  K  V  N  S  S  P  N  V  V  V  I  V  E  D  Y  Y  N  P  V  V  K  G  E  I  T  I  K  S

4201  AAATAAATATTATTAAAAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  4250
       K  *
```

PsfuA2M-3

```
   1  CCGAACTTATAAATAAGGCAGACTGCCCCAAAACTTAGTTCACTACAGATGTTGGTCTCAGTGAAAACGTTCGTCTCTCTTGGAGTGCTTCTTCTTTGCG  100
                                                       M  L  V  S  V  K  T  F  V  S  L  G  V  L  L  L  C

 101  AAGGAGTCCCTGCTCTCGGAGCACCACAGGAGCTTCTTCCTACTTCGGAGCCTCCCAAATTCGAAGTTGTTAGAGGAGACAATCACACGGATTTGGAGGG  200
      E  G  V  P  A  L  G  A  P  Q  E  L  L  P  T  S  E  P  P  K  F  E  V  V  R  G  D  N  H  T  D  L  E  G
                                                                                       CUB domain
 201  TGACTCTGGGATATTCAAGTCTCCAAACTTCCCTGAGCAGTATCCTGAGAACTCCAATACAACATGGAAGATCACGGTTCCGAAGGATAAAGATGTCGAG  300
        D  S  G  I  F  K  S  P  N  F  P  E  Q  Y  P  E  N  S  N  T  T  W  K  I  T  V  P  K  D  K  D  V  E

 301  CTAACATTTCATCTATTAAATTTAGAGAAGCATCAGTATTGTACCTGGGACTGGGTGAAGATCTACAATGCGGAGGACTATAGTGAGCTTCGCCAGCTAT  400
       L  T  F  H  L  L  N  L  E  K  H  Q  Y  C  T  W  D  W  V  K  I  Y  N  A  E  D  Y  S  E  L  R  Q  L

 401  GCGGCTATCTCCAACAAGATCTCATTCTACACTTCAACACCAGCAGCATTATCGTGGAGCTTCGTTCGGACTCTGTGTCTCAGAGGACAGGGTTCCTGGC  500
      C  G  Y  L  Q  Q  D  L  I  L  H  F  N  T  S  S  I  I  V  E  L  R  S  D  S  V  S  Q  R  T  G  F  L  A

 501  GTCCTGGCGCCCTATCTCACAGGACTCTTCCGAAGAAGACAACTCGGGAGAGCCCAAAGAGGGATATATTTTCACTTTCCCTCAGTCCTTGACCGTCTCA  600
        S  W  R  P  I  S  Q  D  S  S  E  E  D  N  S  G  E  P  K  E  G  Y  I  F  T  F  P  Q  S  L  T  V  S
```

Supplementary Fig. S3 (continued)

```
 601  GACGAGCCATCTCACGAAGAGGCTTGTTTGGAGATTTTCAATGTGAACACCAAAGGCAAAGTATTCGTCAACGTCTTGTCATCGAATAACATTTTGAACC  700
       D  E  P  S  H  E  E  A  C  L  E  I  F  N  V  N  T  K  G  K  V  F  V  N  V  L  S  S  N  N  I  L  N

 701  TGGAAACGGACATTTATAAGGAAATAGACTACGAACCCGGGAGTCCGGAATCTGAGAAAGTGAAATGCTTTGACATCATGTTACCTATTGACTTCTCAGA  800
      L  E  T  D  I  Y  K  E  I  D  Y  E  P  G  S  P  E  S  E  K  V  K  C  F  D  I  M  L  P  I  D  F  S  E

 801  GAAATACGCCATCATTCAAATCAACGGAACCTTTGAGGATAACGACTATAGGATTCTGAGCTACAAGTCCGTAAAGGTACTTAAGAGCTCTGCAAACATG  900
        K  Y  A  I  I  Q  I  N  G  T  F  E  D  N  D  Y  R  I  L  S  Y  K  S  V  K  V  L  K  S  S  A  N  M

 901  CTCATTCAAACAGACAAGAGCGAATATCGTCCCAAACAAAAGGTATTATTTCGTGTCATGAAGTTGGATGATAATTTGAGACCCTCCAAGAACAACGTAC  1000
       L  I  Q  T  D  K  S  E  Y  R  P  K  Q  K  V  L  F  R  V  M  K  L  D  D  N  L  R  P  S  K  N  N  V

1001  TCAACAAAGTCTTCGTCAAATCACCCAGCCGACAAACCCTCGCACAGTTCATCAATCACTCTTTAGACTCCAGAGGCATTGGGCAATTTGAATTCCAACT  1100
      L  N  K  V  F  V  K  S  P  S  R  Q  T  L  A  Q  F  I  N  H  S  L  D  S  R  G  I  G  Q  F  E  F  Q  L

1101  GGACGAGGAACCAGAGTTGGGAAACTGGCAAATTATGGTAGAAAATGACGTAGTTGGGGGAGAGGAGGACTCTTCTGTTGAGCTAGCCTACTTCGTCGTG  1200
        D  E  E  P  E  L  G  N  W  Q  I  M  V  E  N  D  V  V  G  G  E  E  D  S  S  V  E  L  A  Y  F  V  V

1201  AACGAGGCCGTTCTTCCCAAATTTGAAGTACAAATCAATGGTCCTAAAGCCATTTTGAGTGAGGATGAAAAAGTAGACTTTGAAATTTGTGGCATTTACA  1300
       N  E  A  V  L  P  K  F  E  V  Q  I  N  G  P  K  A  I  L  S  E  D  E  K  V  D  F  E  I  C  G  I  Y

1301  CTCACGGTAGTAAAGTGAAGGGATCTGTGGAAGTCTTCTTCGAACACAAGTACACGAAGGGTTCTGCCTACTGGAGGGCTCCAATGTACTCATCTAATTT  1400
      T  H  G  S  K  V  K  G  S  V  E  V  F  F  E  H  K  Y  T  K  G  S  A  Y  W  R  A  P  M  Y  S  S  N  F

1401  TACCAAGAAGATCGTCATAAAAGACGATGATACATGCGGAACTGTGAGCTTGAACCAAACAGAGATATCCCTTCTCTCCATTGATGGGGATCGGGATTTC  1500
        T  K  K  I  V  I  K  D  D  D  T  C  G  T  V  S  L  N  Q  T  E  I  S  L  L  S  I  D  G  D  R  D  F

1501  TCCATCAACGCAAAATTAACTGAAGAAGGGACTGACATTGTTCAAGAGAGTTCTTCCAATGAAAAAGTCCTCTTTACTCGTGCGAAAATTGATTTCGGAG  1600
       S  I  N  A  K  L  T  E  E  G  T  D  I  V  Q  E  S  S  S  N  E  K  V  L  F  T  R  A  K  I  D  F  G

1601  ACAGCTCCTATGAACATATTCTTGGTGAGTATCCCTACGTCATATTTTTCAAAGTCGTTGAACACGGAACCAACCAGGCCCTGAAGGGAGAAAAGGTTGA  1700
      D  S  S  Y  E  H  I  L  G  E  Y  P  Y  V  I  F  F  K  V  V  E  H  G  T  N  Q  A  L  K  G  E  K  V  E

1701  GATTTGCTCAATTCTATGGAAGGATATAACTCAATTCACTCAATTTGTGTCTCGAAGCAACTTCTACAGTTTTGACGAGTTTCACCTCTATGATTTGGGT  1800
        I  C  S  I  L  W  K  D  I  T  Q  F  T  Q  F  V  S  R  S  N  F  Y  S  F  D  E  F  H  L  Y  D  L  G

1801  CAAAGGCTCATCGATATTAAGTTCCAGGAGACTTGTATCAATGAAACAACTTCAGAGGTAGATGGATCTCTCACACTCGGTATTCCACTCACGGGACTTC  1900
       Q  R  L  I  D  I  K  F  Q  E  T  C  I  N  E  T  T  S  E  V  D  G  S  L  T  L  G  I  P  L  T  G  L

1901  CGGCCAACGTTACAAAAATGAGCGTGAAGATCACTGCACTCAACTTCAAGGCCAACGAGACTACTAGAATGAAGGTATCATCCGATAAATTTGATGTAAC  2000
      P  A  N  V  T  K  M  S  V  K  I  T  A  L  N  F  K  A  N  E  T  T  R  M  K  V  S  S  D  K  F  D  V  T

2001  TCTCACTCATTTGAGCAACGACACCATCGGAAGAGACGCCCATCGACTCACGATCCGTGAAAAGCTGATTGAGGAAAGAAAAGTAAAATTCGATTGTGAG  2100
        L  T  H  L  S  N  D  T  I  G  R  D  A  H  R  L  T  I  R  E  K  L  I  E  E  R  K  V  K  F  D  C  E

2101  GGTGCTTCAAACTTCACAGTCTACTTCCAGGGTAAATTGGACGAGACCGTGGACTTGAATTACGTCGTTTCCTCCGGCGGATCCTTGATTTCCTCAGGAA  2200
       G  A  S  N  F  T  V  Y  F  Q  G  K  L  D  E  T  V  D  L  N  Y  V  V  S  S  G  G  S  L  I  S  S  G

2201  ACCATCAAGTCTTGGTTGACTCCAACGACACGGCGCAATACTTAGATGGACTCAAGCAATATGATGAACTAGGCTACCTCAATTCCTCTGTTCAGTCCAG  2300
      N  H  Q  V  L  V  D  S  N  D  T  A  Q  Y  L  D  G  L  K  Q  Y  D  E  L  G  Y  L  N  S  S  V  Q  S  S

2301  TGATGACCATGTTTTGAAATCTTATGTCATTTCCTTGACTCGCCCATTCCCAATGGAAGGAAAAGTGACCAAGAAAATTAAACTCCTAGTTTATACCCGA  2400
        D  D  H  V  L  K  S  Y  V  I  S  L  T  R  P  F  P  M  E  G  K  V  T  K  K  I  K  L  L  V  Y  T  R

2401  GACATTGAAACTGGAAAGACACTCAGTCATGTTCGAGAATATGATTCGGAGTCCTGTGCTCCCAAGCCCAAACTGGAATGGAGCCCGGCTGAAACTAATC  2500
       D  I  E  T  G  K  T  L  S  H  V  R  E  Y  D  S  E  S  C  A  P  K  P  K  L  E  W  S  P  A  E  T  N

2501  CTGGACATTCCGTTTCTTTGAAAATAAAATACAAGCCAAAGTCCCTCTGCGCTTACAGCGTCATTGATAAGTCCGCTGATTTGATCGAAAACCCAAACAA  2600
      P  G  H  S  V  S  L  K  I  K  Y  K  P  K  S  L  C  A  Y  S  V  I  D  K  S  A  D  L  I  E  N  P  N  K

2601  AATAACGACCGACTCGATTCAACAATTGAAGGAAAACCTGGCGAGCAAGCGCATTGTATATGATCAAGTCTTTCCCCCTGAATGCAAAGACGCCATGCAT  2700
        I  T  T  D  S  I  Q  Q  L  K  E  N  L  A  S  K  R  I  V  Y  D  Q  V  F  P  P  E  C  K  D  A  M  H

2701  TTGTTCAAAGCCATTCAAAGTATTGGCCTCTACATCATGAGTGATAAGTTGATTCAAGACCCATCATGCAACACAGTCGTGGACTCAATCAACTTTGATC  2800
       L  F  K  A  I  Q  S  I  G  L  Y  I  M  S  D  K  L  I  Q  D  P  S  C  N  T  V  V  D  S  I  N  F  D
                                                                                 Bait region
2801  CCAATGAATCCGACTACTCATCCAAACCCGTGCCCGTGGCCTTTGTCAGTTCCCTCGGAGCAGCGTCTCCCGCAATCGAAATGGACTCTGTTAATAGGGT  2900
      P  N  E  S  D  Y  S  S  K  P  V  P  V  A  F  V  S  S  L  G  A  A  S  P  A  I  E  M  D  S  V  N  R  V

2901  GGAAGACGACTTCGACGGCTTGAGTCAACATAAGCAGGGACAAGTTCTCGTGTCGCCTCTTCAAGCTCCATCTGTGGAAATCCGAGACTACTTCCCCGAA  3000
        E  D  D  F  D  G  L  S  Q  H  K  Q  G  Q  V  L  V  S  P  L  Q  A  P  S  V  E  I  R  D  Y  F  P  E

3001  ACATGGCTCTTCGACTTGGTTGACCTGGAGGACAATGAGGAAGCAGTCATTGAGAAAGAGGCACCCCATACAATTACCACTTGGGTGGCCGAGGCCTTCT  3100
       T  W  L  F  D  L  V  D  L  E  D  N  E  E  A  V  I  E  K  E  A  P  H  T  I  T  T  W  V  A  E  A  F

3101  GCTCTAATTTGGAAACCGGCCTCTCCGTTGCCGAAGTTGCTAACCTGAAAGCCTCTCAGGATTTCTTTGCTGACTTGAATGCTCCCTACGCTGCAAAACG  3200
      C  S  N  L  E  T  G  L  S  V  A  E  V  A  N  L  K  A  S  Q  D  F  F  A  D  L  N  A  P  Y  A  A  K  R

3201  AGGGGAAGTTCTCCAAGTTAACGCCACCGTTTTTAATAAAGTGGAAGGTCCCCTTCCAATGAAGATAAGCATTGAGTCCTCTGCACACTATAAGGTTGTG  3300
        G  E  V  L  Q  V  N  A  T  V  F  N  K  V  E  G  P  L  P  M  K  I  S  I  E  S  S  A  H  Y  K  V  V

3301  AATAAATCAGAAGAAATCGTTTGTGTCAACCCAGGCGGAAATATCCCTGTGGACTTTTATGTGGAGATGAATACGTTAGACGTCGTAAATGTAACGGTGA  3400
       N  K  S  E  E  I  V  C  V  N  P  G  G  N  I  P  V  D  F  Y  V  E  M  N  T  L  D  V  V  N  V  T  V

3401  AGGCTGAAATTATCAAGAACGAGGACTGCGGTCTTGTTGCCGAGAATTCCATTGGTTTTACTGATGTTCTTAAAAAAGCTATCCATGTTCGTCCAGAGGG  3500
      K  A  E  I  I  K  N  E  D  C  G  L  V  A  E  N  S  I  G  F  T  D  V  L  K  K  A  I  H  V  R  P  E  G

3501  ATTTCCCAAGGAAGAGGTGAAGTCTTACTTTTTATGTGGAGATCAAAAGGATACCAAGCTTGAAGATATGATCTTACCCGAGAGCAATCTCGTTGAAGAT  3600
        F  P  K  E  E  V  K  S  Y  F  L  C  G  D  Q  K  D  T  K  L  E  D  M  I  L  P  E  S  N  L  V  E  D
```

Supplementary Fig. S3 (continued)

```
3601  TCTGTTCGAGCATGGTTTGGAATCTCAGGGGACATCATGGCGCCTGCAGTTAAGAACTTGGACAATTTAGTCGCCCTTCCAACAGGATGCGGGGAGCAAA  3700
       S  V  R  A  W  F  G  I  S  G  D  I  M  A  P  A  V  K  N  L  D  N  L  V  A  L  P  T  G  C  G  E  Q
                                                                                              Thioester site
3701  CCATGATTCGAATGGTTCCCAATATCTACTTGTTGGACTACCTCAAGTCAATTGGAAAGTCCTTGCCTCAATTGGAAGAGAAGGCCCGACGGTACATTCA  3800
      T  M  I  R  M  V  P  N  I  Y  L  L  D  Y  L  K  S  I  G  K  S  L  P  Q  L  E  E  K  A  R  R  Y  I  Q

3801  AACTGGATTCGATCGTCAAAACAGAAATTTCCGTCATACAGATGGAGCTTACTCAATTTGGGGCCCTAAGGACAGCGAAGAGGATGGGTCCATGTGGCTC  3900
        T  G  F  D  R  Q  N  R  N  F  R  H  T  D  G  A  Y  S  I  W  G  P  K  D  S  E  E  D  G  S  M  W  L

3901  ACTGCCTACGTTGTCAAGGCCTTCTCACAGGCCTCCAAGTACATTGAAGTCGATAAAAACCTCCTACAGAAATCCGTCAAGTGGATTGTCAAGCGACAGG  4000
       T  A  Y  V  V  K  A  F  S  Q  A  S  K  Y  I  E  V  D  K  N  L  L  Q  K  S  V  K  W  I  V  K  R  Q

4001  ACTCGGAAACGGGATGCTTCCACAACGAGGGCTACGCCTACTCCATCAACTCTCCAAGAGAAACCCTCACTTCTCACGTGCTCGTAACTTTATTCGAAGC  4100
      D  S  E  T  G  C  F  H  N  E  G  Y  A  Y  S  I  N  S  P  R  E  T  L  T  S  H  V  L  V  T  L  F  E  A

4101  CAAATATACGGCAGACTTGGGTGATGCCGTGGACTCCAAGGTGATTCACAAGGCTCTTCGATGCTTAAAAGTATCGCAAAAACGGAAAGCGGATGATGAG  4200
        K  Y  T  A  D  L  G  D  A  V  D  S  K  V  I  H  K  A  L  R  C  L  K  V  S  Q  K  R  K  A  D  D  E

4201  GAAAGTGAGGAGTCTGTTGAGGCTGAAGAATCCGAAGAGAAGGATGAGCAAAAGGACCCCAGCGAATCCCTTTACGCCCTTTCCTTAAAAACATACGCCG  4300
       E  S  E  E  S  V  E  A  E  E  S  E  E  K  D  E  Q  K  D  P  S  E  S  L  Y  A  L  S  L  K  T  Y  A

4301  TTCATCTTTTGGAGACGAAGGCAGACTTGAAGGAGGAAATTGCTGAAGCCTTTGAGGAGATCAAGGGTGAGGAGCTCTTCAAGGAACTGATGGAGCGATC  4400
      V  H  L  L  E  T  K  A  D  L  K  E  E  I  A  E  A  F  E  E  I  K  G  E  E  L  F  K  E  L  M  E  R  S

4401  TTCTCGAGACGACAAGGGATTCCTCAACTGGAATAGCAATAGCTCCAAATCTCGCTCCGTGGAGATGACGGCCTACAATGTGATGACGCTTCTCTTTAAC  4500
        S  R  D  D  K  G  F  L  N  W  N  S  N  S  S  K  S  R  S  V  E  M  T  A  Y  N  V  M  T  L  L  F  N

4501  GACAAACGAGTGGATGCCCTAAGTGCAATCCGATGGATATCGGGCTTCAGGAATCAGATGGGAGGATTCATCTCCACACAGGATACAGTAGTGGCTCTTC  4600
       D  K  R  V  D  A  L  S  A  I  R  W  I  S  G  F  R  N  Q  M  G  G  F  I  S  T  Q  D  T  V  V  A  L

4601  AGGCCATCAGCAGCTACTCTACATTGGTATTTGAGAACAATACGAGTCTGGATGTTCAATTCTTCAACGAAACATCGTTGGTTGAGGACTTTTCCATCGA  4700
      Q  A  I  S  S  Y  S  T  L  V  F  E  N  N  T  S  L  D  V  Q  F  F  N  E  T  S  L  V  E  D  F  S  I  D

4701  CGAGGACAACAAGCTGCTCTTCAAGAGAATCAAAATCGACTCTCTCCGTGATTTAAAAGTCAGTAGCACGGGCAAGGGATGCTATACCATTTCGACTATG  4800
        E  D  N  K  L  L  F  K  R  I  K  I  D  S  L  R  D  L  K  V  S  S  T  G  K  G  C  Y  T  I  S  T  M

4801  GTTCGATACAACATCAACAATGAGACGGAAGAGAACGCCAAATTTATCATTCTGGCAGAGGGAAATAGCAGCTATCTTCATATTTGCGCCAGCTATATTG  4900
       V  R  Y  N  I  N  N  E  T  E  E  N  A  K  F  I  I  L  A  E  G  N  S  S  Y  L  H  I  C  A  S  Y  I

4901  GAGAAAAGGAAAAGACAAACATGGTTCTCTTGGAAGTGGAGCTTCTGAGTGGCTACGAAGTCGACGAAGACTCCTTGTCCAGCCTTCTGAACGAAGTGGA  5000
      G  E  K  E  K  T  N  M  V  L  L  E  V  E  L  L  S  G  Y  E  V  D  E  D  S  L  S  S  L  L  N  E  V  D

5001  CTCCGGCGTGGAGAAGTACGAGGTGAACCAGGAGGATCGGACCTTTGTGCTCTACTTCAATGGGCTGTCTAAGGGGGAGAATCACTGTTGGGACTTGGAG  5100
        S  G  V  E  K  Y  E  V  N  Q  E  D  R  T  F  V  L  Y  F  N  G  L  S  K  G  E  N  H  C  W  D  L  E

5101  CAGAAGCAAGTGAGTCAAGTTGATAACTTGAAGCCGGCTATCATCAAGATCTTCGACTACTACTCCCAGGAAGACAGCTACTCAACGACCTACATTATTT  5200
       Q  K  Q  V  S  Q  V  D  N  L  K  P  A  I  I  K  I  F  D  Y  Y  S  Q  E  D  S  Y  S  T  T  Y  I  I

5201  AATTATTCTCCACCAATGTATTCTTTGTTGTTGCTTGATTTTTTCATTACTTTTTTAATTATTCACGCTGTAGTCACTTATAATCTTTGAGACGTTCTAA  5300
      *

5301  GGAACGTTTCCCAATCCAATCATCAGCATTTTCAAGAAATTGCATTCACGTCTACGTATTTTCCCTACAAAATACATATTTTGAAAAATATTTATTTTTT  5400
5401  TAAAGCGATTATATCCACAACTCAATTCAACATATTTCCAGAGTCATGCACATTCAATGATGCTAAAAAGTTATATTTCCGAATTTACCGTCCATAGTAC  5500
5501  ATAATATGTTTATATACGTTAAATTTGCAATAAAAATAAAGTTACTTTATCTACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  5582
```

PsfuiTEP/CD109

```
   1  TGAATTAGTTTGAAGTCGACAGTGAAATGATTCATATCCAATCCCAACTTTAAAGGTGACTGCAACCTACGTCGTCGTCTTGAAAAAGGAAGGAGTCTTT  100
 101  GTATGACTACGTGTGTGTGTGTGTGAATGATGAAGATATCCAAGGTCGCATCTATATTTTTCCTGTCACTGAGTATAGCCCTCATCCAGGCCGAAAAGCC  200
                                 M  M  K  I  S  K  V  A  S  I  F  F  L  S  L  S  I  A  L  I  Q  A  E  K  P

 201  AACCTACACAATCATAGCTCCAGATGTGATCCGTCCAAACACGGATTTCCTCATTGCTGTTTCTACTCACGACATATCATCGGATCAGGATGTTCTCCTC  300
        T  Y  T  I  I  A  P  D  V  I  R  P  N  T  D  F  L  I  A  V  S  T  H  D  I  S  S  D  Q  D  V  L  L

 301  ACCATCAAGGGCCAGAGTGAGTCAGAGGGGAACGTGGAAATTAGCAAAGAGACCACGGTGAGACCGGACGAAACCCAAATCGTTCGAATGGTTATTGGTA  400
       T  I  K  G  Q  S  E  S  E  G  N  V  E  I  S  K  E  T  T  V  R  P  D  E  T  Q  I  V  R  M  V  I  G

 401  ATCTGGGCGAGGGAAAGTATGTTCTTCATGCCAGAGGAAATAATCCAATCGCCTTTGATGAAACGCAAACGTTAAAGTACATTCATCAGGGATACTCCGT  500
      N  L  G  E  G  K  Y  V  L  H  A  R  G  N  N  P  I  A  F  D  E  T  Q  T  L  K  Y  I  H  Q  G  Y  S  V

 501  TTTTGTTCAAACTGACAAGGCCATTTACCGGCCGGGAAATGTTGTCAAATTCCGGGTAATAGTGGTGACTCCAAAGCTTAAGCCAAGCGTCGTAGGCTCC  600
        F  V  Q  T  D  K  A  I  Y  R  P  G  N  V  V  K  F  R  V  I  V  V  T  P  K  L  K  P  S  V  V  G  S

 601  ATCGACGTGGGCGTTTTGGATGGGAAAGGCCATTTGGTTCGCAAGTGGGATCGCGTTTTCACAAAAGGAGGTGTTTTTGCAGAGGAATTTCTTATTGATG  700
       I  D  V  G  V  L  D  G  K  G  H  L  V  R  K  W  D  R  V  F  T  K  G  G  V  F  A  E  E  F  L  I  D

 701  ACGATCCCGTTAGAGGAGACTGGAACATAACTGTGGACGTGAGTGGACAAAGGTTCTCCAAGTCTTTCCAGGTCGTCGAGTACGTTCTTCCTCAATTCAT  800
      D  D  P  V  R  G  D  W  N  I  T  V  D  V  S  G  Q  R  F  S  K  S  F  Q  V  V  E  Y  V  L  P  Q  F  I

 801  TGTGGACATCGACATTCCAAAGTATGGTTTGTTTGATGAAGGACTCACAACGGCCAAAATAAAAGCATATTACTCTGTTGGAGTACCTGTGGAGGGCGAA  900
        V  D  I  D  I  P  K  Y  G  L  F  D  E  G  L  T  T  A  K  I  K  A  Y  Y  S  V  G  V  P  V  E  G  E

 901  GCCACCGTCTCCATATTTCCAAAGTATAAGTCTGGACTCCTTCAACCAATATTCATCCGCCCTGTTCGACAAGTTGTGCCCATTAAAGGGGAAGTTGATG  1000
       A  T  V  S  I  F  P  K  Y  K  S  G  L  L  Q  P  I  F  I  R  P  V  R  Q  V  V  P  I  K  G  E  V  D

1001  TAAGCTTCAACATTGCTAAGGAACTAAATATTCGGGGTGATTATTCTCGAGAAGTCGTTTTCGATGTTGAAATCAAGGAAAGTCGTACTGATCGAGTTCA  1100
      V  S  F  N  I  A  K  E  L  N  I  R  G  D  Y  S  R  E  V  V  F  D  V  E  I  K  E  S  R  T  D  R  V  Q
```

Supplementary Fig. S3 (continued)

```
1101  AAATAATACTGCTCTCTATCCCATGTACAGATATGACTACAAATTAGAGATGGTACGAACTGCGGATGCATACAAACCCGGAATGCCCTACACTGCCTAC  1200
        N  N  T  A  L  Y  P  M  Y  R  Y  D  Y  K  L  E  M  V  R  T  A  D  A  Y  K  P  G  M  P  Y  T  A  Y

1201  ATCAAAGTCGCAAAACAGGATGACACTCCTATTGTTGATAAAAATGTTCCCATTCAGCTAAAATGGGGCTTTTCTAATGAGCCCTCCAATGAGTACAACA  1300
       I  K  V  A  K  Q  D  D  T  P  I  V  D  K  N  V  P  I  Q  L  K  W  G  F  S  N  E  P  S  N  E  Y  N

1301  CAACCACCATCTACTCAGACAAAAATGGAATTGTGACCCTCGTTCTGGATACACCAAGCGTAAACGACACGGATGCAGTCGTGCTAGGCATTGAAGCCTC  1400
      T  T  T  I  Y  S  D  K  N  G  I  V  T  L  V  L  D  T  P  S  V  N  D  T  D  A  V  V  L  G  I  E  A  S

1401  CTATAAAAACTTGACTCAATGGTTTTCTACTATTCCCAGGGCAGAATCCCGATCTGGACTCTACTTAAGATCAAATCTAGCAACTAAGAATCCCAAAGTT  1500
        Y  K  N  L  T  Q  W  F  S  T  I  P  R  A  E  S  R  S  G  L  Y  L  R  S  N  L  A  T  K  N  P  K  V

1501  GGGGAAACCATTTCAATCGATGTCTCATCAAGTGAAAATATTGATTCTATCACCTATTTAGTGTTTGGAAGAGGAAAACTGGCACTTGGTATCACCCGAA  1600
       G  E  T  I  S  I  D  V  S  S  S  E  N  I  D  S  I  T  Y  L  V  F  G  R  G  K  L  A  L  G  I  T  R

1601  AAGCCAGTGGTACAATGGAAAATCAAATTTCATTTCGTGCTACCAGTGATATGTCCCCAGTTTGCCGAGTTATAGTGTATTACATTACGTCAATTAGCGG  1700
      K  A  S  G  T  M  E  N  Q  I  S  F  R  A  T  S  D  M  S  P  V  C  R  V  I  V  Y  Y  I  T  S  I  S  G

1701  AGAAATCATTGCCGACTCCATGGACTTTGAGGTTGAAGGAATACTCACCAACTTTGTTCAAATTGCTTCAACCAAAAAAGAAACTTTTGCTTCAAGCGAT  1800
        E  I  I  A  D  S  M  D  F  E  V  E  G  I  L  T  N  F  V  Q  I  A  S  T  K  K  E  T  F  A  S  S  D

1801  GTAACCATAAATATCAAAAGTAAACCAAACTCCTTCATTGGAATATTGGCTGTTGACAAGAGTGTTCGATCTCTCAAAGGAGGGCATGATGTGCTGTTGA  1900
       V  T  I  N  I  K  S  K  P  N  S  F  I  G  I  L  A  V  D  K  S  V  R  S  L  K  G  G  H  D  V  L  L

1901  AAGAAGTCACCGATGAGCTTCGACGCTATGATACAGCAAACACACCTGATTTCTTCCCATGGTTTAAAATCATTCAACCCAAAGAAGGCTCTCTCTCCTG  2000
      K  E  V  T  D  E  L  R  R  Y  D  T  A  N  T  P  D  F  F  P  W  F  K  I  I  Q  P  K  E  G  S  L  S  W

2001  GCATACAGGCTCTCTCAACTCTGAAGATACATTCAGCAAATCTGGAACCATCATTTTTACAAATGGAGAACTCGAGAAAGGCTATTTAGAGGAGGAGGAA  2100
        H  T  G  S  L  N  S  E  D  T  F  S  K  S  G  T  I  I  F  T  N  G  E  L  E  K  G  Y  L  E  E  E  E

2101  TCCTCAAACGTTATTGAAAATGAAGTTCTTCGAACACGAGCAAAGAACCGCCCCTTCGGACGTCCAATCCCGAGACCCGGTGTTCCAACTTTAAACCCTG  2200
       S  S  N  V  I  E  N  E  V  L  R  T  R  A  K  N  R  P  F  G  R  P  I  P  R  P  G  V  P  T  L  N  P

2201  ATAAAGGACCTGGATTGGAGTATGAATCAGCTACAAGACCTCCTTTGGAAGGACCCTACGCCTTTTCTCGTTTCCCTAGGCCGTTGGACAACATTCCAAA  2300
      D  K  G  P  G  L  E  Y  E  S  A  T  R  P  P  L  E  G  P  Y  A  F  S  R  F  P  R  P  L  D  N  I  P  K

2301  AATATACTTGAAGAACGACTTGCCCGAAACATGGCTATTCCTCAATGCAACAACGGACTCCGATGGGAGAGCCTCCATTCCCGTTAAGGCTCCTGAAACG  2400
        I  Y  L  K  N  D  L  P  E  T  W  L  F  L  N  A  T  T  D  S  D  G  R  A  S  I  P  V  K  A  P  E  T

2401  ACAAATACCACTTGGATTATATCCGGCTTTGCCCTTGATGATCTTCATGGAATGGGTATCACTGAGCAGTTCGGATCCCTGGAAGTGTTTCAGCCTTTTT  2500
       T  N  T  T  W  I  I  S  G  F  A  L  D  D  L  H  G  M  G  I  T  E  Q  F  G  S  L  E  V  F  Q  P  F

2501  ACGTCCAAGTGGATCTTCCTCCATCCATTAAATTGGGGGAAACGCTGAGTGTACAGATGGTTGTCTATAACTACTTGAAGGAAACCATCAGTGCTACGGT  2600
      Y  V  Q  V  D  L  P  P  S  I  K  L  G  E  T  L  S  V  Q  M  V  V  Y  N  Y  L  K  E  T  I  S  A  T  V

2601  TGTACTAGAGGGAACGGAGGAAGAAATTGTTTTTGGGGAAGCAGATCCTTATTCCCTGAGGGAGGACACTCAAATTGGAAGTAGAGTTTCTCAAGAAAAA  2700
        V  L  E  G  T  E  E  E  I  V  F  G  E  A  D  P  Y  S  L  R  E  D  T  Q  I  G  S  R  V  S  Q  E  K

2701  AGTATAAGTGTTCGTCCTGGACGAGGAACAATCGTTTCTTTCGTGATAACTCCTTTAGTTACGGGGGATATTCAATTGAGAATCAAGGCTGAAGGCAAAA  2800
       S  I  S  V  R  P  G  R  G  T  I  V  S  F  V  I  T  P  L  V  T  G  D  I  Q  L  R  I  K  A  E  G  K

2801  GTGGTGACAGCAACATGGGCAAAGAAGATATTTTCAAGACACTGAGTGTGAGGTCTGAAGGTGAAGTCATCAAAAGAAACAAGGGATTTCTCCTTGACTT  2900
      S  G  D  S  N  M  G  K  E  D  I  F  K  T  L  S  V  R  S  E  G  E  V  I  K  R  N  K  G  F  L  L  D  F

2901  TGACAAGAGCTCAGACTTTGACAAAAATATATCGATCGACATTCCATTTAATGCAATTCCGGGCTCTGAAAAAGTATATTTATCCCTAATGGCTGATCCA  3000
        D  K  S  S  D  F  D  K  N  I  S  I  D  I  P  F  N  A  I  P  G  S  E  K  V  Y  L  S  L  M  A  D  P

3001  TTAAGCTCAGCCATGAACCATCTGGATAAATTGATAAGCTATTCGACAGGATGTGGAGAACAAAACATGATTCGTCTTCTTCCAACGTTGGCTATTTTAG  3100
       L  S  S  A  M  N  H  L  D  K  L  I  S  Y  S  T  G  C  G  E  Q  N  M  I  R  L  L  P  T  L  A  I  L
                                                      Thioester site
3101  AGTACATTGATAAAATGAACATTCCTGCCAACCTTCAAAGGAATAATGCAATCAAAACAATGGAACTTGGCTATCAACGAGAGCTGACCATGAGACTGCA  3200
      E  Y  I  D  K  M  N  I  P  A  N  L  Q  R  N  N  A  I  K  T  M  E  L  G  Y  Q  R  E  L  T  M  R  L  Q

3201  AGATGGTTCTTTCTCCTTCTTTGGACAAGAAATGGAAGAAAAAGGTTCAACTTGGGTGACAAGTTTGGTCATTGGAAACTTTATAAAGGCCTCAAAATAC  3300
        D  G  S  F  S  F  F  G  Q  E  M  E  E  K  G  S  T  W  V  T  S  L  V  I  G  N  F  I  K  A  S  K  Y

3301  ATTGACATAGATCCTGACGTGATTCAATTGGGACTTGAATGGCTTGCAAATGTTCAAAATGATGAGGGTGCCTTCTCTGAAGGTGGAAACATATATTTTA  3400
       I  D  I  D  P  D  V  I  Q  L  G  L  E  W  L  A  N  V  Q  N  D  E  G  A  F  S  E  G  G  N  I  Y  F

3401  AAGATCTACAAGAGAGCAACATTGGACTTACTGCGTTTGTTGTGAGTGCACTTGCTTCCGAAAAAGATAAACTAGATGCTAATTCGAAGAATGCACTGAA  3500
      K  D  L  Q  E  S  N  I  G  L  T  A  F  V  V  S  A  L  A  S  E  K  D  K  L  D  A  N  S  K  N  A  L  N

3501  CAGAGGAATTTCGTACCTTGCTAAGCATTGGAAGGATGAAAGCAGCAAGGCGGAGGAGAACCCTTTTATTCTGAGCATAATCACCTATGCTCTATCAAAA  3600
        R  G  I  S  Y  L  A  K  H  W  K  D  E  S  S  K  A  E  E  N  P  F  I  L  S  I  I  T  Y  A  L  S  K

3601  ACGGACCATCCAGACACCGGGGACGCATACAACGTTCTCAAATCATTTGCCAAGAGAAATGAAACAATGGGATGGGAGTGGATTGAGGCCAAAGTTCCAC  3700
       T  D  H  P  D  T  G  D  A  Y  N  V  L  K  S  F  A  K  R  N  E  T  M  G  W  E  W  I  E  A  K  V  P

3701  CGGAACTTGAAGATAATCACTGGTTTAACCAAAGCAATTCCATCAACAATCAAATAACTGCCTACTATGCCCTTGCTACCCTCAAAAACTCTCCGGAGGG  3800
      P  E  L  E  D  N  H  W  F  N  Q  S  N  S  I  N  N  Q  I  T  A  Y  Y  A  L  A  T  L  K  N  S  P  E  G

3801  GGCCTATCCTGTCATGAACTGGCTTGTTGCTCAACAAAACTCCATGGGCGGCTTTGCTTCCACTCCGGACACCTATGTTGGAATTGAAGCTCTTAAGGAA  3900
        A  Y  P  V  M  N  W  L  V  A  Q  Q  N  S  M  G  G  F  A  S  T  P  D  T  Y  V  G  I  E  A  L  K  E

3901  TATGATCTAACACTCAACATCCCTCAGAGAAAGTCCAAAATCACTGTTGACTACTCCTATTTAGACAACTCTCGTTCAAGGGAAATCGATCCCGAATCCC  4000
       Y  D  L  T  L  N  I  P  Q  R  K  S  K  I  T  V  D  Y  S  Y  L  D  N  S  R  S  R  E  I  D  P  E  S

4001  TCACAATCTCCAAAAGAATAGTTCTACCCTCAGATATCAGAAATCTCACCATGAAGGCTCAAGGAGCCGGAGCGGTGGGAGTTGTGAACGTTGAGTATTC  4100
      L  T  I  S  K  R  I  V  L  P  S  D  I  R  N  L  T  M  K  A  Q  G  A  G  A  V  G  V  V  N  V  E  Y  S
```

Supplementary Fig. S3 (continued)

```
4101  CTACAACTTGAACGTCACTGCCAGTTGGCCCTCTTTTGTTTTAAATCCCCAGCTCCTGGATCCACCAAAGGCTCAATTCATTCTTAACTCCTGCGTCAAC  4200
        Y  N  L  N  V  T  A  S  W  P  S  F  V  L  N  P  Q  L  L  D  P  P  K  A  Q  F  I  L  N  S  C  V  N

4201  TTCATATTTTATAGTAACATTGAGGCCAGCAACATGGCCGTGATGGAAATTGACCTTCCAAGTGGATACACGGTAGACGCAACCTCTCTCCCCAGTCTCA  4300
       F  I  F  Y  S  N  I  E  A  S  N  M  A  V  M  E  I  D  L  P  S  G  Y  T  V  D  A  T  S  L  P  S  L

4301  AGCGCTATCAAGGCGTGACTCGAGTTGAGGCCACTCAAAGAAACACTCAAGTTATCATTTACTTCAAAAATATTGGCAAGAGCGAGGTTTGCCCCACGAT  4400
      K  R  Y  Q  G  V  T  R  V  E  A  T  Q  R  N  T  Q  V  I  I  Y  F  K  N  I  G  K  S  E  V  C  P  T  I

4401  CCTGGGCTTCCGGACATATGGAGTCGCAAATCAACGCCCGGCTCTCATCAAGGTGTATGACTACTATGATCAATCCCGTGTGGCCCGCATGTTCTATCAA  4500
        L  G  F  R  T  Y  G  V  A  N  Q  R  P  A  L  I  K  V  Y  D  Y  Y  D  Q  S  R  V  A  R  M  F  Y  Q

4501  GTAGGCCCTGATAACATATGCCATATTTGTAATGGAGAAGACTGTCCTCAAGATGGATGCCCCGAAAAAGAGGGCTACACCTTGTTTGGACCCTCATATG  4600
       V  G  P  D  N  I  C  H  I  C  N  G  E  D  C  P  Q  D  G  C  P  E  K  E  G  Y  T  L  F  G  P  S  Y

4601  ATGCCAACGTGGATCCAATTCAAAACTACCCCTCATCGGCATCTTTGATTTTCTCATTAAAATTTAATTTCCTCTTAGCAGCACTTCTTCCCGCATTTAT  4700
      D  A  N  V  D  P  I  Q  N  Y  P  S  S  A  S  L  I  F  S  L  K  F  N  F  L  L  A  A  L  L  P  A  F  I

4701  CATAATCGAGACTGCTCTGAGCCCAAGATTTTAATTTTTACAATTTCTTCTTTCTACTAAACTGTTAACTCTCGTCTCCACTTACTACTACTGTAGACTA  4800
        I  I  E  T  A  L  S  P  R  F  *

4801  ATACCAAATCCTCTCTTATCATGATTATTTAACTCAATTCACAATTGCATTCGTCTGATCTAACCTTCGATTCGTCAAAATATAATACATATATGTATAT  4900
4901  TTTTTTATCGTAATTTCCTAGTCCGATATTATATTGTGTACCTTAATAATACTATGAATAATATGTTAATTGTGACAAACGAGGGGGGAGGGGTGGGTTT  5000
5001  TTTAATTTAAAAAAATCATAATGATTATAAATTGCAAAG  5039
```

Supplementary Fig. S3 (continued)

CaspC3

```
   1 ATTATGAACGGGATGACTATGGCCCTCATTGCTACTTACTACCTCGACACGACCGACCAGTGGGACAAAGTGGGAGTTGAGCGCAGACATGAAGCAGTTG 100
      I  M  N  G  M  T  M  A  L  I  A  T  Y  Y  L  D  T  T  D  Q  W  D  K  V  G  V  E  R  R  H  E  A  V

 101 ATCATATAAAAACAGGCTACATAAAACAGCTTGCTTTTCAGCAGGAATGGGATGGTTCATTTATTACATTTAAGCGTCCTTTACGGCAGACAGGTAGCAC 200
     D  H  I  K  T  G  Y  I  K  Q  L  A  F  Q  Q  E  W  D  G  S  F  I  T  F  K  R  P  L  R  Q  T  G  S  T

 201 A  201
```

CaspA2M

```
   1 AATAGCGAGATGGTGTTACAGGGGGATAGACACCTTTCCCTAACTGGGGGACTAGGTCTTCTGCTACTACTATCAGTGCTTCTATTACAAATATTCGGGG  100
               M  V  L  Q  G  D  R  H  L  S  L  T  G  G  L  G  L  L  L  L  L  S  V  L  L  L  Q  I  F  G

 101 CCCACGGTGCTCCGCAACAGGATACATCACCACCCCAGCCTCCATCCCCGCCTCCTCTACCCAAACACAAAGTTGTCCGTGGTGAGAATCATACGGATTT  200
     A  H  G  A  P  Q  Q  D  T  S  P  P  Q  P  P  S  P  P  P  L  P  K  H  K  V  V  R  G  E  N  H  T  D  L
                                                                                       CUB domain
 201 GGAAGGAGACTCAGGAATATTCAAATCACCAAACTTCCCCAGTTTTTACCCACAGAACTCCAATGTGACGTGGAAGATAACAGTTCCTAAGGATAAGGAT  300
       E  G  D  S  G  I  F  K  S  P  N  F  P  S  F  Y  P  Q  N  S  N  V  T  W  K  I  T  V  P  K  D  K  D

 301 GTGGAGCTGACGTTTCATCTTTTGAATTTGGAAAAGAACCAGTATTGTTCCTGGGACTGGATTAAAGTCTTCGATGCCGAGACCCACAAAGAACTTCGTC  400
      V  E  L  T  F  H  L  L  N  L  E  K  N  Q  Y  C  S  W  D  W  I  K  V  F  D  A  E  T  H  K  E  L  R

 401 AACTATGTGGCTACCTGCAGCAGGATCTCATACTACAATTTAACACAAGCAGCATCATCGTGGAGTTCCGCTCGGACTCAGTGTCCCAACGAACGGGATT  500
     Q  L  C  G  Y  L  Q  Q  D  L  I  L  Q  F  N  T  S  S  I  I  V  E  F  R  S  D  S  V  S  Q  R  T  G  F

 501 CCTTGCCTCCTGGAGGTACGCTACCCATATGGACTCAGAAAAGGAAGAAGAGAAGAAGGAGAAAGAGAAAGGCTACATCCTCACGATTCCCCAGTCCATT  600
       L  A  S  W  R  Y  A  T  H  M  D  S  E  K  E  E  E  K  K  E  K  E  K  G  Y  I  L  T  I  P  Q  S  I

 601 ACGGCCTCAGAGGAGGCTGACCCAGAGGAGGCATGCCTTGAGATCTTCAACGTGGACTCCAAGGGAAAAGTCTTCATCAACGTCTTCTCCTCCGAAAATG  700
      T  A  S  E  E  A  D  P  E  E  A  C  L  E  I  F  N  V  D  S  K  G  K  V  F  I  N  V  F  S  S  E  N

 701 TCCTGAGTAGAGACACAAACATTTATCACGTCATCGACTATGAGCCAGAGAGCTTTGAGAGGGTACAATGCTTCGATCTAACTCTCCCCTCAAACTTCTC  800
     V  L  S  R  D  T  N  I  Y  H  V  I  D  Y  E  P  E  S  F  E  R  V  Q  C  F  D  L  T  L  P  S  N  F  S

 801 AGGACAATACGCCATTCTTCAAATCAACGGCACCTTCCAAGACGAAGAAAACTACCGGATCCTCAGCTACAAGTCCGTCAAAGTCATCCGCAGTTCTCCC  900
       G  Q  Y  A  I  L  Q  I  N  G  T  F  Q  D  E  E  N  Y  R  I  L  S  Y  K  S  V  K  V  I  R  S  S  P

 901 CAGGTCCTCATTCAGACGGACAAAAAAGAATATCGCCCCAAGCAGACTGTGTTATTCCGAGTCATTAAACTGGATGAGAGTCTTAAACCCTCGAAGAGCA  1000
      Q  V  L  I  Q  T  D  K  K  E  Y  R  P  K  Q  T  V  L  F  R  V  I  K  L  D  E  S  L  K  P  S  K  S

1001 GGGTTTTGAACAGGATCTTTGTAAAGTCCCCCAGCCGTCAGATCCTCTCGCAGCACTTGAATGTGTCTCTGGGCTCCAGGGGCATTGGGCAATTTGAGTA  1100
     R  V  L  N  R  I  F  V  K  S  P  S  R  Q  I  L  S  Q  H  L  N  V  S  L  G  S  R  G  I  G  Q  F  E  Y

1101 TTTGCTTGATGAGGAACCCGAGCTTGGGAACTGGGAAATAACCGTTGAGAATGATGCTGAGAGCGAGAAAGAAGAAGCCTCCTCGGCATCTCCCGCCTTT  1200
       L  L  D  E  E  P  E  L  G  N  W  E  I  T  V  E  N  D  A  E  S  E  K  E  E  A  S  S  A  S  P  A  F

1201 GTGGTAAATGAGGCCGTTTTGCCCAAGTTTGAGGTGGTGATCAAGTCCCCCGCCGTCATTCTTGGAGAGGATGAAAAGGTTGAATTTGAGATTTGTGGAA  1300
      V  V  N  E  A  V  L  P  K  F  E  V  V  I  K  S  P  A  V  I  L  G  E  D  E  K  V  E  F  E  I  C  G

1301 TCTACACACATGGGAGTAAAGTGCGGGGTGAACTCGAGGTATTTTTCGATCACAGATATCGACAAGCATCCTACTGGAGGGCGCCAACATACTCTTCCAA  1400
     I  Y  T  H  G  S  K  V  R  G  E  L  E  V  F  F  D  H  R  Y  R  Q  A  S  Y  W  R  A  P  T  Y  S  S  N

1401 CTTTACAAAGAAAGTTCATATTAAGGATGATAACGCCTGTGGGACCGTCAGCCTCAATCGCACGGAAATATCTCTTCTCTCCAAACACAAGGAAACTCCT  1500
       F  T  K  K  V  H  I  K  D  D  N  A  C  G  T  V  S  L  N  R  T  E  I  S  L  L  S  K  H  K  E  T  P

1501 TTCACCATTAAGGCCACACTAACCGAGGAAGGGACTGCGATCCAGCAAGAGTCCAGGTCTGAGGGAAAAGTCACTTTCACTAAAGCGGAAATTGATTTCG  1600
      F  T  I  K  A  T  L  T  E  E  G  T  A  I  Q  Q  E  S  R  S  E  G  K  V  T  F  T  K  A  E  I  D  F

1601 GAGACAGTGCTACAGAGCATATCCTTGGGGAATATCCCTACGTCGTGTTCCTAAGACTTGTTGAGCATGGAAGTCAGAAGCCCCTAAAGGGAGAAAAGGT  1700
     G  D  S  A  T  E  H  I  L  G  E  Y  P  Y  V  V  F  L  R  L  V  E  H  G  S  Q  K  P  L  K  G  E  K  V

1701 AGAGGTCTGCTCTCGCTTATGGCGAGACCTCAACAACTTCCAAAATCTAATTTCCTCAAGCCAGTTCTATTCCTTTGATGAAGATCAGTATTATGACTTG  1800
       E  V  C  S  R  L  W  R  D  L  N  N  F  Q  N  L  I  S  S  S  Q  F  Y  S  F  D  E  D  Q  Y  Y  D  L

1801 GGTAAAAAACTGCTTAACATTAAGTTCAAGGATACATGTCTTAACCTGACGACTTCCGTGGATGACGGATCTCTCACATTGGGAATCCCCCTGAGCGGAG  1900
      G  K  K  L  L  N  I  K  F  K  D  T  C  L  N  L  T  T  S  V  D  D  G  S  L  T  L  G  I  P  L  S  G

1901 TTCCACTCAATGTAACCAAGCTGGGCATTAAGGTAACGGCTCTGAGCTTCATGGCGAATGAAACAACTCGCATGGAGATGACCTCGGGCAAGTTTGATGT  2000
     V  P  L  N  V  T  K  L  G  I  K  V  T  A  L  S  F  M  A  N  E  T  T  R  M  E  M  T  S  G  K  F  D  V

2001 GAGTCTTACACATTTAGACAAGGCAGAGGGAGCTGACTCCCATCGGCTCACCATTCGTGAAAAACTCGTCGAGGAAAGGAAGATTAAGTTTGACTGCGAG  2100
       S  L  T  H  L  D  K  A  E  G  A  D  S  H  R  L  T  I  R  E  K  L  V  E  E  R  K  I  K  F  D  C  E

2101 GGGGCATCAAATTTCACAGTTTACTTCCAGGGTAAAATGGATGAAACTGTTGATTTGAACTATATTGTCTCTTCTGGAGGATCCTTGATTGCCTCTGGGA  2200
      G  A  S  N  F  T  V  Y  F  Q  G  K  M  D  E  T  V  D  L  N  Y  I  V  S  S  G  G  S  L  I  A  S  G

2201 ATCATAGAGTGCTCGTTGACTCCAATGATACCATGAAATACTTGGATGGCCTTTCTCTCTACCAAGAGTTGGGACACGTCAATTCTTCATCAGAGGTCCC  2300
     N  H  R  V  L  V  D  S  N  D  T  M  K  Y  L  D  G  L  S  L  Y  Q  E  L  G  H  V  N  S  S  S  E  V  P

2301 CTCTGGCAAAAATGTTCTCAAATCTCACGTCATTTCCCTCAGCCGTCCTTTCCCATTGGAGGGAAAGGTTACGAAAATTATTAAACTTCTCATTTATACT  2400
       S  G  K  N  V  L  K  S  H  V  I  S  L  S  R  P  F  P  L  E  G  K  V  T  K  I  I  K  L  L  I  Y  T

2401 CGCGACATTGAGACCGGGGAGACTCTTAGCCACGTACGAGAATATGAGACAGAGTCCTGTGCTCCCAAACCCGTTCTGGAGTGGAGCCCAGAGGAAACAA  2500
      R  D  I  E  T  G  E  T  L  S  H  V  R  E  Y  E  T  E  S  C  A  P  K  P  V  L  E  W  S  P  E  E  T
```

Supplementary Fig. S4

```
2501 ATCCTGGAAAGGAAGTATCTCTCAAGATCAAATCAAAGCCCCAATCGCTTTGCGCCTATAGCGTCATCGATAAATCCGCTGATCTCATTGAAAATCCTAA  2600
     N  P  G  K  E  V  S  L  K  I  K  S  K  P  Q  S  L  C  A  Y  S  V  I  D  K  S  A  D  L  I  E  N  P  N

2601 CAAGATGACCTCGGAGCGGATTCAGCAGTTGAGAGAAGACCTATCTGCGAGGCGGATCCCCTATGATCAAGTATCATCCCCTTCCGAATGCATGGACGCC  2700
       K  M  T  S  E  R  I  Q  Q  L  R  E  D  L  S  A  R  R  I  P  Y  D  Q  V  S  S  P  S  E  C  M  D  A

2701 AAGCACTTATTCAAGGCCATGGAGAACATCGGCCTTTATATCCTCAGTGATACCCTCATTCAAGACCCATCTTGCAATACAGTCATCGACTCCGCCAACT  2800
      K  H  L  F  K  A  M  E  N  I  G  L  Y  I  L  S  D  T  L  I  Q  D  P  S  C  N  T  V  I  D  S  A  N
                                                                                            Bait region
2801 TTGACGGCCCTGAAGAGGAAGATCCCTTCAAGCCTGCTCCGGAGGCGCTCTTTGGATCCACTCAATCCGCATTTGCCCCGGGAATGAGCACCGCCAATCG  2900
     F  D  G  P  E  E  E  D  P  F  K  P  A  P  E  A  L  F  G  S  T  Q  S  A  F  A  P  G  M  S  T  A  N  R

2901 CTTAGACGCAGACTATGATAACTACAGTGAAAAGAAGGAGGGACAAGTATCTCTCTCTCAGGTACGTGGGGGATCAAGTCTTGAGGTGCGGGATTACTTC  3000
       L  D  A  D  Y  D  N  Y  S  E  K  K  E  G  Q  V  S  L  S  Q  V  R  G  G  S  S  L  E  V  R  D  Y  F

3001 CCTGAAACCTGGTTGTTTGATCTAGTAGATCTTGAGGATGGAGAAGAGGCCGTTATAAAGAGAGAAGCTCCACACACCATCACAACTTGGATTGCAGAGG  3100
      P  E  T  W  L  F  D  L  V  D  L  E  D  G  E  E  A  V  I  K  R  E  A  P  H  T  I  T  T  W  I  A  E

3101 CCTTTTGCTCCCATCCAGACTCTGGACTCTCCGTTGCAAGGCGTGCTTCTCTTAAGGCCTATCAAGACTTCTTTGCAGACTTGAACGCTCCTTACTCCGT  3200
     A  F  C  S  H  P  D  S  G  L  S  V  A  R  R  A  S  L  K  A  Y  Q  D  F  F  A  D  L  N  A  P  Y  S  V

3201 CAAACGAGGGGAGATTCTTCAAATCAACGCCACAGTATTTAATAAAGTAAAAAGCCCACTCCCTATGAAGTTGAGCGTAGAGGAATCACCCGAACAGGCC  3300
       K  R  G  E  I  L  Q  I  N  A  T  V  F  N  K  V  K  S  P  L  P  M  K  L  S  V  E  E  S  P  E  Q  A

3301 TATGAGATTATAGAAAAGACGCATGAAGTTGTATGTGTCAACGCGGGCGGAAACGTTCACGTGGACTTTTACATTCGAGCTAATGAGTTAGACGTCGTCA  3400
      Y  E  I  I  E  K  T  H  E  V  V  C  V  N  A  G  G  N  V  H  V  D  F  Y  I  R  A  N  E  L  D  V  V

3401 ACATCACAATCAAGGCCGAGATTGTGAAGGATGAGGCATGTGGCTCGGTTTCAGATGAATCCATTGGCTTTACAGATATCCTTAAAAAGGCCATTCATGT  3500
     N  I  T  I  K  A  E  I  V  K  D  E  A  C  G  S  V  S  D  E  S  I  G  F  T  D  I  L  K  K  A  I  H  V

3501 TAAGCCAGAGGGTTTCCCACAGGAAGTCACCAAGTCCTATTTCCTCTGCGGGGAACAGGTAGATACTCCTCTTGAGGATATCGTTGCCCCAGAGGAGGGG  3600
       K  P  E  G  F  P  Q  E  V  T  K  S  Y  F  L  C  G  E  Q  V  D  T  P  L  E  D  I  V  A  P  E  E  G

3601 CTCGTGAAGGACTCTGTTCGCTTCTGGGCGGGTCTCTCTGGAGACATAATGGCCCCAGCCGTCAAGAACTTGGACAGCCTTGTGGCCCTCCCGACGGGCT  3700
      L  V  K  D  S  V  R  F  W  A  G  L  S  G  D  I  M  A  P  A  V  K  N  L  D  S  L  V  A  L  P  T  G
                                                                                                  Thioester site
3701 GTGGGGAGCAAACTATGATTCGCATGGTGCCCAATATCTACCTTCTCGACTACTTAAAATCCATTGGGAAGTCTCTTCCTGAGTTGGAGACGAAGGCCAA  3800
     C  G  E  Q  T  M  I  R  M  V  P  N  I  Y  L  L  D  Y  L  K  S  I  G  K  S  L  P  E  L  E  T  K  A  K

3801 GGGATACATTCAGACTGGATTTGACCGTCAAAACCGGAAGTTCCGTCATCGTGATGGGGCATACTCCATTTGGGGACCTAAGGACTCTGAGGAAGACGGC  3900
       G  Y  I  Q  T  G  F  D  R  Q  N  R  K  F  R  H  R  D  G  A  Y  S  I  W  G  P  K  D  S  E  E  D  G

3901 TCGATGTGGCTCACTGCCTATGTCGTCAAGGCATTTTCACAGGCCTCCAAGTACATTAAGGTAGACGAAAAGCTTCTCAAGAAGTCTGTTAAATGGATCA  4000
      S  M  W  L  T  A  Y  V  V  K  A  F  S  Q  A  S  K  Y  I  K  V  D  E  K  L  L  K  K  S  V  K  W  I

4001 TCAAGCGCCAGGACTTTGAGACGGGCTGTTTCCACAATGAGGGCTATGCTTACTCCGTTAATTCCCCAAGGGAGACATTGACCTCTCATGTACTGGTCAC  4100
     I  K  R  Q  D  F  E  T  G  C  F  H  N  E  G  Y  A  Y  S  V  N  S  P  R  E  T  L  T  S  H  V  L  V  T

4101 TCTTTTTGAAGCAAAATACACTGCTGGGCTAGGAAATGCTATTGACTCCAAGGTTTTAAACAAGGCCCTCCGATGCCTAAAGAACTCTCAAAAGGAGGAA  4200
       L  F  E  A  K  Y  T  A  G  L  G  N  A  I  D  S  K  V  L  N  K  A  L  R  C  L  K  N  S  Q  K  E  E

4201 ATCAATGAGGAAGCTAACGAGGAGGATGCTAATGCAGGAGCAAAGAATAAAACCAAGTCCTTGTATGCTCTTTCCCTCGAAACCTATGCAGTTCATCTTC  4300
      I  N  E  E  A  N  E  E  D  A  N  A  G  A  K  N  K  T  K  S  L  Y  A  L  S  L  E  T  Y  A  V  H  L

4301 TGGAAACACAAGCCGAGCTCAAGAAGGAGATTGGCGACGCCTTTGATGAGATCAAGGGGGACAAGCTTTTCAAGGAGCTCTTGGAGCGATCTGAGAGAGA  4400
     L  E  T  Q  A  E  L  K  K  E  I  G  D  A  F  D  E  I  K  G  D  K  L  F  K  E  L  L  E  R  S  E  R  D

4401 CGAAAAGGGGTTCCTCCATTGGGGACCCGGCAATGATTCCTCTCTCTCAATGTCCGTTGAGATGACGGCTTATAACGTGATGAGCCTTCTCTTCAGAGAT  4500
       E  K  G  F  L  H  W  G  P  G  N  D  S  S  L  S  M  S  V  E  M  T  A  Y  N  V  M  S  L  L  F  R  D

4501 GAACTAACTGATGCTCTGAGTGCGATTCGATGGATGTCGACGCATCGGAATCAGAGAGGGGGCTTTGTGTCCACACAGGATACTGTTGTTGCCCTGCAGG  4600
      E  L  T  D  A  L  S  A  I  R  W  M  S  T  H  R  N  Q  R  G  G  F  V  S  T  Q  D  T  V  V  A  L  Q

4601 CCATTAGCCGCTATTCTTCTCTTGTGTATGAGAATGAAACCAGCCTGGATGTGAAGCTATTCAATGGAACGGGACTGTTGAAGGACTTCAAAATTGATGA  4700
     A  I  S  R  Y  S  S  L  V  Y  E  N  E  T  S  L  D  V  K  L  F  N  G  T  G  L  L  K  D  F  K  I  D  E

4701 GGACAATAAACTGCTCTTCAAAAGGATGAAAGTGGATGGTCTTGTAGATCTAAAGGTCAGCAGCGAGGGCAAGGGCTGCTATGCCCTCTCCACCATGGCC  4800
       D  N  K  L  L  F  K  R  M  K  V  D  G  L  V  D  L  K  V  S  S  E  G  K  G  C  Y  A  L  S  T  M  A

4801 CGCTACAACATCAACAACCAAACGGATGAGTCCCCCAAGTTCCAAATCCGAACGGACGGAAACAGCTCGTACCTCCACATCTGCGCCAGCTACATCGGAG  4900
      R  Y  N  I  N  N  Q  T  D  E  S  P  K  F  Q  I  R  T  D  G  N  S  S  Y  L  H  I  C  A  S  Y  I  G

4901 ACCGTGAGCGAACCAACATGGCCCTCCTGGAAGTGGAGCTTCTCAGCGGCTACGAAGTCCTCGAGGACTCCCTGGCCCTACTCCTCAACGAAGTCGAATC  5000
     D  R  E  R  T  N  M  A  L  L  E  V  E  L  L  S  G  Y  E  V  L  E  D  S  L  A  L  L  L  N  E  V  E  S

5001 CGGAGTGGAAAAATACGAAATCATGAACGGTGACGATGATAAAGACGAGGACCGAAAATTCGTTCTCTACTTCAATGCCTTGGAGCAAGGAAGAAACGCC  5100
       G  V  E  K  Y  E  I  M  N  G  D  D  D  K  D  E  D  R  K  F  V  L  Y  F  N  A  L  E  Q  G  R  N  A

5101 TGTTGGGACGTGGAGCTCAAGAGAGTCAACTTAGTCAAGGACCTCAAGCCCGCCATTATTAAAATCTACGACTACTATGCACAGGAAGACGCCTTCTCAA  5200
      C  W  D  V  E  L  K  R  V  N  L  V  K  D  L  K  P  A  I  I  K  I  Y  D  Y  Y  A  Q  E  D  A  F  S

5201 CGTCATATATTATTTAATACCCCTACATCAACACTTGTAAAAAAAGGGGATACTCATTTTTTTTAATTCTTAGTAACATCTTGTCAAATAGCGTTTTATA  5300
     T  S  Y  I  I  *

5301 AATTCATGGAGGGGAAAAAAAGATGTGTCAATCATTGGAATTAAGATAAATGTTATCAAGTATATAAAATTAATGTCATTAGACAGTGTAGCTCATATAA  5400
```

Supplementary Fig. S4 (continued)

CaspA2M-2

```
   1  GTCTCTTTGAGCATAGTGCCGTTTGTGTGTATTCCTCTACTTGGAGTCTTGGGACAAACGCCCTTTCAAGTCCCTCCTTCCGAACCAAAAGTTATAAGAG  100
       V  S  L  S  I  V  P  F  V  C  I  P  L  L  G  V  L  G  Q  T  P  F  Q  V  P  P  S  E  P  K  V  I  R

 101  GTAAAAATCATACAGATTTGGAAGGGTACGACGGGATTTTCAAGTCACCCAACTTTCCCAATGACTATCCCCAAAACTCAAACGTCACATGGAAGATAAC  200
      G  K  N  H  T  D  L  E  G  Y  D  G  I  F  K  S  P  N  F  P  N  D  Y  P  Q  N  S  N  V  T  W  K  I  T
                                                   CUB domain
 201  CGTTCCGGAAGGGCAATTTGTTCAATTAACATTCCACATTTTGTTTCTTGAAAACAATCAGTTTTGCTCATGGGATTGGATCAAAATCTATGATGCCTCG  300
        V  P  E  G  Q  F  V  Q  L  T  F  H  I  L  F  L  E  N  N  Q  F  C  S  W  D  W  I  K  I  Y  D  A  S

 301  AATTATACAGAGCTTCGCCAACTTTGTGGTTTTATTCGACAAGATCTTATTTTAACATTTAATACAAGTAGCATCGTTATAGAGTTTCATTCTGATTCCG  400
       N  Y  T  E  L  R  Q  L  C  G  F  I  R  Q  D  L  I  L  T  F  N  T  S  S  I  V  I  E  F  H  S  D  S

 401  TGTCTCAAAGACGTGGTTTCCTTGTCTCTTGGAAACCCGTTTCTCCAGAAAATAAGGATGGAGAGGTTGAAAACAGACGTGAAGGTTACATGCTTACCTT  500
      V  S  Q  R  R  G  F  L  V  S  W  K  P  V  S  P  E  N  K  D  G  E  V  E  N  R  R  E  G  Y  M  L  T  L

 501  GCCTCAGTCATTTACTGCATCTGATGAAACTACCCCTGAGGAGATGTGCCTGGAGATATTCAATGTAAGGTCTAAGGGCAAAGTTTTCATAAATGTATTT  600
        P  Q  S  F  T  A  S  D  E  T  T  P  E  E  M  C  L  E  I  F  N  V  R  S  K  G  K  V  F  I  N  V  F

 601  ACTTCTCAAAGTGTAATTAGTGAAGACACAGACATTTATAAAGAAATCAACTATGATCCCGAAGTCAACGGGGACAAGCTTCAATGCTTTAACATGAAAA  700
       T  S  Q  S  V  I  S  E  D  T  D  I  Y  K  E  I  N  Y  D  P  E  V  N  G  D  K  L  Q  C  F  N  M  K

 701  TACCAACAGATTTTAGAGAAAAACAAGCTATTGTCCAAATTAATGGAACTTTTGAAGATTCAGATTATCGTATTTTAAGTTATAAGGCTGTAAAAGTTTT  800
      I  P  T  D  F  R  E  K  Q  A  I  V  Q  I  N  G  T  F  E  D  S  D  Y  R  I  L  S  Y  K  A  V  K  V  F

 801  TAAAGTCTCTGTGCAAAATCTCATTCAAACGGATAAGCATGAGTATCGTCCTAAACAATTAGTCAAATTTAGAGTGCTCTCCATGAATGGAAATTTAAAA  900
        K  V  S  V  Q  N  L  I  Q  T  D  K  H  E  Y  R  P  K  Q  L  V  K  F  R  V  L  S  M  N  G  N  L  K

 901  CCATCAAAAAATAAAATGTTGGATGAAGTGTATGTCAAGTCACCAAGTCGACAAGTTCTTGCTCAATACAAAAATGTAGAATTAGATCCTCGAGGTATTC  1000
       P  S  K  N  K  M  L  D  E  V  Y  V  K  S  P  S  R  Q  V  L  A  Q  Y  K  N  V  E  L  D  P  R  G  I

1001  ATCATTTTGAGTATCTTCTAGATGAAGAGCCAGAATTAGGATCCTGGGAAATTTCTGTAATAACAAATGGTGCTGATGATCTTGACTCATCAGTCGAATT  1100
      H  H  F  E  Y  L  L  D  E  E  P  E  L  G  S  W  E  I  S  V  I  T  N  G  A  D  D  L  D  S  S  V  E  L

1101  AGCATCATTCATTGTCAATGAGGCTGTTCTTCCTAAATTTGAAGTTCAAATAAATAGTCCTAAAGTTATTTTCAGTGAAAGTGAAAAAGTAAATTTTGAG  1200
        A  S  F  I  V  N  E  A  V  L  P  K  F  E  V  Q  I  N  S  P  K  V  I  F  S  E  S  E  K  V  N  F  E

1201  ATTTGTGGTATTTATACTCATGGAAGTAAAGTTAAAGGATCCGTCGAAGTCTTTTTTGAACATAAGTACAAAACGGCAAACTATTGGAGAGCACCAACTA  1300
       I  C  G  I  Y  T  H  G  S  K  V  K  G  S  V  E  V  F  F  E  H  K  Y  K  T  A  N  Y  W  R  A  P  T

1301  TTACATCAAACTTTACGAAGAAATCCATTATTTCACAAGATAATGCCTGTACGACTGTATCTATTAATAATACTGAGCTTGCGAAACTTTCCTTACAAAT  1400
      I  T  S  N  F  T  K  K  S  I  I  S  Q  D  N  A  C  T  T  V  S  I  N  N  T  E  L  A  K  L  S  L  Q  M

1401  GGAAACTCCATTCCTTGTCAAAGCAAAGTTAATTGAAGAGGGCACTGATATTGTTCAAGAGGCTTCTAAACAAGAAAAAGTGATTTTTACTCATGCTGAA  1500
        E  T  P  F  L  V  K  A  K  L  I  E  E  G  T  D  I  V  Q  E  A  S  K  Q  E  K  V  I  F  T  H  A  E

1501  TTCAATTTTGGGGATAGCTCTTCAGAGCATATCATTGGAGAATTTCCTTTCGTTATGTATTTAAAGCTTCAGGAGCATGTTACTAAAAAACCGCTAAAAG  1600
       F  N  F  G  D  S  S  S  E  H  I  I  G  E  F  P  F  V  M  Y  L  K  L  Q  E  H  V  T  K  K  P  L  K

1601  GAGCCAATGTTGAAATATGTTCTAGATTATGGAGGGACGTCAAAGATTTCACTCAGCACATCAATTCAAGACAATTCTATTCATTCGATGAAGATGATTA  1700
      G  A  N  V  E  I  C  S  R  L  W  R  D  V  K  D  F  T  Q  H  I  N  S  R  Q  F  Y  S  F  D  E  D  D  Y

1701  CTTTGAATTGGGTAAAAAATTGATTAATATTAAATTTAAAGAAACTTGCATCAACCAAACAACATCTGAAGAAGATGGATCAATAACCTTTGGCATTCCG  1800
        F  E  L  G  K  K  L  I  N  I  K  F  K  E  T  C  I  N  Q  T  T  S  E  E  D  G  S  I  T  F  G  I  P

1801  TTGAATGGCGTCCCAATCAATGTGACGAAATTGAGCGTTAAAGTAACTGCTTTGGATTTTGAGGCGAACCAAACAACAAGAATGAAGCTAACTCAAAATA  1900
       L  N  G  V  P  I  N  V  T  K  L  S  V  K  V  T  A  L  D  F  E  A  N  Q  T  T  R  M  K  L  T  Q  N

1901  AACACGATGTAATTTTGACTCATCTTAGTAATGACACGATTGGAAAAGATGCTCATCGTCTTACAATTCGTGAAAAGCTGATCGAAGATAAAAAGGTTAA  2000
      K  H  D  V  I  L  T  H  L  S  N  D  T  I  G  K  D  A  H  R  L  T  I  R  E  K  L  I  E  D  K  K  V  K

2001  ATTTGACTGTGAGGGTACATCAGAATTCACAGTTTATTTCCAAGGCCAAATGGGAAGTTCGGTTGACTTAAATTATGTTGTGTCATCTGGTGGTTCTTTG  2100
        F  D  C  E  G  T  S  E  F  T  V  Y  F  Q  G  Q  M  G  S  S  V  D  L  N  Y  V  V  S  S  G  G  S  L

2101  ATTACTTCTGGAAACAAGTTGGTCCTAGTCGATTCTAATGATACGTCCAAGTATCTAGATGGACTTGTTCAAATGGATGGACTTGGCTATGTGAATTCTT  2200
       I  T  S  G  N  K  L  V  L  V  D  S  N  D  T  S  K  Y  L  D  G  L  V  Q  M  D  G  L  G  Y  V  N  S

2201  CAATAGATGCAAGCAATAATGTACTCAAGTCTCATGTAATTTCTTTGAATAGACCATTTCCCAATGAAGGAAAAGTAACAAAAATGATTAATTTGCTTGT  2300
      S  I  D  A  S  N  N  V  L  K  S  H  V  I  S  L  N  R  P  F  P  N  E  G  K  V  T  K  M  I  N  L  L  V

2301  TTATACAAGAGATCTTGAAAGTGGGAAAACTCTTAGCTTTGTTAGAAAATACGAATCCGAATCATGTGCCGAAAAACCTAAAATTGAATGGAATCGTCAT  2400
        Y  T  R  D  L  E  S  G  K  T  L  S  F  V  R  K  Y  E  S  E  S  C  A  E  K  P  K  I  E  W  N  R  H

2401  GAGTCTAATCCCGGTCACAAGGTTTCTCTTAATATCAAATACCAACCAAAATCTCTTTGCGCATATAGCGTTATTGATAAATCTGCAGATTTAATAGAAA  2500
       E  S  N  P  G  H  K  V  S  L  N  I  K  Y  Q  P  K  S  L  C  A  Y  S  V  I  D  K  S  A  D  L  I  E

2501  ACCCAAATAAGATAACTTCTGTAAAAATACAAGAAGTGAGAGAGAAATTGGCTGAAAAACGTATTGTAAGAGACGGCGTTTCATCTGATAAATGTAAAAA  2600
      N  P  N  K  I  T  S  V  K  I  Q  E  V  R  E  K  L  A  E  K  R  I  V  R  D  G  V  S  S  D  K  C  K  N

2601  TGCTGAGCATTTGTTTCAGTCAATGGAGACCCTCGGGTTATTTGTTTTGAGCGACAAGTTGATTCAGGATCCTGCATGTAATAGTGTAGTTGATTCCAAT  2700
        A  E  H  L  F  Q  S  M  E  T  L  G  L  F  V  L  S  D  K  L  I  Q  D  P  A  C  N  S  V  V  D  S  N
                                                                                         Bait region
2701  AATTTCTTCAATAATGAAGGCCAATATGGTCAATATAAGCCTACACCCATTGCATTTAGTTCAGCTGCAGCTCCTGCATCTGGCTTTAGCCAGGTAAACA  2800
       N  F  F  N  N  E  G  Q  Y  G  Q  Y  K  P  T  P  I  A  F  S  S  A  A  A  P  A  S  G  F  S  Q  V  N

2801  GACTTGAGGCTGATATTGAAGATCAAAGCCATCATAAGCAAGGGCAGGTTTTACTTTCTGTACAAGAGCAAATTGTAAAGATACGTGACTACTTTCCAGA  2900
      R  L  E  A  D  I  E  D  Q  S  H  H  K  Q  G  Q  V  L  L  S  V  Q  E  Q  I  V  K  I  R  D  Y  F  P  E
```

Supplementary Fig. S4 (continued)

```
2901  AACATGGCTTTTTGATATCGTTGACTTAGAAAATAAAACTGAAGTTACAATAGAAAAAGAAGCTCCTCACACAATTACAACATGGGTTGCAGACGCTTAT  3000
        T  W  L  F  D  I  V  D  L  E  N  K  T  E  V  T  I  E  K  E  A  P  H  T  I  T  T  W  V  A  D  A  Y

3001  TGTTCCAACTTAGAAAGTGGATTTTCCGTCGCAGATAGGTCAGAATTAAAAGTTACTCAAGATTTCTTTGCTGACCTAAATTCTCCATACTCTGCAAAGC  3100
       C  S  N  L  E  S  G  F  S  V  A  D  R  S  E  L  K  V  T  Q  D  F  F  A  D  L  N  S  P  Y  S  A  K

3101  GTGGAGAAATTTTGCGAGTAAATGCTACTGTGTTCAATAAAGTTGAAGGGCCTCTCCCAATGAAATTAAGTATTCAAGCATCAGAAGGTTATTCTTTAGT  3200
      R  G  E  I  L  R  V  N  A  T  V  F  N  K  V  E  G  P  L  P  M  K  L  S  I  Q  A  S  E  G  Y  S  L  V

3201  CAACAAATCCGAAGAGATAACTTGTGTAAATTATGGTGGAAATGTCCCAGTGGATTTTTTTATTCAGCTGAATAAATTGGGTCAAGTCAACTTGACAGTC  3300
        N  K  S  E  E  I  T  C  V  N  Y  G  G  N  V  P  V  D  F  F  I  Q  L  N  K  L  G  Q  V  N  L  T  V

3301  AAGGCAGAAATTGTTAAAGATGAGTCCTGTGGAAGCATTTCAGAAGAATCTGTTGGCTATACCGATACTTTAGTTAAATCGATCAATGTTCGTCCAGAAG  3400
       K  A  E  I  V  K  D  E  S  C  G  S  I  S  E  E  S  V  G  Y  T  D  T  L  V  K  S  I  N  V  R  P  E

