

# Protein coevolution and physicochemical adaptation in the APAF-1/apoptosome: structural and functional implications

José A Sánchez-Borbón<sup>1,2\*</sup>, Steven E Massey<sup>3</sup>, José D Hernández-Martich<sup>1,2</sup>

<sup>1</sup>Universidad Autónoma de Santo Domingo, Santo Domingo, Dominican Republic; <sup>2</sup>Instituto de Investigaciones Botánicas y Zoológicas, Universidad Autónoma de Santo Domingo, Santo Domingo Dominican Republic; <sup>3</sup>Department of Biology, University of Puerto Rico, San Juan, PR, United States

## Abstract

The apoptosome is involved in the mitochondrial pathway of apoptosis, consisting of APAF-1, caspase 9 and cytochrome c, forming a heptamer that activates effector caspases, causing cell death. This protein complex has also been characterized in *Drosophila melanogaster* (DARK) and *Caenorhabditis elegans* (CED-4). Here we present an evolutionary guided *in silico* characterization of the APAF-1/apoptosome. The evolutionary history of the apoptosome was determined, taking the possible orthologs of the APAF-1 version, and executing a protein coevolution and a positive selection analysis, to make structural and functional inferences and identify residues under destabilizing changes, respectively. Results suggests that the APAF-1/apoptosome is not unique to vertebrates, but also some basal invertebrates could possess orthologous copies. New possible versions were also detected in other taxa. Not all insects and other arthropods have the DARK version, just as not all nematodes have the CED-4 version. In the APAF-1 version, amino acid clusters with coevolution signal located in the interior, gave more insights on new potential interactions, allowing us to infer a more detailed model that includes allosterism, of how cytochrome c associates with  $\beta$  propellers during APAF-1 activation, as well as interactions essential for nucleotide exchange, activation of CASP9, the molecular timer and other pathways in mitochondria to induce apoptosis. Residues on the surface under destabilizing changes have guided the protein complex in adaptations necessary for conformational changes, interactions and folding.

## Introduction

Under certain physiological and pathological conditions, multicellular organisms must eliminate unnecessary or undesirable cells through regulated mechanisms, to maintain homeostasis. Regulated cell death is the product of the activation of signal transduction, which can be pharmacologically or genetically modulated (Galluzi et al., 2018). Among its main mechanisms is apoptosis (Kerr et al., 1972). One of the main pathways of apoptosis is the intrinsic or mitochondrial, which can be initiated by perturbations of both the extracellular and intracellular environments (Galluzi et al. 2018). These perturbations cause the permeabilization of the outer

membrane of the mitochondria due to the formation of BAX / BAK pores, releasing pro-apoptotic proteins such as cytochrome c (CYCS). Once released into the cytoplasm, CYCS associates with APAF-1, which associates with CASP9, forming a heptamer called apoptosome. Through the apoptosome, CASP9 activates CASP3, which in turn activates caspases 6 and 7, causing cell death (Elmore, 2007). The protein complex has also been characterized in the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, which have the apoptosomes made up of DARK and CED-4, respectively. Despite being homologous, they have differences in their structure and functional dynamics. The *Drosophila* apoptosome is an octamer of DARK (Yu et al., 2006). Despite having the regulatory region, CYCS is not necessary for the activation of the protein, and there is another, as yet unknown, ligand that associates with  $\beta$  propellers (Dorstyn et al., 2002). In *C. elegans* the protein complex is formed by an octamer of CED-4 (Qi et al., 2010). CYCS is also not involved in its activation, due to the absence of  $\beta$  propellers in it (Lettre and Hengartner, 2006).

One of the phenomena by which proteins evolve is coevolution. This process occurs when a substitution of an amino acid located in a specific site, generates a compensatory substitution of another amino acid located in another site (Altschuh et al., 1987; Altschuh et al., 1988; Taylor and Hatrick, 1994). These series of mutations can occur in speciation events. There are various evolutionary forces or constraints that influence how these compensatory changes occur. One of them is the structural restriction due to orthosteric interactions between residues (Göbel et al., 1994; Shindyalov et al., 1994). The other type of restriction is functional, in which the amino acids under coevolution may be in distant positions in the protein, this due to their binding to the same ligand or due to allosteric interactions (Süel et al., 2003; Anishchenko et al., 2017). The amino acids under coevolution can be located in the same protein or in different proteins: intraprotein and interprotein coevolution, respectively (Pazos et al., 1997).

Changes in amino acids are caused by mutations in their codons. In synonymous mutations, a nucleotide substitution generates a different codon, but it still codes for the same amino acid. In contrast, in the non-synonymous mutations, the new codon produced codes for an amino acid different from the original one (Pál et al., 2006). When non-synonymous changes produce a beneficial genetic variant, it quickly spreads through the population. This is known as positive selection (Pál et al., 2006). In some cases of positive selection, the substitutions can code for amino acids with physicochemical properties very different from the original, being "selected" by drastic evolutionary pressures, and producing radical changes in the structure and function of that protein region. All this contributes to the new protein being able to adapt physicochemically to new cellular microenvironments. McClellan et al. (2005) call this phenomenon "destabilizing positive selection".

In this research we carried out an *in silico* characterization of the APAF-1/apoptosome, through evolutionary analysis of its sequences and three-dimensional structures. Focusing on the sequences homologous to humans, the evolutionary history of the apoptosome in Metazoa was determined taking as reference the protein sequences of the apoptosomes formed by APAF-1, DARK and CED-4, and identifying their putative orthologs in other taxa. From these sequences, the possible orthologues of the APAF-1/apoptosome were taken, identifying amino acids with coevolution signal and those under destabilizing positive selection. We use this information to make structural and functional inferences about the protein complex.

## Results

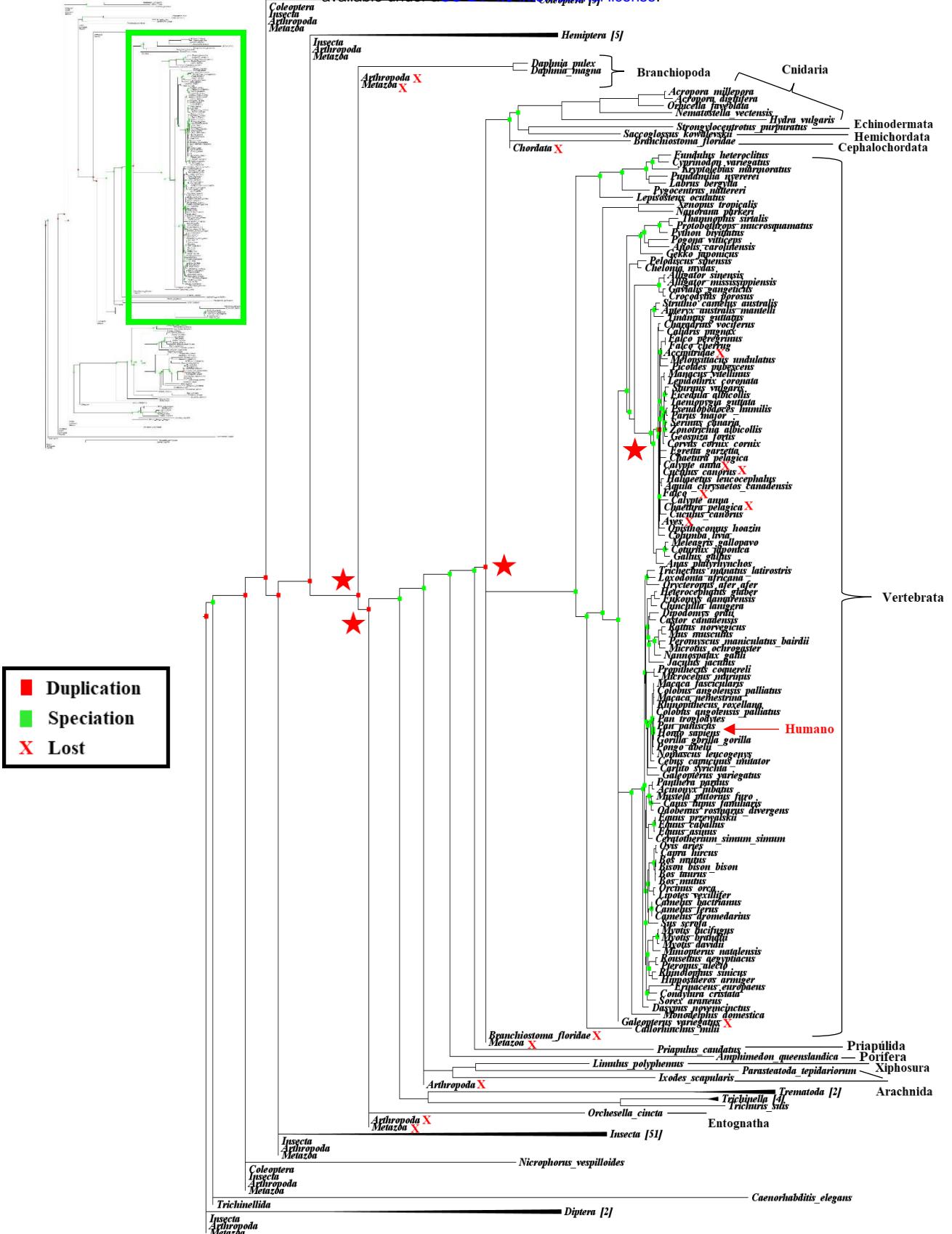
Homologous sequences similar to APAF-1/apoptosome in Metazoa were searched in GenBank (Supplementary file 1), detecting putative orthologs to human sequences. For this, similarity analyses were made through the Ortho-Profile method (Szklarczyk et al., 2012), in which the human sequences against invertebrates were subjected to a series of BLASTs and HMMs in three stages with increased sensitivity, sequence against sequence, profile against sequence and HMM against HMM. If in one of the phases, the sequences show reciprocity with e-values less than 0.01, they are considered orthologs. Orthology was also detected through reconciliation of phylogenies, also using as reference the DARK (*D. melanogaster*) and CED-4 (*C. elegans*) apoptosomes.

### Proteins orthologous to human sequences identified by similarity

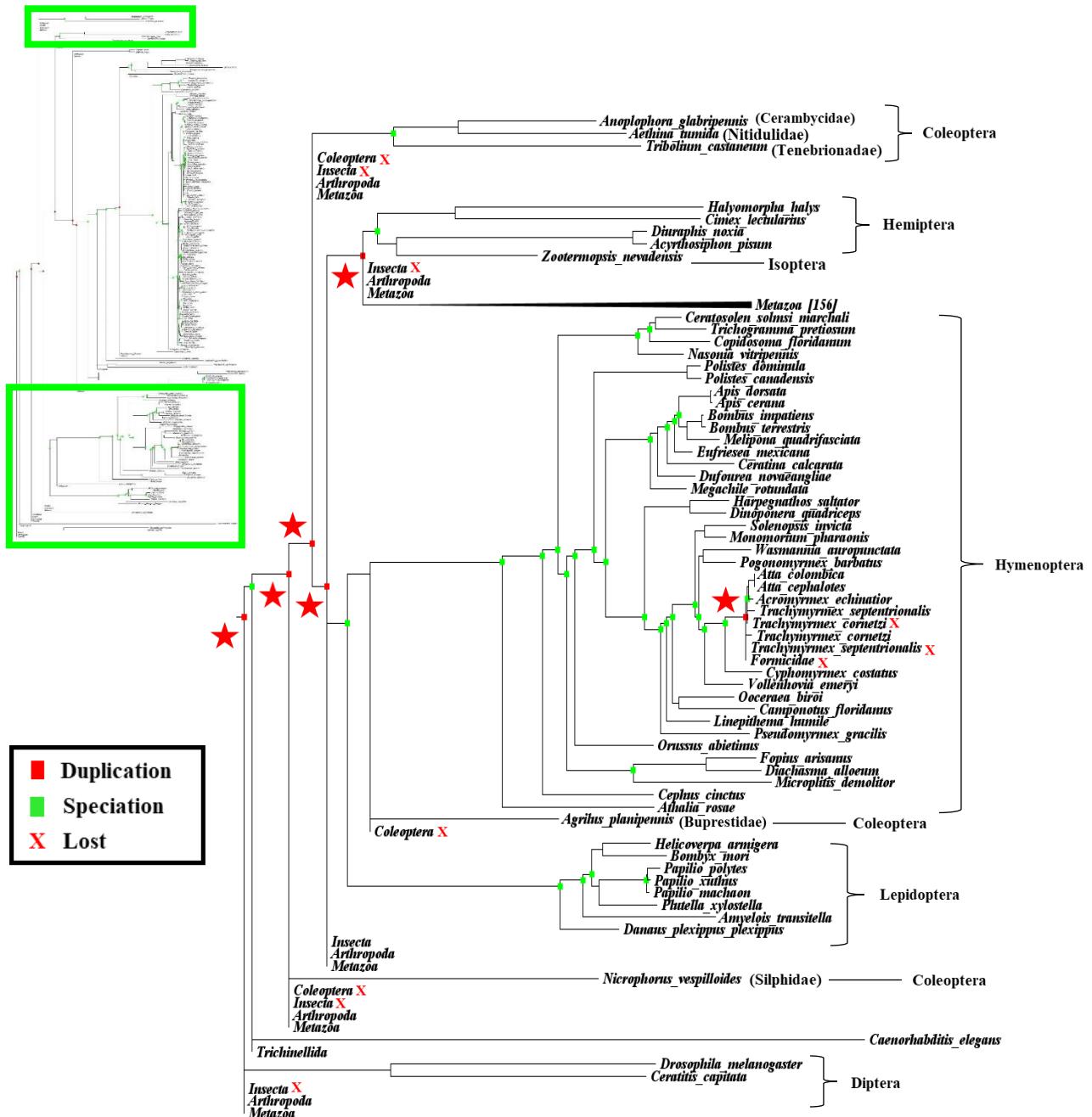
Of the homologous proteins in invertebrates, only 20 passed the test (Supplementary file 2). In the first phase, six proteins were detected; three in the cnidarians *Acropora millepora*, *Hydra vulgaris* and *Orbicella faveolata*; the echinoderm *Strongylocentrotus purpuratus*; the hemichord *Saccoglossus kowalevskii* and the cephalochord *Branchiostoma floridae*. In the second phase, 14 proteins were identified; one in the poriferous *Amphimedon queenslandica* and the other 13 within Protostomia; three in the nematodes *Trichinella nativa*, *T. papuae* and *T. zimbabwensis*; 10 in arthropods distributed in two branchiopods, *Daphnia magna* and *D. pulex*; the arachnid *Parasteatoda tepidariorum*; the entognatha *Orchesella cincta* and six insects, *Amyelois transitella*, *Athalia rosae*, *Neodiprion lecontei*, *Ooceraea biroi*, *Trachymyrmex cornetzi* and *Trichogramma pretiosum*. In the third phase, for CASP9, in *Lingula anatina* and *S. purpuratus*, possible orthologs were detected, but they did not coincide with the homologous proteins.

