Integrated Proteomics Network Analysis and Phylogenetic Analysis on Small Spotted Catshark CD8 Alpha

Avantika Sanjay[1], Uma Kumari[2]
[1]*Project Trainee at Bioinformatics Project and Research Institute, Noida - 201301, India*
[2]*Senior Bioinformatics Scientist, Bioinformatics Project and Research Institute, Noida - 201301, India*

***Abstract*- The aim of this project is to comprehensively analyze the CD8α protein of the small spotted catshark (Scyliorhinuscanicula) using integrated proteomic network analysis and phylogenetic techniques. Using bioinformatics tools such as RasMol, PyMOL, InterPro, Jalview, T-Coffee, SAVES Server, PROCHECK, ERRAT, COBALT and BioPython etc we performed sequence scanning, structure visualization, network analysis and modeling. validation Our findings revealed conserved regions and critical functional motifs of the CD8α protein, highlighted by the evolutionary conservation observed in multiple sequence alignments and phylogenetic analysis. Hydrogen bond analysis and molecular weight calculations provided additional structural insights, while network analysis highlighted potential protein-protein interactions. Structural models validated using ERRAT and Ramachandran plots confirmed the accuracy and reliability of our predictions. This integrated approach not only advances our understanding of CD8α in the small-spotted catshark, but also sets a precedent for similar studies in other species, making important contributions to the fields of immunology and developmental biology**

Keywords: Homology modelling, Sequence similarity, CD8 alpha, Sequence Alignment, Errat, Biopython

I. INTRODUCTION

Cartilaginous fish are the oldest living group of animals with an adaptive immune system based on immunoglobulins (Ig), T-cell receptors (TCR), and Major Histocompatibility Complex (MHC) molecules [1]. This system forms the cornerstone of vertebrate immunity, providing both specificity and memory in immune responses. During the 1960s, pivotal studies on fish immune responses to antibodies were conducted, with lampreys and sharks being the first species examined [2]. These early investigations laid the foundation for subsequent research into the unique aspects of the cartilaginous fish immune system [3].

Shark immunoglobulin (Ig) genes, for instance, are organized in a distinctive cluster configuration, as opposed to the translocon organization found in mammals and other vertebrates. In this cluster configuration, each cluster contains variable (V), diversity (D), joining (J), and constant (C) gene segments [4]. This unique organization suggests a divergent evolutionary pathway in the development of the immune system in sharks. Additionally, sharks possess conventional TCRs of both alpha/beta and gamma/delta types, mirroring the TCR translocon organization found in higher vertebrates [5]. Unlike higher vertebrates, cartilaginous fish lack lymph nodes, thus adaptive immune responses occur primarily in the spleen, with evidence of lymphocytic activity also observed in the gills, intestine, and other tissues [6,7].

Cartilaginous fish utilize recombination mechanisms akin to those in mammals to generate their primary Ig repertoire. This involves recombination-activating genes (RAG1 and RAG2) functioning within lymphopoietic organs (such as the epigonal and Leydig organs), and terminal deoxynucleotidyl transferase (TdT) introducing nucleotide diversity at junctions during rearrangement [8]. However, recombination is predominantly confined within a single cluster, with inter-cluster recombination being a rare occurrence [9].

Orthologs of CD8 alpha (CD8α) and CD8 beta (CD8β), referred to as ScCD8α and ScCD8β, have been identified in the small-spotted catshark

(Scyliorhinus canicula). Both ScCD8α and ScCD8β feature an extracellular immunoglobulin superfamily (IgSF) V domain, consistent with previously characterized CD8 proteins in other species [10]. CD8+ cytotoxic T cells, expressing CD8α, are pivotal in the adaptive immune response, crucially limiting the proliferation of intracellular pathogens like viruses by targeting and killing infected cells. These cells are activated by peptides presented by MHC class I molecules (p/MHC I) and secrete cytotoxins such as perforin and granzymes to induce apoptosis in target cells. The CD8αα-p/MHC I complexes interact with TCR/CD3 complexes, facilitating signal transduction in T cells. CD8 acts as a co-receptor, stabilizing the interaction between the TCR and MHC I on antigen-presenting cells (APCs), ensuring effective T cell activation [11,12].

