Supplementary Figures

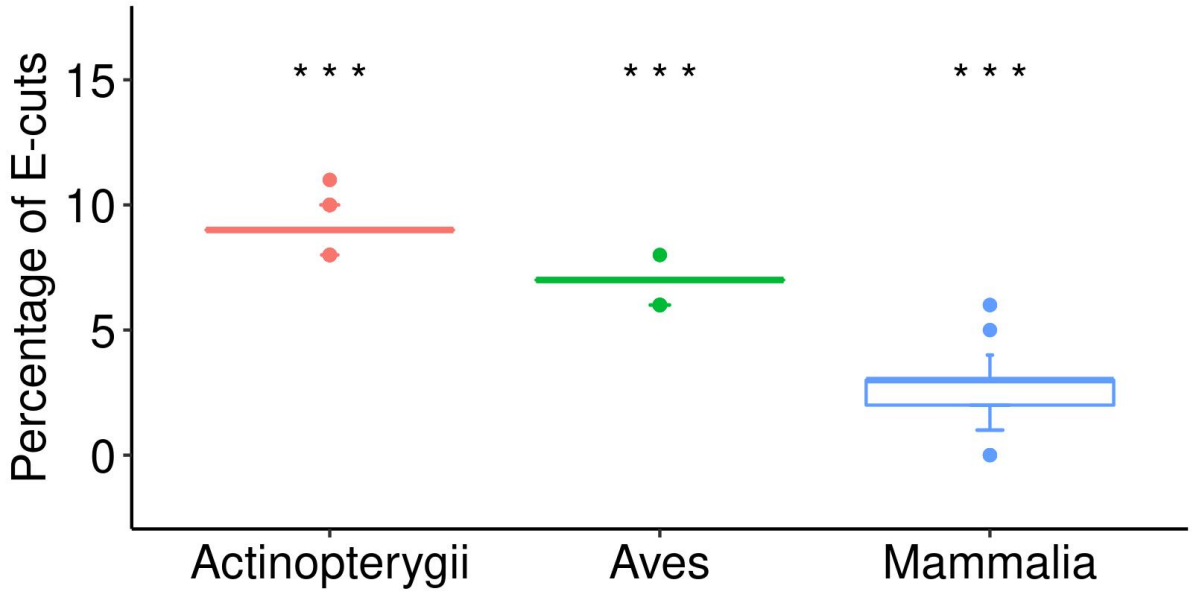
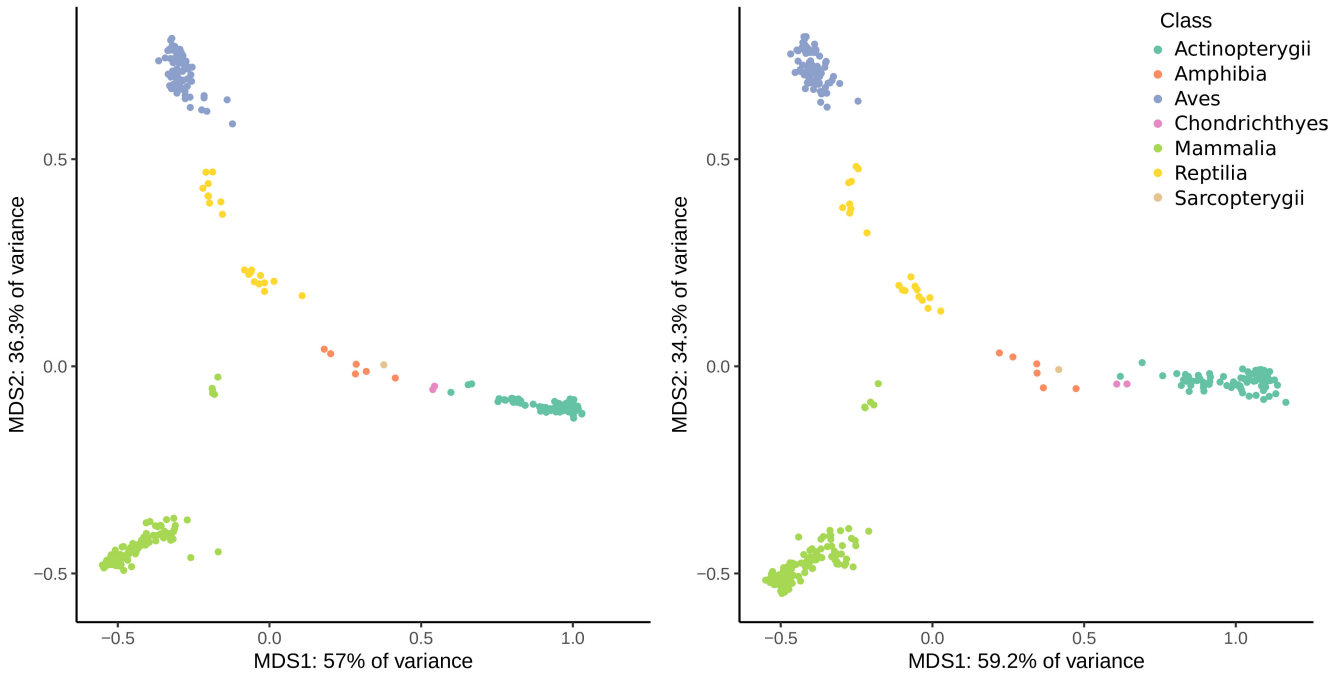