### Proteins orthologous to human sequences identified by reconciliation of phylogenies

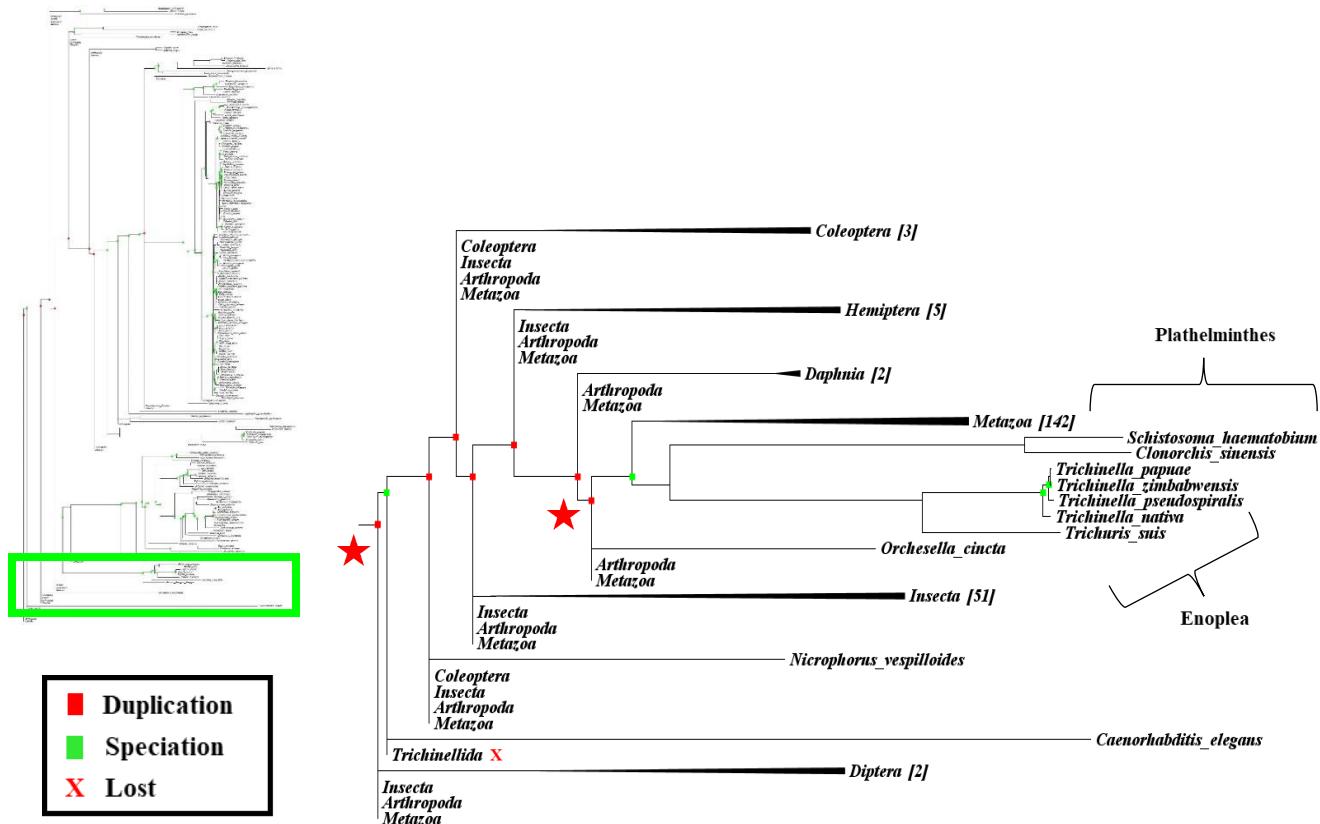
The APAF-1 phylogeny reconciled with its species tree shows 10 duplications. Close to *H. sapiens* there are four duplication events (Figure 1). One that gives rise to two speciation events; to entognatha and another clade that includes Porífera, Priapúlida, Xiphosura, Plathelminthes, Enoplea and Arachnida. Branchiopoda arises from another duplication. Upon reaching deuterostomes, another duplication occurs with two speciation events. One goes to Cnidaria, Echinodermata, Hemichordata and Cephalochordata and the other to Vertebrata. Taking into account that in the unreconciled phylogeny, these phyla are found in the same clade (Figure 1-figure supplement 1), we interpret that the sequences involved in this duplication event would be co-orthologous. In birds there is a duplication where Falconiformes, Passeriformes and Apodiformes diverge. In insects there are six duplication events (Figure 2). One of them gave rise to Isoptera and Hemiptera. Another duplication gives rise to the coleopteran families Cerambycidae, Nitidulidae and Tenebrionidae. Within coleoptera another duplication originates the family Silphidae. A duplication is in the node that originates Hymenoptera and Lepidoptera. Within Hymenoptera, in *Trachymirmex* there is a duplication in between *T. cornetzi* and *T. septentrionalis*. The clade of Hymenoptera and Lepidoptera also includes the beetle *Nicrophorus vespilloides* of the family Buprestidae. Diptera originates from a unique duplication event. In relation to Nematoda (Figure 3), Enoplea and Plathelminthes arise from the same duplication event. *C. elegans* originates from a unique duplication. The analysis of conserved domains in Pfam in the nematodes *Trichuris suis*, *T. nativa* and *T. papuae* and Plathelminthes weakly detected WD40 repeats that form the  $\beta$  propellers (Figure 3-figure supplement 1), so we decided to do more detailed analysis of conserved domains in the NCBI, confirming that unlike *C. elegans*, these species have the WD40 repeats (Figure 3-figure supplement 2).



**Figure 1.** Emphasis on regions close to the human sequence of the APAF-1 phylogeny reconciled with its species tree. The stars indicate the four nodes that show duplication in the evolution of the protein during speciation events. The other nodes are collapsed. See also Figure 1-figure supplement 1.



**Figure 2.** Emphasis on insects of the APAF-1 phylogeny reconciled with its species tree. The stars indicate the six nodes that show duplication in the evolution of the protein during speciation events in insects. The nodes of the other taxa are collapsed. See also Figure 2-figure supplement 1.



**Figure 3.** Emphasis on nematodes of the APAF-1 phylogeny reconciled with its species tree. The stars highlight the two nodes that show duplication in protein evolution during speciation events in Nematoda. The Enoplea clade also includes Plathelminthes. The nodes of the other taxa are collapsed. See also Figure 3-figures supplements 1 and 2.

The phylogeny of CASP9 reconciled with its species tree presents six duplications (Figure 4). *H. sapiens* is included in a duplication event at the node where deuterostomes originate with a speciation event encompassing echinoderms, *B. floridae*, and other vertebrates. In the latter there are three duplication events; one at the node where reptiles and birds emerge with two speciation events, one towards Squamata and the other towards Crocodylia, Testudines and Aves. A duplication occurs at the node where the Hominidae family originates with two speciation events, one going towards *Pongo abelii*, *H. sapiens*, *Pan Paniscus* and the other towards *P. troglodytes*. The other duplication event within vertebrates is at the node common to Eulipotyphla, separating *Sorex araneus* from the rest of the order. The clade of the deuterostomes also includes Porifera, the arachnid *P. tepidariorum* and the mollusk *Octopus bimaculoides*. *Branchiostoma belcheri* is in a unique speciation event. The *Aplysia californica* mollusk and the lingulate *L. anatina* are in the same speciation event. At the node where Diptera arises there is a duplication with two branches of speciation. One goes to *D. melanogaster* and the other covers the other Diptera *D. kikkawai*, *D. takahashii* and *Bactrocera oleae*. Lepidoptera arises from a unique duplication event far removed from other insects. Plathelminthes are in a unique speciation event.

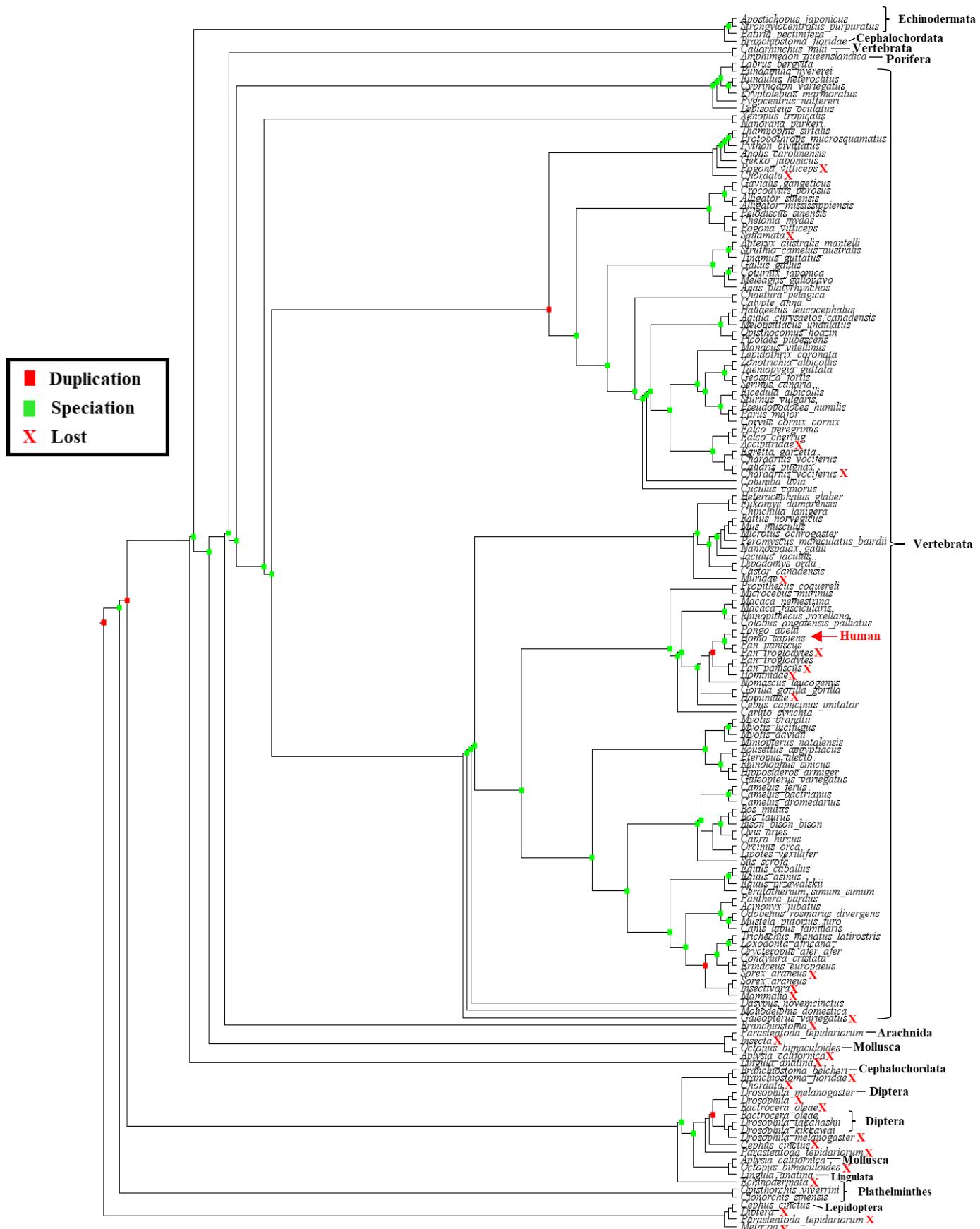
The automatically generated species trees in the NCBI taxonomy database are polytomous. Another aspect to take into account is that these trees are cladograms, not phylogenies as such, so the branches do not have the real evolutionary distances. Taking the previous statements as caveats, they could have generated the following possible errors in the reconciliations; the duplication events in both proteins within vertebrates. As a support of this, APAF-1 has not been duplicated in vertebrates (Zmasek et al., 2007). The duplication event in APAF-1/DARK homologues found within Hymenoptera, and coleopterans outside their respective clade. The APAF-1 of Porifera and Priapulida should originate in the same duplication event as cnidarians, echinoderms, hemicordates, and cephalochords. CASP9 of *O. bimaculoides* and *P. tepidariorum* must be in a clade separate from that of vertebrates. If the phylogenies of species had been made by ourselves, the results would be more accurate. There are also the following possible errors that we believe are the result of false positives during the search for CASP9 homologs; the possible cause of *B. belcheri* being in a node other than *B. floridae* is that it is identified as caspase 7 in the database. The sequences of the Diptera, *D. kikkawai*, *D. takahashii* and *B. oleae* according to the analysis of conserved domains are not well characterized, so we suggest that in this case they may also be caspases other than CASP9/DRONC. This may be the reason why these sequences originate from a node other than *D. melanogaster*.

The results of the first phase of Ortho-Profile were very similar to the unreconciled phylogeny of APAF-1. In the second phase, only the APAF-1 of *A. queenslandica* matches the phylogeny (Supplementary file 2 and Figure 1-supplement figure 1). Considering the low precision of the second phase, we think that the third phase would have even less precision because it has more flexible restrictions, which suggests that they are mostly false positives.

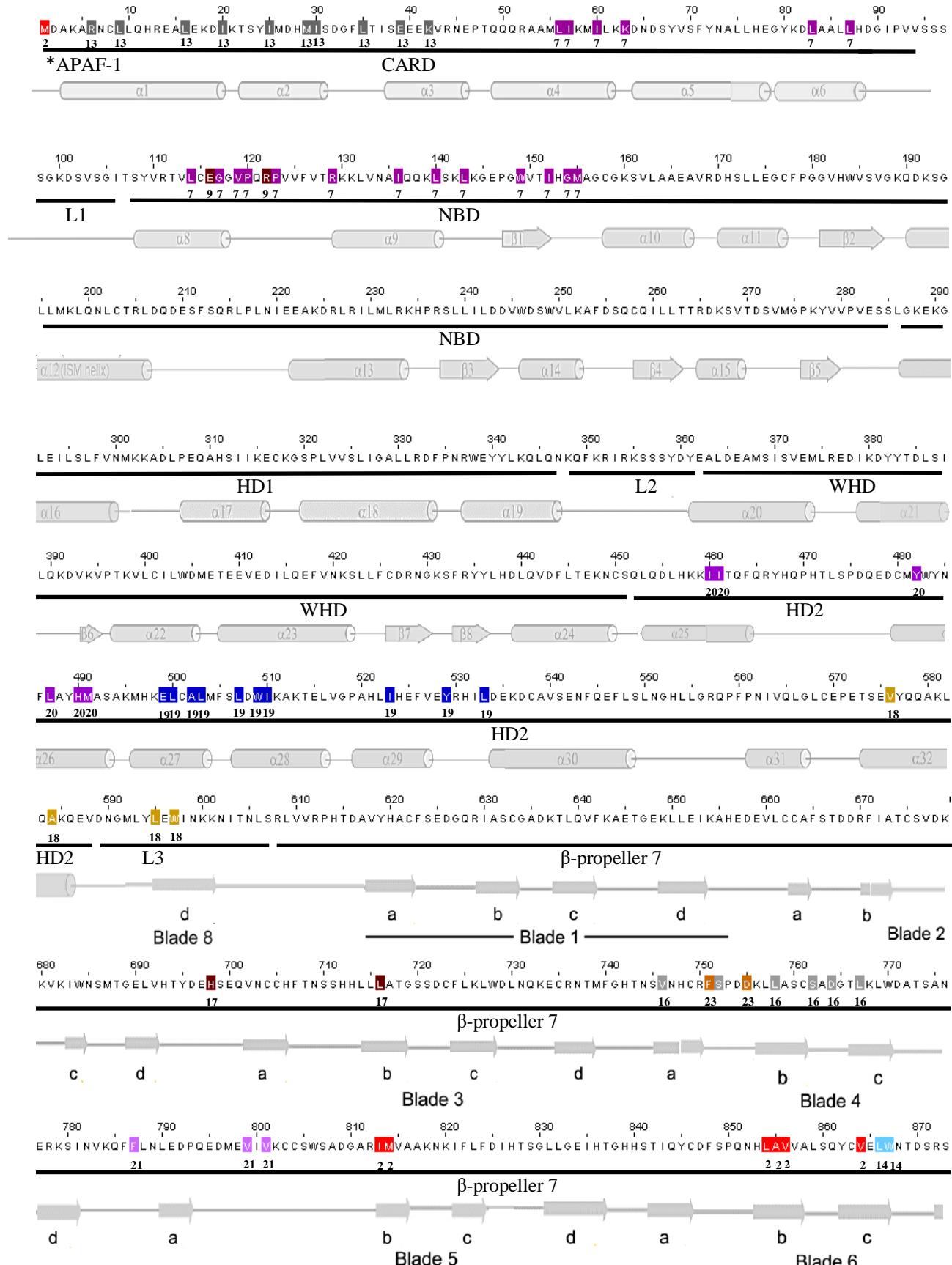
From the putative orthologs detected, we selected the species that have the two proteins available, to apply protein coevolution and destabilizing positive selection analyses. In invertebrates these were the poriferous *A. queenslandica*, the echinoderm *S. purpuratus* and the cephalochord *B. floridae*, plus the 115 species of vertebrates including the reference species *H. sapiens*, making a total of 119 species. For each of them, the CYCS sequences were obtained to complete the proteins that make up the APAF-1/apoptosome to include them in the analyses (Supplementary file 1).

