Readers Insight
Publisher of open access books and journals
PJMLS
Print ISSN: 2707-4471. Online ISSN: 2707-448X

Research Article	Pak-Euro Journal of Medical and Life Sciences
DOI: 10.31580/pjmls.v7isp2.3114	Copyright © All rights are reserved by Corresponding Author
Vol. 7 No. Sp. 2, 2024: pp. S379-S386	
www.readersinsight.net/pjmls	**Revised:** November 30, 2024 **Accepted:** December 11, 2024
Submission: July 03, 2024	**Published Online:** December 31, 2024

ANALYSIS OF DIFFERENTIALLY EXPRESSED PROTEINS IN RHEUMATOID ARTHRITIS USING BIOINFORMATICS TOOLS

Qudsia Khalid Luni[1], Muhammad Ali Khan[1], Hafiz Muhammad Ali[2], Safia Mengal[3], Irfan Shahzad Sheikh[1], Asad Ullah[1], Muhammad Tauseef Asmat[1], Sumaira Akram[1], Fayyaz Ahmed[4], Mohammad Zahid Mustafa[1]*

[1]**Center for Advanced Studies in Vaccinology and Biotechnology (CASVAB),** University of Balochistan Quetta, Pakistan
[2]**Faculty of Veterinary and Animal Sciences,** The Islamia University of Bahawalpur, Bahawalpur, Pakistan
[3]**Institute of Public Health,** Government of Balochistan, Health Department, Quetta, Pakistan
[4]**Livestock and Dairy Development Department,** Government of Balochistan, Quetta, Pakistan
***Corresponding Author:** Prof. Dr. Mohammad Zahid Mustafa.
E. mails: drzahid.casvab@um.uob.edu.pk

Abstract
Rheumatoid arthritis (RA) is a chronic autoimmune disorder that affects millions globally, primarily targeting small and large joints, leading to tendon weakening and bone erosion. Affecting approximately 0.5%–1% of the population, the precise pathological mechanisms of RA remain unclear. This study investigates differentially expressed proteins involved in RA pathogenesis using bioinformatics tools. Proteomics data were obtained from the synovial fluid and serum samples of RA patients and healthy individuals. Synovial fluid was used as an antigenic extract, and proteins were separated via SDS-PAGE, followed by transfer to a nitrocellulose membrane. Immunoblotting revealed specific protein bands, and comparisons between RA patient blots and controls highlighted distinct differences. A total of 35 serum samples from RA patients were analyzed. The application of bioinformatics has enabled the identification of potential biomarkers and therapeutic targets for RA, enhancing our understanding of its molecular basis.
Keywords: *Autoimmunity, Bioinformatics, Biomarkers, Proteomics, Rheumatoid arthritis*

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disease that primarily affects small joints and progressively involves larger joints, leading to the weakening of ligaments and tendons and the destruction of joint bones (1). Beyond joint involvement, RA can extend to affect multiple organ systems, including the skin, eyes, kidneys, lungs, and heart. Being an autoimmune condition, the patient's immune system erroneously targets healthy cells, resulting in widespread inflammation and tissue damage (2).

Early diagnosis and prompt treatment are crucial in managing RA to prevent irreversible damage. Standard therapeutic regimens include glucocorticoids and disease-modifying antirheumatic drugs (DMARDs), which are effective in reducing disease activity and slowing progression (3). If left untreated, RA can lead to significant disability and even premature death. One of the hallmark features of RA is synovitis—the inflammation of the synovial membrane—which leads to joint swelling, pain, and bone deformity (4). Epidemiologically, RA is more prevalent in women than men, with a reported female-to-male ratio of approximately 3:1, and it affects 0.5–1% of the global population (5).

A detailed pathological description of RA is provided by Fassbender, who characterized the disease through three pathological mechanisms: exudative inflammation, a proliferation-destructive process, and enzymatic collagenolysis. These mechanisms contribute to tissue necrosis, including that of the myocardium, blood vessels, and ocular structures. Millions worldwide are afflicted by RA, a disease marked by synovial inflammation and progressive joint degeneration that causes pain, stiffness, and impaired mobility (6). Although the etiology of RA involves genetic and environmental contributors, the exact biochemical pathways and molecular mechanisms underlying the disease remain incompletely understood, and no definitive cure currently exists (7).

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent advancements in proteomics—the large-scale study of proteins—have allowed for a deeper understanding of diseases like RA. Proteomic approaches have revealed differentially expressed proteins that are involved in the pathogenesis of RA, providing insights into disease mechanisms and potential therapeutic targets (8). Analyzing such vast proteomic datasets requires specialized bioinformatics tools, which combine elements of computer science, statistics, and molecular biology to interpret complex biological data (9). These tools have proven invaluable in identifying novel biomarkers for early diagnosis and in discovering potential molecular targets for therapy (10).