3401  GGTTCCCTAAGGAAGTTGTTAAATCTTACTTTTTATGTGGTGAGCAGAAAGACTCGCCTCTTGAAGATACGCTCTTGCCAGATGGAAATCTTGTTGATGA  3500
      G  F  P  K  E  V  V  K  S  Y  F  L  C  G  E  Q  K  D  S  P  L  E  D  T  L  L  P  D  G  N  L  V  D  D

3501  TTCTACTCGAGCATGGTTCGGAATTTCTGGAGACATAATGGCTCCTGCGGTCAAAAACTTAGATGGACTTGTTGCTCTTCCAACTGGCTGTGGAGAACAA  3600
        S  T  R  A  W  F  G  I  S  G  D  I  M  A  P  A  V  K  N  L  D  G  L  V  A  L  P  T  G  C  G  E  Q
                                                                                           Thioester site
3601  ACAATGATTAGAATGGTTCCAAATATTTATCTCTTGGATTACTTAAAATCGATTGGAAAGTCAGTACCAACTTTAGAATCAAAAGCAAAGAGATATATAC  3700
       T  M  I  R  M  V  P  N  I  Y  L  L  D  Y  L  K  S  I  G  K  S  V  P  T  L  E  S  K  A  K  R  Y  I

3701  AGGCTGGGTATGATCGTCAAAATAGAAAATTCCGTCATAGAGATGGAGCTTACTCGATTTGGGGACCCAAAAGCAGCGAAGAAGACGGTTCTATGTGGCT  3800
      Q  A  G  Y  D  R  Q  N  R  K  F  R  H  R  D  G  A  Y  S  I  W  G  P  K  S  S  E  E  D  G  S  M  W  L

3801  CACAGCATATGTTGTAAAAGCATTTTCCCAAGCGTCTAAATATATTGATGTAGATTTGAAACTCTTAAAAAAATCAGTCAAATGGATATTAAAAAGACAA  3900
        T  A  Y  V  V  K  A  F  S  Q  A  S  K  Y  I  D  V  D  L  K  L  L  K  K  S  V  K  W  I  L  K  R  Q

3901  GATTTCGAAACTGGATGCTTCCATAATGAAGGATATGCATATTCTGTCAATTCTCCGAGAGAAACGCTTACATCCCATGTCTTAGTTACATTATTTGAAG  4000
       D  F  E  T  G  C  F  H  N  E  G  Y  A  Y  S  V  N  S  P  R  E  T  L  T  S  H  V  L  V  T  L  F  E

4001  CTAAATACACTGCAAAATTAGGAAACTCCATTGATTCCAAAGCCATTCACAAAGCTCTGAGATGTATTAAGATATCTCAAGCTCCAAAGGATTCAGCAGA  4100
      A  K  Y  T  A  K  L  G  N  S  I  D  S  K  A  I  H  K  A  L  R  C  I  K  I  S  Q  A  P  K  D  S  A  E

4101  AACTAATGAATCTCCAGACGAATCAATTGAAGAGATGGACGACGCCAACAAATCCGATCAAGGTGGTGAATCCTTGTACGCTCAATCTCTTAAGACATAC  4200
        T  N  E  S  P  D  E  S  I  E  E  M  D  D  A  N  K  S  D  Q  G  G  E  S  L  Y  A  Q  S  L  K  T  Y

4201  GCTATTCATCTCCTTGAAACCAAGCCTGAACTAAAGAAAGAAATTGGAGAAGAATTCGAAGAGATCAAAGCAAAAAGTCTCTTCAATGAATTAAAGGAGA  4300
       A  I  H  L  L  E  T  K  P  E  L  K  K  E  I  G  E  E  F  E  E  I  K  A  K  S  L  F  N  E  L  K  E

4301  AGTCTAAGAAAGATGACAAGGGATATCTATTTTGGCAAACGAACAATTCTCTGTCAAGATCAGTTGAAATGACTGCTTATAATGTCATGACTCTTCTTTT  4400
      K  S  K  K  D  D  K  G  Y  L  F  W  Q  T  N  N  S  L  S  R  S  V  E  M  T  A  Y  N  V  M  T  L  L  F

4401  TAATGGAAAGCACAATGATGCACTGAGTGCCATTCGTTGGATTTCTGGCCATAGAAATGAAAGAGGTGGATTTATTTCTACTCAAGATACAGTTGTTGCT  4500
        N  G  K  H  N  D  A  L  S  A  I  R  W  I  S  G  H  R  N  E  R  G  G  F  I  S  T  Q  D  T  V  V  A

4501  CTTCAGGCCATCAGTATGTACTCAACTATGGTGTATGAGAATGATACAAGTTTAAGTGTACAATTCTCGAACAAAACATCAGAAATAGATTCTTTTGATC  4600
       L  Q  A  I  S  M  Y  S  T  M  V  Y  E  N  D  T  S  L  S  V  Q  F  S  N  K  T  S  E  I  D  S  F  D

4601  TTAATGAGGACAACAAGCTTTTGTTTAATAGAATTAAAATTAATGATCTTCGCAACTTAAAAGTTAGCAGTACTGGCAAGGGCTGTTACACATTTTCCAC  4700
      L  N  E  D  N  K  L  L  F  N  R  I  K  I  N  D  L  R  N  L  K  V  S  S  T  G  K  G  C  Y  T  F  S  T

4701  TATGGTTCGATACAATGTGAAGGACGAGAAGGACAAAAACGCCAAGTTTCTCATTCGTGCTGACGCCAACAAAACAGCTCTTCATATTTGTTCTAGCTAC  4800
        M  V  R  Y  N  V  K  D  E  K  D  K  N  A  K  F  L  I  R  A  D  A  N  K  T  A  L  H  I  C  S  S  Y

4801  ATTGGTGAGAAATCGGTAACGAATATGGTTTTGCTAGAAGTAGAGCTTCTTAGCGGTTATGATATAGTCGAAAGTTCCTTGGAAGTACTATTAAATGAAG  4900
       I  G  E  K  S  V  T  N  M  V  L  L  E  V  E  L  L  S  G  Y  D  I  V  E  S  S  L  E  V  L  L  N  E

4901  TACAGTCAGGTGTTGAAAAGTATGAAGTCTTAAAGGACGAAAGAAAATTTGTTCTTTATTTTAATGGATTCAAGAAAGAAGAGAATCATTGTTGGAATTT  5000
      V  Q  S  G  V  E  K  Y  E  V  L  K  D  E  R  K  F  V  L  Y  F  N  G  F  K  K  E  E  N  H  C  W  N  L

5001  GGAAGTCAAAGAAGTCAGTCAAGTAGAAAATCTTAAACCCGCTCTCATTAAAATATTTGACTACTATTCTCAAGAGGATACTTTTACAACCACCTATAAT  5100
        E  V  K  E  V  S  Q  V  E  N  L  K  P  A  L  I  K  I  F  D  Y  Y  S  Q  E  D  T  F  T  T  T  Y  N

5101  ATCTAATAATTCTTTTCACTCACTAAGAAATGAACTTATTCATCATGCTGTAGTTATTTTATTATTATCATATGCAAGTGTTTTTATCTATGAAAAAAGA  5200
       I  *

5201  TTACACAAATTCTGAGATTTTATGTTCATGCTACTCGTCCTTAAAAGGACTTATTTATTTTTCCAAATACTTATTTCCAAATAATTGAAATCTTGTAATT  5300
5301  AGTCAACAATCTAGCTTCATAATATTGATCTGACTAGACAACTTTGAAATGAGTGTTTCATTATTATATGCATTTTTGTCATCAATTAAATAAATAAAGT  5400
5401  CATTCGGAT  5409
```

CaspiTEP/CD109-1

```
   1  TTACTTACTTACTTACAAGAGGCTCAGAGAAGAGTGAGTATTGGTGGTGGGGACCAAAAGACTGGCGAAAAAAAAGTAAAGTGTACGCCTCCACATTTCT  100
 101  ATGATTTGAATTAGTTTAAAGCTGAGTGTGTGAAATGAATCAAGTTGGATCATAAGTGAATGAATAGTGATAATTATAAAAAGTTTAAAAAGAGTTTTGA  200
                                                                  M  N  S  D  N  Y  K  K  F  K  K  S  F  D

 201  CAGGATGAGGGACTTCAAGGTCATCCTCTACTTATTACCAGTGTTCTACCTCTCAGTGGGTCTCTCAGAGGCCAAGAAGCCAACGTATAGCATCATAGCC  300
        R  M  R  D  F  K  V  I  L  Y  L  L  P  V  F  Y  L  S  V  G  L  S  E  A  K  K  P  T  Y  S  I  I  A

 301  CCAAATGTCATACGTCCTAATACGGACTTCCTAATTGCAGTCTCTACTCATGGCATATCCTCAGATCAGGATGTGCTTCTCACAATTAAGGGTCAGAGTG  400
       P  N  V  I  R  P  N  T  D  F  L  I  A  V  S  T  H  G  I  S  S  D  Q  D  V  L  L  T  I  K  G  Q  S

 401  AATCCGAAGGGAACGTTGAGATTAGCAAGGAGACCACAGTGAGACCGGATGAGACTCAGATCGTCCGGATGGTCGTGGGTAACCTTGGTGAGGGGAAGTA  500
      E  S  E  G  N  V  E  I  S  K  E  T  T  V  R  P  D  E  T  Q  I  V  R  M  V  V  G  N  L  G  E  G  K  Y
```

Supplementary Fig. S4 (continued)

```
 501  CGCTCTACATGCCCGTGGAAACAGTCCTCTCGCCTTTGACGAGACTCAGAAGCTAAACTACATTCATCAAGGATACTCCGTCTTTGTGCAAACGGACAAG  600
        A  L  H  A  R  G  N  S  P  L  A  F  D  E  T  Q  K  L  N  Y  I  H  Q  G  Y  S  V  F  V  Q  T  D  K

 601  GCCATCTACAGACCCGGGAACGCCGTCAAGTTTCGAGTGATTGTTGTCACGCCAAAATTAAAACCAAGTGTTTTAGGGTCCATTGACGTGGGCGTTTTGG  700
       A  I  Y  R  P  G  N  A  V  K  F  R  V  I  V  V  T  P  K  L  K  P  S  V  L  G  S  I  D  V  G  V  L

 701  ACGGAAAAGGTCATTTAGTTCGTAAATGGGACCGAGTGTTCACACAAGGAGGCGTATTTGCGGAGGAGTTCCTAATCGATGATGATCCTGTGCGAGGAGA  800
      D  G  K  G  H  L  V  R  K  W  D  R  V  F  T  Q  G  G  V  F  A  E  E  F  L  I  D  D  D  P  V  R  G  D

 801  CTGGAATATTACGGTTGACGTCAGCGGTCAAAAGTTTAGCAAGTCCTTTCAAGTCGTGGAGTACGTCCTCCCACAGTTTATTGTGGACATTGATATCCCA  900
        W  N  I  T  V  D  V  S  G  Q  K  F  S  K  S  F  Q  V  V  E  Y  V  L  P  Q  F  I  V  D  I  D  I  P

 901  AAATACGGTTTATTCGACGAAGGACTCACCACAGCCAAGATTAAGGCGCATTACTCATTCGGTGTCCCTGTAGAGGGTGAAGCCACGGTCTCCATCTTCC  1000
       K  Y  G  L  F  D  E  G  L  T  T  A  K  I  K  A  H  Y  S  F  G  V  P  V  E  G  E  A  T  V  S  I  F

1001  CAAAGTACAAGTCTGGTATCCTACAGCCTCTGTTTACCCGCCCCGTGAGGAAAGTTGTTCCCATCAAGGGCGAAGTGGACGTGAGTTTTAACATTGCCAA  1100
      P  K  Y  K  S  G  I  L  Q  P  L  F  T  R  P  V  R  K  V  V  P  I  K  G  E  V  D  V  S  F  N  I  A  K

1101  GGAGTTGAACATCCGTGATGACTACTCTCGAGAGGTCATTTTCGATGTGGAAATCAAGGAAAGCCTCACGGATCGAGTTCAAAATAACACTGCTCTCTAC  1200
        E  L  N  I  R  D  D  Y  S  R  E  V  I  F  D  V  E  I  K  E  S  L  T  D  R  V  Q  N  N  T  A  L  Y

1201  CCCATGTACAGATATGACTACAAACTAGAAATGGTTCGAACTGCTGATGCCTATAAACCCGGCATGCCTTACTCTGCCTACATCAAAGTGGCGAAACAGG  1300
       P  M  Y  R  Y  D  Y  K  L  E  M  V  R  T  A  D  A  Y  K  P  G  M  P  Y  S  A  Y  I  K  V  A  K  Q

1301  ACGACACCCCTATCCGCGATAAGAACATTCCCATTAAACTAAAATGGGGCTTCTCTAATGAACCCTCTGAGGACTACAACACGACAACCATCTACTCCGA  1400
      D  D  T  P  I  R  D  K  N  I  P  I  K  L  K  W  G  F  S  N  E  P  S  E  D  Y  N  T  T  T  I  Y  S  D

1401  CAGCAACGGCATCGTTACCCTCGCCTTGGAAACACCCACCGTAAATGATACTGAAGCCGTCGTCCTGGGTATTGAGGCTTCCTACAAGAACCTGACACAG  1500
        S  N  G  I  V  T  L  A  L  E  T  P  T  V  N  D  T  E  A  V  V  L  G  I  E  A  S  Y  K  N  L  T  Q

1501  TGGTTTTCCACGATTCCCCGCGCCGAGTCAAGATCCGGGCTTTACCTACGATCTAAACTAGTGACTACGAGCCCCAAGGTGGGTGAAACAATATCCATTG  1600
       W  F  S  T  I  P  R  A  E  S  R  S  G  L  Y  L  R  S  K  L  V  T  T  S  P  K  V  G  E  T  I  S  I

1601  ATGTGATATCCAATGAAGGCATTGACTCCTTAACGTATTTAGTTTTTGGTAGAGGCAGACTTGCTCTTGGTGTGACACGAGGTGCAAGTGGAACAAAGGA  1700
      D  V  I  S  N  E  G  I  D  S  L  T  Y  L  V  F  G  R  G  R  L  A  L  G  V  T  R  G  A  S  G  T  K  E

1701  AAATCAGGTATCCTTTCGTGCAACAAGTGACATGTCTCCCATCTGTCGAGTGATAGTGTACTATGTGACTCCAATCAGCGGAGAAATAATAGCAGATTCC  1800
        N  Q  V  S  F  R  A  T  S  D  M  S  P  I  C  R  V  I  V  Y  Y  V  T  P  I  S  G  E  I  I  A  D  S

1801  ATGGACTTTGAAGTTGAGGGTATACTTACGAACTTTGTGGAGATTTCATCAACCAAAAAAGAGACTTATGCAACGAGTGATGTGACAATTAATATTAAGA  1900
       M  D  F  E  V  E  G  I  L  T  N  F  V  E  I  S  S  T  K  K  E  T  Y  A  T  S  D  V  T  I  N  I  K

1901  GCAAGCCCAATTCTTTCATCGGAATCCTGGCTGTCGATAAAAGTGTACGTTCCCTAAAGGGAGGACATGATGTTTTACTAAAAGAAGTGACGGAAGAACT  2000
      S  K  P  N  S  F  I  G  I  L  A  V  D  K  S  V  R  S  L  K  G  G  H  D  V  L  L  K  E  V  T  E  E  L

2001  TCGGCGCTATGACACAGTCAAGACGCCAAACTTTTTCCCCTGGTTTCGTATCATAAAGCCTAAAGATGGCTCCCTCTCTTGGCATACGGGCTCATTAAAC  2100
        R  R  Y  D  T  V  K  T  P  N  F  F  P  W  F  R  I  I  K  P  K  D  G  S  L  S  W  H  T  G  S  L  N

2101  TCTGAGGACACCTTCCTTAAATCTGGGACCCTCATTTTCACAAATGGGGAGCTTCAAAAGAGCTATACCGAGGATGACTCAACAAATGCCATAGAGACTG  2200
       S  E  D  T  F  L  K  S  G  T  L  I  F  T  N  G  E  L  Q  K  S  Y  T  E  D  D  S  T  N  A  I  E  T

2201  AGGTTCTTCGAACCCGAGCAAATAATCGGCCCTTCGGACGCCCTATTCCTAGACCCGGCGTTCCAACCTTAAACCCTGATAAGGGACCAGGGTTAGAGTA  2300
      E  V  L  R  T  R  A  N  N  R  P  F  G  R  P  I  P  R  P  G  V  P  T  L  N  P  D  K  G  P  G  L  E  Y

2301  TGAGTCCGCAACAAGACCCCCCTTGGAAGGCCCTTATGCCTTCTCTCGTTTTCCAAGACCTTTGGACAATATCCCAAAAATATATTTGAAGAACGACTTA  2400
        E  S  A  T  R  P  P  L  E  G  P  Y  A  F  S  R  F  P  R  P  L  D  N  I  P  K  I  Y  L  K  N  D  L

2401  CCTGATACCTGGCTTTTCCTGAACGCAACCACAGACTCAGAGGGAAGAGCCTCCATCCCCGTTAAGGCTCCAGAAACAACCAACACGACCTGGGTCATTT  2500
       P  D  T  W  L  F  L  N  A  T  T  D  S  E  G  R  A  S  I  P  V  K  A  P  E  T  T  N  T  T  W  V  I

2501  CTGGATTCGCCCTAGATGACCTCCATGGCATGGGTATCACCCAGCAATTCGGGTCCCTGGAAGTCTTCCAACCCTTTTACGTTCAAGTTGACATGCCTCC  2600
      S  G  F  A  L  D  D  L  H  G  M  G  I  T  Q  Q  F  G  S  L  E  V  F  Q  P  F  Y  V  Q  V  D  M  P  P

2601  CTCCATTAAGCTTGGCGAAACACTCAGCGTCCAAATGGTTGTATATAACTACCTGAAGGAAGACATCAGTGCCTCAGTCGTTCTAGAGGGTCGAGAAGAA  2700
        S  I  K  L  G  E  T  L  S  V  Q  M  V  V  Y  N  Y  L  K  E  D  I  S  A  S  V  V  L  E  G  R  E  E

2701  GAATTTGTCTTTGGAGAGGCAGACCCATATTACTCAGACGAAGACTATCAGATTGGGAGCATCGTGTCTCAAGAAAAGGAGGTCGGAATCCGCCCGGGCC  2800
       E  F  V  F  G  E  A  D  P  Y  Y  S  D  E  D  Y  Q  I  G  S  I  V  S  Q  E  K  E  V  G  I  R  P  G

2801  GTGGAACCATCGTTTCCTTTTTACTCACCCCCCTGGTAACAGGTGATATAAGGATGAGGATTAAGGCTGAAGGAAAGAGTGGGGATAGCAATGTAGGCTA  2900
      R  G  T  I  V  S  F  L  L  T  P  L  V  T  G  D  I  R  M  R  I  K  A  E  G  K  S  G  D  S  N  V  G  Y

2901  TGAAAACATTGTAAAGGTACTCCGTGTCGAGTCTGAAGGGGAGGTTATGCGCCGAAACAAGGGTTATCTCCTGGACTTCAACGGAGGCTCCGACTTTGAT  3000
        E  N  I  V  K  V  L  R  V  E  S  E  G  E  V  M  R  R  N  K  G  Y  L  L  D  F  N  G  G  S  D  F  D

3001  AAAAATATATCAATTGATATTCCCTTTAATGCTGTTCCTGGATCCGAAAAAGTATATGTGTCCATGATGGCGGATCCTCTCAGCTCAGCAATGAACCATA  3100
       K  N  I  S  I  D  I  P  F  N  A  V  P  G  S  E  K  V  Y  Y  V  S  M  M  A  D  P  L  S  S  A  M  N  H

3101  TGGAGAAGCTTATACATTACCCTACGGGCTGTGGAGAACAAAATATGATTCGCCTTGTTCCGGTGTTGTCCATATTGGAATACATTGATGAAAAGAAAAT  3200
      M  E  K  L  I  H  Y  P  T  G  C  G  G  E  Q  N  M  I  R  L  V  P  V  L  S  I  L  E  Y  I  D  E  K  K  I
                                 Thioester site
3201  AACGGCCCCTCTTCAGAGAAACACTGCAATCAAAACAATGGAACTCGGTTATCAACGGGAACTCACCATGCGATTGGACGACGGATCGTTTTCCTTTTTC  3300
        T  A  P  L  Q  R  N  T  A  I  K  T  M  E  L  G  Y  Q  R  E  L  T  M  R  L  D  D  G  S  F  S  F  F

3301  GGGAAGGACATGGACGAAAAAGGGTCCACTTGGGTGACGGCCTTGGTCATTGGAACGTTTTTAAGGGCCTCAAAGTATATTGACATTGATCCCAATGTCT  3400
       G  K  D  M  D  E  K  G  S  T  W  V  T  A  L  V  I  G  T  F  L  R  A  S  K  Y  I  D  I  D  P  N  V

3401  TACAGTCCGGACTGGAATGGCTGGCAAAGGCTCAAAATGACGAAGGAGCATTCTCTGAGAGTGGAAACATTTACTTCCGTCACCTACAGGAAAGCAACAT  3500
      L  Q  S  G  L  E  W  L  A  K  A  Q  N  D  E  G  A  F  S  E  S  G  N  I  Y  F  R  H  L  Q  E  S  N  I
```

Supplementary Fig. S4 (continued)

```
3501  TGGACTCACAGCCTTTGTTGTAGGAGCCTTGGCGTCTCAAAAGGAAATGCTAGACGCCAACTCCAAGAATGCACTTAACCGTGGAATTTCCTTCTTGGCC  3600
        G  L  T  A  F  V  V  G  A  L  A  S  Q  K  E  M  L  D  A  N  S  K  N  A  L  N  R  G  I  S  F  L  A

3601  AAACATTGGAAGGATGAAAGCAGCTCAGCGGAAGAGAATCCCTATATTTTTAGCGTAATTACGCATGCTTTGGTCAAAACAGACCATCCAGAGGCCGGAG  3700
       K  H  W  K  D  E  S  S  S  A  E  E  N  P  Y  I  F  S  V  I  T  H  A  L  V  K  T  D  H  P  E  A  G

3701  ATGCCTACAACGTCCTCAAATCCTTCGCCAAGAGAAATGAGACCAACGGATGGGAGTGGATCGAGGCTAAAATCCCTGAAGCCCAACAAGAGAATCACTG  3800
      D  A  Y  N  V  L  K  S  F  A  K  R  N  E  T  N  G  W  E  W  I  E  A  K  I  P  E  A  Q  Q  E  N  H  W

3801  GTTCAACACCTCCAACTCCATAAACAATCAAATCACAGCCTACTACGCCCTTTCTACTCTCAAAATATCTCCAGATCGAATCATCCCCATCATGAACTGG  3900
        F  N  T  S  N  S  I  N  N  Q  I  T  A  Y  Y  A  L  S  T  L  K  I  S  P  D  R  I  I  P  I  M  N  W

3901  CTCATTACCCAGCAGAACTCCATGGGAGGCTTTGCCTCTACCCCAGATACCTATGCTGCAATCGAGGCCCTACACAACTACGATCTCAACCTCCCTCAGA  4000
       L  I  T  Q  Q  N  S  M  G  G  F  A  S  T  P  D  T  Y  A  A  I  E  A  L  H  N  Y  D  L  N  L  P  Q

4001  GACGATCCAAAATATCCGTGGATTATTCCTACCTGGACAACTCCCGCTCCAGAGAAATCGATCCCGAGTCCCTCACCATATTACAAAGAAGAGTACTCCC  4100
      R  R  S  K  I  S  V  D  Y  S  Y  L  D  N  S  R  S  R  E  I  D  P  E  S  L  T  I  L  Q  R  R  V  L  P

4101  TGCCTACATTCGAAATCTCACCGTGAAGGCAACGGGAGCCGACGCGGTGGGTGTTATTAATGTTGAATATGCCTACAATTTGAACGTGACGGCAAGCTAT  4200
        A  Y  I  R  N  L  T  V  K  A  T  G  A  D  A  V  G  V  I  N  V  E  Y  A  Y  N  L  N  V  T  A  S  Y

4201  CCCTCCTTCGTTCTTAACCCACAGCTCCTAGACCCTCAAAAAAAGGCCCAATTCATTCTCAACGCCTGCGTCAATTACATATTCTACAGCGACTCCGACG  4300
       P  S  F  V  L  N  P  Q  L  L  D  P  Q  K  K  A  Q  F  I  L  N  A  C  V  N  Y  I  F  Y  S  D  S  D

4301  CCAGTAATATGGCTGTCATGGAGATCGACCTCCCCAGCGGATACACCGTGGACGCCACCTCACTCCCAAGTCTCAAGCGCTACCAGGGTGTGACCCGCGT  4400
      A  S  N  M  A  V  M  E  I  D  L  P  S  G  Y  T  V  D  A  T  S  L  P  S  L  K  R  Y  Q  G  V  T  R  V

4401  AGAGGCCAGCCGTCGCAACACCCGAGTCATACTCTACTTCAAACATATCGGGAAAAGTGAGGTTTGTCCCACGATCCTTGGGTTCAGGACCTTCCGAGTC  4500
        E  A  S  R  R  N  T  R  V  I  L  Y  F  K  H  I  G  K  S  E  V  C  P  T  I  L  G  F  R  T  F  R  V

4501  GCAAATCAACGCCCAGCTCTCATCAAGGTCTACGACTACTATGACCAATCCAAGGTGGCTAGAATGTTCTATCAAGTTGGGCCAGACAATATTTGCCATA  4600
       A  N  Q  R  P  A  L  I  K  V  Y  D  Y  Y  D  Q  S  K  V  A  R  M  F  Y  Q  V  G  P  D  N  I  C  H

4601  TATGCACGGGAGAGGACTGTCCAGAGGACGGATGCCCAGAAGGAGAGTCTTACGATTTCTTCGGCCCTACATACAACGCCAACGTGGATCCCATCCAGGA  4700
      I  C  T  G  E  D  C  P  E  D  G  C  P  E  G  E  S  Y  D  F  F  G  P  T  Y  N  A  N  V  D  P  I  Q  D

4701  TTTCTCAGCGGCCGCCCAAGTCCTCCCCATCTATTCTCTTACATTTGCCATCTCCTTATTTGCCATGCTAAACTACTAGACGTCGTCATTAAAATCAAGC  4800
        F  S  A  A  A  Q  V  L  P  I  Y  S  L  T  F  A  I  S  L  F  A  M  L  N  Y  *

4801  CTCCTTTCACAAGCAGTCATTTGTTTTACAAAAATACATCCATACTAAAATGTTAATTATGTCTTGCTAATTCACTCATTAGACTAATCATATATTATTT  4900
4901  TTCTGAGCATTTCATTTTTAACTCCTGTTAATAACTACAAACATATATATATCATAACCTTTAATTCGTAAAAAAAATACATATTTAATTTACTAGTCTA  5000
5001  AATAACTTATTCTTCTTCGTTGTACCTAATTATTTTCTCTTTGTATTTGTCTAATTACTCTAGGGGGGGCATTATTATATAAGATGAAATGAGGGATTCA  5100
5101  AGGCTCATAAAGTGATATGAATATTTTGCAAGACTAATTATTCTATGTGATATTTATTTTATGTACATTCATAATATACCTATATCAAAATAAAA  5195
```

CaspiTEP/CD109-2

```
   1  CATGCTAGAGGGAATTTTCCTATAGCGTTTGATCATACTCAAAAGTTAATCCACATCAATCAGGGCTACTCTGTCTTTGTTCAAACGGATAAAGCAATAT  100
       H  A  R  G  N  F  P  I  A  F  D  H  T  Q  K  L  I  H  I  N  Q  G  Y  S  V  F  V  Q  T  D  K  A  I

 101  ACAGGCCAGGGAATGTTGTCAAATTCCGGGTTATTGTTGTTACTCCTAAACTAAAGCCAAGTGTAGTTGGATCCATTGATGTCGGAGTTTTGGATGGAAA  200
      Y  R  P  G  N  V  V  K  F  R  V  I  V  V  T  P  K  L  K  P  S  V  V  G  S  I  D  V  G  V  L  D  G  K

 201  AGGTCATTTAGTTCGTAAATGGGATAGGGTTTTCACGACGGGTGGAGTTTTTGCTGAGGAATTTCTCATAGATGAAGACCCAGTTCGAGGGGATTGGAAC  300
        G  H  L  V  R  K  W  D  R  V  F  T  T  G  G  V  F  A  E  E  F  L  I  D  E  D  P  V  R  G  D  W  N

 301  ATAACGGTGAATGTAAGTGGACAAAAGTTCACCAAATCCTTTCAAGTTGTAGAGTACGTTCTTCCACAATTTATCGTTGACATTGGGATTCCAAAGTATG  400
       I  T  V  N  V  S  G  Q  K  F  T  K  S  F  Q  V  V  E  Y  V  L  P  Q  F  I  V  D  I  G  I  P  K  Y

 401  GTTTATTTGATGAAGGACTTACATCTGCCAAAATTAAGGCATATTATCCGTTCGGAGTCCCTGTGAAAGGGGAGGCCACTGTTTCTATATTTCCAAAATA  500
      G  L  F  D  E  G  L  T  S  A  K  I  K  A  Y  Y  P  F  G  V  P  V  K  G  E  A  T  V  S  I  F  P  K  Y

 501  TAAATCTGGATTGCTTCAACCCATTTTTATTCGTCCAGTCAGAAAAGTCGTACCAATTAATGGCGAAGTAGATGTTAGTTTCAACATTGCCAAGGAATTA  600
        K  S  G  L  L  Q  P  I  F  I  R  P  V  R  K  V  V  P  I  N  G  E  V  D  V  S  F  N  I  A  K  E  L

 601  AATATAAGAGATGATTATTCTAGAGAAGTTATTTTTGATGTTGAAATCAAAGAAAGTTTGACGGATCGAGTACAAAATAATACAGCCCTATATCCAATGT  700
       N  I  R  D  D  Y  S  R  E  V  I  F  D  V  E  I  K  E  S  L  T  D  R  V  Q  N  N  T  A  L  Y  P  M

 701  ACAAATATGATTATAAACTGGAAATGGTTCGAACTGCTGATGCGTATAAGCCTGGTATGCCTTACACTGCATACATCAAAGTTGCAAAACAAGATGATAC  800
      Y  K  Y  D  Y  K  L  E  M  V  R  T  A  D  A  Y  K  P  G  M  P  Y  T  A  Y  I  K  V  A  K  Q  D  D  T

 801  ACCGATTAGTGATAAAAATGCGCAAATTAAATTAAAATGGGGATTTTCTAATGAACCTTCCGAGAAATACAATACAACAACTTTGTTGGCTAATAAAAAC  900
        P  I  S  D  K  N  A  Q  I  K  L  K  W  G  F  S  N  E  P  S  E  K  Y  N  T  T  T  L  L  A  N  K  N

 901  GGAATTGTAACTCTCATCTTGGATACTCCTGATGTGAATGATACAGACGCCGTAGTGTTAGGAATTGAAGCATCATATAAAGATTTAACTCAATGGTTTT  1000
       G  I  V  T  L  I  L  D  T  P  D  V  N  D  T  D  A  V  V  L  G  I  E  A  S  Y  K  D  L  T  Q  W  F

1001  CTACCATTCCTAGGGCTCAATCAAGATCTGGACTATACTTGAGATCCAAATTGATTACGAAAAATCCCAAAGTAGGAGAGACCATCTCAGTCGAAGTTAC  1100
      S  T  I  P  R  A  Q  S  R  S  G  L  Y  L  R  S  K  L  I  T  K  N  P  K  V  G  E  T  I  S  V  E  V  T

1101  TTCAAGTGAAAATATTGATTCTATATCATATTTAGTCTTTGGACGGGGACAACTTGTCTTAGCCATTACTCGTAAAGCCAGTGGAACAAATGAAAATAAA  1200
        S  S  E  N  I  D  S  I  S  Y  L  V  F  G  R  G  Q  L  V  L  A  I  T  R  K  A  S  G  T  N  E  N  K

1201  ATATCATTCAGAGCGACAAGTGATATGTCTCCAATTTGTCGGATTATTATATACTATGTAACTTCAATCAGTGGAGAAATTATATCAGACTCAATGGACT  1300
       I  S  F  R  A  T  S  D  M  S  P  I  C  R  I  I  I  Y  Y  V  T  S  I  S  G  E  I  I  S  D  S  M  D

1301  TTGAAGTTGATGGAATACTAACCAATTTTGTAAGCATCTCGTCAACTAAAAAGGTTACTTCAGCAGCAAACGACGTGACAATAAATATCAAGAGTAAACC  1400
      F  E  V  D  G  I  L  T  N  F  V  S  I  S  S  T  K  K  V  T  S  A  A  N  D  V  T  I  N  I  K  S  K  P
```

Supplementary Fig. S4 (continued)

```
1401  TAATTCATTTGTTGGCATTCTAGCCGTAGACAAGAGTGTGAGATCTCTCAAAGGAGGTCACGATGTTCTTTTAAAGGAAGTTACCGATGAATTAAGAAGA  1500
        N  S  F  V  G  I  L  A  V  D  K  S  V  R  S  L  K  G  G  H  D  V  L  L  K  E  V  T  D  E  L  R  R

1501  TATGACACTGCAAACACTCCGAATTTTTTCCCTTGGTTCAAAGTTATTAAACCCAATGAAGGTTCTTTGTCCTGGCACACAGGTTCTTTAAATTCCGAAG  1600
       Y  D  T  A  N  T  P  N  F  F  P  W  F  K  V  I  K  P  N  E  G  S  L  S  W  H  T  G  S  L  N  S  E

1601  ATACTTTTGTAATGTCTGGTACTATTATATTTACAAATGGAATATTAGAAAAACGTTATTCGGAGGAGTCCACAAATGTAATTGAGACAGAGGTACAGAG  1700
      D  T  F  V  M  S  G  T  I  I  F  T  N  G  I  L  E  K  R  Y  S  E  E  S  T  N  V  I  E  T  E  V  Q  R

1701  AACTTTAGCTAAAAATAGACCTTTCGGACGTCCAATTCCAAGACCTGGCGTTCCAACTTTAAATCCTGATAAGGGTCCTGGATTAGAATATGAATCAGCC  1800
        T  L  A  K  N  R  P  F  G  R  P  I  P  R  P  G  V  P  T  L  N  P  D  K  G  P  G  L  E  Y  E  S  A

1801  ACCAGACCTCCATTAGAAGGTCCTTATGCTTTTTCACGCTTCCCAAGGCCATTAGACAATACTCCAAAGATTTACTTAAAGAATGATTTGCCCGAAACTT  1900
       T  R  P  P  L  E  G  P  Y  A  F  S  R  F  P  R  P  L  D  N  T  P  K  I  Y  L  K  N  D  L  P  E  T

1901  GGCTTTTTCTTAATGCAACAACTGATTCGAATGGAAGAGCATCTATTCCAGTGAAGGCTCCCGAAAAAACAAACACAACGTGGATTATATCTGGGTTTGC  2000
      W  L  F  L  N  A  T  T  D  S  N  G  R  A  S  I  P  V  K  A  P  E  K  T  N  T  T  W  I  I  S  G  F  A

2001  TTTGGATGATCTTCATGGAATGGGTATTACTGAGCAATTTGGATCTTTGGAAGTCTTTCAACCATTTTATGTCCAAGTGGATCTTCCTCCTTCAGTTAAA  2100
        L  D  D  L  H  G  M  G  I  T  E  Q  F  G  S  L  E  V  F  Q  P  F  Y  V  Q  V  D  L  P  P  S  V  K

2101  CTAGGCGAAACTTTGAGTGTGCAAATGGTTGTGTATAATTACTTGAAGGAGAGTATTCAGGCATCAGTAACTTTAGAAGGGCGTGAAGAAGAATTTGTTT  2200
       L  G  E  T  L  S  V  Q  M  V  V  Y  N  Y  L  K  E  S  I  Q  A  S  V  T  L  E  G  R  E  E  E  F  V

2201  TTGGTGAAGCTCAGCCGGATTTATCCAATGCTTATCCTAAAAATGAACAGATAGGCAGCCTAATATCTCAAGAAAAAGAAGTGGGTATCAAGCCTGGCAG  2300
      F  G  E  A  Q  P  D  L  S  N  A  Y  P  K  N  E  Q  I  G  S  L  I  S  Q  E  K  E  V  G  I  K  P  G  R

2301  AGGAACCATTGTATCATTCCTCATTACTCCGCTCTTATCGGGGAATATCAGAATGCGAATAAAAGCAATAGGGAAAAGTGGTACCAGTAATGTAGGCAAT  2400
        G  T  I  V  S  F  L  I  T  P  L  L  S  G  N  I  R  M  R  I  K  A  I  G  K  S  G  T  S  N  V  G  N

2401  GACAATGTTGTCAAAACTTTGACCGTCACATCAGAAGGGGAAATCATGCATAGAAATAAGGGGTATCTCCTTGACATGAGGGACGCCTCAGAGTTTAAAA  2500
       D  N  V  V  K  T  L  T  V  T  S  E  G  E  I  M  H  R  N  K  G  Y  L  L  D  M  R  D  A  S  E  F  K

2501  AAAATGTGACGATCGACATTCCGTTTAATGCAGTCCCTGGATCTGAAAATGTATTCATTTCCTTGATGGCAGATCCACTCAGTTCAACAATGAATAACTT  2600
      K  N  V  T  I  D  I  P  F  N  A  V  P  G  S  E  N  V  F  I  S  L  M  A  D  P  L  S  S  T  M  N  N  L

2601  GGAGAGATTGATTCGATATCCTAGTGGATGTGGTGAACAGAATATGATAAGACTCGTTCCAACTCTTGCTATCTTAGAGTATATTGATAAACTGGGAATA  2700
        E  R  L  I  R  Y  P  S  G  C  G  E  Q  N  M  I  R  L  V  P  T  L  A  I  L  E  Y  I  D  K  L  G  I
                                Thioester site
2701  TCTGCCACTTATCAAAGAAATGAGGCTCTAAAAACTATGGAATTGGGATATCAACGAGAACTAACCATGAGAATGATGGATGGCTCCTTCTCATTTTTTG  2800
       S  A  T  Y  Q  R  N  E  A  L  K  T  M  E  L  G  Y  Q  R  E  L  T  M  R  M  M  D  G  S  F  S  F  F

2801  GAAAGGGAATTGATGAAAGTGGTTCAACCTGGGTGACCGCATTGGTGATTGGAAATTTCTTAAAAGCTTCAAAATATATTGATATTGACCCTGTTGTCAT  2900
      G  K  G  I  D  E  S  G  S  T  W  V  T  A  L  V  I  G  N  F  L  K  A  S  K  Y  I  D  I  D  P  V  V  I

2901  ACAATCTGGACTTGATTGGCTTTCAAAAATTCAAAATGAAGAAGGTGCATTCCCCGAAAAAGGCAAGACATATTTTAGTGATTTACAAGAAAATGAATTG  3000
        Q  S  G  L  D  W  L  S  K  I  Q  N  E  E  G  A  F  P  E  K  G  K  T  Y  F  S  D  L  Q  E  N  E  L

3001  GGGCTCACTGCTTTTGTTGTAAGTGCCTTGGCTTCTCAAAAAGAAGATTTAGATGCCAATTCTAAGAATGCCTTGAATAGAGGAATTTCGTTTGTTGCAA  3100
       G  L  T  A  F  V  V  S  A  L  A  S  Q  K  E  D  L  D  A  N  S  K  N  A  L  N  R  G  I  S  F  V  A

3101  AACATTGGAAGGATGTCAGTACTCAAGCTGAAGAGAATCCCTATATTATGAGTCTTGTTGCTCATGCTCTTGCTAAAACCGATCATCCTGATTCAAATGA  3200
      K  H  W  K  D  V  S  T  Q  A  E  E  N  P  Y  I  M  S  L  V  A  H  A  L  A  K  T  D  H  P  D  S  N  D

3201  TGCACTTAACATTCTTAAATCTGTTTCCAAAAGAAACGAGAGTATGGGTTGGGAATGGCTTGAAGCTAAGATACCTGAAGAAGTTGAAAATAACCATTGG  3300
        A  L  N  I  L  K  S  V  S  K  R  N  E  S  M  G  W  E  W  L  E  A  K  I  P  E  E  V  E  N  N  H  W

3301  TATAATCAGTCAAATTCATTGAATAATCAGATGACTGCGTACTACGCCTTGGCCACTTTAAAAGTCTCTCCAGAGAAGGCATCTCCTATCATGGCATGGC  3400
       Y  N  Q  S  N  S  L  N  N  Q  M  T  A  Y  Y  A  L  A  T  L  K  V  S  P  E  K  A  S  P  I  M  A  W

3401  TAGTTGCTCAAAAAAATTCTATTGGAGGGTTTGCATCAACTCCTGATACTTATATTGGTATTGAGGCCCTTAAAGAATTTGACTTAACTCTCAACATTCC  3500
      L  V  A  Q  K  N  S  I  G  G  F  A  S  T  P  D  T  Y  I  G  I  E  A  L  K  E  F  D  L  T  L  N  I  P

3501  AAAAAGCACTTCAAAAATAGACGTGAATTATTCTTATTTGGATAATTCTCGTAGTAGGGAAATTAATCCTGAAGCATCAACCATCTCAATAAGACGAGTC  3600
        K  S  T  S  K  I  D  V  N  Y  S  Y  L  D  N  S  R  S  R  E  I  N  P  E  A  S  T  I  S  I  R  R  V

3601  CTACCCTCAAACACACGCAACATAACTTTAGAAGCCAAAGGAACGAATGCTATTGGGGTCATTAACGTGGAATATGCTTATAGTTTGAATGTAACGGCAA  3700
       L  P  S  N  T  R  N  I  T  L  E  A  K  G  T  N  A  I  G  V  I  N  V  E  Y  A  Y  S  L  N  V  T  A

3701  GTTGGCCTTCTTTTGTTTTGAATCCTCAGCTCCTGCCTGCTCCACCAGCACAACTGATCCTTAACTCTTGCGTTAATTACATATTTTATAGCAACAAAGA  3800
      S  W  P  S  F  V  L  N  P  Q  L  L  P  A  P  P  A  Q  L  I  L  N  S  C  V  N  Y  I  F  Y  S  N  K  E

3801  GTCATCTAACATGGCTGTCATGGAGATTGATTTGCCTAGTGGCTACACAGTCGATGCGAATTCTCTACCGAGTTTGAAGCGTTATGAAGGAGTTTCCCGG  3900
        S  S  N  M  A  V  M  E  I  D  L  P  S  G  Y  T  V  D  A  N  S  L  P  S  L  K  R  Y  E  G  V  S  R

3901  GTTGAATCGAGTTTACGAAATACTCGAGTGATTATTTACTTTCAAAATATTCGCAAAAGTGAGGTATGCCCAACGATTTTAGGATTCAGAACCTATGGTG  4000
       V  E  S  S  L  R  N  T  R  V  I  I  Y  F  Q  N  I  R  K  S  E  V  C  P  T  I  L  G  F  R  T  Y  G

4001  TATCTAATCAGCGACCAGCGCTCATCAAGGTCTATGATTATTATGATCAATCACGAGTTGCTAGAATGTTTTATCAAGTTCGTCCAGCTACAATTTGCGA  4100
      V  S  N  Q  R  P  A  L  I  K  V  Y  D  Y  Y  D  Q  S  R  V  A  R  M  F  Y  Q  V  R  P  A  T  I  C  E

4101  AATTTGTACTGGAAATGACTGTCCAGAAGATGGTTGTGAAGACAAATCTTTTTACCCAATTTATAATCCAACTTATCATGCCAATGTGGATCCTAATAAT  4200
        I  C  T  G  N  D  C  P  E  D  G  C  E  D  K  S  F  Y  P  I  Y  N  P  T  Y  H  A  N  V  D  P  N  N

4201  GAAGCGGGGGCAACTTCAAAAATGTCAGTGTCTTTATTTTTGATGGCCTCCTCATTGTGCTGTACTTTAGTCAAGATGATCTACGTCTAGTAAGCAATTA  4300
       E  A  G  A  T  S  K  M  S  V  S  L  F  L  M  A  S  S  L  C  C  T  L  V  K  M  I  Y  V  *

4301  CTTCGACTCCATTGTTTTAACCTTTAAACATTCATTGGGAAAAATCATAATCATAATTTTCAAACAGTTAAAATACATACATGTAATAAATATACCTATA  4400
4401  TCTATCTATCTATCTATCTAAAAATATATTTAAGAGAAAATATAAGTGAATCCATC  4456
```

Supplementary Fig. S4 (continued)

PaspC3

```
   1 ACAGGAGGAGACATCCCTGTGACGGTTAAAGTAATGAATCATCCAACCAAGAACAGAGAGCTGGCAACAAAATCTGTGACCCTCACCAGTGCAAAGGACT  100
      T  G  G  D  I  P  V  T  V  K  V  M  N  H  P  T  K  N  R  E  L  A  T  K  S  V  T  L  T  S  A  K  D

 101 TCCAAGCACTTGCACAAATATCGATCCCGGCCGATGGCTTCAGCAGGGATCCCAATGTGAAGCAATATGTTTACTTACAAGCTCAGTTCCCAGACCGGCT  200
     F  Q  A  L  A  Q  I  S  I  P  A  D  G  F  S  R  D  P  N  V  K  Q  Y  V  Y  L  Q  A  Q  F  P  D  R  L

 201 GCTGGAGAAAGTCGTCTTGGTGTCCTTCCAGTCTGGGTACATCTTCATCCAGACTGACAAGACCCTGTACACCCCTGAAAGCACAGTTCTTTTCAGGATG  300
       L  E  K  V  V  L  V  S  F  Q  S  G  Y  I  F  I  Q  T  D  K  T  L  Y  T  P  E  S  T  V  L  F  R  M

 301 TTTGGAGTGACGCCTCGCATGGAACCTGTGGAGAGGGACGACAAAACTGATGCCTCTATCGCTATTGAGATTGTGACACCTGATGGCATCATTTTACCAC  400
      F  G  V  T  P  R  M  E  P  V  E  R  D  D  K  T  D  A  S  I  A  I  E  I  V  T  P  D  G  I  I  L  P

 401 TTGATCCAGTCTTCTTGAAATCGGGGCTGTACTCTGGAGATTACAAACTTGGTGACATTGTCAGTTTTGGAGTGTGGAAAGTGGTGGCAAAGTTCCAGAG  500
     L  D  P  V  F  L  K  S  G  L  Y  S  G  D  Y  K  L  G  D  I  V  S  F  G  V  W  K  V  V  A  K  F  Q  S

 501 CAACCCACAGCAGACCTTTGCTGCAGAGTTTGAGGTCAAAGAATACGTGCTGCCCAGTTTTGAAGTGAAACTGTTGCCTGTGAGTCCCTTCTACTACGTG  600
       N  P  Q  Q  T  F  A  A  E  F  E  V  K  E  Y  V  L  P  S  F  E  V  K  L  L  P  V  S  P  F  Y  Y  V

 601 GGCAGTCCGGAGCTCACCGTCAACATCAAAGCCACGTATCTGTTTGGTGAAGAGGTGGATGGTACTGCATACGTAGTATTCGGAGTTGTGGAAAACAAAC  700
      G  S  P  E  L  T  V  N  I  K  A  T  Y  L  F  G  E  E  V  D  G  T  A  Y  V  V  F  G  V  V  E  N  K

 701 AGAAAAGGGGATTCCAAAGTTCTCTTCAGAGAGTACCGATTAAAGATGGTAAAGGAGATGCCACACTGAAGATAGAGCACATCACAAACACGTTCCCAAA  800
     Q  K  R  G  F  Q  S  S  L  Q  R  V  P  I  K  D  G  K  G  D  A  T  L  K  I  E  H  I  T  N  T  F  P  N

 801 TATCAATGAGCTGGTGGGGACTAATATATATGTAGCAGTCAGCGTGTTGACGGAAAGCGGCAGTGAAATGGTGGAGGCCGAGCTGCAAAACATCCAGATT  900
       I  N  E  L  V  G  T  N  I  Y  V  A  V  S  V  L  T  E  S  G  S  E  M  V  E  A  E  L  Q  N  I  Q  I

 901 GTCACATCGCCGTACACCATCACCTTCAAGAAAACGCCCAAGTATTACAAGCCAGGAATGTCCTTCGATGTGGCGGTGGAAGTTGTAAATCCTGATGGTA  1000
      V  T  S  P  Y  T  I  T  F  K  K  T  P  K  Y  Y  K  P  G  M  S  F  D  V  A  V  E  V  V  N  P  D  G

1001 CTCCTGCAGACGGAATTTCTGTGGTGATTAAACCAGGCCCGGTGAATGGGAATACCGCAGCTAACGGCATGGCAAGGCTCACCATCAACACCATGGGAGG  1100
     T  P  A  D  G  I  S  V  V  I  K  P  G  P  V  N  G  N  T  A  A  N  G  M  A  R  L  T  I  N  T  M  G  G

1101 AAATAACAACTTACAGATCACTGCCGAGACCAATGATCCTCGGCTTTCTCCTGAAAGACAAGCAAAAGCCACCATGGTAGCTGTCCCGTACCAGACAAGC  1200
       N  N  N  L  Q  I  T  A  E  T  N  D  P  R  L  S  P  E  R  Q  A  K  A  T  M  V  A  V  P  Y  Q  T  S

1201 AGCAACAGCTACATCCACATAGGTGTGGACGCAGCCGAGCTGAAATTAGGAAACAATCTGAAAGTCAACCTCAACCTCAACAATAATGCAAATCGTGACA  1300
      S  N  S  Y  I  H  I  G  V  D  A  A  E  L  K  L  G  N  N  L  K  V  N  L  N  L  N  N  N  A  N  R  D

1301 TCACTTACCTGATCCTGAGCAGAGGCCAGCTGGTGAAATCTGGTCGATACAAGACAAGAGGCCAAGTACTGATTTCCCTGATAGTTCCCGTTACAAAAGA  1400
     I  T  Y  L  I  L  S  R  G  Q  L  V  K  S  G  R  Y  K  T  R  G  Q  V  L  I  S  L  I  V  P  V  T  K  E

1401 GATGCTGCCATCATTCCGCATCGTAGCCTACTACCACACAAATAACAACGAAGTGGTGTCCGATTCCATTTGGGTGGATGTCCAGGATGACTGCATGGGT  1500
       M  L  P  S  F  R  I  V  A  Y  Y  H  T  N  N  N  E  V  V  S  D  S  I  W  V  D  V  Q  D  D  C  M  G

1501 TCGCTGAGGCTAGAACCAACAAGACCAGTTCCATCTTATGAACCTCGCAAGATGTTTGGTCTGAAGGTGATTGGAGACCCAACAGCCACAGTGGGACTGG  1600
      S  L  R  L  E  P  T  R  P  V  P  S  Y  E  P  R  K  M  F  G  L  K  V  I  G  D  P  T  A  T  V  G  L

1601 TGGCAGTCGACAAAGGCGTCTACGTTCTCAACAACAAGCACCGTCTCACACAGAAGAAGGTTTGGGACATTGTGGAAAAGTACGACACGGGTTGCACACC  1700
     V  A  V  D  K  G  V  Y  V  L  N  N  K  H  R  L  T  Q  K  K  V  W  D  I  V  E  K  Y  D  T  G  C  T  P

1701 AGGTGGAGGGAAGAACAGCATGAGTGTGTTCTTCGATGCTGGGCTGTTGTTTGAAACCAGCACAGCTGTGGGTACTCCCTACAGACAAGAGCTGAAATGT  1800
       G  G  G  K  N  S  M  S  V  F  F  D  A  G  L  L  F  E  T  S  T  A  V  G  T  P  Y  R  Q  E  L  K  C

1801 GCAGCTCCCAGCAGGAGGAAACGAAGCACTACTATTACGGACACCATCACCAGTTTAACGAGTCGGTATAACGACACTCTGCAACGGGACTGTTGTAGGG  1900
      A  A  P  S  R  R  K  R  S  T  T  I  T  D  T  I  T  S  L  T  S  R  Y  N  D  T  L  Q  R  D  C  C  R
                 β-α cleavage site
1901 ATGGCATGGCAGACACTCCAGTTTCATACTCGTGTGAGCGACGAAGCGAATACATTGACGATGGTCCATCTTGTGTCGAAGCCTTTGTGCACTGCTGCAA  2000
     D  G  M  A  D  T  P  V  S  Y  S  C  E  R  R  S  E  Y  I  D  D  G  P  S  C  V  E  A  F  V  H  C  C  K
                                     ANA domain
2001 GGAGATGGAAAAAGAGCGAGCTGAAAAAAGAGAGGAAAACCTTCAACTGGCTCGCAGTGAAGATGATGACAACAGTTACATGGACAGCAGTGAAATTGTT  2100
       E  M  E  K  E  R  A  E  K  R  E  E  N  L  Q  L  A  R  S  E  D  D  D  N  S  Y  M  D  S  S  E  I  V

2101 TCTCGAACACAATTCCCAGAAAGTTGGTTGTGGTTGGATATCAATTTGCCTCCTTGTCCTAGAAATACCCCCAACTGTGAAACCACATCGTTTGTGAGAA  2200
      S  R  T  Q  F  P  E  S  W  L  W  L  D  I  N  L  P  P  C  P  R  N  T  P  N  C  E  T  T  S  F  V  R

2201 ATGTTCCTCTGCAAGATTCAATCACGACCTGGCAGTTCACTGGCATCAGTCTATCTAGAACTCTAGGGATCTGTGTTGGTGATCCATTGGAGGTCATTGT  2300
     N  V  P  L  Q  D  S  I  T  T  W  Q  F  T  G  I  S  L  S  R  T  L  G  I  C  V  G  D  P  L  E  V  I  V

2301 CAGGAAAGAATTCTTCATTGATCTCAAACTCCCCTACTCTGCTGTCCGAGGAGAGCAGATCGAAATAAAGGCCATCCTCCACAACTACAGCCCAGATGAT  2400
       R  K  E  F  F  I  D  L  K  L  P  Y  S  A  V  R  G  E  Q  I  E  I  K  A  I  L  H  N  Y  S  P  D  D

2401 GTCACTGTGCGTGTCGATCTGACTGAGGAGCAGCATGTGTGCAGTTCAGCCTCGAAACGTGGAAGATACCGCCAGGAAGTTAGAGTTGGGCCTGAAACCA  2500
      V  T  V  R  V  D  L  T  E  E  Q  H  V  C  S  S  A  S  K  R  G  R  Y  R  Q  E  V  R  V  G  P  E  T

2501 CACGATCTGTGCCCTTCATCATTATTCCCATGAAGGAAGGACAATATCGCATCGAGGTCAAAGCGGCTGTCAAAGATTCTTCGCTCAACGATGGCATCAT  2600
     T  R  S  V  P  F  I  I  I  P  M  K  E  G  Q  Y  R  I  E  V  K  A  A  V  K  D  S  S  L  N  D  G  I  I

2601 TAAGACACTGAGGGTGGTGCCAGCAGGAGTACTGGTTAAGGTTCCTCTGGTTTTACCGTTAGACCCAGCAAAACAAGGTGTTGGCGGTACACAGAAAGAA  2700
       K  T  L  R  V  V  P  A  G  V  L  V  K  V  P  L  V  L  P  L  D  P  A  K  Q  G  V  G  G  T  Q  K  E

2701 ATAATCAACAGTGCAATTCCTAAGAAAGATATTGCTCCAAACACACCCACGAGCACACAGATCTCTGTTTCAGGTAAAGAACAGGTATCAACACTAGTGG  2800
      I  I  N  S  A  I  P  K  K  D  I  A  P  N  T  P  T  S  T  Q  I  S  V  S  G  K  E  Q  V  S  T  L  V

2801 AGAAGGCCATCAGTGGGAACTCAATGGGCAGTCTGATAGTCCAGCCCTCAGGTTGTGGAGAGCAGACCATGATCCGAATGACCCTCCCTGTCATTGCAAC  2900
     E  K  A  I  S  G  N  S  M  G  S  L  I  V  Q  P  S  G  C  G  E  Q  T  M  I  R  M  T  L  P  V  I  A  T
                                                        Thioester site
```

Supplementary Fig. S5

```
2901  CACATACCTGGATACGACTAACCAGTGGGAAAAAGTGGGCTTTCAGAAACGTGCCGAAGCTCTTCAACACATTAAGACTGGCTATGAGAGGATGCTCACT  3000
        T  Y  L  D  T  T  N  Q  W  E  K  V  G  F  Q  K  R  A  E  A  L  Q  H  I  K  T  G  Y  E  R  M  L  T

3001  TTCCGTAAAGGCGACGGATCTTTTGCTGTGTATCCTAAACACGTGAGCAGCACCTGGCTGACAGCTTATGTTACAAAAGTGTTTGCTATGGCCCACAATC  3100
       F  R  K  G  D  G  S  F  A  V  Y  P  K  H  V  S  S  T  W  L  T  A  Y  V  T  K  V  F  A  M  A  H  N

3101  TGGTGGCTGTGCAAACCGACATTCTGTGCGGCGCTGTCAGGTATCTCATCCTGAAGACGCAGCAGCCCGATGGCGTATTCAGAGAAATTGGGAAGGTGTA  3200
      L  V  A  V  Q  T  D  I  L  C  G  A  V  R  Y  L  I  L  K  T  Q  Q  P  D  G  V  F  R  E  I  G  K  V  Y

3201  CCATGGAGAGATGATTGGTGATATACGTGGTGGAGATTCAGATGCTTCCATGACAGCCTTCTGCGTCATTGCTATGCAAGAGGCACGTGCACTATGTGCG  3300
        H  G  E  M  I  G  D  I  R  G  G  D  S  D  A  S  M  T  A  F  C  V  I  A  M  Q  E  A  R  A  L  C  A
   Catalytic histidine
3301  GCCAGCGTTAATAGTCTATCAGGTGCGGTAGACAAATCAGTGTCCTACCTGGAGAAGCGTCTGCCCGGCCTCACCAACCCATATGCTGTTGCCATGACAT  3400
       A  S  V  N  S  L  S  G  A  V  D  K  S  V  S  Y  L  E  K  R  L  P  G  L  T  N  P  Y  A  V  A  M  T

3401  CATACGCGCTGGCCAATGAAAAGAAACTGAACCGGGAGATCCTCTACAAGTTTGCTTCAACAGAATTCTCCCACTGGCCAACACCTAAGGGACATGTTTA  3500
      S  Y  A  L  A  N  E  K  K  L  N  R  E  I  L  Y  K  F  A  S  T  E  F  S  H  W  P  T  P  K  G  H  V  Y

3501  CACCTTGGAAGCCACAGCTTATGCTCTTCTTGCACTTGTCAAGGCTAAGGCATATGAAGATGCCAAACCTATTGTAAGATGGTTTAACAGACAGCAGAAG  3600
        T  L  E  A  T  A  Y  A  L  L  A  L  V  K  A  K  A  Y  E  D  A  K  P  I  V  R  W  F  N  R  Q  Q  K

3601  GTGGGAGGAGGTTATGGATCAACTCAGGCAACTATAATGGTGTACCAGGCTGTAGCTGAGTACTGGGCCAATGCTCAAGAACCAGAGTATGATTTGAATG  3700
       V  G  G  G  Y  G  S  T  Q  A  T  I  M  V  Y  Q  A  V  A  E  Y  W  A  N  A  Q  E  P  E  Y  D  L  N

3701  TGGACGTATTGTTACCAGGCAGGTCAAGGCCTGAAAAGTTCAAGTTCACCCGTGACAACCAGTATACCACAAGATCATCTAAAATCAATGACATCAATCA  3800
      V  D  V  L  L  P  G  R  S  R  P  E  K  F  K  F  T  R  D  N  Q  Y  T  T  R  S  S  K  I  N  D  I  N  Q

3801  AAATGTGACTGTGACTGCCACAGGCTCAGGAGAAGCAACATTTACAATGGTGTCGTTGTACTACGCTCTCCCTAAAGAAAAGGAAAGTGATTGTCAGAAG  3900
        N  V  T  V  T  A  T  G  S  G  E  A  T  F  T  M  V  S  L  Y  Y  A  L  P  K  E  K  E  S  D  C  Q  K

3901  TTTGACTTGAAAGTGGAGCTCATCCCAGACAAATTGGGGGATGATGAGAAGATATACAAACTAAAAATAAATGTTTTGTATAAGAGCCGGGAGCATGATG  4000
       F  D  L  K  V  E  L  I  P  D  K  L  G  D  D  E  K  I  Y  K  L  K  I  N  V  L  Y  K  S  R  E  H  D

4001  CAACAATGTCAATCCTTGATATTGGTTTGCTAACTGGCTTCACTGTTAACACAAATGACCTGGACAGGCTGGCCAAAGGACGCTCACGCATCATCTCCAA  4100
      A  T  M  S  I  L  D  I  G  L  L  T  G  F  T  V  N  T  N  D  L  D  R  L  A  K  G  R  S  R  I  I  S  K

4101  ATATGAGATGAACACAGCCCTGTCAGAGAGAGGCTCACTCATCATCTATCTGGACAAGGTTTCTCACACTCGATCAGATGAGATCACATTCAGGATTCAT  4200
        Y  E  M  N  T  A  L  S  E  R  G  S  L  I  I  Y  L  D  K  V  S  H  T  R  S  D  E  I  T  F  R  I  H

4201  CAGACTCTAAAAGTAGGCGTCCTACAACCAGCCGCTGTTTCTGTCTACGAATACTATGACCAAACACCATGCGTAAAGTTCTATCACCCAGAGAGAAAAG  4300
       Q  T  L  K  V  G  V  L  Q  P  A  A  V  S  V  Y  E  Y  Y  D  Q  T  P  C  V  K  F  Y  H  P  E  R  K

4301  CTGGACAGCTCCTGAGGCTCTGCAGAAATGATGAATGCACGTGTGCCGAAGAGAACTGCAGCATGCAGAAAAAAGGCAAAATTGACAATGATGAGCGCAC  4400
      A  G  Q  L  L  R  L  C  R  N  D  E  C  T  C  A  E  E  N  C  S  M  Q  K  K  G  K  I  D  N  D  E  R  T
                                                                        C345C domain
4401  GGCTAAGGCCTGTGAGACTGAAGAGAACAGCAAAATAGATTTCGTGTATAAAGTACTGCTGGAGGACTTTAAAGCTGACTTATCCACTGATATTTACGTA  4500
        A  K  A  C  E  T  E  E  N  S  K  I  D  F  V  Y  K  V  L  L  E  D  F  K  A  D  L  S  T  D  I  Y  V

4501  GTGAAGATAGTGGAAGTAATCAAGGAAGGAAGTTACGACGTGGGTCCTCTGAACAAAAAGCGTGCATTCCTCAGTTATCCACACTGCAGAGAGTCTTTAG  4600
       V  K  I  V  E  V  I  K  E  G  S  Y  D  V  G  P  L  N  K  K  R  A  F  L  S  Y  P  H  C  R  E  S  L

4601  ATCTGAAGACAGGCAAAACCTACCTTATCATGGGAACATCCAGAGATATTCACAGAGATGAGCAAAATAACTCATACCAGTATGTGCTCGGTGAAAGGAC  4700
      D  L  K  T  G  K  T  Y  L  I  M  G  T  S  R  D  I  H  R  D  E  Q  N  N  S  Y  Q  Y  V  L  G  E  R  T

4701  CTGGGTTGAGTACTGGCCCACTGAAGACGAGTGCCAACTTGATGAACACAGACCAACGTGTTTCGGTATGGAGGAGCTGGTGCAGCAGTATGCACTCTTT  4800
        W  V  E  Y  W  P  T  E  D  E  C  Q  L  D  E  H  R  P  T  C  F  G  M  E  E  L  V  Q  Q  Y  A  L  F

4801  GGATGTCTGCAGTAGTGAACCTTTACAAGTCAAATGTTCTCTTCAAAAATAAATTGTATTCATTGTTCTTGTGTTAATTTTTCACTGAACTGAGCTGTGG  4900
       G  C  L  Q  *

4901  AAGTCGTGTTCGACTGCCCTAAATAAAATTCAGATGATTCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  4972
```

PaspA2M

```
   1  GAGCGGGGATTTTCGCTATCCTAGTCTGTTCCTCCTCTCGTCGTTGATTTTGTTGGATAAAAGTGTGAGTAGTAAGTGAGTGAGTGTGTGTATATATTTA  100
 101  TATACATACATACATACATATATATTGGAAAGAGAGGTTAGATAGAGGGATCATTTGATATGCGGAACTCGGTCGGGGCACTTTGGGCCTTTTGCCTATT  200
                                                                  M  R  N  S  V  G  A  L  W  A  F  C  L  F

 201  CCAGGCGGCCTTTTCATCTTCCCCGCCTTCCATAAGCCCGGAGGTACTCGAGAAGTTCTTGGCGGGTAAAAATCATACGAATCTTCGAGGATCATCCGGG  300
        Q  A  A  F  S  S  S  P  P  S  I  S  P  E  V  L  E  K  F  L  A  G  K  N  H  T  N  L  R  G  S  S  G
                                                                                            CUB domain
 301  ATTTTCAGATCTCCCAACTTTCCTGAGGAATATCCCGCTAATTCAAACTTTACCTGGAAAATCGATGTTCCGGAAGGAGAATTCCTTCAGCTCTCTTTCC  400
       I  F  R  S  P  N  F  P  E  E  Y  P  A  N  S  N  F  T  W  K  I  D  V  P  E  G  E  F  L  Q  L  S  F

 401  ATCTTTTTAACGTGGAAAGAAGCGATCATTGTCAGTGGGATTGGGTGAACATTTATGATTACAATAACAAGACTCAATTAAGGCAGCTATGCGGTTATCT  500
      H  L  F  N  V  E  R  S  D  H  C  Q  W  D  W  V  N  I  Y  D  Y  N  N  K  T  Q  L  R  Q  L  C  G  Y  L

 501  CCAGCAAGACCTCGTTATAATCTTCAATACAAGCAGCATTTTATTGGAGCTTCACTCCGACTCTGTTATTCACAAAACCGGCTTTTTAGCTTCGTGGAAA  600
        Q  Q  D  L  V  I  I  F  N  T  S  S  I  L  L  E  L  H  S  D  S  V  I  H  K  T  G  F  L  A  S  W  K

 601  TCCGTTACTGCTACAGAAGAAGAAAAAGAAGATTCAGAGAGTGATAAGAAAAAGGAAGGTTATATTCTCACATTCCCTCAATCCTTTACCTCTTCCAAAG  700
       S  V  T  A  T  E  E  E  K  E  D  S  E  S  D  K  K  K  E  G  Y  I  L  T  F  P  Q  S  F  T  S  S  K

 701  AGACCTCGGCTGAAAAGACTTGCCTTCAGATTTTCAACGTTCATACGAAGGGGAAAGTATTCATAAATGTATTTTCGTCTGAAAATATTCTGTCTGGAAA  800
      E  T  S  A  E  K  T  C  L  Q  I  F  N  V  H  T  K  G  K  V  F  I  N  V  F  S  S  E  N  I  L  S  G  N

 801  TCCGGTTCTATATAAGGAGATAGATTATGATCCATCTGTTCAAGGAGAAAAACTACAATGTTTCGATATTGAATTGCCAACTGATTACTCAATGAAAAAA  900
        P  V  L  Y  K  E  I  D  Y  D  P  S  V  Q  G  E  K  L  Q  C  F  D  I  E  L  P  T  D  Y  S  M  K  K
```

Supplementary Fig. S5 (continued)

```
 901  TACGCCATCGTTAATATTAATGGTACATTTGAGGAGTCTGATTACAGTATATTAAGCTACAAGACCCTAAGAGTTCTTAAAGTCCCTGTCAAGACCCTTA  1000
       Y  A  I  V  N  I  N  G  T  F  E  E  S  D  Y  S  I  L  S  Y  K  T  L  R  V  L  K  V  P  V  K  T  L

1001  CTCAAACGGATAAGGGAGAATATCGACCGGAACAAACCGTATTATTCCGCGTTCTTAAATTGAATGGAGAATTAAAGCCATCCGAAAATCGAACAATTGA  1100
      T  Q  T  D  K  G  E  Y  R  P  E  Q  T  V  L  F  R  V  L  K  L  N  G  E  L  K  P  S  E  N  R  T  I  D

1101  TGAAATTTACGTCAAGTCTCCTTTGGGGAGGACGATGGCTCAGTACAAAAATGTAACCTTGAGCCCGCAAGGACTCGGTCAATTCGAATATACTTTAGAT  1200
        E  I  Y  V  K  S  P  L  G  R  T  M  A  Q  Y  K  N  V  T  L  S  P  Q  G  L  G  Q  F  E  Y  T  L  D

1201  AAAGAACCAGAGTTGGGACCTTGGCAAATCGTTGTTCGAGTTTTTGGAAATGACACTCACGAGCCTGTTGAAGATATCGCAATTTTTGTTGTAAATGAAG  1300
       K  E  P  E  L  G  P  W  Q  I  V  V  R  V  F  G  N  D  T  H  E  P  V  E  D  I  A  I  F  V  V  N  E

1301  CTGTACTTCCTAAATTTCAAGTAAAAGTCGAGGCTCCAAAGGCTATTTTAGCTGAGAGCGAAGAAGTCAAATTTGAGATTTGCGGGATCTACACTCACGG  1400
      A  V  L  P  K  F  Q  V  K  V  E  A  P  K  A  I  L  A  E  S  E  E  V  K  F  E  I  C  G  I  Y  T  H  G

1401  AAGTAAGGTTAAAGGAACAGTGAAGGCTCATTTCGAACACAAATACAGAGTTGGTACTTACTGGAGAGCTCCTTATAAGACCTTTAATTTAACAAAAATC  1500
        S  K  V  K  G  T  V  K  A  H  F  E  H  K  Y  R  V  G  T  Y  W  R  A  P  Y  K  T  F  N  L  T  K  I

1501  GTGGAAATAAAAGATGACGACTCATGCGCCACTGTAATACTAAATTCGACTGAGCTGGGAGATTTATCTAAAAAGCCCGATGAACCATTTACTCTCACGG  1600
       V  E  I  K  D  D  D  S  C  A  T  V  I  L  N  S  T  E  L  G  D  L  S  K  K  P  D  E  P  F  T  L  T

1601  CAAACTTAACAGAACACGGAACTGACATTGTACAAAGTCATACACTTACTTCTAAAATCACATCACCCACGCAAAAATCGATTTCCAAGATAGCGCTTCC  1700
      A  N  L  T  E  H  G  T  D  I  V  Q  S  H  T  L  T  S  K  I  T  S  P  T  Q  K  S  I  S  K  I  A  L  P

1701  ACGCACATTGCCGGTGGATTCCCTTATGTTCTCGATTTCCGTGTTTTGGATCATTCTGATTCTCCTCTTAAAGGATCGAAAGTTAGACTATGTTCCAGAC  1800
        R  T  L  P  V  D  S  L  M  F  S  I  S  V  F  W  I  I  L  I  L  L  L  K  D  R  K  L  D  Y  V  P  D

1801  TTTTTTCATGATATAGAATCCTTCAGATCATTCATCCAAGATAAATACAGGAATTTATATTCCGCCGACGAAGATCAATTTTATGGACTTGGAAAGGTCC  1900
       F  F  H  D  I  E  S  F  R  S  F  I  Q  D  K  Y  R  N  L  Y  S  A  D  E  D  Q  F  Y  G  L  G  K  V

1901  TCGTAAAAATTTTGTTCAAAGAAACGTGCATAAATGTCACGACATCTTCAGACGATGGATCAGTCAAACTTGGAATTCCTCTAAGCGGCCTCTCTACCAA  2000
      L  V  K  I  L  F  K  E  T  C  I  N  V  T  T  S  S  D  D  G  S  V  K  L  G  I  P  L  S  G  L  S  T  N

2001  CGTAACAAAATTAAGCCTAAAAGTAACTGCACTGGAATTCCATGCCAACGAAACTACGAGAATGAAAAATACAGAATCTAAACATGAAGTTATCTTAAGC  2100
        V  T  K  L  S  L  K  V  T  A  L  E  F  H  A  N  E  T  T  R  M  K  N  T  E  S  K  H  E  V  I  L  S

2101  CATATTGGAAATGATAGCGCACTTAATGATCTTCATAGATTGGCAATTCGAGAGAGACTTATTGATGATACGAAAATTAAATTTGATTGCGAAGGAACCA  2200
       H  I  G  N  D  S  A  L  N  D  L  H  R  L  A  I  R  E  R  L  I  D  D  T  K  I  K  F  D  C  E  G  T

2201  CTACTTTTACCGTTTACTTCCAGGCAGAAATGGGAGACGTAGTCGATGTTAATTATGTAATATCTTCAGGAGGATCTTTGATTCATTCTGGAGTTAAACG  2300
      T  T  F  T  V  Y  F  Q  A  E  M  G  D  V  V  D  V  N  Y  V  I  S  S  G  G  S  L  I  H  S  G  V  K  R

2301  AGTTATGGTGAATTCGAATGACACAGATAAATATTTAGAAGGGTTAATAAGCGAAGATAGTCTAGGGTACGTTAACTTAACAGATTCTCCTGCTAGTGTA  2400
        V  M  V  N  S  N  D  T  D  K  Y  L  E  G  L  I  S  E  D  S  L  G  Y  V  N  L  T  D  S  P  A  S  V

2401  TTAAAAGCCTTTGTCTTATCAGTGGATCGTCCCTTTCCAAACGAACACAAAATCACTAAGCTTCTTAAAATTTTAGTATTTACTCGAGACGTAAATTCTG  2500
       L  K  A  F  V  L  S  V  D  R  P  F  P  N  E  H  K  I  T  K  L  L  K  I  L  V  F  T  R  D  V  N  S

2501  GCTCAACTCTTAGCTCTATTAGAGAATATGAAACTGAATCCTGTAATGCTAAGCCCTCACTTTCTTTCTCCCCATCTAAAACTCGACCCTCAGATCTTGT  2600
      G  S  T  L  S  S  I  R  E  Y  E  T  E  S  C  N  A  K  P  S  L  S  F  S  P  S  K  T  R  P  S  D  L  V

2601  TTCTTTGAGTATTGGGAGTAAACCAAACTCATTATGTGGCTACAGCGTTATTGATAAATCCGCAGATCTGGTTGAAAATCCCAATGAAATAACGAGCAAA  2700
        S  L  S  I  G  S  K  P  N  S  L  C  G  Y  S  V  I  D  K  S  A  D  L  V  E  N  P  N  E  I  T  S  K

2701  AAGTTCCAAAATATAAAGGAGGATCTGGCCAAAAAACGAATTATAACTGAGGGAATTTACGGAGATCGATGCAAAAACGCTCAATACTTGTTCCAAGCTG  2800
       K  F  Q  N  I  K  E  D  L  A  K  K  R  I  I  T  E  G  I  Y  G  D  R  C  K  N  A  Q  Y  L  F  Q  A

2801  CTCAACAAATTGGTCTTTTCATATTGAGTGATACTCTTGTTCATGATCTTAAGTGTGATACTGTTGTTGATTCTAGTGGTTTTGGAAACATTAACAAGGA  2900
      A  Q  Q  I  G  L  F  I  L  S  D  T  L  V  H  D  L  K  C  D  T  V  V  D  S  S  G  F  G  N  I  N  K  D
                                                                                     Bait region
2901  TTTCGAATATTATGATGAACCCGTAGCTTTCGTCGCAGGTCCTCCGGTACAATTCCAATCTGTGAGCTTAAATGCTAGTCCTCCAGTCAATCGTTTTGAA  3000
        F  E  Y  Y  D  E  P  V  A  F  V  A  G  P  P  V  Q  F  Q  S  V  S  L  N  A  S  P  P  V  N  R  F  E

3001  GAAGCCGGAGCGGGCTCAGCAAGACCCGTTCAACAAAATAAACAAGGAGATATATCTTTATCTGTACCATCTCTGGAGATCCGGGATTACTTTCCAGAAA  3100
       E  A  G  A  G  S  A  R  P  V  Q  Q  N  K  Q  G  D  I  S  L  S  V  P  S  L  E  I  R  D  Y  F  P  E

3101  CCTGGTTATTCGATACTGTCTTGTTCTCTTCAGAAGACAGCGAAATCGTAGTGAAGGTAAACGTACCTCATACTCTCACAACATGGGTAGCGGACGCTTT  3200
      T  W  L  F  D  T  V  L  F  S  S  E  D  S  E  I  V  V  K  V  N  V  P  H  T  L  T  T  W  V  A  D  A  F

3201  CTGCTCTCATACGGAAGATGGTTTTGCTGTTGCGGAACGTGCAGAGTTAAAAGTATCGCAAGACTTTTTTGTAGATCTTAATGCACCCTACTCCGCCAAG  3300
        C  S  H  T  E  D  G  F  A  V  A  E  R  A  E  L  K  V  S  Q  D  F  F  V  D  L  N  A  P  Y  S  A  K