Figure S1. Distribution of P1 glutamate cleavage sites in three vertebrate classes**.** The percentage of glutamate in the P1 position of orthologous caspase targets was calculated for each of 328 Vertebrate species (Table S3), and distribution of these values among species was plotted for three classes: Actinopterygii, Aves and Mammals. Chondrichthyes, Sarcopterygii, Amphibia and Reptilia were excluded from this analysis because of incomparably small number of species per a class (1-20) in respect to the three former classes (82-132). Statistical significance was evaluated using ANOVA followed by Tukey's post-hoc test. Asterisks indicate significance levels *** $p < 0.001$.



(a) **(b)**

Figure S2. Principal coordinates analysis (PCoA) of cleavage sites and 60 amino acid sequences. A phylogenetic analysis was performed on cleavage sites and orthologous sequences obtained from pBLAST using distance matrices, for all species with well-represented proteomes, including humans. The distance between each pair of species was calculated as a pairwise comparison of common elements: cleavage sites (8 amino acids) or full query and hit sequences (60 amino acids). To compare species by cleavage sites, we calculated average Hamming distances. For long sequence comparison, we first performed multiple alignments and then calculated average Hamming distances on aligned sequences. In both cases, the obtained values were normalized by the maximum length of the sequence (8 and 60 respectively). As a result, two distance matrices, where each cell corresponds to comparison between two organisms, were obtained: for cleavage sites and for 60 amino acid orthologous sequences. Visualization of differences between species was performed using the principal coordinate analysis (PCoA) in R software. (**a**) PCoA for 60 amino acid sequences (93.3% of variance are explained by Multidimensional Scale 1 and Multidimensional Scale 2). (**b**) PCoA for 8 amino acid cleavage sites (93.5% of variance are explained by Multidimensional Scale 1 and Multidimensional Scale 2). The same distance matrices were used to plot dendrograms by hierarchical clustering (Figures S3 and S4).

Figure S3. Hierarchical clustering of cleavage sites. A phylogenetic analysis was performed on cleavage sites using distance matrices, for all species with well-represented proteomes, including humans. The distance between each pair of species was calculated as a pairwise comparison of 8 amino acids cleavage sites using average Hamming distances. The obtained values were normalized by the maximum length of the sequence (8 amino acids). As a result, we acquired a distance matrix, where each cell corresponds to a comparison between two organisms. The same matrix was used to visualize differences between species by principal coordinates analysis (Figure S2). The dendrogram was plotted using R software. Colors represent seven classes of vertebrates: purple – Mammalia, blue – Aves, orange – Reptilia, green – Amphibia, black – Sarcopterygii, red – Actinopterygii, yellow – Chondrichthyes. Each “leaf” of the tree corresponds to a species, the order is given in parentheses.