Bioinformatics, as a multidisciplinary field integrating mathematics, computer science, and biology, supports the storage, retrieval, and interpretation of biological information (11), bioinformics tools have been used to predict the physicochemical properties, transmembrane domains, various secondary structures and 3D structure of proteins (1-14). Various software tools are now used to determine the functional and structural characteristics of genes and proteins, saving both time and research costs. Among the widely used platforms is ExPASy (Expert Protein Analysis System), which enables users to calculate protein properties such as isoelectric point (pI), molecular weight, transmembrane domains, and 3D structure (15). Hosted by the Swiss Institute of Bioinformatics (SIB), ExPASy provides comprehensive tools for analyzing primary and secondary protein structures, patterns, and molecular modeling (16).

For example, ProtScale is used to assess the hydrophobicity and secondary structure of proteins based on their amino acid sequences. Another tool, ProtParam, estimates physicochemical properties such as amino acid composition, theoretical pI, molecular weight, extinction coefficient, instability index, estimated half-life, aliphatic index, and atomic composition from protein sequences (17). The integration of these bioinformatic resources in RA research allows for the systematic identification and evaluation of protein biomarkers using data derived from synovial fluid and serum samples of both RA patients and healthy individuals.

MATERIALS AND METHODS

SAMPLE COLLECTION

A total of thirty-five serum samples were obtained from patients clinically diagnosed with rheumatoid arthritis (RA). Additionally, twenty serum samples were collected from healthy individuals to serve as controls; these were pooled together to form a single representative control sample (n = 20). All blood samples were collected using sterile falcon tubes and centrifuged at 1200 rpm for 5–10 minutes. The separated serum was transferred into sterile microcentrifuge tubes and stored at −80°C until further analysis.

SDS-PAGE ANALYSIS

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate proteins based on their molecular weight. Cleaned notched and flat glass plates were assembled on a casting stand after being wiped with 70% ethanol and dried. Two gels were prepared: a resolving gel (12 %) and a stacking gel (4 %). After polymerization and well formation, the denatured protein samples were carefully loaded into the wells. Electrophoresis was conducted by applying an electric current across the gel. Following separation, protein bands were visualized using the Coomassie Brilliant Blue staining method.

BIOINFORMATIC ANALYSIS

To evaluate the physicochemical and structural characteristics of proteins identified in RA samples, several bioinformatic tools were employed. The ProtScale tool was used to analyze transmembrane domains, while ProtParam was utilized to predict various secondary structure parameters and physiochemical properties, including molecular weight and isoelectric point. PSIPRED was used for secondary and 3D structure prediction. Additionally, tools from the ExPASy (Expert Protein Analysis System) and UniProt platforms were used to retrieve protein sequences and determine detailed functional and structural attributes.

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

RESULTS AND DISCUSSION

The protein profiling of serum samples obtained from patients with rheumatoid arthritis (RA) was conducted using SDS-PAGE (Fig. 1). Out of the 35 patient samples analyzed, five proteins—Cerpin-H1, Interleukin-6 (IL-6), T-cell surface glycoprotein CD3, Calgranulin A, and Serum amyloid A1—were found to be differentially expressed (Table I). These proteins were further analyzed through various bioinformatic tools to investigate their physicochemical properties, transmembrane domains, secondary structures, and three-dimensional models (Table II).

Fig. 1. Representative SDS-PAGE pattern of differentially expressed proteins

Table I. Names, accession numbers, sequence length and range of detected proteins

S. No.	Proteins	Accession number	Sequence length (Amino acids)	Range
1	Cerpin-H1	P50454	400	19-418
2	Interkeukin-6 (IL-6)	P05231	212	30-212
3	T-cell Surface Glycoprotein(D3)	P20963	164	22-164
4	Calgranulin-A (S100-A8)	P05109	93	1-93
5	Serum amyloid (A-1)	P954630	252	1-252

PREDICTION OF TRANSMEMBRANE DOMAINS USING PROTSCALE

Understanding whether a protein contains transmembrane domains provides crucial insight into its biological function, localization, and behavior during isolation and purification. Transmembrane domains are specific regions of proteins that span across the phospholipid bilayer of cell membranes. These domains consist of amino acid sequences that interact with the fatty acyl groups of membrane phospholipids, allowing the protein to be embedded in the membrane with intracellular and extracellular domains extending on either side (18).

In the present study, several computational tools were employed to analyze the structure and characteristics of the identified RA-related proteins. The ProtScale tool was used to evaluate key physicochemical properties such as molecular weight, extinction coefficient, accession number, and aliphatic index. The respective molecular weights of the five identified proteins were approximately 47 kDa, 208 kDa, 164 kDa, 108 kDa, and 116 kDa. Accession numbers for each protein were confirmed using the UniProt database.