3301  AGAGGTGAGATCCTCCAGATCAATGCAACGGTATTTAATAAAGTTGAAGGACCTCTTCCCATGAAAATAAGTATCCTTCCATCGGAAGGTCATTATTCCG  3400
       R  G  E  I  L  Q  I  N  A  T  V  F  N  K  V  E  G  P  L  P  M  K  I  S  I  L  P  S  E  G  H  Y  S

3401  TCATTAATAACTCTGATGTAAATATCTGCGTTCAAAGTAAAGGGAATGAAGTTATTGGATTTTTCGTGGAATTTCAAAAGCTTCATGAGGTCAATGTGAC  3500
      V  I  N  N  S  D  V  N  I  C  V  Q  S  K  G  N  E  V  I  G  F  F  V  E  F  Q  K  L  H  E  V  N  V  T

3501  TATTTCTGCCGAAATAGTCCAAGATTCTTCTTGCGGTGATGTTTCCGAATCCTCTATTGGTATTTCCGACTCATTAAAGAAACCAATCATTATTAAGCCC  3600
        I  S  A  E  I  V  Q  D  S  S  C  G  D  V  S  E  S  S  I  G  I  S  D  S  L  K  K  P  I  I  I  K  P

3601  GAAGGATTTCCACATGAAGAAGTGAGATCTTACTTCTTATGTGGTGAGCAAAATGATACAGTCTTGGAGGATGTGAATTTACCCGAAGATGATTTAGTCG  3700
       E  G  F  P  H  E  E  V  R  S  Y  F  L  C  G  E  Q  N  D  T  V  L  E  D  V  N  L  P  E  D  D  L  V

3701  ACGGTTCTGTGAGAGCGCGTTTCTCTATTTCGGGAGATATAATGGCTCCCGCAGTAAAGAATTTGGGCCGACTGGTCTCTCTTCCTACTGGATGTGGTGA  3800
      D  G  S  V  R  A  R  F  S  I  S  G  D  I  M  A  P  A  V  K  N  L  G  R  L  V  S  L  P  T  G  C  G  E
                                                                                                 Thioester site
3801  ACAAACGATGATCCGAATGGTTCCTAACATTTATCTGCTTGACTATCTCAAATCTGTAGGAAAGAATTTGCCCGAATTGGAAACCAAAGCTCTTAAACAT  3900
        Q  T  M  I  R  M  V  P  N  I  Y  L  L  D  Y  L  K  S  V  G  K  N  L  P  E  L  E  T  K  A  L  K  H
```

Supplementary Fig. S5 (continued)

```
3901  ATGAATCTGGGCTACGATCGACAGAATAAAAAATTCCGTCATAAAGATGGGTCTTATTCCATATGGGGTCCGTCGAACACCGAAAGTGAGGGCTCGATGT  4000
       M  N  L  G  Y  D  R  Q  N  K  K  F  R  H  K  D  G  S  Y  S  I  W  G  P  S  N  T  E  S  E  G  S  M

4001  GGTTAACTGCCTATGTAGTCAAAGCCTTTTCTCAAGCCTCTAAATATATTGATGTCGATAAGAATCTTCTCAAAGAATCTGTGAGATGGATTTTGGAGAG  4100
      W  L  T  A  Y  V  V  K  A  F  S  Q  A  S  K  Y  I  D  V  D  K  N  L  L  K  E  S  V  R  W  I  L  E  R

4101  ACAACATCGCGAAACCGGGTGTTTTAGAAACGAAGGATATGCATATTCCGTCAACAGTCCAAGAGAAACCCTAACATCTCACGTTCTAATCACACTTTTC  4200
        Q  H  R  E  T  G  C  F  R  N  E  G  Y  A  Y  S  V  N  S  P  R  E  T  L  T  S  H  V  L  I  T  L  F

4201  GAAGCAAGATACACAGCGAATTTGTCTGAAATTATTAACTCCAAAATATTCTACAGGGCCCTCAAATGTCTGAGCGATTCTTTAAACGAAGAACCCGAAA  4300
       E  A  R  Y  T  A  N  L  S  E  I  I  N  S  K  I  F  Y  R  A  L  K  C  L  S  D  S  L  N  E  E  P  E

4301  AACCTGAGCTCGAAGAAGAGAATAAAAAAAATAAGCAAAAATCTATTTATACACAGTCCCTTAAAACTTACGCCATACATCTAATTGAAACAAAGGAAGA  4400
      K  P  E  L  E  E  E  N  K  K  N  K  Q  K  S  I  Y  T  Q  S  L  K  T  Y  A  I  H  L  I  E  T  K  E  D

4401  TTTGAAAAAGGAAATCGGAGAAAATTTTGAGGGAATCAAAGCCGATACCCTCTTCAAGAACCTATTAGACGCCTCCAAAAGAGACGACCAAAATCTTCTT  4500
        L  K  K  E  I  G  E  N  F  E  G  I  K  A  D  T  L  F  K  N  L  L  D  A  S  K  R  D  D  Q  N  L  L

4501  CATTGGAAAAATAATTCCTCAAAGTCTAGATCTGTGGAAATGACGGCCTATAATATCATGACACTGCTCTTCAATGATGAATTAATCGATGCTCTTAGTG  4600
       H  W  K  N  N  S  S  K  S  R  S  V  E  M  T  A  Y  N  I  M  T  L  L  F  N  D  E  L  I  D  A  L  S

4601  CGATTCGCTGGATATCTAAGTATCGAAACGGACGAGGAGGATTCATATCTACTCAAGACACAGTAGTAGCATTAGAAGCGATCAGTAAATATTCATCAAG  4700
      A  I  R  W  I  S  K  Y  R  N  G  R  G  G  F  I  S  T  Q  D  T  V  V  A  L  E  A  I  S  K  Y  S  S  R

4701  AGTCTTTGAAAACGCAACTGATCTTTCCGTGGACTTGTATAACACGACCGGTACAATTGAGTCCTTCCATATCGTCGAGGACGATAAAATTCTTTTGAGA  4800
        V  F  E  N  A  T  D  L  S  V  D  L  Y  N  T  T  G  T  I  E  S  F  H  I  V  E  D  D  K  I  L  L  R

4801  AGAATTGACGTAGATGAAGTTAAAGATTTTAAGGTCTCGAGCGTCGGTAAAGGTTGTTATTCCTTCCAAACTAGCATCAAATACAATACCAAATCTAATG  4900
       R  I  D  V  D  E  V  K  D  F  K  V  S  S  V  G  K  G  C  Y  S  F  Q  T  S  I  K  Y  N  T  K  S  N

4901  AAAAAGAAGAAGAAAAGTTTTTCCTTAAAGCAGAAACCAACGAAACCTTTGTTAATGTCTGCGGATCTTATAAAGGTCAATCTTCGGTTACCAACATGAT  5000
      E  K  E  E  E  K  F  F  L  K  A  E  T  N  E  T  F  V  N  V  C  G  S  Y  K  G  Q  S  S  V  T  N  M  I

5001  TATTCTCGAAGTGGAGCTCTTGAGTGGATTCGAAGTTGTAGAGTCTTCGATAGAAATCCTCTTGAATGAAATTGACTCTGGAGTACAAAAATACGAAATT  5100
        I  L  E  V  E  L  L  S  G  F  E  V  V  E  S  S  I  E  I  L  L  N  E  I  D  S  G  V  Q  K  Y  E  I

5101  GAAGAGAAAGAAGGAAAGTTCGTATTGTATTTTGATAATATGAAAAAAGATGAGCTCAAATGTTGGGACTTCGAAGTCAAAAGAGTTAGTCCCGTCGAGA  5200
       E  E  K  E  G  K  F  V  L  Y  F  D  N  M  K  K  D  E  L  K  C  W  D  F  E  V  K  R  V  S  P  V  E

5201  ATCTCAAACCCGCTCTAGTTAAAATCTACGACTATTATTCCCAAGAAGATTCCTTTACTACTAGTTACACAATTTAAAAAAAAGTCTTATTATATTTCTT  5300
      N  L  K  P  A  L  V  K  I  Y  D  Y  Y  S  Q  E  D  S  F  T  T  S  Y  T  I  *

5301  TTTTCTTGATGCTTGATCGAAAAAAAAAAAA  5330
```

Supplementary Fig. S5 (continued)

HaadA2M-2

```
   1  CCAGTCAGGAAAGAGGGAGACGCTGAAGCCCGTTGATAACAACTGATCGGAACCGGCTCTAGGAAGTCGGTATTTTTTGAAGCACATCATTGTTATTTAA  100
 101  AAGGATATCAAGGCGCGGGTTTTTAAATTGGATTACATTTTAGAATGGGGATTCATTCAATAGTTTGGACTATAGTGTCTTTCCTCTACTTAATTAATGC  200
                                                   M  G  I  H  S  I  V  W  T  I  V  S  F  L  Y  L  I  N  A

 201  AGAAACTGAATGCAAAAAAGATGGTTACATTTTTACATGTCCGAGATCTCTCAAAACAGGCGCAAGCAACCAGATGCAACTTCGTCGATATGGGGAACTG  300
        E  T  E  C  K  K  D  G  Y  I  F  T  C  P  R  S  L  K  T  G  A  S  N  Q  M  Q  L  R  R  Y  G  E  L

 301  GATGCAGGAGAGTTTAAAATAACAGTAACATATATGAATTCTATAAATCAAAATGAGACAATTGCAACGGAACAGACATTCGACATTCCAGAAGGTGAAG  400
       D  A  G  E  F  K  I  T  V  T  Y  M  N  S  I  N  Q  N  E  T  I  A  T  E  Q  T  F  D  I  P  E  G  E

 401  CCGACACATTAATGACGGTATTTCTGGAACCTATTGAAAATTATGTTTTCAACGGCAAGGTCACAATCAATGGTACATTTGATGACTATGTCATAGGAGG  500
      A  D  T  L  M  T  V  F  L  E  P  I  E  N  Y  V  F  N  G  K  V  T  I  N  G  T  F  D  D  Y  V  I  G  G

 501  AGTCGAGAAAGTATATTTTAGTTCTTCTAGAGATGACATCGTGTTTATTCAGACGGACAAGCCATTGTATAAAGAAGGACAAACTGTGAAATTTAGAGTT  600
        V  E  K  V  Y  F  S  S  S  R  D  D  I  V  F  I  Q  T  D  K  P  L  Y  K  E  G  Q  T  V  K  F  R  V

 601  TTAAGAGTTGATAAGACTTTGCGACCTTCAGTAAAAGATCAAGCAAGCATATGGGTCGAGGATCCAGCAGGAACACGATTATTTCAGTGGAAAAATATTT  700
       L  R  V  D  K  T  L  R  P  S  V  K  D  Q  A  S  I  W  V  E  D  P  A  G  T  R  L  F  Q  W  K  N  I

 701  CAATGGAAAGAGGCATGAAGCAATTTGAATTTCCTCTGGCTGACGAACCAGTTTTAGGTAACTGGAAAATATCAGTCTCTTTTCAAGGAGAGGTAACCAC  800
      S  M  E  R  G  M  K  Q  F  E  F  P  L  A  D  E  P  V  L  G  N  W  K  I  S  V  S  F  Q  G  E  V  T  T

 801  GACTACTTTTGAAGTCAAGCAATACGTTCTTCCAACATTTGATGTAAAAATTACTTTACCATCGTTTGTATTGTCGAATGCTGAAGAAATTCCCATCACA  900
        T  T  F  E  V  K  Q  Y  V  L  P  T  F  D  V  K  I  T  L  P  S  F  V  L  S  N  A  E  E  I  P  I  T

 901  GTATGTGCAAAGTATACTTACGGAAAGCCTGTTAAAGGAGTTTTACGCCTGAATACTTCCCTAGAAATGTTTTCATGGGGAGATGAAAAATTTCCAACTA  1000
       V  C  A  K  Y  T  Y  G  K  P  V  K  G  V  L  R  L  N  T  S  L  E  M  F  S  W  G  D  E  K  F  P  T

1001  TAGAATACGAAGGAAAGATTAATGGGTGTTTCGACTATGTAATAAATGTATCCATGGTTGAAACAGAAGATTACTACAGATACAGAAGAATACAGATTGT  1100
      I  E  Y  E  G  K  I  N  G  C  F  D  Y  V  I  N  V  S  M  V  E  T  E  D  Y  Y  R  Y  R  R  I  Q  I  V

1101  GGCTAGTGTAGAAGAATCTGGTACAGGTATTGAAAGAAATGAGACGCAGTATGTGCAAAGGCAATATTCACCACTTTCACTCTCTTTTAACCGTGACCAA  1200
        A  S  V  E  E  S  G  T  G  I  E  R  N  E  T  Q  Y  V  Q  R  Q  Y  S  P  L  S  L  S  F  N  R  D  Q

1201  AAACAGTTTTACAAGCCAGGCTTACCTTACAATGGAAAACTTTTCGTTAAAAATCCAGATGATACCCCAGCAGCTGACGAAGGGATACAACTATGTTACA  1300
       K  Q  F  Y  K  P  G  L  P  Y  N  G  K  L  F  V  K  N  P  D  D  T  P  A  A  D  E  G  I  Q  L  C  Y

1301  CAGTCAATAAAGAAAGAGTAGTAATGGACGGGATGTGGAAAGCTACTAGGACAGTGAAATTTTGCCAGAACTACACATCGGATGACAATGGTGTCATTGA  1400
      T  V  N  K  E  R  V  V  M  D  G  M  W  K  A  T  R  T  V  K  F  C  Q  N  Y  T  S  D  D  N  G  V  I  E

1401  GTTCGTTATTCCAAGACAAAACACTGATTCTATTGATATAAACGTAGAGGCTAAATCATTGAAATATGCAAAAGATAATCAAAAATCAGGAACTCATCGC  1500
        F  V  I  P  R  Q  N  T  D  S  I  D  I  N  V  E  A  K  S  L  K  Y  A  K  D  N  Q  K  S  G  T  H  R

1501  GAAGGCTCTCTAAACCAGCCGCAGACTAGCATGTCACTCTCACCATGGTACTCACCATCCGGGAGCTTTATTCAACTACAACAAGTTCAAGAAACACTTC  1600
       E  G  S  L  N  Q  P  Q  T  S  M  S  L  S  P  W  Y  S  P  S  G  S  F  I  Q  L  Q  Q  V  Q  E  T  L

1601  TTTGCGGAACCAAAAATACTTTGAAAGTTCTCTTTACATCAGCGGAAGACGAGGATTATACATTTTACTATCAGGTTCTTAAACAAGGTAGAGTTGTAAA  1700
      L  C  G  T  K  N  T  L  K  V  L  F  T  S  A  E  D  E  D  Y  T  F  Y  Y  Q  V  L  K  Q  G  R  V  V  K

1701  GAAAGGATCCATAGAAAAGTCTTTTTCAACAAAAGATGATGTTGCGGACTTATATGAAGATGAGTATAAAGTTATTGATGACGTTGAAATGCAAATTGTA  1800
        K  G  S  I  E  K  S  F  S  T  K  D  D  V  A  D  L  Y  E  D  E  Y  K  V  I  D  D  V  E  M  Q  I  V

1801  CCACCTGTTGACCAGTCTGAAATTCCCAAGTCATCTGAACCTAAAGAAGAAGAATGTAAATCTGCAAAAGAAGCAAGATATGTACCTCCTGTCGGCGAAG  1900
       P  P  V  D  Q  S  E  I  P  K  S  S  E  P  K  E  E  E  C  K  S  A  K  E  A  R  Y  V  P  P  V  G  E

1901  TGGATATTGATCTCGATATTGATGCTTCCTGGTCTCCAACTTTCCATTTGTTGGTGTATTATATTCGTGATGATAGAGAAACCATCGCAGATTCACAAAA  2000
      V  D  I  D  L  D  I  D  A  S  W  S  P  T  F  H  L  L  V  Y  Y  I  R  D  D  R  E  T  I  A  D  S  Q  K

2001  ATTCAATGTTGAGAAATGCTTCAAAAATCAAGTGAAACTACAATTTGGAGATGACGTCAAACAGCCAGGTACTAAAACATCTATTAGAGTTACATCTTCT  2100
        F  N  V  E  K  C  F  K  N  Q  V  K  L  Q  F  G  D  D  V  K  Q  P  G  T  K  T  S  I  R  V  T  S  S

2101  CCCAATTCTTTGTGCGGCCTTAAAGTGGTTGATAAAAGTGTAGCATTGATGAATTCTGAAGATCAACTCACTCCTGAAAAAGTTTTCAGAGCTCTAGAAT  2200
       P  N  S  L  C  G  L  K  V  V  D  K  S  V  A  L  M  N  S  E  D  Q  L  T  P  E  K  V  F  R  A  L  E

2201  CTTTGGACACTAGCATGTACTATGGAATCAATCATTGTAATGAAAAAATTCGACAACCAGGTTTGTATTCTGCATCTAGTAAATATCTGCCGCGTCCTCC  2300
      S  L  D  T  S  M  Y  Y  G  I  N  H  C  N  E  K  I  R  Q  P  G  L  Y  S  A  S  S  K  Y  L  P  R  P  P

2301  TCAACCTTGGTCATCTTCATCTTACGAGGATTCGCTTGCTGCTTTTGAAAATGCTGGATTTTTGGTAATAAGTGACTTAATTTTGTTTACTCGGCCCTGT  2400
        Q  P  W  S  S  S  S  Y  E  D  S  L  A  A  F  E  N  A  G  F  L  V  I  S  D  L  I  L  F  T  R  P  C

2401  AAATCTAGAGGTGGAGGAGGTAACATCGCTTACGAGACAGGTTATGGAGGCGCAGTGGCTATGGCATCGACTGCAAGAAGACCCCCAGCTTCTCCGGCAA  2500
       K  S  R  G  G  G  G  N  I  A  Y  E  T  G  Y  G  G  A  V  A  M  A  S  T  A  R  R  P  P  A  S  P  A
                                                Bait region
2501  TGGCACCAGTAGCTGCTGATAAGATGGGAGAATTTTCCACCAAGTCTGTCGTAGATGTGCGAGATTATTTCCCAGAAACATGGCTATTTGACTTAAAATT  2600
      M  A  P  V  A  A  D  K  M  G  E  F  S  T  K  S  V  V  D  V  R  D  Y  F  P  E  T  W  L  F  D  L  K  L

2601  GACTGAAGAAGATGGCGTATACCTGGCAAAGGAAAAAATTGCCACATACTATTACCGAGTGGGTAGGAAGTGCTGTATGCATCAACGATGAAGATGGTCTC  2700
        T  E  E  D  G  V  Y  L  A  K  E  K  L  P  H  T  I  T  E  W  V  G  S  A  V  C  I  N  D  E  D  G  L

2701  GGATTATCGAACACAACAAGTATTAAAGGCTTCCAAGCTTTCTTCATCTCAATGACTCTACCTTATTCTGTAATTAGAGGAGAATCATTCTGGATTACTA  2800
       G  L  S  N  T  T  S  I  K  G  F  Q  A  F  F  I  S  M  T  L  P  Y  S  V  I  R  G  E  S  F  W  I  T

2801  TTTCTGTTTTCAGCTATGTCGAAGATCCTCTACCAATCACAGTTACTCTGGATAACCTTGAAGGATTTGAAATTGTTAGTGAATCAATAGATGGTGATAT  2900
      I  S  V  F  S  Y  V  E  D  P  L  P  I  T  V  T  L  D  N  L  E  G  F  E  I  V  S  E  S  I  D  G  D  I

2901  CTGCGTTCAACCAGGATCTAGTAACAATCTAAAAATACAGCTTAAAGGAAAAACATTAGGCAGCAACAACATAACGGTGCATGCAGAAAGCGCATCATCT  3000
        C  V  Q  P  G  S  S  N  N  L  K  I  Q  L  K  G  K  T  L  G  S  N  N  I  T  V  H  A  E  S  A  S  S
```

Supplementary Fig. S6

```
3001  AGTGATGTTTGTGGAAGTGACAGTATTTCAGACGCTGTTGCCAAAGATTCTATAAGAAAGCCTGTCATTGTTGAAGCCGAAGGTTGGCCAGTAGAAGAAA  3100
       S  D  V  C  G  S  D  S  I  S  D  A  V  A  K  D  S  I  R  K  P  V  I  V  E  A  E  G  W  P  V  E  E

3101  TCGAAAGTGTGCTCTTTTGTCCTAAAGATGAAGAAAATGATGTTTTCAAAAAGACTTTGACATTAAACGAACCAGAAGATGTCGTACCAGATTCATCGCG  3200
      I  E  S  V  L  F  C  P  K  D  E  E  N  D  V  F  K  K  T  L  T  L  N  E  P  E  D  V  V  P  D  S  S  R

3201  AGCATATTTAGATTGTTCAGGTAATGTATTGGGTAAATGCTTAGACAACTTGGAAAATCTCGTATCATTACCAACTGGATGTGGAGAGCAGAATATGGTG  3300
        A  Y  L  D  C  S  G  N  V  L  G  K  C  L  D  N  L  E  N  L  V  S  L  P  T  G  C  G  E  Q  N  M  V
                                                                Thioester site
3301  AAGTTTGCTCCAAACGTTGTTGCTATGAAAATGTTAATTAATACTAATCAGCTTAGTGATAAAACAAAGGATAGAATTGTAAGGAACCTTAACACAGGGT  3400
       K  F  A  P  N  V  V  A  M  K  M  L  I  N  T  N  Q  L  S  D  K  T  K  D  R  I  V  R  N  L  N  T  G

3401  CCCAGCGACAGATGAAGTTCAAACATCCCGATGGTTCCTTTAGTGCGTTTGGCACAAGAGATAAACAAGGAAGTATGTTCCTCACTGCATTTGTTTTACG  3500
      S  Q  R  Q  M  K  F  K  H  P  D  G  S  F  S  A  F  G  T  R  D  K  Q  G  S  M  F  L  T  A  F  V  L  R

3501  CTACTTCTCTGAAGCATCACAATATATCACAATTGACAATGCCACGATTTCTGAAATGCAAAAATGGATCACATCAAAACAAAAGGATGATGGTTGTTTC  3600
        Y  F  S  E  A  S  Q  Y  I  T  I  D  N  A  T  I  S  E  M  Q  K  W  I  T  S  K  Q  K  D  D  G  C  F

3601  CCAGACGTTGGAAAGATCATCGACAGAGGCTTTCAAGGAGCCATTGAAAAAGACAAATCTGATGGCACTATAACTGCGTATGTTCTTGCTTCTCTTCGGA  3700
       P  D  V  G  K  I  I  D  R  G  F  Q  G  A  I  E  K  D  K  S  D  G  T  I  T  A  Y  V  L  A  S  L  R

3701  TATCAAATTATCAAAACCAGACAGTTCTTGATAAAGCTCTGTCCTGCCTTAGCAACAGCCAGGATTCTAGTCTGTATGCAACCTTCTTGTATGCTTATGC  3800
      I  S  N  Y  Q  N  Q  T  V  L  D  K  A  L  S  C  L  S  N  S  Q  D  S  S  L  Y  A  T  F  L  Y  A  Y  A

3801  AGAAGCTCTCTCTGATAAAAAAGATTCGGCAAAAGAACGTATCGAGTCAGCCAAAGACAGAGCTATAACAAAAGGAAAAGAGGTGTATTATCACGACGTG  3900
        E  A  L  S  D  K  K  D  S  A  K  E  R  I  E  S  A  K  D  R  A  I  T  K  G  K  E  V  Y  Y  H  D  V

3901  AATGCGACTAAATCTCAGGATATAGAGACTTCTTCGTATGCAATTTTATCTATCTTGAACTCTGATGGATCAGCTTCAGATGCTTTGCCGATAGTCCAAT  4000
       N  A  T  K  S  Q  D  I  E  T  S  S  Y  A  I  L  S  I  L  N  S  D  G  S  A  S  D  A  L  P  I  V  Q

4001  ATCTTACAAAAAATATGAATCCTCGAGGTGGTTTCTTCTCCACACAGGACACATGTGTAGGTTTGGAAGCTCTCGGTCAGTTTTCGGAAATGACATTCAA  4100
      Y  L  T  K  N  M  N  P  R  G  G  F  F  S  T  Q  D  T  C  V  G  L  E  A  L  G  Q  F  S  E  M  T  F  K

4101  AGATGAAGTCGATATTACTATCACTGCTACTGGAGATATAGAAAAGAACATTGAAATTACAGAAGACGAAAAATTGCTTGTGAAAAGATACAAAGTAAAC  4200
        D  E  V  D  I  T  I  T  A  T  G  D  I  E  K  N  I  E  I  T  E  D  E  K  L  L  V  K  R  Y  K  V  N

4201  GAAGTACCATCTGAAATAAATATAGAAGCTACGGGTTCAGGATGCGCTGTCATTCAACAAATATTCAGGTACAATTCGAAGACTTCCCCTGAAAAGAGGA  4300
       E  V  P  S  E  I  N  I  E  A  T  G  S  G  C  A  V  I  Q  Q  I  F  R  Y  N  S  K  T  S  P  E  K  R

4301  GTTTCCATCTTGAGGCTCTTGGAAAATGTTCCGATGATGATTGCAAGAAAGCCACAATTTCCTTGTCTTTCAGTTATATACCAGAGGGCAAAAAGACTGG  4400
      S  F  H  L  E  A  L  G  K  C  S  D  D  D  C  K  K  A  T  I  S  L  S  F  S  Y  I  P  E  G  K  K  T  G

4401  AATGAGTGTTTTAGAAGTAAAGATGGTAACCGGCATGTCTCCAGTAAAAGATAGTCTTGAAAAGCTCCTTGGGGATAAAAGATCAAAAGTAATGCGGTAC  4500
        M  S  V  L  E  V  K  M  V  T  G  M  S  P  V  K  D  S  L  E  K  L  L  G  D  K  R  S  K  V  M  R  Y

4501  GATGTCGAAGATAATACTGTTGTAATGTACTTCAATCAGGTAGAAAATGAAGAGATGAATATCGCCTTTGATGTTGAGAAGGTTGTAGAAGTAGAAAATA  4600
       D  V  E  D  N  T  V  V  M  Y  F  N  Q  V  E  N  E  E  M  N  I  A  F  D  V  E  K  V  V  E  V  E  N

4601  CTCAACCGGGCATCGTAAAGCTCTATGATTACTACAATAGGGATGTTTCCAGCAGTACCAACTATTCTTTTTGTGGAAAATCCGAGTCATGTTCGACAGA  4700
      T  Q  P  G  I  V  K  L  Y  D  Y  Y  N  R  D  V  S  S  S  T  N  Y  S  F  C  G  K  S  E  S  C  S  T  E

4701  GCCATAAATTTTCTTTTTTTGTCAGCGTTGAAATATAAGATTTGCGTTAATAGAGAAATTGATTTTGCAATCGAGAGACTGAATTTTTGTAAACGTGTTG  4800
        P  *

4801  GACTGGATTGGTGCTCTATCAGTATTTTTTTGTTTGTTTGTTTTTTCTCTTACAACGCACACAGATGACACCCTTAAAAAAAATATTTGTTCTTATTGGA  4900
4901  AATTTGTATCTCCTGAATTTTCTTGCATGAAAAATGAGTCTAATAAAAAAATATATATATTTTGTAAAAAAAAAA  4975
```

HaadiTEP/CD109-2

```
   1  ATTATGTTGAAACAAATGTTTTCAATTTAACACACTATTATGTTTTCTATTTAAATTCGCACACCCATTTTAAACGTTTACTTGTTAAACACAATGCTTC  100
                                                                                                    M  L

 101  TACTTTCATCTTCCGTGTTTATTATCCTTAGCTACATTGTTTTAAAAACTTATTGTCAAAGAACCTCTCCTATATATACTGTCACTGCTCCAGCAAAATT  200
      L  L  S  S  S  V  F  I  I  L  S  Y  I  V  L  K  T  Y  C  Q  R  T  S  P  I  Y  T  V  T  A  P  A  K  L

 201  GCGACCAGATATAGTATATCATGTAAGTGTTACTCTTCATGACTCCCCTGCAGATGTGGATTTTAATGTTCAAATAATTGGCATATCAGAAGATTTGTCC  300
        R  P  D  I  V  Y  H  V  S  V  T  L  H  D  S  P  A  D  V  D  F  N  V  Q  I  I  G  I  S  E  D  L  S

 301  CCTGTTAATGTTGTGAAGGATGTTCACTTAGAATCAAAGCAAACGAAAGTTGTTGATTTTGAGCTTTTTGGATGGAAACCAGGTAACTATACTCTTGAAG  400
       P  V  N  V  V  K  D  V  H  L  E  S  K  Q  T  K  V  V  D  F  E  L  F  G  W  K  P  G  N  Y  T  L  E

 401  TTATTGGACAAGGTGGCCTAACTGTCAGAAATTCAACACAATTAACTTTCGAGCATAAAAGTCATTCTGTTTTTATACAAACAGACAGACCTGTATACAA  500
      V  I  G  Q  G  G  L  T  V  R  N  S  T  Q  L  T  F  E  H  K  S  H  S  V  F  I  Q  T  D  R  P  V  Y  K

 501  GCCTGGACAAATTGTTCAATTTCGGGTTATTGTTGTTGATCCATATTTAGTTCCAAGTCCTAATGCATTATTTATTGATTTACTTGTAAAAGATGGACAT  600
        P  G  Q  I  V  Q  F  R  V  I  V  V  D  P  Y  L  V  P  S  P  N  A  L  F  I  D  L  L  V  K  D  G  H

 601  GGTAATAGTATTCAGGAGTGGAAGAATATTCCTACAAGAACAGGTCTTGCCTCAGCTGAATTTACTCTTGCAAATCAACCTGTGTTAGGTGAATGGGAAG  700
       G  N  S  I  Q  E  W  K  N  I  P  T  R  T  G  L  A  S  A  E  F  T  L  A  N  Q  P  V  L  G  E  W  E

 701  TGCATGTTGATGTTGAAGGCCAAAAATTCAAGAAACCATTTACTGTAGCTGAATTTATTTTACCAACTTATGAAGTTGAGATCGATTTGCCAGTGTATGT  800
      V  H  V  D  V  E  G  Q  K  F  K  K  P  F  T  V  A  E  F  I  L  P  T  Y  E  V  E  I  D  L  P  V  Y  V

 801  TACTTATAATAAATCTGATGTAGTTGCTGTTATCAAAGCTATGCATTCATATGGTAAGCCTGTTAAAGGTGAAATTACACTGACAGTTGCTCCACGGACA  900
        T  Y  N  K  S  D  V  V  A  V  I  K  A  M  H  S  Y  G  K  P  V  K  G  E  I  T  L  T  V  A  P  R  T

 901  AGATATAACAAGTTAACTGTTCGGCCTTATGAATCATTTCAGACAAAAGCTAAGATCGATGGAGTGGTTGAAATTTACTTGAATTTACTGAATGATTTGT  1000
       R  Y  N  K  L  T  V  R  P  Y  E  S  F  Q  T  K  A  K  I  D  G  V  V  E  I  Y  L  N  L  L  N  D  L
```

Supplementary Fig. S6 (continued)

```
1001  CGCTGAGAACTGATTTTTTTAGAAGAGAAATAGAATTTTTTGTACTTGTTGAAGAAGAAGAAACTGGCCATAAATATAACTCTACTAACACTATGTGGAT  1100
      S  L  R  T  D  F  F  R  R  E  I  E  F  F  V  L  V  E  E  E  E  T  G  H  K  Y  N  S  T  N  T  M  W  I

1101  TTATGATAAAGAAATTAAATTAGAATTAATAAGAACCTCTGAGACATTTAAACCAGGTTTAAAATACACCGCATTTTTAAAAGTTGCCTATCAAGATGAC  1200
        Y  D  K  E  I  K  L  E  L  I  R  T  S  E  T  F  K  P  G  L  K  Y  T  A  F  L  K  V  A  Y  Q  D  D

1201  ACTCCAGTCTCTGATAGTAGAGGCCAGTTGGAACTAAAATATGGCTACTCAGTGCGGGAAGATGAATGGAAAACAGAACTATATACAGTTCCTCGAAATG  1300
       T  P  V  S  D  S  R  G  Q  L  E  L  K  Y  G  Y  S  V  R  E  D  E  W  K  T  E  L  Y  T  V  P  R  N

1301  GTCTTATAAAATTGGAATTTTTACCACCTAATCAAGATGGTGTAAACTTTTTAAACATGAGGGCTGTATATCATGGCCATATATATTACTTAGATCGTAC  1400
      G  L  I  K  L  E  F  L  P  P  N  Q  D  G  V  N  F  L  N  M  R  A  V  Y  H  G  H  I  Y  Y  L  D  R  T

1401  TGATGCTGCTCAGTCTCCCAGCGGAAATTATATCCAGGCAATTCTTGTTACACAGAATCCAAAAGTGTTAGAAAATATTGAAATAGAAGTCAATGCAACT  1500
        D  A  A  Q  S  P  S  G  N  Y  I  Q  A  I  L  V  T  Q  N  P  K  V  L  E  N  I  E  I  E  V  N  A  T

1501  GAAGAACTTAATCACCTTGTATATAAGGTTATAGGCAGGGGAAATATTGAGATTGGTAGAACAATTCCTGTACCAAATCAAAAAGAATATCGATTTAGCT  1600
       E  E  L  N  H  L  V  Y  K  V  I  G  R  G  N  I  E  I  G  R  T  I  P  V  P  N  Q  K  E  Y  R  F  S

1601  TTAGAGCTCCAAGTTCATTAGCTCCTCAAGCTAGAGTTCTTGTTTATTATGTCAGAGCCACAAATAATGAAATTGTAGCTGATTCAGTTAGTTTTGATGT  1700
      F  R  A  P  S  S  L  A  P  Q  A  R  V  L  V  Y  Y  V  R  A  T  N  N  E  I  V  A  D  S  V  S  F  D  V

1701  TGAAGGGCTCTTTAGAACATCTGTTACTGTTAGTTCAAATGTTAAAGAAGTTCAACCTGGTCGCCAAGTTAATCTTCGACTGCAAACAACTCCTAACTCT  1800
        E  G  L  F  R  T  S  V  T  V  S  S  N  V  K  E  V  Q  P  G  R  Q  V  N  L  R  L  Q  T  T  P  N  S

1801  CTTGTGGGTGTTTTGGGAGTTGATCAAGGTATTTTAAAATTGAAATCAGGAAACGACATTACTTTACCTGAGGTGATTGAGGATCTTGAAACATATGATG  1900
       L  V  G  V  L  G  V  D  Q  G  I  L  K  L  K  S  G  N  D  I  T  L  P  E  V  I  E  D  L  E  T  Y  D

1901  GTGGACAACGTACTAAGTATAGACCTCCCTGGTTTCGACGCCGAAGAAGATCCCTTTCATGGCCTGGATCTAAATCTGCCGGCTTGTTATTTGAGGATTC  2000
      G  G  Q  R  T  K  Y  R  P  P  W  F  R  R  R  R  R  S  L  S  W  P  G  S  K  S  A  G  L  L  F  E  D  S
                                          β-α cleavage site
2001  TGGTTTTGTAATCATGACAAATGCTTTTCTTTTAAATACTGGAAATGAAGTATCAACAGAAAATGTTATCCGCATTGATGAAAACAGCAATAACGCTCCT  2100
        G  F  V  I  M  T  N  A  F  L  L  N  T  G  N  E  V  S  T  E  N  V  I  R  I  D  E  N  S  N  N  A  P

2101  ATTCAGCCACCATCTGAATTACCTGAAGCTATCGTTCCTGAAGGTCGCTTAGTAATCAGGAAAATGTATCCAGAAACATGGTTATGGGTTAATACTACCA  2200
       I  Q  P  P  S  E  L  P  E  A  I  V  P  E  G  R  L  V  I  R  K  M  Y  P  E  T  W  L  W  V  N  T  T

2201  CAGGGAATGATGGAGCAGCTTCTATTTCTCATGCTGTACCAGATTCCATTACTTCTTATACTATCAGTGCTTTTGCTATTCATCCTGTTGATGGTTTAGG  2300
      T  G  N  D  G  A  A  S  I  S  H  A  V  P  D  S  I  T  S  Y  T  I  S  A  F  A  I  H  P  V  D  G  L  G

2301  TATTGCAACATCCAATTCTCAAATAACTACATACCGTCCTTTCTTTATTACTATGAGTTTACCATATTATGTATTAATGGGTGAAGACTTGGCAATTCAA  2400
        I  A  T  S  N  S  Q  I  T  T  Y  R  P  F  F  I  T  M  S  L  P  Y  Y  V  L  M  G  E  D  L  A  I  Q

2401  GTTGTAGTTTTTAACTATAATGATAAACCAATACAGGCTGAAATCACCATGGAAAATCGAAAAAGAGAGTTCGATTTCACTGCTGCTGGTCAAGAAAGTG  2500
       V  V  V  F  N  Y  N  D  K  P  I  Q  A  E  I  T  M  E  N  R  K  R  E  F  D  F  T  A  A  G  Q  E  S

2501  TGTATTCACCAGATCAAAATCAAAGAACTAAAATAGTTCATATTCCTCCATCAGATGGAGTTCCAGTCTCATTTTTGATTATACCAAAAAAAGTTGGTTA  2600
      V  Y  S  P  D  Q  N  Q  R  T  K  I  V  H  I  P  P  S  D  G  V  P  V  S  F  L  I  I  P  K  K  V  G  Y

2601  CATTGAGATAAGAGTATCTGCCTCTACGTCTGTCGCTGGAGATTCTTTAACAAAGCGACTTCTTGTTCAGCCTGAAGGTTCAACCCAATATTTCAATAAG  2700
        I  E  I  R  V  S  A  S  T  S  V  A  G  D  S  L  T  K  R  L  L  V  Q  P  E  G  S  T  Q  Y  F  N  K

2701  GCATTTTTGATCGATACACGTAATCCAAGTTCTCCTAGCAAAATGAACATTTCTACAATCATTCCTAAAAATGCCATAAGAGACTCTGGAAAAATCATTG  2800
       A  F  L  I  D  T  R  N  P  S  S  P  S  K  M  N  I  S  T  I  I  P  K  N  A  I  R  D  S  G  K  I  I

2801  TATCTGCTGCAGCTGATTTAATGGGACCAAGTATAAAAAATCTAGATAAACTTCTATATATGCCAAATGGATGTGGCGAGCAGAACTTAGTCACCATTGT  2900
      V  S  A  A  A  D  L  M  G  P  S  I  K  N  L  D  K  L  L  Y  M  P  N  G  C  G  E  Q  N  L  V  T  I  V
                                                                           Thioester site
2901  ACCCAGAGTTATTGCACTGGAGTACTTAGCAAGATCTAACAGATTAACTGAAAATATTAGAGCCAAAGCCATTGCAAATCTTCGCAAAGGTTACCAAAGA  3000
        P  R  V  I  A  L  E  Y  L  A  R  S  N  R  L  T  E  N  I  R  A  K  A  I  A  N  L  R  K  G  Y  Q  R

3001  CAACTGACCTATAAAAGAGATGATGGTTCTTTTAGTACATTTGGTGAAAGGGATCGTAGTGGCAGTACATGGTTAACTGCATATGCCATTAAGTCACTTT  3100
       Q  L  T  Y  K  R  D  D  G  S  F  S  T  F  G  E  R  D  R  S  G  S  T  W  L  T  A  Y  A  I  K  S  L

3101  CTCAAGCCAACAAATACATTTACATTGATCCTGACATTTTAGATAAGGGAATACAATGGATCATGTCCAAGCAGAGTTCAGATGGTTCTTTTGAAGAACC  3200
      S  Q  A  N  K  Y  I  Y  I  D  P  D  I  L  D  K  G  I  Q  W  I  M  S  K  Q  S  S  D  G  S  F  E  E  P

3201  AGGAGAGGTTCACCACAAAGCTTTACAAGGTGGTGAACATGGAGCAGCTTTAACTGCTTTTGTCTTAGTATCTCTTTATGAAGCTAAAGCACAAAATAAA  3300
        G  E  V  H  H  K  A  L  Q  G  G  E  H  G  A  A  L  T  A  F  V  L  V  S  L  Y  E  A  K  A  Q  N  K
           Catalytic histidine
3301  TATGGCAATGAATTGAGTCAAGCTCAAAGATATGTTGAAAGGGAATTGGCATCCAGTTCAAATCCTTATGTTGTCTCTATTCTTTGTTACACTCTTCATT  3400
       Y  G  N  E  L  S  Q  A  Q  R  Y  V  E  R  E  L  A  S  S  S  N  P  Y  V  V  S  I  L  C  Y  T  L  H

3401  TATTAAACAGTGCTTCTAGAGATCGAGCTTTCCAGATGTTATTGGATTTAGCTGAAAGAAAAGATGATGTTGTATACTGGGATAATAAAGAGAATCAGGT  3500
      L  L  N  S  A  S  R  D  R  A  F  Q  M  L  L  D  L  A  E  R  K  D  D  V  V  Y  W  D  N  K  E  N  Q  V

3501  TAATACCACAGATAAGCAATCAGACTACTGGTTTTTAGCACCATCTATTGATATTGAAACAGCAGCATATGCCATTCGCACATATGCTCTTAGACTGGAT  3600
        N  T  T  D  K  Q  S  D  Y  W  F  L  A  P  S  I  D  I  E  T  A  A  Y  A  I  R  T  Y  A  L  R  L  D

3601  CCATCTGGAGCATTACCTGTTCTTACTTGGCTTATTACTAAACAAAACAGAAAAGGAGGATTTTCATCTACTCAGGATACAGTAGTTGCTCTACATGCTA  3700
       P  S  G  A  L  P  V  L  T  W  L  I  T  K  Q  N  R  K  G  G  F  S  S  T  Q  D  T  V  V  A  L  H  A

3701  TTAGTGAAATAGCTCCCTTTATTTCACCTCCTGTATCTAATATAAATGTTAAATTTATGTACCCAGATGGACAACAAGATATGCAGGTTACAAGTTCTAG  3800
      I  S  E  I  A  P  F  I  S  P  P  V  S  N  I  N  V  K  F  M  Y  P  D  G  Q  Q  D  M  Q  V  T  S  S  R

3801  ACCATTAGATGTTCATGAAATTGAAATACCATCGGATGTTCCATATGTAGAAGTTGAAACTTCTGGTTCTGGTGTTGCTGTAGTTCAAGTCTCTTGGTCT  3900
        P  L  D  V  H  E  I  E  I  P  S  D  V  P  Y  V  E  V  E  T  S  G  S  G  V  A  V  V  Q  V  S  W  S

3901  TTTAATCTGGCTGTATCTGGAGAAGCTCCACAGTTCTTTTTGAATGCTTTGTTAGATAAAACATCTACTGCAAGTTATTTACAACTGAGTATTTGTACAC  4000
       F  N  L  A  V  S  G  E  A  P  Q  F  F  L  N  A  L  L  D  K  T  S  T  A  S  Y  L  Q  L  S  I  C  T
```

Supplementary Fig. S6 (continued)

```
4001   ACCAAAGAGAAAGAAGGAATGATACTAGTAATATGGCTGTCATGGAAGTTGGCTTACCTTCTGGCTATGTGGCTGATGTTGATGCCTTACCTAGTGTTCT  4100
       H  Q  R  E  R  R  N  D  T  S  N  M  A  V  M  E  V  G  L  P  S  G  Y  V  A  D  V  D  A  L  P  S  V  L

4101   GCAAATTCCTAAGGTGAAGAGAGTTGAAACACAATTGCAAGATACTGGTGTTGTAATCTATTTTGATAGATTGGATAGAGAAGAATCATGTGTTACTGTT  4200
         Q  I  P  K  V  K  R  V  E  T  Q  L  Q  D  T  G  V  V  I  Y  F  D  R  L  D  R  E  E  S  C  V  T  V

4201   CCAGCGCATCGCATTCATAAAGTAGCTCACCAACGACGAGCTCCTGTCAAAGTGTACGATTTTTATAGTCAAGCTAAAAGTGCTCGAATGTTTTACCGCC  4300
        P  A  H  R  I  H  K  V  A  H  Q  R  R  A  P  V  K  V  Y  D  F  Y  S  Q  A  K  S  A  R  M  F  Y  R

4301   CTCATAAAACTGTACTGTGTGATATATGTGATGATGAAGATTGTGGGAATGGTTGTTTCACAGAAACAATAACAGAAGAATCAGGGTTCAAAAGTTCTGG  4400
       P  H  K  T  V  L  C  D  I  C  D  D  E  D  C  G  N  G  C  F  T  E  T  I  T  E  E  S  G  F  K  S  S  G

4401   AGAAATAGTTCTTATAAAATTTTATTTGGTAGTTATAGCATCTGTTCTGATTTGCTTAGTGTAGAAATTAATATATCTAAAAAAATTAAAAATGAAATTC  4500
         E  I  V  L  I  K  F  Y  L  V  V  I  A  S  V  L  I  C  L  V  *

4501   CAGGACATTCATTTTTGCTGAATGTTAACATTTTAGATGTTCTGTTATTGACTAGGTCAATAAATTAGTTCTAAGATTATATATTTATAAAAAAATAAAT  4600
4601   AAATCTAATTAAAAAGTAAGTTAAATAATTAAAATATTTTACCATGTAATTATTATTTTACAATAGAATGT  4671
```

Supplementary Fig. S6 (continued)

EpspA2M

```
    1  AGAAACTTTTGTGTGGTAAATAAATCGCACTTTTTGTATTACAATAATTAAGAATGAAATATGTGGGATTAGCCTGGCGATTGACTCAATTCCAAAACAA  100
  101  AAAAATTACAACCGGATATCAAATAAATTGAAGTTTCAATCTACAAATTGTGAAACTGGATGTATGAAATCATTTGGAAATTGGATGTACTATCGAATGA  200
  201  CACCGTAATTCATACTTGGGATCAATATAAAGGTCATTTAAGGAGAAGTAGGGTAAACACTATTTTCTAGGAAAAATATCGCTTGCGTCATTAAAAGACG  300
  301  CATTAGACATTTTAATCAATTGACGTTTCTCGACACTCCAATGTATAAAATGAATAATACTTTCTGATTTCGATTTTATAAGCGAGTTACTTGAAGTTTG  400
  401  AATTAAGTAGCATTACTTAATTTAAAATAATTTGTTTTAAAATGAAATTATTTAAAACTTTGTGCTTTGTTTTGCATTTTCTTTCGATTGGTTTTATAGC  500
                                                 M  K  L  F  K  T  L  C  F  V  L  H  F  L  S  I  G  F  I  A

  501  AAACAGTATTCGCGCTCAAGATAGTGTGGAAAATCAGAAGAAGGGTTTTATTTTAACTGCTCCTTCAATTTTTAAGACTAACGTAGACGAAAATGTTTGC  600
         N  S  I  R  A  Q  D  S  V  E  N  Q  K  K  G  F  I  L  T  A  P  S  I  F  K  T  N  V  D  E  N  V  C

  601  ATTTCATTTCAAGAAATTCACGGCGATGGGGAAATTCGAGTTTTTATCCAAGGTTCAACTACTAACACAACTCTGGCCATAGCCATTCAAAAAATAAAAA  700
        I  S  F  Q  E  I  H  G  D  G  E  I  R  V  F  I  Q  G  S  T  T  N  T  T  L  A  I  A  I  Q  K  I  K

  701  ATGGGGCAAATGTATGTTTTTCTCTGCACGTGAAGGTCACAACGGATCTGGAAGGTGTCATCCTCGTCGATGGCTGGTTTCAAGACGAAAATTATAGATT  800
       N  G  A  N  V  C  F  S  L  H  V  K  V  T  T  D  L  E  G  V  I  L  V  D  G  W  F  Q  D  E  N  Y  R  F

  801  TAACTCTACTGAAAAAATCATTATAAAAAAACAGCCAGAAATTACTATAATTCAATCTGATAAGCCATTATACAAACCAGGGCAAAAAGTGCTTTTTCGA  900
         N  S  T  E  K  I  I  I  K  K  Q  P  E  I  T  I  I  Q  S  D  K  P  L  Y  K  P  G  Q  K  V  L  F  R

  901  ATTTTGCGTGTTGATTCATATTTTTTACCTAGCCGCAGCAAAATAAATTCAGTTTGGATTGAAAATCCTTCTCAAGTGAGGGTGGCTCAATGGCTATCGC  1000
        I  L  R  V  D  S  Y  F  L  P  S  R  S  K  I  N  S  V  W  I  E  N  P  S  Q  V  R  V  A  Q  W  L  S

 1001  TACCTGCTTCAAATGGAATTGCCGATCTCAACTTTCAATTGTCAGTAGAACCGCAATTAGGAGTTTGGAATATATTTGTTGAAGACGAGTTTAAAAACAG  1100
       L  P  A  S  N  G  I  A  D  L  N  F  Q  L  S  V  E  P  Q  L  G  V  W  N  I  F  V  E  D  E  F  K  N  R

 1101  GGTTAATAAAACATTTAAAGTTGAGGAATATGTATTGCCAAAATTTTCTGTTGAAATTACTTCCTCTTCCGTCCTTCTTTCCAATTTTGAAAACTATAGT  1200
         V  N  K  T  F  K  V  E  E  Y  V  L  P  K  F  S  V  E  I  T  S  S  S  V  L  L  S  N  F  E  N  Y  S

 1201  TGGAAAGTTTGCGCACATTACACTTATGGCGAGCCAGTTGCTGGAAAACTTAAAGCCACAGTTGCTTTGGAAAATGTTCGGCATTATTTTGGTATGCGGG  1300
        W  K  V  C  A  H  Y  T  Y  G  E  P  V  A  G  K  L  K  A  T  V  A  L  E  N  V  R  H  Y  F  G  M  R

 1301  CACGTATTCTTCCAAAAAACTTTACTTTAGAAAGCGAGTTTTTCAAGTGCAAATTAATCAACATTTCAAAATCAGATATTGTTGCGGATTCAACCGGTTC  1400
       A  R  I  L  P  K  N  F  T  L  E  S  E  F  F  K  C  K  L  I  N  I  S  K  S  D  I  V  A  D  S  T  G  S

 1401  TATTTATGGTGATCAAATAAAACTCTATGCAGATATCATTGAACACGGAACTGATGTGATTATGAGTGCTTCTAAACTTGTATCGGTGAAAAGAACTGCT  1500
         I  Y  G  D  Q  I  K  L  Y  A  D  I  I  E  H  G  T  D  V  I  M  S  A  S  K  L  V  S  V  K  R  T  A

 1501  GTGAATCTGTCTTTGAAATCGCGAGATTACTTTAAACCAGGTTTACAATACAAAGGAAAAATTCAAGCAGAATACCACGACGGAAGTCCAGCAGAAAATA  1600
        V  N  L  S  L  K  S  R  D  Y  F  K  P  G  L  Q  Y  K  G  K  I  Q  A  E  Y  H  D  G  S  P  A  E  N

 1601  AAAATTTAAACATTGAATTGCAAGCTTATAATGCTGAAAACACTAAATTAGCAACCGAGATAATTAATGTAAAAACAGATCTTTTAGGAATGTCTTTTTT  1700
       K  N  L  N  I  E  L  Q  A  Y  N  A  E  N  T  K  L  A  T  E  I  I  N  V  K  T  D  L  L  G  M  S  F  F

 1701  CGATTTGTTTTCATTGCCTCAAAATTCGGAAAGAATTACATTGAAAGCTACAGCCACCGATTACAAAACCCAAACATCTATGGGCGAAGTGAGATATAGC  1800
         D  L  F  S  L  P  Q  N  S  E  R  I  T  L  K  A  T  A  T  D  Y  K  T  Q  T  S  M  G  E  V  R  Y  S

 1801  AATTTTATTCAACCTCCGTTCTATGCAACAATCAAAAGTTGGTATTCGCCGAGTAACAGCTTTGTACAAATTGACCAACAGCTCCAATCATTACAATGTA  1900
        N  F  I  Q  P  P  F  Y  A  T  I  K  S  W  Y  S  P  S  N  S  F  V  Q  I  D  Q  Q  L  Q  S  L  Q  C

 1901  ACGAGTTAAACTTAATAACTGTCAATTTTGCAAATATAAATCAATCGGAAATTATTGTTTTCTACCAGTTGATATCTCGAGGGAATATCTTAAGCCAAAA  2000
       N  E  L  N  L  I  T  V  N  F  A  N  I  N  Q  S  E  I  I  V  F  Y  Q  L  I  S  R  G  N  I  L  S  Q  N

 2001  TCAATTTCAATTAAGTTTTAATAACTTTACACAAACATATCAAATACCAATAATGGTTGAACCAAAAATGGCTCCAAAACTTCAGATTTTGGTTTATTAT  2100
         Q  F  Q  L  S  F  N  N  F  T  Q  T  Y  Q  I  P  I  M  V  E  P  K  M  A  P  K  L  Q  I  L  V  Y  Y

 2101  ATTCGCTATGATGGGGAAATTGTGAGTGATAGCGAAACAATAAACGTTAACCATTGCTACAGTAATGAGGTTAGAATGGATTTTAAAGATGAGAAAATTT  2200
        I  R  Y  D  G  E  I  V  S  D  S  E  T  I  N  V  N  H  C  Y  S  N  E  V  R  M  D  F  K  D  E  K  I

 2201  TGCCTGGATCTTTGTCTTCTCTTTCTTTATCTGCAGAAACCGGAAGCTTATGCGCTGTTAGTGTTACAGACAAGAGCATTGAGCTACTCTCAGGTCATGT  2300
       L  P  G  S  L  S  S  L  S  L  S  A  E  T  G  S  L  C  A  V  S  V  T  D  K  S  I  E  L  L  S  G  H  V

 2301  GTTTGATGGACATAAGGTTTTTAGCATGATTGACGATAAACTGCCACAAATAAATAATAATAATCAATGTCCACAGAACAATCATCCACGTTATTATATA  2400
         F  D  G  H  K  V  F  S  M  I  D  D  K  L  P  Q  I  N  N  N  N  Q  C  P  Q  N  N  H  P  R  Y  Y  I

 2401  TTTAATAGTGCGGTGACGAATTCGAAAGCGGCGTTTGATGATATTGGAATGCTGGTGATAACAAATTTGGAATTAAAATTACATCCGTGCAAAGCTCGAG  2500
        F  N  S  A  V  T  N  S  K  A  A  F  D  D  I  G  M  L  V  I  T  N  L  E  L  K  L  H  P  C  K  A  R

 2501  CTGTCGTTGCCCGGCCGGGGGAAAGCCGTCAGGCGGGGAGGCGTCGTGGGGGGAGCTCTCGAAATAGAGAAGGTGTATTATTGTCCATGGCATCTCCTTT  2600
       A  V  V  A  R  P  G  E  S  R  Q  A  G  R  R  R  G  G  S  S  R  N  R  E  G  V  L  L  S  M  A  S  P  F
                                                Bait region
 2601  TCCTATGGCTTCAGAAAGTGACACTAAAGTCACTCACGAAATATCAATGGTAACATCGCCAGAAATCTCTTTACCGGCTGTTGAAATAAGAGATTTTTTT  2700
         P  M  A  S  E  S  D  T  K  V  T  H  E  I  S  M  V  T  S  P  E  I  S  L  P  A  V  E  I  R  D  F  F

 2701  CCTGAAACTTGGATTTGGGAATTGCAAACAATTCAAAATAATAGTAAAGTTGTAATTGATCGAGAAATACCTCACTCAATTACTGAATGGTCAGGAAACA  2800
        P  E  T  W  I  W  E  L  Q  T  I  Q  N  N  S  K  V  V  I  D  R  E  I  P  H  S  I  T  E  W  S  G  N

 2801  TGTTTTGCATGTCAGAAAAGAGTGGATTAGGAGTTTCCCCAAGGACTTCGATCAAATCCTTTCAACCCTTCTTCCTTTCCTACACCATGCCGTATTCCGT  2900
       M  F  C  M  S  E  K  S  G  L  G  V  S  P  R  T  S  I  K  S  F  Q  P  F  F  L  S  Y  T  M  P  Y  S  V

 2901  CAAAAACGGCGAAACCATTCCAATAACTGTCTCAATATTTAATTATCTTTCGGGTTGCTTTCCGGTTGTTGTAAAGTTAGAAGAAAGTCAATTTTTTGTG  3000
         K  N  G  E  T  I  P  I  T  V  S  I  F  N  Y  L  S  G  C  F  P  V  V  V  K  L  E  E  S  Q  F  F  V

 3001  ACGGATGTTAATTCAACGGAAGTCAAATTGTGTCTTTGCGGCAGTAAATCTTATTCCCTCAAATTCCTTATTCAACCTTTGAAAATTGGCAACTTAAACG  3100
        T  D  V  N  S  T  E  V  K  L  C  L  C  G  S  K  S  Y  S  L  K  F  L  I  Q  P  L  K  I  G  N  L  N

 3101  TGACAGTCCGCGCTCATTCGATTACCGAATTTGATATCTTATTATGTGAAATCGAGAGCACTCTTCAAAACGTTGTGGCTTTTGATGCTATTACTAAATC  3200
       V  T  V  R  A  H  S  I  T  E  F  D  I  L  L  C  E  I  E  S  T  L  Q  N  V  V  A  F  D  A  I  T  K  S
```

Supplementary Fig. S7

```
3201  ACTTCTTGTAAAAGCTCCAGGTTTTCCTCAAGAAACTACTCAATCAAATTGGATCTGCACCAATGATTTTGAAAATGGCAGCAAGACCCTTGAGTACTCA  3300
        L  L  V  K  A  P  G  F  P  Q  E  T  T  Q  S  N  W  I  C  T  N  D  F  E  N  G  S  K  T  L  E  Y  S

3301  TTAGAATTGCCTGAAGACGTCATCGAAGGTTCAGCGAGAGCTTTTATTAGTGTTACAGGAGATTTATTAGGGCCAACAATCAGCGGGTTAGATCATCTGG  3400
       L  E  L  P  E  D  V  I  E  G  S  A  R  A  F  I  S  V  T  G  D  L  L  G  P  T  I  S  G  L  D  H  L

3401  TGAAAATGCCCACGGGATGCGGAGAACAAAACATGGTATTATTCGTTCCGAATATTTACGTGTTGCAATATCTTGAAGTGACGCACCAATTGACCATCAA  3500
      V  K  M  P  T  G  C  G  E  Q  N  M  V  L  F  V  P  N  I  Y  V  L  Q  Y  L  E  V  T  H  Q  L  T  I  N
                    Thioester site
3501  TATTAAATTGAAATCCATATCTTACATGGAATTAGGTTATCAACGAGAACTGAACTACAAACGAGATGACGGATCTTACAGTGCGTTTGGAAAGTCTGAT  3600
        I  K  L  K  S  I  S  Y  M  E  L  G  Y  Q  R  E  L  N  Y  K  R  D  D  G  S  Y  S  A  F  G  K  S  D

3601  GCTGAAGGAAGTATATGGCTGACTTCATTTGTGATTAAATCTTTTGCACAAACGAAATCAATTATTTATATCGACCAACACGAAATTGATGATGGAATCA  3700
       A  E  G  S  I  W  L  T  S  F  V  I  K  S  F  A  Q  T  K  S  I  I  Y  I  D  Q  H  E  I  D  D  G  I

3701  AGTTCATTGTGAAAAACCAATTAGAAGATGGCTGCTTTAAATCAGTTGGAAGCGTTATTCATAAAGAACTTCAAGGAGGAACTGGTCACGGACAATCGTT  3800
      K  F  I  V  K  N  Q  L  E  D  G  C  F  K  S  V  G  S  V  I  H  K  E  L  Q  G  G  T  G  H  G  Q  S  L
                                                         Catalytic histidine
3801  GACTCTGTTTATATCAATTTCTTTGCTTGAAGCTGGGTTGTCGCCTTCAGGAAAAATTATTAAAAAAGCATTTAAATGCTTTGAAAACCACTCTACCATG  3900
        T  L  F  I  S  I  S  L  L  E  A  G  L  S  P  S  G  K  I  I  K  K  A  F  K  C  F  E  N  H  S  T  M

3901  AATGTTTATGAACTTTCCATGAAAGCTTACGCAGCTGCACTCGCAAACAAAACTGTTTTAGCAAAATCGATAATTAAAACATTAAAACAACTTGCAATTC  4000
       N  V  Y  E  L  S  M  K  A  Y  A  A  A  L  A  N  K  T  V  L  A  K  S  I  I  K  T  L  K  Q  L  A  I

4001  AGAAAGATAATTTGATGTATTGGAGCAACAGTGGATCTGGTTCAGACGCTCTAGATATCGAGACAACGTCTTATGTTTTGTTGGCATTGGCTACAACTGA  4100
      Q  K  D  N  L  M  Y  W  S  N  S  G  S  G  S  D  A  L  D  I  E  T  T  S  Y  V  L  L  A  L  A  T  T  D

4101  CTCGGGAGAAGATATGAAGATGGCCATCGATATTGTGAGATGGTTGACTAAGCAACGCAATGCCTACGGGGGTTTTAAATCAACACAGGACACGGTGTTG  4200
        S  G  E  D  M  K  M  A  I  D  I  V  R  W  L  T  K  Q  R  N  A  Y  G  G  F  K  S  T  Q  D  T  V  L

4201  GCACTTCAGGCGTTGGTCAAATTTATTAGCCAAACGCCTAAAATAAATGCTAATTTAAGCCTTGCGCTTGAAGCAAATGATTTCACAGATGAAATGTTTG  4300
       A  L  Q  A  L  V  K  F  I  S  Q  T  P  K  I  N  A  N  L  S  L  A  L  E  A  N  D  F  T  D  E  M  F

4301  TCACTGAGGAAAATCGACTTTTGATGCAAACAAGAGACATTTCTGTTTTGCCTAATATGCTTGATGTGCAAATATCAGGAAAAGGGTGTTCCTTGATTCA  4400
      V  T  E  E  N  R  L  L  M  Q  T  R  D  I  S  V  L  P  N  M  L  D  V  Q  I  S  G  K  G  C  S  L  I  Q

4401  GGTTACTTTAAGATATAATATTCCAAGTGCCCAACTGTTACCGGCATTTACTTTAAACATATCTACTTCATCAAATGAAATATCGAACGATATAAATCAC  4500
        V  T  L  R  Y  N  I  P  S  A  Q  L  L  P  A  F  T  L  N  I  S  T  S  S  N  E  I  S  N  D  I  N  H

4501  CGACAAACTATTAAAATCTGCACAAAGTATGATGGAGCAGATAACAAATCAAACATGGCTGTTGTGGAGATACAAATGATAACAGGCTTTGAAGCACTCC  4600
       R  Q  T  I  K  I  C  T  K  Y  D  G  A  D  N  K  S  N  M  A  V  V  E  I  Q  M  I  T  G  F  E  A  L

4601  ATTCACATTTAGAAAAAATTGAAAACCTCCAATATTGATTTAAAGAGATGGGAAAGTGAAAATGACGGAACAGTGGAACTTTACTTTGATCAATTGGAAAG  4700
      H  S  H  L  E  K  L  K  T  S  N  I  D  L  K  R  W  E  S  E  N  D  G  T  V  E  L  Y  F  D  Q  L  E  S

4701  TGAATTAAAGTGCTTTGATGTGATCGTGGAGGAGAAATTGAAAGTGCAAAATCGTAAACCAGCGTTGATCAAAGTGTACGACTACTACAAAACAGATGAG  4800
        E  L  K  C  F  D  V  I  V  E  E  K  L  K  V  Q  N  R  K  P  A  L  I  K  V  Y  D  Y  Y  K  T  D  E

4801  TCGGTATCCAAGGAATATTTTCTTGAATCTTGAACAACCTCCTTTGGTCTATTAACGCAGATTTAATTAAAATAATGAAAGAATTTCTATGGTAGTTTAA  4900
       S  V  S  K  E  Y  F  L  E  S  *
4901  TCAATATTTTTTGGAATTCTTAATGCAAAACTTTGCTTTTATTTTTCATTACATTTTGTTTATACAATTTTTTTTTTGCATTTAATTTAACACTAGTTTG  5000
5001  TTAATTTACGTTGCCACAATTCGTTTCAGATAATAAAAAATATCAAGCGGCTAATTTAAGCTACTATAAAAAGTGATATACAGTAATTTTGATTGTGAAA  5100
5101  TGTACTTTTAGTTTAAAAATAGTGCTGCAAATTACTCCATTAAATTTCAATTCAAGTTAAAA  5162
```

EpspiTEP/CD109-1

```
   1  GCAGATTGACGCGCAGCTCAAGACGCAATAAAAAGTTGATGTTCTTATTTTGAAAAAGAAAGAAAGACGAAAAAAAAAAAAAAAGAAATAATAATTTTTT  100
 101  TTTCTCCTAAGTGCATACAATGCGTTAGTAGGGAAACCCAATGAGTGAGTTTGTCCTGACGTCACACGCATAGGCCGGTATATATAATAATATAATTTTC  200
 201  GTCTCGTCACTCAGTCGTCACTTTTCGCTCGTGATAAGGAAACGTATTGAGGTCGGTGGACGTCACCGGACCAACCACTAACATTCGCTGTACGCAATCA  300
 301  TTTTGACGGCATATATCTAGTTTAGCAGCGCTCAACATTTTCAGTACTTGCGGAGAAATTGCGCGGGGAACACGATCTTCGGAGTTTCAGAGACGACTCG  400
 401  GACTAATTAGGATAATTAAAATGAACACGCCGAATGGTCTGTTGGCTGTTCTTCTCTTGGGAATATGCGCTGCTACGTCTCATGCTGAAAGGACGTACAC  500
                           M  N  T  P  N  G  L  L  A  V  L  L  L  G  I  C  A  A  T  S  H  A  E  R  T  Y  T

 501  CGTGACAGCGCCGAAAAATGTTCGCGTGGGAACACCGTATCAAGTTGTCGTGTCCATCCATAACTCACCTGAGGACGTTGAAATATTTGCAAATTTGAGT  600
        V  T  A  P  K  N  V  R  V  G  T  P  Y  Q  V  V  V  S  I  H  N  S  P  E  D  V  E  I  F  A  N  L  S

 601  TGCAGCTCGGATGACAACTCGAAGCCGCAATTAGTCACGGGTTCAATTACAGCGCCCAACGAAATCACCAAACTGCTCACGTTACCTATACCTGATTGGT  700
       C  S  S  D  D  N  S  K  P  Q  L  V  T  G  S  I  T  A  P  N  E  I  T  K  L  L  T  L  P  I  P  D  W

 701  GGAAACCTGGCAATTGCGAACTCACTGTTACCGGAGATAAAGGAATTGTTTTTAAAAGATCGGCCAGCTTAGGTTTCAATTCCAAAACTTCGTCAGTTTT  800
      W  K  P  G  N  C  E  L  T  V  T  G  D  K  G  I  V  F  K  R  S  A  S  L  G  F  N  S  K  T  S  S  V  F

 801  TATTCAGACTGATAAAGCCATTTATCAACCTGGGCAATTAGTGCAATTTCGCGTGCTTGTTGTGGATCCCAATCTTAAACCTTATGCAAGCGATGAGTTA  900
        I  Q  T  D  K  A  I  Y  Q  P  G  Q  L  V  Q  F  R  V  L  V  V  D  P  N  L  K  P  Y  A  S  D  E  L

 901  ACTGTTTTTATCACGGATGCTCAAGGAAACCGAATCAAACAATGGAATAACGCAACTCTGAAATCCGGAATCTTTTCCGGTGAGTTGCAGTTATCTGATC  1000
       T  V  F  I  T  D  A  Q  G  N  R  I  K  Q  W  N  N  A  T  L  K  S  G  I  F  S  G  E  L  Q  L  S  D

1001  AGCCTGTCCTTGGAGACTGGAGCATCAACGCCCAACTCAACGACGCGAGGGCTAACAAACAAGTTAGCGTTGCCGAATATGTTTTACCTAAATTTGAGGT  1100
      Q  P  V  L  G  D  W  S  I  N  A  Q  L  N  D  A  R  A  N  K  Q  V  S  V  A  E  Y  V  L  P  K  F  E  V

1101  CACCGTTCGCCCACCAACACTGGCCATATTTAATGACACCGAATTAATTGTGGGCGTCGAAGCCAAATACACGTACGGAAAACCGGTTAAAGGCAAATTG  1200
        T  V  R  P  P  T  L  A  I  F  N  D  T  E  L  I  V  G  V  E  A  K  Y  T  Y  G  K  P  V  K  G  K  L

1201  GTTTTGAACACCACGGAAAGTTATTGCCGAAGTCCTTATTCCAATTATTGCGGGTCAACTCCGGTCGTAGTCCGAACCAATATTGATGGAATAGCAAATG  1300
       V  L  N  T  T  E  S  Y  C  R  S  P  Y  S  N  Y  C  G  S  T  P  V  V  V  R  T  N  I  D  G  I  A  N
```

Supplementary Fig. S7 (continued)

```
1301  TGAAGATTCCATTATCTCAATTTAACTTTCCGGACTATTATCGCAATAGTCTGGCAAATTTGGATTTTTTGGCGGTCGTCACTGAAGATTTAACCGGAAG  1400
      V  K  I  P  L  S  Q  F  N  F  P  D  Y  Y  R  N  S  L  A  N  L  D  F  L  A  V  V  T  E  D  L  T  G  R

1401  AATGATGAACGCATCCGCCGCCGGAAATATTTATTCCAAAAGAGAGAAAATCGACGTTGTTCAATCGTCTAACTCTTTTAAACCCGGTCTCCCTCACACG  1500
        M  M  N  A  S  A  A  G  N  I  Y  S  K  R  E  K  I  D  V  V  Q  S  S  N  S  F  K  P  G  L  P  H  T

1501  TACAAAATTAAACTGCAATTGCAAGATGGAACTCCCGTGACCAAGGCTGACTCTTTGTTGACTGTGAAAACCTCGAGTTCTCATGGCAAACCGGAAGTAG  1600
       Y  K  I  K  L  Q  L  Q  D  G  T  P  V  T  K  A  D  S  L  L  T  V  K  T  S  S  S  H  G  K  P  E  V

1601  TTACCAATTACACAGTTCCAGCTAGCGGAATTGTTTCCGTGGAAGCGTTTCCTGATGAAGATGCAGAGTTTCTTCGTCTGAATGCTGATTACAAAGATGT  1700
      V  T  N  Y  T  V  P  A  S  G  I  V  S  V  E  A  F  P  D  E  D  A  E  F  L  R  L  N  A  D  Y  K  D  V

1701  TTCCGGAAGTGCGTACGCGAACAAAGCCCAGTCCATAAGTAGCAAGTACCTTCAATTGAGTCTTCACAATGAAAACGAAATCGAACCTAAAGTTGGCGAC  1800
        S  G  S  A  Y  A  N  K  A  Q  S  I  S  S  K  Y  L  Q  L  S  L  H  N  E  N  E  I  E  P  K  V  G  D

1801  GTAGTTCAATTGGATGTGAATGGAACCTTTTACATTTCTCGTTTGGATTACGAAGTTGTAGCAAGAGGCAAAATTATTACTTCCGGTTCGTTAAAATTCG  1900
       V  V  Q  L  D  V  N  G  T  F  Y  I  S  R  L  D  Y  E  V  V  A  R  G  K  I  I  T  S  G  S  L  K  F

1901  ACAAGGATGCCAAATCGCACTCTTTCCCGTTGAACATAACTCAGGATATGGCGCCACGAGCCCGAGTGGTGGCTTATTACGTTTCTTCTTGCGGAGAAGT  2000
      D  K  D  A  K  S  H  S  F  P  L  N  I  T  Q  D  M  A  P  R  A  R  V  V  A  Y  Y  V  S  S  C  G  E  V

2001  TGTGGCTGACAGTTTGGATTTCACTGTTAATGGAGTTTTTCAAACACCGGTGGGTTTGCATACAAGCGAGAACAAAACAAAGCCTGGAGCTCCCATCGAA  2100
        V  A  D  S  L  D  F  T  V  N  G  V  F  Q  T  P  V  G  L  H  T  S  E  N  K  T  K  P  G  A  P  I  E

2101  GTGACAGTTGACACCTTGCCAAACTCAACTGTGGGTCTGTTGGCTATTGATCAGAGTGTTTTGCTATTGAAATCGGGCAACGACCTTAATCGGAACGAAA  2200
       V  T  V  D  T  L  P  N  S  T  V  G  L  L  A  I  D  Q  S  V  L  L  L  K  S  G  N  D  L  N  R  N  E

2201  TTATTAATGATTTGGGAGATTATGAAAGCGGTTGGAGACCTAGTGCTTTCGACAGGAAGAAGCGATCGATCTGGCGTCCACCTGGAGACACGGTTACTCA  2300
      I  I  N  D  L  G  D  Y  E  S  G  W  R  P  S  A  F  D  R  K  K  R  S  I  W  R  P  P  G  D  T  V  T  Q
                                                        β-α cleavage site
2301  ACTGTTTGACACAGCCGGTTTGGTTTTTTTCTCCAATGGACTTTTCCAAAAGCAACCCGACTTCAACTACGGTCCATATCCCGTGCGCTTTAATAGTTTT  2400
        L  F  D  T  A  G  L  V  F  F  S  N  G  L  F  Q  K  Q  P  D  F  N  Y  G  P  Y  P  V  R  F  N  S  F

2401  GGTGGTGGAGCTGGTGCAAGTTTTGCTGAATCAGATTCGGTGCAATTTCAAGCTTCGACAGCGCCAGCCTCGTTCAGCGGCGCGGGCGCTAACAATGACA  2500
       G  G  G  A  G  A  S  F  A  E  S  D  S  V  Q  F  Q  A  S  T  A  P  A  S  F  S  G  A  G  A  N  N  D

2501  AACCAAAACTACGTCAATCATTTCCTGAAACGTGGATTTGGACTATGCCAATCGCAGGTCCTGATGGAAAAGCTGTTTTAAAAACAACGGTGCCGGATAC  2600
      K  P  K  L  R  Q  S  F  P  E  T  W  I  W  T  M  P  I  A  G  P  D  G  K  A  V  L  K  T  T  V  P  D  T

2601  TATCACTTCCTGGCAGTTATCCGCTTTTGCCATGGATGATGAAAATGGTTTAGGAATGGCTGATGGTCCATCAAAGGTTGAGGTGTTCCGTCCATTTTTT  2700
        I  T  S  W  Q  L  S  A  F  A  M  D  D  E  N  G  L  G  M  A  D  G  P  S  K  V  E  V  F  R  P  F  F

2701  GTTACTCTTAATTTGCCGTATTCGGTTGTTCGAGGAGAGTCTGTTGCTGTTCAAGCTTTAGTTTTCAATTACATGAAGGAGGATGTTGAGGCGGAAGTGA  2800
       V  T  L  N  L  P  Y  S  V  V  R  G  E  S  V  A  V  Q  A  L  V  F  N  Y  M  K  E  D  V  E  A  E  V

2801  CTTTGGAGAATCTGAACGACCAATTTGAATTGACCGGTTTATTGAACCGAGTGGATGTTGACCATAATGCTACCTCCGAGAAAAAGACCGTCAAAGTTAA  2900
      T  L  E  N  L  N  D  Q  F  E  L  T  G  L  L  N  R  V  D  V  D  H  N  A  T  S  E  K  K  T  V  K  V  K

2901  AGCTGGCGATGGTTCCTCCGTTTCTTTCCTCATCACTCCCAAAGTGGTTGGGCCAATCGACCTCACCGTTTCGGCCGTTTCTTCAAAAGCCGGAGATGCC  3000
        A  G  D  G  S  S  V  S  F  L  I  T  P  K  V  V  G  P  I  D  L  T  V  S  A  V  S  S  K  A  G  D  A

3001  TTGAATAAAAAGCTGCTCGTTAAGGCTGAAGGTTCTCCCCAATACTTTAACAAAGCCGTTTTAGTGGATTTACGAAACACCAGCAATTTTAAATCCCAAG  3100
       L  N  K  K  L  L  V  K  A  E  G  S  P  Q  Y  F  N  K  A  V  L  V  D  L  R  N  T  S  N  F  K  S  Q

3101  TAGAAGTCAATATCCCTCCTTTTGCAGTTAAAGATTCCGAACACGTGGAAGTATCCGCCATCGGTGACATCATGGGACCTACTGTCAATAATTTGGACAA  3200
      V  E  V  N  I  P  P  F  A  V  K  D  S  E  H  V  E  V  S  A  I  G  D  I  M  G  P  T  V  N  N  L  D  K

3201  ATTGATTAAAATGCCTTATGGTTGCGGCGAACAAAATATGATCAATTTTGTTCCAAACATTGCCGTTTCCGACTACCTCAACACCACCAATCAATTTAGC  3300
        L  I  K  M  P  Y  G  C  G  E  Q  N  M  I  N  F  V  P  N  I  A  V  S  D  Y  L  N  T  T  N  Q  F  S
                         Thioester site
3301  GATAAATTGAGAACAAAAGCCATCAAATTCATGGAAGCTGGTTACCAACGCGAATTGACTTATAAAAGACCAGACGGTTCGTTTAGCGCTTTCGGAACAA  3400
       D  K  L  R  T  K  A  I  K  F  M  E  A  G  Y  Q  R  E  L  T  Y  K  R  P  D  G  S  F  S  A  F  G  T