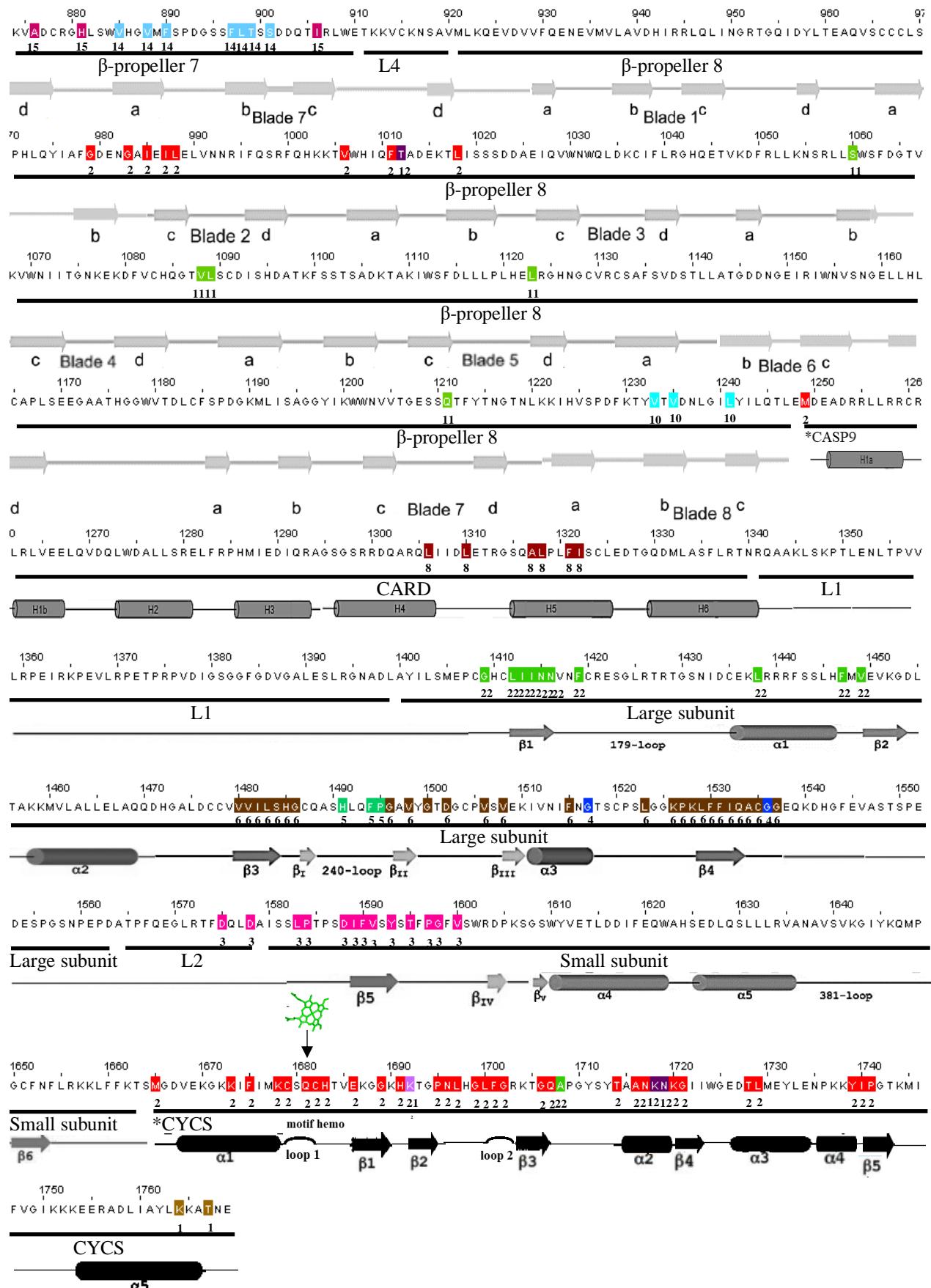
### **Amino acids with coevolution signal in the APAF-1/apoptosome**

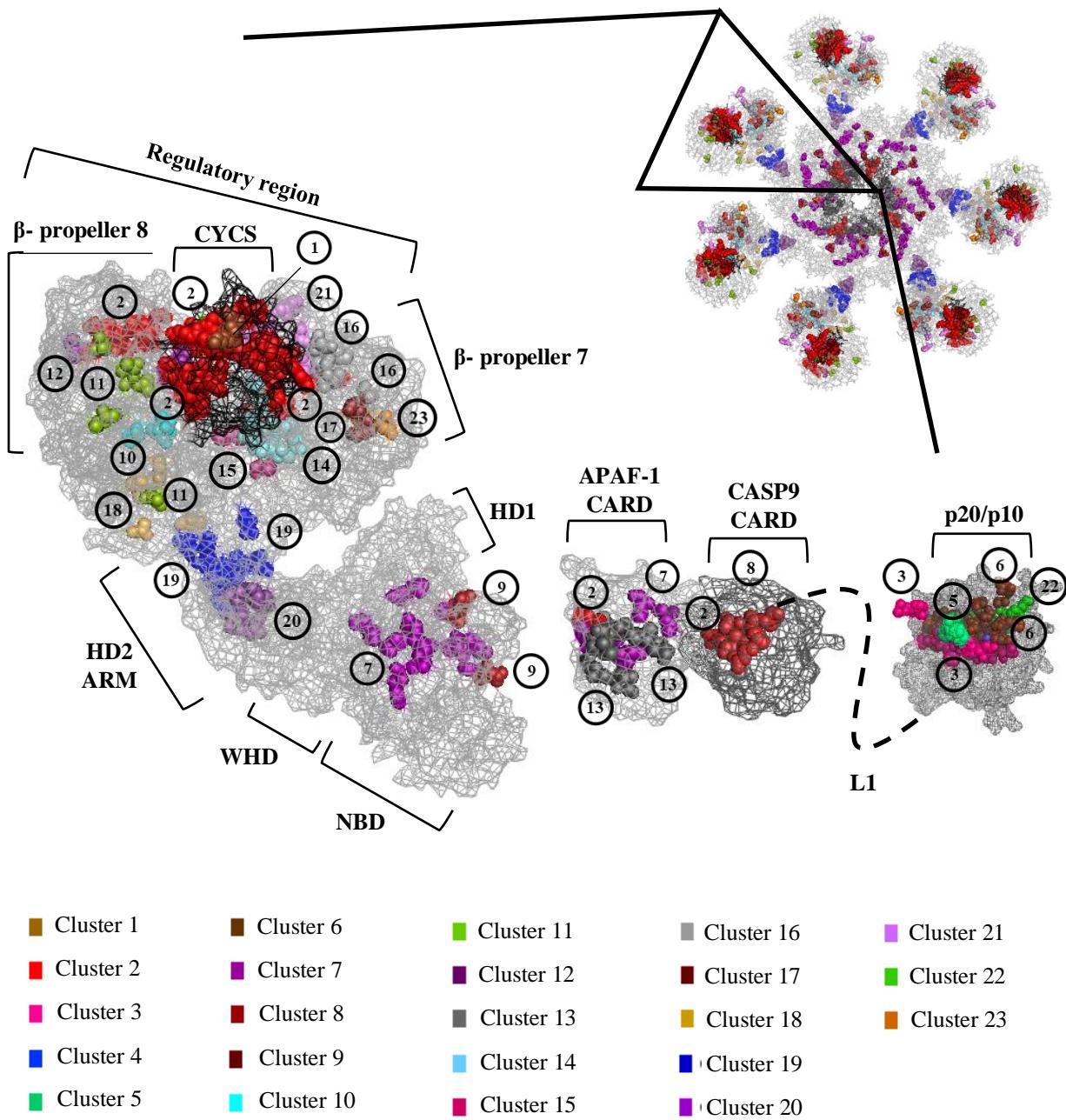
The concatenated alignment with the proteins of the selected species is 5,426 positions long including gaps; together with its respective phylogeny were uploaded to the BIS2Analyzer server (<http://www.lcqb.upmc.fr/BIS2Analyzer/>) (Oteri et al., 2017), which detects amino acids with a coevolution signal. The program detected 23 amino acid clusters distributed in the different protein domains (Figure 5). Clusters 21, 22, and 23 have symmetric and environmental values of 0.55, 0.83, 0.55, 0.65, and 0.73, 0.5, respectively. Clusters 1 to 20 have values 1, 1, with a perfect coevolution signal (Supplementary files 3 and 4 and Figure 5-table supplement 1).



**Figure 4.** CASP9 phylogeny reconciled with its species tree. The six nodes that show duplication in the evolution of the protein during speciation events in invertebrates and vertebrates are shown. See also Figure 4-figure supplement 1.







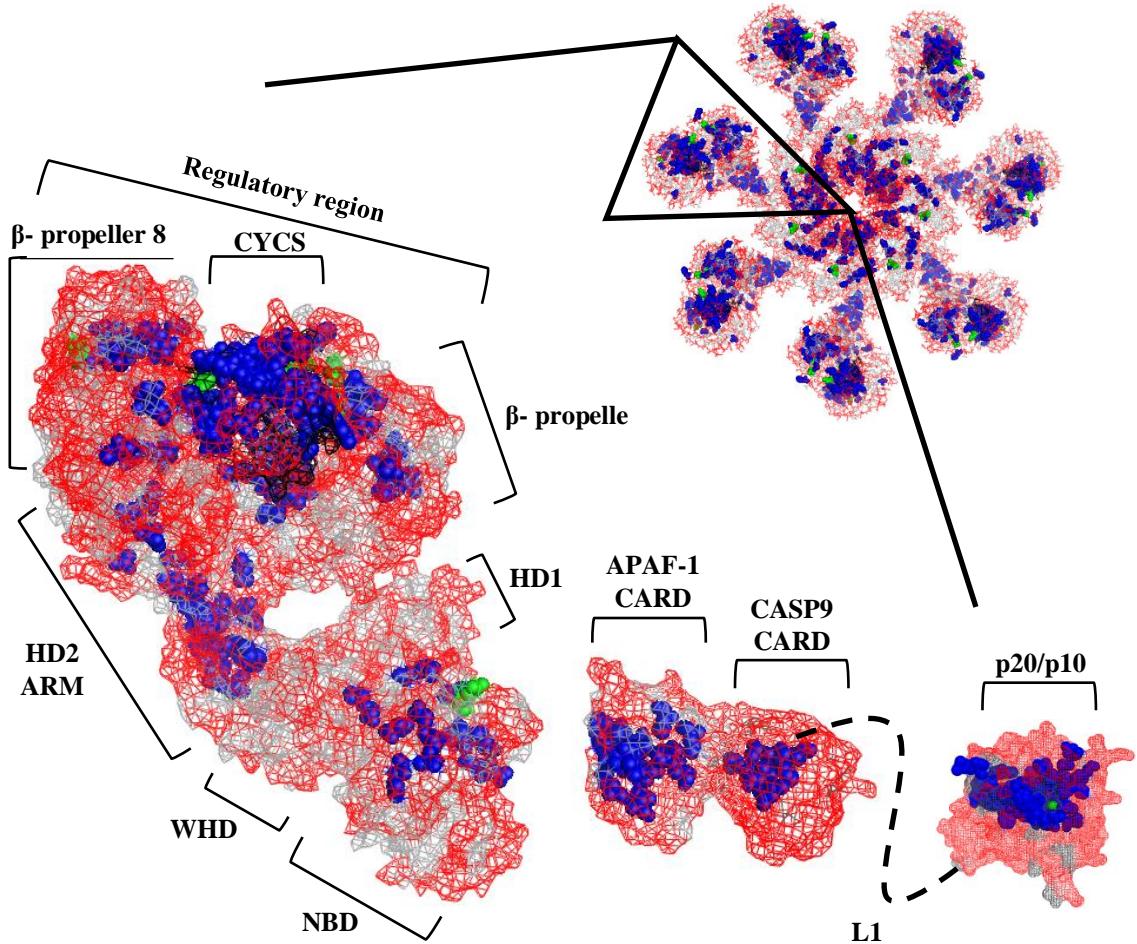
## Amino acids affected by destabilizing positive selection in the APAF-1/apoptosome

A concatenated alignment with the coding DNA of the components of the apoptosome in the selected species and their respective phylogeny was supplied to TreeSAAP (Woolley et al., 2003), to detect amino acids under positive selection. The program identified 3,929 codon substitutions during phylogeny. During these substitutions, taking as reference the 1,769 amino acids in the concatenated human sequences, 1,164 residues are under positive selection in all categories of magnitude (1-8), distributed in all protein domains. Of these, 996 residues are under destabilizing changes in the most radical categories 6, 7 and 8. In the three-dimensional structures, residues under coevolution are located inside the proteins, close to active sites and other regions involved in conformational changes, specifically in the regulatory region, the arm, the nucleotide exchange zone, the CARD domains of both proteins and the catalytic domain of CASP9, while those that are under destabilizing changes are found in the periphery (Figure 6). A few residues were detected in both analyzes; on the APAF-1, R/122, V/799, V/801, T/1012 and CYCS K/1692 and K/1718.

The destabilizing substitutions have significantly affected nine physicochemical properties in the following protein domains, taking into account regions under strong selection (protein regions with a large agglomeration of affected residues): helical contact area (*Ca*) (Richmond and Richards, 1978) (APAF-1 HD2,  $\beta$  propellers 7 and 8, CASP9 and CYCS), compressibility (*K0*) (Gromiha and Ponnuswamy, 1993) (APAF-1  $\beta$  propellers 7 and 8, CASP9 and CYCS), hydrophobicity by thermodynamic transfer (*Ht*) (Zimmerman et al., 1968) (APAF-1 NOD, HD2 and  $\beta$  propeller 7, CASP9 p20/p10 and CYCS), chromatographic index (*RF*) (Zimmerman et al., 1968), which affects the majority of residues in the protein complex (Table 1) (APAF-1 CARD, NOD, HD2,  $\beta$  propellers 7 and 8, CASP9 and CYCS), refractive index ( $\mu$ ) (Prabhakaran and Ponnuswamy, 1979) (APAF-1 WHD, HD2,  $\beta$  propellers 7 and 8, CASP9 p20/p10 and CYCS), molecular weight (*Mw*) (Fasman, 1976) (APAF-1 HD2 and  $\beta$  propeller 7, CASP9 CARD), bulkiness (*Bl*) (Zimmerman et al., 1968) (APAF-1  $\beta$  propeller 8 and CASP9), partial specific volume (*V0*) (Gromiha and Ponnuswamy, 1993) (APAF-1  $\beta$  propeller 7, CASP9 CARD) and molecular volume (*Mv*) (Grantam, 1974) (APAF-1  $\beta$  propeller 8, CASP9 and CYCS) (See sliding windows in Figure 6-figure supplement 1). The same amino acid can be involved in several categories of magnitude, tendencies and physicochemical properties at the same time, because homologues to that residue can have substitutions of different natures at different nodes (Supplementary file 5).

