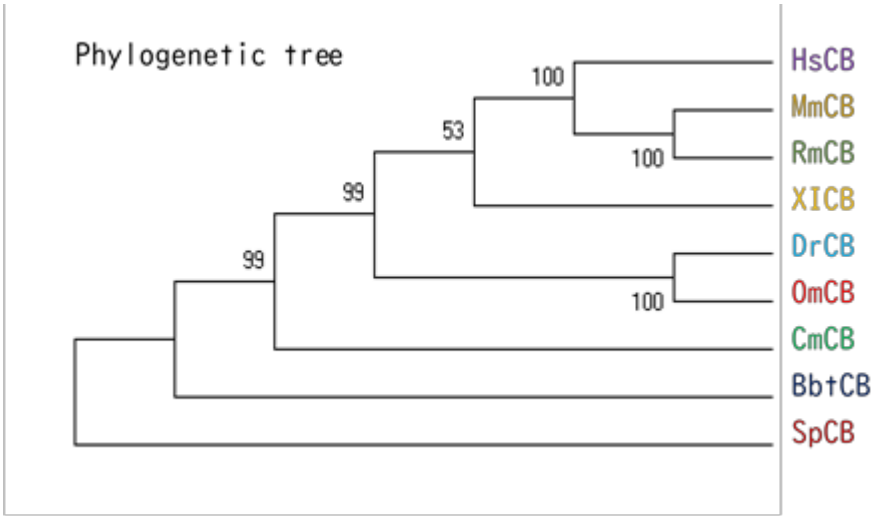

Figure 1: The unrooted phylogenetic tree of cathepsin B from various species constructed by the neighbor-joining method in MEGApackage. Bootstrap majority consensus values on 1000 replicates are indicated at each branch point in percent. Sequence obtained from NCBI are: Human, *Homo sapiens* (**HsCB**: NP 001899.1) ; Mouse, *Mus musculus* (**MmCB**: NP 031824.1) ; Rat, *Rattus norvegicus* (**RnCB**: NP 072119.2) ; African clawed frog, *Xenopus laevis* (**XlCB**: NP 001079570.1) ; Zebrafish, *Danio rerio* (**DrCB**: NP 998501.1) ; Rainbow trout, *Oncorhynchus mykiss* (**OmCB**: XP 021441036.1) ; Elephant shark, *Callorhinchus milii* (**CmCB**: XP 007882610.1) ; Lancelet, *Branchiostoma belcheri tsingtauense* (**BbtCB**: AAQ83887.1) Sea Urchin, *Strongylocentrotus purpuratus* (**SpCB**: XP_787947.3, XP 011661731.1, undeleted connection)

Sequence	# #	#	#
HsCB	[95].IKEI RDQGSCGSCW AFGAVEAI.	[77].RPPCTGE.	[70].GEMMGGH
MmCB	[95].IGQI RDQGSCGSCW AFGAVEAI.	[77].RPPCTGE.	[70].GDMMGGH
RmCB	[95].IAQI RDQGSCGSCW AFGAVEAM.	[77].RPPCTGE.	[70].GDVMGGH
XlCB	[95].IREI RDQGSCGSCW AFGAVEAI.	[77].RPACKGE.	[71].GEELGGH
DrCB	[94].LKEI RDQGSCGSCW AFGAAEAI.	[76].RPPCSGE.	[71].GSPVGGH
OmCB	[94].LKEI RDQGSCGSCW AFGAAEAI.	[76].RPPCTGE.	[71].GSAVGGH
CmCB	[93].TRQI RDQGSCGSCW AFAAVGAI.	[77].RPACSGE.	[70].GEMLGGH
BbtCB	[95].IKEV RDQGSCGSCW ALAAVEAM.	[76].RPACGKL.	[70].GAELGGH
SpCB	[93].IKEV RDQGSCGSCW AFGAVEAI.	[76].KGPCQGE.	[70].GGVLGGH

Figure 2: The results of the sequence alignment used to construct the phylogenetic tree in Figure 1. Because the cathepsin B sequence is too long, the amino acids far from the protein action site are omitted by using brackets and numbers, and the amino acids at the end of the sequence are also omitted. The black letters indicate that the locus is highly conserved, and the letters of different colors represent different kinds of amino acids. The pound sign is used to label the action sites according to the results of molecular docking.