3401  GTGATAAGAACGGAAGTACTTGGTTAACTGCTTTTGTTGTGCGATCTTTCAAGCAAGCCAAGCCATACATCTCGGTAGATGACAACGTTATCGATGCTTC  3500
      S  D  K  N  G  S  T  W  L  T  A  F  V  V  R  S  F  K  Q  A  K  P  Y  I  S  V  D  D  N  V  I  D  A  S

3501  TTTAAGATACTTAAAATCCACTCAAAAGGAAAATGGTAGCTTTGTTGAAAACGGAGAAGTGCATAACAAACGACTTCAGGGCGGCGCTGCTGGTGGTTTG  3600
        L  R  Y  L  K  S  T  Q  K  E  N  G  S  F  V  E  N  G  E  V  H  N  K  R  L  Q  G  G  A  A  G  G  L

3601  TCGTTGACAGCTTACGTCACGTTAGCATTTTTGGAAAACTCTGATAAAACGGAATACAGCAACGTTACAAAACGCGCGGTTAAGTTTCTCGAAGATGAAG  3700
       S  L  T  A  Y  V  T  L  A  F  L  E  N  S  D  K  T  E  Y  S  N  V  T  K  R  A  V  K  F  L  E  D  E

3701  TGGACACTGTGGAAGATCCATATGAACTGGCTATTGTTAGCTATGCTCTTCATAAAGCCGGAAGTCCCGCTAAGGACGCCGCGTTTCAACGTTTCCTTAA  3800
      V  D  T  V  E  D  P  Y  E  L  A  I  V  S  Y  A  L  H  K  A  G  S  P  A  K  D  A  A  F  Q  R  F  L  K

3801  AAAGGCTGAAAAGAAAGGCGACAATGTGTTTTGGTCCAAACCGTTAGAGAAACCACCAACCAATAGCTCTTATTTCTATTACCCACCTCCATCTGTGGAT  3900
        K  A  E  K  K  G  D  N  V  F  W  S  K  P  L  E  K  P  P  T  N  S  S  Y  F  Y  Y  P  P  P  S  V  D

3901  ATCGAAATGACCGCCTATGCTCTACTCACTCACTTGGAAAGAAACTTGATCACCGAAGCAGTTCCAATCATGAAATGGTTGATTACCCAAAGAAACGAAA  4000
       I  E  M  T  A  Y  A  L  L  T  H  L  E  R  N  L  I  T  E  A  V  P  I  M  K  W  L  I  T  Q  R  N  E

4001  ATGGCGGATTCTCTTCAACTCAAGATACTGTGGTTGGAATTCAATCTTTGGCTGGCATAGCTTCTCACATCACTTCACCAGATGGTGCTAAAATGGAATT  4100
      N  G  G  F  S  S  T  Q  D  T  V  V  G  I  Q  S  L  A  G  I  A  S  H  I  T  S  P  D  G  A  K  M  E  L

4101  AGACTTCGAATACGACGGTCGCCACAAAAAGTTAACATTGGACAAAGGCAATGCTATGGTTCTTCAACGAGAAGAGTTGCCTTCGGAAACTCGAGAAGTT  4200
        D  F  E  Y  D  G  R  H  K  K  L  T  L  D  K  G  N  A  M  V  L  Q  R  E  E  L  P  S  E  T  R  E  V

4201  AACGTCGGAGCTGAAGGAAAAGGTTTTGGAATTGTCCAAGTCACGTGGTCTTACAATTTGGACAATAAAACTCAATCGCCAGTGTTTGAAATTTCTCCAA  4300
       N  V  G  A  E  G  K  G  F  G  I  V  Q  V  T  W  S  Y  N  L  D  N  K  T  Q  S  P  V  F  E  I  S  P
```

Supplementary Fig. S7 (continued)

```
4301  AAGTCAAACAGATTGATAAAGACTCATTTGACCTCACTGTTTGCGCAATTTACGAAGACGTGGATAAAAGCAGCAATATGGCGGTCATCGAAATGAATTT  4400
      K  V  K  Q  I  D  K  D  S  F  D  L  T  V  C  A  I  Y  E  D  V  D  K  S  S  N  M  A  V  I  E  M  N  L

4401  GCCATCTGGTTACGTAGTTGAAGGGGAATCCCTACCTCAAGTGGGTCAACACGACCGTCTGAAGAGAGTTGAAACTGTGGACGGTGGAACCAAAGTTCAA  4500
        P  S  G  Y  V  V  E  G  E  S  L  P  Q  V  G  Q  H  D  R  L  K  R  V  E  T  V  D  G  G  T  K  V  Q

4501  TTGTATTACGACGAGATGAATGATGTCGAGGTTTGTCCTCATATAACTGCCTACCGAACATTCCCAGTGGCCAATGTGAAGCCGGCCGCCGTGTCGGTCT  4600
       L  Y  Y  D  E  M  N  D  V  E  V  C  P  H  I  T  A  Y  R  T  F  P  V  A  N  V  K  P  A  A  V  S  V

4601  ACGATTATTACGATAACGACCAACGAGCGGAAACGTTTTACAATGCTCCAAGCAGCGATTTATGCGACATTTGCACCGGAAGTGAATGCAGCTCAAAGTG  4700
      Y  D  Y  Y  D  N  D  Q  R  A  E  T  F  Y  N  A  P  S  S  D  L  C  D  I  C  T  G  S  E  C  S  S  K  C

4701  CAAAAAATCTTAAATCTATTTTTTTCTGATCAAAAACTTTTTTTAGACACCCATCACACAGTTATCTTATAAAGTCTATTACTGTATTATATTGATAGTT  4800
        K  K  S  *

4801  TAACATATCAGTGTTAAAAATGTGCTCGATTTATTCCAATCTTAAAACAATTGGCTTTTCGTTAATTCATTATTACACATTTCAATGGATTTTAAACGTT  4900
4901  TATTCCTTTTAAAAATTGTTTTTATGAGAAAAAAAAATTAAAAAAATACTTAATT  4954
```

EpspiTEP/CD109-2

```
   1  TAAACAAACATTTAAGTTTTTTTGGACTTGAAGAAGTTCGGATTTTGCGGTCATTTTCGCGATTGATCGGGCGGATTTTTTTGTCGTCGAAGGAGGAGTG  100
 101 CTCAAATATTAAGAGGAGGAAGATAATGAATTCTCATTGACATCTGAAATTGATTGTCAGCAACAACAATGGCACACACGTTCTGCAGGCTGCTGTTCGC  200
                                                                          M  A  H  T  F  C  R  L  L  F  A

 201  TTTGGTCGCCGTTTTATTGTCGGTGACTCAGCTGTCCTCGGCCGTCGGCACGTACACGCTCGTGGGACCTAAAGTTTTAAGGCCGGGACTGAGCTATCAA  300
        L  V  A  V  L  L  S  V  T  Q  L  S  S  A  V  G  T  Y  T  L  V  G  P  K  V  L  R  P  G  L  S  Y  Q

 301  GTCAGCGTGTCCATTCACGACACCAGAGAGCCGGTCAAAGTCACCGTGTCCATCACCGGACAAGGTTTGGCCGGGGTCAACTATGGACAGACCACCGAAG  400
       V  S  V  S  I  H  D  T  R  E  P  V  K  V  T  V  S  I  T  G  Q  G  L  A  G  V  N  Y  G  Q  T  T  E

 401  CCCTCATCCAATCGGGAGAAACGCAGATTTTAAAGTTTGAAATTGGTGAATGGGGTCCAGGTAGATACAATATTTCAATTTCCGGAAGTGGAGGTTTACG  500
      A  L  I  Q  S  G  E  T  Q  I  L  K  F  E  I  G  E  W  G  P  G  R  Y  N  I  S  I  S  G  S  G  G  L  R

 501  ATTCTCAAACACTACCGAGATCGAGTACCAGCACAAAAGTTACTCCGTTTTTATTCAGACCGATAAGGCGGTTTATCAACCGAATCAGATGGTTCATTTC  600
        F  S  N  T  T  E  I  E  Y  Q  H  K  S  Y  S  V  F  I  Q  T  D  K  A  V  Y  Q  P  N  Q  M  V  H  F

 601  AGGGCAATTGTGGTGAACCCCATGTTGAGACCGACTGTTTCGGGAGCCATAGAAGTTTATATAACGGATGGACAGGGAAATCGAGTGAAACAATGGAGGA  700
       R  A  I  V  V  N  P  M  L  R  P  T  V  S  G  A  I  E  V  Y  I  T  D  G  Q  G  N  R  V  K  Q  W  R

 701  GAGTTTTCACCACCAAAGGTGTGTTTTCTGGAAGTTTGCAATTGTCTGATCAACCGGTTTTGGGCAATTGGAACATTACTGTTTTCGTATCGGATCAACA  800
      R  V  F  T  T  K  G  V  F  S  G  S  L  Q  L  S  D  Q  P  V  L  G  N  W  N  I  T  V  F  V  S  D  Q  Q

 801  GTATACGAAATCATTTACAGTTATGGAATACGTGTTACCTACATATGAAGTGACAGTGCAACTTCCTCCCTACGCCACATTTAACGATACGTCGGTAGTT  900
        Y  T  K  S  F  T  V  M  E  Y  V  L  P  T  Y  E  V  T  V  Q  L  P  P  Y  A  T  F  N  D  T  S  V  V

 901  GCCACTGTAACTGCTAGATATACGTATGGAAAACCGGTGAAGGGAGAATTGACGCTGGTTATGTCACCCAAAGTTCATTCCCCCGTTATTCAAGCATGGG  1000
       A  T  V  T  A  R  Y  T  Y  G  K  P  V  K  G  E  L  T  L  V  M  S  P  K  V  H  S  P  V  I  Q  A  W

1001  TGATTCCGCCTGTCAGAAAACTCGCCACGATTGATGGCGTTGCCGACGTTAGTTTGGATTTGAGCCACTTTCGCTTTGAAGACGATTTTCAGAGAGATAT  1100
      V  I  P  P  V  R  K  L  A  T  I  D  G  V  A  D  V  S  L  D  L  S  H  F  R  F  E  D  D  F  Q  R  D  I

1101  TCGAGTGGAAGCAATTGTAAGAGAGACTCTGACCGGACAGAGACAAAACTCTTCAGCCATCTTGTCTCTTTATAGTCATCGAGTAAAATTAGATCTGATT  1200
        R  V  E  A  I  V  R  E  T  L  T  G  Q  R  Q  N  S  S  A  I  L  S  L  Y  S  H  R  V  K  L  D  L  I

1201  AAAACCGCCGAGACGTTCAAACCTGGCTTGAAATTCACCGCTTATTTGAAAGTGGCGTTGCAAGACGACATTCCCATACGTGACGATGTAAACAAAGTTA  1300
       K  T  A  E  T  F  K  P  G  L  K  F  T  A  Y  L  K  V  A  L  Q  D  D  I  P  I  R  D  D  V  N  K  V

1301  GAGTGAAATTCGGATACAATTATAACGAATCGAGCCACGAAGTTCGGGAGTATCAGATCCCCCGCAATGGAATTATTCAGTTAGATTTTTATCCTCCGCG  1400
      R  V  K  F  G  Y  N  Y  N  E  S  S  H  E  V  R  E  Y  Q  I  P  R  N  G  I  I  Q  L  D  F  Y  P  P  R

1401  ATCACCTGAAGCTAATGTTTTGATCATTTTGGCTGAATATTTGGACATTAAACAACCTTTTCCAGGAATCGAATCGTCGCTTTCCCTAAGCAATAATTAC  1500
        S  P  E  A  N  V  L  I  I  L  A  E  Y  L  D  I  K  Q  P  F  P  G  I  E  S  S  L  S  L  S  N  N  Y

1501  ATTCAAGCTTTACTATTAACAGACAATCCTAGAGTTGCGGATGAAGTTGAGGTTCAGATCAATACAACCGAAAGAATTATGAGTTTCGTTTTTCAAGTTT  1600
       I  Q  A  L  L  L  T  D  N  P  R  V  A  D  E  V  E  V  Q  I  N  T  T  E  R  I  M  S  F  V  F  Q  V

1601  ACGGTCGAGGAAATCTAGCTCTTGCGCAGACAATTCCGATGAACAATGAAAAATATGTGCGTTTCAAGTTCCGAGTCTCATCTCGAATGGCTCCTAAAGC  1700
      Y  G  R  G  N  L  A  L  A  Q  T  I  P  M  N  N  E  K  Y  V  R  F  K  F  R  V  S  S  R  M  A  P  K  A

1701  CCGATTCCTGGCTTATTACGTGAGGGGAGACGGTGAGGTGGTGGCCGATTCTCTCAATTTCTACGTGGAAGGTGTTTTCCAAACTCCCGTTGCTGTGGGG  1800
        R  F  L  A  Y  Y  V  R  G  D  G  E  V  V  A  D  S  L  N  F  Y  V  E  G  V  F  Q  T  P  V  A  V  G

1801  GTGAGTGCCAATCGCACCGGCCCGGGCACATTGGTAGAGGTCAAAGTCAACACTAAGCCTAACGCTTTTGTGGGAATATTAGGAGTGGATCAAAAAGTTT  1900
       V  S  A  N  R  T  G  P  G  T  L  V  E  V  K  V  N  T  K  P  N  A  F  V  G  I  L  G  V  D  Q  K  V

1901  TGTTATTGAAATCAGGGAACGATATAACAAGAGACGACGTTCTGAAAGAATTGGTATCATACGACGGTGGAGCGGATAAAAAATTAGATGACTTTTACGC  2000
      L  L  L  K  S  G  N  D  I  T  R  D  D  V  L  K  E  L  V  S  Y  D  G  G  A  D  K  K  L  D  D  F  Y  A

2001  CCGATACTATTGGTCTCCTGGTACAATCACAGCAAGCCAGGTTTTCGAGGATGCCGGAGTGGTTATTGTTACAAATGGTAACGTGTTTCAGTACTACCCC  2100
        R  Y  Y  W  S  P  G  T  I  T  A  S  Q  V  F  E  D  A  G  V  V  I  V  T  N  G  N  V  F  Q  Y  Y  P

2101  AGACTTCTTTACCGATCCAATATCGGTCCTGACAGTGATCTTGAATACGCTCCCGACAATTCTCTTGGCAACAGTATTTTTGTTGATTCTGACTCACCTG  2200
       R  L  L  Y  R  S  N  I  G  P  D  S  D  L  E  Y  A  P  D  N  S  L  G  N  S  I  F  V  D  S  D  S  P

2201  ACTCTATCAGATTAAGACAGCATTTCCCAGAAACTTGGCTGTGGAATAACACCGTCGCCCAACAAGATGGAGTGGCCACAATTAGCGCAACTGCACCTGA  2300
      D  S  I  R  L  R  Q  H  F  P  E  T  W  L  W  N  N  T  V  A  Q  Q  D  G  V  A  T  I  S  A  T  A  P  D
```

Supplementary Fig. S7 (continued)

```
2301  TACCATCACTTCCTGGATCGTAAGCGCTTTTGCTATTGACAGTATCACTGGTTTAGGTGTTGTTCGAGCACCGGCGAAGTTAACGATTTTCCGCCCGTTT  2400
        T  I  T  S  W  I  V  S  A  F  A  I  D  S  I  T  G  L  G  V  V  R  A  P  A  K  L  T  I  F  R  P  F

2401  TTCGTGTCTGTTGTTCTACCTTATTCGGTGATTCGAGACGAAGCGGTGGCCATTCAGGTGGTTGTGTTCAATTACATGGCGGAAGATGTTAAAGCAACAG  2500
       F  V  S  V  V  L  P  Y  S  V  I  R  D  E  A  V  A  I  Q  V  V  V  F  N  Y  M  A  E  D  V  K  A  T

2501  TTACTCTAAAAAATCAAAAAGGACAGTTTGAGTTTGCTAATTTTGAGACGGGCGCAAACGAAGTTGTAGATAACTCGCAGCCAAAATACAAAAGAGTTCA  2600
      V  T  L  K  N  Q  K  G  Q  F  E  F  A  N  F  E  T  G  A  N  E  V  V  D  N  S  Q  P  K  Y  K  R  V  Q

2601  AGTGAAAGCCGGAGATGGAACATCCATTTCTTTTATGATTATTCCTAAACTAATAGGTTCAATCGACATCGATGTTACGGCTCAAAGTGATCGTGCCGGC  2700
        V  K  A  G  D  G  T  S  I  S  F  M  I  I  P  K  L  I  G  S  I  D  I  D  V  T  A  Q  S  D  R  A  G

2701  GATCGAGTTCTCAAAAAACTCCTCGTCAAGCCTGAAGGAACGCAATTGAACTTCAATAAGGCTGTTTTAGTAGATTTAAGGTCAACAAATTCATTTACAA  2800
       D  R  V  L  K  K  L  L  V  K  P  E  G  T  Q  L  N  F  N  K  A  V  L  V  D  L  R  S  T  N  S  F  T

2801  GCACCGTTAAAGTCGGAATTCCAAACTATGCAGTTGAAGGTTCAGCCCGCGTGGAGATTTCTGTAATCGGTGACATGTTAGGACCGGCCATTAACAACTT  2900
      S  T  V  K  V  G  I  P  N  Y  A  V  E  G  S  A  R  V  E  I  S  V  I  G  D  M  L  G  P  A  I  N  N  L

2901  AGATAGCCTTTTGCGTATGCCATTTGGATGCGGCGAACAAAACATGGTCTATTTTGTGCCCAATATTGTCGTTGCTCGATATTTGGAATTTGTCAACCGC  3000
        D  S  L  L  R  M  P  F  G  C  G  E  Q  N  M  V  Y  F  V  P  N  I  V  V  A  R  Y  L  E  F  V  N  R
                                Thioester site
3001  TTGCAGGACACTATTAAGCAAAAAATACTAAGTCATTTGGAAACCGGTTATCAACGAGAATTAACTTATAAACGCGAAGATGGCTCATTTAGTGCTTTTG  3100
       L  Q  D  T  I  K  Q  K  I  L  S  H  L  E  T  G  Y  Q  R  E  L  T  Y  K  R  E  D  G  S  F  S  A  F

3101  GCAAACAAGATAAAATGGGCAGTACATGGTTGACCGCATTTGTTGCCAAATCATTCCACGATGCTCGCAATTACGTTACAATCGATGAGAACGTGGTGAA  3200
      G  K  Q  D  K  M  G  S  T  W  L  T  A  F  V  A  K  S  F  H  D  A  R  N  Y  V  T  I  D  E  N  V  V  N

3201  CGGTGCTCTCAAATGGTTAGCGATGAAACAGGAAGGAGATGGCTCCTTCCCCGAAGTTGGTTCTGTCAGTCACAAAGCTATGCAGGGTGGTTCTGCCAAC  3300
        G  A  L  K  W  L  A  M  K  Q  E  G  D  G  S  F  P  E  V  G  S  V  S  H  K  A  M  Q  G  G  S  A  N
                                                                    Catalytic histidine
3301  GGTTTAGCGTTGACCGCTTACGTTTTGATCGCTTTTTTGGAGACTGGAGCTCAACAAACGCAGGAATTTAGCAGCGTTGTAGATAAAACCCGACAAAATA  3400
       G  L  A  L  T  A  Y  V  L  I  A  F  L  E  T  G  A  Q  Q  T  Q  E  F  S  S  V  V  D  K  T  R  Q  N

3401  TCGAGCAAGTGATGCTTACCGTGGAAGATTCTTACGCTTTGGCAATTGCCACTTATGCTTTGCATCTTGCCGGAAGCCAAGAGCGAGAAACTGCTTTCCG  3500
      I  E  Q  V  M  L  T  V  E  D  S  Y  A  L  A  I  A  T  Y  A  L  H  L  A  G  S  Q  E  R  E  T  A  F  R

3501  AAAACTAGATCAAAGATCAATTACCAACGATGAACAAAAGTTCTGGACTGGAGTGGTTGATAATAACGACGTGCTCGATCCTCCGGAAAATGATGAACCT  3600
        K  L  D  Q  R  S  I  T  N  D  E  Q  K  F  W  T  G  V  V  D  N  N  D  V  L  D  P  P  E  N  D  E  P

3601  AAAAAATCAAAAGCGGCTGTTAGAGAAGCCAATTCTTATGACGTGGAAATGACGTCATACGCTTTGATGACTTACATTCTAAGAAATGATATCGACAATG  3700
       K  K  S  K  A  A  V  R  E  A  N  S  Y  D  V  E  M  T  S  Y  A  L  M  T  Y  I  L  R  N  D  I  D  N

3701  CTCTGAAAATTGCTAAATGGTTAATTTCTCAAACAAATTCCAATGGTGGATTCACTTCTACTCAAGATACGGTGATTGGCATTCAAGCTTTGGCCAAATT  3800
      A  L  K  I  A  K  W  L  I  S  Q  T  N  S  N  G  G  F  T  S  T  Q  D  T  V  I  G  I  Q  A  L  A  K  L

3801  AGCCCAACAAGTAAATTCACCACGTAGAAGCGTTGACGTTACTTTCAGTTTTAATAACTCGACTAAAACTGTTCCAATCAATAATGAAAACGCGATGATT  3900
        A  Q  Q  V  N  S  P  R  R  S  V  D  V  T  F  S  F  N  N  S  T  K  T  V  P  I  N  N  E  N  A  M  I

3901  CTTCACAAATATTCGATTCTCGATCCCGTTGATGAAGTTCGATTCACAGCTCGAGGAAGTGGTTTTGCCATTGCTCAGGTATCGTATTCATACAACGTGA  4000
       L  H  K  Y  S  I  L  D  P  V  D  E  V  R  F  T  A  R  G  S  G  F  A  I  A  Q  V  S  Y  S  Y  N  V

4001  ATATCACTCAAGAAAAGCCTTCGTTCATCATTAACCCGTTAGTGGATAGAAATTCCAACAGAAATTTATTGGTGATCAACGCATGCACTTCATACACTCG  4100
      N  I  T  Q  E  K  P  S  F  I  I  N  P  L  V  D  R  N  S  N  R  N  L  L  V  I  N  A  C  T  S  Y  T  R

4101  AAATGGCAGCAGCAACATGGCCGTGGTCGAGGTGGACTTGCCCTCTGGTTATACGGTGGACAAGGACTCGCTTCCGGCTTTGCTGAAAATGAAGGACATT  4200
        N  G  S  S  N  M  A  V  V  E  V  D  L  P  S  G  Y  T  V  D  K  D  S  L  P  A  L  L  K  M  K  D  I

4201  AAACGAGTAGATACAAAGGCCGGAGACTCCGGTGTTGTCATCTACTTCGATAAGTTGTCGAAAGGAGAAGAAGTTTGCCCGACGGTGAAAGCCTTCAGAA  4300
       K  R  V  D  T  K  A  G  D  S  G  V  V  I  Y  F  D  K  L  S  K  G  E  E  V  C  P  T  V  K  A  F  R

4301  CCTTTAAAGTGGCCAAACAGAAGCCAACCGCGGTCAAAGCTTACGATTACTACGATCAGAGTCGAAGCGCTCGGTCGTTTTATCAAGCGTTACCAGCCAC  4400
      T  F  K  V  A  K  Q  K  P  T  A  V  K  A  Y  D  Y  Y  D  Q  S  R  S  A  R  S  F  Y  Q  A  L  P  A  T

4401  CCTTTGTGATATTTGCGATGACGACGAATGCGAAAGAGGAAGTTGCCCGAAGAATGCGGGAATGTTCGAGGATTATACGGGCAATTCCGCCGCTAGTTTT  4500
        L  C  D  I  C  D  D  D  E  C  E  R  G  S  C  P  K  N  A  G  M  F  E  D  Y  T  G  N  S  A  A  S  F

4501  ACATTGACTTTGATTAACGTCTTCACTTTCATCATTCCTGTGTGGCTGATATCGAATTTTAGTTGAGGTATTTACGTATTTTTTTTTACCTTTTTTTCGA  4600
       T  L  T  L  I  N  V  F  T  F  I  I  P  V  W  L  I  S  N  F  S  *

4601  GAGTTAAAACATGGCTGGATCATAATCGTACAAATATTTCTTTTTTTTCAAAAATAGTTGTTTGAATTGAAAAAGATATTTTTTGAGAAAAAAACTTAAG  4700
4701  TCTTTTGAGATTAGTAGTTGCGTTACATACGTAATTCTAAAAATATTATTTTTTAGCAATTGTAAATTATTTTGATCGTGAACTTCTACAGCTATATACG  4800
4801  CACCATCAATTATTTTTCATTCAAATTATCCAACATCAACTCGCATACATTTTAACGATGGCAAAACACTTTTAAAATTGACTTCCGATTTTTAATTTGC  4900
4901  CATTCTATCATAAAAATTACAGCTTTAAAATATGTTTTTATTTTTAATATTATATTTTTATTAACCGATATTTCCGCTGTAAAAAACTAGCATTTGATTG  5000
5001  TTATTTTAAAATATAGAGTTTTAATTTACA  5030
```

Supplementary Fig. S7 (continued)

VaspA2M-1

```
   1  GTAGTTGACGCGAGCGTCAGGCTGCTGTCCTCCAACAATCAGTTCAACAAAGACAACCTCCTCAACAGCAATCCATCACTTCATTTAAATCAAAGCAGCG  100
       V  V  D  A  S  V  R  L  L  S  S  N  N  Q  F  N  K  D  N  L  L  N  S  N  P  S  L  H  L  N  Q  S  S

 101  CGCCAATACAACAGAATGATTTTGAGTATTGCTTAGAGAAAAAGAAAGAAAGTGAAAGAAAGAGCAAATCTTCAATTGAAGAAAGTTTCTACTACAGCCT  200
      A  P  I  Q  Q  N  D  F  E  Y  C  L  E  K  K  K  E  S  E  R  K  S  K  S  S  I  E  E  S  F  Y  Y  S  L
                                                                            Bait region
 201  GAATTCTGCAGACTCGATCACTGCGTTCGACATTGCCGGATTGACTGTTATGAGTGACTTGGATTTGGAAACGCGACTGTGTCCAGAAGGCACAACGTCA  300
        N  S  A  D  S  I  T  A  F  D  I  A  G  L  T  V  M  S  D  L  D  L  E  T  R  L  C  P  E  G  T  T  S

 301  GTTAGTGCTGTAGAAGGAAACTCTACGTCTTTTCTACCTGGAGACGGGGGATCAATTCTTGGCATCACAACAAGAATTAGACAAGATTTTCCTGAAACAT  400
       V  S  A  V  E  G  N  S  T  S  F  L  P  G  D  G  G  S  I  L  G  I  T  T  R  I  R  Q  D  F  P  E  T

 401  GGCTATTTGAGTTACAATATATGGGTAAAGAACAAATGTTTCACAAGCATGTGACGGTACCAGATACGATTACAGAGTGGATTGGTTCCAGTTACTGCCT  500
      W  L  F  E  L  Q  Y  M  G  K  E  Q  M  F  H  K  H  V  T  V  P  D  T  I  T  E  W  I  G  S  S  Y  C  L

 501  TTCTCCAACCAAAGGCCTTGGCATGAGTGACGAAGCCACTTTCAATTCTGTAATTTCTTTCTTCATTGATTACACTCCCCCTTACGCAGTAAAGAGAGGA  600
        S  P  T  K  G  L  G  M  S  D  E  A  T  F  N  S  V  I  S  F  F  I  D  Y  T  P  P  Y  A  V  K  R  G

 601  GAGTCATTTGAACTTCCCGTCTACATTTACAGTTACCTCAACTGGACCTCGTCGGTATTTGTGCATCTTGAAAAAAGCACAAAAAATCTGTACCAACTGA  700
       E  S  F  E  L  P  V  Y  I  Y  S  Y  L  N  W  T  S  S  V  F  V  H  L  E  K  S  T  K  N  L  Y  Q  L

 701  CCGATGGGCTTGACTTGAAGATTGCTTCTATTCCACCCAGAGGGAAAAGTGTTGTCACCTTTGAAATCACGCCAACTAAAATTGGTGAAATAGAAATTGG  800
      T  D  G  L  D  L  K  I  A  S  I  P  P  R  G  K  S  V  V  T  F  E  I  T  P  T  K  I  G  E  I  E  I  G

 801  AGTCTATGCAACTGGGAATTTCATTATTTTCGATGGCAACAAAGTGAACCATTCTATCATTACTGATGCAGTAAGAAAAAAGATCTTAGTTAAACCAGAC  900
        V  Y  A  T  G  N  F  I  I  F  D  G  N  K  V  N  H  S  I  I  T  D  A  V  R  K  K  I  L  V  K  P  D

 901  GGATTTCCTCAGGAAAATCACGACGTTCAATTGCTGTGTACAAAAGATAACGAAAAACTCGCAAAAAATCTGAATTGTTCTCCACCAAAGAACATAGTAG  1000
       G  F  P  Q  E  N  H  D  V  Q  L  L  C  T  K  D  N  E  K  L  A  K  N  L  N  C  S  P  P  K  N  I  V

1001  AAGATTCAGCAAGAGCTACCTACGTTGTCACCGGTGTTTTTGGACCAATACTCGATAACCTGGATTCGTTCATTCAGCTTCCAATGGGTTGCGGAGAACA  1100
      E  D  S  A  R  A  T  Y  V  V  T  G  V  F  G  P  I  L  D  N  L  D  S  F  I  Q  L  P  M  G  C  G  E  Q
                                                                                             Thioester site
1101  GATTATGGCAGCAATGGTTCCCAGCCTCTATGTGCTAACGTACCTGGACAGTATTGGAGCAGCCGAACCACAGCTACGGGATAAGGCCTTCGAAACTATA  1200
        I  M  A  A  M  V  P  S  L  Y  V  L  T  Y  L  D  S  I  G  A  A  E  P  Q  L  R  D  K  A  F  E  T  I

1201  CTATTAGGGTATCAGCAGCAACTGACTTACAAGAATCCAGATGGCTCTTACAGTATTTTCAGCGGACTACCAAACCCGAGGGGATCGATTTGGCTAACAG  1300
       L  L  G  Y  Q  Q  Q  L  T  Y  K  N  P  D  G  S  Y  S  I  F  S  G  L  P  N  P  R  G  S  I  W  L  T

1301  CGTTCGTAGTGAAGTCATTCAGTCGAACAAAAACGTACATTTCGATCAACGATACAGTGATTGACAGCAGTAAGAAATTTATCCTGAAAAATCAAGATGA  1400
      A  F  V  V  K  S  F  S  R  T  K  T  Y  I  S  I  N  D  T  V  I  D  S  S  K  K  F  I  L  K  N  Q  D  E

1401  AGATGGCTGCTTTCTGGTAGTTGGGAGAGTCTTCAGCAGCGGTTTGAGGGGTGGTGTCGAGAAAGGTGTTCCAAAGGAGACGCTGACTGCCTACATAATT  1500
        D  G  C  F  L  V  V  G  R  V  F  S  S  G  L  R  G  G  V  E  K  G  V  P  K  E  T  L  T  A  Y  I  I

1501  ATCGCGCTGCTGGAGGCAGGGGAAGATCCATCAAGTCAGCACATGAAGGACGCTCTACAGTGTATCAAGAAGATGCCCAAATACGCCTTTGCGCATGCTC  1600
       I  A  L  L  E  A  G  E  D  P  S  S  Q  H  M  K  D  A  L  Q  C  I  K  K  M  P  K  Y  A  F  A  H  A

1601  TAGCTGCATACGCTTGGTTTTTGGCGTATCAAAAACACACATTGCTCTCGAAAACAAGCAATTCTTCTGAAGCACTGGAACTGAAGAAAACTGGAGAACG  1700
      L  A  A  Y  A  W  F  L  A  Y  Q  K  H  T  L  L  S  K  T  S  N  S  S  E  A  L  E  L  K  K  T  G  E  R

1701  TTTCCTGAACAAACTTCTTAAAACGGCGAAAAAAGAGGGGCAGAAGATGTGGTGGCAACATCCAATTAAGTACCGTGATGCGAGGGTTGAGACGGCAGGA  1800
        F  L  N  K  L  L  K  T  A  K  K  E  G  Q  K  M  W  W  Q  H  P  I  K  Y  R  D  A  R  V  E  T  A  G

1801  TACGCTCTACTAGCATTTCTGACGGCAGATCCGTTTGATTTGAAGACTATCCGTCCGATTGCTCGTTATTTGATATCTCAGAGAAACATTAAGGGAGGAT  1900
       Y  A  L  L  A  F  L  T  A  D  P  F  D  L  K  T  I  R  P  I  A  R  Y  L  I  S  Q  R  N  I  K  G  G

1901  TTTACACAACACAGGATACGGTAGTGGCTTTGGAAGCGCTGACTAAATATTCAATGCAAGCAGACGAAGTTCCGTTAAAGAATATTACAGTAAGATGTGC  2000
      F  Y  T  T  Q  D  T  V  V  A  L  E  A  L  T  K  Y  S  M  Q  A  D  E  V  P  L  K  N  I  T  V  R  C  A

2001  AGCAACTACGTCATTCTTAACTCACGTTATAAGTCCGAAAAATCGGCTGGTCACATACAAAACAGATGTTGAGCCCTTCACTGGAAACGTTGTTGTGCAA  2100
        A  T  T  S  F  L  T  H  V  I  S  P  K  N  R  L  V  T  Y  K  T  D  V  E  P  F  T  G  N  V  V  V  Q

2101  GGAGAAGGATGCGTCGTTGCTCAGTGTTCGGTTAAATACAGCACACCAAAAGCAAGAAACATCAAAGGTTTCGAAGTGAATGTCACTCGAGAATGTATAG  2200
       G  E  G  C  V  V  A  Q  C  S  V  K  Y  S  T  P  K  A  R  N  I  K  G  F  E  V  N  V  T  R  E  C  I

2201  AAGGAGGAATTAAACCAACGGTGCAACTAGACATATGTGTTAGCTACGTATTAGACGATGGAGAATCAAATATGGCAATAGTAGAAGTAAACCTGCTCTC  2300
      E  G  G  I  K  P  T  V  Q  L  D  I  C  V  S  Y  V  L  D  D  G  E  S  N  M  A  I  V  E  V  N  L  L  S

2301  AGGCTACTTAGACAGTCCTTACTCTCTGAACGATCTTTACAGCGGTGGCGTAGTGAGCCGATTTGAACCTGGCGAAGATACCGGCGGCACAGCCAATTTA  2400
        G  Y  L  D  S  P  Y  S  L  N  D  L  Y  S  G  G  V  V  S  R  F  E  P  G  E  D  T  G  G  T  A  N  L

2401  TATTTCGACAGCTTCAATAAAACGGAGACCTGCTTTTCCATCCAACAGTTTCAAATAAACAAAGTATTCAATCTGAAACCAGCATTTGTAAAAGTATATG  2500
       Y  F  D  S  F  N  K  T  E  T  C  F  S  I  Q  Q  F  Q  I  N  K  V  F  N  L  K  P  A  F  V  K  V  Y

2501  ATTACTATCGACCAGAAATCACGGCAAGTGAGAGCTATTCTGCCTGTGAGGATGCCCCGTAAACTGGATGTCTTGATTGAAACACTCTTCTGGAGAAAGC  2600
      D  Y  Y  R  P  E  I  T  A  S  E  S  Y  S  A  C  E  D  A  P  *

2601  CGTATCTTGCTTCATCAGATTACTCTGCGAACAAAAAGTCTGCTACTCCTCCGTTGGCATTAACCTGTACAACATTGTTACATAAATTATTTCTTGTTAT  2700


2701  TTTTGCCGCTGTTATTGTTACATGGCAACAC  2731
```

Supplementary Fig. S8

VaspA2M-2

```
   1  CCTGATACCATTACAGAATGGATTGGTTCTGCATACTGCATCTCACCTGTCAGCGGCTACGGAGAATCGCAACCCAACTCCATCAACGTTTTTGTTCCAT  100
       P  D  T  I  T  E  W  I  G  S  A  Y  C  I  S  P  V  S  G  Y  G  E  S  Q  P  N  S  I  N  V  F  V  P

 101  TCTTCATCGATTATTCCTTACCATATTCAATCAAACGAGGAGAAACTTTCGAGTTCAAAGTGTCCGTGTTCAGCTATGTAGACGAGGCATCACCGATTAA  200
      F  F  I  D  Y  S  L  P  Y  S  I  K  R  G  E  T  F  E  F  K  V  S  V  F  S  Y  V  D  E  A  S  P  I  K

 201  ACTTATATTAGAGCCATCTGTAGATTTTGCTGTGGATGACGAAGAATACGAGAAGGACTTGTGGTTAAATTCTAATGAACATGAAGTAGTAATATTCAAT  300
        L  I  L  E  P  S  V  D  F  A  V  D  D  E  E  Y  E  K  D  L  W  L  N  S  N  E  H  E  V  V  I  F  N

 301  GTGACAGCCACAACTATCGGAAGCATAGAGCTTATCACTTATGCAATGACATCAATTGGTAATGAACCAGAATGGGGATATGGAAATTCTACTATTACAT  400
       V  T  A  T  T  I  G  S  I  E  L  I  T  Y  A  M  T  S  I  G  N  E  P  E  W  G  Y  G  N  S  T  I  T

 401  ACCATAGTGATGCCGTCAAGAAATCTGTGATTGTGGAGCCTGAGGGAATTTCAGAAGAGTACACTCAGTCAGAGTTCATGTGTGTTCAAGAAAGTGATAC  500
      Y  H  S  D  A  V  K  K  S  V  I  V  E  P  E  G  I  S  E  E  Y  T  Q  S  E  F  M  C  V  Q  E  S  D  T

 501  AACGATGGTCCGGCTGTTTGCTTGTGAAGCACCACAAAATAGTGTTGCTGATTCCTCTAGAGCTTCCTTCTCTGTGACTGGAGATCTTATGGGACCAGTA  600
        T  M  V  R  L  F  A  C  E  A  P  Q  N  S  V  A  D  S  S  R  A  S  F  S  V  T  G  D  L  M  G  P  V

 601  TTAAATAACTTGGAATCGCTTATTCGCTTCCCAACAGGATGTGGAGAACAGAACATGGCCAAGTTTGTTCCGAATATTCATGTGTTGTCCTATTTTAAAT  700
       L  N  N  L  E  S  L  I  R  F  P  T  G  C  G  E  Q  N  M  A  K  F  V  P  N  I  H  V  L  S  Y  F  K
                                           Thioester site
 701  CTACAAAATTCAATGACAACACCTTGAAAACAAAGTGCATAGAGCACATACGAGAAGGTTACCAAAGAGAACTGAAGTATAAGCACAGCGACGGCTCCTT  800
      S  T  K  F  N  D  N  T  L  K  T  K  C  I  E  H  I  R  E  G  Y  Q  R  E  L  K  Y  K  H  S  D  G  S  F

 801  TAGTGCGTTCGGTTCGTCCGATGCATCGGGTTCGTTGTGGTTGACGGCATTCGTTTTGAAGTCATTTGGACAAGCTCTGCCCTACATCTTTGTTGATAAG  900
        S  A  F  G  S  S  D  A  S  G  S  L  W  L  T  A  F  V  L  K  S  F  G  Q  A  L  P  Y  I  F  V  D  K

 901  AAGATCATGAATAAGTCGGTTAGATGGATTAAAAGCAAACAACTTGAGAACGGATGTTTTCCGGAAGTGGGTTCTATGTTCAATCAAGGTTTACAGGGTG  1000
       K  I  M  N  K  S  V  R  W  I  K  S  K  Q  L  E  N  G  C  F  P  E  V  G  S  M  F  N  Q  G  L  Q  G

1001  GTTTGCAAGACAACCCTACGTTGCTTACTGCATATGTGCTGATAGCCCTCCTAGAAGCAGATCAAGATGAAGAGGAAGCCCTTCCTTGTCTCAACCCACA  1100
      G  L  Q  D  N  P  T  L  L  T  A  Y  V  L  I  A  L  L  E  A  D  Q  D  E  E  E  A  L  P  C  L  N  P  Q

1101  GCATATACCAAACACACCCTATGCTCTCGCTCTCGCCTCATATGCTTGGCTGCTCTCCTACAAAAGAGAGAAAGACGCTGGAAACTCGTCAGAAGATGGT  1200
        H  I  P  N  T  P  Y  A  L  A  L  A  S  Y  A  W  L  L  S  Y  K  R  E  K  D  A  G  N  S  S  E  D  G

1201  ATTGAGAAACTTTTAGTTGGACAAGAACTTCTGGATAAGTTAATGCAACTAAGTCACGTTAACACAGGCATGGTGTGGTGGCAGCAACAAAACTGCAGCG  1300
       I  E  K  L  L  V  G  Q  E  L  L  D  K  L  M  Q  L  S  H  V  N  T  G  M  V  W  W  Q  Q  Q  N  C  S

1301  AATGTAACACAGTTGACGTAGAGACAGCAGGATATGCCATGATGGCACTGATAATAGGCTCTCCTGAAGAAATGTCAACTATCAGAGGCATAGCTCGATG  1400
      E  C  N  T  V  D  V  E  T  A  G  Y  A  M  M  A  L  I  I  G  S  P  E  E  M  S  T  I  R  G  I  A  R  W

1401  GCTTGTTAAACACAGAAACCCTAACGGAGGATTCTATTCAACGCAGGATACCGTGGTTGCACTAGAAGCATTAACGAAACTAGCAATGCTGTCTGACGAA  1500
        L  V  K  H  R  N  P  N  G  G  F  Y  S  T  Q  D  T  V  V  A  L  E  A  L  T  K  L  A  M  L  S  D  E

1501  ATACCATTAAAACAAGTACAAGTGAAAACTCAGTTGCAACATACCTGGAACATAACTGAAGAAAACAGATTGGTAACTCACTTGAAGGAAGGCATGACAT  1600
       I  P  L  K  Q  V  Q  V  K  T  Q  L  Q  H  T  W  N  I  T  E  E  N  R  L  V  T  H  L  K  E  G  M  T

1601  CGTTTCTGTCTGACGTCAGCATTTCTGGTGAAGGATGTGTTCTAGCACAGTGTTCGGTTCGATATAATGTACTTGAAGCCAGTGAAAGGAGCAGTGGATT  1700
      S  F  L  S  D  V  S  I  S  G  E  G  C  V  L  A  Q  C  S  V  R  Y  N  V  L  E  A  S  E  R  S  S  G  F

1701  CTCACTAGAGATAAAGGGTTTACCGGCAGAGGAAAGCTGTGAAAGCAGAGAGTTACTGGTTTGTGTATCTTTCGAAGGAATTTCCAACATGGCTGTGGTC  1800
        S  L  E  I  K  G  L  P  A  E  E  S  C  E  S  R  E  L  L  V  C  V  S  F  E  G  I  S  N  M  A  V  V

1801  GAGATGCACTTGCCTTCAGGATTTGCTCCCGTCGCAGAAAAACTCGATTATTTAATGAACGACGTGTATTTTGCATTAATGAAATGGGAGTTGGATAAGA  1900
       E  M  H  L  P  S  G  F  A  P  V  A  E  K  L  D  Y  L  M  N  D  V  Y  F  A  L  M  K  W  E  L  D  K

1901  ATAAACTACACTTAT  1915
      N  K  L  H  L
```

VaspA2M-3

```
   1  GTAATCGTACTAGTGAATGGCGGGAATTAAATGAAAAACCTGAATTGGGTTATATTGTTGTGCTGTATGTACCTGTTTGTGCAAGGAAGCCACGGGAAAC  100
                                       M  K  N  L  N  W  V  I  L  L  C  C  M  Y  L  F  V  Q  G  S  H  G  K

 101  GGGGTTACATATTCGTTGCACCAAAACAGTTTCAGCGAGAGAGCAACGAGACCATGTGTATTTTGTTCAATAACTTGGAGGAAGCTGCTGCAAACATCAC  200
      R  G  Y  I  F  V  A  P  K  Q  F  Q  R  E  S  N  E  T  M  C  I  L  F  N  N  L  E  E  A  A  A  N  I  T

 201  AATCACCCTGTATGATGACAAGAATGAAACAAGTTCTACTAAAAACTTCAATATGACAGGATCTTCATCTTGGCATTGCATCAGCGACTTTCAAGTTCCT  300
        I  T  L  Y  D  D  K  N  E  T  S  S  T  K  N  F  N  M  T  G  S  S  S  W  H  C  I  S  D  F  Q  V  P

 301  GCTATTGAATCATACGGGGGTTATCTGAATCTTCGTATATCATTCCCAAGTCTTCCATTGGAGCACGCAGTTGATGAAACAAGAGAAGCCACCATTCTTT  400
       A  I  E  S  Y  G  G  Y  L  N  L  R  I  S  F  P  S  L  P  L  E  H  A  V  D  E  T  R  E  A  T  I  L

 401  TTTCCAGAAACCACGGACCCACTACTTTCCTAGAAACAGATAAAGACAAGTACAAGTATGGGCAAACAGTGAAGTTTAGAATCCTCTCCCTCGACTTTGA  500
      F  S  R  N  H  G  P  T  T  F  L  E  T  D  K  D  K  Y  K  Y  G  Q  T  V  K  F  R  I  L  S  L  D  F  D

 501  CAACAAGCCCTTTACGGATACTATTCCAGAAGTATGGATAGAAGATCCAGGACTTAACCGTGTTGCTCAGTGGAAGGATGTAGCAATGAAGCATGGGCTG  600
        N  K  P  F  T  D  T  I  P  E  V  W  I  E  D  P  G  L  N  R  V  A  Q  W  K  D  V  A  M  K  H  G  L

 601  AACCAACTCAGTTATACACTCTCCAGTGATCCACCAAAGGGACACTGGACGATAAGGGTGAAGGGCATTCATGGAGAAAAAAGTATAAACTTTGAAGTTG  700
       N  Q  L  S  Y  T  L  S  S  D  P  P  K  G  H  W  T  I  R  V  K  G  I  H  G  E  K  S  I  N  F  E  V

 701  CCGAATATGTACTGCCAAAGTTTACCTTGAGCATTAAAGGGCAAAGAGATAGTAATAACATCTACGATACAGACAAGAATTACACCTGGACAATCTGTGC  800
      A  E  Y  V  L  P  K  F  T  L  S  I  K  G  Q  R  D  S  N  N  I  Y  D  T  D  K  N  Y  T  W  T  I  C  A
```

Supplementary Fig. S8 (continued)

```
 801  TAAATATACACATGGTGAGGATGTGACAGGAACTATGGACGTATCATTTACTAAGCGGTCATATAGTTGGTGGCGAAATGATCAAGAAAAGGATAAATAC  900
        K  Y  T  H  G  E  D  V  T  G  T  M  D  V  S  F  T  K  R  S  Y  S  W  W  R  N  D  Q  E  K  D  K  Y

 901  GTTGCAGTCTATAACAAAACGAGGTTTTCAGGCTGTAAAGAAGTGACTTTAACAAAGGAAGAATACGGAGGTGATGGAGGTGACCGAAAACACTTTGGGT  1000
       V  A  V  Y  N  K  T  R  F  S  G  C  K  E  V  T  L  T  K  E  E  Y  G  G  D  G  G  D  R  K  H  F  G

1001  TGGAACTGAAGGCTAGCGTAGAAGAGGATGGGACTGGAGTCACAGTTAACGATTCCATCGATCACTATCATTCGGTGAATTCGGAGCTGGTGATGAATGA  1100
      L  E  L  K  A  S  V  E  E  D  G  T  G  V  T  V  N  D  S  I  D  H  Y  H  S  V  N  S  E  L  V  M  N  E

1101  GAACGATAACCACTTTAAACCTGGACTGTTTTATAGTGGGGAGATACAATTAAGTGATTCGGAAAGGATTGGTCTCGGACAAAGAAAGTTTGAACTCTGT  1200
        N  D  N  H  F  K  P  G  L  F  Y  S  G  E  I  Q  L  S  D  S  E  R  I  G  L  G  Q  R  K  F  E  L  C

1201  GTGGAATATCTCAAGAAAAATTGTGCCAACTTTACCTCAGATGCTGACGGAAAGTTTCGTTATGTAATTCAACCTGTGGATGAGCACGATTTCAACAGTA  1300
       V  E  Y  L  K  K  N  C  A  N  F  T  S  D  A  D  G  K  F  R  Y  V  I  Q  P  V  D  E  H  D  F  N  S

1301  CCAGACAATGGTTGGATATGAAGATAAAGCTGTTTAAAGAGAACGGGTCTGTTTCTCGGAGATCATACTCCAGTGTTCAAGTATGGCACTCCAATGCTTA  1400
      T  R  Q  W  L  D  M  K  I  K  L  F  K  E  N  G  S  V  S  R  R  S  Y  S  S  V  Q  V  W  H  S  N  A  Y

1401  CAAGTTCATTCGGTTCTTAAAGCAGGAAGACGGCTCCAAACTCAAGTGCAACAGAACGTACCAGACCGACCTCTACTACTCACCGCCACCAGAAACAGTA  1500
        K  F  I  R  F  L  K  Q  E  D  G  S  K  L  K  C  N  R  T  Y  Q  T  D  L  Y  Y  S  P  P  P  E  T  V

1501  ACAGCTAGTATAGCGGTAATGTCTTCAGGGCGTATCGTAAGAACTACTTCGATTTTACTTAGTCCAACTGATTCAAAGGCATCCGTGCCCACCGCCGATC  1600
       T  A  S  I  A  V  M  S  S  G  R  I  V  R  T  T  S  I  L  L  S  P  T  D  S  K  A  S  V  P  T  A  D

1601  ACGTAATATCACTAGACAATCAAACCAACATACCAGATGTACTGCACCAGACGTTTCCACTAGAGATCGGAACAGATGTCAGCGAGAAGTTCAACGTTCT  1700
      H  V  I  S  L  D  N  Q  T  N  I  P  D  V  L  H  Q  T  F  P  L  E  I  G  T  D  V  S  E  K  F  N  V  L

1701  TGTCCATTACACCCTGCCAGATGGAGAAATAGTAGCAGCAGGGACGACATTTGAGGTTGAAACATGTCTAGAGCATGAGGTGAGACTGAAATGGAGTAAA  1800
        V  H  Y  T  L  P  D  G  E  I  V  A  A  G  T  T  F  E  V  E  T  C  L  E  H  E  V  R  L  K  W  S  K

1801  GACAAACCGTCACCTGGTGAGGACATCAACCTCCAGATCGATGCTGCCCCTCTTTCTGTATGTGGAATCTCTGTTGTCGATAGAAGCGTTAAGCTACTCT  1900
       D  K  P  S  P  G  E  D  I  N  L  Q  I  D  A  A  P  L  S  V  C  G  I  S  V  V  D  R  S  V  K  L  L

1901  CTTCAAACAATCAAATCAAAAAAGAAAATCTTTTGAAAAGAAGGTCCTGGCTCCACACATATCAGTATACCTACCCACCACAACAACGAGACTGGGAATA  2000
      S  S  N  N  Q  I  K  K  E  N  L  L  K  R  R  S  W  L  H  T  Y  Q  Y  T  Y  P  P  Q  Q  R  D  W  E  Y

2001  CTGCCGAAAGAAAAACGAAGCCGAAATGGAGAAATTAGAATCCATTTCAAATAAGAATAAAAGAAGTTTATTTTGGGGCCCTTACTACGGAAGCAGTGCT  2100
        C  R  K  K  N  E  A  E  M  E  K  L  E  S  I  S  N  K  N  K  R  S  L  F  W  G  P  Y  Y  G  S  S  A

2101  GATTCTATTACAGCCTTTGACATTGCTGGATTGATAGTGATGACGGATTTATTGCTGGAAACTCGGCCATGCCATACAGTTCACGGCCCTGCGATTCTTC  2200
       D  S  I  T  A  F  D  I  A  G  L  I  V  M  T  D  L  L  L  E  T  R  P  C  H  T  V  H  G  P  A  I  L
                                                                                          Bait region
2201  GACGGCCTGTCATGGCAGAAGTGGCACCTACTGTAGAAATGGCCCTTGATGCCCAACCTGTTGATGCGGAATATTATGGAGGGGGTGCTTCACCGGGACC  2300
      R  R  P  V  M  A  E  V  A  P  T  V  E  M  A  L  D  A  Q  P  V  D  A  E  Y  Y  G  G  G  A  S  P  G  P

2301  ATCTCAAAAGGAATTGACCAGAATACGACAAGACTTTCCTGAAACGTGGCTGTTTGATCTGGAGTATATGGGGGATGCAAATTCCATCGTAAAGCAGGTT  2400
        S  Q  K  E  L  T  R  I  R  Q  D  F  P  E  T  W  L  F  D  L  E  Y  M  G  D  A  N  S  I  V  K  Q  V

2401  AAGGTCCCTGACACGATCACAGAATGGGTTGGATCTGGTTACTGCCTTTCATCTTCTGAGGGGCTTGGTATTAGTGAGGAAGCTTCTTTCATCTCTTTTA  2500
       K  V  P  D  T  I  T  E  W  V  G  S  G  Y  C  L  S  S  S  E  G  L  G  I  S  E  E  A  S  F  I  S  F

2501  CTCCATTTTTCATTGATTACAACCTTCCCTATTCCGTGAAGCGAGGAGAAACATTTGAACTCAAAGTATCTGTCTTTAATTATCTAGGAATCAGTGCGCC  2600
      T  P  F  F  I  D  Y  N  L  P  Y  S  V  K  R  G  E  T  F  E  L  K  V  S  V  F  N  Y  L  G  I  S  A  P

2601  GGTACTGCTGGAGTTGCAACAAAGTAACGATTACGATGTGGCTGATGGAGTCTACAATAAGACGGTTATGATTCCACCAAATAAAAGTCAAGTCATTACG  2700
        V  L  L  E  L  Q  Q  S  N  D  Y  D  V  A  D  G  V  Y  N  K  T  V  M  I  P  P  N  K  S  Q  V  I  T

2701  TACCTCCTGACCGCTACCAACATAGGAGATGTGAACATCACATCCTTTGCATATTTATCACCTGTAAATAGCTCGGATCCGACTTATGCTAATATCTCCT  2800
       Y  L  L  T  A  T  N  I  G  D  V  N  I  T  S  F  A  Y  L  S  P  V  N  S  S  D  P  T  Y  A  N  I  S

2801  TTACCAGTGATGCCATAAGAAAAACATTGCTAGTGGAGCCAGAAGGGTTCCCACAAGAAGAACATCAAGTAGAATTCATGTGTGTACAAGATGGTGCGCA  2900
      F  T  S  D  A  I  R  K  T  L  L  V  E  P  E  G  F  P  Q  E  E  H  Q  V  E  F  M  C  V  Q  D  G  A  Q

2901  ACTTTCCAAACTTCTAGACTGTTCTCCTCCACCTCTGATTGTACCAGATTCTGCAAAAACTATATATTCTGTCACAGGAGATCTCATGGGACCTTTGCTC  3000
        L  S  K  L  L  D  C  S  P  P  P  L  I  V  P  D  S  A  K  T  I  Y  S  V  T  G  D  L  M  G  P  L  L

3001  ACTAATATAGATTCGTTCATCCAATTTCCGACTGGATGTGGAGAGCAAAACATGGCAACTTTCGTTCCCAACATTCATGCGCTGACGTATTTAGACAGTA  3100
       T  N  I  D  S  F  I  Q  F  P  T  G  C  G  E  Q  N  M  A  T  F  V  P  N  I  H  A  L  T  Y  L  D  S
                                        Thioester site
3101  TGGGTGCAGCAGAGCCTGAATTGAGGGAAAAGAGCAAGAAACACATGGAAGAAGGATACCAACGTCAACTTAAATACAAGCACAGCGATGGATCGTATAG  3200
      M  G  A  A  E  P  E  L  R  E  K  S  K  K  H  M  E  E  G  Y  Q  R  Q  L  K  Y  K  H  S  D  G  S  Y  S

3201  TGCTTTTGGCGAAAGCAGTAGGACTGAGATTGGATCGATCTGGCTGACCGCGTTTGTTCTGAAATCATTCAGTCAAGCAGCGCCTTATACATTCATAGAC  3300
        A  F  G  E  S  S  R  T  E  I  G  S  I  W  L  T  A  F  V  L  K  S  F  S  Q  A  A  P  Y  T  F  I  D

3301  GAGAAGGTCACCAATCAAAGCAAGGAATTTATATATAAGCACCAGATGAAGGACGGATGTTTTCAGAAAGTTGGAAAAGTGTTTCATCGGGCCTTAAAAG  3400
       E  K  V  T  N  Q  S  K  E  F  I  Y  K  H  Q  M  K  D  G  C  F  Q  K  V  G  K  V  F  H  R  A  L  K
                                                                                    Catalytic histidine
3401  GTGGTGTCGGCCAAAGCAATGATACACGGGTTTTAACAGCATACATAATCACTGCACTACTGGAGGCTGGGGAGAAACCTGAAAGCGAGCCCATGTTGAG  3500
      G  G  V  G  Q  S  N  D  T  R  V  L  T  A  Y  I  I  T  A  L  L  E  A  G  E  K  P  E  S  E  P  M  L  R

3501  AGCATTACAGTGCATAAAAGAAATGCCAGACCAACCATATGCTCAGGCGCTTGGCGCATACGCTTTTCTGTTAGCTAATGGAAGAACACAGTGGGAGAGT  3600
        A  L  Q  C  I  K  E  M  P  D  Q  P  Y  A  Q  A  L  G  A  Y  A  F  L  L  A  N  G  R  T  Q  W  E  S

3601  TCCAACACTACAACTTTATTACAGGAGAGTTCCAACACTACAACTGTATTACAGGAGAGTTCCAACACTACAACTGTATTACAGGAGAGTTCCAACACTA  3700
       S  N  T  T  T  L  L  Q  E  S  S  N  T  T  T  V  L  Q  E  S  S  N  T  T  T  V  L  Q  E  S  S  N  T

3701  CTACTTTATTACAGGAGGATTCTAACCCAACAATTTCACTTAATTTCCAAGAAGCCGGAGAATCTTTCCTATCACAGCTCCTACAAACAGCAAAAAAGGA  3800
      T  T  L  L  Q  E  D  S  N  P  T  I  S  L  N  F  Q  E  A  G  E  S  F  L  S  Q  L  L  Q  T  A  K  K  E
```

Supplementary Fig. S8 (continued)

```
3801  ATCTCAGTTACTGTGGTGGGAACAGGAAGCTGGATGCAAATACTGTCGAGACAGTAATGTTGAGATTGCTGCTTACGCTCTCATGGCTTTCCTCCTGTCT  3900
        S  Q  L  L  W  W  E  Q  E  A  G  C  K  Y  C  R  D  S  N  V  E  I  A  A  Y  A  L  M  A  F  L  L  S

3901  GATCCTACTGGCTGGAGCAACCTTCGACCTATCGCCAGATATTTAATTTCTCAGAGGAACACTCATGGAGCATTCTATTCTACACAGGACACCGTGGTAG  4000
       D  P  T  G  W  S  N  L  R  P  I  A  R  Y  L  I  S  Q  R  N  T  H  G  A  F  Y  S  T  Q  D  T  V  V

4001  CTCTGGAAGCTCTGACTAAGTACGCGATGCAGGCTGAACAAGTTGAACTCAAACAAGTAGCAGTAACGTGTGTGTCATGCGCAGTAACGCATACTGTAAA  4100
      A  L  E  A  L  T  K  Y  A  M  Q  A  E  Q  V  E  L  K  Q  V  A  V  T  C  V  S  C  A  V  T  H  T  V  N

4101  TGCAGAGAACAGACTGATTACAGATCAGACAAGTGTCGGGGAGAGCTTGACTGGCAACATTGTCGTCCAAGGCCACGGCTGCGCGATTGCTCAGTGCTCC  4200
        A  E  N  R  L  I  T  D  Q  T  S  V  G  E  S  L  T  G  N  I  V  V  Q  G  H  G  C  A  I  A  Q  C  S

4201  ATCAAGTACAACATACCAGAAGCAAAACCAAGCGTAGCCTTCAACATTAAAGCATCTGGTTGGCCTTTAGATGAACCTGCTCGCTGTAGATACTCGGAGC  4300
       I  K  Y  N  I  P  E  A  K  P  S  V  A  F  N  I  K  A  S  G  W  P  L  D  E  P  A  R  C  R  Y  S  E

4301  TGCTCATATGCATCAGCTATGTGCTTGACACTGGAGAATCAAACATGGCAGTCGTGGAAATAGGCATGGCATCTGGCTACAAACCTGAATTACTATCCCT  4400
      L  L  I  C  I  S  Y  V  L  D  T  G  E  S  N  M  A  V  V  E  I  G  M  A  S  G  Y  K  P  E  L  L  S  L

4401  TGATGAGCTTCAGAACAATCTTGATATAGATTTAAAACGATACGAGGTTGAGGAGAATAAAGTAAATCTGTATTTTGATTACTTTGATCAAAATCAGAAA  4500
        D  E  L  Q  N  N  L  D  I  D  L  K  R  Y  E  V  E  E  N  K  V  N  L  Y  F  D  Y  F  D  Q  N  Q  K

4501  TGCTTCACAATTAGAATGGATCAAGACACAGTTGTGACCAATCCGAAACCAGCGAATGTGAAAGTATACGATTACTATCAGACGGAACTCACCACCAGCA  4600
       C  F  T  I  R  M  D  Q  D  T  V  V  T  N  P  K  P  A  N  V  K  V  Y  D  Y  Y  Q  T  E  L  T  T  S

4601  CGAGTTACTCCATTTGTAATGATGCGCCTATTCCTGCACCTCCTCCACCTGGCATTTAACGATGTTGAAGATCTAATGTGAAACGTGGCCTATTAATCAC  4700
      T  S  Y  S  I  C  N  D  A  P  I  P  A  P  P  P  P  G  I  *

4701  GTGATAAATCAAGTCTGCGACTTAGCTACATATTGAAATATTTATCAAATATAAATGAGTTTCATCTTCAAAAAAA  4776
```

VaspiTEP/CD109

```
   1  GAAGGCGAATGGAAATACTGGAGAAGGGAAGAACCAATTGGAGACACGCTGTATAAGCCTAAATCACTCGATATCGAAATGACGTCATACGCCCTCCTGT  100
       E  G  E  W  K  Y  W  R  R  E  E  P  I  G  D  T  L  Y  K  P  K  S  L  D  I  E  M  T  S  Y  A  L  L

 101  TGTATGTGGAAAGAGATGATACTGCAAACGCAGTGCCCAGTATGAAATGGTTGATAGCACGAAGAAATTCACAGGGAGGATTCTTTTCGACACAGGACAC  200
      L  Y  V  E  R  D  D  T  A  N  A  V  P  S  M  K  W  L  I  A  R  R  N  S  Q  G  G  F  F  S  T  Q  D  T

 201  AGTTGTGGGAATCCAAGCTCTGGCGACTTTAGCTACGGTTTTGGGATCAGGAGTTACGGATTTGAATGTTGAGTTTGAATATGAAACAGGAATCAAAAAC  300
        V  V  G  I  Q  A  L  A  T  L  A  T  V  L  G  S  G  V  T  D  L  N  V  E  F  E  Y  E  T  G  I  K  N

 301  GCTAAAATAGACCAAGAAAATGTGGTTATGCTTCAGACATTTGAGCTTCCAAGTAACACTTCCAATGTTACAATCCGAGCATCAGGAAGAGGCGTGGGTT  400
       A  K  I  D  Q  E  N  V  V  M  L  Q  T  F  E  L  P  S  N  T  S  N  V  T  I  R  A  S  G  R  G  V  G

 401  TGGTGCAGGTATCATGGTCTTACAATGTTCAAGTGTCAGCTGCTAACCCTGCGTTTACGTTGGATCCTCAGGTTTCTCGTGTTTCTACAAAGAATTATTT  500
      L  V  Q  V  S  W  S  Y  N  V  Q  V  S  A  A  N  P  A  F  T  L  D  P  Q  V  S  R  V  S  T  K  N  Y  L

 501  GAGAGTCAGCGTGTGCACGGGATATCATTACAAAGGCGACACCAATATGGCAGTGATGGATGTATCCTTACCTAGTGGCTATACGGTTGACGAAGATGCA  600
        R  V  S  V  C  T  G  Y  H  Y  K  G  D  T  N  M  A  V  M  D  V  S  L  P  S  G  Y  T  V  D  E  D  A

 601  ATTCTTTCTTTATATCGATACAGCTCCAACATAAAAAGAGTAGAGGAACGAGATGGAAAGACCGGCATAGTTATATACTTTGACAAGCTGAATAACACTG  700
       I  L  S  L  Y  R  Y  S  S  N  I  K  R  V  E  E  R  D  G  K  T  G  I  V  I  Y  F  D  K  L  N  N  T

 701  AAGTTTGTCCTACTGTAAATGCGCATCGTACTTATCCCGTAGCAGACCAAAAGCCAGCTCCGATAGTAGTGTATGACTACTACGAAAAAACACGAGAAGC  800
      E  V  C  P  T  V  N  A  H  R  T  Y  P  V  A  D  Q  K  P  A  P  I  V  V  Y  D  Y  Y  E  K  T  R  E  A

 801  CCGAAGGTTCTACTCGGCATTAACAGCTGATGTATGTGAAGTGTGCGATGGAAAGGATTGCGAGAAGTACAAATGCAAAGGCACCCCTGGCTACAATGAG  900
        R  R  F  Y  S  A  L  T  A  D  V  C  E  V  C  D  G  K  D  C  E  K  Y  K  C  K  G  T  P  G  Y  N  E

 901  AGGACGGGAGGAGCATCGTCGCTTGTCCTTTGCTGGTTCCTCATCACTGCTGCCTTGCTCCTTCATTTGTAATGGCATTGCATTAACGAGAATGTATAGC  1000
       R  T  G  G  A  S  S  L  V  L  C  W  F  L  I  T  A  A  L  L  L  H  L  *

1001  GATTCTGATTATGATTTCACTTTTTTCAAATTTGTGTTTCCAAATATTTCGCAAGTAATAGTCGAGGAAAAATTAAACCTTCTAAAATTTATAAGTAGGA  1100
1101  ATTGTATTAACTAATTCCATATTAAATCTCTTTTTCTC  1138
```

Supplementary Fig. S8 (continued)

```
                                                                                 /      sea lice A2M CUB domain
AmspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadC3-1         : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadC3-2         : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
ScsuC3-1         : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
ScsuC3-2         : -------------------------------------------------------------------------------------------------------------------------------------------------MKVPS :    5
ScsuC3-3         : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
PsfuC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC3           : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
HosaC4           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaC5           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
AmspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
AmspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
AmspA2M-3        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
AmspA2M-4        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
ScsuA2M          : ---------------------------------------------------------------------------------------------------------------------------------------------------MAR :    3
NinoA2M          : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
EpspA2M          : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : ----------------------------------------------------------------------------------------------------------------------------------------------------MK :    2
PsfuA2M-1        : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
PsfuA2M-2        : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
PsfuA2M-3        : ---------MLVSVKTFVSLGVLLLCEGVPALGAPQELLPTSEPP------KFEVVRGDNHTDLEGDSGIFKSPNFPEQYPENSNTTWKITVPKDKDVELTFHLLNLEKHQYCTWDWVKIYNAEDYSELRQLCGYLQQDLILHFNTSSII :  134
CaspA2M-1        : MVLQGDRHLSLTGGLGLLLLLSVLLLQIFGAHGAPQQDTSPPQPPSPPPLPKHKVVRGENHTDLEGDSGIFKSPNFPSFYPQNSNVTWKITVPKDKDVELTFHLLNLEKNQYCSWDWIKVFDAETHKELRQLCGYLQQDLILQFNTSSII :  150
CaspA2M-2        : ---------------VSLSIVPFVCIPLLGVLGQTPFQVPPSEPK---------VIRGKNHTDLEGYDGIFKSPNFPNDYPQNSNVTWKITVPEGQFVQLTFHILFLENNQFCSWDWIKIYDASNYTELRQLCGFIRQDLILTFNTSSIV :  126
PaspA2M          : -------------MRNSVGALWAFCLFQAAFSSSPPSISPEVLEK---------FLAGKNHTNLRGSSGIFRSPNFPEEYPANSNFTWKIDVPEGEFLQLSFHLFNVERSDHCQWDWVNIYDYNNKTQLRQLCGYLQQDLVIIFNTSSIL :  128
HosaA2M          : ---------------------------------------------------------------------------------------------------------------------------------------------------MGK :    3
AmspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadiTEP/CD109-1 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HaadiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------ML :    2
ScsuiTEP/CD109-1 : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
ScsuiTEP/CD109-2 : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
ScsuiTEP/CD109-3 : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
ScsuiTEP/CD109-4 : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
NinoiTEP/CD109-1 : -----------------------------------------------------------------------------------------------------------------------------------------------------M :    1
NinoiTEP/CD109-2 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
NinoiTEP/CD109-3 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
EpspiTEP/CD109-1 : ---------------------------------------------------------------------------------------------------------------------------------------------------MNT :    3
EpspiTEP/CD109-2 : ---------------------------------------------------------------------------------------------------------------------------------------------------MAH :    3
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
CaspiTEP/CD109-1 : -----------------------------------------------------------------------------------------------------------------------------------------MNSDNYKKFKKSF :   13
CaspiTEP/CD109-2 : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
HosaiTEP/CD109   : -----------------------------------------------------------------------------------------------------------------------------------------------MQGPPLL :    7
```

Supplementary Fig. S9

```
                                        /
AmspC3           : -MAKNLFFIICVCMLPCSIFS------------------QADYFVVAPNMF----RFGVPETVSINIHGIQGSVDV-EVFLQDSPGRGFTFSQKVVKVENEKSSQVSVLVKPSDIPDVVVTDPSARIYVYLVVKCTSPKLT--------- :  117
HaadC3-1         : -MDMWSCAVFLLVVGVFGHTC------------------AQQIYVVAPNTL----RLNSDETIAVAIDGNIG-AVV-SVFVQDHPGKVKNISQTLVAVQPGQPELFKIQLNSQNFPPNFLSGPGFPKYVSLTAVFPD------------- :  112
HaadC3-2         : -MVLSLHLVFLIFV-LIGKIT------------------AQNYYVAVPNSL----RLNSDEFIGVAADGNQG-GTV-NIYVQDYPGKIKNITITTDQIDPGTPKLFKVRLDTENFPYGFLSSPGFPKYVRLTAHFQG------------- :  111
ScsuC3-1         : MGFLVSLSVVVLA-FSQCIN-------------------GLQYSVTAPNTF----RIGVQETVGVAITNSPTPVQV-EIFIHDRTPQKKVIDSQKLTLQNDKPQITTLLLRAQDVPADQKDQPHIFVDLAVKESTNKFY----------- :  114
ScsuC3-2         : DNLLSIFFIALIETYAVEYVS------------------CINIVVSAPNVI----RFGVEETVAVSILNHPSQVQV-DVCLHDHPLKISTISCQRISLYEGKPKIVNVVLHPENIPEQQRSDANFFVLLVVRAVSGEFTQ---------- :  122
ScsuC3-3         : KGPLLILTFLLLSSLVQCNNK------------------NIQYTVAAPNTF----RVGVAETVSVVVTGNREPVRV-DIFLKDNTANKGIIVSQSLTISDGKPQITTLLLRPEDIPKSQKNYRNIFVYLNVTESTGKFN----------- :  117
PsfuC3           : MRSSPLWLLASLAFTSLSSPA---------------D-GAAMKVMMAPNLL----RVGTPEKIFVECQDCTGAEIPVEIKVMNHPTKTETLKSTR--VRLNSGNHFQALGELTVPTA--SFSRDPSAKEYVVYLEAHFP----------- :  114
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ----------------------------------------------------------------------TGGDIPVTVKVMNHPTKNRELATKS--VTLTSAKDFQALAQISIPAD--GFSRDPNVKQYVVYLQAQFP----------- :   64
HosaC3           : GPTSGPSLLLLLLTHLPLALG------------------SPMYSIITPNIL----RLESEETMVLEAHDAQGDVPV-TVTVHDFPGKKLVLSSEKTVLTP-----ATNHMGNVTFTIPANREFKSEKGRNKFVTVQATFGT--------- :  114
HosaC4           : --MRLLWGLIWASSFFTLSLQ------------------KPRLLLFSPSVV----HLGVPLSVGVQLQDVPRGQVV-KGSVFLRNPSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPEVQLVAHSP---------- :  115
HosaC5           : ---MGLLGILCFLIFLGKTWG-----------------QEQTYVISAPKIF----RVGASENIVIQVYGYTEAFDA-TISIKSYPDKKFSYSSGHVHLSSENKFQN----SAILTIQPKQLPGGQNPVSYVYLEVVS------------- :  108
AmspA2M-1        : --MVAIMKIIFLAATCL-------------AFTCLVN-AERGYLLTVPKVL----DSGTSERLCLTLTDVK-GPGT-VTIRL--------LQDRIDTAIAEKSIT--YPRDDS----CFFLEIPNVKTRNARIEIKG------------- :  101
AmspA2M-2        : ------MKLILLAVACL-------------AFYSKAN-AERGYLFTAPKVL----DAGTSERLCLTLTDVK-GPGT-VTVSL--------LKEKIDEVIAETSVN--FPRDDM----CFFIEIRNTKLTRGNLKIQG------------- :   97
AmspA2M-3        : --MMKTMHTVYLAAIFL-------------VSVCQVG-AERSYLLTAPKVL----DAG-SERLCLTLTDIE-GPGT-VNVRL--------LHEN-DTTLSETSIT--YPRHEP----CFFLKIPNIKSTNAYIEVNG------------- :   99
AmspA2M-4        : ------MKLLFVTSIFI-------------AFSFPIN-AERGYLLTAPKEF----HAGSSERLCLTLTDVK-ERGQ-VRVRL--------LQERKDIAIAEKLIN--YPSEDP----CFLLEIPNTNFIKGRFEIKGRFGSVQYDFDSDD :  110
HaadA2M-1        : -MDNLTRFLVLCCTLCS-------------LTIAVRG-SKGNFILTAPRAI----DAGSIVYFTLTVFDIP-QGGT-VTLRL--------THLISNVLIAESRVS--VHNNYN--TWVEMNVPPMSSDISATLHIIG------------- :  104
HaadA2M-2        : -MGIHSIVWTIVSFLYL-------------INAETEC-KKDGYIFTCPRSL----KTGASNQMQLRRYGEL-DAGE-FKITVTYM----NSINQNETIATEQTFD--IPEGEADTLMTVFLEPIENYVFNGKVTING------------- :  110
ScsuA2M          : GTRTDPKKFFLFLVCFV-----------FCTRAAEEQ-DKRGFVLTAPKHL----LAETVEHICLFFHNIN-YDGE-IWLELLS---------ENSTVISTSSQK--IFKGKGE---CIEMFIPNLFIGNAKLSVRG------------- :  108
NinoA2M          : ------MKLHHILVIFL-----------VSASANCDA-VKKGYILTAPVVF----DANSSETICIAFQGIS-SNGT-VNVELK----------TLETILSQTQQD--ILNGESK---CFQLKVPIGNFQNVVLAVNG------------- :   98
EpspA2M          : ------MKLFKTLCFVLHFLSIGFIANSIRAQDSVEN-QKKGFILTAPSIF----KTNVDENVCISFQEIH-GDGE-IRVFIQG--------STTNTTLAIAIQK--IKNGANV---CFSLHVKVTTDLEGVILVDG------------- :  111
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : NLNWVILLCCMYLFV------------------QGSH-GKRGYIFVAPKQF----QRESNETMCILFNNLEEAAAN-ITITLYD-------DKNETSSTKNFNMT----GSSSWH--CISDFQVPAIESYGGYLNLR------------- :  102
PsfuA2M-1        : LNSKYVSLTLLLNFLLLISPN---------VFPPRRY-PQSLLKVYRPINV----YPGQKNFYYGFSKNINCE---SGSGIKSEIKITNGIDNFPLIFPTDIEETPKNRKLIDKFEVQPIKVGNFCFIAFEKILPDGFDSYSINLIQD-- :  132
PsfuA2M-2        : ELYYQNKLACRVVQEMRPSMR-----------NNQIY-EALIKHGFSRPWI----QDN-QYLNLR----------IKVSGRMIFYSKKVVHAER-------------------------------------------------------- :   68
PsfuA2M-3        : VELRSDSVSQRTGFLASWRPISQ-D--SSEED-NSGE-PKEGYIFTFP-QSLTVSDEPSHEEACLEIFNVNTK-GK-VFVNVL---SSNNILNLETDIYKEIDYE---PGSPSEKVKCFDIMLPIDFSEK-YAIIQI------------- :  256
CaspA2M-1        : VEFRSDSVSQRTGFLASWRYATHMD--SEKEE-EKKE-KEKGYILTIP-QSITASEEADPEEACLEIFNVDSK-GK-VFINVF---SSENVLSRDTNIYHVIDYE---PES-FERVQCFDLTLPSNFSGQ-YAILQI------------- :  272
CaspA2M-2        : IEFHSDSVSQRRGFLVSWKPVSPEN--KDGEV-ENR---REGYMLTLP-QSFTASDETTPEEMCLEIFNVRSK-GK-VFINVF---TSQSVISEDTDIYKEINYD---PEVNGDKLQCFNMKIPTDFREK-QAIVQI------------- :  247
PaspA2M          : LELHSDSVIHKTGFLASWKSVTATEEEKEDSESDKK---KEGYILTFP-QSFTSSKETSAEKTCLQIFNVHTK-GK-VFINVF---SSENILSGNPVLYKEIDYD---PSVQGEKLQCFDIELPTDYSMKKYAIVNI------------- :  253
HosaA2M          : NKLLHPSLVLLLLVLLPT---------------DASVSGKPQYMVLVPSLL----HTETTEKGCVLLSYLNETVTV-SASLESVR--------GNRSLFTDLEA-------ENDVLHCVAFAVPKSSS---NEEVMF------------- :  102
AmspiTEP/CD109   : ----MTLAIIIIFIFLYPSV-----------------ESQGYYTVVAPKVL----RPDTRYHIGVSIYNTTST-----VHVAVQLTG------NLRVSSELDVRG-------GDTGLVTFQIGNWSAG-VYKLEVVG------------- :   93
HaadiTEP/CD109-1 : -MLLKLWAFIALIALVASQE-----------------SESSVYTVIAPSKL----RPNIPFHVSASAHNVQKP-----LDMKIAIEGPADSGQYNNVEKRITLNS-------GETQILNFEIGEWSSG-NYSLIVTG------------- :  102
HaadiTEP/CD109-2 : LLSSSVFIILSYIVLKTYCQ-----------------RTSPIYTVTAPAKL----RPDIVYHVSVTLHDSPAD-----VDFNVQIIGISEDLSPVNVVKDVHLES-------KQTKVVDFELFGWKPG-NYTLEVIG------------- :  105
ScsuiTEP/CD109-1 : VKLTCFTIILLFLTFSGGNG-------------------TSYYAITAPEIL----RPNALYRVFVTVFEVTDP-----VIVEATLKPASH--LINENLTAIASINS------GKTEEINFQVGNWNHQ-LYRLEIKG------------- :  101
ScsuiTEP/CD109-2 : DKVIYFFLLFAFASSVEKN--------------------QPFYTVTAPNTI----RANTPYQVVVTVAYVNKP-----VNVVVSLNPVHVNVTEDEELRVVGTLNN------GETKTLELKIGNLNVR-AYELIVQG------------- :  102
ScsuiTEP/CD109-3 : TFNYRIFIGLIFFVFIFQNV-----------------LAIGHYMIVAPKIL----RPALNYHVSVSVHNVTQP-----VRIDITIGGISDSGVLVSLPQQVFLES-------EKSQVVNFQIGMWGMG-KYSMKAIG------------- :  104
ScsuiTEP/CD109-4 : DKIIIFSFILGSLYAVQGL--------------------DPLYTVTAPKTL----RANSPYHVVITVHQAPEP-----VNVEISLKPTEN---VTEELKEKDVISN------GETKKLQLNVGDWNYP-SYRLNVQG------------- :   99
NinoiTEP/CD109-1 : ANVVFATALLVIVLAAVHPS-----------------AAERTYTVTAPKNI----RIGTPYQVVASIHNTPEE-----VQIFANLSCTSDDIDKPQIASGSVTVPN------EATKMLNLTIPDWWKPGSCELTVSG------------- :  106
NinoiTEP/CD109-2 : --MGELYRILISLLAVIHLT-----------------AAAGTYTLVGPKIL----RPGLSYQVSVSIHDTREP-----VKVSVSITGAGLAGVNYGQTTEALIQS-------GETQILKFEIGEWGPG-KYNISISG------------- :  101
NinoiTEP/CD109-3 : -MALLKISLVTTLLVLLPYV-----------------SGECFYLATATNQL----RPNTDFNVAVTVQCSGNDTNA-VFQVTLSLTSVSDGSTTAITSSSANAQSTV---TSGSTVIIPLKIGANWNANQYILSINGK------------ :  112
EpspiTEP/CD109-1 : PNGLLAVLLLGICAATSH--------------------AERTYTVTAPKNV----RVGTPYQVVVSIHNSPED-----VEIFANLSCSSDDNSKPQLVTGSITAPN------EITKLLTLPIPDWWKPGNCELTVTG------------- :  105
EpspiTEP/CD109-2 : TFCRLLFALVAVLLSVTQLS-----------------SAVGTYTLVGPKVL----RPGLSYQVSVSIHDTREP-----VKVTVSITGQGLAGVNYGQTTEALIQS-------GETQILKFEIGEWGPG-RYNISISG------------- :  106
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ---MMKISKVASIFFLSLSI-------------ALIQAEKPTYTIIAPDVI----RPNTDFLIAVSTHDISSD-----QDVLLTIKGQSESEGNVEISKETTVRP-------DETQIVRMVIGNLGEG-KYVLHARG------------- :  104
CaspiTEP/CD109-1 : DRMRDFKVILYLLPVFYLSV-------------GLSEAKKPTYSIIAPNVI----RPNTDFLIAVSTHGISSD-----QDVLLTIKGQSESEGNVEISKETTVRP-------DETQIVRMVVGNLGEG-KYALHARG------------- :  120
CaspiTEP/CD109-2 : -------------------------------------------------------------------------------------------------------------------------------------HARG------------- :    4
HosaiTEP/CD109   : TAAHLLCVCTAALAVAPGP----------------------RFLVTAPGII----RPGGNVTIGVELLEHCPSQ----VTVKAELLKTASN-LTVSVLEAEGVFEK-----GSFKTLTLPSLPLNSADEIYELRVTGRT----------- :  110
```

Supplementary Fig. S9 (continued)