**Table 1. Number of destabilizing residues divided into positive and negative tendencies**

Helical contact area ( <i>Ca</i> )	(+) 382 (-) 423
Compressibility ( <i>K0</i> )	(+) 747 (-) 546
Thermodynamic transfer hydrophobicity ( <i>Ht</i> )	(+) 27 (-) 31
Chromatographic index ( <i>RF</i> )	(+) 809 (-) 747
Refractive index ( $\mu$ )	(+) 277 (-) 268
Molecular weight ( <i>Mw</i> )	(+) 160 (-) 157
Bulkiness ( <i>Bl</i> )	(+) 227 (-) 236
Partial specific volume ( <i>V0</i> )	(+) 325 (-) 348
Molecular volume ( <i>Mv</i> )	(+) 187 (-) 166



- Amino acids with coevolution signal
- Amino acids under destabilizing positive selection
- Amino acids with coevolution signal and destabilizing positive selection at the same time

**Figure 6. Amino acids under destabilizing positive selection mapped in a single unit of the APAF-1/apoptosome.** A single unit is shown within the PDB 5JUY heptamer in “mesh” format, which allows visualization of residues in the interior. The catalytic domain of CASP9 PDB 1JXQ was attached, representing L1 with intermittent lines. The residues under destabilizing positive selection in categories 6, 7 and 8 are shown as red sticks, whose z scores are + 3.09 and - 3.09. To facilitate the visualization of these amino acids, the 23 clusters under coevolution were colored blue. Green spheres indicate residues detected in both analysis. It can be seen that amino acids under coevolution are in the interior, close to active sites and other regions involved in conformational changes, while those that are under destabilizing positive selection are located in the periphery. See also Figure 6-figure supplement 1 and Supplementary file 5.

## Discussion

### Evolutionary history of the apoptosome in Metazoa

#### The APAF-1/apoptosome arises in Porifera and is present in other basal invertebrates

Zmasek et al. (2007), proposed that the common ancestor of Cnidaria - Bilateria had a protein similar to human APAF-1. Zmasek and Godzik (2013), proposed that the CARD domain arose at the base of Metazoa, putting *A. queenslandica* as a model species. This is confirmed by our results having the APAF-1 sequence of this poriferous the CARD domain. Similarity and phylogenetic analyses suggest that not only the CARD domain as such, but also the APAF-1 version of vertebrates was already present in Porifera (Figure 1-figure supplement 1 and Supplementary file 2), so this phylum would be the common ancestor at the base of Metazoa, where the APAF-1/apoptosome arose. CASP9 from *A. queenslandica* is also included within vertebrates (Figure 4 and Figure 4-figure supplement 1). This codivergence in the evolutionary history of both proteins would also be evidence that the interaction and therefore the coadaptation between them already existed since the emergence of Metazoa and that as a result of this, they have coevolved until reaching the vertebrates. Similarity analyzes and unreconciled phylogenies of APAF-1 in the phyla Cnidaria, Echinodermata and the subphylum Cephalochordata (Figure 1-figure supplement 1 and Supplementary file 2), coincide with those of other investigations in which genes and proteins have been detected in these phyla similar to APAF-1 of vertebrates, specifically in the cnidarians *N. vectensis* (Zmasek et al. 2007) and *A. acropora* (Moya et al. 2016), in the echinoderm *S. purpuratus* (Robertson et al. 2006; Zmasek et al. 2007) and in the cephalochord *B. floridae* (Zmasek et al. 2007). In the phylogeny generated by Zmasek et al. (2007), some of these species have several copies of the gene, each having exclusive copies of its group and a copy similar to APAF-1 of vertebrates. In *S. purpuratus*, all copies are found in the clade of vertebrates. Based on this; as in the reconciled phylogeny these phyla are in a separate speciation event from that of vertebrates, we suggest that the sequences in these basal invertebrates would be co-orthologous to the vertebrate proteins; apparently in this scenario with all these genes duplicated at the base of Metazoa, the only gene similar to APAF-1 in turn also began to duplicate and in the course of evolution when the gene system was simplified, vertebrates inherited one of those copies.

#### Several taxa have apoptosomes other than APAF-1, DARK, and CED-4

In the similarity analyses performed by Menze et al. (2010), the APAF-1 of the crustacean *D. pulex* was close to the human sequence, showing that not all species in the phylum Arthropoda must be similar to DARK of *D. melanogaster*. Our phylogenetic analyses agree with this result, in the sense that the proteins of the genus *Daphnia* are found in a node close to that of vertebrates, but not within the same clade. The origins of the genus *Daphnia* and the entognate *O. cincta* in unique duplication events confirms the existence of apoptosomes different from those formed by APAF-1, DARK and CED-4 in Arthropoda (Figure 1).

A caspase similar to human CASP9 was detected in the mollusc *M. provincialis* (Romero et al. 2011). Our results differ in this aspect, since molluscs have multiple origins distant from those of vertebrates (Figure 4), but we also did not rule out the possibility that in *M. provincialis* there is an apoptosome similar to that of vertebrates, which suggests that within Mollusca there are

different versions of the protein complex that are common to Brachiopoda and Chelicerata, so the common ancestor of Lophotrochozoa and Ecdyzosoa probably had all these versions.

### **Not all insects have the DARK/apoptosome**

There is a paradigm that the version of the DARK/apoptosome in *Drosophila* is common to all insects (Lee y Fairlie, 2012). From our phylogenetic results, we speculate that in the course of the evolution of Insecta, APAF-1/DARK duplicated several times, with the different orders acquiring different versions of the protein (Figure 2), which challenges this paradigm. CASP9/DRONC has a similar divergence, where *D. melanogaster* originates from a single node and is distant from the hymenopter *C. cinctus* (Figure 4), suggesting that the different versions of APAF-1/DARK and CASP9/DRONC within Insecta, by interacting and forming different apoptosomes in the different orders, have coevolved in the course of their evolutionary histories.

### **Not all nematodes have the CED-4/apoptosome**

The version of the protein complex of *C. elegans* has been generalized to the other nematodes (Oberst *et al.*, 2008). Our phylogenetic analyzes suggest that the version of APAF-1/CED-4 in nematodes of the genera *Trichinella* and *Trichuris* is different from that of *C. elegans* (Figure 3), which contradicts the mentioned paradigm. The only caspase present in *C. elegans* is CED-3, homologous to CASP3 (Yuan *et al.*, 1993). As APAF-1/CED-4 from Plathelminthes are possible orthologous to *Trichinella* and *Trichuris*, the detection of CASP9-like proteins in Plathelminthes suggests the presence of this protein, as well as other initiator caspases in other nematodes. In the apoptosome characterized in *C. elegans*, CED-4 lacks the WD40 domains that form the  $\beta$  propellers at the carboxyl terminal (Qi *et al.*, 2010). In the genera *Trichinella*, *Trichuris* and the Plathelminthes, the presence of the WD40 domains (Figure 3-figure supplement 2) suggests that, similar to APAF-1, the protein is regulated by CYCS or another ligand. Bender *et al.* (2012), demonstrated that in the Plathelminthe *Schmidtea mediterranea* apoptosis is regulated by MOMP, so it may be dependent on CYCS. They suggested that the mitochondrial pathway originated before vertebrates. They assumed that the mitochondrial pathway was lost in some lineages like Nematoda. In contrast to these authors, we suggest that the mitochondrial pathway as well as the WD40 repeats were not lost in most nematodes. Apparently, due to some evolutionary pressure, there was only loss and divergence in the genus *Caenorhabditis*.

Coevolution and destabilizing positive selection analyses were performed on the possible orthologs of APAF-1, CASP9 and CYCS in the species that have the three available proteins (Supplementary file 1), to predict possible aspects of the structure and functional dynamics of the protein complex.

## Structural and functional inferences in the APAF-1/apoptosome, guided by amino acid clusters with coevolution signal

### Regulatory region

#### Allosteric regulation in interaction CYCS and $\beta$ propellers during APAF-1 activation

The first step during the activation of APAF-1 is the association of CYCS with the regulatory region, to cause conformational changes in APAF-1 that release the CARD domain. The area of interaction between CYCS and  $\beta$  propeller 8 is greater than with  $\beta$  propeller 7,  $\sim 1,242 \text{ \AA}^2$  and  $\sim 712 \text{ \AA}^2$ , respectively (Zhou et al., 2015), so CYCS will probably first contact APAF-1 through  $\beta$  propeller 8 (Cheng et al., 2016). One of the interactions identified by Shalaeva et al. (2015), is between CYCS K/54 and  $\beta$  propeller 8 D/1147. In our results, only K/54 was detected in cluster 12, together with CYCS N/55 and  $\beta$  propeller 8 T/1012, located on  $\beta$ -sheet “a” of blade 2 on the back face of  $\beta$  propeller 8 (Figures 5 – A-figure supplement 1 - E). K/54 and N/55 are associated in the CYCS in the area of contact with  $\beta$  propeller 8. After CYCS binding,  $\beta$  propeller 8 rotates slightly upward to approach  $\beta$  propeller 7 (Cheng et al., 2016). This suggests that the interaction of K/54 and N/55 with D/1147, through an allosteric interaction influence T/1012, which causes a conformational change that would initiate this rotation, specifically in the upper part of the protein domain, moving  $\beta$ -sheet “a” of blade 2 forward and up (Figure 6 - A). Therefore, cluster 12 residues would be involved in the initiation of CYCS interaction with APAF-1.

After CYCS binding during APAF-1 activation,  $\alpha$ 29-32 helices in HD2 rotate slightly upward along with  $\beta$  propeller 8 (Cheng et al., 2016). Residues V/576 and A/584 in cluster 18 are located in the  $\alpha$ -helix 32 (Figure 5 – A-figure supplement 1 B and C), suggesting that these are part of the specific amino acids involved in conformational changes at the base of the protein domain during rotation (Figure 6 - B). Residues in HD2 L/593, L/595, W/597 and I/603 follow a hydrophobic corridor on the surface of  $\beta$  propeller 8 (Zhou et al., 2015). Two of these residues, L/595 and W/597 are located in cluster 18. These are in linker 3 (Figure 5 – A), which suggests that the linker is also involved in the rotation (Figure 6 - B).

The amino acids in clusters 10 and 11 are unique to  $\beta$  propeller 8 (Figure 5 – A-figure supplement 1 B and C). The distance between the residues in the structure suggests that they are involved in allosteric changes that cause alterations in  $\beta$ -sheet “d” of blade 4,  $\beta$ -sheets “a” and “d” of blade 5,  $\beta$ -sheet “c” of blade 7 and  $\beta$ -sheet “b” of blade 8, finalizing the conformational changes that rotate upwards to  $\beta$  propeller 8 (Figure 6 - C). Clusters 10 and 11 are independent of CYCS, suggesting that conformational changes in  $\beta$  propeller 8 are also caused by other ligands. The participation of residues V/1235 and L/1241 of cluster 10 in the rotation would initiate the formation of the V shaped cleft at the base.

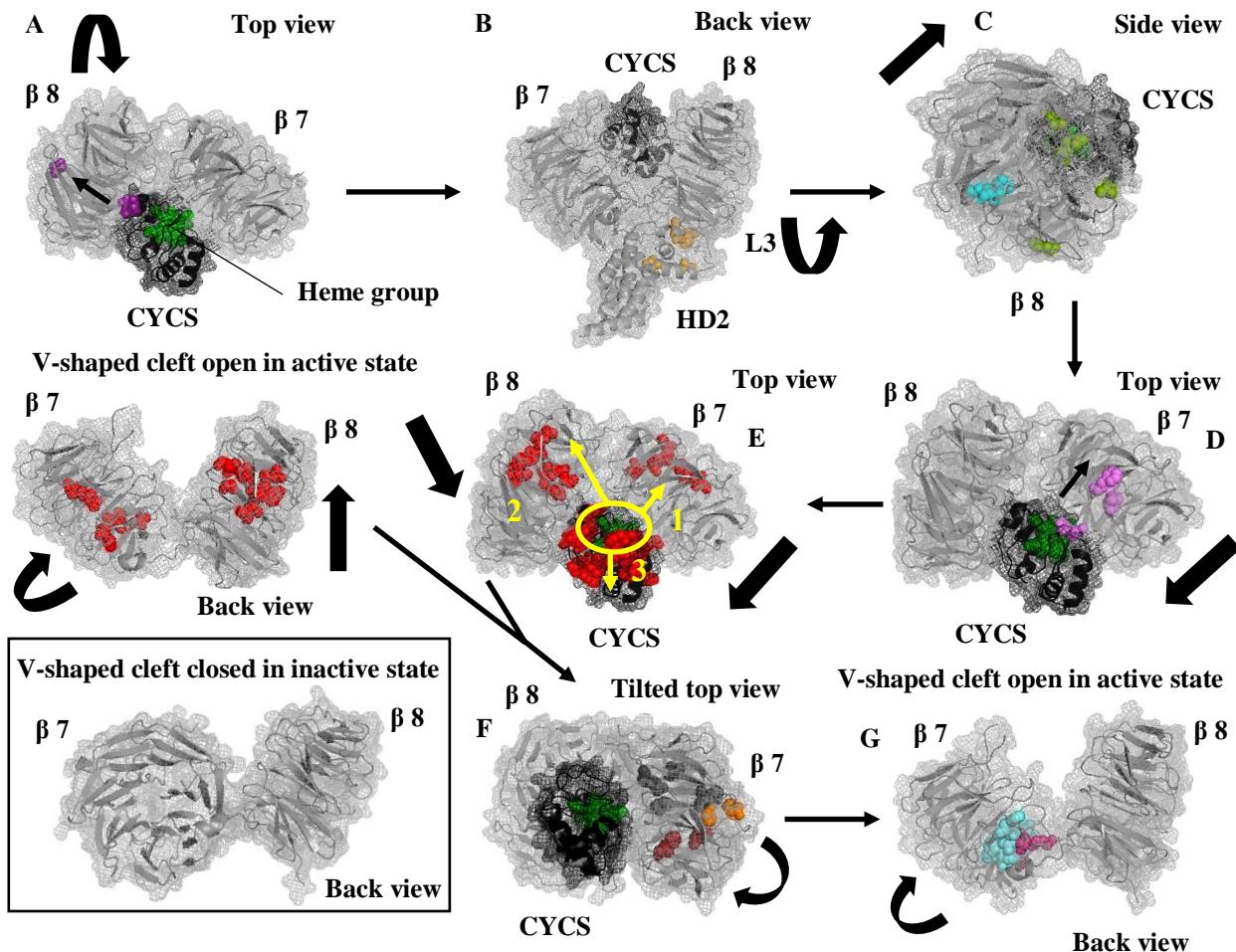
The next step during APAF-1 activation is the contact of CYCS with  $\beta$  propeller 7. Cheng et al. (2016), identified CYCS K/28 as an important residue for the interaction with the regulatory region of APAF-1. Shalaeva et al. (2015), determined that this residue interacts with  $\beta$  propeller 8 E/925. In our results K/28 was located in cluster 21 along with other residues in  $\beta$  propeller 7 (Figure 5 – A-figure supplement 1 - E), which makes more sense since K/28 is on the surface of the CYCS in contact with this domain. This suggests that CYCS contacts  $\beta$  propeller 7 through

these residues. After contact,  $\beta$  propeller 7 spins and rotates to "hug" CYCS in the V shaped cleft (Cheng et al., 2016), which suggests that K/28 through allosteric regulation interacts with F/787, V/799 and V/801, causing alterations in the  $\beta$ -sheet "a" of blade 4, which would initiate the rotation and spin of this protein domain and the formation in the upper region of the V shaped cleft in which the CYCS is coupled (Figure 6 - D).