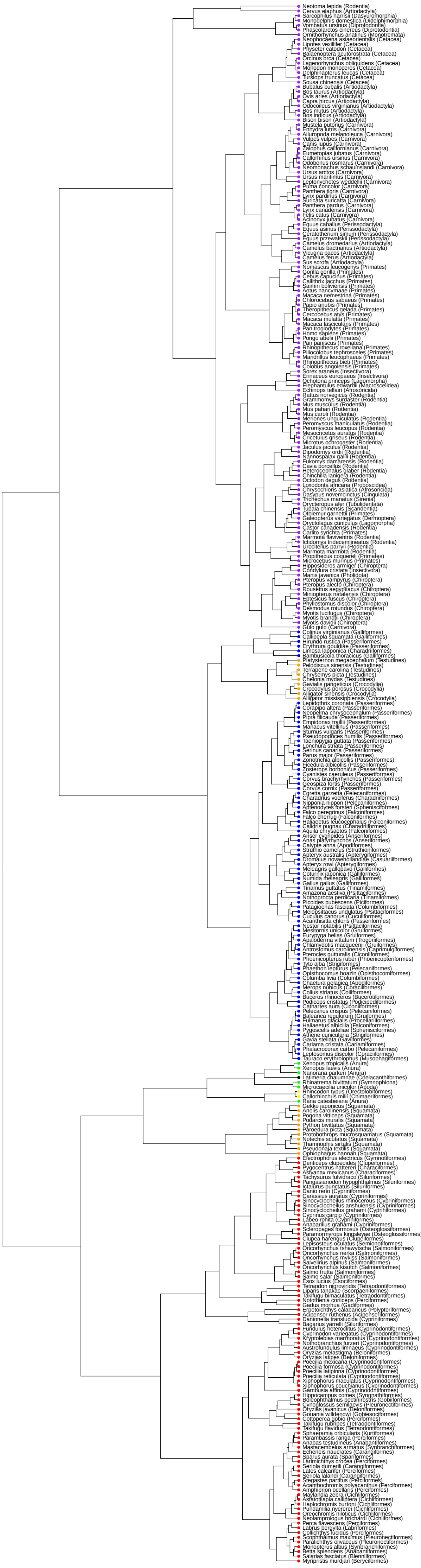


Figure S4. Hierarchical clustering of 60 amino acid sequences. Hierarchical clustering was performed on the distance matrix calculated for PcoA, where each cell corresponds to a comparison between two organisms. A phylogenetic analysis was performed on orthologous sequences obtained from pBLAST using distance matrices, for all species with well-represented proteomes, including humans. The distance between every pair of species was calculated as a pairwise comparison of 60 amino acid queries and hit sequences. To compare these sequences, we first performed multiple alignments and then calculated average Hamming distances on aligned sequences. The obtained values were normalized by the maximum length of the sequence (60 amino acids). As a result, we acquired a distance matrix, where each cell corresponds to a comparison between two organisms. The same matrix was used to visualize differences between species by principal coordinates analysis (Figure S2). The dendrogram was plotted using R software. Colors represent seven classes of vertebrates: purple – Mammalia, blue – Aves, orange – Reptilia, green – Amphibia, black – Sarcopterygii, red – Actinopterygii, yellow – Chondrichthyes. Each “leaf” of the tree corresponds to a species, the order is given in parentheses.