Structurally, CD8α and CD8β are related, with their encoding genes being tandemly linked in both mammalian and teleost genomes [13]. Although CD8αα and CD8αβ bind to p/MHC I with similar affinity and both are recruited to the immunological synapse, their expression patterns differ among immune cell types. CD8α is expressed on γδ T cells, NK cells, certain dendritic cell subsets, and intestinal intraepithelial lymphocytes, whereas CD8αβ expression is restricted to αβ T cells. Notably, ScCD8α can form homodimers or heterodimers with ScCD8β. The ScCD8αα homodimer, like its mammalian counterparts, is primarily stabilized by a hydrophobic core, forming a dimeric structure. Moreover, ScCD8αα possesses a canonical cavity that mediates interactions with p/MHC I across various species [14,15].

In this study, we employ an integrated approach combining proteomic network analysis and phylogenetic techniques to analyze the CD8α protein of the small-spotted catshark. Using an array of bioinformatics tools such as RasMol, PyMOL, InterPro, Jalview, T-Coffee, SAVES Server, PROCHECK, ERRAT, COBALT, and BioPython, we perform sequence scanning, structure visualization, network analysis, and model validation. This comprehensive analysis aims to elucidate conserved regions and critical functional motifs within the CD8α protein, providing insights into its structural and functional roles. Additionally, our phylogenetic analysis seeks to contextualize the evolutionary conservation observed in multiple sequence alignments. This study not only enhances our understanding of CD8α in the small-spotted catshark but also contributes valuable knowledge to the broader fields of immunology and developmental biology.

II. MATERIALS AND METHODS

The amino acid sequence of the small-spotted catshark CD8α protein was retrieved from public databases. InterPro was used to scan the sequence, identifying conserved domains, motifs, and functional regions of the protein. InterPro integrates multiple predictive models to provide a comprehensive functional analysis and classification. Multiple sequence alignments were performed using Jalview, T-Coffee, and ClustalW to examine evolutionary conservation and relationships. Jalview facilitated interactive visualization and editing of alignments, while T-Coffee and ClustalW provided robust alignment algorithms. These tools allowed us to identify conserved regions and infer evolutionary relationships between different species. To construct the CD8α protein interaction network, the STRING database, a repository of known and predicted protein-protein interactions, was utilized. This analysis revealed potential interaction partners and functional pathways involving CD8α in the small-spotted catshark, underscoring its role in the immune system. For structural modeling and visualization, RasMol and PyMOL were employed. These tools enabled the creation of detailed three-dimensional models, visualization of the spatial arrangement of amino acids, and analysis of structural features. Structural information from the Protein Data Bank (PDB) and the Molecular Modeling Database (MMDB) was used as a benchmark for our models, facilitating comparative analysis and validation of the CD8α protein structure. Several validation tools were used to ensure the accuracy and reliability of the structural models. The SAVES server, which includes multiple validation tools, provided a comprehensive assessment of model quality. PROCHECK evaluated the stereochemical properties of the models, ERRAT assessed overall quality, and PDBSum offered detailed analyses of structural properties and comparisons with known structures. These validation steps confirmed the accuracy and reliability of our structural models.

Constraint-based alignment was performed using COBALT, which allowed the incorporation of additional biological information into the sequence alignment. This enhanced the accuracy of our evolutionary analyses and provided a more detailed view of the evolutionary history of the CD8α protein. BioPython, a versatile bioinformatics library, was employed to automate various data processing and analysis tasks. This tool streamlined our workflow, ensuring reproducibility and efficiency in handling large datasets and performing complex analyses [16,17,18,19,20,21,22,23,24,25,26].

III. RESULT AND DISCUSSION

Using RasMol, we visualized the three-dimensional structure of the small spotted catshark (PDB ID: 8HXS) CD8α protein. This tool allowed us to study the detailed spatial conformation of the protein, including the arrangement of its alpha helices, beta sheets, and loops. In addition, we focused on identifying the hydrogen bonds within the structure that are crucial to maintaining its stability and integrity. By highlighting these hydrogen bonds, we gained insight into the interactions that stabilize the three-dimensional shape of the protein and their possible functional implications for the small-spotted catshark's immune system. This analysis provided the basis for subsequent structural validation and functional predictions. It can also be used for active site identification for cartilaginous fish (8HXS).