```
AmspC3           : ---------------FKKEALVLVSPHNGFV-FVQTDKPIYTPRQEVRMRIIPLD---QDLTYMKSP-----FRLRILNPQNITVRRIDYDEGNK----QGIITEIYKFPRQAIVGIWGISVEHGHE-LSTTTYVSFEVKKY--VLPTFD :  236
HaadC3-1         : ---------------FRKELIIPVSNQSGYI-FIQTDKPIYTPKERASIRIIPLN---EDSRPSEER-----FLLQIRNPKNIIVEEKYFNDKNR-KLDKAFASHVYRFPTYPVLGEWSATVRYGHD-LEQNATVRFELEEY--VLPTFT :  234
HaadC3-2         : ---------------YYKELLIPINEKYGHM-FIQTDKPIYTPKDRVLFRIMHLN---ESALPTDET-----FQIQIRNPNNSIVDQKLFDGKRKSELKKVSVTHVYKIPTYPVIGEWSATVQYGYN-LKQNTTVHFLVQEY--VLPMFT :  234
ScsuC3-1         : -----------------KEMTIPVTKYSGYV-FVQTDKPIYLPNQRVHIRLFYLD---ENLLPLTGD-----LTLEVKNPNG----SRVLYKENLPATPSGITEASFKFPSSPVFGNWTVTAFYGYK-KAARTTVRFEVKDY--VLPTFS :  231
ScsuC3-2         : -----------------KQAVIPVMKYSGYV-FVQTDKPLYLPNQKVHIRVLRLD---EKLLPKNDV-----VRLEIKNPKD----MIVRQQELMPN-TKGIREAVFTISQVPILGNWSVSVHYGFK-FVAKTTVRFEIKEY--VLPTYS :  238
ScsuC3-3         : -----------------KEERIPVTKYSGYL-FIQTDKPLYTSSDSVHIRIMYVD---EKLIPITEE-----VKLEVKNPND----TIVYYKESLPCKPNGFTEVVFKLSSSPVFGNWSVSVSYGYE-MAAKTTVGFEVKQY--VLPTFS :  234
PsfuC3           : --------------DRVLEKVVMVSFQAGYI-FIQTDKPLYTPDSKVQYRLFAVTPGMEPVERDNSSKAGASVAIEIVTPEN------IVLTQDPVS-MIGIHSGDYTLGDVVSFGVWKIVAKFQSN-PQLSFSAEFEVKEY--VLPSFE :  239
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : --------------DRLLEKVVLVSFQSGYI-FIQTDKTLYTPESTVLFRMFGVTPRMEPVER--DDKTDASIAIEIVTPDG------IILPLDPVFLKSGLYSGDYKLGDIVSFGVWKVVAKFQSN-PQQTFAAEFEVKEY--VLPSFE :  188
HosaC3           : ---------------QVVEKVVLVSLQSGYL-FIQTDKTIYTPGSTVLYRIFTVN---HKLLPVGRT-----VMVNIENPEG--IPVK--QDSLSSQNQLGVLPLSWDIPELVNMGQWKIRAYYENS-PQQVFSTEFEVKEY--VLPSFE :  233
HosaC4           : -----WLKDSLSRTTNIQGINLLFSSRRGHL-FLQTDQPIYNPGQRVRYRVFALD---QKMRPSTDT-----ITVMVENSHGLR------VRKKEVYMPSSIFQDDFVIPDISEPGTWKISARFSDG-LESNSSTQFEVKKY--VLPNFE :  242
HosaC5           : -------------KHFSKSKRMPITYDNGFL-FIHTDKPVYTPDQSVKVRVYSLN---DDLKPAKRET-----VLTFIDPEGS-----EVDMVEEIDHIGIISFPDFKIPSNPRYGMWTIKAKYKED-FSTTGTAYFEVKEY--VLPHFS :  228
AmspA2M-1        : -----NFQLDRYSFE--SETAVSITSRSTLA-FIQTDKAVYKPGQKVQFRILTVN---HLLEPIS----TEISKVYINNPTGV----RIAQWLGVKNEN-GLIQLDMQLSDEPTLGKWKIYVQLQG---ASKV-QEFEVDEY--VLPKFE :  225
AmspA2M-2        : -----KFDSDDYSFE--SESPVSIASRSTLT-YVQTDKAVYKPGQKVQFRILTVN---HLLEPLP----TEISKVYVLNPNGI----RVAQWLGIKNEN-GLIQLDLQLSDESARGQWKINAQVKG---STIS-QDFEVDEY--VLPKFE :  221
AmspA2M-3        : -----SFESDQYHFG--SKTGIVIAPLAPVT-VVQTDRAVYKPGQKVQFRILNVN---HLLEPVA----NEITKVYIHNPNEV----RVAQWLGIQNEN-GLVQLDMQLSDEPTLGLWKIYVEIQG---HKRE-QGFEVSEY--VLPKFE :  223
AmspA2M-4        : DYLDYSFDSDQYTFD--SETKVSIKSRSTLT-FVQTDKAVYKPGQKVQFRILTVN---HVLEPDS----TEISKIFIENPKGI----RIAQWLAVKNEN-GLIQLDMQLSDEAPKGTWKISAQVQG---LTKE-QSFEVDEY--VLPKFE :  239
HaadA2M-1        : -----SFSSDYHIEA--SQNIHIRHNT-ILT-FIQTDKPLYKPGQTIRFRVLPMD---NQLKPLDA---NTMGDIWIEDPSGI----RVAQWNHEQFTE-GIKQFELPLSGEPPLGTWNIHAFINQ---VTTS-QTFIVKKY--VLPKFD :  228
HaadA2M-2        : -----TFDDYVIGG---VEKVYFSSSRDDIV-FIQTDKPLYKEGQTVKFRVLRVD---KTLRPSV----KDQASIWVEDPAGT----RLFQWKNISMER-GMKQFEFPLADEPVLGNWKISVSFQG---EVTT-TTFEVKQY--VLPTFD :  233
ScsuA2M          : ---IFPTEDGSDPYNIHQTKSVIIKHYNSLV-FIQTDKPVYKPGEKIRFRILHVT---MDLKPIS----DEIPSIWIEDPSGV----RISQWLNEKPEL-GLIDLQMSLSTEPILGKWNIKASIG----KLRKTQSFDVEEY--VLPKFE :  236
NinoA2M          : ---SFENSD----YTFSGSKPVEIKQVKDAT-FIQTDKAVYKPGQKVQFRILRLT---SKYLSSA----TKIKTVWIENPFQL----RLFQWVSPQSRS-GLIDLNFQLSEEPQLGEWHIFAEYET---GDKVNQKFKVEEY--VLPKFS :  223
EpspA2M          : ---WFQDEN----YRFNSTEKIIIKKQPEIT-IIQSDKPLYKPGQKVLFRILRVD---SYFLPSR----SKINSVWIENPSQV----RVAQWLSLPASN-GIADLNFQLSVEPQLGVWNIFVEDEF---KNRVNKTFKVEEY--VLPKFS :  236
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : --ISFPSLPLEHAVDETREATILFSRNHGPTTFLETDKDKYKYGQTVKFRILSLD---FDNKPFT----DTIPEVWIEDPGLN----RVAQWKDVAMKH-GLNQLSYTLSSDPPKGHWTIRVKGIHG----EKSINFEVAEY--VLPKFT :  232
PsfuA2M-1        : ------------KKEFSFGNRIFTKNIKPFT-LLQLNKPIFRPGENLKIWLKVFYP---KEMEFLIRNKN-----TKKSSMNKVKVNILDSKVRKVDFFESDMNNDMFVEHKINRE-SFLGQWKIVVFLNG---VVIETKRF--KVDQYL :  255
PsfuA2M-2        : -RMMSYIGHHNNIIYTKSMNIKMRKLS-KLT-FFELDKKYYRPGESIKVWIKQVDHSQLVDLKKMTKNSCFGGKGASNMKLPELTLNIEYPNGMKFEKKKIGNTNQDLFKFEIPSE-TLMGKWKAYLEYGK----EKSYKYF--EVKYFT :  208
PsfuA2M-3        : ----NGTFEDND-YRILSYKSVKVLKSS-ANMLIQTDKSEYRPKQKVLFRVMKLD---DNLRPSKN---NVLNKVFVKSPSRQ----TLAQFINHSLDSRGIGQFEFQLDEEPELGNWQIMVENDVVGGEEDSSVELAYFVVEAVLPKFE :  390
CaspA2M-1        : ----NGTFQDEENYRILSYKSVKVIRSS-PQVLIQTDKKEYRPKQTVLFRVIKLD---ESLKPSKS---RVLNRIFVKSPSRQ----ILSQHLNVSLGSRGIGQFEYLLDEEPELGNWEITVENDAESEKEEASSASPAFVVEAVLPKFE :  407
CaspA2M-2        : ----NGTFEDSD-YRILSYKAVKVFKVS-VQNLIQTDKHEYRPKQLVKFRVLSMN---GNLKPSKN---KMLDEVYVKSPSRQ----VLAQYKNVELDPRGIHHFEYLLDEEPELGSWEISVITNG-ADDLDSSVELASFIVEAVLPKFE :  380
PaspA2M          : ----NGTFEESD-YSILSYKTLRVLKVP-VKTLTQTDKGEYRPEQTVLFRVLKLN---GELKPSEN---RTIDEIYVKSPLGR----TMAQYKNVTLSPQGLGQFEYTLDKEPELGPWQIVVRVFG-NDTHEPVEDIAIFVVEAVLPKFQ :  386
HosaA2M          : ----LTVQVKGPTQEFKKRTTVMVKNEDSLV-FVQTDKSIYKPGQTVKFRVVSMD---ENFHPLN----ELIPLVYIQDPKGN----RIAQWQSFQLEG-GLKQFSFPLSSEPFQGSYKVVVQKKSGGR---TEHPFTVEEF--VLPKFE :  230
AmspiTEP/CD109   : ----------SRGLDFRNSTEIKFVARSLNV-FIQTDKSVYRSSQKVQFRAIILD---KSLWPRRTA-----VEIYITDADGN----RVKHYRGLNA-RLGLVSEELQLPDQPVLGVWIIHVVASG----QEYKKSFSVAEY--VLPGFY :  213
HaadiTEP/CD109-1 : ----------EGGMTFRNQTILTYEHKSYSV-FIQTDKAIYKPGQLVQFRVLVVT---PHLLPSVTG----AINIFITDASGN----RIKQWNRLFT-TKGISSGELLLSEQPVLGDWTINVDVLG----QMFKKSFTVAEY--VLPNFE :  223
HaadiTEP/CD109-2 : ----------QGGLTVRNSTQLTFEHKSHSV-FIQTDRPVYKPGQIVQFRVIVVD---PYLVPSPNAL---FIDLLVKDGHGN----SIQEWKNIPT-RTGLASAEFTLANQPVLGEWEVHVDVEG----QKFKKPFTVAEF--ILPTYE :  227
ScsuiTEP/CD109-1 : ----------SGGITFRETATVKQSFKIFSI-FIQTDKGIYTPSQTVYFRIIVTK---PSLLPHNPN----GLVVYINDPEAN----RIKQWTNVTF-THGIYSGKFQLADQVNFGKWKISADFQG----QTESVHFTVEEY--VLPKFE :  222
ScsuiTEP/CD109-2 : ----------SGGIDFNQTESVKINENPFSV-FIQTDKGIYQPGQIVHFRVIATK---PVLLPYNPE----DIEIYVNDPKYN----RIKQWRNVTF-THGVYSGEFQLSDQVNFGGWMINVNHKAN---ILRSTSITVEEY--VLPKFE :  224
ScsuiTEP/CD109-3 : ----------SGGLQFANETDLSYELKSYLV-FIQTDKAIYKPGQPVRIRVIVVS---PSLRPAGTE----PLDMYITDGQGN----RIKQWKRAFT-SRGVFNAEMQLSNEPVLGDWNITVLIQD----QVYKKSFTVAEY--VLPSFE :  225
ScsuiTEP/CD109-4 : ----------SGGIVFNKTETVNFNSKSYSV-FIQTDKAIYQPGQIVHFRVIVTN---STLQPISP-----ETVIYVSDSQGN----RIKQWMNVTF-DRGVYSGDLQLSDQVILGDWRISVDTQK----VLQTKYFTIAEY--VLPTFE :  219
NinoiTEP/CD109-1 : ----------DKGIVFKRSASLGFNSKTSSV-FIQTDKAIYQPGQLVQFRVLVVD---PNLRPSQTD----DLNVFITDAQGN----RIKQWNKAQL-KSGIFSGELQLSDQPVLGDWVINAQANG----SSATKAVTVEEY--VLPKFE :  227
NinoiTEP/CD109-2 : ----------SGGLTFFNTTEIEYQHKSYSV-FIQTDKAIYQPNQIVHFRAIVVN---PMLRPTVSG----AIEVFITDGQGN----RVKQWRRVFT-SKGVFSGELKLSDQPVLGNWNITVFVSD----QQYTKSFIVMEY--VLPTYE :  222
NinoiTEP/CD109-3 : ----------STSLSFNRTLSLNFNSKQFSL-ILQTDKGIYQAGQTVNFRLFCFN---ADLLPYPNQ----TVEIYVTDAAGN----RIQQWNVPIP-TSGIYANSFPLSTQPVLGTWSINANSGD----QQDSAQILISQY--VLPRFQ :  233
EpspiTEP/CD109-1 : ----------DKGIVFKRSASLGFNSKTSSV-FIQTDKAIYQPGQLVQFRVLVVD---PNLKPYASD----ELTVFITDAQGN----RIKQWNNATL-KSGIFSGELQLSDQPVLGDWSINAQLND----ARANKQVSVAEY--VLPKFE :  226
EpspiTEP/CD109-2 : ----------SGGLRFSNTTEIEYQHKSYSV-FIQTDKAVYQPNQMVHFRAIVVN---PMLRPTVSG----AIEVYITDGQGN----RVKQWRRVFT-TKGVFSGSLQLSDQPVLGNWNITVFVSD----QQYTKSFTVMEY--VLPTYE :  227
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ----------NNPIAFDETQTLKYIHQGYSV-FVQTDKAIYRPGNVVKFRVIVVT---PKLKPSVVG----SIDVGVLDGKGH----LVRKWDRVFT-KGGVFAEEFLIDDDPVRGDWNITVDVSG----QRFSKSFQVVEY--VLPQFI :  225
CaspiTEP/CD109-1 : ----------NSPLAFDETQKLNYIHQGYSV-FVQTDKAIYRPGNAVKFRVIVVT---PKLKPSVLG----SIDVGVLDGKGH----LVRKWDRVFT-QGGVFAEEFLIDDDPVRGDWNITVDVSG----QKFSKSFQVVEY--VLPQFI :  241
CaspiTEP/CD109-2 : ----------NFPIAFDHTQKLIHINQGYSV-FVQTDKAIYRPGNVVKFRVIVVT---PKLKPSVVG----SIDVGVLDGKGH----LVRKWDRVFT-TGGVFAEEFLIDEDPVRGDWNITVNVSG----QKFTKSFQVVEY--VLPQFI :  125
HosaiTEP/CD109   : ----------QDEILFSNSTRLSFETKRISV-FIQTDKALYKPKQEVKFRIVTLF---SDFKPYKTS-----LNILIKDPKSN----LIQQWLSQQS-DLGVISKTFQLSSHPILGDWSIQVQV-ND---QTYYQSFQVSEY--VLPKFE :  230
```

Supplementary Fig. S9 (continued)

```
AmspC3           : VKIVVPA----TIMEDTEEIMGKVIARYVYDKPVFGAVRLRYGVKAVNG-DISYLPDDTKVVIKKIRGEATFSIPM--KMVTNKFKSFPEG------------HKFHVEATVVEGVSGKEFMEISEDAYFQKSPYIISFKKS--VNTFKI :  365
HaadC3-1         : VELKVPD----VVLPKDETIQIEVKANYVYGKKVKGIVTFRLGVKGDHSPDAVFFAVITPKELDDGSYVLRIKTDD--LRRHKDIGWFPEIE----------GSHLVVEATVADAATGNKETTIDSKGRFSKTPFLISFKRC--LKDFKP :  366
HaadC3-2         : IQLETPE----IILPKDEEISIKVSARYVYGKKVKGMVSFRLGIKKPTDSETKFVAIIGSRELTDGFYNLRLNVQE--LLEKSNIESFAEIQ----------NSHLIVEALVTDIATGNEEEVVDSKCRFTDTPFSISFKRS--LEDFRP :  366
ScsuC3-1         : VKIKSQK----VVLKTDNLVKVDMIAEYVYGKPVEGFVNYKFAIRKPSGSIHSIGGHSNLKLR-DGKSTITIR-KS--DIVK-KLQWFPAID----------KSVLIVEAEVIEQATGKRESEYDDSTIFTTSPYVIDLSRS--LNEFKP :  360
ScsuC3-2         : VTIDLPK----VILETQKDLEGKVIAKYVYGKPVQGFVNFKYSIKNSVGVQILIGMRRNVKLS-SGEIQFKIPLEK--EIKAKNLPWFPNIE----------KSRLIMEAEAIEQASGKRETTIVDNTVFTTSPYKISFKNS--LKSFKP :  369
ScsuC3-3         : VTITPRK---YFVLSTDDAIIGDIKAEYVYGKAVQGAVTYKYYIRQPSGIQSPIGKLVHYPLNKDGTSTYNIP-KR--VIDENRIPWFPAID----------KSTVIVEAEVTEKATGKKESAINDDTIFTTTPYMIRFHRS--LKEFKP :  366
PsfuC3           : VKLTPS--NPFFYVDSPS-LAVDIRATYLFGEDVSGVAYVVFGVSKEDRK--YGLPASLQRVS--IGGGTGQATLT--REQITQTFRN---------VEELVGGSIFVTVSVLTDNGGEMVEAEFRNIQIVKSPYAINFKKTP--KFFKP :  369
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : VKLLPV--SPFYYVGSPE-LTVNIKATYLFGEEVDGTAYVVFGVVENKQK--RGFQSSLQRVP--IKDGKGDATLK--IEHITNTFPN---------INELVGTNIYVAVSVLTESGSEMVEAELQNIQIVTSPYTITFKKTP--KYYKP :  318
HosaC3           : VIVEPTE--KFYYIYNEKGLEVTITARFLYGKKVEGTAFVIFGIQ-DGEQRISLPESLKRIPIEDGSGEVVLSRKV--LLDGVQNLRAEDLV----------GKSLYVSATVILHSGSDMVQAERSGIPIVTSPYQIHFTKT--PKYFKP :  366
HosaC4           : VKITPGKPYILTVPGHLDEMQLDIQARYIYGKPVQGVAYVRFGLLDEDGKKTFFRGLESQTKLVNGQSHISLSKAE--FQDALEKLNMG--------ITDLQGLRLYVAAAIIESPGGEMEEAELTSWYFVSSPFSLDLSKT--KRHLVP :  380
HosaC5           : VSIEPEYN--FIGYKNFKNFEITIKARYFYNKVVTEADVYITFGIREDLK--DDQKEMMQTAMQNTMLINGIAQVT--FDSETAVKELSYYS-----LEDLNNKYLYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATP--LFLKP :  365
AmspA2M-1        : VSVTP-P---KYLLANSETATWKICARYTYGKDVQGTLTAEMNYSGYAW-------E--KR--GPPLSSLTTKKIN--GCYDWIVHVR---DAKWNMSSVS-YKSIIVKADVKESGTDIEFSSSSTTSIGHDVLKINM-TPKNGEIYFRP :  353
AmspA2M-2        : VTITP-P---TFLSSNMDIATWKICARYTYGKGVQGTLKAKLEY-VTYS-------Y--ERDRDSFPETNLEVKIN--GCHDLTVYAK---TMKWNTRKMH-YKSLKLNAEVEEEGTDIKFSSSSSMSISHEPLFLNF-KNWDRQKYFHP :  350
AmspA2M-3        : VTVTP-P---TTLFPDSDSATWKICAKYTYGKGVEGVLTAEIENTATKY-------W--R----IPMKRNLETKIN--ECYNWTLRKE---DSFWNGIDLD-FGVINVSAKVKESGTDIEFSSSSSTKVERWPYTIEG-RSNEDRYYFRP :  349
AmspA2M-4        : VKVVP-P---RFLLPNMDSATWKICAKYTYGKNVQGTSTIVIGYRNTPL-------N--EGNIAHYPHEKYEVKVD--GCYELTVNKT---VMRWYSALVD-SKYIGVTSSVKEDGTGVEFTDSSVTLLHETPLKIDF-EARNNEKYFRP :  369
HaadA2M-1        : VSIKP-P---AVIMADAQTIPIEVCAKYTYGKYVEGSLKAKVTYKKLWT-------FM-YRDQRTIPSVEHQAELP--GCHTFQVNTD---DLLMQTEEFG-GKELEIFAEVTENVTGIVRNATTSFEISHQKVFLEFLRDN---DYYKP :  357
HaadA2M-2        : VKITL-P---SFVLSNAEEIPITVCAKYTYGKPVKGVLRLNTSLEMFS------------WGDEKFPTIEYEGKIN--GCFDYVINVS---MVET-EDYYR-YRRIQIVASVEESGTGIERNETQYVQRQYSPLSLSFNRDQ--KQFYKP :  358
ScsuA2M          : VKISP-P---PFILANQLNAVWNVCAHYSYGKPVQGYAVIKAVLG------------SIEFPTK--TVAQYEGKIT--GCHSIIGKIQ---QKLRSYG----PFQIEIHAEVTELGTNATMTALAQSKIHLEALILDMSSYMP--FYFKP :  357
NinoA2M          : VQITP-P---AVILSNFKLFTWKVCAYYTYGEPVVGKIKANITG----------------FPNS--NQIQVEDEFR--ECRTMNVSVE---QLNLNNGRIPRNVIVKFYAEITEYGTGVTMHDFKQVPVKTNPLILSLDAKQ----YYKP :  342
EpspA2M          : VEIIS-S---SVILSNFENYSWKVCAHYTYGEPVAGKLKARVAMENVRH-------HYRRMPRNPPKNFTLESEFF--KCKLINISKS---DIVADSTGSIYGDQIKLYADIIEHGTDVIMSASKLVSVKRTAVNLSLKSRD----YFKP :  366
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : LSIKGQR-DSNNIYDTDKNYTWTICAKYTHGEDVTGTMDVSFTKRSY-------SWWRNDQEKDKYVAVYNKTRFS--GCKEVTLTKE----EYGGDGGDRKHFGLELKASVEEDGTGVTVNDSIDHYHSVNSELVMNENDN----HFKP :  364
PsfuA2M-1        : MPTLEVSLKKIAEKNMTVTYEMCAEDNENRPIIGIAKTKFCIRRRSMS-----------------CNDKIFKFQNEK-CMKKDFEINS----------------NDLDIYSHFLQVDFIESVTGIKQR---RKAYSNSGFFGDFVVKKYQ :  368
PsfuA2M-2        : VPRYSVTQKLIPLSMNEVSVEACAKMSNARQFKGTVEIKSCINSRR-------------------CVTKNISFSTTK-CVTAKIHFKN--------------TYFSKWGVNFSAEFKYFEYKGAEPIIKKLSCYYRSSRIVARPYQIIKL :  324
PsfuA2M-3        : VQING-P---KAILSEDEKVDFEICGIYTHGSKVKGSVEVFFEHKYTK------GSAYWRAPMYSSNFTKKIVIKDDDTCGTVSLNQT-----EISLLSIDGDRDFSINAKLTEEGTDIVQESSSNEKVLFTRAKIDFGDSS--YEHILG :  523
CaspA2M-1        : VVIKS-P---AVILGEDEKVEFEICGIYTHGSKVRGELEVFFDHRYRQ-------ASYWRAPTYSSNFTKKVHIKDDNACGTVSLNRT-----EISLLSKHKETPFTIKATLTEEGTAIQQESRSEGKVTFTKAEIDFGDSA--TEHILG :  539
CaspA2M-2        : VQINS-P---KVIFSESEKVNFEICGIYTHGSKVKGSVEVFFEHKYKT-------ANYWRAPTITSNFTKKSIISQDNACTTVSINNT-----ELAKLSLQMETPFLVKAKLIEEGTDIVQEASKQEKVIFTHAEFNFGDSS--SEHIIG :  512
PaspA2M          : VKVEA-P---KAILAESEEVKFEICGIYTHGSKVKGTVKAHFEHKYRV-------GTYWRAPYKTFNLTKIVEIKDDDSCATVILNST-----ELGDLSKKPDEPFTLTANLTEHGTDIVQSHTLTSKI-TSP-TQKSISKIA-LPRTLP :  517
HosaA2M          : VQVTV-P---KIITILEEEMNVSVCGLYTYGKPVPGHVTVSICRKYSD-----ASDCHGEDSQAFCEKFSGQLNSH--GCFYQQVKTKV--FQLKR---KEYEMKLHTEAQIQEEGTVVELTGRQSSEITRTITKLSFVKVDS---HFRQ :  361
AmspiTEP/CD109   : VKVKLSP---SFVTYDNPLVRATVSATYNYGKPVKGTVTLTVIP---KHR-TTSISVRPLDSYQTILPLNKEVTHE--INVRRVLNLIT----------DNLKREIEFMAIVEEELTGRRYNGSNSIFIYNDPVKLELIKT---SQSFKP :  341
HaadiTEP/CD109-1 : VQVSLP----PYITYSKPEFVATVNAKYTYGKNVKGRVKLIVKP---SLR-YGYLANE-NKPVVTEAEIDGSIDIP--VNVVRDLGLKE----------DTLTLEIDVIAEVEEYLTKRKYNASSVIKVYDKEIKIDVVKS---AESFKP :  349
HaadiTEP/CD109-2 : VEIDLP----VYVTYNKSDVVAVIKAMHSYGKPVKGEITLTVAP---RTR-YNKLTVRPYESFQTKAKIDGVVEIY--LNLLNDLSLRT----------DFFRREIEFFVLVEEEETGHKYNSTNTMWIYDKEIKLELIRT---SETFKP :  354
ScsuiTEP/CD109-1 : VIIQLP----PFVTWDDTDVTAFIEARYTYGKPVKGQLTLKVTV---PFFWWTRD--SEKRIAVFNTSIDGKAEVK--LNIVRDLLLNQT--------YSFTR-EIKFHAKVTEALTGKQMNATNELQIHLSKYKLN-FSN---LNKFKP :  348
ScsuiTEP/CD109-2 : VNIHLP----KFIISNDTDLTAFVDARYTYGKPVRGKLVLNISD---YYCDWPCSY-WSVKPLSLNTSIDGKAELN--IN-LKDLEIPEW--------FRHGN-RIKFSATVTDQLTGRQMIGMNELRIFSSKYKLN-FET---PPRFKP :  350
ScsuiTEP/CD109-3 : VRVDLP----PFVTFNASDMVATVRAKYTYGKPVKGKVVLMVTP---LVR-SPKIRTYYTDPLRRTAEIDGKVDIH--FNLFSDLNLKD----------DYHR-MIRFEAIVTEAVTERRENATNTMGIFKYKHKVELVKL---SETFKP :  351
ScsuiTEP/CD109-4 : VIIHLP----TFVTFNESDVVATVEAKYTYGKPVKGEVTLNVTD---YYCQWPCIS-SAVKPFALKSSIDGKADMK--LNLVKELNLPDW--------YRYGSKRFTFIATVTEALTKRQQNGTNDLNLYSDKYKLN-FDT---PDSFKP :  347
NinoiTEP/CD109-1 : VTVRPP----ALAIFNDSEIIVGVDAKYTYGKPVKGKVMVNVTE---AYCRSPFSSYCGSTPVVVRTNIDGTANVK--IP-ISQFNFPDY--------YKNSQASLDFLAVVTEDLTGRTMNASAQGIIYSLREKIDIVQS---SNSFKP :  356
NinoiTEP/CD109-2 : VTVQLP----SYATFNDTSVIATVTARYTYGKPVKGELTLVMSP---KVH-SPVLQAWLIPPVRKLAMIDGIADVP--LDVS-EFRFED----------DYQR-DIRVEAIVREALTGQRQNASAILTLHRHRVKLELIKT---AETFKP :  347
NinoiTEP/CD109-3 : VTVQPP----SYITSTTSDFPVTISAKYTYGEPVIGTFLLNLTANGCYYINNQYNKCIYPTPYVVTGSLTGGTATV--TVKLSALNLGN---------LLKQGTSINVLGVVTEAGTQIPQNTTVPLQLYTYATQLLATGA---N-TFKP :  364
EpspiTEP/CD109-1 : VTVRPP----TLAIFNDTELIVGVEAKYTYGKPVKGKLVLNTTE---SYCRSPYSNYCGSTPVVVRTNIDGIANVK--IP-LSQFNFPDY--------YRNSLANLDFLAVVTEDLTGRMMNASAAGNIYSKREKIDVVQS---SNSFKP :  355
EpspiTEP/CD109-2 : VTVQLP----PYATFNDTSVVATVTARYTYGKPVKGELTLVMSP---KVH-SPVIQAWVIPPVRKLATIDGVADVS--LDLS-HFRFED----------DFQR-DIRVEAIVRETLTGQRQNSSAILSLYSHRVKLDLIKT---AETFKP :  352
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : VDIDIP----KYGLFDEGLTTAKIKAYYSVGVPVEGEATVSIFP---KYK-SGLLQPIFIRPVRQVVPIKGEVDVS--FNIAKELNIRG----------DYSR-EVVFDVEIKESRTDRVQNNTALYPMYRYDYKLEMVRT---ADAYKP :  351
CaspiTEP/CD109-1 : VDIDIP----KYGLFDEGLTTAKIKAHYSFGVPVEGEATVSIFP---KYK-SGILQPLFTRPVRKVVPIKGEVDVS--FNIAKELNIRD----------DYSR-EVIFDVEIKESLTDRVQNNTALYPMYRYDYKLEMVRT---ADAYKP :  367
CaspiTEP/CD109-2 : VDIGIP----KYGLFDEGLTSAKIKAYYPFGVPVKGEATVSIFP---KYK-SGLLQPIFIRPVRKVVPINGEVDVS--FNIAKELNIRD----------DYSR-EVIFDVEIKESLTDRVQNNTALYPMYKYDYKLEMVRT---ADAYKP :  251
HosaiTEP/CD109   : VTLQTP----LYCSMNSKHLNGTITAKYTYGKPVKGDVTLTFLP---------LSFWGKKKNITKTFKINGSANFS--FNDEEMKNVMDSSNGLSEYLDLSSPGPVEILTTVTESVTGISRNVSTNVFFKQHDYIIEFFDY---TTVLKP :  362
```

Supplementary Fig. S9 (continued)

```
AmspC3           : GLIVYIQVLVQFASKAPAP-G---------------------------VDVVFTVKTDDGVA---------LAITNNKRRSDNSGYVEVK------------------ISVP-NTKKLTVEVVTDDVKLKA-NQQGKGTFVMLPYASP-- :  456
HaadC3-1         : GLTSVFEADVTYVDGSPAG-G---------------------------VRTKIQAVADNGKK---------LFIAVSSAQSDENGKVSFE------------------IQPELHHKSVTITLETDDPRYE--GSQAKGIFEQHAYQSG-- :  457
HaadC3-2         : GHTSVIEADISYANGEPAE-G---------------------------ILTMIRANAGNGQA---------IEIKQDSAISDKEGKVSFE------------------VRPQLHHNEILVTLIVVDKEY-----LSEASFKQHAFQSA-- :  454
ScsuC3-1         : GVPYQVQVDVRLVNNQPVN-DR--------------------------IPVTVNARAKKGGN-------FQALGKKPDLRTDVQGRVMFQ------------------FDTDENFEELVIEVQTKDEAVGD--NQAKTTLSVIRYNTPV- :  455
ScsuC3-2         : GFPYQVQVEVLLFDNQQIP-KP--------------------------VTVEISATAFKKDYS------SRQIGKIAPEVTDEFGRAVFQ------------------FDTDGTDTRIVITAKTNDKNYEE-KNQANQTFTGYQFSSP-- :  465
ScsuC3-3         : GVPYQLQVDVHHINNRPLK-YK--------------------------VPVIISGSAKKSGN-------AVKTIFTKNLQTDLHGRVMFQ------------------VDTEDGFQELNIQVETADQEIGN--NQAKEKFVVVRYKTPF- :  461
PsfuC3           : GMSFDIAVEVENPDKSPAR-G---------------------------VKVVVNPGNVEGY--------TAANGLAMLTINSDAGSRELA-----------------INARTDDSLLTRERQATER--MVA-YPYKTTGNNYIHISADTT :  463
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : GMSFDVAVEVVNPDGTPAD-G---------------------------ISVVIKPGPVNGN--------TAANGMARLTINTMGGNNNLQ-----------------ITAETNDPRLSPERQAKAT--MVA-VPYQTSSNSYIHIGVDAA :  412
HosaC3           : GMPFDLMVFVTNPDGSPAY-R---------------------------VPVAVQGED------------------TVQSLTQGDGVAKLS------------------INTHPSQKPLSITVRTKKQELSE-AEQATRTMQALPYSTVGN :  451
HosaC4           : GAPFLLQALVREMSGSPAS-G--------------------------IPVKVSATVSSPGS-----------VPEVQDIQQNTDGSGQVS---------IPIIIPQTISELQLSVSAGSPHPAIARLTVAA-PPSGGPGFLSIERPDSRP :  482
HosaC5           : GIPYPIKVQVKDSLDQLVG-GVPVTLN------------------AQTIDVNQETSDLD---------------PSKSVTRVDDGVASFVLN------LPSGVTVLEFNVKTDAPDLPEENQAREGYRAIA-YSSLSQSYLYIDWTDNHK :  474
AmspA2M-1        : GLPYFGELNTKKPDGQPAS-DVLILVC---------------YRVSNTKECRNFTSDSNG---------------IVKFSIPPQKPEVLS-----ISVDATAPNYKPEYYNEN-KYRVKVRQPHSSKTLRP-WFSPSGSFMDVKPVRKVI :  465
AmspA2M-2        : GLPYHGQLHVTLPNGQNAP-DELIQVC---------------YEA-IVSSCRNFTSDSHG---------------IINFTVPPQGAEATY-----VKIKALAPNHPSEFYPNDRYAQIKIRQPSAENSLTP-WYSPSGSFLKLKPVLGTS :  462
AmspA2M-3        : GLPYFGEFVVKKPDKEPAA-DVLILVC---------------YEVNNTKECRNFTSDDKG---------------IIKFTIPPQKPEVVA-----ISVEATLALFESEHYNNQ-SYQEKLYQPQQQIPLAP-WYSPSGSFLDVKPVLDIL :  461
AmspA2M-4        : GLPYFGKLVVKHPDKQAAF-NALILIC---------------HKTKHLTECRNFTSDING---------------MIKFTIPPQKPEVQS-----LQISASAALYKSEKVGG--ELGRLIQQPSGEMTLKP-WYSPTGSFMDVKPVHDVI :  480
HaadA2M-1        : KMPYVGQLEAKNPDGTPAL-SEKIQICVTLEGK----------------QCLIFTSDKNG---------------LIGFAIKP-SPVSRD-----IRVEATTINYED-VHYSSTFWTRKLRKPTATMILSP-WYSPSWSYLQIQPTTEEF :  467
HaadA2M-2        : GLPYNGKLFVKNPDDTPAA-DEGIQLCYTVNKER-VVMDGMWKATRTVKFCQNYTSDDNG---------------VIEFVIPRQNTDSID-----INVEAKSLKYAKDNQKSGTHREGSLNQPQTSMSLSP-WYSPSGSFIQLQQVQETL :  485
ScsuA2M          : GLPFHGKIKVTTPDHNSAS-NISVEISLETRT---------KRGISQIVASESFKSDDNG---------------IISFITIPPVVSDTENVIIKAKILPVSSDSGPSDHDHSYFLPNHNFGPTSTISAPI-WYSPSGSFLQIHRPADNL :  481
NinoA2M          : TLNYSGIVKVKNNDGSPAA-SVRLNVNFHSE----------GRTYETPKISKQYTTNEEG---------------IVDFAIHFPKESNDY---YTANIVAVDYYVSDSLT----------MHPTLSKTVKR-WWSPTQSYIQIEKHREPL :  452
EpspA2M          : GLQYKGKIQAEYHDGSPAE-NKNLNIELQAYN---------AENTKLATEIINVKTDLLG---------------MSFFDLFSLPQNSEK---ITLKVTATDYKTQTSMSGMSYSNF---IQPPFYATIKS-WYSPSNSFVQIDQQLQSL :  484
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : GLFYSGEIQLSDSERIGLG-QRKFELCVEYL----------------KKNCANFTSDADG-----------------KFRYVIQPVDEHDFNSTRQWLDMKIKLFKENGS--------VSRRSYSSVQVWH-SNAYKFIRFLKQEDGSKL :  471
PsfuA2M-1        : YTSNDKSGERLKIDLITKF-AEAPILRLDLK-----------------EDLTIKFKIKSAN---------------YEFEKKILKGKSELK--------IPISENSFKVIG----KNKFDEIVFN----EYFKCDWDKLNYKLEKIGGEC :  469
PsfuA2M-2        : TAYKNFYGLIHFKTLKMGA-YRLELISSKNK------------------DLIWKSSFTYS-------------KDLKTFFVKQLNEQAYLH--------IHFNNKLIEKR---HISVSSVKFMIN--------YDESQMEKCVENFKVKI :  423
PsfuA2M-3        : EYPYVIFFKVVEHGTNQALKGEKVEICSILWKDITQFTQFVSR--SNFYSFDEFHLYDLGQR-------------LIDIKFQETCINETTSEV-DGSLTLGIPLTGLPANVTKMSVKITALNFKANETTRM-KVSSDKFDVTLTHLSNDT :  656
CaspA2M-1        : EYPYVVFLRLVEHGSQKPLKGEKVEVCSRLWRDLNNFQNLISS--SQFYSFDEDQYYDLGKK-------------LLNIKFKDTCLNLTTSVD-DGSLTLGIPLSGVPLNVTKLGIKVTALSFMANETTRM-EMTSGKFDVSLTHLD-KA :  671
CaspA2M-2        : EFPFVMYLKLQEHVTKKPLKGANVEICSRLWRDVKDFTQHINS--RQFYSFDEDDYFELGKK-------------LINIKFKETCINQTTSEE-DGSITFGIPLNGVPINVTKLSVKVTALDFEANQTTRM-KLTQNKHDVILTHLSNDT :  645
PaspA2M          : VDSLMFSISVFWIILILLLKDRKLDYVPDFFHDIESFRSFIQDKYRNLYSADEDQFYGLGKV-------------LVKILFKETCINVTTSSD-DGSVKLGIPLSGLSTNVTKLSLKVTALEFHANETTRM-KNTESKHEVILSHIGNDS :  652
HosaA2M          : GIPFFGQVRLVDGKGVPIP-NKVIFIRGNEAN-----------------YYSNATTDEHG----------------LVQFSINTTNVMGTS-------LTVRVNYK--DRSPCYGYQWVSEEHEEAHHTAY-LVFSPSKSFVHLEPMSHE :  467
AmspiTEP/CD109   : GLVYKAFLKVSRQDNTPLN-LPNG------------------------ALTLKYAYNYK-----------PGSTRS-DRYRIPTNGLIELN--------FFPPLSKD--TVTIFTKADFNGKEYDLAYVDK--AYSPSNTYMQITLNTPF :  442
HaadiTEP/CD109-1 : GLKYTAHLKVCYQDDTPVT-VTNGD-----------------------QIVLKYGYSYD-----------ERNWDS-RRIAVPNNGLLSVD--------FFPPLNPN--ITSFGMSAEFRGYQYHLGNIEA--AMSPSNSFIQVSLRTDN :  451
HaadiTEP/CD109-2 : GLKYTAFLKVAYQDDTPVS-DSRG------------------------QLELKYGYSVR-----------EDEWKT-ELYTVPRNGLIKLE--------FLPPNQDG--VNFLNMRAVYHGHIYYLDRTDA--AQSPSGNYIQAILVTQN :  455
ScsuiTEP/CD109-1 : GLPYTVYLNLVLQDGTPFS-D-NVH-----------------------LITVNCTFHKKDIWDRFSVYDHSKNVMNTMNFSVPIDGRVKIQ--------ILPPKSSK----SITFKAKFLDVFTVSLARRA-PMHPNSDRFIQLLPPLTV :  460
ScsuiTEP/CD109-2 : GLTFTTYLIVRLQDDTPVV-DDEVN-----------------------KVKVTYSFAHH-------------GKVHEMDLPISKNGKIKIE--------FVPPESAS----MISLEAHYKDVNSYSRIDKS---YSLSNKYIQIVSNVEK :  448
ScsuiTEP/CD109-3 : GLKFSVIFKVATQDDIPVM-DEIN------------------------PLIIRYGYTYE-----------ESIYDR-LELKIPKNGTVPLD--------LYPPFADN--VNQIIIIAEYKDVKQQFPPIRR--AESPSNTYIQAVLTTLQ :  452
ScsuiTEP/CD109-4 : GLLFTTYLNVLLQDGTPII-D-DVN-----------------------NVTINYFYSWN-------------EASQTLKFPVPKDGKIKLE--------LVPPESAE----IIRLSASFIEASSYSTVNRA---QSLSERFLQLSLITEN :  444
NinoiTEP/CD109-1 : GLPHTYKIKLQLQDGTPVT-N-ADA-----------------------QLTVKTSSSYN-------------KPDAVTNYTIPANGVVTVD--------AFPDEDAD----FLRLNAEYKQVSSSAFANKA---QSISNKYLQLSLHQPK :  453
NinoiTEP/CD109-2 : GLKFTAYLKVATQDDMPIR-DDVN------------------------MVKVRFGYNYN-----------ESTHET-KEYRIPRSGIIQLD--------FYPPRSPE--TNVLVILAEYLDVKQPFPGIES--SLSLSHNYIQAFLLTEN :  448
NinoiTEP/CD109-3 : LLPYTITISLTYPDNTPVV-NPTGQ-----------------------GIISITFSDTN--------------TPNNQTFTFGQDGTATIT--------VTPPLDAQS----LNYQANYTNLTSYGSASAA---AALSSKFISITATGQN :  461
EpspiTEP/CD109-1 : GLPHTYKIKLQLQDGTPVT-K-ADS-----------------------LLTVKTSSSHG-------------KPEVVTNYTVPASGIVSVE--------AFPDEDAE----FLRLNADYKDVSGSAYANKA---QSISSKYLQLSLHNEN :  452
EpspiTEP/CD109-2 : GLKFTAYLKVALQDDIPIR-DDVN------------------------KVRVKFGYNYN-----------ESSHEV-REYQIPRNGIIQLD--------FYPPRSPE--ANVLIILAEYLDIKQPFPGIES--SLSLSNNYIQALLLTDN :  453
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : GMPYTAYIKVAKQDDTPIVDKNVP------------------------IQLKWGFSNEPSN----------E-YNT-TTIYSDKNGIVTLV--------LDTPSVND--TDAVVLGIEASYKNLTQWFSTIPRAESRSGLYLRSNLATKN :  455
CaspiTEP/CD109-1 : GMPYSAYIKVAKQDDTPIRDKNIP------------------------IKLKWGFSNEP-----------SEDYNT-TTIYSDSNGIVTLA--------LETPTVND--TEAVVLGIEASYKNLTQWFSTIPRAESRSGLYLRSKLVTTS :  471
CaspiTEP/CD109-2 : GMPYTAYIKVAKQDDTPISDKNAQ------------------------IKLKWGFSNEP-----------SEKYNT-TTLLANKNGIVTLI--------LDTPDVND--TDAVVLGIEASYKDLTQWFSTIPRAQSRSGLYLRSKLITKN :  355
HosaiTEP/CD109   : SLNFTATVKVTRADGNQLT-LEER--RNNVVITV-----------TQRNYTEYWSGSNSG--------NQKMEAVQKINYTVPQSGTFKIE--------FPILE---DSSELQLKAYFLGSKSSMAVHS---LFKSPSKTYIQLKTRDE- :  475
```

Supplementary Fig. S9 (continued)

```
AmspC3           : --NEAYLGIKPLDASRKLQKG-----VQFDGTVIVNPWQHVNRVMFSVISRGKVLVTGKTATKDGK----------------------------------------------------NTLERTFTFKVTEQMVPSIRVIVFANY----- :  542
HaadC3-1         : --QDAYIAIARSSP-QKLKPG-----HMYGKEIHFHPAG-IQDIYYMVTARGKILSMNKLPSG-------------------------------------------------------AYKNQRVEFQIGHDMVPSFRVVVFAHH----- :  538
HaadC3-2         : --TNSYIALDRMS--QTLKAN-----DTFKRIIHYDPIG-LSNIYYLVISKGRIVMMNKLPDG-------------------------------------------------------KYGDQHVSFKITYDMVPNFRILIFAHH----- :  534
ScsuC3-1         : --SKTYVWIAAPHEGTLFQVG-----KTFQTQVTVYPQERQMKLMYMVVSRGKILMMNETETRG------------------------------------------------------EAVVRTIQFPVTVDMSPSFRLIVYFIK----- :  539
ScsuC3-2         : --SNSFIWLQSPKEGLRFKVG-----KTLHTTATLRSSDASQRLYYMVVNRGNILAMNQTEAQ-------------------------------------------------------SAVIRTLHVPITTEMVPSFRLVVFVIE----- :  548
ScsuC3-3         : --HNSYIWIKVPEHGRYFQVD-----KSFQTVVTVYPADEQTKLFFMVVSRGIILLMNETEAKG------------------------------------------------------KYLVRTIQFPVTEDMSPSFRLIAYIIK----- :  545
PsfuC3           : ELEKGENFKINLNIQRAMTGK--------------------MDITYLIMSRGQLVKSGRHTLT-------------------------------------------------------NQVLVSLIIPITKEMLPSFRVVAFYHTNDNK- :  537
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ELKLGNNLKVNLNLNNN--AN--------------------RDITYLILSRGQLVKSGRYKTR-------------------------------------------------------GQVLISLIVPVTKEMLPSFRIVAYYHTNNNE- :  484
HosaC3           : SNNYLHLSVLRTELRPGETLN-----VNFLLRMDRAHEAKIRYYTYLIMNKGRLLKAGRQVREP------------------------------------------------------GQDLVVLPLSITTDFIPSFRLVAYYTLIGASG :  542
HosaC4           : PRVGDTLNLNLRAVGSGATFS---------------------HYYYMILSRGQIVFMNREPKR---------------------------------------------------------TLTSVSVFVDHHLAPSFYFVAFYYHGDHP- :  553
HosaC5           : ALLVGEHLNIIVTPKSPYIDK-------------------ITHYNYLILSKGKIIHFGTREKFS------------------------------------------------------DASYQSINIPVTQNMVPSSRLLVYYIVTGEQT :  551
AmspA2M-1        : SCDEEVPLDVVYTTN-----------------------GEDITLNYQVMSRGRIVDYGMKSYKFSLDDYKEDSDVIRNKTEGSSRTKRTIIVPDIDKTP-------------KYRLPRHIGKFSLPIKIKAEMAPIARVLVYYIRSD-G- :  577
AmspA2M-2        : SCNEEVPLDVFYTTS-----------------------GEDIQLHYQVMSRGRIVTHGMKSYKFNENDYQDDSYTIQN-ENVNERQKR---FYSEKVN---------------YTLPKHIGKFSLPIKISAHMSPVARVLIYYIRTD-G- :  568
AmspA2M-3        : SCDTEVPLNIMYTTN-----------------------GEDIILNYQVLSRGRIIDYGKKSYKFNIDDYNEDHSVIRNETEDSGKVKR-SIDPSINVN---------------YSLPQHIGKFSLPIQIKAEMAPITRVFVYYIRPD-G- :  570
AmspA2M-4        : SCGEEVPLEVLYTTS-----------------------GEDITLHYEVMSRGRIVDYGKKMYKYNANDYKEDAFVVRNTMNISNEKKE---LTNETVN---------------YILPGSIGKFHLPIKMKAEMSPIARVLVYYIRSN-S- :  587
HaadA2M-1        : ECDKAQKVTVQYTAER----------------------GSTIKFYHQVLSKGRIVQQGSHQRTFYSKVEEIQYDFEEQIVKKGSSSGNSS--------------------------ETEIGEFILAFDTRATMSPISRLLIFYVRDD-K- :  567
HaadA2M-2        : LCGTKNTLKVLFTSAE----------------------DEDYTFYYQVLKQGRVVKKGSIEKSFSTKDDVADLYEDEYKVIDDVEMQIVPPVDQSEIPKSSEPKEEECKSAKEARYVPPVGEVDIDLDIDASWSPTFHLLVYYIRDD-R- :  611
ScsuA2M          : PCNGEYSFTVFYSELS-----------------------QNIHLSYQVISRGNIIYQSSKLQILEENSLNFSNSWHLLPSRNKLSFNF----------------------------------FTIPIPLTPKMTPVSRLLVFYVRDD-G- :  572
NinoA2M          : QCVRMFSAFVNFRPLTSD-------------------VDGTLVIYYLLISR-NDITQ---------NKLKFEVSNGLLRDTSTTSF-------------------------------------SIPIVINPDMSPKFHLIVYYIRED-G- :  534
EpspA2M          : QCNELNLLTVSFANIN----------------------QSEIIVFYQLISRGNILSQ---------NQFQLSFNN------FTQTY-------------------------------------QIPIMVEPKMAPKLQILVYYIRYD-G- :  558
VaspA2M-1        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : KCNRTYQTDLYYSPPP-----------------------ETVTASIAVMSSGRIVRTTSILLSPTDSKASVPTADHVISLDNQTNIPDVL-------------------------------HQTFPLEIGTDVSEKFNVLVHYTLPD-G- :  565
PsfuA2M-1        : SNVFKFKQFGRKKNIK----------------------VMLSYKKYAILDEDESNFEEKCINEDCFKIFEDEIQLEN---------------------------------------------TFIFDNKFYGINSQAYDKSFEVIYDDK- :  551
PsfuA2M-2        : YSIHKIHQCMQFLTYN--------------------------ELITLKPMLVASNSNMYLY-----------------------------------------------------------TVNISLKHFRHIQLNGTLLFADFTS--QK- :  485
PsfuA2M-3        : IGRDAHRLTIREKLIEERKVKFDCEGASNFTVYFQGKLDETVDLNYVVSSGGSLISSGNHQVLVD-SNDTAQYLDGLKQYDELGYLNSSVQS------------------SDDHVLKSYVISLTRPFPMEGKVTKKIKLLVYTRDIETG- :  786
CaspA2M-1        : EGADSHRLTIREKLVEERKIKFDCEGASNFTVYFQGKMDETVDLNYIVSSGGSLIASGNHRVLVD-SNDTMKYLDGLSLYQELGHVNSSSEVP-----------------SGKNVLKSHVISLSRPFPLEGKVTKIIKLLIYTRDIETG- :  802
CaspA2M-2        : IGKDAHRLTIREKLIEDKKVKFDCEGTSEFTVYFQGQMGSSVDLNYVVSSGGSLITSGNKLVLVD-SNDTSKYLDGLVQMDGLGYVNSSIDA-------------------SNNVLKSHVISLNRPFPNEGKVTKMINLLVYTRDLESG- :  774
PaspA2M          : ALNDLHRLAIRERLIDDTKIKFDCEGTTTFTVYFQAEMGDVVDVNYVISSGGSLIHSGVKRVMVN-SNDTDKYLEGLISEDSLGYVNLTDSP--------------------ASVLKAFVLSVDRPFPNEHKITKLLKILVFTRDVNSG- :  780
HosaA2M          : LPCGHTQTVQAHYILNG----------------GTLLGLKKLSFYYLIMAKGGIVRTGTH--------------------------------------------------GLLVKQEDMKGHFSISIPVKSDIAPVARLLIYAVLPT-G- :  549
AmspiTEP/CD109   : ---PQVDQEVEVLVNCTSQLPQ---------------------YVYQVIARGNILRTRSVRPPG-------------------------------------------------------GNSHSFKVQMTDNMAPLVRIVVYFTRDD-G- :  511
HaadiTEP/CD109-1 : ---PTVDKEVELEINATEPLNQ---------------------LVYEVLGRGDIVLAGAIDIPN-------------------------------------------------------VKSYRFSLPVSYKMAPKARVVVFYVRKE-NN :  521
HaadiTEP/CD109-2 : ---PKVLENIEIEVNATEELNH---------------------LVYKVIGRGNIEIGRTIPVPN-------------------------------------------------------QKEYRFSFRAPSSLAPQARVLVYYVRAT-NN :  525
ScsuiTEP/CD109-1 : ---PQIGEKIELLLQSTHTLDD----P----------------LILQVMGRGKILLTNNTAGNK-------------------------------------------------------SKIQIIRFTVTEEMAPAVHVIIFSLTNN-G- :  530
ScsuiTEP/CD109-2 : ---PKIGDEVEFFISSTRPLED----F----------------LTVEVIGRGKILYTHNIPANG-------------------------------------------------------QKLETFKLKLTKEMAPEIRVIVHYVTSC-G- :  518
ScsuiTEP/CD109-3 : ---PKANEEVIITVNSTVRLNT---------------------FNVVVMGRGDIVFAETVNAGG-------------------------------------------------------ERSIKVKFFITRAMAPNIRFIVYYTDLS-G- :  521
ScsuiTEP/CD109-4 : ---PKIGDEVELLVNATKNLED----P----------------LILEIIGRGKILHTENIPGSK-------------------------------------------------------TNHQKISFKLLPEMAPKIRVIVYYTTPC-G- :  514
NinoiTEP/CD109-1 : DKDPKVGETVQLDINGTFYISR---------------------LEYEVVARGKIVASGSFKFDKD------------------------------------------------------AKSHSFPLNLTQDMAPRVRVVAYHVSSC-G- :  526
NinoiTEP/CD109-2 : ---PRVADEVEVQINTTERIMN---------------------FVFQVYGRGNLALAQTIPMNN-------------------------------------------------------EKYVRFKFRVTPRMAPKARFLAYYARPD-G- :  517
NinoiTEP/CD109-3 : ---YTVGQSLTVSVLASFTMQN----Y----------------INYLVVGKGLLLAAKTLPIPNS------------------------------------------------------GTQFSFNFPLTYEMAPSAAVAVFYITDC-G- :  532
EpspiTEP/CD109-1 : EIEPKVGDVVQLDVNGTFYISR---------------------LDYEVVARGKIITSGSLKFDKD------------------------------------------------------AKSHSFPLNITQDMAPRARVVAYYVSSC-G- :  525
EpspiTEP/CD109-2 : ---PRVADEVEVQINTTERIMS---------------------FVFQVYGRGNLALAQTIPMNN-------------------------------------------------------EKYVRFKFRVSSRMAPKARFLAYYVRGD-G- :  522
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ---PKVGETISIDVSSSENIDS---------------------ITYLVFGRGKLALGITRKASG-------------------------------------------------------TMENQISFRATSDMSPVCRVIVYYITSISG- :  525
CaspiTEP/CD109-1 : ---PKVGETISIDVISNEGIDS---------------------LTYLVFGRGRLALGVTRGASG-------------------------------------------------------TKENQVSFRATSDMSPICRVIVYYVTPISG- :  541
CaspiTEP/CD109-2 : ---PKVGETISVEVTSSENIDS---------------------ISYLVFGRGQLVLAITRKASG-------------------------------------------------------TNENKISFRATSDMSPICRIIIYYVTSISG- :  425
HosaiTEP/CD109   : --NIKVGSPFELVVSGNKRLKE---------------------LSYMVVSRGQLVAVGK------------------------------------------------------------QNSTMFSLTPENSWTPKACVIVYYIEDD-G- :  540
```

Supplementary Fig. S9 (continued)

```
AmspC3           : KGHLIADSIWMDVEDTCQHEIETSITPQFDLK----RPGDEGSIVIKA-AKKTTV-GLLSVDKAVYVLRNK--GLLSASKVYKTMEGYDIGCGAG------------------------------------------------------- :  629
HaadC3-1         : KDELIADSLKIDVERECNPEVQVAVTPEFGEK----EPGNNGKIVIRG-TKGTYV-GLLGVDEAVYALSKK--DILTKAKVFNKLATHDLGCGPG------------------------------------------------------- :  625
HaadC3-2         : DNQLLVDSLKFDVENECNPKAEVTIHPEYSIQ----EPGNAGKIIIRGNSKGTFV-GLLGVDEAVYALSKK--DILTKAKIFKRLASHDLGCGPG------------------------------------------------------- :  622
ScsuC3-1         : NGKIIADSLNIDVERTCKYNNGK-FSVTADSSSGFFSPNQEVTFKITG-EPDSVV-GIGAVDEAVYLLNDR--DVLTRDKMFKELSKHDLGTGPG------------------------------------------------------- :  629
ScsuC3-2         : KYYLVADSMQFEVERICKYNEGK--GLIIEASTPLASPGENINFKIKG-EEESYV-GLLGVDEAVYVLNKQ--DLLTKEKMFRELRNHDLGKGPG------------------------------------------------------- :  637
ScsuC3-3         : DNKIIADSVQIEVERVCKYNGGKGFSIKTDRKAGIAVPGAAVNFIITG-EQDSFI-GLSAIDEALYFLNNR--SVFTKEKMFREIQKYDLGFGPG------------------------------------------------------- :  636
PsfuC3           : ---VVSDSIWVDVKDSCMGTL--RLEP-VRPAPSF-EPRRMFGLRVTG-DPGATV-GLAAVDKGIFVLNNK--HRLTQKKIWDIVEKYDTGCTPG------------------------------------------------------- :  621
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ---VVSDSIWVDVQDDCMGSL--RLEP-TRPVPSY-EPRKMFGLKVIG-DPTATV-GLVAVDKGVYVLNNK--HRLTQKKVWDIVEKYDTGCTPG------------------------------------------------------- :  568
HosaC3           : QREVVADSVWVDVKDSCVGSL--VVKSGQSEDRQP-VPGQQMTLKIEG-DHGARV-VLVAVDKGVFVLNKK--NKLTQSKIWDVVEKADIGCTPG------------------------------------------------------- :  630
HosaC4           : ---VANSLRVDVQAGACEG----KLELSVDGAK-QYRNGESVKLHLET-DSLALV-ALGALDTALYAAGSKSHKPLNMGKVFEAMNSYDLGCGPG------------------------------------------------------- :  638
HosaC5           : -AELVSDSVWLNIEEKCGN----QLQVHLSPDADAYSPGQTVSLNMAT-GMDSWV-ALAAVDSAVYGVQRG--AKKPLERVFQFLEKSDLGCGAG------------------------------------------------------- :  637
AmspA2M-1        : ---EVIAAFATLDVMPCFLN----KAALTFEK--KSIKPGMSAKYKIAAT-AKSLCAVGVVDKSTHLLKTS--NQITGDKIFKILKGFDSNRNTRPKFINTDEYCRSKPKE--------------------------------------- :  676
AmspA2M-2        : ---EVVAASTSLDVMPCFAN----KASFTFEK--DSVKPGEAAKYKISAA-PKSLCAVGVVDKSVHLLKSD--NQITSEKIFKVLKSFDTGRYTYPTLIDDSKYCKERLEGKIPTESPFDSTTQRPEPINPFGFQGLRLDELVSSTSSEA :  706
AmspA2M-3        : ---EVVASYTTLKVMPCFVN----KASFTFEK--KSIKPGVSAKYKIRAS-PKSLCAIGVVDKSSNLLKTG--HQITAERLFEIMKAFDVNVYNLPVVANNQAYCQEKYKDS-------------------------------------N :  671
AmspA2M-4        : ---EVISAFATLDVMPCFLN----KVSLNFEK--NSIKPGTLARYKISAS-AKSLCAVGVVDKSVHLLKSS--NQITVEKIFNILKSFDAT--------NDNVHCRR-------------SRRRR------------------------- :  679
HaadA2M-1        : ---EVVADSRKFRIKKCLQN----KVSLHFRH--EQQYPNTEATMLLSAS-PSSLCGIHMVDKSIRLLEDD--TTFNTDKLFKIMESYDTGKDPTEFPGICIEDSKEDKP---------------------------------------- :  665
HaadA2M-2        : ---ETIADSQKFNVEKCFKN----QVKLQFGD--DVKQPGTKTSIRVTSS-PNSLCGLKVVDKSVALMNSE--DQLTPEKVFRALESLDTSMYYGINHCN---EKIRQPGL--------------------------------------- :  707
ScsuA2M          : ---EVVADSLKFEIEKCTEN----DVSLNFDT--QRVIPATHTEISIKAS-PYSLCAVGIVDRAVHFLRAN--NQLTLTKIFNGLSAFDITKDSLPEQSK-VKYCHQHFDSYP------------------------------------- :  672
NinoA2M          : ---EIVSDSETFDVEKCFSN----EVRMSFNE--EKILPGSPASIHLSAD-ANSLCALSVSDTSVDLLTT---------RDFSEEKVFHTVINNLPSMKVPETPCHHHQRTSG------------------------------------- :  628
EpspA2M          : ---EIVSDSETINVNHCYSN----EVRMDFKD--EKILPGSLSSLSLSAE-TGSLCAVSVTDKSIELLSG---------HVFDGHKVFSMIDDKLPQINN-NNQCPQNNHPR-------------------------------------- :  650
VaspA2M-1        : -----------------------------------------------------------VVDASVRLLSSN--NQFNKDNLLNSNPSLHLNQSSAPIQQNDFEYC--------------------------------------------- :   44
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : ---EIVAAGTTFEVETCLEH----EVRLKWSK--DKPSPGEDINLQIDAA-PLSVCGISVVDRSVKLLSSN--NQIKKENLLKRRSWLHTYQYTYPPQQRDWEYCRKKNEAEMEK----------------------------------- :  668
PsfuA2M-1        : ------CLTKYQLDSSESK-------GITFSENN--VKPGQSIDLKMELSKKPDYCLIASYDKNLVNLVKKEQIAIYLKTIKDSLKNYLIGNLG-------------------------------------------------------- :  630
PsfuA2M-2        : ---ITIKSLNNCLIRSQTN------FGKIYFNKS-KVYPGDALKLNANFIEKPKLCFINSFDQNLENMDVKQVGTLYSKYFSILINYSNRSLGMN------------------------------------------------------- :  570
PsfuA2M-3        : ---KTLSHVREYDSESCAPK----PKLEWSPAE---TNPGHSVSLKIKYK-PKSLCAYSVIDKSADLIENP--NKITTDSIQQL-KENLASKRIVYDQVFPP-EC--------------------------------------------- :  876
CaspA2M-1        : ---ETLSHVREYETESCAPK----PVLEWSPEE---TNPGKEVSLKIKSK-PQSLCAYSVIDKSADLIENP--NKMTSERIQQL-REDLSARRIPYDQVSSPSEC--------------------------------------------- :  893
CaspA2M-2        : ---KTLSFVRKYESESCAEK----PKIEWNRHE---SNPGHKVSLNIKYQ-PKSLCAYSVIDKSADLIENP--NKITSVKIQEV-REKLAEKRIVRDGVSSD-KC--------------------------------------------- :  864
PaspA2M          : ---STLSSIREYETESCNAK----PSLSFSPSK---TRPSDLVSLSIGSK-PNSLCGYSVIDKSADLVENP--NEITSKKFQNI-KEDLAKKRIITEGIYGD-RC--------------------------------------------- :  870
HosaA2M          : ---DVIGDSAKYDVENCLAN----KVDLSFSP--SQSLPASHAHLRVTAA-PQSVCALRAVDQSVLLMKPD--AELSASSVYNLLPEKDLTGFPGPLNDQDDEDC--------------------------------------------- :  642
AmspiTEP/CD109   : ---EIVADGLSLDLEKIFEN----QISFTAGPG--VLRPREKVRISLNTD-PNSMVGLMGIDQRNLVLDPG--NDITQNDVIRSLEGFDSGKKDSDQQM--------------------------------------------------- :  598
HaadiTEP/CD109-1 : ---EIVADAINFDVAGVFRT----PVSVKADVK--DTKPGGLVNVSVETK-PNAVVGLLGIDQSVLLLKSG--NDITQNDVITELETYDGGKKKK-Y----------------------------------------------------- :  605
HaadiTEP/CD109-2 : ---EIVADSVSFDVEGLFRT----SVTVSSNVK--EVQPGRQVNLRLQTT-PNSLVGVLGVDQGILKLKSG--NDITLPEVIEDLETYDGGQRTK-Y----------------------------------------------------- :  609
ScsuiTEP/CD109-1 : ---LVMADSILFGVKGLFKT----PVSVNVVPK--SAKPGTTMEVSVKTN-PHAFVALSAVDQSVLLLNKG--NDLKSSRVLSKLRNIESSNALKFYDD--------------------------------------------------- :  617
ScsuiTEP/CD109-2 : ---EVIAEGLNLGVEGVFKT----LVELNVDPK--SAKPGAPMEVSVKTN-PNAFVGLSAVDQSVLLLRKG--NDLTTGELLSDLRKYEIGSQYQFYYG--------------------------------------------------- :  605
ScsuiTEP/CD109-3 : ---EVVADGISFEVEGVFQN----YVNIESNSK--DVKPKDTVNLQISTN-PNSFVGLLGIDQSVLILKTG--NDISQQEILYQLDEFDPGKQPY------------------------------------------------------- :  604
ScsuiTEP/CD109-4 : ---EVVADAIDFGVEGIFKT----PVKVNVNPN--STKPGSEIDVSVQTN-PNAFIGLSAIDQSVLLLKKG--NDITTKEVLTDLQNYEIGDRSPFQFD--------------------------------------------------- :  601
NinoiTEP/CD109-1 : ---EVVADSLDFTVDGVFQT----PVALHTSEN--RTKPGAPMEVTVSTL-PNATVGLLAIDQSVLLLKTG--NDLNRNEIVNDLNDYESGWKPSPYD---------------------------------------------------- :  612
NinoiTEP/CD109-2 : ---EIVADSLNFFVEGVFQT----PVIVGVSAN--RTSPGSPVEVRVETK-PNAYVGILGVDQKVLLLKSG--NDITRDDVLKELVSYDSGATKS------------------------------------------------------- :  600
NinoiTEP/CD109-3 : ---EIVGDSFNINIDGAIRT----KVQLQLSTP--ETTPGSNITITVDTA-PNSVVTMTGVDQSVLLLNAG--NDVTVTEIANQLAQYQVQQSGP---I--------------------------------------------------- :  616
EpspiTEP/CD109-1 : ---EVVADSLDFTVNGVFQT----PVGLHTSEN--KTKPGAPIEVTVDTL-PNSTVGLLAIDQSVLLLKSG--NDLNRNEIINDLGDYESGWRPSAFD---------------------------------------------------- :  611
EpspiTEP/CD109-2 : ---EVVADSLNFYVEGVFQT----PVAVGVSAN--RTGPGTLVEVKVNTK-PNAFVGILGVDQKVLLLKSG--NDITRDDVLKELVSYDGGADK-------------------------------------------------------- :  604
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ---EIIADSMDFEVEGILTN----FVQIASTKK--ETFASSDVTINIKSK-PNSFIGILAVDKSVRSLKGG--HDVLLKEVTDELRRYDTANTPD------------------------------------------------------- :  608
CaspiTEP/CD109-1 : ---EIIADSMDFEVEGILTN----FVEISSTKK--ETYATSDVTINIKSK-PNSFIGILAVDKSVRSLKGG--HDVLLKEVTEELRRYDTVKTPN------------------------------------------------------- :  624
CaspiTEP/CD109-2 : ---EIISDSMDFEVDGILTN----FVSISSTKK--VTSAANDVTINIKSK-PNSFVGILAVDKSVRSLKGG--HDVLLKEVTDELRRYDTANTPN------------------------------------------------------- :  508
HosaiTEP/CD109   : ---EIISDVLKIPVQLVFKN----KIKLYWSKV--KAEPSEKVSLRISVTQPDSIVGIVAVDKSVNLMNAS--NDITMENVVHELELYNTGYYLGMFMN--------------------------------------------------- :  628
```

Supplementary Fig. S9 (continued)

