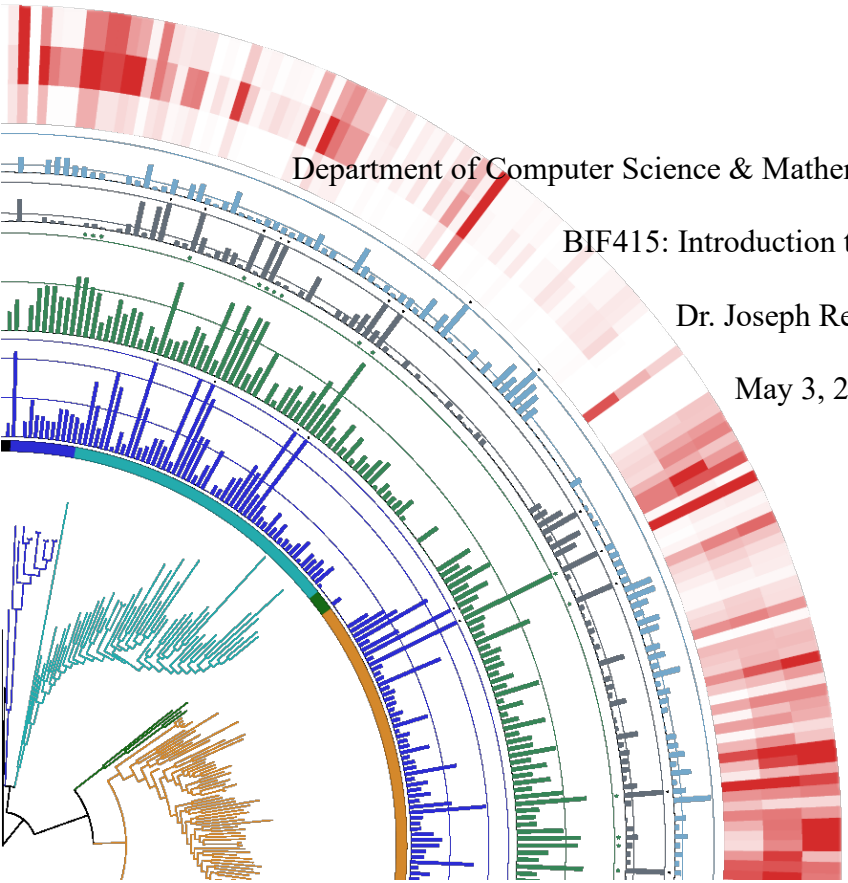
Decoding CTLA-4 protein in Ig-like V-type Domain:
An Exciting Journey of Phylogenetic Analysis and Conservation

Joudy Allam, Lea Hassan Kassem, Nour Saad



Department of Computer Science & Mathematics, Lebanese American University

BIF415: Introduction to Bioinformatics

Dr. Joseph Rebehmed

May 3, 2023

CTLA-4 – Ig-Like V-Type Domain

Introduction

As we dive deeper into the world of immunology, one protein that has piqued our interest is CTLA-4. Cytotoxic T-Lymphocyte protein 4 is an inhibitory receptor, found in many species. It assists in the development of immunological tolerance and the prevention of autoimmunity by regulating T-cell activity. It has been proven to be the root of various autoimmune disorders such as Lupus, and cancer. Unlocking the secrets of this protein's evolutionary history and conservation will help us grasp the understanding of its function. In this study, we embark on a journey of discovery, using cutting-edge computational techniques and phylogenetic analysis to reveal the interesting world of CTLA-4.

Ig-like V-type Domain

The only domain in the CTLA-4 protein is Ig-like V-type. It extends from amino acids 39 to 140. This domain is essential for regulating the immune response and maintaining immune homeostasis. It allows high affinity binding to the B7 protein on antigen-presenting cells (APCs) with the co-stimulatory receptor CD28 on T-cells, which is necessary for T-cell inhibition. Moreover, dysfunction or loss of CTLA-4 can result in the development of autoimmune diseases, such as type 1 diabetes, as well as immune dysregulation in cancer.

Uncovering Homologous Domains

Using the SwissProt database, a BLAST search was conducted to identify homologous domains to our protein of interest. The parameters were set to a maximum of 50 target sequences and an expected threshold of 0.001. However, the results only aligned 12 sequences originating from a variety of species, such as mammals, birds and fish.

Upon analyzing the results, we observed a coverage range from 80% to 100% and a sequence identity range from 30.95% to 100%. These high percentages highlight the similar structure and function shared between the 12 sequences. The high degree of conservation could be explained by the CTLA-4's crucial role in regulating the immune response, where the Ig-like V-type domain is essential for its function.