Fig.1 H-bond on 63 (in sample 8HXS)

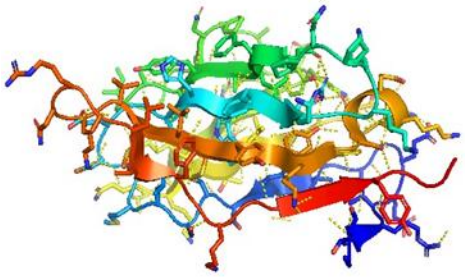

Fig.2 Active site identification

Polar contacts can show valuable insights into molecular interactions. Polar contacts are pivotal in molecular interactions by facilitating hydrogen bonds and electrostatic interactions between biomolecules. Hydrogen bonds, formed between polar groups like hydroxyl and amino residues, stabilize protein structures and enhance specificity in molecular recognition. Electrostatic interactions, such as salt bridges between oppositely charged groups, also contribute to protein stability and ligand binding affinity. Understanding these polar interactions provides crucial insights into protein function, molecular recognition processes, and drug design strategies, offering pathways for advancing biomedical and biotechnological applications.

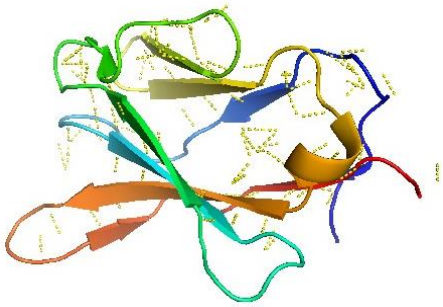

Fig.3 Polar contacts

We utilized PyMOL to visualize the helices and loops within the structure of the CD8α protein from the small-spotted catshark. PyMOL is a powerful molecular visualization tool that allows us to generate detailed three-dimensional representations of the protein's secondary structure elements. Specifically, we focused on identifying and analyzing the alpha helices, spiral-shaped secondary structures stabilized by hydrogen bonds between amino acids, and the loops connecting these helical segments. This visualization provided insights into the spatial arrangement and organization of these structural

features, aiding in understanding how they contribute to the overall architecture and function of the CD8α protein.

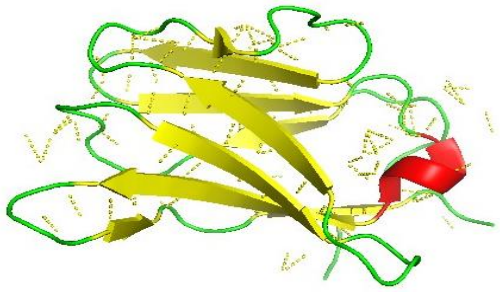

Fig.4 Helix (red) Sheet (yellow) Loop (green)

We utilized the STRING database to obtain a root-mean-square deviation (RMSD) value of 5.239 to compare the structures of 2NOU and 8HXS. RMSD measures the average distance between the atoms (or residues) of superimposed proteins, indicating their structural similarity or difference. This analysis provided quantitative data on the spatial alignment and similarity of the CD8α protein structures from different sources, offering insights into their structural conservation and potential functional implications in molecular interactions and biological processes.

TABLE 1: RESULT OF STRING

SAMPLE CODE	OTHER	RMSD	RESULT
8HXS	2NOU	5.239	MODERATE

The structural model of the small-spotted catshark CD8α protein was validated using the ERRAT program. The ERRAT plot provided an overall quality factor of 93.407, indicating a high level of accuracy in the model. As shown in the ERRAT2 plot (Figure 5), the majority of residues have error values below the 95% rejection threshold, suggesting a well-validated structure. Specifically, the analysis identified a small region around residues 20-30 that exhibited higher error values, yet these remained below the critical 99% rejection threshold.

High-resolution structures typically produce overall quality factors around 95% or higher, and lower resolution structures (2.5 to 3Å) average around 91%. The obtained overall quality factor of 93.407 for our model signifies that the structural predictions are reliable and consistent with good resolution structural models. This high-quality factor supports the validity of the modeled CD8α protein structure, providing confidence in subsequent structural and functional analyses.