```
                                                     β-α cleavage site                                                                       /       ANA domain
AmspC3           : ---GGQNSANVLEKSGAIILTSSTITNEKRDHFSCN--DHSRPKR-ALEEHSAKYTDAFTK-------------------------------------------------------------ACCIMGQRPTKFSVSCQVR------AQI :  706
HaadC3-1         : ---GGITVNSVLGNAGVTIG--TRVFSPTESHS-CVEIK--RRKREAAMDIVKTYMG-TDR-------------------------------------------------------------YCCSLGLSEDKYRRTCQER------SNV :  699
HaadC3-2         : ---GGLTINDVLGKAGIILG--TNVYAPSHSYS-CND-RRKRWKRDLQWDIVEMYKG-FEK-------------------------------------------------------------ECCSLGLESDTYGRSCKER------SNA :  697
ScsuC3-1         : ---GGINPAVVFKNAGILMLSNNEIGEHGRKEGITQ--PKERRKRSLLEKVEEYSGQ--AA-------------------------------------------------------------ICCKFGQFEGPVDMDCADR------AAM :  705
ScsuC3-2         : ---GGISTEAVLRDSGIIILSSVYIGEHGREESLIQ--SQSRKKRSLPDKVNEYSGK--AA-------------------------------------------------------------ICCRMGQFEGPQHLNCTSR------ATM :  713
ScsuC3-3         : ---GGIDPAAVFKNAGILILSNSHIGRHGRTEGIDH--SHHRMKRSLQSKIDEYFGN--AA-------------------------------------------------------------ICCRYGQFEGPKGMNCTTR------AAK :  712
PsfuC3           : ---GGKDSMSVFFDAGLLFESDTASGTAYRQESKCPT--VERRKRATSRMDVITSLASEYDGT--------------------------------------------------------PVKECCVEGMKLLPVSYSCEVRS-------- :  702
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ---GGKNSMSVFFDAGLLFETSTAVGTPYRQELKCAA--PSRRKRSTTITDTITSLTSRYNDT--------------------------------------------------------LQRDCCRDGMADTPVSYSCERRS-------- :  649
HosaC3           : ---SGKDYAGVFSDAGLTFTSSSGQQTAQRAELQCPQP-AARRRRSVQLTEKRMDKVGKYPKE---------------------------------------------------------LRKCCEDGMRENPMRFSCQRRTRFISLGE- :  718
HosaC4           : ---GGDSALQVFQAAGLAFSDG-DQWTLSRKRLSCPKEKTTRKKRNVNFQKAINEKLGQYASP--------------------------------------------------------TAKRCCQDGVTRLPMMRSCEQRAARVQ-QP- :  726
HosaC5           : ---GGLNNANVFHLAGLTFLTNANADDSQENDEPCKE--ILRPRRTLQKKIEEIAAKYKHS----------------------------------------------------------VVKKCCYDGAC-VNNDETCEQRAARISLGP- :  722
AmspA2M-1        : ------------------------PRPTFIPFPIRGG-LR-RSKRRSMPFWRHR--PTTEYVDASMAFEASGLTLLTTNDVN-TRPCRRRYPEILFDG-----GIPIRR-KTVAFSAPN-VLLAQSG--PRPTPRPTPRP---------- :  778
AmspA2M-2        : PTTEAEEPGSETTSVAPTTEEPTSTEFVDKQPEGAPPSVF-RKKR-SSPPRFYSDSYSSSYVDALLAFESSGVLAMSDRSLE-SRLCSFFQRR--HLVYASPGFGG----------NGVQKFGIHDDE----KGAGVPSA---------- :  827
AmspA2M-3        : STRPPPIRPRIIRTIIPAPRPIPIPPRLVPRPVPPSQTVF-RPQ-YSYFWLNRYDQSKLKLVDASMAFETSGLTFLTSNDVN-TRPCS---PEDIYDGRYTFPPPPIRRPQPVRRPRPSPRPIPVAP--ARPLPRPLPAP---------- :  803
AmspA2M-4        : ----------------TRSDARIVRS----------VPFAPRPSRTEFAD---APMSFEFSGLTYLTDNAVGTLA----------PCRRIIRY--RKPVNRIQEDS----------RGRNQISARRGA----NRPSEIAE---------- :  764
HaadA2M-1        : -----------------------------------------RMPRNMIFQDTFPR-SGRPYVDARQAFEEAGMTVITDLKLK-SYHCT------------------------YYELPIPLMLPDSR--YEDTEP---------------- :  730
HaadA2M-2        : ------------------------------YSASSKY-LP-R--------PPQPWSS-SSYEDSLAAFENAGFLVISDLILF-TRPCKS---------------------RGGGGNIAYETGYGGA--VAMAST---------------- :  776
ScsuA2M          : -----------------------------------LHPFVDNS-------LIRVMEVESGYADAATAFDDTGVIIMSDLTFD-TRPCVDVTGIMALARTFAMPISLEFQ-------QPGPPGP--------------------------Q :  746
NinoA2M          : ----------------------------------GFYFKN------------------SNVMNSRKAFDTLGVVTLTDLSLD-DSSCKFVKFPNRVRHWGMHRYYDEMLIPGPVGASLSNRRMGASGGIPFKLASSTRGG---------- :  715
EpspA2M          : ------------------------------------YYIFN-----------------SAVTNSKAAFDDIGMLVITNLELK-LHPCKARAVVARPGESRQAGRRRGGS--------SRNREG--------------------------V :  712
VaspA2M-1        : ----------------------------------LEKKKESER--KSKSSIEESFYYSLNSADSITAFDIAGLTVMSDLDLE-TRLCPEGTTSVSAVEGNSTSFL------------PGDGGSILGI----------------------- :  122
VaspA2M-2        : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
VaspA2M-3        : ----------------------------------LESISNKNK------RSLFWGPYYGSSADSITAFDIAGLIVMTDLLLE-TRPCHTVHGPAILRRPVMAEVAPTVE---MALDAQPVDAEYYGGGA--------------------- :  753
PsfuA2M-1        : --------------------------------------------------------MSLMKDLDFLYETVLRSNQAIYMK---------------------------------------------------------------------- :  654
PsfuA2M-2        : ------------------------------------------------------------LYKPIIYYPSYKRPRPMRNQ---------------------------------------------------------------------- :  590
PsfuA2M-3        : ----------------------------------------------------------KDAMHLFKAIQSIGLYIMSDKLIQ-DPSCNTVVDSINFDPNESDY-SSKPVPVAFVSSLGAASPAIEMDS---------------------- :  944
CaspA2M-1        : ----------------------------------------------------------MDAKHLFKAMENIGLYILSDTLIQ-DPSCNTVIDSANFDGPEEED-PFKPAPEALFGST-QSAFAPGMST---------------------- :  960
CaspA2M-2        : ----------------------------------------------------------KNAEHLFQSMETLGLFVLSDKLIQ-DPACNSVVDSNNFFNNEGQYGQYKPTPIAFSSAA---APASGFSQ---------------------- :  930
PaspA2M          : ----------------------------------------------------------KNAQYLFQAAQQIGLFILSDTLVH-DLKCDTVVDSSGFGNINKDF-EYYDEPVAFVAGPPVQFQSVSLNASPP------------------- :  941
HosaA2M          : ----------------------------------INRHNVYIN---GITYTPVSSTNEKDMYS---FLEDMGLKAFTNSKIRKPKMCPQL--------------------QQYEM----HGPEGLR------------------------ :  704
AmspiTEP/CD109   : -----------------------------------ILQLPVRRGRALFY---------PGSLSAASVFDDAGVTVMSNGIVNSFDR-------------------------------------RKSKSAPPGE----------------G :  651
HaadiTEP/CD109-1 : -----------------------------------WP--YYRRKRSL---------WWPGSATAHDVFDDSGVVVLSNGLLYRFMP------------------------------------MIMYRSFRPEE----------------D :  657
HaadiTEP/CD109-2 : -----------------------------------RPPWFRRRRRSL---------SWPGSKSAGLLFEDSGFVIMTNAFLLNTGN-----------------------------------EVSTENVIRIDE----------------N :  664
ScsuiTEP/CD109-1 : --------------------------------------------QEG----KFLDP-----ITTLELFKDAGLLLLTNGYQMYSTP-----------------------------------FFWNRERFPEREMPVMPG----------R :  669
ScsuiTEP/CD109-2 : -------------------------------QRT-PLAYCLPPRPL-------------WAATTSEMFETSGLLLFTNGLLIPQRQ-----------------------------------SSYG----------YSPH----------Q :  655
ScsuiTEP/CD109-3 : ----------------------------------------------SKEDLYYNSVWFPGSATASEVFKEAGFITLSNAMIYSIFP----------------------------------YLMYRS---FAET----------------V :  655
ScsuiTEP/CD109-4 : -------------------------------DYVSPRHYSLRPWRPR-------------SSSTLELFTNVGLIFLTNGLLARYPY-----------------------------------SGYGGFAGGAGGISLRPA----------Q :  662
NinoiTEP/CD109-1 : -----------------------------------------RKKRSI---------WRPPGDTITQLFDNAGLVFLSNGLFQKQPE-----------------------------------YNYGYYRPDTRIRPMFAM----------A :  667
NinoiTEP/CD109-2 : ----------------------------------------------AFDDFYDRYLWSPGTITASQVFEDAGVVIVTNGNVFQYYP----------------------------------RILYRANTGFDSD----------------- :  653
NinoiTEP/CD109-3 : --------------------------------------------YFGSSFIPGIIRRPYPNGGLDQIFQACGLSYVTNGLYFKQQIFFG---------------------------------PLPMFKFAAGG----------------A :  673
EpspiTEP/CD109-1 : -----------------------------------------RKKRSI---------WRPPGDTVTQLFDTAGLVFFSNGLFQKQPD-----------------------------------FNYGPYP--VRFNSFGGG----------A :  664
EpspiTEP/CD109-2 : ----------------------------------------------KLDDFYARYYWSPGTITASQVFEDAGVVIVTNGNVFQYYP----------------------------------RLLYRSNIGPDSD----------------- :  657
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : ----------------------------------------------FFPWF-KIIQPKEGSLSWHTGSLNSEDTFSKSGTIIFTNGEL----EKGYSEEEESSNV----------IENEVLRTRAKNRPFG------------RPIPRPG :  685
CaspiTEP/CD109-1 : ----------------------------------------------FFPWF-RIIKPKDGSLSWHTGSLNSEDTFLKSGTLIFTNGEL----QKSYTEDD-STNA----------IETEVLRTRANNRPFG------------RPIPRPG :  700
CaspiTEP/CD109-2 : ----------------------------------------------FFPWF-KVIKPNEGSLSWHTGSLNSEDTFVMSGTIIFTNGIL----EKRYSEE--STNV----------IETEVQRTLAKNRPFG------------RPIPRPG :  583
HosaiTEP/CD109   : -------------------------------------------------------------------SFAVFQECGLWVLTDANLTK-------------------------------DYIDGVYDNA---------------------E :  659
                                                                                                   /            A2M bait region
```

Supplementary Fig. S9 (continued)

```
                                          /
AmspC3           : LKVHLTEKYPDKLNCADAFLSCCSSKSVIMARSGAN--------------------EALQFVDDIDAKESQSIVRSSFPESWLYVAEHMGESTEKK---------IATHLPHSITTW-VFQALSVSPNRGLCIAKP-QEVIVRQTVFLDV :  825
HaadC3-1         : VQKYLDG---EYPTCAKAFMECCIYGLRNGLADMK--VMRQGIALGRMGGENLEEPEFIDAEFEED-FEKQLTVRKDFRETWIYEDVTIGPDNREE---------LGVSLPHSITTW-VLQAVSVSPTHGICVAEP-QKLVSFKKIFLHL :  832
HaadC3-2         : VEHYLIDEE-EYINCSTIFLDCCLYKQRIELDSWRNEKMRSYMIVGKAFEDDDIGLIDIDPELEDD-MERKSTRRTYFPETWIFEDLTIGPDHKEE---------LRATLPHSITTW-VIQAVSVSPTHGICIAEP-KKIVSFQKIFLHL :  834
ScsuC3-1         : IYEKIGEKF----NCSTAFLDCCENKLKYMAAN--------PGRANQEVEDEKP----INELIEMVEADTLKHIRHIFPETWFFNTLQIGDGNECKGEQGQ--CITKFNAPHSITTW-VIQGIAVSKTTGMCVAEP-LKITVFKKTFVQL :  835
ScsuC3-2         : IEDSIGEKH----NCSVAFLDCCQHAEEIRKTFGSG-----VGRSLDEEDESDPNFADIMQVIETFEQETLDNIRRYFPETWILDIFQITNSECKREDLSV--CEKTYTAPHSITTW-MVQGFGLSRTTGLCIADP-IRIPVFKPMFVEL :  850
ScsuC3-3         : IKERMGEKF----NCSEAFLDCCEHTEDSILT---------FGRNFRYEKVEKP----VENLVDELKEDDIKNIRRFFPEKWLFDTYHIGNQKDCKNEKDI--CIVSATAPDSITKW-VVQAVGISKKTGMCIAEP-LELIVFKSMFVQL :  841
PsfuC3           : EYIVDD------APCAAAFLHCCKAMETERVERQE----------------DNLQLARSEEDD-SYIDSDEITSRSNFPESWLWLDQVLPTCPQNTPNCRSTSFEKTVPLQDSITTW-QFIGISLSKTYGICVAEP-LEVIVRKEFFIDL :  827
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : EYIDDG------PSCVEAFVHCCKEMEKERAEKRE----------------ENLQLARSEDDDNSYMDSSEIVSRTQFPESWLWLDINLPPCPRNTPNCETTSFVRNVPLQDSITTW-QFTGISLSRTLGICVGDP-LEVIVRKEFFIDL :  775
HosaC3           : -------------ACKKVFLDCCNYITELRRQHARASHLGLARS----------------NLDEDIIAEENIVSRSEFPESWLWNVEDLKEPPKNG----ISTKLMNIFLKDSITTW-EILAVSMSDKKGICVADP-FEVTVMQDFFIDL :  833
HosaC4           : -------------DCREPFLSCCQFAESLRKKSRDKGQAGLQRALEILQEE-----------DLIDEDDI--PVRSFFPENWLWRVETVDRFQILT-----------LWLPDSLTTW-EIHGLSLSKTKGLCVATP-VQLRVFREFHLHL :  837
HosaC5           : -------------RCIKAFTECCVVASQLRANISHKDMQLGRLHMKTLLPVS------------------KPEIRSYFPESWLWEVHLVPRRKQLQ-----------FALPDSLTTW-EIQGVGISN-TGICVADT-VKAKVFKDVFLEM :  827
AmspA2M-1        : TPRPTQRPRPAPPLFESSTTSSDDDLT-DDAVN------------------------------------EAVEVRTYFPETWLWDLE-VVGDDGYS--------NKEAEIPHTITEW-VGSMFCTSKSKGLGISAP-SAIRAFQPFFVSY :  880
AmspA2M-2        : --ASGFGSGGGGGVN------------------------------------------------------EAVEVRTYFPETWLWDLE-VVGDAGYT--------DKEAEIPHTITEW-VGSMFCTSKTNGLGISSP-TAIKAFQPFFVSY :  910
AmspA2M-3        : QVRPIPRPPPVPKVQPIPRPPLRRPASPDDGVN------------------------------------EAVEVRNYFPETWLWDLE-VVGDDGIT--------NKEAEIPHTITEW-TGSMFCTSKTDGLGISPS-AAIKSFQPFFVSY :  906
AmspA2M-4        : --VLNADDDDNAGVN------------------------------------------------------EAAEVRAYFPETWLWDLE-IVGDGGYA--------DKEAEIPHTITEW-VGSMFCTSKAKGLGVSPP-ATIKAFQPFFVSY :  847
HaadA2M-1        : ---------------------QFITRVEALPS-------------------------------------NSEEIRSFFPETWLWELH-SVDSTGET--------AIKRQLPHTITEW-VGGAVCVHPKTGLGIWDI-SSVTTFQPFFIDF :  811
HaadA2M-2        : ---ARRPPASPAMAPVAA-----DKMGEFSTK-------------------------------------SVVDVRDYFPETWLFDLK-LTEEDGVY--------LAKEKLPHTITEW-VGSAVCINDEDGLGLSNT-TSIKGFQAFFISM :  870
ScsuA2M          : RQSREFSPPLFELPVITFRNKEKQIFEKKQKK-------------------------------------SAVEIRNFFPETWLWELH-KIGNSGQQ--------KVPLKVPHSITHW-VGNAFCISSYAGIGVALP-AHLKAFQPFFMTY :  848
NinoA2M          : SSRPIVRQYSESVPQDISGVRHVDQLEVEETSL------------------------------------PAVELRDHFPETWIWDIV-DLGVDGQQ--------SVERNIPHSITKW-TGDAFCMSPLNGLGMSSK-ASIKSFQPFFLTY :  818
EpspA2M          : LLSMASPFPMASESDTKVTHEISMVTSPEISL-------------------------------------PAVEIRDFFPETWIWELQ-TIQNNSKV--------VIDREIPHSITEW-SGNMFCMSEKSGLGVSPR-TSIKSFQPFFLSY :  814
VaspA2M-1        : ----------------------------------------------------------------------TTRIRQDFPETWLFELQ-YMGKEQMF--------HKHVTVPDTITEW-IGSSYCLSPTKGLGMSDE-ATFNSVISFFIDY :  191
VaspA2M-2        : --------------------------------------------------------------------------------------------------------------PDTITEW-IGSAYCISPVSGYGESQP-NSINVFVPFFIDY :   38
VaspA2M-3        : ------------SPGPSQK--------------------------------------------------ELTRIRQDFPETWLFDLE-YMGDANSI--------VKQVKVPDTITEW-VGSGYCLSSSEGLGISEE-ASFISFTPFFIDY :  830
PsfuA2M-1        : -----------KDNSVKVRKH--------------------------------------------------------FPESWIFEKYSEFP--------------IKIKAPDSITKY-QTTGICFQKMIWVSPTD--TSLTISQDFYSYL :  720
PsfuA2M-2        : -NREKLMGDGLDNLNLNLVQGVILREY--------------------------------------------------FPESWLFKSIKSFP--------------LSITAPHTITTY-KTSAVCFFEKSIWISST-YALLKVHKDYYVTV :  673
PsfuA2M-3        : -VNRVEDDF--------DGLSQHKQGQVLVSPLQ---------------------------------AP-SVEIRDYFPETWLFDLV-DLE-DNEE--------VIEKEAPHTITTW-VAEAFCSNLETGLSVAEV-ANLKASQDFFADL : 1039
CaspA2M-1        : -ANRLDADY--------DNYSEKKEGQVSLSQVR---------------------------------GGSSLEVRDYFPETWLFDLV-DLE-DGEE--------VIKREAPHTITTW-IAEAFCSHPDSGLSVARR-ASLKAYQDFFADL : 1056
CaspA2M-2        : -VNRLEADI--------EDQSHHKQGQVLLS--V---------------------------------QEQIVKIRDYFPETWLFDIV-DLE-NKTE--------TIEKEAPHTITTW-VADAYCSNLESGFSVADR-SELKVTQDFFADL : 1024
PaspA2M          : -VNRFEEAGAGS--------ARPVQQNKQGDISL---------------------------------SVPSLEIRDYFPETWLFDTV-LFSSEDSE--------VVKVNVPHTLTTW-VADAFCSHTEDGFAVAER-AELKVSQDFFVDL : 1038
HosaA2M          : ---VGFYES-DVMGRGHARLVHVEEPHT-------------------------------------------ETVRKYFPETWIWDLV-VVNSAGVA--------EVGVTVPDTITEW-KAGAFCLSEDAGLGISST-ASLRAFQPFFVEL :  796
AmspiTEP/CD109   : VYDFKIVPFGAARPPDTPQRPRTPFIN------------------------------------------------IDLPPTWLWWNRT-VGSDGSA--------SLESHVPENMTSW-IISAFSISPTNGLALAQNSAKVTVFERFFVKL :  743
HaadiTEP/CD109-1 : ILEEHQQLDGDSMAYNSGSYGGKPRVR------------------------------------------------KHFPETWIWDLSP-AGPDGKM--------LLSKKAPDTITSW-IISAFAVDPVTGLGIAPDTTKVTVFRPFFVKL :  749
HaadiTEP/CD109-2 : SNNAPIQPPSE-LPEAIVPE-GRLVIR------------------------------------------------KMYPETWLWVNTT-TGNDGAA--------SISHAVPDSITSY-TISAFAIHPVDGLGIATSNSQITTYRPFFITM :  754
ScsuiTEP/CD109-1 : P-GVPAMDSLEITSRLSPENMPYEPPSALPMP------------------------------------KEPLHVRSYFPETFLWTNET-ASNNGTV--------HIKTTVPDTITSY-YINAFAMDDVNGIGLAEQPAKLQIFRPFFVTL :  772
ScsuiTEP/CD109-2 : SYGMEYDSSLAAPERVGVAQTSYEPA---------------------------------------------MRVRSYFPETFLWTNAS-TNDDGSL--------IIKATVPDTITSY-FINAFAVDDKTGIGLSDQPAKLQIFRPFFVTL :  750
ScsuiTEP/CD109-3 : DQTSDISLLN-----VDESINANLR--------------------------------------------------MHFPETWLWENLT-AGPDGKA--------VITREVPDTITSW-MITAFSVDMITGLAVTESPTKITVFRPFFVSM :  740
ScsuiTEP/CD109-4 : AFATTALGSPPPPARR-PPNSGLVEP---------------------------------------------SRVRTYFPETFLWINTT-SNEEGIF--------NIKTAAPDTITSY-FINAFAIDNENGLGLSDQPAKLQIFRPFFVTM :  756
NinoiTEP/CD109-1 : SLGAPAPAAAPRPSTTMRPESNFKVES-----------------------------------------REKPTIRKSFPETWIWSTEM-SGSDGQA--------VLKTIVPDTITSW-QLSAFAMDDENGLGMADGPQKVEVFRPFFVTL :  766
NinoiTEP/CD109-2 : -LEYAPDNSLGNSIFVNSNSPDSVKLR------------------------------------------------QHFPETWLWNGTA-AKQDGLA--------TISAVAPDSITSW-IVSAFAIDSITGLGVVRAPAKLTIFRPFFVSV :  744
NinoiTEP/CD109-3 : LATMDMAPAAPSVASDGGNSVQTTIPR------------------------------------------------TYFPETWIFQTFQ-TNDSNTT--------VLPFTVPDTITSW-VITSVVMNPVSGLCLTNAPTKLKVFKSFFLVA :  765
EpspiTEP/CD109-1 : GASFAESDSVQFQASTAPASFSGAGAN-----------------------------------------NDKPKLRQSFPETWIWTMPI-AGPDGKA--------VLKTTVPDTITSW-QLSAFAMDDENGLGMADGPSKVEVFRPFFVTL :  763
EpspiTEP/CD109-2 : -LEYAPDNSLGNSIFVDSDSPDSIRLR------------------------------------------------QHFPETWLWNNTV-AQQDGVA--------TISATAPDTITSW-IVSAFAIDSITGLGVVRAPAKLTIFRPFFVSV :  748
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : VPTLNPDKGPGLEYESATRPPLEGPYAF------------------------------SRFPRPLDNIPK-IYLKNDLPETWLFLNAT-TDSDGRA--------SIPVKAPETTNTTWIISGFALDDLHGMGITEQFGSLEVFQPFYVQV :  795
CaspiTEP/CD109-1 : VPTLNPDKGPGLEYESATRPPLEGPYAF------------------------------SRFPRPLDNIPK-IYLKNDLPDTWLFLNAT-TDSEGRA--------SIPVKAPETTNTTWVISGFALDDLHGMGITQQFGSLEVFQPFYVQV :  810
CaspiTEP/CD109-2 : VPTLNPDKGPGLEYESATRPPLEGPYAF------------------------------SRFPRPLDNTPK-IYLKNDLPETWLFLNAT-TDSNGRA--------SIPVKAPEKTNTTWIISGFALDDLHGMGITEQFGSLEVFQPFYVQV :  693
HosaiTEP/CD109   : YAERFMEENEGHIVDIHDFSLGSSPHVR-----------------------------------------------KHFPETWIWLDTN-MGYRIYQ--------EFEVTVPDSITSW-VATGFVISEDLGLGLTTTPVELQAFQPFFIFL :  752
                                                                                                    /
```

Supplementary Fig. S9 (continued)

```
AmspC3           : NMPYSVVRNEEIEIKVNVFN-YLSTTVPVTVYMYGVDGLCTGAEAGK---------------RTDRKLIRVPKNDAVSTGFFIIPLRIGTYKLRFVALSEAKT-------------------------DVVEKTLHVRAEGSQQFKDSST :  934
HaadC3-1         : NLPYSVVRNEQVEIQATVFN-YGNTKIGAVVYMYGAKDLCSGTQAGE---------------KSERKRLIIEGQSAATVTFPVIPLKAEDFVIKVVALTPAGS-------------------------DVVQRTLHVVAEGVTKEIDIPI :  941
HaadC3-2         : NLPYSIVRNEQVEIQATVFN-YDHRPIRAVVYMYGTEGLCSSTQPGQ---------------KSERQYVTVEGQSAATVTFPVMPLKAEVIPIKVVALSSLGS-------------------------DVIVRELNVVPEGITRFINIPI :  943
ScsuC3-1         : SLPPVAIRGEQIEVLATVFN-YEPEDLDVSVYMFGVEGVCMGAGPGE---------------RTEIRKLKVPANGASSATFSVMPLEVSEYQLRVAALSYTSS-------------------------DAVQKVLRVVPEGARVEKSISF :  944
ScsuC3-2         : NLPPAAVLGEQIEVVATVFN-YGQESLKVTVYMYGVEGICMGAAAGE---------------KSPVRQVEVSANSATSVSFPVMPLEVSEYPLRVVALSWRAN-------------------------DAIEKKLRIVPEGVTKDKSLSF :  959
ScsuC3-3         : SLPAVAMRGEQIEILATVFN-YDSSDLDVFVYLYGVEGLCTGAIPGE---------------RTEPKRIKVLGNSASTVTFPIMPLTVSDFNIQVLASAGHLY-------------------------DAVIKVLKVVHEGIPTEKTISF :  950
PsfuC3           : RLPYSAVKGEQIEVKAILHN-YSPDIITVRVELLEEEDICSSASK---------------RGR------YRQEVEVGEHSTRAVPFVIIPMKEGQFSIHVKAAVRDSSL------------------RDGIEKKLRVVPAGVLVKTPVVV :  937
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : KLPYSAVRGEQIEIKAILHN-YSPDDVTVRVDLTEEQHVCSSASK---------------RGR------YRQEVRVGPETTRSVPFIIIPMKEGQYRIEVKAAVKDSSL------------------NDGIIKTLRVVPAGVLVKVPLVL :  885
HosaC3           : RLPYSVVRNEQVEIRAVLYNYRQNQELKVRVELLHNPAFCSLATTKR----------RHQ-QT-----VTIPPKSSLSVPYVIVPLKTGLQEVEVKAAVYHHFIS-----------------------DGVRKSLKVVPEGIRMNKTVAV :  944
HosaC4           : RLPMSVRRFEQLELRPVLYN-YLDKNLTVSVHVSPVEGLCLAGGGG-------------LAQQ-----VLVPAGSARPVAFSVVPTAATAVSLKVVARGSFEFP----------------------VGDAVSKVLQIEKEGAIHREELVY :  946
HosaC5           : NIPYSVVRGEQIQLKGTVYN-YRTSGMQFCVKMSAVEGICTSESPVIDHQGTK----SSKCVR-----QKVEGSSSHLVTFTVLPLEIGLHNINFSLETWFG-------------------------KEILVKTLRVVPEGVKRESYSGV :  942
AmspA2M-1        : SLPYSVIRNEKVPITVSVFN-YLSECLPIKLKLEKNDGFTLLSNSY-------------------THSLCVCGGKPAIHNFRILPTSLGEVNLTVHSFSFDDKNYEVCSQDKVSSKVV--------ARDAITKPLLVEPEGFPQVITESL : 1002
AmspA2M-2        : ALPYSVVRKEKVPIIVSVFN-YLSECLPIQLKLEKSDEFTLLSDSY-------------------THRMCVCGGQPATHRFRILPTGLGEVNLTVYSHSFNDANNEVCSKDKKASTLE--------ARDAITKPLLVEPEGFPQESTEST : 1032
AmspA2M-3        : TLPYSVIRNEKVPVIVTVFN-YLPECLPIELRLEESEDFELLSN-N-------------------THRMCVCSG-PATHRFRIQPTDLGKINLTVHSDSFVDATHEVCPEDGGASTLV--------ARDAITKPLLVEAEGFPQESIQSV : 1026
AmspA2M-4        : TLPYSVIRKEKVPIVVSVFN-YLPECLPIQIKLQKNEGFTLLSNNY-------------------THSMCVCGGKPATHQFRILPTSLGEVNLTVYSHSFNDSNKKICPEDKNASALI--------ARDAITKPLLVEPEGFPQELTESV :  969
HaadA2M-1        : HLPYSVIRGESFPLVVTVFN-YLSECLPIKLSLEPSDDYTLLTELR-------------------FQKTCVCGGQSSSVSFPVRPATLGMVNFTVYGYSIEQDDE-ACGNEI-TARLS--------ARDAITKEILVEAEGFPKEDVFNY :  931
HaadA2M-2        : TLPYSVIRGESFWITISVFS-YVEDPLPITVTLDNLEGFEIVSESI-------------------DGDICVQPGSSNNLKIQLKGKTLGSNNITVHAESASSSD--VCGSDS-ISDAV--------AKDSIRKPVIVEAEGWPVEEIESV :  989
ScsuA2M          : TLPYSVKRGETMKLLVSLSN-YLQECLPIKATLQESKVFSIASETT-------------------FQKICLCKEETKTIQFLIRPKSIGKMNITVYAYTSTKN-ICDSNTPDPVSNEQ--------ARDAVTKSLLVEPEGFPKEDTWTS :  969
NinoA2M          : TMPYSVKRGESFPLTVSIFN-YLSRCFPVVIILEESNAYIMDGNTS-------------------SITLCLCGSKSYSHKYYIKPLETGNINITVRAHSTKLEKVTVCDNQEATQNVV--------AFDAMTKSLLVKSHGFPIEITESN :  940
EpspA2M          : TMPYSVKNGETIPITVSIFN-YLSGCFPVVVKLEESQFFVTDVNST-------------------EVKLCLCGSKSYSLKFLIQPLKIGNLNVTVRAHSITEFDILLCEIESTLQNVV--------AFDAITKSLLVKAPGFPQETTQSN :  936
VaspA2M-1        : TPPYAVKRGESFELPVYIYS-YLNWTSSVFVHLEKSTKNLYQLTDG-----------------LDLKIASIPPRGKSVVTFEITPTKIGEIEIGVYATGNFI-----IFDGNKVNHS--------IITDAVRKKILVKPDGFPQENHDVQ :  310
VaspA2M-2        : SLPYSIKRGETFEFKVSVFS-YVDEASPIKLILEPSVDFAVDDEEY------------------EKD-LWLNSNEHEVVIFNVTATTIGSIELITYAMTSIGNEPEWGYGNSTITY----------HSDAVKKSVIVEPEGISEEYTQSE :  158
VaspA2M-3        : NLPYSVKRGETFELKVSVFN-YLGISAPVLLELQQSNDYDVADGVY-------------------NKTVMIPPNKSQVITYLLTATNIGDVNITSFAYLSPVN----SSDPTYANISF--------TSDAIRKTLLVEPEGFPQEEHQVE :  948
PsfuA2M-1        : NLPKY-LIGSEVALITA-TH-YLNKTMIGNGKNYVKFDLEVNKDVTEVQNNKE---FSFCTNKNDLNIHSAIFLEYAKEFGTLKMRLNSQIFIDSVKCG--------GSKSD--EKANILYDIVEKKIEVFPSGEKTEIVKSGLICTNQL :  854
PsfuA2M-2        : KAPAFVYIQNTFALRVNTFL-KQSTLQNIRVILKVSVSSSDFQILES-------------------KQKVFDQKSNEVKVNTFMLKALKYSKGLKVIVS------AKFYLINGFALK---------YEDAVKVSIPSIPNGDRQELVQTV :  788
PsfuA2M-3        : NAPYAAKRGEVLQVNATVFNKVEG-PLPMKISIESS--AHYKVVNKS------------------EEIVCVNPGGNIPVDFYVEMNTLDVVNVTVKAEIIKNED--CGLVAENSIGFT----------DVLKKAIHVRPEGFPKEEVKSY : 1156
CaspA2M-1        : NAPYSVKRGEILQINATVFNKVKS-PLPMKLSVEESPEQAYEIIEKT------------------HEVVCVNAGGNVHVDFYIRANELDVVNITIKAEIVKDEA--CGSVSDESIGFT----------DILKKAIHVKPEGFPQEVTKSY : 1175
CaspA2M-2        : NSPYSAKRGEILRVNATVFNKVEG-PLPMKLSIQAS--EGYSLVNKS------------------EEITCVNYGGNVPVDFFIQLNKLGQVNLTVKAEIVKDES--CGSISEESVGYT----------DTLVKSINVRPEGFPKEVVKSY : 1141
PaspA2M          : NAPYSAKRGEILQINATVFNKVEG-PLPMKIILPSE--GHYSVINNS------------------DVNICVQSKGNEVIGFFVEFQKLHEVNVTISAEIVQDSS--CGDVSESSIGIS----------DSLKKPIIIKPEGFPHEEVRSY : 1155
HosaA2M          : TMPYSVIRGEAFTLKATVLN-YLPKCIRVSVQLEASPAFLAVPVEK------------------EQAPHCICANGRQTVSWAVTPKSLGNVNFTVSAEALESQEL-CGTEVPSVPEHG--------RKDTVIKPLLVEPEGLEKETTFNS :  918
AmspiTEP/CD109   : ILPHSVILGETLSVQVVVFN-YNDRPAQVEVTMENKGG---FEFTTVEDDPSI------RRATRMRKATTVPAQEGKATYFMIKPNRLGYIDIKVFARS-------------------------SFAGDGAQDKLLVKPTGGPQYFNRPI :  858
HaadiTEP/CD109-1 : NLPYAIIRSESLYVEVIVFN-YMKKTMKADVTLENRKGEFEFASTSNE-VSVP---------KEMTKTVDVPAQDGVAVKFLITPKKLGYIDLKVTAQA-------------------------ENVGDAIVKKLLVKPEGSPQFFNKAL :  863
HaadiTEP/CD109-2 : SLPYYVLMGEDLAIQVVVFN-YNDKPIQAEITMENRKREFDFTAAGQE-SVYS------PDQNQRTKIVHIPPSDGVPVSFLIIPKKVGYIEIRVSAST-------------------------SVAGDSLTKRLLVQPEGSTQYFNKAF :  871
ScsuiTEP/CD109-1 : NLPYSVVRGETLALQALVFN-YMTENLMVWLTLENENNEFVLINLENEMNVDE---------NSKYIEIQVKSGEGASVFFYIVPKKIGYIDIKIAARS-------------------------HIAADAIHKKLLVKAEGIPMYNSKTV :  887
ScsuiTEP/CD109-2 : NLPYSVIRGETLGLQALVFN-YMTEDIEAVVTLDNEQQEFTIIDKENEITDGS------ATVNEISKTVKVASGDGVSVLFYVKPKKLGYLNVKVTARS-------------------------KTAGDAILKKLLVKAEGKTVYETRGV :  868
ScsuiTEP/CD109-3 : NLPYSVIRGEAIAVQAIVFN-YMKQTVSVTVTLEN-THQFDFVTVEDEINEVF------N-ANLKSKRIAVKSGEPESVYFMIIPKELGFIDIKITAET-------------------------ERAGDAMLKKLLVKPEGIQQKFNKAF :  856
ScsuiTEP/CD109-4 : NLPYSVVRGETLSLQALVFN-YMKEDLEAEVTLSIEDGSFTFVDLENEINDDA-----NAKQNFITKTVKVKSGDGTSVYFYIVPKKLGHLDLKMTAKT-------------------------NVAADALIRKLLVKAEGMPVYVNKAS :  875
NinoiTEP/CD109-1 : NLPYSIVRGESVAIQALVFN-YMKEDVEAEVTLDNPNDQFELTGLLNRVDADH-------NATSEMKTVKVKAGDGTSVSFLITPKVVGPIDLTVTAKS-------------------------SKSGDALNRKLLVKAEGSPQFFNKAV :  883
NinoiTEP/CD109-2 : VLPYSVIRDEAVAIQIVVFN-YMAEEAKATVTLKNQREQFDFATFEMGPNEVL------DSSQPKYKKVHVKSGEGTTVSFMIIPKVLGYIDIEVSAQS-------------------------ERAGDRVLKKLLVKPEGTQINLNKAV :  862
NinoiTEP/CD109-3 : NLPYSVIRGESMVIQVLVFN-YLATDLSQVVVTLTQDSNLPPQFTFVDPATTG---------TSASKTISVKANNTATVSFTITPLVVGYTKLNINAKS-------------------------GQAGDALVAQLNVKPEGAPQFFNKAV :  880
EpspiTEP/CD109-1 : NLPYSVVRGESVAVQALVFN-YMKEDVEAEVTLENLNDQFELTGLLNRVDVDH-------NATSEKKTVKVKAGDGSSVSFLITPKVVGPIDLTVSAVS-------------------------SKAGDALNKKLLVKAEGSPQYFNKAV :  880
EpspiTEP/CD109-2 : VLPYSVIRDEAVAIQVVVFN-YMAEDVKATVTLKNQKGQFEFANFETGANEVV------DNSQPKYKRVQVKAGDGTSISFMIIPKLIGSIDIDVTAQS-------------------------DRAGDRVLKKLLVKPEGTQLNFNKAV :  866
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : DLPPSIKLGETLSVQMVVYN-YLKETISATVVLEGTEEEIVFGEADPYSL---REDTQIGSRVSQEKSISVRPGRGTIVSFVITPLVTGDIQLRIKAEGKSGDSN--------------------MGKEDIFKTLSVRSEGEVIKRNKGF :  921
CaspiTEP/CD109-1 : DMPPSIKLGETLSVQMVVYN-YLKEDISASVVLEGREEEFVFGEADPYYS---DEDYQIGSIVSQEKEVGIRPGRGTIVSFLLTPLVTGDIRMRIKAEGKSGDSN--------------------VGYENIVKVLRVESEGEVMRRNKGY :  936
CaspiTEP/CD109-2 : DLPPSVKLGETLSVQMVVYN-YLKESIQASVTLEGREEEFVFGEAQPDLSNAYPKNEQIGSLISQEKEVGIKPGRGTIVSFLITPLLSGNIRMRIKAIGKSGTSN--------------------VGNDNVVKTLTVTSEGEIMHRNKGY :  822
HosaiTEP/CD109   : NLPYSVIRGEEFALEITIFN-YLKDATEVKVIIEKS-DKFDILMTSNEINATGHQ------QT-----LLVPSEDGATVLFPIRPTHLGEIPITVTALS-------------------------PTASDAVTQMILVKAEGIEKSYSQSI :  864
```

Supplementary Fig. S9 (continued)

```
                                                                                                             Thioester site
AmspC3           : VLDPSNYNGRPGSAIWIQKP-DGYSYFLDSKNRKLVVSLPII--PPVNSIAETEIATINIIGDSFGPTVKTAFTKNELGKFINKPRGCGEQ-NMMLMAPCLYTMKYLAATGKI---GVN--EE-KTGYGWIRLGYERQLN-FRRKDGGYS : 1073
HaadC3-1         : KLDPTNQQRRQKRHIETEL--YSDHIDPTQNLQVTAVKLS----APEGFVPGTASCSITALGDMYGPAVQTSINNPDA--LFQKPRGCGEQ-NMMYLAPTLYALRYLKVTGKL---TAA--AE-ESGYEFIRHGYGNQLA-FRKEDGSYA : 1075
HaadC3-2         : ALDPTNLQKRKKRSIEDSY--ISDHIDPTQNTQATTVKLDFC--LPEDFVPGTESCAISAIGDVFGPAVQTATENPEG--LVQLPYGCGEQ-NMKWLAPTLYAMKYLKVTGKL---TPH--IE-EKGYQYMREGYNNQLN-YKKGNGAYA : 1079
ScsuC3-1         : TLDPGGIYSKRPRRQADHEGTIKDVYDEVLRKQLITVDLD----LPNNHIPGTEKCFISVVGDPVGQAVNTTLSGLGG--FLKMPTGCGEQ-TMIALGPLVYTMSYLKKTKQM---TAN--IE-DTGYKYLWGGYSLQQK-YRKADGSYA : 1080
ScsuC3-2         : FLDPSGLIRNKHRKREVTP---AGVIEYSTNKQKMKINLT----LPENYVPGTEECFVSVIGDSMGSVVSTSLKGLDQFLVAAGPHACGEQ-TLVKLAPLVYTINYLKKTKQL---TSS--FE-SKGYSYISQSYDQQMK-YRKADGSFS : 1094
ScsuC3-3         : PLDPEGKINKRKKRAIDQS--ISEVYNEVQKRQEITVDVT----FPHDHIKGTEKCFVNFIGDPVGQAVNVTLSGVEEE-FLKLPQGCGEQ-TMIKLAPLVSTMHYLKKTNQF---SAT--AE-KKGYDLIWKGYDNMQK-FKKNDGSYS : 1085
PsfuC3           : TLDPIKKGQDGVQT----------ETINSAIPKTDVAPNTP------TSTQISITGKENLAGLSSGNNAEYAG---G--TLIYQPSGCGEQ-NMIHMTLPVIAATYLDKTKQW---EAVGFQKRNEAIKHIQTGLNNQQA-YAKKDGSFA : 1061
CaspC3           : ---------------------------------------------------------------------------------------------IMNGMTMALIATYYLDTTDQ--WDKVGVERRHEAVDHIKTGYIKQLAFQQEWDGSFI :   55
PaspC3           : PLDPAKQGVGGTQK----------EIINSAIPKKDIAPNTP------TSTQISVSGKEQVSTLVEKAISGNSMG-----SLIVQPSGCGEQSTMIRMTLPVIATTYLDTTNQW---EKVGFQKRAEALQHIKTGYERMLT-FRKGDGSFA : 1009
HosaC3           : RTLDPERLGREGVQKEDIP-----------------PAD-----LSDQVPDTESETRILLQGTPVAQMTEDAVDAERLKHLIVTPSGCGEQ-NMIGMTPTVIAVHYLDETEQW---EKFGLEKRQGALELIKKGYTQQLA-FRQPSSAFA : 1067
HosaC4           : ELNPLDHRG--------------RTLEIPGNSDPNMIPDG--------DFNSYVRVTASDPLDTLG--SEGALSPGGVASLLRLPRGCGEQ-TMIYLAPTLAASRYLDKTEQW---STLPPETKDHAVDLIQKGYMRIQQ-FRKADGSYA : 1067
HosaC5           : TLDPRGIYG----------------TISRRKEFPYRIPLD-------LVPKTEIKRILSVKGLLVGEILSAVLSQEGINILTHLPKGSAEA-ELMSVVPVFYVFHYLETGNHWNIFHSDPLIEKQKLKKKLKEGMLSIMS-YRNADYSYS : 1067
AmspA2M-1        : LFCP-SEHG-----------------NRFKKDFELSLPD--------DLVEGSARAFLYVSGDIMG-PSLSGLEN-----LVTRPTGCGEQ-NMIKFAPIIFVTQYLQGVGSL---TPE--TK-KKSLEFMKVGYQREL-TYRHNDGSYS : 1112
AmspA2M-2        : LFCP-SEYQ-----------------NGFKKSFELMLPD--------DLVEGSARAFLSVSGDIMG-PSLSGLEK-----LVARPTGCGEQ-NMIRFAPNIFVMQYLQGTSSL---TPE--IE-KKALDFMRIGYQREL-NYRHDDGSYS : 1142
AmspA2M-3        : LFCP-SEHQ-----------------NGFKKAFELMLPD--------DLVQGSARAFLHVTGDIMG-PSLSGLER-----LVRFPTGCGEQ-NMVLFAPNIFVIQYLQGIDSL---TPE--VE-NKALGFMRAGYQREL-NYRHGDGSYS : 1136
AmspA2M-4        : LFCP-SEHQ-----------------NGFEQDFEFMLPH--------DLVEGSARAFFSITGDIMG-PSLSGLER-----LVALPIGCGEQ-NMIRFAPNIFVMQYLQGTNNV---TKE--IE-KKALKFMKTGYQRQL-NYRHSDGSYS : 1079
HaadA2M-1        : FICP-ENTN-----------------GSFATEIPLLLPD--------DVIMDSARAYMTITGDVMG-PSIKGLKK-----LVSLPFGCGEQ-NMVLFVPNIFVLDYLTSTEKL---TDD--IK-EECLHNMKTGYQREL-QYKHSDGSYS : 1041
HaadA2M-2        : LFCPKDEEN-----------------DVFKKTLTLNEPE--------DVVPDSSRAYLDCSGNVLG-KCLDNLEN-----LVSLPTGCGEQ-NMVKFAPNVVAMKMLINTNQL---SDK--TK-DRIVRNLNTGSQRQM-KFKHPDGSFS : 1100
ScsuA2M          : LICLNDSES-----------------NNDTEFEDVIKESVVLSISDTQVVPGSVRAYITVIGDTMG-PSLQGLDH-----LVRLPVGCGEQ-NMVLFAPNIYVLHYLKNTNQL---TTA--ME-NKIISHLKTGYQREL-TYRREDGSFS : 1088
NinoA2M          : WICVKDN---------------------STTFEHTLELS-------DDVIKGSERAHVSVTGDLMG-PTISGLDH-----LVRLPTGCGEQ-NMVLFVPNIYVLHYLNASLQL---TSD--IK-SKAISNMEKGYQREL-NYKRKDGSYS : 1048
EpspA2M          : WICTNDFE------------------NGSKTLEYSLELP-------EDVIEGSARAFISVTGDLLG-PTISGLDH-----LVKMPTGCGEQ-NMVLFVPNIYVLQYLEVTHQL---TIN--IK-LKSISYMELGYQREL-NYKRDDGSYS : 1047
VaspA2M-1        : LLCTKDNEK-----------------LAKNLNCSPPK-N---------IVEDSARATYVVTG-VFG-PILDNLDS-----FIQLPMGCGEQ-IMAAMVPSLYVLTYLDSIGAA---EPQ--LR-DKAFETILLGYQQQL-TYKNPDGSYS :  418
VaspA2M-2        : FMCVQESDTT----------------MVRLFACEAPQNS---------VADSS-RASFSVTGDLMG-PVLNNLES-----LIRFPTGCGEQ-NMAKFVPNIHVLSYFKSTKFN---DNT--LK-TKCIEHIREGYQREL-KYKHSDGSFS :  268
VaspA2M-3        : FMCVQDGAQ-----------------LSKLLDCSPPP----------LIVPDSAKTIYSVTGDLMG-PLLTNIDS-----FIQFPTGCGEQ-NMATFVPNIHALTYLDSMGAA---EPE--LR-EKSKKHMEEGYQRQL-KYKHSDGSYS : 1057
PsfuA2M-1        : N--------------------------QVIDLPIDQ------------IVPKSLKMSVTLATDGTQ-MISENLKN-----MIRQPSGCGEQ-NIALVIPSISALSYLQNANIV-MEEEKKEKLKLKATKFILQGYVRQL-KYQHKNMGFS :  957
PsfuA2M-2        : FLCKDK------------------LYSGVLDFSSIKSI----------IPSTMSIKAEAAADDMQL-TMSH-LSS-----LLLQPMGCGEQ-NIGRVGPSIAVLIYLINHQMV--HMKS---MKAKAIRYIRLGFIRQQ-KYIHGNYSYS :  896
PsfuA2M-3        : FLCGDQKDTK--------------------LEDMILPESN--------LVEDSVRAWFGISGDIMA-PAVKNLDN-----LVALPTGCGEQ-TMIRMVPNIYLLDYLKSIGKS---LPQ--LE-EKARRYIQTGFDRQNRNFRHTDGAYS : 1265
CaspA2M-1        : FLCGEQVDTP--------------------LEDIVAPEEG--------LVKDSVRFWAGLSGDIMA-PAVKNLDS-----LVALPTGCGEQ-TMIRMVPNIYLLDYLKSIGKS---LPE--LE-TKAKGYIQTGFDRQNRKFRHRDGAYS : 1284
CaspA2M-2        : FLCGEQKDSP--------------------LEDTLLPDGN--------LVDDSTRAWFGISGDIMA-PAVKNLDG-----LVALPTGCGEQ-TMIRMVPNIYLLDYLKSIGKS---VPT--LE-SKAKRYIQAGYDRQNRKFRHRDGAYS : 1250
PaspA2M          : FLCGEQNDTV--------------------LEDVNLPEDD--------LVDGSVRARFSISGDIMA-PAVKNLGR-----LVSLPTGCGEQ-TMIRMVPNIYLLDYLKSVGKN---LPE--LE-TKALKHMNLGYDRQNKKFRHKDGSYS : 1264
HosaA2M          : LLCPSG--------------------GEVSEELSLKLPP--------NVVEESARASVSVLGDILG-SAMQNTQN-----LLQMPYGCGEQ-NMVLFAPNIYVLDYLNETQQL---TPE--VK-SKAIGYLNTGYQRQL-NYKHYDGSYS : 1026
AmspiTEP/CD109   : LIDRRS-----------------AGGEPLSVDVELNIPRT--------VIRGSEKIEVTAIADVMG-PVIENLGD-----LLRIPRGCGEQ-NMVNFVPNILVINYLKGNNRL---AND--IY-SKAIGNIGTGYLREL-TYKHDDGSFS :  969
HaadiTEP/CD109-1 : LVDLRD-----------------PSSKKIDQTVKINVPGD--------AVPGSTSVMLSAIGDLLG-PTVNNLDK-----LLRMPQGCGEQ-NMLNFVPNIVISKYLKRVNRL---TPA--IE-KRSLHFMESGYQREL-TYRRNDGSFS :  974
HaadiTEP/CD109-2 : LIDTRN-----------------PSSP-SKMNISTIIPKN--------AIRDSGKIIVSAAADLMG-PSIKNLDK-----LLYMPNGCGEQ-NLVTIVPRVIALEYLARSNRL---TEN--IR-AKAIANLRKGYQRQL-TYKRDDGSFS :  981
ScsuiTEP/CD109-1 : LADLR-------------------KQSQFQDKIQILMPKN--------FVDGSERIEISATRDIMG-TAINNIDQ-----LLRMPSGCGEQ-NMLNFVPNIVIMDYFTATNRL---TPQ--IE-DKAIRFMESGYQREL-TYKRKEGSFS :  996
ScsuiTEP/CD109-2 : VADLR-------------------KFTQIGEEIHLNFPSD--------TVEDSERIEVSAISDIMG-TTINNLHN-----LLRMPGGCGEQ-NLLRFVPNIVIADYLTNTNQL---TPA--LK-EKAIRYKETGFQGQL-IYRRYDDSFS :  977
ScsuiTEP/CD109-3 : LVDLQ-------------------TPSVFNAYVNVDIPKH--------VVSGSEKIEISAIADVMG-PTINNFDD-----LLQMPFGCGEQ-NMMRFVPNIVVLEYLSNTEQL---TDS--LR-SKAILNMETGYQREL-VYKRDDGSFS :  965
ScsuiTEP/CD109-4 : VADLR-------------------EQSQFNEKVKIEFPED--------YVKDSENIEISAISDIMG-TTVSNIDK-----LLKMPYGCGEQ-NMLNFVPNIAITDYLNETNKL---TPE--IK-EKTIRFMESGYQRQL-TYKRTNNSFS :  984
NinoiTEP/CD109-1 : LVDLR-------------------NSSNFKEDVEINIPPF--------AVKDSEHVEVSAIGDIMG-PTVNNLDK-----LIKMPYGCGEQ-NMVNFVPNIAVADYLNTTNQF---TDK--LK-NKAIKYMEAGYQREL-TYKRSDGSFS :  992
NinoiTEP/CD109-2 : LIDLR-------------------SSNSFTTNVKVDIPNY--------AVPGSGRVEVGIIGDILG-PAINNLDS-----LLRMPFGCGEQ-NMVYFVPNIVIGRYLEYVNRL---QDS--IK-QKILTHLETGYQREL-TYKREDGSFS :  971
NinoiTEP/CD109-3 : LVDLR-------------------NTSTFSTTITIDVPTT--------AVPGSTFIQFSGQGNIMA-PVIANLNG-----LIQLPTGCGEQ-NMVNFVPDIVVYNFLKNIDLL---TDD--IK-NKIINYLIIGYQTEL-TYKHYDGSFS :  989
EpspiTEP/CD109-1 : LVDLR-------------------NTSNFKSQVEVNIPPF--------AVKDSEHVEVSAIGDIMG-PTVNNLDK-----LIKMPYGCGEQ-NMINFVPNIAVSDYLNTTNQF---SDK--LR-TKAIKFMEAGYQREL-TYKRPDGSFS :  989
EpspiTEP/CD109-2 : LVDLR-------------------STNSFTSTVKVGIPNY--------AVEGSARVEISVIGDMLG-PAINNLDS-----LLRMPFGCGEQ-NMVYFVPNIVVARYLEFVNRL---QDT--IK-QKILSHLETGYQREL-TYKREDGSFS :  975
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : LLDFD-------------------KSSDFDKNISIDIPFN--------AIPGSEKVYLSLMADPLS-SAMNHLDK-----LISYSTGCGEQ-NMIRLLPTLAILEYIDKMN-I---P-AN-LQRNNAIKTMELGYQREL-TMRLQDGSFS : 1030
CaspiTEP/CD109-1 : LLDFN-------------------GGSDFDKNISIDIPFN--------AVPGSEKVYVSMMADPLS-SAMNHMEK-----LIHYPTGCGEQ-NMIRLVPVLSILEYIDEKK-I---T-AP-LQRNTAIKTMELGYQREL-TMRLDDGSFS : 1045
CaspiTEP/CD109-2 : LLDMR-------------------DASEFKKNVTIDIPFN--------AVPGSENVFISLMADPLS-STMNNLER-----LIRYPSGCGEQ-NMIRLVPTLAILEYIDKLG-I---S-AT-YQRNEALKTMELGYQREL-TMRMMDGSFS :  931
HosaiTEP/CD109   : LLDLTDN-----------------RLQSTLKTLSFSFPPN--------TVTGSERVQITAIGDVLG-PSINGLAS-----LIRMPYGCGEQ-NMINFAPNIYILDYLTKKKQL---TDN--LK-EKALSFMRQGYQREL-LYQREDGSFS :  975
```

Supplementary Fig. S9 (continued)

```
                                                                         Catalytic histidine
AmspC3           : AGG----TRKSSTWLTAFVMKCFCQAD----AFINVDETVGCTGLQWLVNSQQQDGSFVDKHPVYHREMTGGIKG-----SFAMTAFVLMSLMECTTCKIEGLEITRQ----------QAIAFLEKGIQTK------------------- : 1181
HaadC3-1         : AYQ----SKPTSTWLTAFVIKVFCQAT----ELIHIDDDVVCSGVKWLIKNQERDGSYVEKHPMYHVDMMGGVQG-----KTPMTAFALIALEECKCDEEH-LLLSKK----------RAVAYLENHLGEV------------------- : 1182
HaadC3-2         : AFK----DRAENTWLTAFVMKVFCQAT----ELIYIDEKDICSGVEWLINNQQMDGSYIENYPLYEPKLMVGMHN-----KVPLTAFVLIALEECKCEPDR-LNMAKL----------KAVAYLEDHLHRV------------------- : 1186
ScsuC3-1         : VWA----TYASSTWLTAFVAKVFCQAS----AFIPIPPENINTALEWLVKRQHANGIFAEAYKVHHREMTGGIQG-----DVTLTAYVLISLLECNKYDSVTKKTAVN----------RAILFLEQNIPHL------------------- : 1188
ScsuC3-2         : LWT----FTESGTWLTAYVLKVFCQAYSSN--IFSIPNYIRCDGIKWLLTKQYTNGEFFPLSYRIYGDQKIG-----------LTAFVLITMLECDSCYSAEKNLAVS----------RATAYLRSRLDSI------------------- : 1198
ScsuC3-3         : IFT----ESPSSTWLTAFVIKVYCQAS----EFIQIPLENIGTAVEWLIKHQSSSGKFFDDYKVHSNTIAGGLNG-----DVTLTAYVLIAMMECNRYNSVTKKTAVQ----------RAITFLEKNIDHT------------------- : 1193
PsfuC3           : VWP----KYQSSTWLTAYVAKVFSMANS-LV-AVPGGQ-ICNAIKYLILTSQKPDGRFEERGRMYHTEMIGDVTGI--DADVSMTAFCVIAMQESGPLCTASLTSLPGS-------IAKAVGFLEQRLPQV------------------- : 1176
CaspC3           : TFK-RPLRQTGST----------------------------------------------------------------------------------------------------------------------------------------- :   67
PaspC3           : VYP----KHVSSTWLTAYVTKVFAMAHN-LV-AVQTDI-LCGAVRYLILKTQQPDGVFREIGKVYHGEMIGDIRGG--DSDASMTAFCVIAMQEARALCAASVNSLSGA-------VDKSVSYLEKRLPGL------------------- : 1124
HosaC3           : AFV----KRAPSTWLTAYVVKVFSLAVN-LI--AIDSQVLCGAVKWLILEKQKPDGVFQEDAPVIHQEMIGGLRN-NNEKDMALTAFVLISLQEAKDICEEQVNSLPGS-------ITKAGDFLEANYMNL------------------- : 1183
HosaC4           : AWL----SRGSSTWLTAFVLKVLSLAQEQVG---GSPEKLQETSNWLLSQQ-QADGSFQDLSPVIHRSMQGGLVG--NDETVALTAFVTIALHHGLAVFQDEGAEPLKQRVEAS-ISKASSFLGEKASAGLLG----------------- : 1189
HosaC5           : VWK----GGSASTWLTAFALRVLGQVNKYVE---QNQNSICNSLLWLVENYQLDNGSFKENSQYQPIKLQGTLP--------VEARENSLYLTAFTVIGIRKAFDICPLVKIDT-ALIKADNFLLENTLPA------------------- : 1182
AmspA2M-1        : AFG--QSDAEGSSWLTAFVVKSFAQARNLID---IDAIDLKKSTDWLLSKQ-QSDGCFPFIGMVHHSDMKG---GVGKSVPTALTAYT---VISILESETPISQQKLNK---------AFECISQQTDP--------------------- : 1220
AmspA2M-2        : AFG--KSDAEGSSWLTAFVVKSFAQARQFID---IDPVDLKKSTDWLLSKQ-QADGCFPFIGMVHHQDMKG---GVGKSVPTALTAYT---VISLLEAETPISQDKLDK---------AFECISKQTDP--------------------- : 1250
AmspA2M-3        : AFG--ESDPEGSSWLTAFVVKSFAQARHLID---IDPVDLKKSADWLLSKQ-QDDGCFPFIGMVHHQAMRG---GVGRNEPTALTAYT---VIAILESETPISQDKLDK---------AFGCISRRTSP--------------------- : 1244
AmspA2M-4        : AFG--ENDPEGSSWLTAFVVKSFAQARQFID---IDPIDLKKSTEWLLSKQ-QADGCFPFIGMVHHQDMKG---GIGKNVPTALTAYV---VISLLESETPIGQDQLDK---------AFECITKETAP--------------------- : 1187
HaadA2M-1        : AFG--ASDKEGSLWLTAFVLRSFGQARRFMN---VDENDLSATRSWILKKQ-FENGCFIPSGTVLNKEMKGGL-SSSEQSLAPLTAYV---LISLLESDMEKHDTLVVK--------NALKCLESEKQP--------------------- : 1152
HaadA2M-2        : AFG--TRDKQGSMFLTAFVLRYFSEASQYIT---IDNATISEMQKWITSKQ-KDDGCFPDVGKIIDRGFQG-A-IEKDKSDGTITAYV---LASLRISNYQNQTVLDK----------ALSCLSNSQDS--------------------- : 1208
ScsuA2M          : AFG--KSDREGSIWLTAFVVKSFAQAREFIF---IDDTIMDESISWITNKQ-MENGCYSFVGKVLHKDMKGGV-SDSESSFAALTAYIT---IALLEAGIANDSKPIVN---------AFFCLKAEKEP--------------------- : 1198
NinoA2M          : AFG--NSDMEGSIWLTAFVVKSFAQAKSFIY---VNSEELNEGIQYILRNQ-LKNGCFKSTGRLLHKNMQGGV-AKSNASSQALTLYII---IALLETGYDPSSDVIKR---------AFNCMDEDFHP--------------------- : 1158
EpspA2M          : AFG--KSDAEGSIWLTSFVIKSFAQTKSIIY---IDQHEIDDGIKFIVKNQ-LEDGCFKSVGSVIHKELQGGT-GHG----QSLTLFIS---ISLLEAGLSPSGKIIKK---------AFKCFENHSTM--------------------- : 1153
VaspA2M-1        : IFSG-LPNPRGSIWLTAFVVKSFSRTKTYIS---INDTVIDSSKKFILKNQ-DEDGCFLVVGRVFSSGLRGGVEK-GVPKET-LTAYII---IALLEAGEDPSSQHMKD---------ALQCIKKMPKY--------------------- :  528
VaspA2M-2        : AFG--SSDASGSLWLTAFVLKSFGQALPYIF---VDKKIMNKSVRWIKSKQ-LENGCFPEVGSMFNQGLQGG----LQDNPTLLTAYVL---IALLEA-DQ-DEEE------------ALPCLNPQHIP--------------------- :  370
VaspA2M-3        : AFGESSRTEIGSIWLTAFVLKSFSQAAPYTF---IDEKVTNQSKEFIYKHQ-MKDGCFQKVGKVFHRALKGGV--GQSNDTRVLTAYII---TALLEAGEKPESEPMLR---------ALQCIKEMPDQ--------------------- : 1168
PsfuA2M-1        : AFG--ESDGNASTWLTAHVLETFNEASKYFKKDENFDYAILKSSFRFLKKVLGGKQCVKENGRVLHRDIIS-----NDEYARYEHLNLIGKIITALSSLNKNKNAHPHFEYNKKAIKQTLENVLTCVTENYKKEFSK------------- : 1087
PsfuA2M-2        : AFG--VRDKFGSTWLTAHVLEYFSLAMQYTDVNNDDMLNLIPKMLHFFSTMKQTIGFL-EKGRVIHRDFYE-----FSQSKIDLYKKINLNAKILDSLNQLEKALS---VYKIKQFNRTISMLINHSLKSLASLLSSVK----------- : 1024
PsfuA2M-3        : IWGPKDSEEDGSMWLTAYVVKAFSQASKYIE---VDKNLLQKSVKWIVKRQDSETGCFHNEGYAYSINSPRETLTSHVLVTLFEAKYTA----DLGDAVDSKV---IHK---------ALRCLKVSQKRKADDEESEE-SVEAEESEEKD : 1395
CaspA2M-1        : IWGPKDSEEDGSMWLTAYVVKAFSQASKYIK---VDEKLLKKSVKWIIKRQDFETGCFHNEGYAYSVNSPRETLTSHVLVTLFEAKYTA----GLGNAIDSKV---LNK---------ALRCLKNSQKEEINEEANEE-DANAG------ : 1408
CaspA2M-2        : IWGPKSSEEDGSMWLTAYVVKAFSQASKYID---VDLKLLKKSVKWILKRQDFETGCFHNEGYAYSVNSPRETLTSHVLVTLFEAKYTA----KLGNSIDSKA---IHK---------ALRCIKISQAPKDSAETNESPDESIEEMDDAN : 1381
PaspA2M          : IWGPSNTESEGSMWLTAYVVKAFSQASKYID---VDKNLLKESVRWILERQHRETGCFRNEGYAYSVNSPRETLTSHVLITLFEARYTA----NLSEIINSKI---FYR---------ALKCLSDSLNEEPEKPELEE-----------E : 1384
HosaA2M          : TFGERYGRNQGNTWLTAFVLKTFAQARAYIF---IDEAHITQALIWLSQRQ-KDNGCFRSSGSLLNNAIKG-----GVEDEVTLSAYIT---IALLEIPLTVTHPVVRN---------ALFCLESAWKTAQEGDH--------------- : 1140
AmspiTEP/CD109   : SFG--RTDNSGSTWLTAFVLKTFKQAKEHIPDM-IDDEVVKAAMSWLSKQL-QFNGTFAEPGSVLNKELQG-----GAKQGLPLTAYVLIALHENREY--YDSVAGA-----LDSARQAQSSLEQNYASI-------------------- : 1083
HaadiTEP/CD109-1 : AFG--NSDKNGSTWLTAFVVRSFHQAKEFME---IDEDIMSNSLEWLAKQQ-RADGSFGEPGEVHHKAMQG-----GSGSKSALTAYVLLAFLENQA------QKTF-----GQEMDKASKYLSKELKNS-------------------- : 1082
HaadiTEP/CD109-2 : TFG--ERDRSGSTWLTAYAIKSLSQANKYIY---IDPDILDKGIQWIMSKQ-SSDGSFEEPGEVHHKALQG-----GE-HGAALTAFVLVSLYEAKA------QNKY-----GNELSQAQRYVERELASS-------------------- : 1088
ScsuiTEP/CD109-1 : AFG--ENDSKGSTWLTAFVVKSFIQAKKYIS---IDENVIRESLIWLSRQK-NANHLFPEVGTVLHKEMQG-----GSSHGLGLTAYVLSAFLESKLNDVDSSD-RF-----HPLIDNSINILENYLEEL-------------------- : 1109
ScsuiTEP/CD109-2 : GFG--NRDSSGSTWLTAFVVKSFVQADKYIT---IDHTIIRSALNWLAKTQ-SPNGSFSEIGNVFYKSMQG-----GSSQDLGLTAYVLSAFLESKRNGFNEDQYLI-----ENVINTTVSVLERDLESI-------------------- : 1091
ScsuiTEP/CD109-3 : AFG--SRDENGSTWLTAFVVKSFGQALPYIT---IDEKVIENSLKWLALQQ-AENGSFPEVGSINNKEIQG-----GSSDGIALTAYVLLAFFQNKG----ETSH-------GPVVMKALQLLEKELEAL-------------------- : 1073
ScsuiTEP/CD109-4 : AFG--NSDKVGSTWLTSFVVKSFIQAKKYIT---IDENVIHSSLMWLSHQQ-FPNGSFPEVGTVFHKAMQG-----GSSQGLGLTAYVLSAFLESKMSGLDSSKELV-----DSVISSSIEILEKDLDSI-------------------- : 1098
NinoiTEP/CD109-1 : AFG--NSDKNGSTWLTAFVVRSFKQAKPYIS---VDDNVIDASLKFLQSSQ-AENGSFPENGEVHSKRLQG-----GAAGGLALTAYVVLAFLE------NSNKAEY-----HNVTKRAVKYLEDEIDQV-------------------- : 1100
NinoiTEP/CD109-2 : AFG--KQDKMGSTWLTAFVAKSFHQARSYIT---IDDNVIANALKWLASKQ-QGDGSFPEVGTVSHKAMQG-----GSANGLALTSYVLISFLENGA----QQSQEL-----NTVVEKARQNIEQVIQTV-------------------- : 1081
NinoiTEP/CD109-3 : AFG--NSDPSGSTWLTAFVVKSFAMAQTIIT---IDPNVIQAAKTFLVSTQ-QTDGSFAEVGKILDNIQGG-----SAISNCSLTAFVLIAFLENPS-----PGNTY-----VNVTTKATAYLENCLTQN-------------------- : 1098
EpspiTEP/CD109-1 : AFG--TSDKNGSTWLTAFVVRSFKQAKPYIS---VDDNVIDASLRYLKSTQ-KENGSFVENGEVHNKRLQG-----GAAGGLSLTAYVTLAFLE------NSDKTEY-----SNVTKRAVKFLEDEVDTV-------------------- : 1097
EpspiTEP/CD109-2 : AFG--KQDKMGSTWLTAFVAKSFHDARNYVT---IDENVVNGALKWLAMKQ-EGDGSFPEVGSVSHKAMQG-----GSANGLALTAYVLIAFLETGA----QQTQEF-----SSVVDKTRQNIEQVMLTV-------------------- : 1085
VaspiTEP/CD109   : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PsfuiTEP/CD109   : FFG-QEMEEKGSTWVTSLVIGNFIKASKYID---IDPDVIQLGLEWLANVQ-NDEGAFSEGGNIYFKDLQESNI--GLTA-FVVSALASEKDKLDA-----NSKNAL-----NRGISYLAKHWKDESSKA-------------------- : 1142
CaspiTEP/CD109-1 : FFG-KDMDEKGSTWVTALVIGTFLRASKYID---IDPNVLQSGLEWLAKAQ-NDEGAFSESGNIYFRHLQESNI--GLTA-FVVGALASQKEMLDA-----NSKNAL-----NRGISFLAKHWKDESSSA-------------------- : 1157
CaspiTEP/CD109-2 : FFG-KGIDESGSTWVTALVIGNFLKASKYID---IDPVVIQSGLDWLSKIQ-NEEGAFPEKGKTYFSDLQENEL--GLTA-FVVSALASQKEDLDA-----NSKNAL-----NRGISFVAKHWKDVSTQA-------------------- : 1043
HosaiTEP/CD109   : AFG--NYDPSGSTWLSAFVLRCFLEADPYID---IDQNVLHRTYTWLKGHQ-KSNGEFWDPGRVIHSELQG-----GNKSPVTLTAYIVTSLLGYRKYQPNIDVQES------------IHFLESEFSRG-------------------- : 1082
```

Supplementary Fig. S9 (continued)

```
AmspC3           : ------RRPYELAVATYALTLAK--SPKHQDAIKLLKDSSIYNKNED---QRQWEKGNGAL------------------------------------------------------------SVEATAYGLLTLVLTNDLP--------- : 1250
HaadC3-1         : ------HTPLAVAVVAYALSLSD--SELRQVANDKLLKLAKYDEDTN---RMYWNTENSAQ------------------------------------------------------------DIETAGYALLNQLLFNDMS--------- : 1251
HaadC3-2         : ------NHPLTASIVAYALTLGN--RALAQEANNNLLRMAKYDEDTN---RMYWTTDDSAH------------------------------------------------------------DIQTAGYALLTQLLLNDME--------- : 1255
ScsuC3-1         : ------QRPYTIAIVSYALALTN--STKRQEANQKLKNIAKFHQESY---TRYWNWDATEFG--------------------------------------------AGPKPWVYQHKPAAVAVETTSYALLAQLAYDDLD--------- : 1273
ScsuC3-2         : ------TNYYTMAIVAYALALNN--DDKANEANEKLKNMSYYSEENN---IRYWSWKKLHDS--------------------------------------------DLYRPWLYRSKPWSGDIEATAYALLTQLQLNNIN--------- : 1283
ScsuC3-3         : ------RQPYNLAIVAYALSLTE--NRRRYDVNEDLKNIAKFISGSN---VRYWNWDSLEFG--------------------------------------------TGNVPWIYQKKADAAAVETTSYALLAQLQFDEID--------- : 1278
PsfuC3           : ------GNPYAAAIASYALANENR-----FNRNILYKHVHQ-GLDHWP--SP---RGR-------------------------------------------------------------VITLETTAYALLALVRAQLFE--------- : 1238
CaspC3           : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PaspC3           : ------TNPYAVAMTSYALANEKK-----LNREILYKFAST-EFSHWP--TP---KGH-------------------------------------------------------------VYTLEATAYALLALVKAKAYE--------- : 1186
HosaC3           : ------QRSYTVAIAGYALAQMG-------RLKGPLLNKFLTTAKDK---NRWEDPGK-------------------------------------------------------------QLYNVEATSYALLALLQLKDFD--------- : 1247
HosaC4           : --------AHAAAITAYALTLTK--APADLRGVAHNNLMAMAQETGD---NLYWGSVTGSQS-------------------------------------NAVSPTPAPRNPSDPMPQAPALWIETTAYALLHLLLHEGKA--------E : 1280
HosaC5           : ------QSTFTLAISAYALSLGDK--THPQFRSIVSALKREALVKGNPPIYRFWKDNLQHKD-------------------------------------------------SSVPNTGTARMVETTAYALLTSLNLKDIN--------- : 1265
AmspA2M-1        : -------NSYTLALSAYAYILAG----RSDMSNKLIDKLFSRAIVEG---TDVHWEA-----------------------------------------------------------ASKSISVELGSYVILTLMKMGGTA------NQG : 1290
AmspA2M-2        : -------NSYTMALTAYAYALAG----RYELANKIIDSLFSRATIQG---TDVYWST-----------------------------------------------------------SSKSISVELGSYVILSLMKMGGAA------NQA : 1320
AmspA2M-3        : -------DSYALALSVYAYTLAG----RYEIANKLLDQLYSHATTEG---ADVYWAA-----------------------------------------------------------DSKSISVEIGSYVILSLMKLGGTA------NQA : 1314
AmspA2M-4        : -------SQYTVALTAYAYTLAG----RYNLTEKLLDDLFSKSSTEG---RGMYWPT-----------------------------------------------------------SSKSVSIELGSYVILSLMKLGGAA------NKV : 1257
HaadA2M-1        : -------NIYVLSLFAYASALAK----ENETYGRYLDELDKRAITKD---YMKYWEP---------------------------------------------------------SSNSKSVAVEIASYYMLARFEMEEAK------ALK : 1224
HaadA2M-2        : -------SLYATFLYAYAEALSD----KKDSAKERIESAKDRAITKG---KEVYYHD---------------------------------------------------------VNATKSQDIETSSYAILSILNSDG--------SAS : 1278
ScsuA2M          : -------DIYTLVLSTYASILAK----DENYTSLLMKRLLGLGISKD---NLLFWEK--------------------------------------------------------QSKKSLALNVEMSAYALLSLVSLGDQE------SIL : 1271
NinoA2M          : -------NTYHLAMQAYAASLAN----KTKLAQQLINTLKQKSVKKG---NMMYWSN--------------------------------------------------------SEVNSRAVDIETTSYVLLTLVKLNTED------NLK : 1231
EpspA2M          : -------NVYELSMKAYAAALAN----KTVLAKSIIKTLKQLAIQKD---NLMYWSN--------------------------------------------------------SGSGSDALDIETTSYVLLALATTDSGE------DMK : 1226
VaspA2M-1        : --------AFAHALAAYAWFLAY----QKHTLLSKTSNSSEALELKK---TGE-RFL---NKLLKTAKKEGQKMWWQHP-----------------------------------IKY-RDARVETAGYALLAFLTADPFD-------LK :  615
VaspA2M-2        : ------NTPYALALASY-AWLLS------YKREKDAGNSSEDGIEKL---LVGQELL---DKLMQLSHVNTGMVWWQQQ----------------------------------NCSECNTVDVETAGYAMMALIIGSPEE-------MS :  459
VaspA2M-3        : --------PYAQALGAYAFLLAN----GRTQWESSNTTTLLQESSNT---TTVLQESSNTTTVLQESSNTTTLLQEDSNPTISLNFQEAGESFLSQLLQTAKKESQLLWWEQEAGCKYCRDSNVEIAAYALMAFLLSDPTG------WS : 1296
PsfuA2M-1        : ------ENIQMMPAVLLARWNYA----LNSLSDRDKSIFNQIEFENK---LKSIVKEITSNIADEKG-------------------------------------------------MQNAEKVEVISLYLMTIDDKE------------ : 1162
PsfuA2M-2        : --------LASMSTVLTARILFA----LTHLKTKTLFSSQMTTSLFA---VSKSQVEIIVGYKK----------------------------------------------GKMTENLSKSNSVEVVAYFMMSRHIMTP----------- : 1101
PsfuA2M-3        : EQKDPSESLYALSLKTYAVHLLETKADLKEEIA---EAFEEIKGEEL---FKELMERSS-------------------------------------------RDDKGFLNWN--SNSSKSRSVEMTAYNVMTLLFND---------KRV : 1484
CaspA2M-1        : -AKNKTKSLYALSLETYAVHLLETQAELKKEIG---DAFDEIKGDKL---FKELLERSE-------------------------------------------RDEKGFLHWGPGNDSSLSMSVEMTAYNVMSLLFRD---------ELT : 1498
CaspA2M-2        : KSDQGGESLYAQSLKTYAIHLLETKPELKKEIG---EEFEEIKAKSL---FNELKEKSK-------------------------------------------KDDKGYLFWQ--TNNSLSRSVEMTAYNVMTLLFNG---------KHN : 1470
PaspA2M          : NKKNKQKSIYTQSLKTYAIHLIETKEDLKKEIG---ENFEGIKADTL---FKNLLDASK-------------------------------------------RDDQNLLHWK--NNSSKSRSVEMTAYNIMTLLFND---------ELI : 1473
HosaA2M          : -----GSHVYTKALLAYAFALAG----NQDKRKEVLKSLNEEAVKKD---NSVHWERPQKPKAPVG---------------------------------------------HFYEPQAPSAEVEMTSYVLLAYLTAQPAPT---SEDLT : 1229
AmspiTEP/CD109   : ------DDTYTLAIISYALQLVN----SPRRDAAFQQLASRA-RFGE---ETMYWS----------------------------------------------ANVRKPNNPEEVFHLPSSSDIEMTSYALMTYTLRGDI---------S : 1163
HaadiTEP/CD109-1 : ------KDPYFVSIVTYAFHLSE----HSEKDLALQKLLSLS-TRGV---DTIHWQRK-------------------------------------------------KDNSVDSYYTPQSQDVEMTAYALLTYSLRGDV---------A : 1159
HaadiTEP/CD109-2 : ------SNPYVVSILCYTLHLLN----SASRDRAFQMLLDLA-ERKD---DVVYWDNK-------------------------------------------ENQVNTTDKQSDYWFLAPSIDIETAAYAIRTYALRLDP---------S : 1171
ScsuiTEP/CD109-1 : ------KTDYDLVLVTYVLHLAN----SSVKDVAFEKMNMRS-KKDG---DKMFWTMSST------------------------------------------SSSDDKLDPFSYKSKPKSVDIEMTSYALMTYSLRNMI---------A : 1193
ScsuiTEP/CD109-2 : ------ESDYDLVFVTYVLHLAD----SPKKDQAFNLMNGRS-KTVG---DTKYWTVPLP------------------------------------------EVNET--YSYAYYYKPRSVDVEITAYALLTYSLRNMV---------A : 1173
ScsuiTEP/CD109-3 : ------TDLYSIVITTYALHVAS----SNFKEVAYQKLQAAA-TVKG---DLRYWQ--------------------------------------------------KPEPTAAGEIEVKSVDVEMTSYALMTYILQNDL---------S : 1149
ScsuiTEP/CD109-4 : ------ESDYDLVFVTYVLHLAN----SSSKDVAFQKMNERS-KTVG---DTKFWTMPLP------------------------------------------EINQS--DPYAYYNRPRSVDVEMTSYALLTYSLRNMI---------A : 1180
NinoiTEP/CD109-1 : ------EDPYELAIATFALEKAK----SPAKDKAFQKLDKKS-VKKG---DVVFWSKPLE------------------------------------------KPPTN---STIFYYPPPPVDIEMTAYVLQSYLERGRI---------T : 1181
NinoiTEP/CD109-2 : ------DDTYPLAIATYVLHLIG----SQERETAFRKLEQRA-VIQD---DQKFWTNAVEMNDAT-----------------------------------EDNTNNEKSKTTKSTSEINSYDVEMTSYALMTYVMRNDI---------D : 1172
NinoiTEP/CD109-3 : ------LTTYDITIVTYALVLAN----SSKSDAAFTKMNSMA-TNQG---DLTYWHYQVP----------------------------------------------KPNGTDWWYYTPPTYDVEATGYALLVNIVRKVTP--------I : 1179
EpspiTEP/CD109-1 : ------EDPYELAIVSYALHKAG----SPAKDAAFQRFLKKA-EKKG---DNVFWSKPLE------------------------------------------KPPTN---SSYFYYPPPSVDIEMTAYALLTHLERNLI---------T : 1178
EpspiTEP/CD109-2 : ------EDSYALAIATYALHLAG----SQERETAFRKLDQRS-ITND---EQKFWTGVVDNNDVL-----------------------------------DPPENDEPKKSKAAVREANSYDVEMTSYALMTYILRNDI---------D : 1176
VaspiTEP/CD109   : --------------------------------------------------EGEWKYW------------------------------------------------RREEPIGDTLYKPKSLDIEMTSYALLLYVERDDT---------A :   42
PsfuiTEP/CD109   : ------EENPFIL--SIITYAL-----SKTDHPDTGDAYNVL-KSFA---KRNETMGWEWI--EAK-------------------------------------VPPELEDNHWFNQSNSINNQITAYYALATLKNS-----------PE : 1224
CaspiTEP/CD109-1 : ------EENPYIF--SVITHAL-----VKTDHPEAGDAYNVL-KSFA---KRNETNGWEWI--EAK-------------------------------------IPEAQQENHWFNTSNSINNQITAYYALSTLKIS-----------PD : 1239
CaspiTEP/CD109-2 : ------EENPYIM--SLVAHAL-----AKTDHPDSNDALNIL-KSVS---KRNESMGWEWL--EAK-------------------------------------IPEEVENNHWYNQSNSLNNQMTAYYALATLKVS-----------PE : 1125
HosaiTEP/CD109   : -----ISDNYTLALITYALSSVG----SPKAKEALNMLTWRAEQEGG----MQFWVSSESKLSD--------------------------------------------------SWQPRSLDIEVAAYALLSHFLQFQT---------S : 1159
```

Supplementary Fig. S9 (continued)