The BLAST results also revealed that the most similar hits were from convergent species, such as primates, while more divergent species like birds and fish showed lower sequence identity. This pattern is consistent with vertebrate evolution through time, with some degree of divergence.

CTLA-4 – Ig-Like V-Type Domain

Multiple Sequence Alignment

```
P09793.1:39-140      QVTQPSVVLASSHGVASFPCEYSPSHNTDEVRVTVLRQTNDQMTEVCATTFTEKNTVGFL
Q9XSI1.1:39-140      HAAQPAVVLASSRGVASFVCEYGSSGNAAEVRVTMLRQAGSQMTEVCAATYTVEDELAFL
Q9MYX7.1:39-140      HVAQPAVVLANSRGVASFVCEYGSAGKAAEVRVTVLRRAGSQMTEVCAATYTVEDELTFL
P42072.1:39-140      HVSQPAVVLASSRGVASFVCEYASSHKATEVRVTVLRQANSQMTEVCAMTYTVENELTFI
P16410.3:39-140      HVAQPAVVLASSRGIASFVCEYASPGKATEVRVTVLRQADSQVTEVCAATYMMGNELTFL
P31043.1:23-121      VAQRPLLIVANRT--ATLVCNYTYNGTGKEFRASLHKGTDSAV-EVCFISWNMT-KINSN
P31041.2:24-123      VKQSPLLVVDSNE--VSLSCRYSYNLLAKEFRASLYKGVNSDV-EVCVGNGNFTYQPQFR
P31042.1:29-123      -----LLVVDNNE--VSLSCRYSYNLLAKEFRASLYKGVNSDV-EVCVGNGNFTYQPQFR
Q28071.1:40-121      -------------------CKYTYNLFSKEFRASLYKGADSAV-EVCAVNGNHSHPLQS-
P10747.1:23-122      VKQSPMLVAYDNA--VNLSCKYSYNLFSREFRASLHKGLDSAV-EVCVVYGNYSQQLQVY
P42069.1:41-123      -------------------CKYTYNLFSKEFRASLYKGADSAV-EVCVVNGNFSHPHQFH
O02757.1:41-123      -------------------CKYTHNFFSKEFRASLYKGVDSAV-EVCVVNGNYSHQPQFY
                                        * *       *.*.:: .  .. : ***

P09793.1:39-140      DYP--FCSGTFNESRVNLTIQGLRAVDTGLYLCKVELMYPPPYF
Q9XSI1.1:39-140      DDS--TCTGTSSGNKVNLTIQGLRAMGTGLYICKVELMYPPPYY
Q9MYX7.1:39-140      DDS--TCTGTSTENKVNLTIQGLRAVDTGLYICKVELLYPPPYY
P42072.1:39-140      DDS--TCTGISHGNKVNLTIQGLSAMDTGLYICKVELMYPPPYY
P16410.3:39-140      DDS--ICTGTSSGNQVNLTIQGLRAMDTGLYICKVELMYPPPYY
P31043.1:23-121      SNKEFNCRGIHDKDKVIFNLWNMSASQTDIYFCKIEAMYPPPY-
P31041.2:24-123      SNAEFNCDGDFDNETVTFRLWNLHVNHTDIYFCKIEFMYPPPY-
P31042.1:29-123      PNVGFNCDGNFDNETVTFRLWNLDVNHTDIYFCKIEVMYPPPY-
Q28071.1:40-121      TNKEFNCTVKVGNETVTFYLQDLYVNQTDIYFCKLEVLYPPPY-
P10747.1:23-122      SKTGFNCDGKLGNESVTFYLQNLYVNQTDIYFCKIEVMYPPPY-
P42069.1:41-123      STTGFNCDGKLGNETVTFYLKNLYVNQTDIYFCKIEVMYPPPY-
O02757.1:41-123      SSTGFDCDGKLGNETVTFYLRNLFVNQTDIYFCKIEVMYPPPY-
                           *      . * : :..: .  *.:*:**:* :*****
```

Figure 1: CLUSTAL Multiple Sequence alignment of the CTLA-4 human protein (P16410) with 11 sequences, according to MUSCLE (3.8). This figure highlights the sequences of high and low conservation between the 12 sequences.

According to the multiple sequence alignments in Figure 1, many amino acids appeared to be conserved in all the sequences, including cysteine (C), tyrosine (Y), proline (P), glutamic acid (E), valine (V), arginine (R), lysine (K), and threonine (T) at positions in respect to P16410: (C56, C83, C103, C129); (Y58, Y127, Y135, Y139); (P136, P137, P138); (E66, E81, E132); (V82, V112); (R68); (K130) and (T124). The conservation of these positions across this protein emphasizes the importance of these amino acids in maintaining the domain's structure and function.