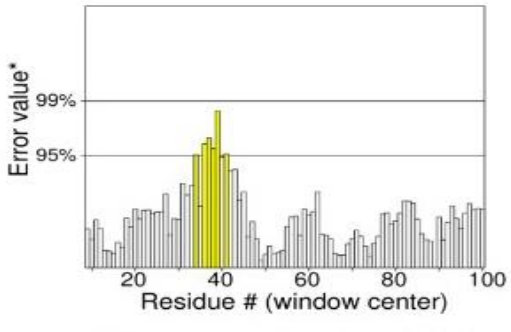

Fig.5 ERRAT plot for the small-spotted catshark CD8α protein model

A Ramachandran plot depicting the dihedral angles Phi (φ) and psi (ψ) of amino acid residues showed that most residues were in preferred regions, indicating correct stereochemistry and confirming the stability of the protein structure. Together, these results confirm the structural integrity and accuracy of our CD8α protein models, supporting their use in further functional analyses.

The structural validation of the small-spotted catshark CD8α protein model was performed using PROCHECK, focusing on the Ramachandran plot analysis. The Ramachandran plot provided a detailed view of the φ (phi) and ψ (psi) dihedral angles of the amino acid residues in the protein structure.

According to the plot statistics (Figure 6):

- 92.0% of the residues are located in the most favored regions [A, B, L], indicating a highly stable and reliable model.
- 8.0% of the residues are in additional allowed regions [a, b, l, p].
- No residues were found in the generously allowed regions [~a, ~b, ~l, ~p], or in disallowed regions, suggesting a well-constrained structure.

In total, 100 residues were analyzed, including 87 non-glycine and non-proline residues, with 7 glycine

residues (shown as triangles) and 4 proline residues. The analysis, based on 118 structures with a resolution of at least 2.0 Å and an R-factor no greater than 20%, suggests that a good quality model typically has over 90% of residues in the most favored regions.

The obtained results from the Ramachandran plot indicate that the CD8α protein model for the small-spotted catshark exhibits excellent stereochemical properties, with a significant majority of residues falling within the most favored regions. This high percentage of favored conformations reinforces the accuracy and reliability of the predicted protein structure.

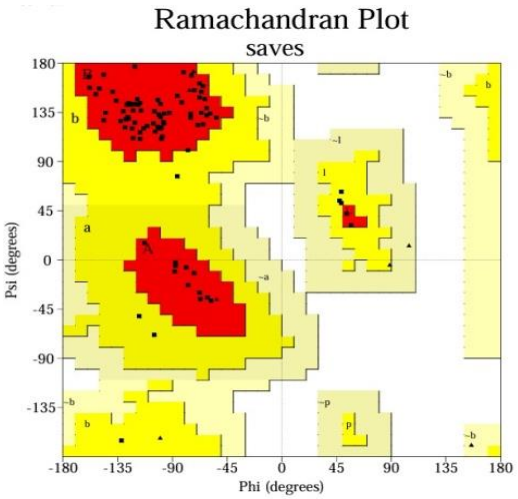

Fig.6 Ramachandran Plot

Using COBALT for boundary-based alignment improved the accuracy of our CD8α protein evolution analysis. By incorporating biological constraints, COBALT produced alignments that better reflected evolutionary relationships. The resulting alignments revealed conserved regions and motifs critical for protein function, highlighting the evolutionary conservation of species. This improved alignment improved our understanding of the structural and functional evolution of CD8α and provided deeper insights into its role in the small spotted catshark immune system.