Now the CYCS is in interaction with both  $\beta$  propellers. Other interactions in the contact area encompassing CYCS and  $\beta$  propellers include: CYCS M/13 and  $\beta$  propeller 7 E/700, CYCS Q/43 and  $\beta$  propeller 8 E/1045, CYCS G/57 and  $\beta$  propeller 8 T/1087, CYCS P/77 and  $\beta$  propeller 8 W/1179 (Zhou et al., 2015), CYCS K/56 and  $\beta$  propeller 8 E/1045 (Shalaeva et al., 2015). The residue CYCS K/56 was also identified by Cheng et al. (2016) as important in the interaction with the APAF-1 regulatory region. Our results reveal that the majority of CYCS residues involved in these interactions are connected through cluster 2 with more residues away from the contact area between both proteins, encompassing areas around the heme group in CYCS (Figure 5 – A-figure supplement 1 – G) and the central and posterior area of the  $\beta$  propellers (Figures 5 – A-figure supplement 1 - E). This suggests that the heme and/or residues within cluster 2 located in the contact zone, through a triple allosteric regulation, influence the other amino acids within the cluster located far from the contact zone, affecting three domains, causing conformational changes in  $\beta$ -sheet "b" of blade 5 and  $\beta$ -sheets "b" and "c" of blade 6 in  $\beta$  propeller 7,  $\beta$ -sheets "b" and "c" of blade 2 and  $\beta$ -sheets "a" and "b" of blade 3 in  $\beta$  propeller 8 and in the upper region of CYCS (Figure 6-E). In APAF-1 the residues in the  $\beta$  propellers far from the contact zone are located longitudinally around the V shaped cleft (Figure 6 - E), which suggests that the conformational changes in these amino acids continue with the formation of the cleft longitudinally, assisting linker 4, causing the CYCS to begin to fit into the cleft.

Clusters 14-17 and 23 are unique of  $\beta$  propeller 7 (Figure 5 – A-figure supplement 1 – B, D and E). Clusters 16, 17, and 23 would come into play first. During conformational changes in the regulatory region,  $\beta$  propeller 7 rotates  $\sim 60^\circ$  towards  $\beta$  propeller 8 (Zhou et al., 2015), "hugging" CYCS making it fit into the V shaped cleft (Cheng et al., 2016), which suggests that this turn is caused by allosterism in amino acids in clusters 16, 17 and 23 causing alterations in  $\beta$ -sheet "d" of blade 2,  $\beta$ -sheet "b" of blade 3 and  $\beta$ -sheets "a", "b" and "c" of blade 4, also contributing to the formation of the V shaped cleft in the upper region (Figure 6 - F).

Clusters 14 and 15 would then come into play, located at the base of the posterior face just below the V shaped cleft (Figure 5 – A-figure supplement 1 – B, D and E). These would cause conformational changes in  $\beta$ -sheets "c" and "d" of blade 6 and  $\beta$ -sheets "a", "b" and "I" of blade 7, assisting linkers 3 and 4, finishing the formation of the V shaped cleft at the base, making the CYCS fully fit into the cleft (Figure 6 - G). Clusters 14-17 and 23 are also independent of CYCS, so the existence of other ligands that contribute to conformational changes in  $\beta$  propeller 7 is possible. In inactive APAF-1, NBD K/192 interacts with  $\beta$  propeller 7 D/616 through a salt bridge, keeping CARD hidden (Reubold et al., 2011). During protein activation, when  $\beta$  propeller 7 spins, it disconnects this salt bridge, releasing CARD (Reubold et al., 2011). The residues in clusters 14 and 15 are close to K/192 and D/616, so they could contribute to breaking this salt bridge. See animated morph with sequence of events described so far in Video 1.



**Figure 6. Tentative model based on clusters with coevolution signal, of sequence of events in regulatory region during APAF-1 activation. A - G** A hierarchical sequence of the order (arrows) is shown in how the clusters cause the conformational changes (thick arrows) during the activation of APAF-1. Clusters are shown separately or in groups with a common function for clarity. The heme group in CYCS appears as green spheres. **A**) (Cluster 12) The interaction between CYCS and the regulatory region of APAF-1 starts when CYCS residues K/54 and N/55 interact directly with D/1147 in β propeller 8; These through allosteric interactions influence T/1012 (indicated by the arrow between the residues), initiating the upward rotation of this protein domain. **B**) (Cluster 18) HD2 residues V/576, A/584 and L3 L/595, W/597 detach β propeller 8 from HD2 in the lower region. **C**) (Clusters 10 and 11) A ligand other than CYCS, by allosteric regulation interacts with β propeller 8 V/1233, V/1235 and L/1241. Independently, another ligand allosterically regulates S/1060, V/1088, L/1089, L/1123 and Q/1211, initiating the formation of the V shaped cleft at the base and finishing the conformational changes that rotate β propeller 8 upwards, allowing interaction with β propeller 7; **D**) (Cluster 21) which occurs when CYCS K/28 interacts with this domain, associating these allosterically with β propeller 7 F/787, V/799 and V/801 (indicated by the arrow between the residues), initiating the rotation and the turning of this protein domain and the formation of the V shaped cleft in the upper region. Then, **E**) (Cluster 2), more interactions are produced that cover the entire length of the contact zone, through a triple allosteric regulation with the participation of heme as a second ligand (circled and indicated by arrows), they influence CYCS K/9, K/14, E/22, G/25, H/27, P/31, N/32, L/33, G/35, L/36, F/37, G/38, G/42, T/50, A/52, N/53, T/64, L/65, Y/75 and I/76, β propeller 7 I/813, M/814, L/854, A/855, V/856 and V/864 and β propeller 8 G/979, G/983, I/985, I/987, L/988, V/1006, F/1011 and L/1018, causing conformational changes that assist L4 to continue the formation of the V shaped cleft, causing CYCS to start to fit into it. The CYCS was removed in order to clearly visualize the cleft. To compare the cleft with the inactive/closed state, the regulatory region is shown in its inactive PDB 3SFZ state. **F**) (Clusters 16,

17 and 23) Some ligand other than CYCS by allosteric interactions regulates  $\beta$  propeller 7 V/746, S/752, L/758, S/762, D/764, L/767, H/698, L/716, F/751 and D/755, which completes the wide turn in which  $\beta$  propeller 7 “hugs” the CYCS, staying this protein between both  $\beta$  propellers. G) (Clusters 14 and 15) Some ligand binds allosterically to  $\beta$  propeller 7 L/866, W/867, V/885, V/888, F/890, F/897, L/898, T/899, S/901, A/876, H/881 and I/906, causing conformational changes in the base of  $\beta$  propeller 7, finishing the formation of the V shaped cleft, making the CYCS fit completely in it. Some of the residues of these clusters break the salt bridge that binds NBD K/192 and  $\beta$  propeller 7 D/616, causing the alterations that release CARD for the activation of CASP9, finishing the conformational changes between CYCS and  $\beta$  propellers during APAF-1 activation. See also Video 1.

## Arm and platform

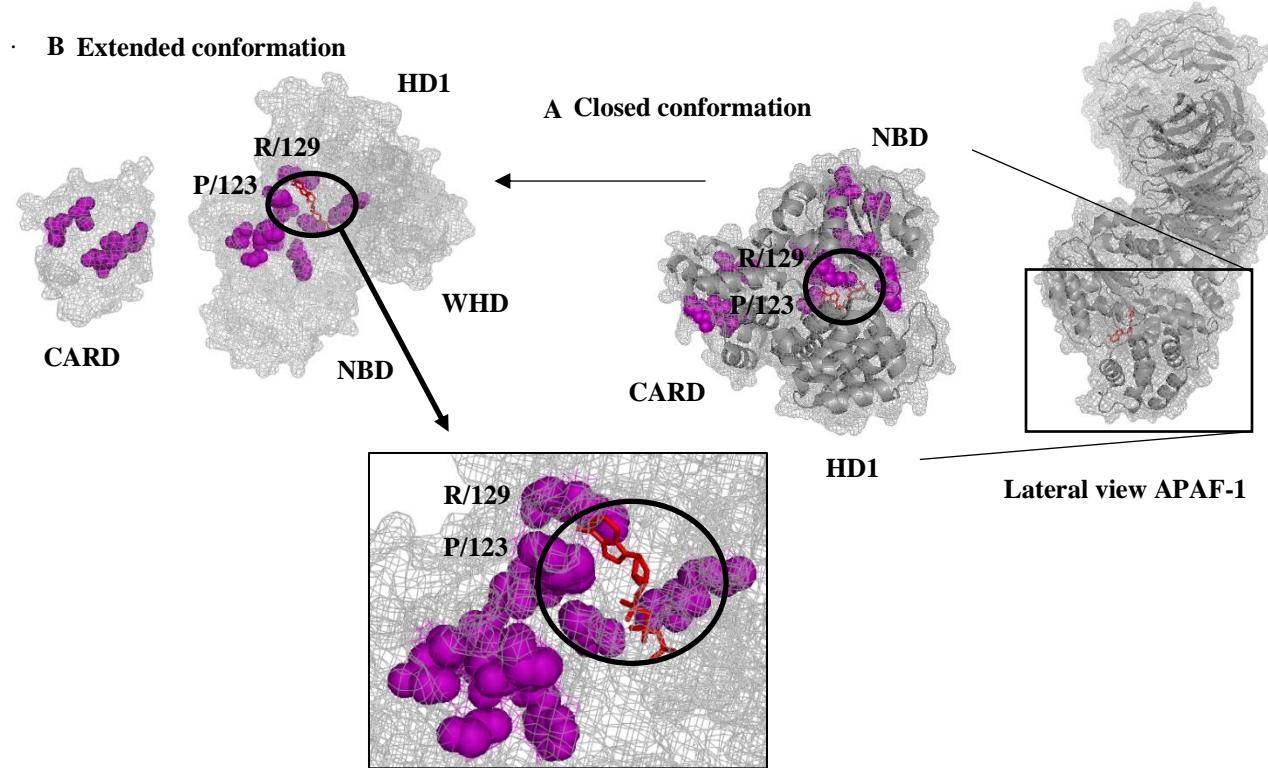
Cluster 19 is unique to the HD2 domain, encompassing residues in  $\alpha$ 27 –  $\alpha$ 29 helices (Figure 5 – A-figure supplement 1 - B). They are probably important to confer stability to the domain during conformational changes.

There are a series of interactions between domains that help maintain APAF-1 in its extended or active conformation, among these WHD-HD2 (Zhou et al., 2015). WHD associates with HD2 through Van der Waals interactions, where residues WHD I/388, V/393, V/399 and L/403 rest on the hydrophobic surface formed by HD2 H/457, I/460, H/490, Y/482, W/483, F/486 and Y/489 (Zhou et al., 2015). Cluster 20 has in common several of the residues in HD2 involved in these interactions, I/460 and I/461 in  $\alpha$ 25, Y/482 and H/490 in the  $\alpha$ 26. Other amino acids in the cluster are L/487 and M/491 in the  $\alpha$ 26 helix (Figure 5 – A-figure supplement 1 - B).

The  $\alpha$ 8 linker helix in NBD associates with a NBD of an adjacent APAF-1 during the oligomerization process (Yuan et al., 2010). In these interactions NBD Q/121 donates a hydrogen bridge to Q/257 of the adjacent NBD (Zhou et al., 2015). E/116 and R/122 (next to Q/121) located in the  $\alpha$ 8 helix of the NBD belong to cluster 9 (Figure 5). Hence we presumed that cluster 9 participates in the oligomerization process.

## Specific residues in NBD and APAF-1 CARD open hydrophobic pocket for exchange and stabilization of nucleotides

At the center of the interface in the interactions between WHD-NBD-HD1, there is an access channel or hydrophobic pocket that stores and keeps ADP stable in the inactive state of APAF-1 (Riedl et al., 2005; Cheng et al., 2016). This is blocked by CARD, but opens when this domain is removed, acquiring its extended conformation, allowing exchange for ATP (Riedl et al., 2005). Two of the amino acids that give structural support stabilizing ADP/ATP are P/123 and R/129 in NBD (Riedl et al., 2005; Cheng et al., 2016). These two residues belong to cluster 7 along with others located in  $\alpha$ 4,  $\alpha$ 6 helices at the carboxyl terminal of the CARD domain; in  $\alpha$ 8,  $\alpha$ 9 helices and in  $\beta$ 1 sheet at the amino terminal of the NBD (Figure 5). Unlike P/123 and R/129, the other amino acids in the cluster probably participate in the conformational changes necessary for the ADP/ATP exchange. Within these changes, the cluster residues located at the carboxyl terminal of the CARD would be essential for blocking and opening the hydrophobic pocket where the nucleotide is found. (Figure 7 and Video 2).

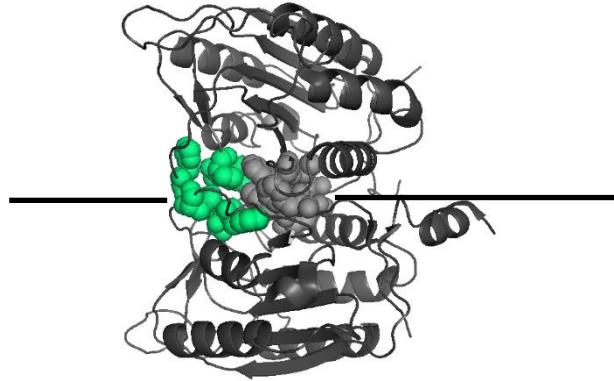


**Figure 7. Role of cluster 7 in hydrophobic pocket aperture for nucleotide exchange.** Side view of APAF-1 in inactive state PDB 3SFZ. **A)** Close-up of the closed hydrophobic pocket that stores ADP (red sticks circled) in structure PDB 1Z6T, showing the residues in cluster 7 located between CARD and NBD. **B)** The residues P/123 and R/129 remain below and above the adenine, giving it stability. The other residues of the cluster would participate in conformational changes that remove CARD, remaining in its extended conformation, opening the hydrophobic pocket (augmented region and indicated with an arrow) for nucleotide exchange. After the removal of the CARD, the open hydrophobic pocket can be seen, the nucleotide being most visible. Structure PDB 5JUY. See also Video 2.

## Disk

### Large subunit of CASP9 also contributes to homodimerization

When catalytic domains homodimerize, they do it by their carboxyl terminals through motif 402 GCFNF 407 in the small subunit (Wu et al., 2016). Renatus et al., (2001) identified in the large subunit a loop of 7 residues inserted at position 240 between the beta 3 and 3A sheets that interact through the interface of the dimer, therefore both the small and large subunits participate in the homodimerization process. Our results corroborate this statement, since of these 7 residues, H/243, F/246 and P/247 that belong to cluster 5 (Figure 5 - A-figure supplement 1 - F), in the dimer structure they are in direct contact (Figure 8).