```
AmspC3           : YAKAVVGWLNANRGFGGVWKSTQD--TVVALQAMSEYAIKSRQPK-----IDMAINITSETDRNFLKS------LRINNENFQDLQTIDINK-----PGGLIFVDAYGAG-IAHLNLRLQYNIPVPPNETCA-FELTIKNEEVEE----- : 1375
HaadC3-1         : RSNSIVNWLNTKQLQSGSFKSTQD--TVVALQAMSEYAIRAQMPS-----INLVANISSSNDRNFHKV------MAFRDDNALVLQDVRIDK-----IGGTVFINTAGHG-MGSLSVKLRYNLNVPVEDICK-FDVKVNVTESKP----- : 1376
HaadC3-2         : KSASIVNWLNTKKLATGAYESTQD--TMIALQAMSQYAIKARMPP-----INLITRISSNNNRFREIV------MALNDQNAQILQDIKINK-----VGGTLLINTMGQG-TGSLSVNVRYNVIVPPEDICK-FDIQVNVTDTKY----- : 1380
ScsuC3-1         : YAHPIVNWLNRQRSASGSFVSTQD--TVMALQALTEYNIKANVPA-----LDIMCNISSSASARVKRT------ILLKKDRPQEIQEIEVP------PKGRLYFDVTGKG-MGTLSLSMRFNVEKNPEDSCH-YDLTITSEEADE----- : 1397
ScsuC3-2         : YSHPIVNWLNQHRSYYGFLSSTQD--SVVTLQALTQYSVKARNPK-----MDMHCNIASTASSTARGA------FHLTSNNPLELYDLKIS------PQADLFVEAEGTG-LASMSLLMRYNVAQEPEKTCK-FHLNITVEEYDD----- : 1407
ScsuC3-3         : YSHPIVNWLTRQRSSGGAFVSTQD--TVITLQALAEYNTKTKIPL-----VDMQCNITSSETSRFKRS------IQLTKDKAQNIEEIEVP------PKGKLYVDVGGKG-IGSMSLSLQYNSEYTPETKCE-YDLIVKTHEYRD----- : 1402
PsfuC3           : DARPVVKWLGRQQKVGGGYGSTQA--TIMVYQAVAEYWTHAQKPE------YNLDVNLGL-PG-RSNPIR----INFNRNNHYTTRTSKFND-----INQDIQVTARGRG-EATLTLVSLYYALPKEKESDCDK-FNVSVKLTEE----- : 1362
CaspC3           : ---------------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PaspC3           : DAKPIVRWFNRQQKVGGGYGSTQA--TIMVYQAVAEYWANAQEPE------YDLNVDVLL-PG-RSRPEK----FKFTRDNQYTTRSSKIND-----INQNVTVTATGSG-EATFTMVSLYYALPKEKESDCQK-FDLKVELIPD----- : 1310
HosaC3           : FVPPVVRWLNEQRYYGGGYGSTQA--TFMVFQALAQYQKDAPDHQ-----ELNLDVSLQLPSRSSKIT------HRIHWESASLLRSEETKE------NEGFTVTAEGKG-QGTLSVVTMYHAKAKDQLTCNKFDLKVTIKPAPETE--- : 1374
HosaC4           : MADQAAAWLTRQGSFQGGFRSTQD--TVIALDALSAYWIASHTTE----ERGLNVTLSSTGRNGFKS-----HALQLNNRQIRGLEEELQFS-----LGSKINVKVGGNS-KGTLKVLRTYNVLDMKNTTCQDLQIEVTVKGHVEYTMEA : 1413
HosaC5           : YVNPVIKWLSEEQRYGGGFYSTQD--TINAIEGLTEYSLLVKQLR-----LSMDIDVSYKHKGALHN-------YKMTDKNFLGRPVEVLLN------DDLIVSTGFGSG-LATVHVTTVVHKTSTSEEVCSFY------LKIDTQDIEA : 1388
AmspA2M-1        : KALGIVRWIARQRNSNGGFVSTQD--TVIALQAFAKVAAILNRN-----KQD--LKVTAEGNGFYK-------EYAVNSTNRLLMQMHRIE-----ELPNIVDVFATGDG-CGLIQTTLKYNKKN-VNASDAFD------LTVVG----- : 1406
AmspA2M-2        : KALSIVRWIARQRNANGGFVSTQD--TVIALQAFAKFAVIHSKN-----KQD--LEVIAEGNNFNQ-------KYAINSTNRLLMQKDKVV-----ELPNIIDVSAVGDG-CGLIQTTLKYNKDN-VNASDALE------LIIVG----- : 1436
AmspA2M-3        : KAMQIVKWITRQRNANGGFVSTQD--TVIALQAFAKFAVHLNRN-----KQD--LEVAVEGNGLNG-------KYAINSTNHLLMQTNKIE-----ELPNIIDVEAVGEG-CGLIQTTLKFNKNI-ANASEVFD------LTVKG----- : 1430
AmspA2M-4        : KAMDIVRWIARQRNSNGGFVSTQD--TVLALQAFAKYSVTLTQN-----AQD--VAVTAKANGFDH-------TYAVKSTNRLLMQTDKIE-----ELPNIVDVQATGSG-CGLIQTTLKYNRNN-VNASEAFE------LSAVG----- : 1373
HaadA2M-1        : SVLPVVRWITHQRNSYGGFISTQD--TVVALQALAKYASYISKN-----PVD--IALAVETDDMTQ-------GFKLDESNKLVTQQLKIV-----DLPTTVDIDAYGDG-CAVVQFSLRYNVEK-VSNTGGLE------LNVNA----- : 1340
HaadA2M-2        : DALPIVQYLTKNMNPRGGFFSTQD--TCVGLEALGQFSEMTFKD-----EVD--ITITATG-DIEK-------NIEITEDEKLLVKRYKVN-----EVPSEINIEATGSG-CAVIQQIFRYNSKT-SPEKRSFH------LEALG----- : 1393
ScsuA2M          : KAQKVFRWITQQRNSHGGFISTQD--TVLALQALSEFAGKFQSN-----ELE--IEISVEAGKLNH-------VFEVNNENKLVQQIIKIP-----EVPTVVDFIALGKG-CSILQTVLKYNVEH-TEGSDAFN------LDIRSE---- : 1388
NinoA2M          : LALKVVRWLTQQRNAYGGFVSTQD--TVLALESIADYLSLIPHG-----SSN--VTIDLEANDLIH-------SFNINDENRLLLQSVPIP-----YLPNVLDIHVSGRG-CALIQVSVKYNVLG-NQSSSVFN------LDIYDV---- : 1348
EpspA2M          : MAIDIVRWLTKQRNAYGGFKSTQD--TVLALQALVKFISQTPKI-----NAN--LSLALEANDFTD-------EMFVTEENRLLMQTRDIS-----VLPNMLDVQISGKG-CSLIQVTLRYNIPS-AQLLPAFT------LNISTS---- : 1343
VaspA2M-1        : TIRPIARYLISQRNIKGGFYTTQD--TVVALEALTKYSMQADEV-----PLKNITVRCAA-TTSFL-------THVISPKNRLVTYKTDV-------EPFTGNVVVQGEG-CVVAQCSVKYSTPK-ARN-IKGF-----EVNVTREC--- :  732
VaspA2M-2        : TIRGIARWLVKHRNPNGGFYSTQD--TVVALEALTKLAMLSDEI-----PLKQVQVKTQL----QH-------TWNITEENRLVTHLKEGM------TSFLSDVSISGEG-CVLAQCSVRYNVLE-ASERSSGF-----SLEIKGLP--- :  575
VaspA2M-3        : NLRPIARYLISQRNTHGAFYSTQD--TVVALEALTKYAMQAEQV-----ELKQVAVTCVS-CAVTH---------TVNAENRLITDQTSVG------ESLTGNIVVQGHG-CAIAQCSIKYNIPE-AKPSVAFN------IKASGWP--- : 1412
PsfuA2M-1        : VELQLISWLLSQQNSRGGYYSSWD--TSLAVRALAAKTVNPIEQ-----SKSVTFSMDGVQKKMQ---------IEPNSKTILIENKFMKEKKSFVIKWTTDLKESKNEC-LVGRVAAIFYSEKSNKKEK---EYFKINHKIIENEKLK- : 1291
PsfuA2M-2        : LYHSCYLYMLRAQNSKGGFISSTD--TVVALRAFAQSLFKTIDK-----DEIISLKILNKTKAKLLK--------PSSKVNFDLNINYYTPPSNRKVSYVWEKKSSSNKC-VIAKVTSVFYVMSPIL----IIEIFSMKSKIIKQDDFR- : 1230
PsfuA2M-3        : DALSAIRWISGFRNQMGGFISTQD--TVVALQAISSYSTLVFEN-----NTSLDVQFFNETSLVED--------FSIDEDNKLLFKRIKID------SLRDLKVSSTGKG-CYTISTMVRYNINN-ETEENAKF------IILAEGNS-- : 1603
CaspA2M-1        : DALSAIRWMSTHRNQRGGFVSTQD--TVVALQAISRYSSLVYEN-----ETSLDVKLFNGTGLLKD--------FKIDEDNKLLFKRMKVD------GLVDLKVSSEGKG-CYALSTMARYNINN-QTDESPKF------QIRTDGNS-- : 1617
CaspA2M-2        : DALSAIRWISGHRNERGGFISTQD--TVVALQAISMYSTMVYEN-----DTSLSVQFSNKTSEIDS--------FDLNEDNKLLFNRIKIN------DLRNLKVSSTGKG-CYTFSTMVRYNVKD-EKDKNAKF------LIRADANK-- : 1589
PaspA2M          : DALSAIRWISKYRNGRGGFISTQD--TVVALEAISKYSSRVFEN-----ATDLSVDLYNTTGTIES--------FHIVEDDKILLRRIDVD------EVKDFKVSSVGKG-CYSFQTSIKYNTKS-NEKEEEKF------FLKAETNE-- : 1592
HosaA2M          : SATNIVKWITKQQNAQGGFSSTQD--TVVALHALSKYGAATFTR-----TGKAAQVTIQSSGTFSS-------KFQVDNNNRLLLQQVSLP-----ELPGEYSMKVTGEG-CVYLQTSLKYNILP-EKEEFPFA------LGVQTLP--- : 1349
AmspiTEP/CD109   : TSLRIMKWLVERRNSLGGFTSTQD--TVIGIQALTMLTNNLNIR-----GSNLEITYSYNENADHNNTFPIRKKINVNDQNSLNMQSR-TLPVTV----RKVRISAEGRG-IAMVQVSWSFNLKVSAP-NPSFG------LNPLVDKV-- : 1291
HaadiTEP/CD109-1 : GSLPILRWLISRQNENGGYSSTQD--TVVGIQALASLAFRLAST-----SISLNVSYSTVDTSN---------VLAINSENAMILQKV-MLPPDT----RSVKVQATGFG-VGIIQVTWQYNIESTGK-VPSFA------LKPVLGKA-- : 1278
HaadiTEP/CD109-2 : GALPVLTWLITKQNRKGGFSSTQD--TVVALHAISEIAPFISPP-----VSNINVKFMYPDGQQ---------DMQVTSSRPLDVHEI-EIPSDV----PYVEVETSGSG-VAVVQVSWSFNLAVSGE-APQFF------LNALLDKT-- : 1290
ScsuiTEP/CD109-1 : EGLPIMRWLLSKQNANGGFQSTQD--TVVGIQALAQIAKEISFSDDD---FHLDVKFFYEGGEKI---------MSLTKDNDLVLYIE-QIPGNV----RQIDIQASGSG-FGIFQVSWSYNVLTLQE-NPPFE------VGIEINNE-- : 1314
ScsuiTEP/CD109-2 : EGLPIMRWLLSKRNAYGGFESTQD--TVVGIQALAEFTKHLYYSD-----SNVQAVFSYDGGANA---------MILTNENALVLHKE-KIPSKV----RDIEVSASGKG-IAVLQVSWSYNVLHTEE-HPAFE------ITLQFDPF-- : 1292
ScsuiTEP/CD109-3 : EAMQILKWLISERNSNGGFKSTQD--TMIAIQALAKLAQRISDP-----QVKITVTFTYSG-QQK--------TFSLNRENAMILQTD-EIPAVE----KNVNISATGYG-FGIVQVSYQYNIMVSKE-FPAFQ------VNPLVDRS-- : 1268
ScsuiTEP/CD109-4 : EGLPIMRWLLTKRNSNGGFESTQD--TVVGIQALAHYAKKISAGDG----SNMKVKFSYKDGEKE---------LELTKENALILHRE-QIPGST----REIDISATGKG-LGILQVSWSYNILTVQE-RPAFE------ILTDVTNE-- : 1300
NinoiTEP/CD109-1 : DAIPIMRWLITQRNENGGFSSTQD--TVVGIQALASMASHITSPDG----AKMDVNFDYGGRSKR---------LTLDKDNAMVLQRE-ELPPET----RNVSIKAEGKG-VGIVQITWSYNLDNKTQ-SPVFE------ITPTVKKA-- : 1301
NinoiTEP/CD109-2 : NALMIAKWLIAQSNSNGGFTSTQD--TVIGIQALATLAEQISSP-----RRNLDITFSFKN-SSS--------KIVINNDNAMILHKS-TLFDDT----DEVRFTARGSG-FAIAQVSYSYNVNITQE-KPSFI------INPLVDRN-- : 1291
NinoiTEP/CD109-3 : NLLPIVRWLISTRNSYGGYSSTQD--TVIATQALGAAAPSFVGAPN----STVTMTVKYGTNTSN---------FNFTKQNQLVLQQQQLLPNTI----PSVQVSANGNG-VAIAQVSYNYYVVINNTGNAAFN------LQLQAYQNSN : 1303
EpspiTEP/CD109-1 : EAVPIMKWLITQRNENGGFSSTQD--TVVGIQSLAGIASHITSPDG----AKMELDFEYDGRHKK---------LTLDKGNAMVLQRE-ELPSET----REVNVGAEGKG-FGIVQVTWSYNLDNKTQ-SPVFE------ISPKVKQI-- : 1298
EpspiTEP/CD109-2 : NALKIAKWLISQTNSNGGFTSTQD--TVIGIQALAKLAQQVNSP-----RRSVDVTFSFNN-STK--------TVPINNENAMILHKY-SILDPV----DEVRFTARGSG-FAIAQVSYSYNVNITQE-KPSFI------INPLVDRN-- : 1295
VaspiTEP/CD109   : NAVPSMKWLIARRNSQGGFFSTQD--TVVGIQALATLATVLGSG-----VTDLNVEFEYETGIKN---------AKIDQENVVMLQTF-ELPSNT--SNVTIRASGRGVG-LVQVSWSYNVQVSAA---NPAFT------LDPQVSRV-- :  161
PsfuiTEP/CD109   : GAYPVMNWLVAQQNSMGGFASTPD--TYVGIEALKEYDLTLNIPQRK---SKITVDYSYLD-NSR--------SREIDPESLTISKRI-VLPSDI----RNLTMKAQGAGAVGVVNVEYSYNLNVTAS-WPSFV------LNPQLLD--- : 1345
CaspiTEP/CD109-1 : RIIPIMNWLITQQNSMGGFASTPD--TYAAIEALHNYDLNL--PQRR---SKISVDYSYLD-NSR--------SREIDPESLTILQRR-VLPAYI----RNLTVKATGADAVGVINVEYAYNLNVTAS-YPSFV------LNPQLLDP-- : 1359
CaspiTEP/CD109-2 : KASPIMAWLVAQKNSIGGFASTPD--TYIGIEALKEFDLTLNIPKST---SKIDVNYSYLD-NSR--------SREINPEASTISIRR-VLPSNT----RNITLEAKGTNAIGVINVEYAYSLNVTAS-WPSFV------LNPQLL-P-- : 1246
HosaiTEP/CD109   : EGIPIMRWLSRQRNSLGGFASTQD--TTVALKALSEFAALMNTERT-----NIQVTVTGPSSPSP-------VKFLIDTHNRLLLQTAELAVVQP----MAVNISANGFG-FAICQLNVVYNVKASGSSRRRRSIQNQEAFDLDVAVKEN : 1290
```

Supplementary Fig. S9 (continued)

```
AmspC3           : --DIKPKALAG--------------------LENDALDQLPEDLQRTVGVQQRADVFENLDNWEEDGRDRVGRQKEASKKVLKIEVCAKFKGPEE-------------ETSMAMIDVGIFTGYAVIKDDLENLLN-EAASPIQQYELSS- : 1488
HaadC3-1         : --EQQEKDPWDEVLNQKGKDMFERFPEDLIRGIKKELDEVPPELAIDPRRRGHVRASLIASPLIKKDRQKRQKPNENSKVKLNITICVRYLG-NK-------------DTEMSIVDAGIFSGFQPVEDDLIWLQD-DHSHLIQRYEKSS- : 1508
HaadC3-2         : --EEIIDD---------------------------------------------VRSHLNSGRDIPLTESR-----PSSKLKYRISICVRYLG-DA-------------DAAMSIVDVGIFSGFQPVKEDLIKIED-EPSQLIQKIELSK- : 1462
ScsuC3-1         : --IIKPVNLKP---------------EFAGNDILPAAVVRSVFDERVQKEKFGYDVEGKDENNPAVDRAGKEIYNVGSQHVIKINICVKYKEKQ--------------NAGMSILDVGLFTGYKPIKEDLITLTM-KKELKVGQFEITD- : 1514
ScsuC3-2         : --IIRPVAPTG---------------ELEGIDIIPENVTRSLFTQNELRDRFG--IRDENEAAEDSDEE----NNGEKVHVVELNICMRYLEKEG-------------NSGMSILDVGLFTGYSMKMEELKNLIR-SIETSLTQFEQNE- : 1519
ScsuC3-3         : --SFQPPNLAE---------------YANG--PIPDNLKEEIINKNVQ-EVFNEAVNRKRDGSHDEEKD----NANNQEHLININICVKYQDKE--------------KAGMSILEIGFLTGYRIDKEELSKLEN-RPKVKC--VETSD- : 1510
PsfuC3           : ---------------------------------------------------------------------QSQGDD---LTYKLTIKVLFKSRDRD--------------ATMSILDIGLLTGFTVDTKDLDLLST-GRARLISKYEMNKA : 1425
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : ---------------------------------------------------------------------KLGDDE---KIYKLKINVLYKSREHD--------------ATMSILDIGLLTGFTVNTNDLDRLAK-GRSRIISKYEMNTA : 1373
HosaC3           : ------------------------------------------------------------------KRPQDAKN------TMILEICTRYRGDQD--------------ATMSILDISMMTGFAPDTDDLKQLAN-GVDRYISKYELDKA : 1437
HosaC4           : N------------------------------------------EDYEDYEYDELPAKDDPDAPLQPVTPLQLFEGRRNRRRREAPKVVEEQESRVHYTVCIWRNGKVGLSGMAIADVTLLSGFHALRADLEKLTSLSDRYVSHFETEGP- : 1520
HosaC5           : S----------------------------------------------------------------------HYRGYGNSDYKRIVACASYKPSREESS---------SGSSHAVMDISLPTGISANEEDLKALVEGVDQLFTDYQIKDG- : 1458
AmspA2M-1        : ----------------------------------------------------------------------NHYRSNCKK--RELRICTRYK-------------LLNERSNMAIVTVKMVSGYIPIKD---NLKQLKNDRTLNLKRYEV- : 1467
AmspA2M-2        : ----------------------------------------------------------------------KADRWNCKR--PQLDICARYK-------------ILGEKSNMAVISVKMISGYIPVKS---LLADLKDVPELNLKRYEV- : 1497
AmspA2M-3        : ----------------------------------------------------------------------RFHRRDCDKDKHTIDICAKYK-------------ITNEKSNMAVISLKMISGYIPVKR---VLNNLKFDEELNLKRYEI- : 1493
AmspA2M-4        : ----------------------------------------------------------------------NTHKQGCKL--RTLDICATYK-------------IPKENSNMAVITVKMISGYIPVKD---SLARLKKDKKLNLKRFEV- : 1434
HaadA2M-1        : ----------------------------------------------------------------------RRR-GSNECNLPSLGICMRYA-------------VHKEKTNMAVLSVKLPSGYVADEW---SLLLLENDKEVQLMRHEI- : 1402
HaadA2M-2        : ----------------------------------------------------------------------KCS-DDD-CKKATISLSFSYI-------------PEGKKTGMSVLEVKMVTGMSPVKD---SLEKLLGDKRSKVMRYDV- : 1454
ScsuA2M          : ----------------------------------------------------------------------NIGTTSAACKRHRLEICARYL-------------LEDEFSNMIVIEIKMVSGFEPDKK---SLAELLEKKDIKLKRWDT- : 1451
NinoA2M          : ----------------------------------------------------------------------SNNSTKNSTKNKILEICTNYQ-------------GSDENSNMVVVEIQMITGYETEGS---HLEKLKNPITQLKRWEIE- : 1411
EpspA2M          : ----------------------------------------------------------------------SNEISNDINHRQTIKICTKYD-------------GADNKSNMAVVEIQMITGFEALHS---HLEKLKTSNIDLKRWESE- : 1406
VaspA2M-1        : ------------------------------------------------------------------------IEGGIKP-TVQLDICVSYV-------------LDDGESNMAIVEVNLLSGYLDSPYS--LNDLYSGGVVSRFEPGED- :  793
VaspA2M-2        : ------------------------------------------------------------------------AEESCES--RELLVCVS-F---------------EGISNMAVVEMHLPSGFAPVAEK--LDYLMNDVYFALMKWELD- :  632
VaspA2M-3        : ------------------------------------------------------------------------LDEPARCRYSELLICISYV-------------LDTGESNMAVVEIGMASGYKPELL---SLDELQNNLDIDLKRYEV- : 1473
PsfuA2M-1        : --------------------------------------------------------------------------------MFKLNIDVEPID-------------SKARSGMLLVEVSIPSGFKVVYNNLDKLIGKGVERPTRYEKYFT- : 1347
PsfuA2M-2        : ---------------------------------------------------------------------------------KVYSVSLSLKK-------------VIKMTGMLLMTIRIPSGYTFNMNYLTNQMKKGYFVKKCEENMKKG : 1286
PsfuA2M-3        : ---------------------------------------------------------------------------------SYLHICASYI-------------GEKEKTNMVLLEVELLSGYEVDED---SLSSLLNEVDSGVEKYEVN : 1656
CaspA2M-1        : ---------------------------------------------------------------------------------SYLHICASYI-------------GDRERTNMALLEVELLSGYEVLED---SLALLLNEVESGVEKYEIM : 1670
CaspA2M-2        : ---------------------------------------------------------------------------------TALHICSSYI-------------GEKSVTNMVLLEVELLSGYDIVES---SLEVLLNEVQSGVEKYEVL : 1642
PaspA2M          : ---------------------------------------------------------------------------------TFVNVCGSYK-------------GQSSVTNMIILEVELLSGFEVVES---SIEILLNEIDSGVQKYEIE : 1645
HosaA2M          : ----------------------------------------------------------------------QTCDEPKAHTSFQISLSVSYT-------------GSRSASNMAIVDVKMVSGFIPLKP----TVKMLERS-NHVSRTEV- : 1410
AmspiTEP/CD109   : ----------------------------------------------------------------------------STKGYLQVSSCI-----NYIPE---------GESGMAVMEFYAPSGYVVDR---SSLSSIRQESIIKRVETYD- : 1347
HaadiTEP/CD109-1 : ----------------------------------------------------------------------------STDDYIELDICT-----KYT---------QEGASNMAVMEVGLPSGFQADS---ETFPAIKKLDKIKRIETQN- : 1334
HaadiTEP/CD109-2 : ----------------------------------------------------------------------------STASYLQLSICT-----HQRER-------RNDTSNMAVMEVGLPSGYVADV---DALPSVLQIPKVKRVETQL- : 1348
ScsuiTEP/CD109-1 : ----------------------------------------------------------------------------NNE--LAVEACVNASRYLYESH--------GET-NMAVMELALPSGYVADK---EHLPHVDARKLIKRVETKD- : 1373
ScsuiTEP/CD109-2 : ----------------------------------------------------------------------------TFG--IVVGACT---KYIYEDG--------GQS-NMAIMEFGFPSGYLVDK---ERLPQLVN--SIKRVETKN- : 1346
ScsuiTEP/CD109-3 : ----------------------------------------------------------------------------STKNRLQLNVCA-----AYGEK--------NGVSNMAVMEVTLPSGYVIDR---DSLPALHRVDEVKRVDVKD- : 1325
ScsuiTEP/CD109-4 : ----------------------------------------------------------------------------NNE--MTVKACT---KYVYEDE--------NES-NMAVMEIGLPSGYVADK---EHLPSIDESKSIKRVETKD- : 1356
NinoiTEP/CD109-1 : ----------------------------------------------------------------------------SDDS-FDLIVCA-----VYDDE--------DHSSNMAVIEMNMPSGYVVDG---ESLPQVGKQDRLKRVDTVD- : 1357
NinoiTEP/CD109-2 : ----------------------------------------------------------------------------SNRNLLVINACT-----SYMQ---------NGSSNMAVVEVDLPSGYTVDK---DSLPALLKMKDIKRVESKT- : 1347
NinoiTEP/CD109-3 : A---------------------------------------------------------------------------NNTNGYYLQICAS-----YQKN---------DSTGMTIITASTLSGYIWDP-----LPVPGSPAQLINVETGN- : 1358
EpspiTEP/CD109-1 : ----------------------------------------------------------------------------DKDS-FDLTVCA-----IYEDV--------DKSSNMAVIEMNLPSGYVVEG---ESLPQVGQHDRLKRVETVD- : 1354
EpspiTEP/CD109-2 : ----------------------------------------------------------------------------SNRNLLVINACT-----SYTR---------NGSSNMAVVEVDLPSGYTVDK---DSLPALLKMKDIKRVDTKA- : 1351
VaspiTEP/CD109   : ----------------------------------------------------------------------------STKNYLRVSVCT-----GYHYK---------GDTNMAVMDVSLPSGYTVDED--AILSLYRYSSNIKRVEERD- :  218
PsfuiTEP/CD109   : ----------------------------------------------------------------------------PPKAQFILNSCV-----NFIFYS------NIEASNMAVMEIDLPSGYTVDA---TSLPSLKR--YQGVTRVEAT : 1403
CaspiTEP/CD109-1 : ----------------------------------------------------------------------------QKKAQFILNACV-----NYIFYS------DSDASNMAVMEIDLPSGYTVDA---TSLPSLKRYQGVTRVEASRR : 1419
CaspiTEP/CD109-2 : ----------------------------------------------------------------------------APPAQLILNSCV-----NYIFYS------NKESSNMAVMEIDLPSGYTVDA---NSLPSLKRYEGVSRVESSLR : 1306
HosaiTEP/CD109   : K---------------------------------------------------------------------------DDLNHVDLNVCT-----SFSGPG---------RSGMALMEVNLLSGFMVP------SEAISLSETVKKVEY--- : 1342
```

Supplementary Fig. S9 (continued)

```
                                                                                                                  /       C345C domain
AmspC3           : ---------RGVVFYLDTISSKQK--TCIKFRAKQTFRVGV-AQRASVVVYDYYKPELTCTAFYAPGETSVIKKSL-CSGN-----KACICPRGGCSACKAFDDVLKLSGNIFQKMTRELLHQTCNM----HDYVWVASVV-NIRLENEL : 1615
HaadC3-1         : ---------RGVVFYLQKVPIAGD--YCFSFHVVRQYIVGN-TQTSVIKVYDYYNPDATCTKFYSPGSNSPMLRTI-CEGG------ICECAEGGCPPNNPFSEISSMS---TAQRREELKTIFCEN----YDYVWKGKLNGRLQKDGGF : 1632
HaadC3-2         : ---------RGVIFYLTKVPSTVK--YCFHFRVLREYIVGN-TQMSYIKVYDYYNPDATCTKSYRPISNSPVIRTI-CEGG------ICQCAEGGCPPRTPFADLEGMQ---VSERRNLLKVKFCED----FDYVWKGELH-YIRKEGGF : 1585
ScsuC3-1         : ---------RSVILYLDEVPINKP--ICLNFRASKEIHVGK-VQPTAVKIYNYYDPDKSCTQFYGPDKGSVMLKKI-CEGK------QCVCVEGICPSCFPFQDIRNIAN--DIDRRMELLNAVCDR---KTDYFWNGTIK-NIREDGSF : 1639
ScsuC3-2         : ---------RSTVLYFDEVPNKER--MCISLRTYQDFHVGK-VQPASVKIYSYYEPSKSCTKFYAPRD--RSPMLTKI-CEGK----QCFCAEGKCPSSTPFKEIRAKIS--DTERRRALLDIACHK-----SQHFIWNVTLE-KITYEN : 1642
ScsuC3-3         : ---------RALILYLEEVPNDRT--ICLDVKLRKFITVGL-VQPTTVKIYNYYKLDKSCTTFYGPDEDSVMLQTI-CEGK------QCRCMEGACPPLNPFKHVWEKNE--EKERNKELLRIICDDKKTKDNFVWLGILK-S------- : 1631
PsfuC3           : LS-----EKGSLIIYLDKVSHTRP--EEISFRIHQKLKVGV-LQPAAVSVYEYY-EETPCVKFYHPERRDGQFLQL-CRNN------ECTCAEENCSMQKKGKIDNSLR-----------TEKSCETTPTNKIDYIYKVKVESVTSELIT : 1548
CaspC3           : ------------------------------------------------------------------------------------------------------------------------------------------------------ :    -
PaspC3           : LS-----ERGSLIIYLDKVSHTRS--DEITFRIHQTLKVGV-LQPAAVSVYEYY-DQTPCVKFYHPERKAGQLLRL-CRND------ECTCAEENCSMQKKGKIDNDER-----------TAKACETEENSKIDFVYKVLLEDFKADLST : 1496
HosaC3           : FS-----DRNTLIIYLDKVSHSED--DCLAFKVHQYFNVEL-IQPGAVKVYAYYNLEESCTRFYHPEKEDGKLNKL-CRDE------LCRCAEENCFIQKS--DDKVTLE--------ERLDKACEPG---VDYVYKTRLV-KVQLSNDF : 1558
HosaC4           : ----------HVLLYFDSVPTSR---ECVGFEAVQEVPVGL-VQPASATLYDYYNPERRCSVFYGAPSKSRLLATL-CSAE------VCQCAEGKCPRQRRALERGLQDEDGY------RMKFACYYPRVEYGFQVKVLREDSRAAFRLF : 1643
HosaC5           : ----------HVILQLNSIPSSD--FLCVRFRIFELFEVGF-LSPATFTVYEYHRPDKQCTMFYSTSNIKIQ--KV-CEGA------ACKCVEADCGQMQEELDLTISAET--------RKQTACKPEIAYAYKVSITSITVENVFVKYK : 1578
AmspA2M-1        : -------DANYVNLYFDYLGNDQ---TCFKMNVEKEIDVED-AKPATITVSDYYTPEKMLEKSYTLPA-DRC------------------------------------------------------------------------------ : 1527
AmspA2M-2        : -------DANYVNFYFDYLSNKQ---TCFALHVEKEIDVED-AKPAIASVYDYYVTELKLEKSYSLP-VVDCKKTPHVEPLDDVPGLVELTTAAAVESTTVDDEQSTTVDDEQSTTVDAEQSTTAGAQE--------------------- : 1614
AmspA2M-3        : -------SNNFVNFYFDHLNNDQ---ICFSIDVEKEIEVEE-TKPATVSVYDYYNSELKLDKSYELPS--TC------------------------------------------------------------------------------ : 1552
AmspA2M-4        : -------DANYVNFYFDFLGNEK---TCFTIHLEKEIDVED-AKPATISVFDYYISELKLEKSYSLPSVANCF----------------------------------------------------------------------------- : 1496
HaadA2M-1        : -------EENVVNLYFEEITNDA---RCFEFHVKSEFEVEN-VMPSIIRLYDYYQPDRQVTKDYSIPS--TCNSTFLPDLTRFPLFKSSEPLHSDFDEFQEFSDTLNGELPEIITPPETTDYQTESRNVSGSDWNSEETPDHLNDSLYPL : 1539
HaadA2M-2        : -------EDNTVVMYFNQVENEE---MNIAFDVEKVVEVEN-TQPGIVKLYDYYNRDVSSSTNYSF-----CGKSESCSTEP-------------------------------------------------------------------- : 1520
ScsuA2M          : -------EGDQLNLYFDQLNAQE---KCFSISITEKVEVKD-TKPAIVTIYDYYQPELFVRKNYSIE---GCNKETLAPFTDEELTELESLIVQGLDDFEKPSSKSSQTSTVSIPGENGKHPVEEIIPDPVYIPPLGTQEKEEFKSSKEQ : 1587
NinoA2M          : -------KDGTVQLYFDELTGLN---KCFHVLVNEIIEVKD-RKPATVKVYDYYQPEKSVSKEYSFLELPAPS----------------------------------------------------------------------------- : 1473
EpspA2M          : -------NDGTVELYFDQLESEL---KCFDVIVEEKLKVQN-RKPALIKVIDYYKTDESVSKEYFLES---------------------------------------------------------------------------------- : 1463
VaspA2M-1        : -------TGGTANLYFDSFNKTE---TCFSIQQFQINKVFN-LKPAFVKVYDYYRPEITASESYSA-----CEDAP-------------------------------------------------------------------------- :  853
VaspA2M-2        : --------KNKLHL---------------------------------------------------------------------------------------------------------------------------------------- :  638
VaspA2M-3        : -------EENKVNLYFDYFDQNQ---KCFTIRMDQDTVVTN-PKPANVKVYDYYQTELTTSTSYSI-----CNDAPIPAPPPPGI----------------------------------------------------------------- : 1542
PsfuA2M-1        : --------ATR-IQLYFWGFGSVNRTRSFQLEFVQ-QVKVKDAALGVIHVEDYYSSEKNAASYYEVPFVNE-----------------NKEVNAQIVEILNNLNDKAIDNLDEHLNKFSEEICEQKLKEEVKVVILKDEKQIGLKLSEND : 1470
PsfuA2M-2        : -------YVKVYLSALGGKVQKIS--VNFELLHEF-KVNSS--PNVVVIVEDYYNPVVKGEITIKSK----------------------------------------------------------------------------------- : 1341
PsfuA2M-3        : Q------EDRTFVLYFNGLSKGE--NHCWDLEQKQVSQVDN-LKPAIIKIFDYYSQEDSYSTTYII------------------------------------------------------------------------------------ : 1713
CaspA2M-1        : NGDDDKDEDRKFVLYFNALEQGR--NACWDVELKRVNLVKD-LKPAIIKIYDYYAQEDAFSTSYII------------------------------------------------------------------------------------ : 1733
CaspA2M-2        : K------DERKFVLYFNGFKKEE--NHCWNLEVKEVSQVEN-LKPALIKIFDYYSQEDTFTTTYNI------------------------------------------------------------------------------------ : 1699
PaspA2M          : E------KEGKFVLYFDNMKKDE--LKCWDFEVKRVSPVEN-LKPALVKIYDYYSQEDSFTTSYTI------------------------------------------------------------------------------------ : 1702
HosaA2M          : -------SSNHVLIYLDKVSNQT---LSLFFTVLQDVPVRD-LKPAIVKVYDYYETDEFAIAEYNAP----CSKDLGNA----------------------------------------------------------------------- : 1474
AmspiTEP/CD109   : -------DETMVAIYFDKIGKE---PVCPTVSAYRVHRVAN-QQQKPVVVYDYYNRAQIARVFYR--LAHVTKKEEIC-DGYEC-DKSNTRRN--------DASVDQRDKNSSVSLNSFSNILILTQVLTFLTFQVLKSIV--------- : 1465
HaadiTEP/CD109-1 : -------GDTNVVIYFDRIDGN---EMCVNVPAFRNHKVAN-QKPVPVKVYDYYDLAKSARMFY---EPRVVDVCSLC-ESTDC-PDTCNKVPG--------------VSSSSIAMYSVALITFATIAVSFFKSIH-------------- : 1440
HaadiTEP/CD109-2 : -------QDTGVVIYFDRLDRE---ESCVTVPAHRIHKVAH-QRRAPVKVYDFYSQAKSARMFY---RPHKTVLCDIC-DDEDC-GNGCFTET--------ITEESGFKSSGEIVLIKFYLVVIASVLICLV------------------ : 1456
ScsuiTEP/CD109-1 : -------GDSVVVIYFDKIG-E---QVCVTAMAERKIMIAD-VKPALVQVYDYYKPEKRGEAFY---NPPALSKCEIC-QNEEC-KQTCDR----------------------------------------------------------- : 1447
ScsuiTEP/CD109-2 : -------ADTVLVIYFDNIGNE---QVCVKIEGYHNIDVKD-LKPAMVQVYDYYEPEKRVEIFY---DLPTKVICETC-VTEEC-MLACSMNS--------------------------------------------------------- : 1423
ScsuiTEP/CD109-3 : -------MDTGIIIYFDKLDN-K--LVCPTIKAYRTFRVAK-QRQTAVYLYDYYDQGKAARYFY---QTPQASLCDIC-EGEECNSSKCRKE-------IEISKESNSNNGCS--KASMQIWTLLTKMILIYYYCH-------------- : 1437
ScsuiTEP/CD109-4 : -------GDSVIVIYFDKIG-E---KVCADAKAYRNNKVAD-LKPALIEVYDYYDLKKRGEKFY---TPPLVTVCDLC-ETDEC-KQKCKK----------------------------------------------------------- : 1430
NinoiTEP/CD109-1 : -------GGTKVQLYYDEINDI---KICPQVTAYRTFPVAN-VRPAAVSVYDYYENERRAESFY---DAPKSDLCDIC-SGNEC-AAKCKKASAKP------------------------------------------------------ : 1437
NinoiTEP/CD109-2 : -------GDSGVVIYFDKLTKGE--EVCPTVKAFRTFKVAK-QKPTAVVAYDYYDQGRSARSFY---QALPASLCDIC-DGDDCTRGNCPKS-------AEMSLDYTNSGDG----LRFNMFVASVSLLLLSVFR--------------- : 1457
NinoiTEP/CD109-3 : -------GNSLVNLYFNNIPQTP-SQVCIILSARQTNVVTG-LKSANVNIYRYYTPADQAAASYNIGTGGPVATTKSG------------------------------------------------------------------------ : 1427
EpspiTEP/CD109-1 : -------GGTKVQLYYDEMNDV---EVCPHITAYRTFPVAN-VKPAAVSVYDYYDNDQRAETFY---NAPSSDLCDIC-TGSEC-SSKCKKS---------------------------------------------------------- : 1430
EpspiTEP/CD109-2 : -------GDSGVVIYFDKLSKGE--EVCPTVKAFRTFKVAK-QKPTAVKAYDYYDQSRSARSFY---QALPATLCDIC-DDDECERGSCPKN-------AGMFEDYTGNSAASFTLTLINVFTFIIPVWLISNFS--------------- : 1465
VaspiTEP/CD109   : -------GKTGIVIYFDKLNNT---EVCPTVNAHRTYPVAD-QKPAPIVVYDYYEKTREARRFY---SALTADVCEVC-DGKDCEKYKCKGT-----PGYNERTGGASSLVLCWFLITAALLLHL------------------------- :  323
PsfuiTEP/CD109   : QRN------TQVIIYFKNIGKS---EVCPTILGFRTYGVAN-QRPALIKVYDYYDQSRVARMFY---QVGPDNICHICNGED-CPQDGCPEKEGYTLFGPSYDANVDPIQNYPSSAS--LIFSLK-------FNFLLAALLPAFIIIETA : 1530
CaspiTEP/CD109-1 : NTR--------VILYFKHIGKS---EVCPTILGFRTFRVAN-QRPALIKVYDYYDQSKVARMFY---QVGPDNICHICTGED-CPEDGCPEGESYDFFGPTYNANVDPIQDFSAAAQV---LPIYSLTFAISLFAMLNY----------- : 1539
CaspiTEP/CD109-2 : NTR--------VIIYFQNIRKS---EVCPTILGFRTYGVSN-QRPALIKVYDYYDQSRVARMFY---QVRPATICEICTGND-CPEDGCEDKSFYPIYNPTYHANVDPNNEAGATSKMSVSLFLMASSLCCTLVKMIYV----------- : 1429
HosaiTEP/CD109   : -------DHGKLNLYLDSVNETQ---FCVNIPAVRNFKVSN-TQDASVSIVDYYEPRRQAVRSYNSEVKLSS--CDLCSDVQGC-R-PCEDGASGSHHHSSVIFIFCFKLLYFMELWL-------------------------------- : 1445
                                                                                                            /     iTEP/CD109 GPI anchor signal sequence
```

Supplementary Fig. S9 (continued)

```
AmspC3           : KFIDVEIDEVIKPGVEGEAQQLRHGSVRTLISYNHCTCPDMFIGKKYFIMGQDGTKWEFE—GVAQYRYLLNTNARIFESKTIR--STSPKYKLMMKAFNWLQKRFIRYK---RGCE--------------------------------- : 1725
HaadC3-1         : LNISFTIADVIKAGIE--KAEDIVDETRTLLARDNCPTANLFPGVTYLILGKDGQNYMNE-NGETWYRYLLDKTSVIHMWTS----AKEAKNKNLQRDLNSVTKALKEN----GCEL-------------------------------- : 1738
HaadC3-2         : FNISFIVNEVIKAGIE--RKDIIESDIRTLLARDNCLAAALRYGTTYLILGKDGEGYMNE-HGHIWYRYVLDQTSIIQQWASYLGDLKDAGSKLLQRDLNQVTKILKED----GCDM-------------------------------- : 1695
ScsuC3-1         : KYFEFEVTDVLKEGVQ--QEREIQDETVTFVGKAACNCPKLTIGEVYMIMGEDGFEYKTD-TGEKKFKYLFSKYTRIYHSRSLR-MIQDQRGQKLQKTFNTIYMRFKR---GERCIH-------------------------------- : 1749
ScsuC3-2         : SFKLFHVNVFHVMKAGIQSNKEVEGETVIFHGRNECRYPEMTEGNSYIVMGTDGYPVTTE-DGEIKFKYMFDRHSRIYHSRSLR-DIADTKGRNLQKTFNSLYNRFVI--RKEGCDN-------------------------------- : 1755
ScsuC3-3         : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PsfuC3           : DVYSMKIETVIKEGTT--DVAPLGKVRSYLSY-KQCKEALGLQTGRSYLLMGTSKDIHKN---GAKYEYVIGEKTWVEYWPTPE-ECQTSAFREVCVGISEMEQQ----YLYFGCPQK------------------------------- : 1655
CaspC3           : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PaspC3           : DIYVVKIVEVIKEGSY--DVGPLNKKRAFLSY-PHCRESLDLKTGKTYLIMGTSRDIHRDE-QNNSYQYVLGERTWVEYWPTED-ECQLDEHRPTCFGMEELVQQ----YALFGCLQ-------------------------------- : 1604
HosaC3           : DEYIMAIEQTIKSGSD---EVQVGQQRTFISP-IKCREALKLEEKKHYLMWGLSSDFWGEK---PNLSYIIGKDTWVEHWPEED-ECQDEENQKQCQDLGAFTES----MVVFGCPN-------------------------------- : 1663
HosaC4           : ETKITQVLHFTKDVKAAANQMRNFLVR------ASCRLRLEPGKEYLIMGLDGATYDLEGHPQYLLDSNSWIEEMP------SERLCRSTRQRAACAQLNDFLQE----YGTQGCQV-------------------------------- : 1744
HosaC5           : ATLLDIYKTGEAVAEKDSEITFIKK-------VTCTNAELVKGRQYLIMGKEALQIKYNFSFRYIYPLDSLTWIEYWP---RDTTCS------SCQAFLANLDEFAEDIFLNGC----------------------------------- : 1676
AmspA2M-1        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
AmspA2M-2        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
AmspA2M-3        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
AmspA2M-4        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
HaadA2M-1        : EVLETSNSIDQREGNISQISTFVDVDHDLDFPDGLEG-----NMPVSVLPPPDFVQPDCPVCSDSFPSNFSAVYCNSAFALKVMKRENNMKTVKIIQDVSFYIDSPKAIKKFGELEYEEECTCTELAEDGKILFIVGSPLSLWNSNGKK : 1683
HaadA2M-2        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
ScsuA2M          : FKNFKSVEWELDFPDGIDGPPPHHHPTPNDLNNGENTT--------------------------------------------------------------------------------------------------------------- : 1625
NinoA2M          : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
EpspA2M          : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
VaspA2M-1        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
VaspA2M-2        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
VaspA2M-3        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PsfuA2M-1        : KIEVVKMPAVKKMSKKIIFVANNGKNQFYPIENELVMKNYLETFMKKSKRCIALSKLFKNVMYLLI----------------------------------------------------------------------------------- : 1536
PsfuA2M-2        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PsfuA2M-3        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
CaspA2M-1        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
CaspA2M-2        : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PaspA2M          : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
HosaA2M          : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
AmspiTEP/CD109   : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
HaadiTEP/CD109-1 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
HaadiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
ScsuiTEP/CD109-1 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
ScsuiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
ScsuiTEP/CD109-3 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
ScsuiTEP/CD109-4 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
NinoiTEP/CD109-1 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
NinoiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
NinoiTEP/CD109-3 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
EpspiTEP/CD109-1 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
EpspiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
VaspiTEP/CD109   : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
PsfuiTEP/CD109   : LSPRF------------------------------------------------------------------------------------------------------------------------------------------------ : 1535
CaspiTEP/CD109-1 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
CaspiTEP/CD109-2 : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
HosaiTEP/CD109   : ----------------------------------------------------------------------------------------------------------------------------------------------------- :    -
```

Supplementary Fig. S9 (continued)

```
AmspC3           : ---------------------------------- :    -
HaadC3-1         : ---------------------------------- :    -
HaadC3-2         : ---------------------------------- :    -
ScsuC3-1         : ---------------------------------- :    -
ScsuC3-2         : ---------------------------------- :    -
ScsuC3-3         : ---------------------------------- :    -
PsfuC3           : ---------------------------------- :    -
CaspC3           : ---------------------------------- :    -
PaspC3           : ---------------------------------- :    -
HosaC3           : ---------------------------------- :    -
HosaC4           : ---------------------------------- :    -
HosaC5           : ---------------------------------- :    -
AmspA2M-1        : ---------------------------------- :    -
AmspA2M-2        : ---------------------------------- :    -
AmspA2M-3        : ---------------------------------- :    -
AmspA2M-4        : ---------------------------------- :    -
HaadA2M-1        : HRIHLTSSVHVLLVPPKQIYSFITEAKSSCANDP : 1717
HaadA2M-2        : ---------------------------------- :    -
ScsuA2M          : ---------------------------------- :    -
NinoA2M          : ---------------------------------- :    -
EpspA2M          : ---------------------------------- :    -
VaspA2M-1        : ---------------------------------- :    -
VaspA2M-2        : ---------------------------------- :    -
VaspA2M-3        : ---------------------------------- :    -
PsfuA2M-1        : ---------------------------------- :    -
PsfuA2M-2        : ---------------------------------- :    -
PsfuA2M-3        : ---------------------------------- :    -
CaspA2M-1        : ---------------------------------- :    -
CaspA2M-2        : ---------------------------------- :    -
PaspA2M          : ---------------------------------- :    -
HosaA2M          : ---------------------------------- :    -
AmspiTEP/CD109   : ---------------------------------- :    -
HaadiTEP/CD109-1 : ---------------------------------- :    -
HaadiTEP/CD109-2 : ---------------------------------- :    -
ScsuiTEP/CD109-1 : ---------------------------------- :    -
ScsuiTEP/CD109-2 : ---------------------------------- :    -
ScsuiTEP/CD109-3 : ---------------------------------- :    -
ScsuiTEP/CD109-4 : ---------------------------------- :    -
NinoiTEP/CD109-1 : ---------------------------------- :    -
NinoiTEP/CD109-2 : ---------------------------------- :    -
NinoiTEP/CD109-3 : ---------------------------------- :    -
EpspiTEP/CD109-1 : ---------------------------------- :    -
EpspiTEP/CD109-2 : ---------------------------------- :    -
VaspiTEP/CD109   : ---------------------------------- :    -
PsfuiTEP/CD109   : ---------------------------------- :    -
CaspiTEP/CD109-1 : ---------------------------------- :    -
CaspiTEP/CD109-2 : ---------------------------------- :    -
HosaiTEP/CD109   : ---------------------------------- :    -
```

Supplementary Fig. S9 (continued)

AmspFB-1

```
   1  GAGCCGAACACGCTAGCACTTCCCTCTCCTCGATACCTTAGGTGATTACGATACCTTCTTGTTAAAGTGGCCATTACATAGCAAGTACTGAAGCTCAAAG  100
 101  AAAGTGTATGTCGAACCAGCTACATCATTTAGTATTCTCATCAGAATTAAAAAAAAATCGCAATTCAGAAATTGGTTTTGAAATGAAATCAATTTTTACG  200
 201  CGGTATAAACTCGTTCGATAAAAAATATTTACGCAGCTCGACGGAAATTTCCCAAATCGCGATGAGTTCCGAGAATTTGGTTCGTACTTCCTCCTCCCCT  300
 301  TACCACTCTTGAATAGGAAAAGGAAGAAAAAAATTGTTTATCCTATTCGATCGTAATAACCTCTTGAGCTTAATCTCGACTTATCTGGAAGCAAACATTA  400
 401  AAATTCGCATTAATGTTTTTGGAGCAGGTAATTTTGTAAGTATTTTCCGTGTTGTCTTCTCCTATTTATGACGCCTTTGTCACACATTTATAGAATTCAC  500
 501  AGCAAGTGTTTTCCTATTTGAGTCAGTTTCCAGGAAGTCAAAGCAAGGAACTGGCAACCTTTTTATTACGATCACTCTTGGATGTGAAGCGATGGACAGC  600
                                                                                                 M  D  S

 601  ATGTTCAGAATATGTCTGAAATTGTTTCTTTTATGTTTTGTCGTTAATTCCGCTTCTTCTGATTGCCATGGGAATTCGCATGCTTTGAAGCTCTATAATG  700
       M  F  R  I  C  L  K  L  F  L  L  C  F  V  V  N  S  A  S  S  D  C  H  G  N  S  H  A  L  K  L  Y  N
                                                                                   SCR1
 701  GAAAGTTTCGAGTCAACAAACTAGTTTCCTACGGAAAAAACAAGGTGGTTGTCGCCATGTGCAATAAAGGGTATGATCTGTTTGGTGATAGGATTCTAAT  800
      G  K  F  R  V  N  K  L  V  S  Y  G  K  N  K  V  V  V  A  M  C  N  K  G  Y  D  L  F  G  D  R  I  L  I

 801  TTGTGATGGAACTAATTGGATAGTTAGACGAGGCAAGACCAAACCTACCTGTCGTGAAAAGAAATGCCCCGAACCGCCTGCAATCGAGAACGGATTCGTT  900
        C  D  G  T  N  W  I  V  R  R  G  K  T  K  P  T  C  R  E  K  K  C  P  E  P  P  A  I  E  N  G  F  V
                                                                                    SCR2
 901  TACTACAGAAAAGGGCAAATAGCAAGACGTGGAGGCACAAGGGTTTTTTACAAATGCCTTGACGGTTTCCAGATCTCCAGCAGCAGACCATCGTTATTTT  1000
       Y  Y  R  K  G  Q  I  A  R  R  G  G  T  R  V  F  Y  K  C  L  D  G  F  Q  I  S  S  S  R  P  S  L  F

1001  GCCTGAAACGTAACAAACGGTGGCGCGGAACTCCCCCAAAATGCATATCTAAAAAGAAATGTCCGGAACTTGGCGATGTTGCCAACGGTCGGAGGACTCC  1100
      C  L  K  R  N  K  R  W  R  G  T  P  P  K  C  I  S  K  K  K  C  P  E  L  G  D  V  A  N  G  R  R  T  P
                                                                                SCR3
1101  AAGCAACTTCGAAGAAGGTTCTGTCGTGCAATACGAATGCAATCAGGACAGTACCATGATCGGTCCATCCGTCATCAAATGTCGATCCGACGCAACTTGG  1200
        S  N  F  E  E  G  S  V  V  Q  Y  E  C  N  Q  D  S  T  M  I  G  P  S  V  I  K  C  R  S  D  A  T  W

1201  AACGACACAATACCAGAATGCAAACCGCAAAATGCTTGCAAACTGCCGACTCAAGATGAAATACCTCATTCTCGGATGCTCTCGGATGTTCCTGGAGCTA  1300
       N  D  T  I  P  E  C  K  P  Q  N  A  C  K  L  P  T  Q  D  E  I  P  H  S  R  M  L  S  D  V  P  G  A
                                                      SCR4
1301  AGTACGTGCTCGCTGATGAATTTGTCACCATGGTATGCGAAAAAGGGTTTCGCATTGAAGGGATGAACTTTTTACAGTGTCTTCAAGGTGGAACTTGGGA  1400
      K  Y  V  L  A  D  E  F  V  T  M  V  C  E  K  G  F  R  I  E  G  M  N  F  L  Q  C  L  Q  G  G  T  W  D

1401  CAGCGACTTTGCTAAATGTTCTGTCGTCCACGGATGTCCTCCCGCTTATCCCATTGCAAATGGCGGCATTATTGAATCCCTCAAAGTTGGTGTGAACGTA  1500
        S  D  F  A  K  C  S  V  V  H  G  C  P  P  A  Y  P  I  A  N  G  G  I  I  E  S  L  K  V  G  V  N  V
                                                   SCR5
1501  AGCCATGGCACGACATTAAACTACTTCTGCAACGAAACGTACAGAATGGTTGGTTCCAGCTGGATCGAATGCGTTGATCATGAATTCGAAGGTTTAGCAT  1600
       S  H  G  T  T  L  N  Y  F  C  N  E  T  Y  R  M  V  G  S  S  W  I  E  C  V  D  H  E  F  E  G  L  A

1601  GGTCCAACAAGCCTCCAAAATGTTTCCCTATTAGATGTTCAGATCATGAAGCCCCTGATAACGGAGTGATATTGTCCAATGGGCCGTATTATGTTGGTGA  1700
      W  S  N  K  P  P  K  C  F  P  I  R  C  S  D  H  E  A  P  D  N  G  V  I  L  S  N  G  P  Y  Y  V  G  D
                                                       SCR6
1701  CGTTGTTTCATACTCCTGTTACGAAGGTTACAAACTTGTTGGTTCAAAGAATCGTACTTGCAAGGATACCGGAACTTGGAGCGGGGCTCTTGCAACTTGT  1800
        V  V  S  Y  S  C  Y  E  G  Y  K  L  V  G  S  K  N  R  T  C  K  D  T  G  T  W  S  G  A  L  A  T  C

1801  GATTCAGAAAATACCGTGTGTCCGAATCCCGGCGTACCGATAAACGGCATGAAAAGTGGTAACCGATATGATGCAGGCGATGTAGTTGTCTTTTCGTGCA  1900
       D  S  E  N  T  V  C  P  N  P  G  V  P  I  N  G  M  K  S  G  N  R  Y  D  A  G  D  V  V  V  F  S  C
                                        SCR7
1901  AACCAACCTACTTTCTTTTCGGCAATGCAACCAGAATTTGTCAGAGTAATGGCTATTGGTCTGGCACTGAAGTTTTCTGCCGCGGTACGAACGAGTTTGA  2000
      K  P  T  Y  F  L  F  G  N  A  T  R  I  C  Q  S  N  G  Y  W  S  G  T  E  V  F  C  R  G  T  N  E  F  D

2001  CGATATTACTCATATAAGCAACAAACTTCTGACGAGGTTCGACATGCTTAAACTCATCAGCAAAAAGTCCTATACAAATTCCGACGCCACCAACAACACG  2100
        D  I  T  H  I  S  N  K  L  L  T  R  F  D  M  L  K  L  I  S  K  K  S  Y  T  N  S  D  A  T  N  N  T

2101  CAAAGCAACTCGTCTGCTCAAGGCAGAACTTTGAATATCAACAATCCCGGTGGATTGAATTTGATATTTATTTTTGATTCGTCTGGAAGTGTTGGAAAAC  2200
       Q  S  N  S  S  A  Q  G  R  T  L  N  I  N  N  P  G  G  L  N  L  I  F  I  F  D  S  S  G  S  V  G  K
                        Factor D cleavage site                            vWFA   Mg²⁺ binding sites
2201  GTGGATTCGAAACGGCCAAGAAATTTGCCTCTACGCTCGTCCAACATATTGGTGTTGGGGCAAACGGAGTTCGTGTCGCGGCTATGACTTTCAGCTCCGA  2300
      R  G  F  E  T  A  K  K  F  A  S  T  L  V  Q  H  I  G  V  G  A  N  G  V  R  V  A  A  M  T  F  S  S  D

2301  CGTCACGGTCAATTTCTACACCAGGGAGTTCCTAACGACCGAGGAGGTGGTTGAACAGATTGGACTCATTCAGTATAATCCCGGTGACACGGCAACGAAT  2400
        V  T  V  N  F  Y  T  R  E  F  L  T  T  E  E  V  V  E  Q  I  G  L  I  Q  Y  N  P  G  D  T  A  T  N
                                                                                          Mg²⁺ binding site
2401  CCAGCATTGATTACGGCTACAACAGAAGTCATACCTGAAGCTGCACGAGCCAGACCGTTGAGCTCATCAGCCGTTTTCCTCATAACGGACGGCCGCGCTA  2500
       P  A  L  I  T  A  T  T  E  V  I  P  E  A  A  R  A  R  P  L  S  S  S  A  V  F  L  I  T  D  G  R  A
                                                                                      Mg²⁺ binding site
2501  ACGTTGGAGGACGACCAAAAGAAGCAGCAGATAGGCTAATTCAGGAATTTGATGTTGAGATTTACGCGATCGGAGTTGGAAGCAACATTCTTGAGGATGA  2600
      N  V  G  G  R  P  K  E  A  A  D  R  L  I  Q  E  F  D  V  E  I  Y  A  I  G  V  G  S  N  I  L  E  D  E

2601  GTTGGCCAGCATTGCCTATTCGAAAGATGGTGATACCGACCGGCATTATATGAAGGTTGAAAGTTTTGCCAAGATGAACGAGATGCTCCAACTTCTCATC  2700
        L  A  S  I  A  Y  S  K  D  G  D  T  D  R  H  Y  M  K  V  E  S  F  A  K  M  N  E  M  L  Q  L  L  I

2701  AATGGAACCATTGACTATTCTGCTTGTGGTTTGCTTCAAAGGCGGAAAAATACTGAAAAGAAAGAACTTGGCGTAGAAAAAGCTGAACCAGGAGCTAGAA  2800
       N  G  T  I  D  Y  S  A  C  G  L  L  Q  R  R  K  N  T  E  K  K  E  L  G  V  E  K  A  E  P  G  A  R

2801  ACAAACGAAGTCTCCGGATAAGAATCGTTGGAGGCGAAACAGTAACCAAGCATTGGCCATGGATGGTGGGACTTTACTATGGATCACCACTACTTGGCCA  2900
      N  K  R  S  L  R  I  R  I  V  G  G  E  T  V  T  K  H  W  P  W  M  V  G  L  Y  Y  G  S  P  L  L  G  H
                                                   SP
2901  CAAGAGTTTGCAATGTGGCGGTTCGCTGATTGCTCCCAATTGGATTCTAACAGCCGCTCATTGTATCAAACTTCAAATAAGTGATGATAAAATCGTCACT  3000
        K  S  L  Q  C  G  G  S  L  I  A  P  N  W  I  L  T  A  A  H  C  I  K  L  Q  I  S  D  D  K  I  V  T
                                                     Protease activate site
3001  TACACAACCAGAAACGTACGCATTCATCTAGGAATCACGAATATCAAAAATCCGGACAATAATAACTTGATCATTGCAACACCCCAAGAATTTGTACTGC  3100
       Y  T  T  R  N  V  R  I  H  L  G  I  T  N  I  K  N  P  D  N  N  N  L  I  I  A  T  P  Q  E  F  V  L

3101  ATCCACAATATGATCCCGCGACAATAAATAACGACGTTGCATTAATCAGGCTTCAAGAACCAGTCACATACAACCCATTCATCAGACCCGTCTGCCTACC  3200
      H  P  Q  Y  D  P  A  T  I  N  N  D  V  A  L  I  R  L  Q  E  P  V  T  Y  N  P  F  I  R  P  V  C  L  P
                            Protease activate site
```

Supplementary Fig. S10

```
3201  CCCAGCACTGGATAAACTGCCGAAAAACAGCTTGCTATATAAAACTGGAGAAACAGCTATTGTGACTGGTTGGGGTCACACTAAAGAGCGAAAAAAACAC  3300
        P  A  L  D  K  L  P  K  N  S  L  L  Y  K  T  G  E  T  A  I  V  T  G  W  G  H  T  K  E  R  K  K  H

3301  GAATCACTCGGCCCGGATGACCTCAAGCAATCTGCAATTCTGAAGAAGTTGTCTGTGCCGATACAAGAAGGAAAAGTTTGCAATGATAGTATTCAAGGAG  3400
       E  S  L  G  P  D  D  L  K  Q  S  A  I  L  K  K  L  S  V  P  I  Q  E  G  K  V  C  N  D  S  I  Q  G

3401  AATTTCGCAAGGGAATTTACACTGACAGCATGCTGTGTGCTGGAATGGGAAAGAAAGGACAAGATGCATGTGCGACAGATAGCGGCGGTCCTCTGCATCA  3500
      E  F  R  K  G  I  Y  T  D  S  M  L  C  A  G  M  G  K  K  G  Q  D  A  C  A  T  D  S  G  G  P  L  H  Q
                                                                         Protease activate site
3501  GCGTCTGGAGGATTCAGAAGGTGATGGCACTTACTACACCCAAATCGGCATCGTTAGTTGGGGTTATGGATGTGCTTTAGAGGGCGAATATGGATTCTAT  3600
        R  L  E  D  S  E  G  D  G  T  Y  Y  T  Q  I  G  I  V  S  W  G  Y  G  C  A  L  E  G  E  Y  G  F  Y

3601  ACAAGACTCACGAAGTTCATTCCTTGGATCGAGGGCGTAACCAACGTTAAATTTTCCAATTGATGATGATTGAAATCAATGATGGACAATTTATTTTTAT  3700
       T  R  L  T  K  F  I  P  W  I  E  G  V  T  N  V  K  F  S  N  *

3701  TTTTTTTTTTGTAATTTTATGTTTTATGCATCGTTTTGTCAATAAAATTTGAAATATTGGAAAATCCAATGGAAAAAAA  3779
```

AmspFB-2

```
   1  CTTAGAGCGCCTTCCGTCATGCGACCTTACGGAAATATTTCCTTGTTGAATTCATTTTTCATTTGTTTGCTTCAATTTCTACATTCTACAGAAGGAAAAT  100
       L  R  A  P  S  V  M  R  P  Y  G  N  I  S  L  L  N  S  F  F  I  C  L  L  Q  F  L  H  S  T  E  G  K

 101  GCCCACCTCTTCTGGGCTCAAGAGTTGAGAATGGACATATCAAGGCTTGGTCTTCGAGAAAACTGCAAGTTATTTGCGATCGAAGCTACATCGTGCATGG  200
      C  P  P  L  L  G  S  R  V  E  N  G  H  I  K  A  W  S  S  R  K  L  Q  V  I  C  D  R  S  Y  I  V  H  G
                          SCR1
 201  CGCTTCGACTTTGTTTTGCAATGGCAGAAGTTGGACCGGGAAAGGCAAACTTCCTCAATGCTCTTTGCGAAGCAGGTTTTGCCAGTCAGCTGGAAACTTG  300
        A  S  T  L  F  C  N  G  R  S  W  T  G  K  G  K  L  P  Q  C  S  L  R  S  R  F  C  Q  S  A  G  N  L
                                                                                                SCR2
 301  AAAAATGGGCGAAAGATTGGAACGATGCAGCATATCGGTGCCGTCGTCCGCTATTCGTGCAATCCTGGCTACACGTTGGTTGGATCGACGTTGCAAGAAT  400
       K  N  G  R  K  I  G  T  M  Q  H  I  G  A  V  V  R  Y  S  C  N  P  G  Y  T  L  V  G  S  T  L  Q  E

 401  GCAAGGCAAGCAGACGCTGGATACCTGCCAAACCACAATGCAAAAGTCCTTCTGAATTGCTACAGGAAACATCAAAGAGGTTCAATAAACAATTTGTCCA  500
      C  K  A  S  R  R  W  I  P  A  K  P  Q  C  K  S  P  S  E  L  L  Q  E  T  S  K  R  F  N  K  Q  F  V  Q

 501  AAGGTTAGCTGTGGACAACAGCACTTTTTCAAGTGGAAGTCGAGTTCTTGGACCAGAAGTTAACAAGATTGGACTGGACCTGGCCTTTCTCATCGACAGA  600
        R  L  A  V  D  N  S  T  F  S  S  G  S  R  V  L  G  P  E  V  N  K  I  G  L  D  L  A  F  L  I  D  R
                                         Factor D cleavage site                  vWFA      Mg²⁺ binding site
 601  TCCAGCAGCATCGATCCTGTCGATTTCAAGATTGGTATCAACTTCCTCAAGGAGCTGGTAGATGAATTTGGAGTGAAAAATGGGGATAACAAATTGAAAG  700
       S  S  S  I  D  P  V  D  F  K  I  G  I  N  F  L  K  E  L  V  D  E  F  G  V  K  N  G  D  N  K  L  K
      Mg²⁺ binding sites
 701  GTGGTACCAGAATAGCGGTGATAACATTCGGGGACAAAGCTGAGATTGTCTTCAACTTTGACGATGCAAGAATCTCGTCTCCGGAAGTAGCCAAACGAAA  800
      G  G  T  R  I  A  V  I  T  F  G  D  K  A  E  I  V  F  N  F  D  D  A  R  I  S  S  P  E  V  A  K  R  K

 801  ACTCGACATGATCCAGGCCAAGGGCGGAAGTACAAATTTGAACGGTGCTCTGGACAAAGTGATCATTCGAGTTAAACCCAGAAGAGAAGTGAAAAGGGCT  900
        L  D  M  I  Q  A  K  G  G  S  T  N  L  N  G  A  L  D  K  V  I  I  R  V  K  P  R  R  E  V  K  R  A
                               Mg²⁺ binding site
 901  TTGTTCATCATGTCTGATGGCAAGCCAAATACTGCAACCAGAAGTACCCCTGAAAACGCGGCACGTCATCTCAAGACAAGATCACTCAATTACGAAATCT  1000
       L  F  I  M  S  D  G  K  P  N  T  A  T  R  S  T  P  E  N  A  A  R  H  L  K  T  R  S  L  N  Y  E  I
                Mg²⁺ binding site
1001  TCACCGTTGGAATCGGACGAGCAGTAAACATGACTCTCTTGCGCCACATGGCCTCGGATCCAAAGATCAACCACAACTTCTACTTCGACAAGTTCTCTGA  1100
      F  T  V  G  I  G  R  A  V  N  M  T  L  L  R  H  M  A  S  D  P  K  I  N  H  N  F  Y  F  D  K  F  S  D

1101  CTTCGAGCGGATCTTGTGGCTCATCAAGAACAAAGCAACTCCAGCACCACCGCCGGGTTTCGACAAATGTGGGTACATACCTCCAGGAAGTTATAGATGG  1200
        F  E  R  I  L  W  L  I  K  N  K  A  T  P  A  P  P  P  G  F  D  K  C  G  Y  I  P  P  G  S  Y  R  W
                                                                                                SP
1201  CCATGGTTGGTTTCTGTTTGGAGTGAAATTCCCGTTGAAGATAGATTCACATTGAAGCAATGCACCGGGACGATAATCTGCCGGAAATGGATTCTTACTT  1300
       P  W  L  V  S  V  W  S  E  I  P  V  E  D  R  F  T  L  K  Q  C  T  G  T  I  I  C  R  K  W  I  L  T

1301  CTGCGCATTGCTTGCATTATTCTGATCGAAGAGACTACAAACCAGTATCCAAGGTTTTCGTGACCGTGGGTCAAAAAGATATTCTCAAGTGGAGTGACAG  1400
      S  A  H  C  L  H  Y  S  D  R  R  D  Y  K  P  V  S  K  V  F  V  T  V  G  Q  K  D  I  L  K  W  S  D  S
     Protease activate site
1401  CAGTGAGAATTTCATTGCTTCCAGAATTATAATACACGAACGTTACAACAACGAGACCTTGGAGAATGACATTGCCCTGATTGAACTCAATGATGTGACT  1500
        S  E  N  F  I  A  S  R  I  I  I  H  E  R  Y  N  N  E  T  L  E  N  D  I  A  L  I  E  L  N  D  V  T
                                                            Protease activate site
1501  GCTCTTGGCGATAAAGTCAAGATAGCCTGTCTTCCACGACCAGAAATTCAATTGTCAAGACGATCATCGCAGAACAGTTTTACTGCTGGATGGGGCAGCA  1600
       A  L  G  D  K  V  K  I  A  C  L  P  R  P  E  I  Q  L  S  R  R  S  S  Q  N  S  F  T  A  G  W  G  S

1601  CGAACGACAAAAATTACGACAGTGCTCCAGTGAAGAGTGCATTCTTCTCCCTCAATGAAGTTGCCCTCACATTGGTGCCGGACAGAGACTGCGCAAGAAG  1700
      T  N  D  K  N  Y  D  S  A  P  V  K  S  A  F  F  S  L  N  E  V  A  L  T  L  V  P  D  R  D  C  A  R  R

1701  AGCTAAATCACTGCAATGCCGTCAGCCTTCAGCTTCCCATTTCTGTGCTGAAACAAGCAATGACTTGATCAAAGCCTGTTCAACCGGTTCCGGAAGTCCA  1800
        A  K  S  L  Q  C  R  Q  P  S  A  S  H  F  C  A  E  T  S  N  D  L  I  K  A  C  S  T  G  S  G  S  P
                                                                                Protease activate site
1801  GTTATCATGTCGGACAGCAGCAGTTACGACAACTACGAGCGTTTGCGTGTCGTTGGCATTTCATCGTCCCATTGCCAGTGTGGCAAACAACGCGGTTATC  1900
       V  I  M  S  D  S  S  S  Y  D  N  Y  E  R  L  R  V  V  G  I  S  S  S  H  C  Q  C  G  K  Q  R  G  Y

1901  ACATGTTCACCAGAGTTCACGATTACATCAAGTGGATTGCATCCAGAACAGATTTCTGCATAGCTGAGCATATCTAAATAGAAAAGGGTGTATAGTGAGT  2000
      H  M  F  T  R  V  H  D  Y  I  K  W  I  A  S  R  T  D  F  C  I  A  E  H  I  *

2001  GTTGTGTAATAAAGCAAATTGTAATCAAA  2029
```

Supplementary Fig. S10 (continued)

HaadFB-1

```
    1  CTCTGTGTTTATGTAAAATCCTTGTAGTTTCTATGAAGGATAAACTCAATGATATTTTGATCAGATGTATATCAATTTTGCTATCTTTTACCACTCGACC  100
  101  TATTGACCGGTAAATCCCCGAACTTCCGCTGCTGAATTCTGGAGAAGGAACCAACCAACTGATGCTTCTCAGCATGACATCACTTGATTATGAGGAAATT  200
  201  TGATAAAACAAAATTCGCCAAAGATCATTACATCAGCGACTGCTGCTTGTTTGGAAGAGAAACTAGATTATAAACGGTTTTTGCATATGACATAATGCAG  300
                                                                                                             M  Q

  301  AAAAATCCAACATTATGCAGCTGTTTTAAAGTAATCATAAGTGTGATTCTTTTTAACTTCATTACATGTGATTGTCCATATCCCAAAAGGGAACTTCCTA  400
        K  N  P  T  L  C  S  C  F  K  V  I  I  S  V  I  L  F  N  F  I  T  C  D  C  P  Y  P  K  R  E  L  P
                                                                                            SCR1
  401  ATGGTCGTTTTGTGTATCAGCAGTTAAAAGATATAAATGAAACTTCTGCTAAAGAGCTGTATATCTTACGTTTTAAATGCTACGCTGGATACCAAAGGAT  500
       N  G  R  F  V  Y  Q  Q  L  K  D  I  N  E  T  S  A  K  E  L  Y  I  L  R  F  K  C  Y  A  G  Y  Q  R  I

  501  AGGACCCAGATACATACAGTGTTTAGATGATGAATGGTCGGATGTAGTACCAACATGTGCAAAAATGCAATGCTCAGATCCTCCTGATGTTGACAAGGCT  600
         G  P  R  Y  I  Q  C  L  D  D  E  W  S  D  V  V  P  T  C  A  K  M  Q  C  S  D  P  P  D  V  D  K  A
                                                                                           SCR2
  601  GAGTTTAAACTCCATTCTGGTGTAGATAAATTCCCAACTATAGGCGCATCTGTTACATATTCATGTAAAATAGGTTATGAATTGAAGAACACAAGTGCTT  700
        E  F  K  L  H  S  G  V  D  K  F  P  T  I  G  A  S  V  T  Y  S  C  K  I  G  Y  E  L  K  N  T  S  A

  701  CAACATTACATTGTGAATTGGACCCTGGGGCAAGCAAGGCAATCTGGAAAGGAGAAATCCCTTTTTGCAAAGAAAAAGAATCTTGTCCGGATCCCGGTGT  800
       S  T  L  H  C  E  L  D  P  G  A  S  K  A  I  W  K  G  E  I  P  F  C  K  E  K  E  S  C  P  D  P  G  V
                                                                                                    SCR3
  801  CTCAAAACATGGTTTTCGGGTAGGAAGCTGCTGTTTTTCAGGAGACATCTTACAATTTTCTTGCAACGAAGATTATGAATTAGTTGGCCAAAAAGAAGTA  900
         S  K  H  G  F  R  V  G  S  C  C  F  S  G  D  I  L  Q  F  S  C  N  E  D  Y  E  L  V  G  Q  K  E  V

  901  CAATGTCTTAGATCCGGTAGTTGGTCATCAGGGCGTCCACTTTGTAAGCCTTTAAGCGATTACTGTCAATTACCTCCCTCAATACCACATGGGATAGTTA  1000
        Q  C  L  R  S  G  S  W  S  S  G  R  P  L  C  K  P  L  S  D  Y  C  Q  L  P  P  S  I  P  H  G  I  V
                                                                                   SCR4
 1001  GTGGTGAAAAAGAAGGAGACTATTTTATACCGTATGATGAGGCGGAAGTCGTTTGTGAGCCCGGATATAAACATGTAGGTCCAACCCAGTTTGTAATGTG  1100
       S  G  E  K  E  G  D  Y  F  I  P  Y  D  E  A  E  V  V  C  E  P  G  Y  K  H  V  G  P  T  Q  F  V  M  C