In addition, many other amino acids were partially conserved. For example, M139 was replaced by T139 as a compatible hydrophobic alternative. While the amino acids are different, their function remains the same.

Furthermore, gaps were inserted to maximize the similarity between all sequences. Gaps can be a sign of insertions or deletions.

Structural and Functional Importance of Conserved Positions

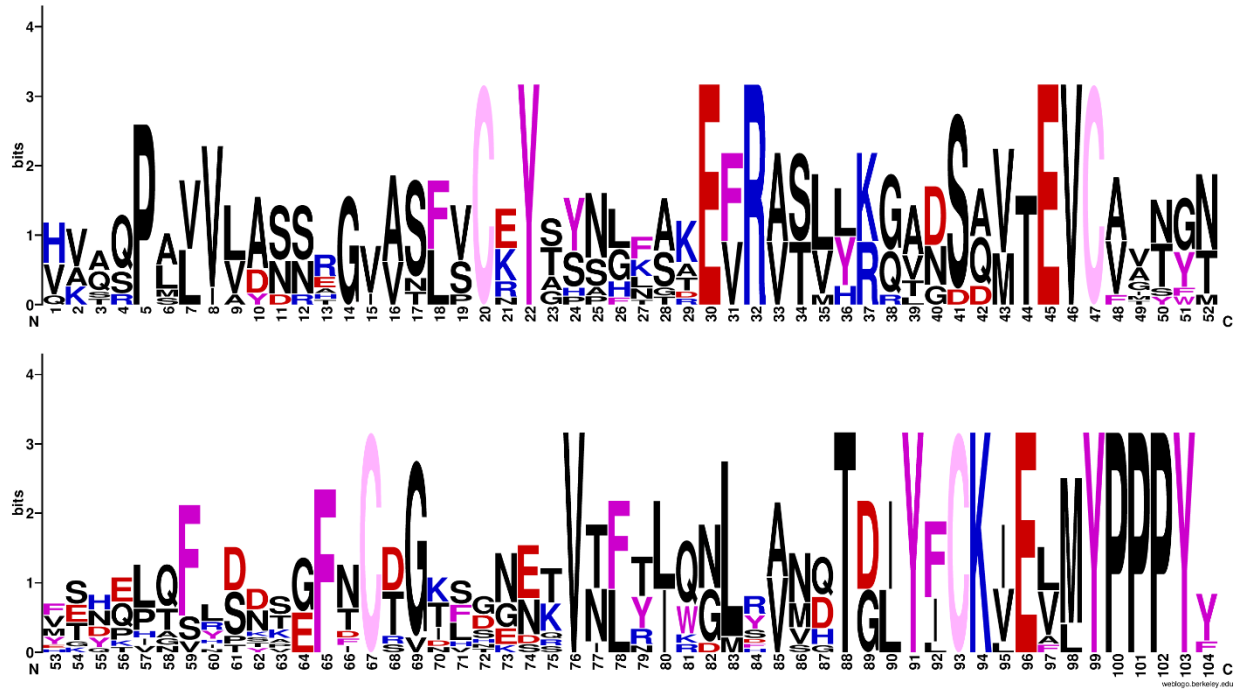


Figure 2: Graphic representation of the MSA using WebLogo, with positively charged amino acids marked in blue, negatively charged in red, cysteine in yellow, and aromatic in purple.

Structurally, these conserved amino acids play a crucial role in stabilizing the three-dimensional structure of the domain. For instance, cysteine residues often form disulfide bonds, which can stabilize the protein's fold, while hydrophobic amino acids like valine and proline might participate in the formation of the protein's hydrophobic core. This conservation suggests that it is essential in maintaining the overall structure of the domain, and any alterations to these positions may disrupt the domain's stability.

Functionally, the preserved positions might be involved in facilitating interactions with other proteins, such as CD80 and CD86. In the case of the CTLA-4 Ig-like V-type domain, these interactions are necessary for its role in regulating immune responses. Arginine and glutamic acid residues, which are respectively, positively, and negatively charged, participate in forming salt bridges and hydrogen bonds with interacting associates. Tyrosine residues, on the other hand, can be phosphorylated and contribute to signaling pathways, emphasizing their functional significance. The conservation of these positions across a variety of species specifies their essential role in maintaining the functionality of the CTLA-4 Ig-like V-type domain throughout evolution. The high degree of conservation proposes that these amino acids have maintained their structure and

function in diverse organisms, highlighting their critical role in immune regulation. Thus, their structural and functional importance is supported by their conservation across different proteins and species, pointing out the crucial role of the CTLA-4 Ig-like V-type domain in immune regulation.