**Figure 8. Residues of cluster 5 participate in CASP9 homodimerization.** Catalytic domains p20/p10 are shown as dimer (delimited by horizontal black line) PDB 1JXQ. Amino acids of Cluster 5 were mapped as lime green spheres located in the large subunit. The residues as gray spheres belong to motif 402 GCFNF 407 (not detected in coevolution analysis) in the small subunit. In the dimer, all these residues are in direct contact, so the large subunit also contributes to homodimerization.

### **Conglomerate residues throughout the catalytic domain would transmit the signal to the active site C/287 of CASP9 and would participate in inhibition mediated by Zinc**

In the process of activation of CASP9, cleavages are caused in E/306, D/315 and D/330 that expose the active site C/287 (Bratton et al., 2001). In our results, residue C/287 located in  $\beta$ -sheet 4 belongs to cluster 6 together with another 25 amino acids in the large subunit, sheets  $\beta$ 1,  $\beta$ 2 of 240-loop,  $\beta$ 3,  $\beta$ 4 and helix  $\alpha$ 3 (Figure 5 - A). Renatus et al. (2001) propose that conformational changes during homodimerization transmit the activation signal to the residues surrounding the active site C/287. In the three-dimensional structure, cluster 6 residues are in direct contact throughout the large subunit (Figure 5-figure supplement 1 - F), and may be those that participate in the conformational changes that transmit this signal. The catalytic domain is also involved in the inhibition of CASP9 mediated by Zinc, having two binding regions to this inhibitor, one of them encompassing the active site C/287 and two residues that in the three-dimensional structure are around it, H/237 and C/239 (Huber and Hardy, 2012). Of these residues, apart from the active site, H/237 belongs to cluster 6 (Figure 5 - A-figure supplement 1 - F), so we can infer that the residues in cluster 6 are involved in both the activation and inhibition mediated by Zinc of CASP9.

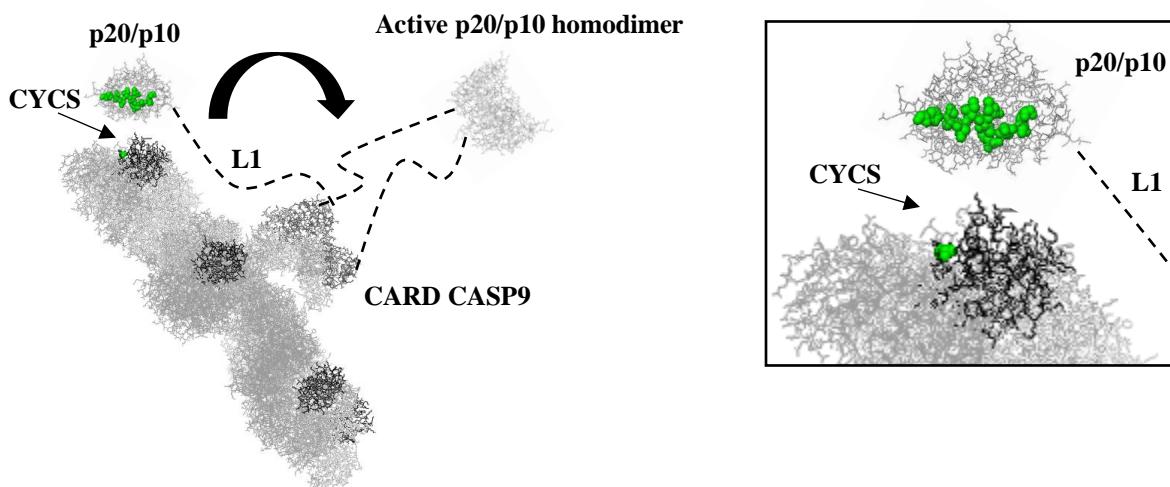
### **The molecular timer**

Wu and Bratton (2017) state that as part of the molecular timer, the cleavage in D/315 during the activation of CASP9 destabilizes the dimers causing them to dissociate from the complex, losing catalytic activity. If the protein complex requires more functioning time, CASP3 avoids the destabilization of the dimers through a positive feedback, carrying out a cleavage at D/330, removing the linker from the intersubunit, which gives more stability to the dimers. In our results, D/330 belongs to cluster 3 along with other residues located in linker 2 and  $\beta$ -sheet 5 (Figure 5 – A-figure supplement 1 - F). Apparently they are responsible for giving stability to CASP9 after cleavage at D/330 by CASP3.

Another of the regulatory mechanisms in these processes to remove caspases from the apoptosome is through a caspase-mediated cleavage in APAF-1 that removes H1 from the CARD domain, turning it into an 84 or 130 kDa fragment, not being able to interact more with caspases (Lauber et al., 2001). Possibly the cleavage occurs in motif 16 LEKD 19 at the amino terminal of CARD, and other amino acids at positions 37-52 and 12-28 may also be involved in the generation of fragments p84 and p130, respectively (Lauber et al., 2001). Several residues inside this range fell within cluster 13: L/16 of motif 16 LEKD 19, along with other residues in  $\alpha$ 1,  $\alpha$ 2 and  $\alpha$ 3 helices (Figure 5), giving a more precise mapping of the specific residues that could be involved in the cleavage process. This suggests that residues in cluster 13 would have an inhibitory function, blocking the functioning of the apoptosome as part of the molecular timer.

### As an alternate mechanism, the catalytic domain of CASP9 could interact with CYCS in the regulatory region to expose active site C/287

Gasperin - Sánchez et al. (2020), made molecular dynamics simulations in order to describe conformational changes in pc9 during its activation, where in all these the active site C/287 was always hidden, so the authors propose a different model than the holoenzyme, in which the catalytic domains interact with other proteins of the apoptosome, causing the exposure of the active site, and then homodimerize. In this model, in one of the conformations, the catalytic domains p20/p10 through linker 1 remain in contact with the regulatory region. In our results, cluster 22 encompasses residues in  $\beta$  sheets 1 and 2 and helix  $\alpha$ 1 in the catalytic domain of CASP9 and CYCS A/44 (Figure 5 - A-figure supplement 1 – F and G), suggesting that this possible interaction between CASP9 and the regulatory region, specifically it would be between the catalytic domain of CASP9 and CYCS. In this hypothetical scenario, after pc9 anchors itself to the apoptosome in the disc, as an alternate mechanism, p20/p10 interacts through linker 1 with CYCS in the regulatory region through these residues, exposing the active site C/287, so that later the catalytic domains can homodimerize (Figure 9).



**Figure 9. Exposure of the active site C/287 in pc9 caused by interaction with CYCS in regulatory region.** Once pc9 is anchored to the disk of the apoptosome (PDB 5JUY), the p20/p10 domains (PDB 1JXQ) interact through L1 with CYCS in the regulatory region, to expose the active site C/287, and then homodimerize.

## **Apoptosome-independent apoptotic functional inferences**

### **CASP9 catalytic domain could be myristoylated to interact with mitochondrial outer membrane**

CASP9 also interacts with the outer membrane of the mitochondria, producing cleavages in the anti-apoptotic BCL-2 proteins, amplifying the release of more pro-apoptotic proteins (Guerrero et al., 2012). Cluster 4 of CASP9 has two glycine residues, G/269 in the  $\alpha$ 3 helix and G/288 in the  $\beta$ 4 sheet of the large subunit, located at each end of the catalytic domain (Figure 5 – A-figure supplement 1 - F). To complement the results of Guerrero et al., 2012, what we propose is that the interaction of CASP9 with the outer membrane of the mitochondrion is carried out through these two residues. In this hypothetical scenario, some caspase or other protease would produce a cleavage in CASP9, exposing the two glycine residues, which are myristoylated, then interacting with the membrane.

### **pc9 could dimerize inside mitochondria**

Inside the mitochondria, the amino terminal of pc9 has been reported to enter the matrix (Milisav and Šuput, 2007). Cluster 8 residues exclusive of pc9/CASP9 are hydrophobic, located within the protein domain (Figure 5-figure supplement 1 – B and E), so they may serve as an anchor point for pc9 to associate with the inner membrane of mitochondria. Once anchored, it is possible that it dimerizes with another anchored pc9 to become activated, with the possibility that it interacts with other substrates including CASP3.

We have as a caveat the following statements. Amino acids with a coevolution signal do not necessarily have to be involved in structural or functional interactions. Other restrictions by which they can be detected are: negative design, codon effects and errors in phylogenetic analyzes (Anishchenko et al., 2017). Another aspect to take into account of these detected residues is that the interaction may not be in that particular protein, but that they could reflect interactions in other members of the protein family. For example, amino acids that coevolve in distant positions in a protein, is because their homologous residues can be in direct contact in other proteins of the family (Anishchenko et al., 2017). Therefore, this restriction is due to amino acids having a common evolutionary history (Chakrabarti and Panchenko, 2009). Another reason why the residues may be distant is due to errors in the determined structure of the protein, which may suggest that these distant amino acids are actually in direct interaction (Anishchenko et al., 2017), which in our specific case could be reflected in some regions of the APAF-1/apoptosome that have been structurally characterized at high resolutions, not being of good quality. For these reasons, certain caution must be taken when taking as valid the structural and functional inferences of the protein complex guided by amino acids with coevolution signals, which are only models. Even so, these potential interactions detected will serve as a guide to be validated experimentally.

## Influence of destabilizing changes in structure and function of the APAF-1/apoptosome

In general, molecular adaptation does not work on the specific amino acids necessary for a catalytic reaction or conformational change. On the contrary, adaptation results from changes in those amino acids that surround active/functional residues to represent a slight change in the way in which these sites perform their function (Baer, 2007), so we presume that the substitutions in amino acids with coevolution signal are more conservative because they are in active sites and involved in conformational changes, so that the functionality of the protein is maintained. On the other hand, the substitutions in amino acids located in the periphery are destabilizing, since their function is not so essential, being more apt to undergo radical changes. In some cases, residues in the interior near active sites may require radical adaptations, these changes being guided in a subtle way by residues in the periphery.

Compressibility is characteristic of globular proteins (Gromiha and Ponnuswamy, 1993), which in turn is characteristic of enzymatic proteins with regulatory function. For this reason, it makes sense that the regions most affected by this property in the apoptosome are globular regions, encompassing in APAF-1 the  $\beta$  propellers that form the regulatory region, CYCS and CASP9 throughout their length, whose function is enzymatic. Compressibility is related to volume, partial specific volume, molecular volume, and molecular weight. These physicochemical properties influence the amino acid packing density of proteins (Karshikoff and Ladenstein, 1998 and Fischer et al., 2004).

The helical contact area and the hydrophobicity by thermodynamic transfer favor the sequestration of residues into the protein (Harris and Pettitte, 2016) and the folding of globular proteins (Dill, 1990). These have served to guide the folding, conglomeration and associations of the hydrophobic amino acids within the affected regions.

Proteins affected by refractive index and chromatographic index possess stability in external fields, which is required during conformational changes and interactions between enzymes (Pace et al., 1990; Vörös, 2004; Hong et al., 2012), such as processes of oligomerization (Folta-Stogniew, 2006). It is relevant to note that one of the regions most affected by these two properties was the platform of the protein complex, which is essential for the oligomerization process. Chromatographic index indicates the most affected residues (Table 1). This makes sense since the apoptosome is very dynamic, having many conformational changes and interactions. The stability given by these properties may be required in the regulatory region for association with CYCS and the conformational changes caused by this interaction during APAF-1 activation; In NOD, during nucleotide exchange and oligomerization during the heptamer formation; on HD2, during the spin to extend the CARD domain of APAF-1; in the CARDs domains of both proteins, during the CARD-CARD interaction for the recruitment of CASP9 in the disk; and in the catalytic domain of pc9 during the cleavage and activation processes of this protein.

Apparently, the globular nature and the dynamic complexity of the apoptosome were the evolutionary constraints on natural selection to favor these significant physicochemical properties that caused destabilizing changes in the protein complex. These drastic changes were essential to keep the apoptosome functioning throughout its evolution.

## Materials and methods

### Obtention of the genes, messenger RNAs and proteins of the human apoptosome

*H. sapiens* sequences were obtained from GenBank (Benson et al., 1993; Sayers et al., 2019). The gene codes are 317 for APAF-1 and 842 for CASP9. For each of these sequences in the section "gene table" there are several isoforms of transcripts. The canonical isoform and its translated protein were selected for each. Here the codes for the messenger RNAs and proteins, respectively, APAF-1: NM\_181861.1 and NP\_863651.1, CASP9: NM\_001229.4 and NP\_001220.2.

### Identification of proteins homologous to the human apoptosome in invertebrates

The sequences were obtained in two ways on 06/25/2017. First, a general search was made in GenBank using as reference the names of the human sequences, selecting the genes within invertebrates. To avoid the exclusion of sequences that could be selected, genes were searched using all synonymous names. The longest protein and its transcript were selected from each gene. Second, within the GenBank reference sequence database (RefSeq), a BLASTP search was run on human protein sequences against the more general invertebrate taxa: Porifera, Coelenterata, Protostomes, and Deuterostomes (excluding vertebrates). A single iteration per taxon was made, selecting the sequences whose scores were greater than 200. For each one its transcript was obtained. As a result, sequences of APAF-1 and CASP9 were obtained in 92 species of invertebrates (Supplementary file 1). Not all species have the two proteins in common. Of these, five sequences were not under the name CED-4 but APAF-1 in the nematodes *Trichinella nativa*, *T. papuae*, *T. pseudospiralis*, *T. zimbabwensis* and *Trichuris suis*, while 59 were not under the name DARK but APAF-1 in six orders of insects, Isoptera, Hemiptera, Coleoptera, Diptera, Lepidoptera and Hymenoptera.

### Identification by similarity of orthologs to the human apoptosome in Metazoa

The Ortho-Profile method (Szklarczyk et al., 2012) was applied to the species within invertebrates in which homologs were detected, designed to detect orthologous genes and proteins in highly divergent sequences. This consists of three phases with an increase in sensitivity in each of them. If the possible ortholog sought is not detected in the first phase, it goes to the second with more sensitivity, and if it is not detected in this, it goes to the third, which has even more sensitivity. In each of the phases, for the similarity to be significant, both sequences must present reciprocity and have an e-value of less than 0.01. In the first phase, a single iteration of reciprocal BLASTP to the human proteins against the selected species was executed, one by one, sequence against sequence, in the RefSeq database. In the second phase, a PSI-BLAST (Altschul et al., 1997) was also applied reciprocally, in the GenBank non-redundant reference database (RefSeq nr), against the remaining species, one by one, profile versus sequence with a profile inclusion threshold equal to 0.005. Up to three iterations were done to see if it detected significant similarities. In the third and final phase, the profile-based HMM was used. On the HMMER online server (<http://www.ebi.ac.uk/Tools/hmmer/>), the algorithm was run against the human sequences against the reference proteome database and against the remaining species one by one.

In the complete GenBank report for each human apoptosome gene, in the general information tab, the putative orthologous genes of 115 species of vertebrates were obtained through the

annotation pipeline (Thibaud-Nissen et al., 2016). For each gene, the protein with the longest length was selected with its respective transcript (Supplementary file 1).