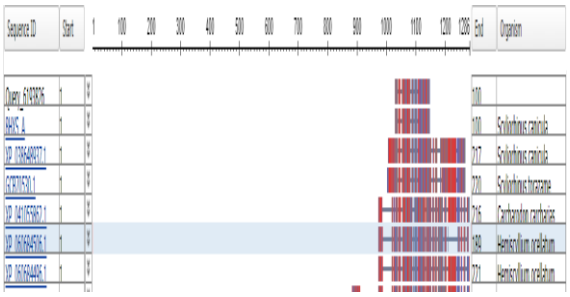

Fig.7 coloration method reflect side chain hydropathy where red represent hydrophobic and blue represent hydrophilic

The COBALT hydropathy assay provided insight into the hydrophobic and hydrophilic regions of the CD8α protein. By identifying these regions, we gained a better understanding of the interaction of the protein with its environment, its folding patterns and potential functional sites. Hydrophobic regions, usually buried in the protein core, contribute to structural stability, while hydrophilic regions, often exposed on the surface, play a key role in molecular interactions and binding activity. This analysis is crucial for predicting protein behavior and interactions in a cellular context.

Using BioPython in our project provided significant advantages for data processing and analysis. Using BioPython'sPDBParser, we efficiently analyzed the 3D structure of the CD8α protein from PDB files, allowing detailed examination of individual atoms, residues and chains in the structure. This enabled deep structural analysis and facilitated insight into protein architecture. In addition, we use the molecular weight function to accurately calculate the molecular weight of the CD8α protein sequence, which is crucial for understanding its biochemical properties. In addition, we automated the search and structuring of PDB files using BioPython'sPDBList, simplifying the workflow and enabling seamless integration of structural data into our analyses. Overall, BioPython's powerful libraries and tools have greatly improved our ability to efficiently perform comprehensive bioinformatics analyses.

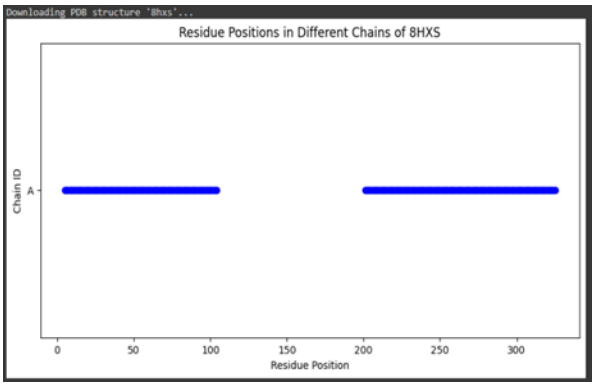

Fig.8 Residue Positions in Different Chains of 8HXS

This plot illustrates the residue positions for chain A of the protein structure 8HXS, obtained from the Protein Data Bank (PDB). The residue positions are plotted along the x-axis, while the chain ID is represented on the y-axis. The figure shows two distinct clusters of residues, suggesting potential structural features or gaps within the protein sequence. The left cluster ranges from residue positions 0 to approximately 150, and the right cluster spans positions 200 to about 325. This discontinuity may indicate a region of missing residues or a flexible linker within the protein.

InterProScan analysis revealed important conserved domains and functional motifs in the CD8a protein sequence. By identifying these key regions, we gained valuable insight into the protein's potential functional roles and evolutionary conservation across species. These data are important for understanding how CD8α affects the immune system of the lesser spotted catshark, directing functional and comparative studies.

Fig.9 INTERPRO SCAN

The sequence analysis of the small-spotted catshark CD8α protein was performed using InterPro, revealing detailed domain architecture and functional classification. The results, visualized in the graphical output (Figure 9), provide comprehensive insights into the protein's structural and functional features.

The representative domains identified include the IG_3c domain (SM00409), which spans a significant portion of the protein sequence, indicating a prominent role in the protein's function.

Within the Family category, the CD8α protein is classified under the CD8_asu family (IPR015468), specifically recognized as the CD8 alpha chain (PTHR10441). This classification confirms the identity and expected functional properties of the protein within the CD8 alpha chain family.

The Domain analysis reveals multiple significant domains:

- Ig_sub (IPR003599) and Ig_like_dom (IPR007110) domains, indicating immunoglobulin-like features.
- Ig_V-set (IPR013106) and V_set (PF07686) domains, suggesting a role in antigen binding.
- Additional domains such as Ig_8 (IPR003585) and Ig_sub2 (IPR003598), which further support the presence of immunoglobulin-like structures.