To predict transmembrane domains, the Kyte & Doolittle hydrophobicity scale was applied via ProtScale. The generated hydrophobicity plots displayed the position of amino acids on the x-axis and their hydrophobicity scores on the y-axis. An optimal window size of 19 amino acids was selected for the

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

analysis. A threshold of 1.5 was used to identify potential transmembrane domains—any peaks above this value were considered as indicative of such domains (Fig. 1).

These predictions support the hypothesis that the proteins identified in RA patients may be involved in membrane-related immune signaling pathways and could serve as potential targets for diagnostic or therapeutic intervention.

Table II. Physiochemical properties of detected proteins on the basis of their degree of hydrophobicity

Name of Protein	Accession number	Molecular weight	Theoretical PI	Number of amino acids	Estimated half-life	Instability index	Aliphatic index	Extinction coefficient
Cerpin-H1	P50454	47KDa	8.81	400	4.4 hours	4290	87.88	36900
Interkeukin-6 (IL-6)	P05231	208KDa	6.21	183	100 hours	57.89	84.81	10220
T-cell Surface glycoprotein (CD3)	P20963	164KDa	9.16	143	0.8 hours	59.12	75.10	13410
Calgranulin-A (S100-A8)	P05109	108KDa	6.50	93	30 hours	25.04	96.45	11460
Serum amyloid A-1	P954630	116KDa	5.89	104	1 hours	26.62	40.67	23950