 1101  TGAAGAAGAAGGAGAATGGGAAGATGAATTTGGAGAATGTACAGAAATAATTTGTGATTTACCTAAACAACTTGTAAACGGAACGATACCCGAGATGCTG  1200
         E  E  E  G  E  W  E  D  E  F  G  E  C  T  E  I  I  C  D  L  P  K  Q  L  V  N  G  T  I  P  E  M  L
                                                                          SCR5
 1201  ACAACGAATCTTACCAGCTTTCCTTACGGCTTCGAGATTACATACTTTTGTGACGATGGATTTCGACTTGTTGGAGGAGACAGTTGGAGAACATGCAGCA  1300
        T  T  N  L  T  S  F  P  Y  G  F  E  I  T  Y  F  C  D  D  G  F  R  L  V  G  G  D  S  W  R  T  C  S

 1301  AAAAAGGCTGGTCTGGAAAAACACCATATTGTGAAGCTATACAATGCCCTGATCCTGGTTTGCCCGAAAATGGATTCAGAACAGGAGACAATTTTGAAGT  1400
       K  K  G  W  S  G  K  T  P  Y  C  E  A  I  Q  C  P  D  P  G  L  P  E  N  G  F  R  T  G  D  N  F  E  V
                                                                    SCR6
 1401  TGGTGCTAAAGTTCGTTTTAAATGTTTTACTGGGTATTATCTCTTGGGATCCTTTGAGCGATACTGCAAACCCAACGGACAATGGAGTGGAGAGCTCTCA  1500
         G  A  K  V  R  F  K  C  F  T  G  Y  Y  L  L  G  S  F  E  R  Y  C  K  P  N  G  Q  W  S  G  E  L  S

 1501  CGCTGTGACACACCAAGTAACTATTGTCCAAATCCCGGAATTCCTGTCAAAGGTTATAAGAACATCTCAAGTTACGAAATGGGAGACAAAGTGGGTTTCC  1600
        R  C  D  T  P  S  N  Y  C  P  N  P  G  I  P  V  K  G  Y  K  N  I  S  S  Y  E  M  G  D  K  V  G  F
                                              SCR7
 1601  ATTGTCAACCAGGATATGTTCAGATAGGATCAGAAGTTAGAGAATGTTTGCCTAACAGAACTTGGTCAGGAACTGAAACCACATGTTTGGGTCCCTATGA  1700
       H  C  Q  P  G  Y  V  Q  I  G  S  E  V  R  E  C  L  P  N  R  T  W  S  G  T  E  T  T  C  L  G  P  Y  D

 1701  CTATGACAACAGTGCCCAAGTAAGAGATGTTTTGAGAGCTAAACTTGCAGAGAAAGCTGAGGAACAAGAGAGAGAAACCCAAAGGTATAGAGAAGCTCTA  1800
         Y  D  N  S  A  Q  V  R  D  V  L  R  A  K  L  A  E  K  A  E  E  Q  E  R  E  T  Q  R  Y  R  E  A  L

 1801  TATGCATCTTGGCATAATAATTCGGGTCCAGTTGGACGAATACTGGACATCAATTTCCCTGGAAGACTCATTTTGTATTTTGCCTTTGATGTCTCAGGTA  1900
        Y  A  S  W  H  N  N  S  G  P  V  G  R  I  L  D  I  N  F  P  G  R  L  I  L  Y  F  A  F  D  V  S  G
                                   Factor D cleavage site                        vWFA        Mg²⁺ binding sites
 1901  GTGTAGGACAACATAATTTCGATAAAAGTATCGAATTTGCCAAAGCCATCGTGAAGAGAGTTGGTATTTCAGAAGCTGGAGCGCGTGCTGGTGCTCTTAT  2000
       S  V  G  Q  H  N  F  D  K  S  I  E  F  A  K  A  I  V  K  R  V  G  I  S  E  A  G  A  R  A  G  A  L  I

 2001  CTTCGGTTCAAAATCTGAAAATATGTTTTTACCGCTATCTTATACTACTACAGAAGAAGTACTTGATGCTCTTGATAAAATAAATTATACAGGTGGAGGA  2100
         F  G  S  K  S  E  N  M  F  L  P  L  S  Y  T  T  T  E  E  V  L  D  A  L  D  K  I  N  Y  T  G  G  G

 2101  ACAGCTGCCAGTTCTGCTTTGAGCTTAATCAGACAAGAAAATATTCCTTTAATCGATGGTGTTCTTGGAAAAAAGAATATAAAATCCATAATTTTCATAC  2200
        T  A  A  S  S  A  L  S  L  I  R  Q  E  N  I  P  L  I  D  G  V  L  G  K  K  N  I  K  S  I  I  F  I
       Mg²⁺ binding site
 2201  TTACTGATGGTAAGGCGAATATGGGAGGAAGTCCAGAAGTAGAGGCCGATCTTCTGAAAAAAGCAGGAGTAGAAATATACTGTATTGGAATCACAGGAAG  2300
       L  T  D  G  K  A  N  M  G  G  S  P  E  V  E  A  D  L  L  K  K  A  G  V  E  I  Y  C  I  G  I  T  G  S
        Mg²⁺ binding site
 2301  TATTGAAAAGGAATCCTTGTATAAAATAGCTTCTACATCTAAAGACTCAAATGGGGAACATCCAAACGTTTTTATTCTTCAGAATTACGCAACGATGTCT  2400
         I  E  K  E  S  L  Y  K  I  A  S  T  S  K  D  S  N  G  E  H  P  N  V  F  I  L  Q  N  Y  A  T  M  S

 2401  TGGCTTGTTCAGGAGATTACAAATGGGACAGTGGATTATGCTCAGTGTGGATTAGGAATGGAAAATGTTGGAAACGAAGCTGCCCGAGGTAGAATTTTAA  2500
        W  L  V  Q  E  I  T  N  G  T  V  D  Y  A  Q  C  G  L  G  M  E  N  V  G  N  E  A  A  R  G  R  I  L
                                                                                                        SP
 2501  ATGGAAAAAAAAAGTATGGAGCCATGGCCATGGATGGCAGCTTTATACATGCCTCATGACAAATTAAATCCCCTTGATACAGAACTCCAGTGTGGCGGTTC  2600
       N  G  K  K  S  M  E  P  W  P  W  M  A  A  L  Y  M  P  H  D  K  L  N  P  L  D  T  E  L  Q  C  G  G  S

 2601  CATAATAAATAATTATTTCATTCTAACAGCTGCTCACTGCATGTACCACCGTGAAGGGAAGAAAAGAAGCAAAAAAGACATCATCGTCAAGTTAGGTTTG  2700
         I  I  N  N  Y  F  I  L  T  A  A  H  C  M  Y  H  R  E  G  K  K  R  S  K  K  D  I  I  V  K  L  G  L
                             Protease activate site
 2701  ACAGATGTAAAGAATGAAACATACGTACAAGAATCTGAAGTGAGCGAAATGTTCATTCATCCTGATTATCGGCCTGCAGGTTCATATGATTATGATATTG  2800
        T  D  V  K  N  E  T  Y  V  Q  E  S  E  V  S  E  M  F  I  H  P  D  Y  R  P  A  G  S  Y  D  Y  D  I
                                                                                            Protease activate site
 2801  CTCTTCTTCTGTTGGACAAGCCTATCGAGTACAATCCATTTGTCAGACCAATTTGCCTTCCTCCTACAGAGCTACCAGAGAATACTCCTTTGTACAGTTC  2900
       A  L  L  L  L  D  K  P  I  E  Y  N  P  F  V  R  P  I  C  L  P  P  T  E  L  P  E  N  T  P  L  Y  S  S

 2901  AGATGAATTTGGCTGGGCAACTGGATGGGGACACGAAGGCGTGGTGTCTGCAGCCGTAAACGAACGTCTGAAGTCAAGCCAAATTCTTAAAGAACTTCTG  3000
         D  E  F  G  W  A  T  G  W  G  H  E  G  V  V  S  A  A  V  N  E  R  L  K  S  S  Q  I  L  K  E  L  L
```

Supplementary Fig. S11

```
3001  CTTCCTATTCAGTCCAAGCAAAGATGCACCCAAAGCCTATTGGACAATAAAGTTGCTACTGATCATTTTACAGACCGTATGTTTTGTGCTGGAGACGGAA  3100
       L  P  I  Q  S  K  Q  R  C  T  Q  S  L  L  D  N  K  V  A  T  D  H  F  T  D  R  M  F  C  A  G  D  G

3101  AAAGGGGAAACGATACCTGTAAAGGTGACAGTGGTGGTCCTTTAATGCAGTCACAGCTGAATTCTGAAGGTTACCTCTTTTGGACTCAAGTAGGCATCGT  3200
      K  R  G  N  D  T  C  K  G  D  S  G  G  P  L  M  Q  S  Q  L  N  S  E  G  Y  L  F  W  T  Q  V  G  I  V
                        Protease activate site
3201  AAGCTGGGGTATTGGATGTGGCAAAGAAAACACTTACGGATACTATACGCATGTGCAAAAGTTTCGATCATGGATAGATTCAACTATTGAAGCAGCGATG  3300
        S  W  G  I  G  C  G  K  E  N  T  Y  G  Y  Y  T  H  V  Q  K  F  R  S  W  I  D  S  T  I  E  A  A  M

3301  GCAGCAATACAATAAAACAATAAAGTCTGTTTAATGAAAAACTCAGATATCAAATCACTATTTTTAATATTGGAATTCCTTGATTCATGAGATATTGCTC  3400
       A  A  I  Q  *

3401  GATGCATAATACCATACATCTAAGTGAATAATTCATGAAAATAACTTTAATGAATTCAACAATATATGAACATAATTTATTTTTTTAAAGTAGATTAGGA  3500
```

HaadFB-2

```
   1  CACGAGATGTTTCTTGATTTGGAGATTAAGGGAGAGAGTGCAAAGTTTCCTTCATCAGTGCTAAAGCAAAAATGAAAATGAAATTCATGTAAAGATGCTT  100
 101  AAAATGAAATAATAATAATGAAACGTAGAAAATTATGTTTGTGGCTGCAGTTCTTTCTGGGACATTTTATGCTGGAATTTACATATTGCGCTCCAAACAT  200
                        M  K  R  R  K  L  C  L  W  L  Q  F  F  L  G  H  F  M  L  E  F  T  Y  C  A  P  N  I

 201  CTACTGTGAATCACCACCTTTCATTCCACACGGTACAGTGTCAGGAATTGAGAGTACAGGCAATTACTTACCGTATGATGAGGTTGAAATCCAGTGTGAT  300
        Y  C  E  S  P  P  F  I  P  H  G  T  V  S  G  I  E  S  T  G  N  Y  L  P  Y  D  E  V  E  I  Q  C  D
                        SCR1
 301  CCTGGTTACAAGTATAGTGGAACAACCGAATTTATTTTATGTGAAGAGGACGGAAAATGGGAAGAAGATTTTGGAGCATGTGTAGAGATATTTTGTGATC  400
       P  G  Y  K  Y  S  G  T  T  E  F  I  L  C  E  E  D  G  K  W  E  E  D  F  G  A  C  V  E  I  F  C  D

 401  CACCGGAACCGCCTGCTAATGGAACTATACCTGAATTGCTGAATTCTAATCTCACATTATTTCCATTCAATTTTGGCATCACATACGAGTGCATACAAGG  500
      P  P  E  P  P  A  N  G  T  I  P  E  L  L  N  S  N  L  T  L  F  P  F  N  F  G  I  T  Y  E  C  I  Q  G
          SCR2
 501  CTATCGATTATTCGGAGGTGACAGCTGGAGGTTTTGTAATAAAAATGGTTGGTCAGGGAAAACACCGCACTGTAAAGAAATTCGATGCCCCGATCCAGGT  600
        Y  R  L  F  G  G  D  S  W  R  F  C  N  K  N  G  W  S  G  K  T  P  H  C  K  E  I  R  C  P  D  P  G
                                                                                                 SCR3
 601  GTACCTGAACATGGCGATCGAGAAGGAAACAGTTTTCAAGTTGGATCCAAAGTTCGCTTCAGGTGCTTTACTGGGTATAATTTGTTGGGATCTTTTGAAA  700
       V  P  E  H  G  D  R  E  G  N  S  F  Q  V  G  S  K  V  R  F  R  C  F  T  G  Y  N  L  L  G  S  F  E

 701  GATTCTGCATGCCGAATGGACAATGGACTGGCGAACTTGCACGTTGTGATGATCCAGGCAATTATTGTCCGAATCCAGGAACGCCCATAGATGGGTTTAA  800
      R  F  C  M  P  N  G  Q  W  T  G  E  L  A  R  C  D  D  P  G  N  Y  C  P  N  P  G  T  P  I  D  G  F  K
                                                                                      SCR4
 801  GAATGGAGCGAGTTACGATATTGGTGACAAAGTGAGCTACCGTTGTCAGCCAGGATACATTTTATTAGGATCAGAGGTTAGAGAATGTTTATCGAACAAA  900
        N  G  A  S  Y  D  I  G  D  K  V  S  Y  R  C  Q  P  G  Y  I  L  L  G  S  E  V  R  E  C  L  S  N  K

 901  ACTTGGTCAGGGACTGAAGCAAAATGTATGGCCCCTTATGACTTTGATAATTATGCGCAAATGACTGAAATGCTGAAGTTCAAAGTGACTGAAAAATTTG  1000
       T  W  S  G  T  E  A  K  C  M  A  P  Y  D  F  D  N  Y  A  Q  M  T  E  M  L  K  F  K  V  T  E  K  F

1001  ATGATGGAAACAGAGAACAGAAACAGTCTCATGAAGCAGAGATTTCGACTTTTGCAACTCTAATATTGGGTGATAAAATTGAGCCAGATTATCAGATGAT  1100
      D  D  G  N  R  E  Q  K  Q  S  H  E  A  E  I  S  T  F  A  T  L  I  L  G  D  K  I  E  P  D  Y  Q  M  M

1101  GCAACATGGTTTGATGAACAAGCATGCTTTTCAGAAACCTTCAGAGGACATTGATTCTATATCTTCATTTAGAACGATAGATGTCAACAACCAACGGAGA  1200
        Q  H  G  L  M  N  K  H  A  F  Q  K  P  S  E  D  I  D  S  I  S  S  F  R  T  I  D  V  N  N  Q  R  R
                                                                     Factor D cleavage site
1201  TTAATTCTCCATTTCGCTTTTGATGTTTCTGGGAGTGTTGGATCGTATTATTTACGGAAAAGCATTGAATTTGCTAAAGCAATTGTAAGGAAGATTGGAA  1300
       L  I  L  H  F  A  F  D  V  S  G  S  V  G  S  Y  Y  L  R  K  S  I  E  F  A  K  A  I  V  R  K  I  G
                  vWFA    Mg2+ binding sites
1301  TTTCAAGAGATGGATCCCGTGCTTCAGCTGTAGTTTTCAGTTCAAAAGCTGAAACTGTATTCTATCCCATGCAAATTGAAACTGAAGAAGAAGTATTAAA  1400
      I  S  R  D  G  S  R  A  S  A  V  V  F  S  S  K  A  E  T  V  F  Y  P  M  Q  I  E  T  E  E  E  V  L  K

1401  GTATTTAGACAATATAACATTTACCGGTGGTGGAACATCAGCCTCCTCTGCTTTTGCTCTGATAAAAGAAAATATTGAATTAGAAGAAGAATCATATTAT  1500
        Y  L  D  N  I  T  F  T  G  G  G  T  S  A  S  S  A  F  A  L  I  K  E  N  I  E  L  E  E  E  S  Y  Y
                                    Mg2+ binding site
1501  TTTGATAGAAATAATAAAGAAGATGTTGTGTTTTTTATTACAGATGGAAAAATAAACATGGGTGGAAATCCAGAAATGGAAGCTAAAATCTTAAAAGAAC  1600
       F  D  R  N  N  K  E  D  V  V  F  F  I  T  D  G  K  I  N  M  G  G  N  P  E  M  E  A  K  I  L  K  E
                                            Mg2+ binding site
1601  GAAATATAGATATATATTGCATTGGAATCACAGGTGACCCTAGGCTAGAAACTTTGTATAAAATTGCATCAACATCAAAATATGGAAATGTGGAACGTTC  1700
      R  N  I  D  I  Y  C  I  G  I  T  G  D  P  R  L  E  T  L  Y  K  I  A  S  T  S  K  Y  G  N  V  E  R  S

1701  GAACGTTTTTATTCTTCAAAACTATGCAACACTATCTCATCTCATTCAAGAGATAACAAATGGAACACTTGATTTCTCTGAATGTGGACTTGGTTTAGAG  1800
        N  V  F  I  L  Q  N  Y  A  T  L  S  H  L  I  Q  E  I  T  N  G  T  L  D  F  S  E  C  G  L  G  L  E

1801  AATATAGATGTTGGAAATCAAGAAGTAACTCAAAATGAAGAAAACAGAAATGATCCATGGCCGTGGATGGCAGCTCTATTTTTCCTAATATCCAAAGATA  1900
       N  I  D  V  G  N  Q  E  V  T  Q  N  E  E  N  R  N  D  P  W  P  W  M  A  A  L  F  F  L  I  S  K  D
                                                 SP
1901  CTTATCAAATACAATGTGGTGGTACCATTATAAAGGAAAATTTTATTTTGACTGCAGCTCATTGTATGTTTCGTAAAGACAAAAGGCGACAGATACTGCA  2000
      T  Y  Q  I  Q  C  G  G  T  I  I  K  E  N  F  I  L  T  A  A  H  C  M  F  R  K  D  K  R  R  Q  I  L  H
                                                     Protease activate site
2001  TCTCAGGCCGAAAGAAGAAATTATAGTAAAAATAGGATTGTTTAGTTTAAATGACGAATCAAAGGTTCAAGAATTCGGAGTCCAGAGGATTTTTATTCAT  2100
        L  R  P  K  E  E  I  I  V  K  I  G  L  F  S  L  N  D  E  S  K  V  Q  E  F  G  V  Q  R  I  F  I  H

2101  GAAAAATATGACCCAGGGAATTATTCGGTACAGTTGTATGACTATGACATTGCTATACTGGAACTGGATGGATCAATTGTATATGATCGACGTATCAGGC  2200
       E  K  Y  D  P  G  N  Y  S  V  Q  L  Y  D  Y  D  I  A  I  L  E  L  D  G  S  I  V  Y  D  R  R  I  R
                                         Protease activate site
2201  CTATATGTCTGCCACCCAGAGATCTGGCGGAAAACAGTCATCTGTACAGTTTTAAGGAATTTGGCTGGGCAGGAGGTTGGGGTCACAATGGTGTACTGAT  2300
      P  I  C  L  P  P  R  D  L  A  E  N  S  H  L  Y  S  F  K  E  F  G  W  A  G  G  W  G  H  N  G  V  L  I

2301  TCCGTATAATCCATATAGTCTGAGTATGCAACTTAATAACAGTCCAAACTTGAAACAGCTACGGATTCCTATCCAGTCCAATGAGCGATGTTCTAAAAGT  2400
        P  Y  N  P  Y  S  L  S  M  Q  L  N  N  S  P  N  L  K  Q  L  R  I  P  I  Q  S  N  E  R  C  S  K  S
```

Supplementary Fig. S11 (continued)

```
2401  TTAGAAAAGAAACTCGATCCAACATTTTTCACAGATCGCATGTTTTGTGCTGGCACCGGAGAAACTGGAGTTGACACTTGCCTTGGTGACAGCGGCGGTC  2500
       L  E  K  K  L  D  P  T  F  F  T  D  R  M  F  C  A  G  T  G  E  T  G  V  D  T  C  L  G  D  S  G  G
                                                                                         Protease activate site
2501  CCCTCATGCAATCACAGATGAATTCGGAAGGCTTTCTTTACTGGACTCAAGTTGGCATCGTAAGCTGGGGAATAGGTTGTGGAATGGAAGACATGTACGG  2600
      P  L  M  Q  S  Q  M  N  S  E  G  F  L  Y  W  T  Q  V  G  I  V  S  W  G  I  G  C  G  M  E  D  M  Y  G

2601  ATTCTATACACACGTGCAAAAATTAAGATCATGGATAGATGAAATAATATACACTCAATAGTGAAAACTCTAAAGACTTGCACATTATTTGCTTTTTTTT  2700
        F  Y  T  H  V  Q  K  L  R  S  W  I  D  E  I  I  Y  T  Q  *

2701  TTTTTATAATTTCAGTGGATTTGATAAAGGAATATAATTACACATTTCATTAATTTTCCAGTTATGTATATTTTTTTTCTTTTAGACACCTATTTTTAAG  2800
```

HaadFB-3

```
   1  TCTAACAAGACTTCCTTCGCAGTAAATAATTGCGGAAAGGCAGGAAGTCGATATTTTTTCATCAGATTCACGCAAGACTTTACGGACAAAATTATGTATT  100
                                                                                                    M  Y

 101  GGTGCCATTCCACAATGTTTACCTTCATTGTATGCCTAACACCTTTACTTGTTACATCCGTTTTGTCAGCTAAATCGAAAACATGTTCACCTTACCCCCA  200
      W  C  H  S  T  M  F  T  F  I  V  C  L  T  P  L  L  V  T  S  V  L  S  A  K  S  K  T  C  S  P  Y  P  H
                                                                                                     SCR1
 201  TGTGAGCCTGGATCATGGTGATATTCATCGAATTGGCAGACAAAAGTATTCCTTCAAATGCAATAATGGATACTTTCTCACATCATCGTCCACAGTTCGT  300
        V  S  L  D  H  G  D  I  H  R  I  G  R  Q  K  Y  S  F  K  C  N  N  G  Y  F  L  T  S  S  S  T  V  R

 301  TGCTTTCGTGGTAACTGGACGTCATCTAAAGCACCAAGGTGCCTCAGGATACAGGGACAATGCGATGAACCTCCTGTAGTGAAACATTCTCTTAGATATG  400
       C  F  R  G  N  W  T  S  S  K  A  P  R  C  L  R  I  Q  G  Q  C  D  E  P  P  V  V  K  H  S  L  R  Y
                                                                                  SCR2
 401  GAGACGAAAGACATGTGGGTGCTAAAGTAAGCTACGTCTGTAAGGATGGATTTACTTTACTGGGCCATTCGCAACTCGTTTGCTCACGGAATGGTCGCTG  500
      G  D  E  R  H  V  G  A  K  V  S  Y  V  C  K  D  G  F  T  L  L  G  H  S  Q  L  V  C  S  R  N  G  R  W

 501  GAATAAACGCGCGCCAACTTGTATGGATGAGTCTGAACCTCTTCAAACTGTAGCTGAGAGACTTAAAAACAGCTTTGTAACGGAACTTGCATATCATTCC  600
        N  K  R  A  P  T  C  M  D  E  S  E  P  L  Q  T  V  A  E  R  L  K  N  S  F  V  T  E  L  A  Y  H  S

 601  TCTGGTACACCTGAAGGACGGCTGTTAGACAGATCTTCTTTGTTGTTGGGGCTGGAGCTGTATCTGCTAATAGACAGATCAAGCAGCATCGACCCTGTTC  700
       S  G  T  P  E  G  R  L  L  D  R  S  S  L  L  L  G  L  E  L  Y  L  L  I  D  R  S  S  S  I  D  P  V
                 Factor D cleavage site                            vWFA      Mg²⁺ binding sites
 701  ACCTTGAAGACGCCAAGAATTTTGTCAAATTTTTGCTAAGAAGATTTGGAGTCAACAACAAACCCAATAATAATAATGGAACTCGGGCATCGGTCCTTGC  800
      H  L  E  D  A  K  N  F  V  K  F  L  L  R  R  F  G  V  N  N  K  P  N  N  N  N  G  T  R  A  S  V  L  A

 801  CTTTGGGACAGAAGTGCAGATAGTTTTCAATATTGATGATACAAATATATCCAATCCACGGATAGCTGCTGCAGCAGTTGACGACATTCACCCTAATGGA  900
        F  G  T  E  V  Q  I  V  F  N  I  D  D  T  N  I  S  N  P  R  I  A  A  A  A  V  D  D  I  H  P  N  G

 901  GGGGGTACTAACATGGAAGGCGCCCTTACAAAAGTACTAATAGACTTCCAGAAGCTGCGTAAAAGGGCGAAGCGAGCCCTCTTTTTGATGACAGATGGTG  1000
       G  G  T  N  M  E  G  A  L  T  K  V  L  I  D  F  Q  K  L  R  K  R  A  K  R  A  L  F  L  M  T  D  G
      Mg²⁺ binding site                                                      Mg²⁺ binding site
1001  AACCAAACATCGCAAACCCTGAAATTACGCCTCAGGATATAGCTCAACAATTAAAGAAATCTCCTAATGATTTCGAAATTTTTACAGTCGGTATCGGGAA  1100
      E  P  N  I  A  N  P  E  I  T  P  Q  D  I  A  Q  Q  L  K  K  S  P  N  D  F  E  I  F  T  V  G  I  G  K

1101  AGGTATTAAGATGAACCTTCTGAACGAACTGGCCAGCGAACCGCCTCTAAGCCACGTATTTATTTTGGAAAATTATCCTGATTTAAATGAAGTCATGAAA  1200
        G  I  K  M  N  L  L  N  E  L  A  S  E  P  P  L  S  H  V  F  I  L  E  N  Y  P  D  L  N  E  V  M  K

1201  ATCATTGAAGACAGCAAACCACCTCCACCACCCATCAGTAAAGACCAATGTGGTTATAACGTCTCTAAAGAAAATCGGCCTTGGTTAGCGACACTTTACA  1300
       I  I  E  D  S  K  P  P  P  P  P  I  S  K  D  Q  C  G  Y  N  V  S  K  E  N  R  P  W  L  A  T  L  Y
                                                                                          SP
1301  TTGGAGGTGCTCCGTTTAAGATGTGCAGTGGTGTGTTGATATGTAACCAGTGGGTACTAACTGCAGCTTCGTGCCTTCAGGATGCTCAGAGTCGAGTTGA  1400
      I  G  G  A  P  F  K  M  C  S  G  V  L  I  C  N  Q  W  V  L  T  A  A  S  C  L  Q  D  A  Q  S  R  V  D

1401  CATGAAAGATGTCTTTGTTGTACTTGGAGAACGGCACTTGTTGAAAGCTGAGCAAGGACAGACCAACTTTTACGTTACTGACGTACAGATCCATCAAAAT  1500
        M  K  D  V  F  V  V  L  G  E  R  H  L  L  K  A  E  Q  G  Q  T  N  F  Y  V  T  D  V  Q  I  H  Q  N

1501  TATAACCCAAAACAAAACAGCATTGAAAATAACTTGGCTGTTCTCAAGCTTAATTTGCCTGCTTCGAGATACAGACCGGCTTGTCTACCACCAACAGACC  1600
       Y  N  P  K  Q  N  S  I  E  N  N  L  A  V  L  K  L  N  L  P  A  S  R  Y  R  P  A  C  L  P  P  T  D

1601  GAGCGATACCTCTACATTTGAATCTGAAGATCAACGCATCCATAACAGGTTGGGGTCGAACTTCAGCATCTAAAATGACATATGCAACGGCAAATGATAT  1700
      R  A  I  P  L  H  L  N  L  K  I  N  A  S  I  T  G  W  G  R  T  S  A  S  K  M  T  Y  A  T  A  N  D  I

1701  AGCATTTGATATGTCATCCTCATCTGTCTCGTTGTCCGAAGAACGGGAGTGTCCTATACGAAGAAGAGTCACAACGCCACTTTGTGCTGGACAAGGATCA  1800
        A  F  D  M  S  S  S  S  V  S  L  S  E  E  R  E  C  P  I  R  R  R  V  T  T  P  L  C  A  G  Q  G  S

1801  AAGACGTGTTTTGCTTTGGTTGGTTCTCCGCTTATGGCAGAAGATTCTTCGACAGGTTTCCAGCACATACTCGGAATATTAGTAGATCGCAGACATTGTT  1900
       K  T  C  F  A  L  V  G  S  P  L  M  A  E  D  S  S  T  G  F  Q  H  I  L  G  I  L  V  D  R  R  H  C

1901  CAAAAAAAGGACAAAATCAATACATCGAACTCACAAGGCATATAGGTTGGATAAATCAAGCGACTTCCAACTGTCAGTTAAAGCACTGGGGTGTTGCCAA  2000
      S  K  K  G  Q  N  Q  Y  I  E  L  T  R  H  I  G  W  I  N  Q  A  T  S  N  C  Q  L  K  H  W  G  V  A  K

2001  ATAAAAGATATCAAATTCAGCACTAGAACACGGGATAAAGAAGCAAGAAGTGTCGGTTATTTACTGATCCTGATATTGTTAACAATACAGACAATGAATA  2100
        *
```

Supplementary Fig. S11 (continued)

ScsuFB-1

```
   1  AAAGGAACTTTCAATTTGGAAAAATTGCGCACAAGAATTTATGCTTTATAAAGCGAACCATAGTTGAAATCATCCTCATTTGGTAAAAGGATAATACTCG  100
 101  TCGAGATTTAATCTTTCTTAATGTCATGAAAAGAAATATTGTGCCCAGTTAAATTTGAAGTCTCAAATACAAAGATTTAAAATTTGTTTTTCTTTGATTT  200
 201  TTTCTTTGATTTTCTTGTCCATGAAATGGAGAATTTTATGAGAAAACTTGTTGCAGTCGGATTTGTAATTATAATTTCAGCGTTACAAGTGAATAGTGAT  300
                                M  E  N  F  M  R  K  L  V  A  V  G  F  V  I  I  I  S  A  L  Q  V  N  S  D

 301  TGTCCTAGACGAAGAACACTTGGGAATGTGATTGTAATTTATCCACCTAATGTTCCACGAACCGGACAACTACCTGAAAACACGATGATTTCTTTTAAAT  400
       C  P  R  R  R  T  L  G  N  V  I  V  I  Y  P  P  N  V  P  R  T  G  Q  L  P  E  N  T  M  I  S  F  K

 401  GTAAACCTGGATACCAATATGTACAAGGGGATAAAACAGTAAGGTGTTTACATGGTCGTTGGGTCGCAAATGGTAAAACAGCTAATACATTTATTTGTGA  500
      C  K  P  G  Y  Q  Y  V  Q  G  D  K  T  V  R  C  L  H  G  R  W  V  A  N  G  K  T  A  N  T  F  I  C  E

 501  AATTAAAGACTGTCCTGCAGTTAACATTGAAAATGGCACAACAGAAGTCATTAACGGTATTCAGATTAAATATTCTTGCAACAGAGGGTTTAAGCTAAAA  600
        I  K  D  C  P  A  V  N  I  E  N  G  T  T  E  V  I  N  G  I  Q  I  K  Y  S  C  N  R  G  F  K  L  K
                                SCR1
 601  GGATTTGCAAATAGTCAGTGCCGACTAAATAAACAAGGAAATTATTATGTCGTACATACGCCAACATGTGAAGCCCTTGCTGAATGTGAAAATCCCAAAG  700
       G  F  A  N  S  Q  C  R  L  N  K  Q  G  N  Y  Y  V  V  H  T  P  T  C  E  A  L  A  E  C  E  N  P  K
                                                                                               SCR2
 701  ACAGTGGAGTTGAATATGAAGGAAAATGTTGTTACGCTGATGACCGTATAAAACTTTCTTGCAGTACAGGTTATAGATTAGAAGGCACTGCAGAAATTAT  800
      D  S  G  V  E  Y  E  G  K  C  C  Y  A  D  D  R  I  K  L  S  C  S  T  G  Y  R  L  E  G  T  A  E  I  I

 801  TTGTCTATCAACTGGCAATTGGTCCTCTTATTTTCCTATTTGTAAATGGATTCAAAAGTGTGAAGATCCTGGAAATTCATCTTTCAGCATTCGGAATGGT  900
        C  L  S  T  G  N  W  S  S  Y  F  P  I  C  K  W  I  Q  K  C  E  D  P  G  N  S  S  F  S  I  R  N  G
                                                                                 SCR3
 901  AGTTGCTGTGAAATCGGAGATATTCTTACCTACACCTGTGAAGAAGGATACGAATTAGTAGATCCTGAGGCTGGAAACTGGATGACTTGTCTACCTGAAG  1000
       S  C  C  E  I  G  D  I  L  T  Y  T  C  E  E  G  Y  E  L  V  D  P  E  A  G  N  W  M  T  C  L  P  E

1001  AAAGATGGAGCGGACCAAAACCTAGATGCAAACCTAAAGGAGCAGTTGTATGCCCAGATATCCCGTTAAAAAAAAACAACTTAAGATTTGAAGATATAAT  1100
      E  R  W  S  G  P  K  P  R  C  K  P  K  G  A  V  V  C  P  D  I  P  L  K  K  N  N  L  R  F  E  D  I  I
                                                                         SCR4
1101  AACCGAGAAAAAGGATGTTTATGTCGCTTGGGACGAATTATATTTACAGTGTAAAGACGGATACACAACGAATGACACCTTGAGTTTATTTTGTCAGGAA  1200
        T  E  K  K  D  V  Y  V  A  W  D  E  L  Y  L  Q  C  K  D  G  Y  T  T  N  D  T  L  S  L  F  C  Q  E

1201  GATGGGCAGTGGGATGAAAAAGTTCCAATTTGCATTCCTGAAAATCCTTGCTTACGTCCACCTATACCAGATAATGGACAAATCGATGAATTGATTGATA  1300
       D  G  Q  W  D  E  K  V  P  I  C  I  P  E  N  P  C  L  R  P  P  I  P  D  N  G  Q  I  D  E  L  I  D
                                                                       SCR5
1301  ATGATCAATTATATTTGCCTATTGGATTCGAAATAAATATTCTTTGCAAAGAAAATCATGAAATTAAGTCTGGTCATTATGGCAGCCGTTGTTTAGGAGA  1400
      N  D  Q  L  Y  L  P  I  G  F  E  I  N  I  L  C  K  E  N  H  E  I  K  S  G  H  Y  G  S  R  C  L  G  D

1401  TAATAAATGGGAAAACGGAATTACAAATTGCACATTGATCGAATGTAGTGATCCAGGAATACCTGATGGTGCTCAAAGATATGGTGATGATTTTCATTTG  1500
        N  K  W  E  N  G  I  T  N  C  T  L  I  E  C  S  D  P  G  I  P  D  G  A  Q  R  Y  G  D  D  F  H  L
                                                                  SCR6
1501  GGAAGTTCTGTTACCTACAAATGCATAGGGAATTTAAATCTTCTTGGCTCAGAAGTAAGAACATGTGAACAAACAAGACGATGGAGTGGTCAAACAACAG  1600
       G  S  S  V  T  Y  K  C  I  G  N  L  N  L  L  G  S  E  V  R  T  C  E  Q  T  R  R  W  S  G  Q  T  T

1601  TATGTGACGCAGGAGATACACATTGTCCCGATCCAGGAATACCTTTACACTCAAGTAGAAAAATTGAAGGATTTAATCACGGAGACAAAATAATTTACTC  1700
      V  C  D  A  G  D  T  H  C  P  D  P  G  I  P  L  H  S  S  R  K  I  E  G  F  N  H  G  D  K  I  I  Y  S
                                              SCR7
1701  ATGTGAGCCAGAAAGTCAAATGATCGGAAATTCTACTAGAATTTGTCTTGAAGATGGCACTTGGTCCGAGGAAGAAGTTATATGTTTAGGTTCAAATGAA  1800
        C  E  P  E  S  Q  M  I  G  N  S  T  R  I  C  L  E  D  G  T  W  S  E  E  E  V  I  C  L  G  S  N  E

1801  TATCCTGATATGGAAGTTGTTGCAAAAGCTCTCTCAAAAACTATGGTTCAAATTTCAAGACAATCAGCTTATAGCAAACCAGTTTTTGGACGCAGTACGA  1900
       Y  P  D  M  E  V  V  A  K  A  L  S  K  T  M  V  Q  I  S  R  Q  S  A  Y  S  K  P  V  F  G  R  S  T
                                                                                      Factor D cleavage site
1901  TTTCTTCTGATCATGCAACTGGTGTTACTATATTCTTCTTGCTGGATGCTTCTGGGAGTATTACTAAAACAGAATTTAAAAAATCTAAAGAACTCGCTAT  2000
      I  S  S  D  H  A  T  G  V  T  I  F  F  L  L  D  A  S  G  S  I  T  K  T  E  F  K  K  S  K  E  L  A  I
                                        vWFA        Mg2+ binding sites
2001  ACATGTAGTTCGTCAGATAGGAATATCTACACATAAAGGAGGAGTACGAATAAGTGTTATAAATTTTAGCCAAAACGTTGAGACAGTTGTAACATGGGCT  2100
        H  V  V  R  Q  I  G  I  S  T  H  K  G  G  V  R  I  S  V  I  N  F  S  Q  N  V  E  T  V  V  T  W  A

2101  GTTGAATCAGTCGAACAAGCTATAACTCGGATTGACAGTATTGAAAAACGAAAAGACGAAGGTACCAACATTGCTAAAGCCTTGAATCATCTAGCTAATG  2200
       V  E  S  V  E  Q  A  I  T  R  I  D  S  I  E  K  R  K  D  E  G  T  N  I  A  K  A  L  N  H  L  A  N
                                                              Mg2+ binding site
2201  AAGTAGCAAGTACTAAAGAAAATGACTTGGTTCGAGATAACAAAAATATAGCATTTATTATATCTGATGGAAATGCTAATGAAGGAGGAAAACCGGAAAA  2300
      E  V  A  S  T  K  E  N  D  L  V  R  D  N  K  N  I  A  F  I  I  S  D  G  N  A  N  E  G  G  K  P  E  K
                                                                 Mg2+ binding site
2301  GGAGGCAAAACTACTTAAGAAATTTGGAGTTAAAATGTTTGCTATTGCAGTTGGGAAAAAAAGAGATAGAGTTATTTTGAAAATCATTACAAATGACGAA  2400
        E  A  K  L  L  K  K  F  G  V  K  M  F  A  I  A  V  G  K  K  R  D  R  V  I  L  K  I  I  T  N  D  E

2401  AATAACATATTTGAATTTGATTCCTATCAAAAACTCTTCGATGTCATCTCACGTACTATAGGAAACCAAACAAATTTCCAACTGTGTGGAATTAATCAGC  2500
       N  N  I  F  E  F  D  S  Y  Q  K  L  F  D  V  I  S  R  T  I  G  N  Q  T  N  F  Q  L  C  G  I  N  Q

2501  CAAACCTAATTAAAAGAAATAAAATTAACCTTAATAAAAATCGGAGGAAACGTTTTATTGACGGTATTGGACGAATTGTTGGCGGGCAAGAAGCCGAATA  2600
      P  N  L  I  K  R  N  K  I  N  L  N  K  N  R  R  K  R  F  I  D  G  I  G  R  I  V  G  G  Q  E  A  E  Y
                                                                                              SP
2601  TTCGTGGCCGTGGATGGCTGCCTTATATCTTAAGAACAAAAATAACAAGTGGGAGTTTAAATGTGGAGGTTCTATCATTCATAGTGATTGGATTTTAACA  2700
        S  W  P  W  M  A  A  L  Y  L  K  N  K  N  N  K  W  E  F  K  C  G  G  S  I  I  H  S  D  W  I  L  T

2701  GCAGCTCACTGCTTTTTTGTAGTTGATGGTTTTCTAGATATAGATGAAGAAAAATCTGAAGAAATTCAGAGTGCAAAATGGAAGGCTTCTGTAGGTAGGC  2800
       A  A  H  C  F  F  V  V  D  G  F  L  D  I  D  E  E  K  S  E  E  I  Q  S  A  K  W  K  A  S  V  G  R
         Protease activate site
2801  TTAATATTTCGGATGAAAATTCGGGACTTTTTTACAAGATAAGCACTATAAAAATACATCAGGAATATGATTCAAAATCTTATGGAAATGATATTGCCCT  2900
      L  N  I  S  D  E  N  S  G  L  F  Y  K  I  S  T  I  K  I  H  Q  E  Y  D  S  K  S  Y  G  N  D  I  A  L
                                                                                       Protease activate site
2901  TTTGAAAGTGGACAAATCTATTGTATTTGACATCTATACGAGTCCTATTTGTCTTCCTCCTTCAATGAATGACATACCTGAAAAGAGTGAAATGTACTCA  3000
        L  K  V  D  K  S  I  V  F  D  I  Y  T  S  P  I  C  L  P  P  S  M  N  D  I  P  E  K  S  E  M  Y  S
```

Supplementary Fig. S12

```
3001  AGCAAATCCGAAGCATATGTTGTTGGCTGGGGTAATACAGAAGTTGCAGATCTTAGTTCCTCACAGAAGCAGAAAAACAGTGATATATTACTACAGCTAA  3100
       S  K  S  E  A  Y  V  V  G  W  G  N  T  E  V  A  D  L  S  S  S  Q  K  Q  K  N  S  D  I  L  L  Q  L

3101  AGCTACCTTTACATGATGATAAAATTTGTAAACAGAAAGTTAACTTTAAAGCATCTCAAGCAGGGAAAGAAATAACACATTACTTGACGCCATCTCCTAA  3200
      K  L  P  L  H  D  D  K  I  C  K  Q  K  V  N  F  K  A  S  Q  A  G  K  E  I  T  H  Y  L  T  P  S  P  N

3201  TTCACTATGTGCTGGTACGGGAGAAGGCAAACAAGATGCTTGTCAAGGAGATAGTGGAGGACCTCTTATGCAACTTATTCACAGCAACAGTAGTTCAAAT  3300
        S  L  C  A  G  T  G  E  G  K  Q  D  A  C  Q  G  D  S  G  G  P  L  M  Q  L  I  H  S  N  S  S  S  N
                                                 Protease activate site
3301  ATTATGAAACGATGGTTTGAAATAGGAATTGTCAGCTGGGGAATTGGTTGTGGATTAGAAGGATATTATGGATATTATACTCATGTGTCTAAATTGAAAC  3400
       I  M  K  R  W  F  E  I  G  I  V  S  W  G  I  G  C  G  L  E  G  Y  Y  G  Y  Y  T  H  V  S  K  L  K

3401  AATGGATCTTAAACGAAACGAATTCAGAGTTTGAAATTGTTGCTTTAAATGTTTGAGATTTCAAATTAACTAAATTCATTTTTTTCTTGAAAGCATGAAA  3500
      Q  W  I  L  N  E  T  N  S  E  F  E  I  V  A  L  N  V  *

3501  ATTCAAAACAAGTAATTAAAGAAATTATTAATTTTAATACGTTATCACTTATAAGAACTAGTGAAATTAATGAGACATATTTAAAACTAATATAAATTTT  3600
3601  AGGTAAATCTGTTTTTCAAAGTATAACTTTAAATTAATGTTTTATCTGAAGTTTTTCATAAACATTAAAAATAGAATTTTAAAAATGTGGAATTTTATCA  3700
3701  ACAAC  3705
```

ScsuFB-2

```
   1  GTAAACTGTATTTAAGATGCTTTCTTTTTGCAGTTAGATTCTTATTGCAATACGAAAAACTCAATCTTAATCATGATTAACAGAGTGGTTTTTATATTTC  100
                                                                                 M  I  N  R  V  V  F  I  F

 101  TCTTTTATTTCTATCCTTTGAAATGTTTTGCAAGAATTTGCCCACCTTTGCCGATTCTAGATCATGGAAAATATGTACGAATCTCTGAAGGAAAAGTTCG  200
      L  F  Y  F  Y  P  L  K  C  F  A  R  I  C  P  P  L  P  I  L  D  H  G  K  Y  V  R  I  S  E  G  K  V  R
                                                          SCR1
 201  AGTAGAGTGTCATCCTTCTTATGAATACAACAATGCATTTGATATTTCAAGACAAACTCAGATAATAACATGTTATAAATCAGGGCAGTGGGATGATCCA  300
        V  E  C  H  P  S  Y  E  Y  N  N  A  F  D  I  S  R  Q  T  Q  I  I  T  C  Y  K  S  G  Q  W  D  D  P

 301  AAGCCAAATTGCTCAAAGAAACATCAGCACTGTCTTCCTCCCAGTGACATTGATCATGGGTTCTTAGTCGGAGATCCTCCTTATCCTCCTGGAAGTGAAG  400
       K  P  N  C  S  K  K  H  Q  H  C  L  P  P  S  D  I  D  H  G  F  L  V  G  D  P  P  Y  P  P  G  S  E
                                                  SCR2
 401  TATATTATCGTTGTAATGAAGGTTACAAAATTGTTGGACCCAGTAAATTGATTTGTAAATATCCTCAATATTTCTGGAATGATAATCCTCCGGAATGTAT  500
      V  Y  Y  R  C  N  E  G  Y  K  I  V  G  P  S  K  L  I  C  K  Y  P  Q  Y  F  W  N  D  N  P  P  E  C  I

 501  AAAAATTAAACCACCACTGCAAGTTGTAGCAGAAAATATAGGGGACAATTTGGTCGACAAATTGACTAAATTAGATGTCGGACGTGTCATACATGGCGAT  600
        K  I  K  P  P  L  Q  V  V  A  E  N  I  G  D  N  L  V  D  K  L  T  K  L  D  V  G  R  V  I  H  G  D
                                                                               Factor D cleavage site
 601  GTTGAATATCTTGGCCTAGATTTATTCCTTGCATTTGATAAATCTAATAGCATTTCTCCACAGCAATTTTTAGAAGGAATTAAATTCGCAAAATTTTTGA  700
       V  E  Y  L  G  L  D  L  F  L  A  F  D  K  S  N  S  I  S  P  Q  Q  F  L  E  G  I  K  F  A  K  F  L
                            vWFA         Mg2+ binding sites
 701  TTAAACAGTTTAACGTGTCAAATAGTGATCAAAAGAAAGTTGGGGGAACTCGGCTGGCAGTATACACATTTGGTAATGATGCTAAAGAGGAAATAAATCT  800
      I  K  Q  F  N  V  S  N  S  D  Q  K  K  V  G  G  T  R  L  A  V  Y  T  F  G  N  D  A  K  E  E  I  N  L

 801  AACAGATACGACCATCACAAGTACAAAAGCTGCTATTAACAAACTAGATCTTATCAGATGTAGTAAATTTTGTGATGGTGCTACAAACATGGCTGATGCT  900
        T  D  T  T  I  T  S  T  K  A  A  I  N  K  L  D  L  I  R  C  S  K  F  C  D  G  A  T  N  M  A  D  A
                                                                                  Mg2+ binding site
 901  TTGAAGAAAATTGGACACATAGCGCCTAAGCAGACTCGCAAAGAGGCAAAAAAAGTGCTTTTCATGACATCAGATGGAGTACCTACAGCAGATCCTAAGT  1000
       L  K  K  I  G  H  I  A  P  K  Q  T  R  K  E  A  K  K  V  L  F  M  T  S  D  G  V  P  T  A  D  P  K
                                                                          Mg2+ binding site
1001  CAAAGGATGTAACTTATTATACAAACAATTTAAAAAAATTGGGATTTGAAATCTATACTGTTGGAATAGGCCAAGATATAGATGAACAGCTGTTGAAAGA  1100
      S  K  D  V  T  Y  Y  T  N  N  L  K  K  L  G  F  E  I  Y  T  V  G  I  G  Q  D  I  D  E  Q  L  L  K  D

1101  TTTGTCTTCAACTCCAATTGAAGAACATATGTTTTTATTGGAAAAATTCAAAGATTTCGCTGAAATTATGGATCTTATTCGGAATGGAACAACTGAACCA  1200
        L  S  S  T  P  I  E  E  H  M  F  L  L  E  K  F  K  D  F  A  E  I  M  D  L  I  R  N  G  T  T  E  P

1201  CCTCCGCCATTGCCAGAACAGTGCGGCTATATCGCAGAAAATTTTTACAAAACAAGAAATTTAGAAACTGGCTTGGCCAAATTAGGCAACTGGCCTTGGT  1300
       P  P  P  L  P  E  Q  C  G  Y  I  A  E  N  F  Y  K  T  R  N  L  E  T  G  L  A  K  L  G  N  W  P  W
                                                                                              SP
1301  TAGCAGCTATTATAGTGAAGGATTACACTTCTGGTGGTTCGAGATTTGCATGCAGTGGCATCCTTATATGTGAAGAATGGGTATTAACTACAGCACAGTG  1400
      L  A  A  I  I  V  K  D  Y  T  S  G  G  S  R  F  A  C  S  G  I  L  I  C  E  E  W  V  L  T  T  A  Q  C

1401  TGTGACTGATGATCTCAATATTCAGTATGAACCAGAGAATGTATATGTTGTTGTCGGAGAAGATAATTTCGAGAAACTTAACGATAAAGAGCAGTTGTTT  1500
        V  T  D  D  L  N  I  Q  Y  E  P  E  N  V  Y  V  V  V  G  E  D  N  F  E  K  L  N  D  K  E  Q  L  F

1501  TATGCTGTGAAAATAATACGTCACCCAAACTTTATACACAACAAAACAAAGGTTCGGAATGATATTGCGTTGATCCAGTTAAATACAAAAGCTGAAATTA  1600
       Y  A  V  K  I  I  R  H  P  N  F  I  H  N  K  T  K  V  R  N  D  I  A  L  I  Q  L  N  T  K  A  E  I
                                                            Protease activate site
1601  ATGATTACGTTCGTACAGCGTGCATTTTACAGTTTGATTTCGATCATTCCTCTAAACCAGGAGAAGTAGGATACATGATTGGTTGGTTGATATCAAGTGA  1700
      N  D  Y  V  R  T  A  C  I  L  Q  F  D  F  D  H  S  S  K  P  G  E  V  G  Y  M  I  G  W  L  I  S  S  E

1701  AGAACAAGTTCGCATTCCACAAGAACCAGTAAAAGTATTTCAAGCTGAACAAGTGGAAATGAAAGTAAACGGACTTGATAATTGTGAGATTATAAAACCA  1800
        E  Q  V  R  I  P  Q  E  P  V  K  V  F  Q  A  E  Q  V  E  M  K  V  N  G  L  D  N  C  E  I  I  K  P

1801  CCTCAGTATTTATGTGATGACTCAATCATAAATGCACATGGAAGGTATGAATCTGATAAAACCTGCAGTGGTTATGACAAAGGAAGTCCCTTCTTAATGA  1900
       P  Q  Y  L  C  D  D  S  I  I  N  A  H  G  R  Y  E  S  D  K  T  C  S  G  Y  D  K  G  S  P  F  L  M

1901  GTCAGGGTAAGAGGATGGCTGCTGTGGGTCTGGCTTCCCATGTCAAGGGTTGCACACTTCAGAGCAAAACAGGGTTTTTTACCCGCATCAGTTCTTACTA  2000
      S  Q  G  K  R  M  A  A  V  G  L  A  S  H  V  K  G  C  T  L  Q  S  K  T  G  F  F  T  R  I  S  S  Y  Y

2001  CGACTGGATTCGTGAAAATAGTGAATTTTGTTCTGATAATCATCAATAAGATTTAAATCAATTTAAGCTTGATCTTCCTTTTTTTATTTCAAACAAACTA  2100
        D  W  I  R  E  N  S  E  F  C  S  D  N  H  Q  *

2101  TTTTTGTGAATAATTCAGATATATGAGATTGCAATTAAAGTATGAATAAATTAAATAAGAAAAAAAAA  2168
```

Supplementary Fig. S12 (continued)

```
                                                                                                          / SCR
AmspFB-1 : MDSMFRICLKLFLLCFVVNSASSD-----------------------------------------------------------------------CHGNSHALKLYNGKFRVNKLVSYGK :   49
TatrFB-2 : MTELATCIIILVSAVGPFLDLAHG-----------------------------------------------------------------------CKFQNAPINGKWEFENPGDAGKKHF :   49
HaadFB-1 : MQKNPTLCSCFKVIISVILFNFITCD---------------------------------------------------------------------CPYPKRELPNGRFVYQQLKDINETS :   51
ScsuFB-1 : MENFMRKLVAVGFVIIISALQVNSDCPRRRTLGNVIVIYPPNVPRTGQLPENTMISFKCKPGYQYVQGDKTVRCLHGRWVANGKTANTFICEIKDCPAVNIENGTTEVINGIQIKYSCNR :  120
TatrFB-1 : MTLIYCLYTIFILVVVGSFIDLVQSAEVTKK------------------------------------------------------------------------------------------ :   31
HaadFB-2 : MKRRKLCLWLQFFLGHFMLEFTYCAPNIY-------------------------------------------------------------------------------------------- :   29
AmspFB-2 : ----LRAPSVMRPYGNISLLNSFFICLLQFLHSTEGK------------------------------------------------------------------------------------ :   33
HaadFB-3 : MYWCHSTMFTFIVCLTPLLVTSVLSAKSKT------------------------------------------------------------------------------------------- :   30
ScsuFB-2 : MINRVVFIFLFYFYPLKCFARI--------------------------------------------------------------------------------------------------- :   22
HosaFB   : MGS-NLSPQ--LCLMPFILGLLSGGVTTTPWSLAWPQGS---------------------------------------------------------------------------------- :   36

                                                  /  / SCR                                                              /   / SCR
AmspFB-1 : NKVVVAMCNKGYDLFGDRILICDGTNWIVRRGKTKPTCREKKCPEPPAIENGFVYYRKGQIARRGG------TRVFYKCLDGFQISSSRPSLFCLKRNKRWRGTPPKCISKKKCPELGDV :  163
TatrFB-2 : PPFYVLSYVCDKGYVQRSTKPYLMCKNGKWDGDIPK-CDRMLCSKPLINNGFFEIKNGNEGFVE------GTTVNFKCHRGHRLVGKSLLMCTFRGSEMRWNGNFPKCKERRKCKNPGVS :  162
HaadFB-1 : AKELYILRFKCYAGYQRIGPRYIQCLDDEWSDVVPT-CAKMQCSDPPDVDKAEFKLHSGVDKFPTIGASVTYSCKIGYELKNTSASTLHCELDPGASKAIWKGEIPFCKEKESCPDPGVS :  170
ScsuFB-1 : GFKLKGFANSQCRLNKQGN----------YYVVHTPTC------------EALAECENPKDSGVEYEGKCCYADDRIKLSCSTGYRLEGTAEIICLSTGNWSSYFPICKWIQKCEDPGNS :  218
TatrFB-1 : -------------------------------------------------------------------------------------------------------------------CKDPGIS :   38
HaadFB-2 : ------------------------------------------------------------------------------------------------------------------------- :    -
AmspFB-2 : ------------------------------------------------------------------------------------------------------------------------- :    -
HaadFB-3 : ------------------------------------------------------------------------------------------------------------------------- :    -
ScsuFB-2 : ------------------------------------------------------------------------------------------------------------------------- :    -
HosaFB   : ------------------------------------------------------------------------------------------------------------------------- :    -

                                                              /     / SCR
AmspFB-1 : ANGRRTPSNFEEGSVVQYECNQDSTMIGPSV---IKCRSDATWNDTIPECKPQNA--CKLPTQDEIPHSRMLSDVPGAKY-VLADEFVTMVCEKG-FRIEGMNFLQCLQGGTWDSDFAKC :  276
TatrFB-2 : PDGKRWGKCCYAGNSLTFSCNDGFDLVGSSR---ILCIRNGSWTSPRPLCKPRNATTCPFPL-DIPKVVMKPWKQGDYY---LPDDELESYCDTGYRLLSGSEYLYCKENGQWDSQFPVC :  275
HaadFB-1 : KHGFRVGSCCFSGDILQFSCNEDYELVGQKE---VQCLRSGSWSSGRPLCKPLSDY-CQLPP-SIPHGIVSGEKEGDYF---IPYDEAEVVCEPGYKHVGPTQFVMCEEEGEWEDEFGEC :  282
ScsuFB-1 : SFSIRNGSCCEIGDILTYTCEEGYELVDPEAGNWMTCLPEERWSGPKPRCKPKGAVVCPDIP-LKKNNLRFEDIITEKKDVYVAWDELYLQCKDG-YTTNDTLSLFCQEDGQWDEKVPIC :  336
TatrFB-1 : PDGERKGHCCFTGNKLIFSCKKGFDLVGNPL---LQCLPSGVWTSPRPLCKPKNAIICPFPP-EIPKVTKTAWKSGDYY---LPGDEVEMYCETSYRLLSGSEYLYCEENGEWDSQFPVC :  151
HaadFB-2 : -------------------------------------------------------CESPP-FIPHGTVSGIESTGNY---LPYDEVEIQCDPGYKYSGTTEFILCEEDGKWEEDFGAC :   88
AmspFB-2 : ------------------------------------------------------------------------------------------------------------------------- :    -
HaadFB-3 : ------------------------------------------------------------------------------------------------------------------------- :    -
ScsuFB-2 : ------------------------------------------------------------------------------------------------------------------------- :    -
HosaFB   : ------------------------------------------------------------------------------------------------------------------------- :    -

            /   / SCR                                                             /  / SCR
AmspFB-1 : SVVHGCPPAYPIANGGIIESLKVGVN-VSHGTTLNYFCNETYRMVG-SSWIECVDHEFEGLAWSNKPPKCFPIRCSDHEAPDNG-VILSNGPYYVGDVVSYSCYEGYKL--VGSKNRTCK :  391
TatrFB-2 : GVT-SCRVPQHLEDGRIVETEYTNLTDVSDGFNLNFICDQNYRLIG-SSWVKCTY-----LGWSSNFPKCQLITCPDPGIPENG-QRKGSGPFNIGDRVTFSCFSDYTI--LGSEERICL :  385
HaadFB-1 : TEI-ICDLPKQLVNGTIPEMLTTNLTSFPYGFEITYFCDDGFRLVGGDSWRTCSK-----KGWSGKTPYCEAIQCPDPGLPENG-FRTGDN-FEVGAKVRFKCFTGYYL--LGSFERYCK :  392
ScsuFB-1 : IPENPCLRPPIPDNGQIDELIDNDQLYLPIGFEINILCKENHEIKSGHYGSRCLGD----NKWENGITNCTLIECSDPGIPDGA--QRYGDDFHLGSSVTYKCIGNLNL--LGSEVRTCE :  448
TatrFB-1 : GVT-ECKVPEGLENGKISERQDTNLTLVPENFELNFICNQNYRLIG-SSWIKCTY-----LGWSSALPTCQLITCPDPGVPENG-LRKGAGPFNIGDRVIFSCFPQYKI--IGSEERICL :  261
HaadFB-2 : VEI-FCDPPEPPANGTIPELLNSNLTLFPFNFGITYECIQGYRLFGGDSWRFCNK-----NGWSGKTPHCKEIRCPDPGVPEHG-DREGNS-FQVGSKVRFRCFTGYNL--LGSFERFCM :  198
AmspFB-2 : ------------------------------------------------------------------------CPPLLGSR---VENGHIKAWSSRKLQVICDRSYIV--HGASTLFCN :   74
HaadFB-3 : ------------------------------------------------------------------------CS-PY-PHVS-LDHGDIHRIGRQKYSFKCNNGYFL--TSSSTVRCF :   71
ScsuFB-2 : ------------------------------------------------------------------------CPPLPILDHGKYVRISEGKVRVECHPSYEYNNAFDISRQTQIITCY :   68
HosaFB   : -----CSLEGVEIKGGSFRLLQEGQALEYVCPSGFYPYPVQTRTCRSTGSWSTLK---TQDQKTVRKAECRAIHCPRPHDFENGEYWPRSPYYNVSDEISFHCYDGYTL--RGSANRTCQ :  146

                         /    / SCR                                                    /
AmspFB-1 : DTGTW-SGALATCDSENTVCPNPGVPINGMKSGN-RYDAGDVVVFSCKPTYFLFGNATRICQSNGYWSGT-EVFCRGTNEFDDITHISNKLLTRFDMLKLISKKSYTNSDATNNTQSNSS :  508
TatrFB-2 : GNGRW-SGRLASCDHQRYYCPDPGVPVNGFKSGN-SYNLGDTVQFSCKAGHSAIGSSNRTCQANHKWSGE-QQFCLEPYFSDRPIDMMMRVTQLLEEKEEEQE--------ELFANLPSS :  494
HaadFB-1 : PNGQW-SGELSRCDTPSNYCPNPGIPVKGYKNIS-SYEMGDKVGFHCQPGYVQIGSEVRECLPNRTWSGT-ETTCLGPYDYDNSAQVRDVLRAKLAEKAEEQERETQRYREALYAS---- :  505
ScsuFB-1 : QTRRW-SGQTTVCDAGDTHCPDPGIPLHSSRKIE-GFNHGDKIIYSCEPESQMIGNSTRICLEDGTWSEE-EVICLGSNEYPDMEVVAKALSKTMVQISRQSAYSKP------------ :  552
TatrFB-1 : GNGRW-SGLLSSCNHPNYYCPDPGVPVGGIKSGS-SYDLRDTINFFCKPGYSFIGSANRTCQMNHTWSGE-QVFCLEPYYSDSPTDALTRLNNVLDEKESEQKPISKISLWHIFKTSSNN :  378
HaadFB-2 : PNGQW-TGELARCDDPGNYCPNPGTPIDGFKNGA-SYDIGDKVSYRCQPGYILLGSEVRECLSNKTWSGT-EAKCMAPYDFDNYAQMTEMLKFKVTEKFDDGNREQKQSHEAEISTFATL :  315
AmspFB-2 : GRSWTGKGKLPQCSLRSRFCQSAGNLKNGRKIGT-MQHIGAVVRYSCNPGYTLVGSTLQECKASRRWIPA-KPQCKSP--SELLQETSKRFNKQFVQRLAVDNSTFS------------ :  177
HaadFB-3 : RGNWT-SSKAPRCLRIQGQCDEPPVVKHSLRYGD-ERHVGAKVSYVCKDGFTLLGHSQLVCSRNGRWNKR-APTCMD--ESEPLQTVAERLKNSFVTELAYHSSGTP------------ :  173
ScsuFB-2 : KSGQW-DDPKPNCSKKHQHCLPPSDIDHGFLVGDPPYPPGSEVYYRCNEGYKIVGPSKLICKYPQYFWNDNPPECIK--IKPPLQVVAENIGDNLVDKLTKLDVG-------------- :  170
HosaFB   : VNGRW-SGQTAICDNGAGYCSNPGIPIGTRKVGS-QYRLEDSVTYHCSRGLTLRGSQRRTCQEGGSWSGT-EPSCQDSFMYDTPQEVAEAFLSSLTETIEGVDAEDGH----------- :  251

                                                         / vWFA
AmspFB-1 : ---------------------------------AQGRTLNINNPGGLN-LIFIFDSSGSVGKRGFETAKKFASTLVQHIGVGAN------GVRVAAMTFSS--DVTVNFYTREFLTTEEV :  586
TatrFB-2 : D--------------------------------RGRMIDLNFPGRLV-IYFVFDASGSVGEKYFYSAIKFAKALVKRMGVKEQ------GTRFGAVSFSS--TVSNSFLPQDYKTVEEV :  572
HaadFB-1 : ---------------------------WHNNSGPVGRILDINFPGRLI-LYFAFDVSGSVGQHNFDKSIEFAKAIVKRVGISEA------GARAGALIFGS--KSENMFLPLSYTTTEEV :  589
ScsuFB-1 : ---------------------------------VFGRSTISSDHATGVTIFFLLDASGSITKTEFKKSKELAIHVVRQIGISTHK----GGVRISVINFSQNVETVVTWAVESVEQAITR :  635
TatrFB-1 : NFGMSTNKVSKMLANKAFEMSSRRIFQMSPRIEPRRRRIDLNFPGRLV-IYFVFDASGSIGRKYFNSAIKFAKGLVTRMGVKEF------GTRFGAVSFSS--TVSASFLPQDYTTEEEV :  489
HaadFB-2 : ILGDKIEPDYQMMQHGLMNKHAFQKPSEDIDSISSFRTIDVNNQRRLI-LHFAFDVSGSVGSYYLRKSIEFAKAIVRKIGISRD------GSRASAVVFSS--KAETVFYPMQIETEEEV :  426
AmspFB-2 : ---------------------------------SGSRVLGPEVNKIGLDLAFLIDRSSSIDPVDFKIGINFLKELVDEFGVKNGDNKLKGGTRIAVITFGDKAEIVFNFDDARISSPEVA :  264
HaadFB-3 : ---------------------------------EGRLLDRSSLLLGLELYLLIDRSSSIDPVHLEDAKNFVKFLLRRFGVNNKPNN-NNGTRASVLAFGTEVQIVFNIDDTNISNPRIA :  258
ScsuFB-2 : ------------------------------------RVIHGDVEYLGLDLFLAFDKSNSISPQQFLEGIKFAKFLIKQFNVSNSDQKKVGGTRLAVYTFGNDAKEEINLTDTTITSTKAA :  254
HosaFB   : ----------------------------GPGEQQKRKIVLDPSGSM-NIYLVLDGSDSIGASNFTGAKKCLVNLIEKVASYGV------KPRYGLVTYATYPKIWVKVSEADSSNADWV :  335
                                            factor D           MG2+ binding
                                            cleavage site      sites

AmspFB-1 : VEQIGL----IQYNPGDTATNPALITATTEVIPEAARA---RPLSSSAVFLITDGRANVG--GRPKEAADRLIQ------------EFDVEIYAIGVGSNILEDELASIAYSKDGD--TD :  683
TatrFB-2 : HSALDK----FNFTEGGTAISLALDYVKTQMIPLSKQT-FQDQAMKSIIFLFTDGKANMR--GDPKQVAKELRE------------EAHAEIYCIALTGDYDIGELKDMASSVKE----- :  668
HaadFB-1 : LDALDK----INYTGGGTAASSALSLIRQENIPLIDGV-LGKKNIKSIIFILTDGKANMG--GSP-EVEADLLK------------KAGVEIYCIGITGSIEKESLYKIASTSKDSNGEH :  689
ScsuFB-1 : IDSIEK------RKDEGTNIAKALNHLANEVASTKEND--LVRDNKNIAFIISDGNANEG--GKPEKEAKLLKKFG-------------VKMFAIAVG-KKRDRVILKIITNDENN---- :  727
TatrFB-1 : LNGLDQ----FDFTEGGTAISSALDFVKTQMIPLSKHT-FADRPMKTIIFLLTDGKANMR--GDPKQVAKELKA------------DVKAEIYSIALTGDYDINKLREVASSKKD----- :  585
HaadFB-2 : LKYLDN----ITFTGGGTSASSAFALIKENIELEEESYYFDRNNKEDVVFFITDGKINMG--GNP-EMEAKILK------------ERNIDIYCIGITGDPRLETLYKIASTSKYGNVER :  527
AmspFB-2 : KRKLDM----IQAKGGSTNLNGALDKVIIRVKPRREVKR--------ALFIMSDGKPNTATRSTPENAARHLKTR-----------SLNYEIFTVGIGRAVNMTLLRHMASDPKIN---- :  357
HaadFB-3 : AAAVDD----IHPNGGGTNMEGALTKVLIDFQKLRKRA-------KRALFLMTDGEPNIA--NPEITPQDIAQQLKKS--------PNDFEIFTVGIGKGIKMNLLNELASEPPLS---- :  353
ScsuFB-2 : INKLDLIR-CSKFCDGATNMADALKKIGHIAPKQTRKEA------KKVLFMTSDGVPTAD--PKSKDVTYYTNNLK----------KLGFEIYTVGIGQDIDEQLLKDLSSTPIEE---- :  351
HosaFB   : TKQLNEINYEDHKLKSGTNTKKALQAVYSMMSWPDDVPPEGWNRTRHVIILMTDGLHNMG--GDPITVIDEIRDLLYIGKDRKNPREDYLDVYVFGVGPLVNQVNINALASKKDNE---- :  449
                         MG2+ binding                       MG2+ binding
                         site                               site
                                                /                            / SP
AmspFB-1 : RHYMKVESFAKMNEMLQLLINGTIDYS-----ACGLLQRRKNTEKKELGVEKAEPGARNKRSLRIRIVGGETVTKHWPWMVGLYYG--SPLLGHKSLQCGGSLIAPNWILTAAHCIKLQI :  796
TatrFB-2 : -HVYILKNYETFDWLVNAVINGTVDYS-----VCGYGMDDVLEEMNK-------------------AGEQRADKPWPWMAAVYYQ----RENFEKLGCGGSIVNKEWILTAAHCFVLKN :  758
HaadFB-1 : PNVFILQNYATMSWLVQEITNGTVDYA-----QCGLGMENVGNEAARG-----------------RILNGKKSMEPWPWMAALYMPHDKLNPLDTELQCGGSIINNYFILTAAHCMYHRE :  787
ScsuFB-1 : --IFEFDSYQKLFDVISRTIGNQTNFQ-----LCGINQPNLIKRNKINLNKNRRKRFIDGIG---RIVGGQEAEYSWPWMAALYLKNKN---NKWEFKCGGSIIHSDWILTAAHCFFVVD :  834
TatrFB-1 : -HVYILKDYETLDWLVNAVVNGTIDYS-----ICGYGMDEVVEEVNK-------------------PGGTRAEKPWPWMAAVYYR----LKENERFRCGGSIVDREWILTAAHCVQNKD :  675
HaadFB-2 : SNVFILQNYATLSHLIQEITNGTLDFS-----ECGLGLENID-VGNQEVT-----------------QNEENRNDPWPWMAALFF---LISKDTYQIQCGGTIIKENFILTAAHCMFRKD :  621
AmspFB-2 : -HNFYFDKFSDFERILWLIKNKATPAP-----PPGFDKCGYIPP----------------------------GSYRWPWLVSVWSEIPVED-RFTLKQCTGTIICRKWILTSAHCLHYSD :  442
HaadFB-3 : -HVFILENYPDLNEVMKIIEDSKPPPPPISKDQCGYNVSK---------------------------------ENRPWLATLYIG------GAPFKMCSGVLICNQWVLTAASCLQDAQ :  432
ScsuFB-2 : -HMFLLEKFKDFAEIMDLIRNGTTEPPPPLPEQCGYIAENFYKTRNLETG--------------------LAKLGNWPWLAAIIVKDYTS--GGSRFACSGILICEEWVLTTAQCVTDDL :  448
HosaFB   : QHVFKVKDMENLEDVFYQMIDESQSLS-----LCGMVWEHRKGT----------------------------DYHKQPWQAKISVIRPS----KGHESCMGAVVSEYFVLTAAHCFT-VD :  531
                                                                                                                         protease
                                                                                                                         active site
```

Supplementary Fig. S13

```
AmspFB-1 : SDDKIVT-----YTTRNVRIHLGITNIKNPDNNNLIIATPQEFVLHPQYDP-----ATINNDVALIRLQEPVTYNPFIRPVCLPPALDKLPKNSLLYKTGETAIVTGWGHTKERKKHESL :  906
TatrFB-2 : DDPKKVE----YLVPANVTVKLGLLNVRN--SSDLKEFEVTDIRLHEKFN-----YTTYDHDIALLKLGRPITYERFIRPVCLPP--AVIPENSTLYQAGQSAFVTGWGHDKRVE----L :  861
HaadFB-1 : GKK---------RSKKDIIVKLGLTDVKN--ETYVQESEVSEMFIHPDYRPAGS----YDYDIALLLLDKPIEYNPFVRPICLPP--TELPENTPLYSSDEFGWATGWGHEGVVS----A :  886
ScsuFB-1 : GFLDIDEEKSEEIQSAKWKASVGRLNISDENSGLFYKISTIKIHQEYDSKS-------YGNDIALLKVDKSIVFDIYTSPICLPPSMNDIPEKSEMYSSKSEAYVVGWGNTEVADLSSSQ :  947
TatrFB-1 : PQRKKNQ----NLVPADIIVKLGVLNVLN--SSDLEEFEVAEIHRNENYN-----FTTYDHDIALLKLDRPVTYKPFVRPICLPP--FNVPENSILYKPGQSAFATGWGYDQRVA----I :  778
HaadFB-2 : KRRQILH----LRPKEEIIVKIGLFSLND--ESKVQEFGVQRIFIHEKYDPGNYSVQLYDYDIAILELDGSIVYDRRIRPICLPP--RDLAENSHLYSFKEFGWAGGWGHNGVLIPYNPY :  733
AmspFB-2 : RRDYKP--------VSKVFVTVGQKDILK-WSDSSENFIASRIIIHERYNNET-----LENDIALIELNDVTALGDKVKIACLPR-----PEIQLSRRSSQNSFTAGWGSTNDK-----N :  538
HaadFB-3 : SRVD----------MKDVFVVLGERHLLKAEQGQTN-FYVTDVQIHQNYNPKQ---NSIENNLAVLKLN--LPASRY-RPACLPP---TDRAIPLHLNLKINASITGWGRTSAS------ :  526
ScsuFB-2 : NIQYE----------PENVYVVVGEDNFEKLNDKEQLFYAVKIIRHPNFIHNK---TKVRNDIALIQLNTKAEINDYVRTACILQ-----FDFDHSSKPGEVGYMIGWLISSEEQVR--- :  547
HosaFB   : DKE------------HSIKVSVGGEK----RDLEIEVVLFHPNYNINGKKEAG-IPEFYDYDVALIKLKNKLKYGQTIRPICLP---CTEGTTRALRLP-PTTTCQQQKEELLP-AQDIK :  629
                                                                     protease
                                                                     active site


AmspFB-1 : GPDDLKQS----------AILKKLSVPIQEGKVCNDS-IQGEFRKGIYTDSMLCAGMGKK--GQDACATDSGGPLHQRLEDSEG--DGTYYTQIGIVSWGYG---CA--------LEGEY : 1000
TatrFB-2 : GHEGVLKG---------IDHLKQIRLPIQNHETCLKS-LEKTKKE--MTDFMICAGDSEG--IVDTCKGDSGGPMAQSLVDDA---EMNYWVQVGIVSWGIG---CK--------LRGHY :  953
HaadFB-1 : AVNERLKS---------SQILKELLLPIQSKQRCTQSLLDNKVATDHFTDRMFCAGDGKR--GNDTCKGDSGGPLMQSQLNSE---GYLFWTQVGIVSWGIG---CG--------KENTY :  981
ScsuFB-1 : KQKNSDIL------LQLKLPLHDDKICKQKVNFKASQAGKEITHYLTPSPNSLCAGTGEGK--QDACQGDSGGPLMQLIHSNSSSNIMKRWFEIGIVSWGIG---CG--------LEGYY : 1048
TatrFB-1 : DEIVPFKR---------VDQLKQIHLPIQSRETCVQS-LENTKDP--MTDVMICAGDGRG--VADTCQGDSGGPLAQSLLDES---GMNYWIQVGIISWGRG---CK--------NRGQY :  870
HaadFB-2 : SLSMQLNN---------SPNLKQLRIPIQSNERCSKS-LEKKLDPTFFTDRMFCAGTGET--GVDTCLGDSGGPLMQSQMNSE---GFLYWTQVGIVSWGIG---CG--------MEDMY :  827
AmspFB-2 : YDSAPVKS----------AFFSLNEVALTLVPDRDCARRAKSLQCRQPSASHFCAETSNDL--IKACSTGSGSPVIMSDSSSYDN-----YERLRVVGISSSHCQCG--------KQRGY :  633
HaadFB-3 : ----------------KMTYATANDIAFDMSSSSVSLSEERECPIRRRVTTPLCAGQGSK-----TCFALVGSPLMAEDSSTG---------FQHILGILVDRRHCS---------KKGQ :  607
ScsuFB-2 : ---------------IPQEPVKVFQAEQVEMKVNGLDNCEIIKPPQYLCDDSIINAHGRYE-SDKTCSGYDKGSPFLMSQGKR-------MAAVGLASHVKG---CT--------LQSKT :  633
HosaFB   : ALFVSEEEKKLTRKEVYIKNGDK-KGSCERDAQYAPGYDKVKDISEVVTPRFLCTGGVSPYADPNTCRGDSGGPLIVHKRSR--------FIQVGVISWGVVD-VCKNQKRQKQVPAHAR :  739
                                                                              protease
                                                                              active site

                         /
AmspFB-1 : GFYTRLTKFIPWIEGVTNVKFSN------ : 1023
TatrFB-2 : GFYTHVAKLRPWIDKVLNS---------- :  972
HaadFB-1 : GYYTHVQKFRSWIDSTIEAAMAAIQ---- : 1006
ScsuFB-1 : GYYTHVSKLKQWILNETNSEFEIVALNV- : 1076
TatrFB-1 : GFYTHVAKLRPWIDKVMNS---------- :  889
HaadFB-2 : GFYTHVQKLRSWIDEIIYTQ--------- :  847
AmspFB-2 : HMFTRVHDYIKWIASRTDFCIAEHI---- :  658
HaadFB-3 : NQYIELTRHIGWINQATSNCQLKHWGVAK :  636
ScsuFB-2 : GFFTRISSYYDWIRENSEFCSDNHQ---- :  658
HosaFB   : DFHINLFQVLPWLKEKLQD---EDLGFL- :  764
```

Supplementary Fig. S13 (continued)