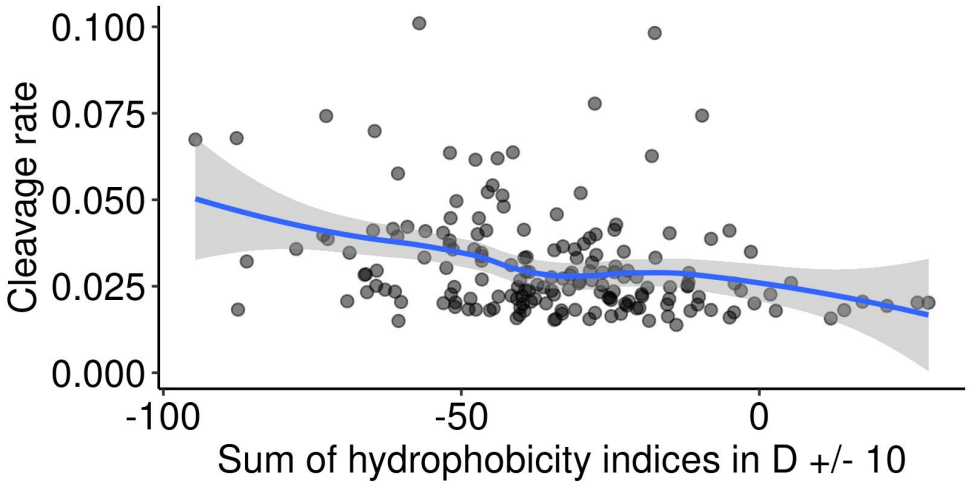


Figure S5. Correlation between hydrophobicity of cleavage site surroundings and cleavage rates for human caspase targets. The cleavage rate is estimated as 1 divided by the elution time of the caspase cleavage product from the chromatographic column loaded with E coli lysate digested by caspase 3 (Agard et al., 2012). The elution time is inversely proportional to the cleavage rate. Therefore, hydrophilic caspase cleavage sites tend to be cut faster than hydrophobic ones. (Kendall tau coefficient = -0.233, p-value = 1.082173e-05). Each point represents one cleavage site (n=176).

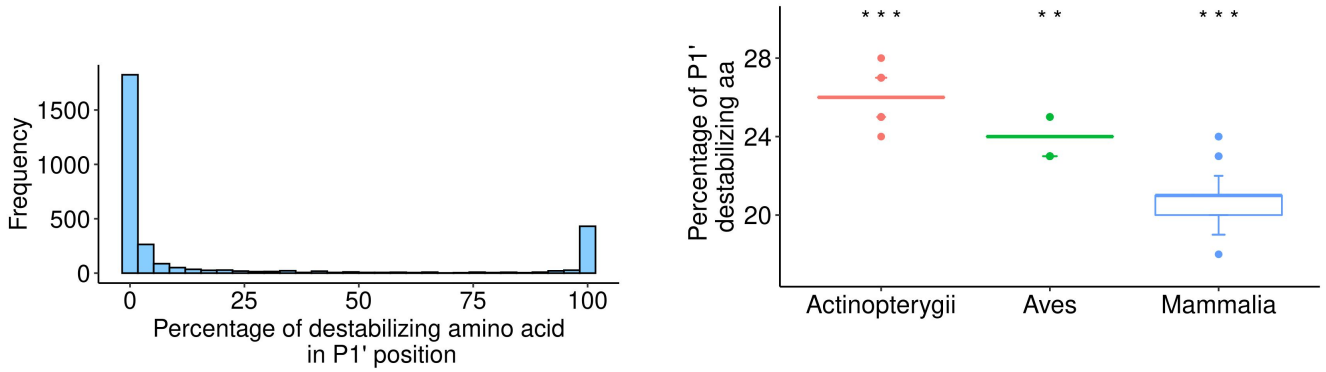


Figure S6. Distribution of P1' destabilizing amino acid residues among cleavage sites and among vertebrates. (**a**) Distribution of the percentage of destabilizing amino acids in the P1' position for each cleavage site. (**b**) The percentage of orthologous targets with destabilizing amino acid in the P1' position was calculated for each of 328 Vertebrate species (Table S7), and the distribution of these values among species was plotted for three classes: Actinopterygii, Aves and Mammals. Chondrichthyes, Sarcopterygii, Amphibia and Reptilia were excluded from this analysis because of incomparably small number of species per a class (1-20) in respect to to the three former classes (82-132). Evaluation of statistical significance was performed using ANOVA followed by Tukey's post-hoc test. Asterisks indicate significance levels ** $p < 0.01$. *** $p < 0.001$.