Tree Topology

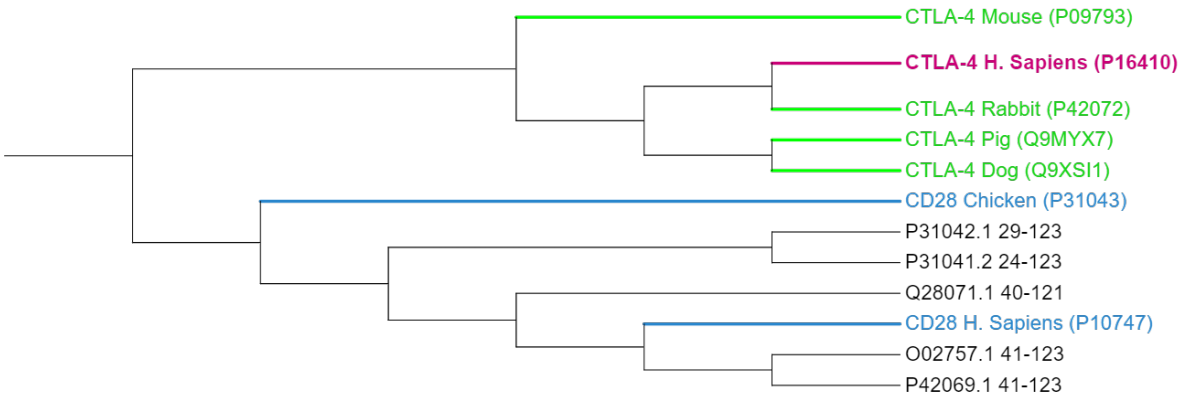


Figure 3: The Cladogram of the Ig-Like V-type domain, according to ITOL.

The most common ancestor of all the leaves diverged into 2 clades: CTLA-4 and CD-28. Our protein of interest, shown in pink, proved to be the closest to the CTLA-4 protein in Rabbits. The pig and the dog share the most common ancestor with the human and the dog indicating that a divergence occurred at a certain point in time. Lastly, this subtree altogether is related to the mouse CTLA-4 protein. These findings highlight the complex relationships between different species and their immune regulatory proteins.

Even though the cladogram shows that CD-28 homo sapiens and CTLA-4 Ig-Like V-Type domain diverge at an early stage, they still work together to ensure the full functionality of the protein.

Conclusion

In conclusion, the study "Decoding CTLA-4 protein in Ig-like V-type Domain" reveals the evolutionary history, conservation, and function of CTLA-4 protein, responsible for regulating the immune response and maintaining immune homeostasis. The study focused on the Ig-like V-type domain, which extends from amino acids 39 to 140 and allows high affinity binding to the B7 protein on antigen-presenting cells (APCs), necessary for T-cell inhibition. The Multiple Sequence

Alignment of SwissProt identified 12 sequences with high conservation, sharing similar structure and function, with the most similar hits from convergent species such as primates. The conserved amino acids, including cysteine, tyrosine, and proline, play a crucial role in stabilizing the three-dimensional structure of the domain and facilitating interactions with other proteins, such as CD80 and CD86. The amino acids are also involved in forming salt bridges, hydrogen bonds, and phosphorylation contributing to signaling pathways essential for the protein's functionality. The findings from this study shed light on the importance of the Ig-like V-type domain of CTLA-4, providing a better understanding of its function and potential implications in autoimmune diseases and cancer. Finally, discovering the secrets of this domain is similar to peeling back the layers of a riddle, exposing its importance in immunological homeostasis and preventing autoimmunity.

References

Krummel MF, Allison JP. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. J Exp Med. 1995;182(2):459-465. doi:10.1084/jem.182.2.459

Paterson AM, Lovitch SB, Sage PT, et al. Deletion of CTLA-4 on regulatory T cells during adulthood leads to resistance to autoimmunity. J Exp Med. 2015;212(10):1603-1621. doi:10.1084/jem.20140884

Zhang Y, Liu S, Liao W, et al. Molecular basis for CTLA-4 ligand binding and regulation of T cell activation. Mol Cell. 2021;81(13):2775-2789.e6. doi:10.1016/j.molcel.2021.05.032

Balaji KN, Schaschke N, Machleidt W, Catalfamo M, Henkart PA. Surface cathepsin B protects cytotoxic lymphocytes from self-destruction after degranulation. J Exp Med. 2002;196(4):493-503. doi:10.1084/jem.20020084

Liang B, Li Y, Li Q, et al. CTLA-4(+) regulatory T cells increased in CML patients and downregulate the expression of TCR ζ chain of T cells. Cancer Immunol Immunother. 2013;62(7):1061-1070. doi:10.1007/s00262-013-1415-2.