### **Identification by phylogeny of orthologs to the human apoptosome in Metazoa**

For the phylogenies, apart from the human sequences, the DARK and DRONC proteins from *D. melanogaster* and CED-4 from *C. elegans* were also included, to be used as a reference and to determine in relation to the apoptosomes formed by APAF-1, DARK and CED-4, where the others are located. In GenBank, for *D. melanogaster* and *C. elegans*, the canonical protein sequences, DARK: NP\_725637.1, DRONC: NP\_524017.1 and CED-4: NP\_001021202.1, were selected. To obtain the architecture of the protein domains of the reference sequences and other invertebrates and vertebrates, analyzes of conserved domains were made in Pfam 32.0 (<https://pfam.xfam.org/>) (Sonnhammer et al., 1997; El- Gebali et al., 2019). Each protein family was aligned separately in MAFFT v7.380 (Katoh et al., 2002; Katoh and Standley, 2013), under the “-auto” mode, which selected the strategies “FFT-NS-i” and the method of iterative refinement, doing two iterations for the APAF-1 alignment, and “L-INS-i” and the iterative refinement method, doing an iteration with information of local pairs, for the CASP9 alignment. The produced alignments were inspected and edited in AliView 1.20 (Larsson, 2014), subsequently removing sequences with amino acids of unknown nature. To increase the quality of the alignments, the poorly aligned regions were removed using trimAL 1.2rev57 (Capella-Gutiérrez et al., 2009). Once edited, they were run separately in phylip format in ProtTest 3.4.2 (Abascal et al., 2005; Darriba et al., 2011), determining the following best models of amino acid substitutions during evolution, “JTT + G + F”, assigning to the gamma distribution the value of 1.639 for the APAF-1 and “JTT + I + G”, assigning to the gamma distribution and proportion of invariant sites the values of 1.186 and 0.028 for the CASP9. Based on these models, the phylogenetic reconstructions were made in phyML 3.1 (Guindon and Gascuel, 2003; Guindon et al., 2010), under the principle of maximum likelihood (Felsenstein, 1981), using the non-parametric method Shimodaira - Hasegawa (Shimodaira and Hasegawa, 1999), generating 3 phylogenetic trees corresponding to each protein family. These were rooted in Seaview 4 (Gouy et al., 2009) and visualized in FigTree 1.4.3 (Rambaut, 2017) (<https://tree.bio.ed.ac.uk/software/figtree>).

Species-level phylogenies were generated automatically in the NCBI taxonomy database (Federhen, 2012). A species tree was generated for each protein phylogeny. For the reconciliation of protein and species phylogenetic trees, Notung Version 2.9 (Chen et al., 2000) was used. All of our protein phylogenies are binary, but our species trees are polytomous, which is why they were reconciled in the program under the parsimony algorithm of gene duplication, transfer, loss, and incomplete lineage sorting (Stolzer et al., 2012). The reconciliations were visualized in Archeopteryx version 0.9921 beta (Zmasek, 2015).

### **Obtention of the CYCS sequences**

Once proteins orthologs to the human sequences were detected and selected, for each species with the two orthologs available, the CYCS sequences were obtained, completing the proteins that make up the apoptosome. Using as reference the protein sequence of the human CYCS (Protein ID: NP\_061820.1), putative orthologous proteins were obtained in vertebrates through the annotation pipeline (Thibaud-Nissen et al., 2016). For invertebrates, a BLASTP search was applied to the human sequence against the selected species (Supplementary file 1). CYCS sequences were aligned in MAFFT v7.380 (Katoh et al., 2002; Katoh and Standley, 2013), under the “-auto” mode,

which selected the strategy "L-INS-i" and the iterative refinement method, iterating with local pair alignment information.

### Protein coevolution analysis

The complete alignments of orthologous sequences (not trimmed with trimAL) were concatenated in Mesquite Version 3.31 (Maddison and Maddison, 2017) (<http://www.mesquiteproject.org>), again determining the best amino acid substitution model and a phylogeny, using ProtTest 3.4.2 (Abascal et al., 2005; Darriba et al., 2011) and phyML 3.1 (Guindon and Gascuel, 2003; Guindon et al., 2010), respectively. In Seaview 4 (Gouy et al., 2009) the phylogeny generated was rooted in the node that separates invertebrates from vertebrates.

For the coevolution analysis, the alignment and its respective phylogeny were supplied to the BIS2Analyzer server (<http://www.lcqb.upmc.fr/BIS2Analyzer/>) (Oteri et al., 2017). The API was calculated for our concatenated alignment in Geneious 11.1.3 (Kearse et al., 2012), obtaining a score of 67.3%, this having a moderate divergence. Therefore, once uploaded to the server, the option "Alphabet-Reduction" was selected, with a maximum dimension (D) of 2. The results were mapped into the concatenated linear sequences and three-dimensional structures available in the database of proteins (PDB) using Jalview 2.10.4 (Clamp et al., 2004; Waterhouse et al., 2009) and PyMOL (DeLano, 2014) (<https://pymol.org/2/>), respectively. Clusters were also mapped on morphs (transformation from inactive to active protein state) in UCSF Chimera 1.13.1 rc (Pettersen et al., 2004).

### Positive selection analysis

In GenBank, from the messenger RNA isoforms of the selected orthologs, only the coding regions were extracted, then aligned in TranslatorX Version 1.1 (Abascal et al., 2010), under the MAFFT algorithm, using as reference the same alignments of the proteins generated in the coevolution part. The alignments were concatenated in Mesquite Version 3.31 (Maddison and Maddison, 2017) (<http://www.mesquiteproject.org>), then executed in phylip format in jModelTest 2.1.10 (Posada, 2008; Darriba et al., 2012), determining the best model for nucleotide substitution during evolution to "TIM3 + G", assigning the gamma distribution a value of 0.7160. Based on this model, the phylogenetic reconstruction was done in phyML 3.1 (Guindon and Gascuel, 2003; Guindon et al., 2010). In Seaview 4 (Gouy et al., 2010) the phylogeny generated was rooted in the node that separates invertebrates from vertebrates.

To detect positive selection, TreeSAAP Version 3.2 (Woolley et al., 2003) was used. The input file with the alignment of the coding DNA in nexus format and the phylogeny in newick format were supplied to the program. A first run was carried out under the predetermined parameters, taking into account the 31 physicochemical properties of the amino acids available in the program, identifying which are the most significant. Using these preliminary data as a reference, another run was made, this time taking into account only the significant physicochemical properties, giving the sliding windows an average length of 20. The definitive results, focusing on the residues under destabilizing positive selection (categories 6 - 8 and z scores: -3.09 and +3.09), were visualized in IMPACT\_S (Maldonado et al., 2014), then mapping the residues in three-dimensional structures in PyMOL (DeLano, 2014) (<https://pymol.org/2/>).

## Acknowledgements

We thank Ruth Bastardo, Altagracia Espinosa, Santo Navarro, Candy Ramírez and América Sánchez of the Institute of Botanical and Zoological Research Prof. Rafael M. Moscoso (IIBZ) of the Autonomous University of Santo Domingo, for their guidance on how to organize voluminous data products of this research. José A. Sánchez - Borbón thanks Dr. Yamixa Delgado - Reyes from the San Juan Bautista School of Medicine, Caguas, Puerto Rico, for teaching the basics of the PyMOL program; Willy Maurer from the Specialized Institute for Higuer Education - Loyola, San Cristobal, Dominican Republic, and Fritz Pichardo - Marcano from the IIBZ, for recommending and providing technical support with the use of some of the programs to run the analyzes.

## Authors' contributions

José A. Sánchez - Borbón, conceptualization, design of the methodology, execution of the analyzes, interpretation of the results, writing of the original draft; Steven E. Massey, conceptualization, methodology design, supervision, original draft writing; José D. Hernández - Martich, supervision, writing of the original draft.

## Conflicts of interest

We declare that we have no conflicts of interest.

## References

1. Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104-2105. doi: [10.1093/bioinformatics/bti263](https://doi.org/10.1093/bioinformatics/bti263)
2. Abascal F, Zardoya R, Telford, M.J. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research* 38(suppl\_2):W7-W13. doi: [10.1093/nar/gkq291](https://doi.org/10.1093/nar/gkq291)
3. Altschuh D, Lesk AM, Bloomer AC, Klug A. 1987. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *Journal of molecular biology* 193(4):693-707. doi: [10.1016/0022-2836\(87\)90352-4](https://doi.org/10.1016/0022-2836(87)90352-4)
4. Altschuh D, Vernet T, Berti P, Moras D, Nagai K. 1988. Coordinated amino acid changes in homologous protein families. *Protein Engineering, Design and Selection* 2(3):193-199. doi: [10.1093/protein/2.3.193](https://doi.org/10.1093/protein/2.3.193)
5. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25(17):3389-3402. doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389)
6. Anishchenko I, Ovchinnikov S, Kamisetty H, Baker, D. 2017. Origins of coevolution between residues distant in protein 3D structures. *Proceedings of the National Academy of Sciences* 114(34):9122-9127. doi: [10.1073/pnas.1702664114](https://doi.org/10.1073/pnas.1702664114)
7. Baer, K.K., 2007. Protein Coevolution and Coadaptation in the Vertebrate bc1 Complex (Master thesis). Provo, UT: Brigham Young University, Provo.
8. Bender CE, Fitzgerald P, Tait SW, Llambi F, McStay GP, Tupper DO, Pellettieri J, Alvarado AS, Salvesen GS, Green DR. 2012. Mitochondrial pathway of apoptosis is ancestral in metazoans.

Proceedings of the National Academy of Sciences 109(13):4904-4909. doi: [10.1073/pnas.1120680109](https://doi.org/10.1073/pnas.1120680109)

9. Benson D, Lipman DJ, Ostell J. 1993. GenBank. Nucleic Acids Research 21(13):2963-2965. doi : [10.1093/nar/21.13.2963](https://doi.org/10.1093/nar/21.13.2963)

10. Bratton SB, Walker G, Srinivasula SM, Sun XM, Butterworth M, Alnemri ES, Cohen GM. 2001. Recruitment, activation and retention of caspases-9 and-3 by Apaf-1 apoptosome and associated XIAP complexes. The EMBO journal 20(5): 998-1009. doi: [10.1093/emboj/20.5.998](https://doi.org/10.1093/emboj/20.5.998)

11. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-1973. doi: [10.1093/bioinformatics/btp348](https://doi.org/10.1093/bioinformatics/btp348)

12. Chakrabarti S, Panchenko AR. 2009. Coevolution in defining the functional specificity. Proteins: Structure, Function, and Bioinformatics 75(1):231-240. doi: [10.1002/prot.22239](https://doi.org/10.1002/prot.22239)

13. Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. Journal of Computational Biology 7(3-4):429-447. doi: [10.1089/106652700750050871](https://doi.org/10.1089/106652700750050871)

14. Cheng TC, Hong C, Akey IV, Yuan S, Akey CW. 2016. A near atomic structure of the active human apoptosome. *Elife* 5:e17755. doi: [10.7554/elife.17755](https://doi.org/10.7554/elife.17755)

15. Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The jalview java alignment editor. Bioinformatics 20(3):426-427. doi: [10.1093/bioinformatics/btg430](https://doi.org/10.1093/bioinformatics/btg430)

16. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27(8):1164-1165. doi: [10.1093/bioinformatics/btr088](https://doi.org/10.1093/bioinformatics/btr088)

17. Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature methods 9(8):772-772. doi: [10.1038/nmeth.2109](https://doi.org/10.1038/nmeth.2109)

18. Dill KA, 1990. Dominant forces in protein folding. Biochemistry 29(31):7133-7155. doi: [10.1021/bi00483a001](https://doi.org/10.1021/bi00483a001)

19. Dorstyn L, Read S, Cakouros D, Huh JR, Hay BA, Kumar S. 2002. The role of cytochrome c in caspase activation in *Drosophila melanogaster* cells. The Journal of cell biology 156(6):1089-1098. doi: [10.1083/jcb.200111107](https://doi.org/10.1083/jcb.200111107)

20. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL. 2019. The Pfam protein families database in 2019. Nucleic acids research 47(D1):D427-D432. doi: [10.1093/nar/gky995](https://doi.org/10.1093/nar/gky995)

21. Elmore S. 2007. Apoptosis: a review of programmed cell death. Toxicologic pathology 35(4):495-516. doi: [10.1080/01926230701320337](https://doi.org/10.1080/01926230701320337)

22. Fasman, GD. 1976. Handbook of Biochemistry and Molecular Biology, 3rd ed. 1976. ed, Proteins – Volume 1. CRC Press.

23. Federhen S. 2012. The NCBI taxonomy database. Nucleic acids research 40(D1):D136-D143. doi: [10.1093/nar/gkr1178](https://doi.org/10.1093/nar/gkr1178)

24. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of molecular evolution 17(6):368-376. doi: [10.1007/BF01734359](https://doi.org/10.1007/BF01734359)

25. Fischer H, Polikarpov I, Craievich AF. 2004. Average protein density is a molecular-weight-dependent function. Protein Science 13(10):2825-2828. doi: [10.1110/ps.04688204](https://doi.org/10.1110/ps.04688204)

26. Folta-Stogniew E. 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors In: Nedelkov D, Nelson R, editors. New and Emerging Proteomic Techniques. Humana Press. pp. 97-112. doi: [10.1385/1-59745-026-X:97](https://doi.org/10.1385/1-59745-026-X:97)

27. Galluzzi L, Vitale I, Aaronson SA, Abrams JM, Adam D, Agostinis P, Alnemri ES, Altucci L, Amelio I, Andrews DW, Annicchiarico-Petruzzelli M, Antonov AV, Arama E, Baehrecke EH, Barlev NA, Bazan NG, Bernassola F, Bertrand MJM, Bianchi K, Blagosklonny MV, Blomgren K, Borner C, Boya P, Brenner C, Campanella M, Candi E, Carmona-Gutierrez D, Cecconi F, Chan FK, Chandel NS, Cheng EH, Chipuk JE, Cidlowski JA, Ciechanover A, Cohen GM, Conrad M, Cubillos-Ruiz JR, Czabotar PE, D'Angiolella V, Dawson TM, Dawson VL, De Laurenzi V, De Maria R, Debatin KM, De Berardinis RJ, Deshmukh M, Di Daniele N, Di Virgilio F, Dixit VM, Dixon SJ, Duckett CS, Dynlacht BD, El-Deiry WS, Elrod JW, Fimia GM, Fulda S, García-Sáez AJ, Garg AD, Garrido C, Gavathiotis E, Golstein P, Gottlieb E, Green DR, Greene LA, Gronemeyer H, Gross A, Hajnoczky G, Hardwick JM, Harris IS, Hengartner MO, Hetz C, Ichijo H, Jäättelä M, Joseph B, Jost PJ, Juin PP, Kaiser WJ, Karin M, Kaufmann T, Kepp O, Kimchi A, Kitsis RN, Klionsky DJ, Knight RA, Kumar S, Lee SW, Lemasters JJ, Levine B, Linkermann A, Lipton SA, Lockshin RA, López-Otín C, Lowe SW, Luedde T, Lugli E, MacFarlane M, Madeo F, Malewicz M, Malorni W, Manic G, Marine JC, Martin SJ, Martinou JC, Medema JP, Mehlen P, Meier P, Melino S, Miao EA, Molkentin JD, Moll UM, Muñoz-Pinedo C, Nagata S, Nuñez G, Oberst A, Oren M, Overholtzer M, Pagano M, Panaretakis T, Pasparakis M, Penninger JM, Pereira DM, Pervaiz S, Peter ME, Piacentini M, Pinton P, Prehn JHM, Puthalakath H, Rabinovich GA, Rehm M, Rizzuto R, Rodrigues CMP, Rubinsztein DC, Rudel T, Ryan KM, Sayan E, Scorrano L, Shao F, Shi Y, Silke J, Simon HU, Sistigu A, Stockwell BR, Strasser A, Szabadkai G, Tait SW, Tang D, Tavernarakis N, Thorburn A, Tsujimoto Y, Turk B, Vanden Berghe T, Vandenabeele P, Vander Heiden MG, Villunger A, Virgin HW, Vousden KH, Vucic D, Wagner EF, Walczak H, Wallach D, Wang Y, Wells JA, Wood W, Yuan J, Zakeri Z, Zhivotovsky B, L. Zitvogel L, Melino G, Kroemer G. 2018. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death. 2018. *Cell Death & Differentiation* 25 (3):486-541. doi: [10.1038/s41418-017-0012-4](https://doi.org/10.1038/s41418-017-0012-4)