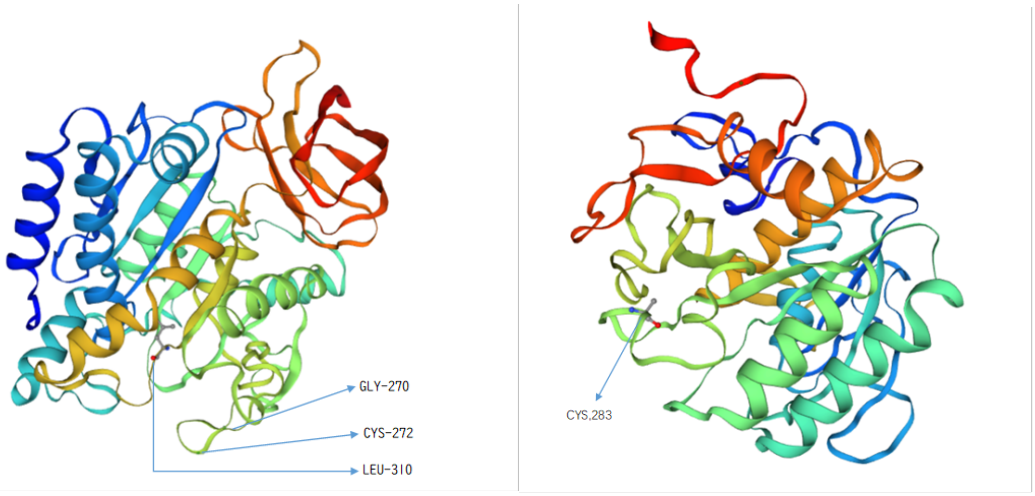

Figure 3 : Three-dimensional modeling of carboxypeptidase Z/N on the left and subtilisin-like protease on the right

Result 1:

The Location of Lancelet on the Species Evolution Map

Because there are more kinds of amino acids than bases, the similarity of amino acid sequences can show the relationship between species to a certain extent. For example, the 67 and 68 amino acids in the amino acid sequence are PN in lancelet and sea urchin, and GP in other species. Similarly, the 106 to 109 amino acids are IKEV in cathepsin B sequence of lancelet, which is highly consistent with sea urchin, but has low similarity with other species. Therefore, lancelet has close evolutionary relationship with invertebrates such as sea urchins. On the other hand, at some amino acid sites, such as 60 and 61 amino acids, sea urchins showed ML, but LC in other species; as 204 to 206 amino acids, TKG in sea urchins and SRP in other species. To some extent, this reflects the relationship between lancelet and vertebrates. Therefore, we infer that amphioxus is a transitional animal between invertebrates and invertebrates on the phylogenetic map by sequence alignment of cathepsin B.

Result 2:

Inference of the reason why cathepsin B can decompose algae toxin in amphioxus cells

The interaction between cathepsin B and microcystin LR is simulated by molecular docking (microcystin LR and microcystin RR are the two main microcystins). The results of molecular docking show that the interaction sites between cathepsin B and microcystin LR were GLY-90, CYX-92, CYX-184, and LEU-261, respectively. The sequence alignment results above were labeled with #. Three of them were highly conserved in the evolution of species, and the amino acid LEU-261 was conserved in the alignment results. Though they were different amino acids, they still belonged to the same class. Although the action sites are highly conserved, the alignment results show that there are differences in the amino acid sequence environment around the action sites. Compared with other species, there are some unique amino acid fragments around the cathepsin B site of lancelet, which are labeled with grey background. We infer that it is these fragments that influence

the spatial structure around the action sites that enable cathepsin B in amphioxus to interact with algae toxin, capture and complete the digestion of algae toxin.

At the same time, by comparing the structures of microcystin LR and microcystin RR, the structural differences between the two kinds of microcystins are only one branch far from benzene ring. The results of molecular docking show that the amino acid sites of cathepsin B interacting with algae toxin generally concentrate around the common branched chains of the two algae toxins, so we judged that the structural differences between the two algae toxins had no decisive effect on the recognition and interaction of the two enzymes and algae toxins. It is also consistent with our experimental results that cathepsin B can independently decompose microcystin LR and microcystin RR.

Result 3:

Study on the Action Mechanism of Carxypeptidase Z/N and Subilisin-like Protease

In the wet lab experiments, we found that carboxypeptidase Z/N and subtilisin-like protein could not decompose microcystins independently, but when the two enzymes were put into the experiment, they could decompose one of the microcystins, microcystin RR. By comparing with the sequence of cathepsin B (Fig. 3), we found that several action sites of cathepsin B could find similar short peptide structures in these two amino acid sequences. GLY-90, CYX-92 and LEU-261 were found in carboxypeptidase Z/N, while CYX-184 was found in subtilisin-like protease. We speculate that both enzymes are involved in the capture and decomposition of microcystin RR at the same time. However, due to the harsh external environment produced by the simultaneous action of the two enzymes, the three-dimensional structure of the acting algal toxin substrates is limited, so the two enzymes can only decompose microcystin RR and cannot interact with microcystin LR.