The Homologous Superfamily category identifies the protein as part of the Ig-like_dom_sf (IPR036179) and Ig-like_fold (IPR013783) superfamilies. This affiliation underscores the protein's evolutionary conservation and functional similarities to other immunoglobulin superfamily members.

Unintegrated segments and specific residues highlighted in the graphical output further delineate the protein's structural elements, including regions involved in antigen binding and heterodimer interface interactions, as indicated by the IgV (cd00099) domain.

These findings from InterPro analysis corroborate the protein's involvement in immune response mechanisms, consistent with its role as a CD8 alpha

chain. The detailed domain architecture and classification enhance our understanding of the CD8α protein's functional properties in the small-spotted catshark.

Fig.10 T-Coffee MSA Analysis

MSA analysis using T coffee revealed highly conserved regions of the CD8a protein, indicating critical functional and structural elements that are conserved across species. The alignment highlighted evolutionary relationships and identified key residues important for maintaining protein integrity and function. This detailed MSA analysis provided a basis for understanding the evolutionary conservation and functional significance of CD8α and guided further structural and functional studies.

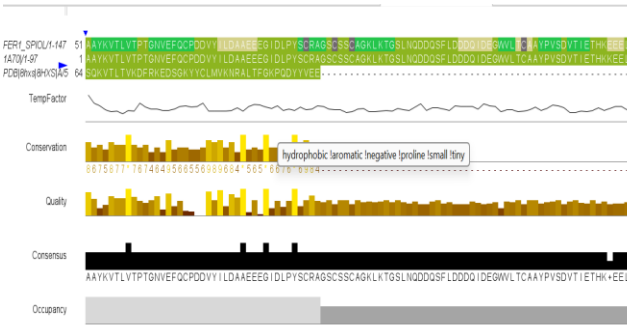

Fig 11:computational analysis of protein disorder prediction in Jalview shading alignment of sequence disorder (blosum 62 with protein disorder prediction either green or brown depending on type of disorder prediction (value over 0.1204 indicate disorder)

Jalview analysis of the multiple sequence alignment of the CD8a protein allowed us to visualize and interpret conserved and variable regions across species. By highlighting these regions, Jalview provided insights into critical functional residues and evolutionary conservation. The tool's interactive features allowed detailed review and modification of alignments, facilitating understanding of protein evolutionary history and functional significance. This analysis was instrumental in identifying key motifs and guiding future structural and functional studies.

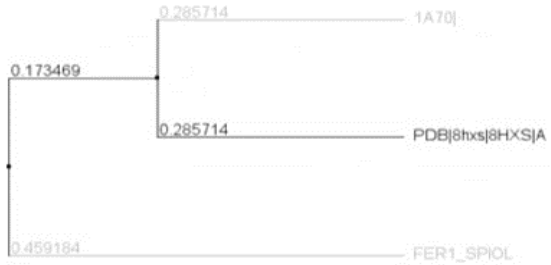

Fig. 12 Distance analysis of the Jalview phylogenetic tree revealed evolutionary relationships and divergence of CD8α protein sequences in different species. This visualization helped determine exactly how closely related the sequences are, providing insight into the protein's evolutionary history.

This phylogenetic tree depicts the evolutionary relationships of the protein structure 8HXS (chain A), as represented by "PDB|8hxs|8HXS|A". The branch lengths indicate the evolutionary distance, with the values provided at the nodes. The protein 8HXS is closely related to the protein structure labeled "1A70", with a branch length distance of 0.285714. Both proteins share a common ancestor with another protein, indicated by a branch length of 0.173469. The most distant relative in this tree is "FER1_SPIOL", with a branch length of 0.459184 from the common ancestor shared by 8HXS and 1A70. This tree highlights the evolutionary divergence and the relationships between these protein structures.

CONCLUSION

In conclusion, our integrated analysis of the speckled catshark protein CD8α provided extensive insights into its structural and functional properties. Using various bioinformatics tools, including sequence scanning, network analysis, structure visualization, and validation, we mapped the conserved regions of the protein, predicted its interactions, and confirmed its structural integrity. RasMol, PyMOL, InterPro, Jalview, T-Coffee, ClustalW, PDB, MMDB,

STRING, SAVES Server, PROCHECK, ERRAT, PDBSum, COBALT and BioPython enabled a versatile approach to study CD8α. Our findings highlight the evolutionary conservation of critical functional domains and the role of the protein in the immune system of the small spotted catshark. This project will advance our knowledge of CD8α and provide a methodological framework for similar studies in other species, contributing to the broader field of immunology and developmental biology.