28. Gasperin-Sánchez H, Benítez-Cardoza CG, Caro-Gómez LA, Rosas-Trigueros JL, Zamorano-Carrillo, A. 2020. Transit of Procaspase-9 towards its activation. New mechanistic insights from molecular dynamics simulations. *Journal of Molecular Modeling* 26(2):24. doi: [10.1007/s00894-019-4285-z](https://doi.org/10.1007/s00894-019-4285-z)

29. Göbel U, Sander C, Schneider R, Valencia A. 1994. Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Bioinformatics* 18(4):309-317. doi: [10.1002/prot.340180402](https://doi.org/10.1002/prot.340180402)

30. Gouy M, Guindon S, Gascuel, O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* 27(2):221-224. doi: [10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259)

31. Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* 185(4154):862-864. doi: [10.1126/science.185.4154.862](https://doi.org/10.1126/science.185.4154.862)

32. Gromiha MM, Ponnuswamy PK. 1993. Relationship between amino acid properties and protein compressibility. *Journal of theoretical biology* 165(1):87-100. doi: [10.1006/jtbi.1993.1178](https://doi.org/10.1006/jtbi.1993.1178)

33. Guerrero AD, Schmitz I, Chen M, Wang J. 2012. Promotion of caspase activation by caspase-9-mediated feedback amplification of mitochondrial damage. *Journal of clinical & cellular immunology* 3(3). doi: [10.4172/2155-9899.1000126](https://doi.org/10.4172/2155-9899.1000126)

34. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* 59(3):307-321. doi: [10.1093/sysbio/syq010](https://doi.org/10.1093/sysbio/syq010)

35. Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* 52(5):696-704. doi: [10.1080/10635150390235520](https://doi.org/10.1080/10635150390235520)

36. Harris RC, Pettitt B. 2016. Reconciling the understanding of 'hydrophobicity' with physics-based models of proteins. *Journal of Physics: Condensed Matter* 28(8):083003. doi: [10.1088/0953-8984/28/8/083003](https://doi.org/10.1088/0953-8984/28/8/083003)

37. Hong P, Koza S, Bouvier ES. 2012. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of liquid chromatography & related technologies* 35(20):2923-2950. doi: [10.1080/10826076.2012.743724](https://doi.org/10.1080/10826076.2012.743724)

38. Huber KL, Hardy JA. 2012. Mechanism of zinc-mediated inhibition of caspase-9. *Protein Science* 21(7):1056-1065. doi: [10.1002/pro.2090](https://doi.org/10.1002/pro.2090)

39. Karshikoff A, Ladenstein R. 1998. Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein engineering* 11(10):867-872. doi: [10.1093/protein/11.10.867](https://doi.org/10.1093/protein/11.10.867)

40. Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14):3059-3066. doi: [10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436)

41. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4):772-780. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010)

42. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199)

43. Kerr JF, Wyllie AH, Currie AR. 1972. Apoptosis: a basic biological phenomenon with wideranging implications in tissue kinetics. *British journal of cancer*: 26(4):239-257. doi: [10.1038/bjc.1972.33](https://doi.org/10.1038/bjc.1972.33)

44. Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276-3278. doi: [10.1093/bioinformatics/btu531](https://doi.org/10.1093/bioinformatics/btu531)

45. Lauber K, Appel HA, Schlosser SF, Gregor M, Schulze-Osthoff K, Wesselborg S. 2001. The adapter protein apoptotic protease-activating factor-1 (Apaf-1) is proteolytically processed during apoptosis. *Journal of Biological Chemistry* 276(32):29772-29781. doi: [10.1074/jbc.M101524200](https://doi.org/10.1074/jbc.M101524200)

46. Lettre G, Hengartner MO. 2006. Developmental apoptosis in *C. elegans*: a complex CEDNario. *Nature Reviews Molecular Cell Biology* 7(2):97-108. doi: [10.1038/nrm1836](https://doi.org/10.1038/nrm1836)

47. Maldonado E, Sunagar K, Almeida D, Vasconcelos V, Antunes A. 2014. IMPACT\_S: integrated multiprogram platform to analyze and combine tests of selection. *PloS one* 9(10):e96243. doi: [10.1371/journal.pone.0096243](https://doi.org/10.1371/journal.pone.0096243)

48. McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Salsbury JK. 2005. Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Molecular biology and evolution* 22(3):437-455. doi: [10.1093/molbev/msi028](https://doi.org/10.1093/molbev/msi028)

49. Menze MA, Fortner G, Nag S, Hand SC. 2010. Mechanisms of apoptosis in Crustacea: what conditions induce versus suppress cell death?. *Apoptosis* 15(3):293-312. doi: [10.1007/s10495-009-0443-6](https://doi.org/10.1007/s10495-009-0443-6)

50. Milisav I, Šuput D. 2007. Procaspsase-9 is attached to the mitochondrial outer membrane in the early stages of apoptosis. *Cellular and Molecular Biology Letters* 12(4):509-522. doi: [10.2478/s11658-007-0020-3](https://doi.org/10.2478/s11658-007-0020-3)

51. Moya A, Sakamaki K, Mason BM, Huisman L, Forêt S, Weiss Y, Bull TE, Tomii K, Imai K, Hayward DC, Ball EE. 2016. Functional conservation of the apoptotic machinery from coral to man: the diverse and complex Bcl-2 and caspase repertoires of *Acropora millepora*. *BMC genomics* 17(1):62. doi: [10.1186/s12864-015-2355-x](https://doi.org/10.1186/s12864-015-2355-x)

52. Oteri F, Nadalin F, Champeimont R, Carbone A. 2017. BIS2Analyzer: a server for co-evolution analysis of conserved protein families. *Nucleic acids research* 45(W1):W307-W314. doi: [10.1093/nar/gkx336](https://doi.org/10.1093/nar/gkx336)

53. Pace C N, Shirley B A, Thomson J A. 1989. Measuring the conformational stability of a protein In: Creighton, TE editor. *Protein Structure, a practical approach*. IRL Press, Oxford. Pp 311-330.

54. Pál C, Papp B, Lercher MJ. 2006. An integrated view of protein evolution. *Nature Reviews Genetics* 7(5):337-348. doi: [10.1038/nrg1838](https://doi.org/10.1038/nrg1838)

55. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. 1997. Correlated mutations contain information about protein-protein interaction. *Journal of molecular biology* 271(4):511-523. doi: [10.1006/jmbi.1997.1198](https://doi.org/10.1006/jmbi.1997.1198)

56. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25(13):1605-1612. doi: [10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084)

57. Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* 25(7):1253-1256. doi: [10.1093/molbev/msn083](https://doi.org/10.1093/molbev/msn083)

58. Prahakaran M, Ponnuswamy PK. 1979. The spatial distribution of physical, chemical, energetic and conformational properties of amino acid residues in globular proteins. *Journal of theoretical biology* 80(4):485-504. doi: [10.1016/0022-5193\(79\)90090-0](https://doi.org/10.1016/0022-5193(79)90090-0)

59. Qi S, Pang Y, Hu Q, Liu Q, Li H, Zhou Y, He T, Liang Q, Liu Y, Yuan X, Luo G. 2010. Crystal structure of the *Caenorhabditis elegans* apoptosome reveals an octameric assembly of CED-4. *Cell* 141(3):446-457. doi: [10.1016/j.cell.2010.03.017](https://doi.org/10.1016/j.cell.2010.03.017)

60. Renatus M, Stennicke HR, Scott FL, Liddington RC, Salvesen GS. 2001. Dimer formation drives the activation of the cell death protease caspase 9. *Proceedings of the National Academy of Sciences* 98(25):14250-14255. doi: [10.1073/pnas.231465798](https://doi.org/10.1073/pnas.231465798)

61. Reubold TF, Wohlgemuth S, Eschenburg S. 2011. Crystal structure of full-length Apaf-1: how the death signal is relayed in the mitochondrial pathway of apoptosis. *Structure* 19(8):1074-1083. doi: [10.1016/j.str.2011.05.013](https://doi.org/10.1016/j.str.2011.05.013)

62. Richmond TJ, Richards FM. 1978. Packing of  $\alpha$ -helices: Geometrical constraints and contact areas. *Journal of molecular biology* 119(4):537-555. doi: [10.1016/0022-2836\(78\)90201-2](https://doi.org/10.1016/0022-2836(78)90201-2)

63. Riedl SJ, Li W, Chao Y, Schwarzenbacher R, Shi, Y. 2005. Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature* 434(7035):926-933. doi: [10.1038/nature03465](https://doi.org/10.1038/nature03465)

64. Robertson AJ, Croce J, Carbonneau S, Voronina E, Miranda E, McClay DR, Coffman, J.A. 2006. The genomic underpinnings of apoptosis in *Strongylocentrotus purpuratus*. *Developmental biology* 300(1):321-334. doi: [10.1016/j.ydbio.2006.08.053](https://doi.org/10.1016/j.ydbio.2006.08.053)

65. Romero A, Estevez-Calvar N, Dios S, Figueras A, Novoa B. 2011. New insights into the apoptotic process in mollusks: characterization of caspase genes in *Mytilus galloprovincialis*. *PLoS One* 6(2):e17003. doi: [10.1371/journal.pone.0017003](https://doi.org/10.1371/journal.pone.0017003)

66. Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I. 2019. GenBank. *Nucleic acids research* 47(D1):D94-D99. doi: [10.1093/nar/gky989](https://doi.org/10.1093/nar/gky989)

67. Shalaeva DN, Dibrova DV, Galperin MY, Mulkidjanian AY. 2015. Modeling of interaction between cytochrome c and the WD domains of Apaf-1: bifurcated salt bridges underlying apoptosome assembly. *Biology direct* 10(1):29. doi: [10.1186/s13062-015-0059-4](https://doi.org/10.1186/s13062-015-0059-4)

68. Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular biology and evolution* 16(8):1114-1114. doi: [10.1093/oxfordjournals.molbev.a026201](https://doi.org/10.1093/oxfordjournals.molbev.a026201)

69. Shindyalov IN, Kolchanov NA, Sander C. 1994. Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? *Protein Engineering, Design and Selection* 7(3):349-358. doi: [10.1093/protein/7.3.349](https://doi.org/10.1093/protein/7.3.349)

70. Sonnhammer EL, Eddy SR, Durbin R. 1997. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function, and Bioinformatics* 28(3):405-420. doi: [10.1002/\(sici\)1097-0134\(199707\)28:3<405::aid-prot10>3.0.co;2-1](https://doi.org/10.1002/(sici)1097-0134(199707)28:3<405::aid-prot10>3.0.co;2-1)

71. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28(18):i409-i415. doi: [10.1093/bioinformatics/bts386](https://doi.org/10.1093/bioinformatics/bts386)

72. Suel GM, Lockless SW, Wall MA, Ranganathan R. 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature structural biology* 10(1):59-69. doi: [10.1038/nsb881](https://doi.org/10.1038/nsb881)

73. Szklarczyk R, Wanschers BF, Cuypers TD, Esseling JJ, Riemersma M, van den Brand MA, Gloerich J, Lasonder E, van den Heuvel LP, Nijtmans LG, Huynen MA. 2012. Iterative orthology prediction uncovers new mitochondrial proteins and identifies C12orf62 as the human ortholog of COX14, a protein involved in the assembly of cytochrome c oxidase. *Genome biology* 13(2):R12. doi: [10.1186/gb-2012-13-2-r12](https://doi.org/10.1186/gb-2012-13-2-r12)

74. Taylor WR, Hatrik K. 1994. Compensating changes in protein multiple sequence alignments. *Protein Engineering, Design and Selection* 7(3):341-348. doi: [10.1093/protein/7.3.341](https://doi.org/10.1093/protein/7.3.341)

75. Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, Pruitt KD, Souvorov A. 2016. P8008 The NCBI Eukaryotic Genome Annotation Pipeline. *Journal of Animal Science* 94(4):184. doi: [10.2527/jas2016.94supplement4184x](https://doi.org/10.2527/jas2016.94supplement4184x)

76. Vörös J. 2004. The density and refractive index of adsorbing protein layers. *Biophysical journal* 87(1):553-561. doi: [10.1529/biophysj.103.030072](https://doi.org/10.1529/biophysj.103.030072)

77. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191. doi: [10.1093/bioinformatics/btp033](https://doi.org/10.1093/bioinformatics/btp033)

78. Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics* 19(5):671-672. doi: [10.1093/bioinformatics/btg043](https://doi.org/10.1093/bioinformatics/btg043)

79. Wu CC, Bratton SB. 2017. Caspase-9 swings both ways in the apoptosome. *Molecular & cellular oncology* 4(2):e1281865. doi: [10.1080/23723556.2017.1281865](https://doi.org/10.1080/23723556.2017.1281865)

80. Wu CC, Lee S, Malladi S, Chen MD, Mastrandrea NJ, Zhang Z, Bratton SB. 2016. The Apaf-1 apoptosome induces formation of caspase-9 homo- and heterodimers with distinct activities. *Nature communications* 7(1):1-14. doi: [10.1038/ncomms13565](https://doi.org/10.1038/ncomms13565)

81. Yu X, Wang L, Acehan D, Wang X, Akey CW. 2006. Three-dimensional structure of a double apoptosome formed by the Drosophila Apaf-1 related killer. *Journal of molecular biology* 355(3):577-589. DOI: [10.1016/j.jmb.2005.10.040](https://doi.org/10.1016/j.jmb.2005.10.040)

82. Yuan J, Shaham S, Ledoux S, Ellis HM, Horvitz HR. 1993. The *C. elegans* cell death gene *ced-3* encodes a protein similar to mammalian interleukin-1 $\beta$ -converting enzyme. *Cell*, 75(4):641-652. doi: [10.1016/0092-8674\(93\)90485-9](https://doi.org/10.1016/0092-8674(93)90485-9)
83. Yuan S, Yu X, Topf M, Ludtke SJ, Wang X, Akey CW. 2010. Structure of an apoptosome-procaspase-9 CARD complex. *Structure* 18(5):571-583. doi: [10.1016/j.str.2010.04.001](https://doi.org/10.1016/j.str.2010.04.001)
84. Zhou M, Li Y, Hu Q, Bai XC, Huang W, Yan C, Scheres SH, Shi Y. 2015. Atomic structure of the apoptosome: mechanism of cytochrome c-and dATP-mediated activation of Apaf-1. *Genes & development* 29(22):2349-2361. doi: [10.1101/gad.272278.115](https://doi.org/10.1101/gad.272278.115)
85. Zimmerman JM, Eliezer N, Simha R. 1968. The characterization of amino acid sequences in proteins by statistical methods. *Journal of theoretical biology* 21(2):170-201. doi: [10.1016/0022-5193\(68\)90069-6](https://doi.org/10.1016/0022-5193(68)90069-6)
86. Zmasek CM, Godzik A. 2013. Evolution of the animal apoptosis network. *Cold Spring Harbor perspectives in biology* 5(3):a008649. doi: [10.1101/cshperspect.a008649](https://doi.org/10.1101/cshperspect.a008649)
87. Zmasek CM, Zhang Q, Ye Y, Godzik A. 2007. Surprising complexity of the ancestral apoptosis network. *Genome biology* 8(10):R226. doi: [10.1186/gb-2007-8-10-r226](https://doi.org/10.1186/gb-2007-8-10-r226)