REFERENCE

[1] Flajnik, M., Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11, 47–59 (2010). https://doi.org/10.1038/nrg2703

[2] Clem, L. W., and P. A. Small Jr. "Phylogeny of immunoglobulin structure and function: I. Immunoglobulins of the lemon shark." The Journal of experimental medicine 125.5 (1967): 893-920.

[3] Lopez et al., 1974

[4] Neely, Harold R., and Martin F. Flajnik. "Emergence and evolution of secondary lymphoid organs." Annual review of cell and developmental biology 32.1 (2016): 693-711.

[5] Rast, Jonathan P., et al. "α, β, γ, and δ T cell antigen receptor genes arose early in vertebrate phylogeny." Immunity 6.1 (1997): 1-11.

[6] Zapata, A., and C. T. Amemiya. "Phylogeny of lower vertebrates and their immunological structures." Origin and evolution of the vertebrate immune system (2000): 67-107.

[7] Tomonaga, Masao, David W. Golde, and Judith C. Gasson. "Biosynthetic (recombinant) human granulocyte-macrophage colony-stimulating factor: effect on normal bone marrow and leukemia cell lines." (1986): 31-36.

[8] Rumfelt, L. L., et al. "The development of primary and secondary lymphoid tissues in the nurse shark Ginglymostomacirratum: B-cell zones precede dendritic cell immigration and T-cell zone formation during ontogeny of the spleen." Scandinavian journal of immunology 56.2 (2002): 130-148.

[9] Malecek, Karolina, et al. "Immunoglobulin heavy chain exclusion in the shark." PLoS biology 6.6 (2008): e157.

[10] Jia Z, Feng J, Dooley H, Zou J and Wang J (2023) The first crystal structure of CD8aa from a cartilaginous fish. Front. Immunol. 14:1156219. doi: 10.3389/fimmu.2023.1156219

[11] Holler, Phillip D., Lukasz K. Chlewicki, and David M. Kranz. "TCRs with high affinity for foreign pMHC show self-reactivity." Nature immunology 4.1 (2003): 55-62.

[12] Wooldridge, Linda, et al. "CD8 controls T cell cross-reactivity." The Journal of Immunology 185.8 (2010): 4625-4632.

[13] Nakanishi, Teruyuki, Yasuhiro Shibasaki, and Yuta Matsuura. "T cells in fish." Biology 4.4 (2015): 640-663.

[14] Wang, Rui, et al. "Expression of GARP selectively identifies activated human FOXP3+ regulatory T cells." Proceedings of the National Academy of Sciences 106.32 (2009): 13439-13444.

[15] Jia Z, Feng J, Dooley H, Zou J and Wang J (2023) The first crystal structure of CD8aa from a cartilaginous fish. Front. Immunol. 14:1156219. doi: 10.3389/fimmu.2023.1156219

[16] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389

[17] Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., ... & Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, 40(W1), W597-W603. https://doi.org/10.1093/nar/gks400

[18] Bailey, T. L., & Gribskov, M. (1998). Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics*, 14(1), 48-54. https://doi.org/10.1093/bioinformatics/14.1.48

[19] Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30(S1), S162-S173. https://doi.org/10.1002/elps.200900140

[20] Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., ... & Yeats, C. (2009). InterPro: The integrative protein signature

database. *Nucleic Acids Research*, 37(suppl_1), D211-D215. https://doi.org/10.1093/nar/gkn785

[21] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H, & Higgins, D. G. (2007). ClustalW and ClustalX version 2.0. *Bioinformatics*, 23(21), 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

[22] McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404-405. https://doi.org/10.1093/bioinformatics/16.4.404

[23] Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3), 779-815. https://doi.org/10.1006/jmbi.1993.1626

[24] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W, & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. https://doi.org/10.1038/msb.2011.75

[25] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[26] Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 54(1), 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3