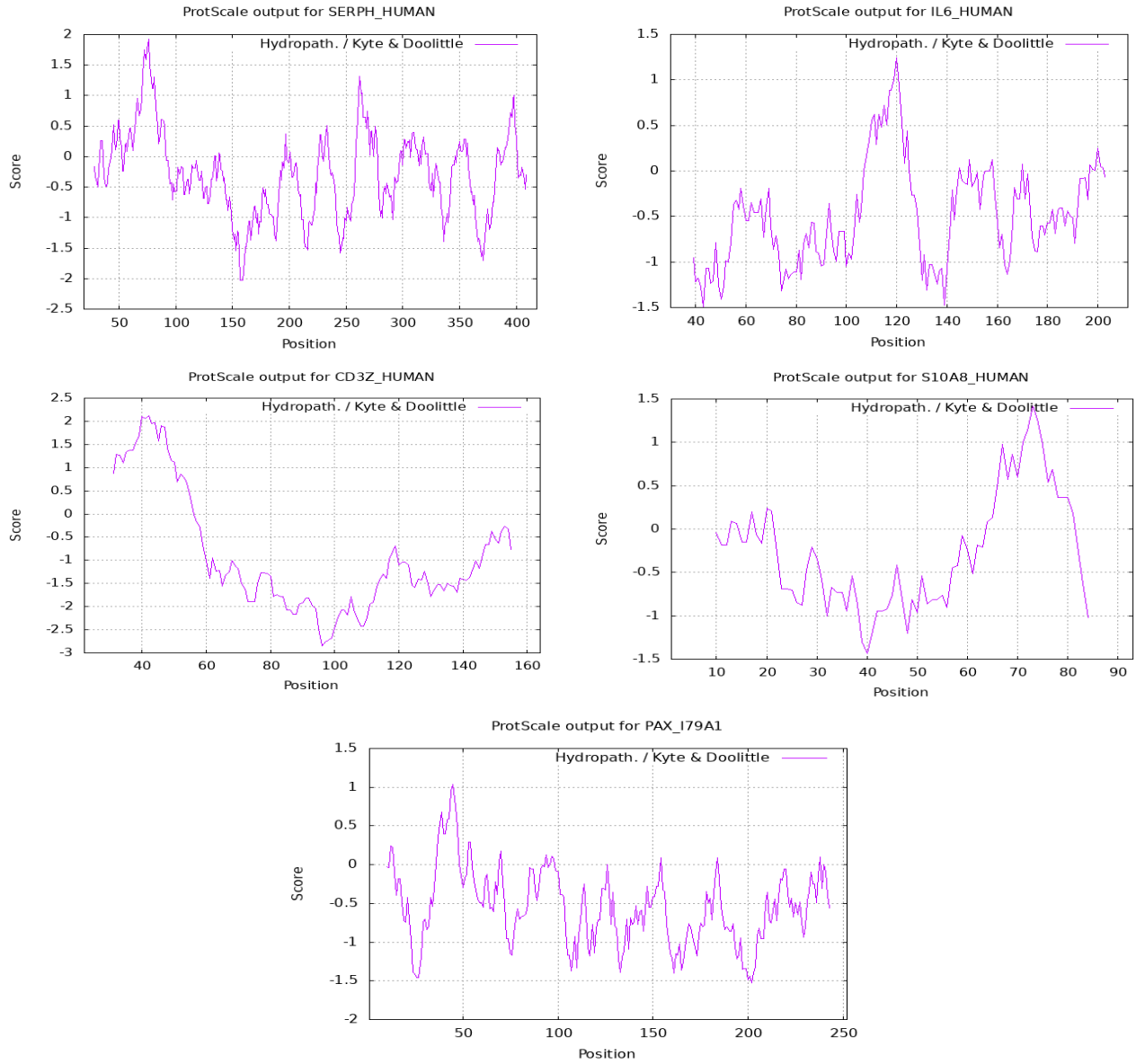

Fig. 1. Putative transmembrane domains of Cerpin-H1, interkeukin-6, T-cell surface glycoproteins, Calgrulin-A and serum amyloid A1 were obtained using hphob. / kyte & Dolittle scale

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

PREDICTION OF SECONDARY STRUCTURE OF PROTEINS

Secondary structure prediction plays a vital role in understanding protein conformation, folding mechanisms, surface-exposed residues, and antigenic regions that may serve as potential therapeutic or diagnostic targets. The primary components of a protein's secondary structure include alpha-helices (α-helices), beta-sheets (β-strands), and random coils, which collectively determine the protein's stability, functionality, and interaction with other biomolecules (19).

In this study, the PSIPRED tool—recognized as one of the most reliable and accurate platforms for predicting protein secondary structures—was employed. Protein sequences corresponding to five differentially expressed proteins with molecular weights of approximately 47 kDa, 205 kDa, 164 kDa, 108 kDa, and 116 kDa were retrieved from the NCBI protein database in FASTA format. These sequences were input into the PSIPRED server to determine their predicted secondary structure patterns (Fig. 2).

Fig. 2. Protein secondary structures of i. Cerpin _ H1, ii. Interleukin-6 IL6, iii. T-cell Surface Glycoprotein (D3), iv. and v. Serum amyloid (A-1)

In the resulting outputs, 'H' represents alpha-helices, 'E' denotes beta-strands (extended sheets), and 'C' corresponds to random coils. PSIPRED also provides a confidence score ranging from 0 to 9 for each residue, where a higher score (closer to 9) indicates greater reliability and stability of the predicted structure. The analysis revealed that all five proteins contained mixed structural motifs, including well-defined α-helices and β-strands interspersed with flexible coil regions. These patterns suggest that the identified RA-

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

related proteins possess both structural stability and conformational flexibility, which are essential for their functional roles in immune modulation, inflammation, and signal transduction.

Notably, coil regions often correspond to epitope-rich sites, making them ideal targets for antibody recognition in autoimmune conditions such as rheumatoid arthritis. Moreover, understanding the secondary structural elements is essential for guiding tertiary structure prediction, epitope mapping, and structure-based drug design.

These findings provide foundational insights into the structural characteristics of proteins implicated in RA pathogenesis, thereby supporting their potential utility in future experimental and therapeutic studies.

PREDICTING PROTEIN 3D STRUCTURE

CERPIN H1

Collagen, the most abundant protein in animals, is a key component of the extracellular matrix in tissues such as skin and bone. CERPIN H1 functions as a molecular chaperone specific to procollagen and is unique among chaperones in that it exclusively recognizes the folded conformation of its target protein. Significantly reduced functional levels of CERPIN H1 have been reported in severe recessive forms of rheumatoid arthritis. A critical Hsp47-binding site is identified by the presence of an arginine residue at the Yaa-position within a Xaa-Yaa-Gly collagen triplet. Two out of the three such binding sites on the collagen triple helix are occupied by Hsp47 molecules. These molecules bind in a head-to-head orientation, establishing extensive interactions with both the leading and trailing strands of the collagen helix. This structural arrangement elucidates how CERPIN H1 contributes to the stabilization of the collagen triple helix and inhibits the formation of collagen bundles (20).

HUMAN INTERLEUKIN-6 (IL-6)

Interleukin-6 (IL-6) is a cytokine composed of 185 amino acids, whose dysregulation is implicated in a range of pathological conditions. Its three-dimensional structure has been elucidated using heteronuclear three- and four-dimensional nuclear magnetic resonance (NMR) spectroscopy. IL-6 comprises five alpha-helices separated by loops of variable lengths. Four of these helices form a classical four-helix bundle, while the fifth helix resides within the CD loop. Although the N-terminal 19 amino acids remain disordered in solution, the remainder of the molecule exhibits a well-defined tertiary structure, sharing several structural motifs common to long-chain four-helix bundle cytokines (21).

T-CELL SURFACE GLYCOPROTEIN (CD3)

The complete structural model of the TCR-CD3 complex includes all extracellular domains and transmembrane helices. The octameric complex exhibits a stoichiometry of 1:1:1:1 for TCRαβ:CD3γε:CD3δε:CD3ζζ. Within the membrane, the CD3 transmembrane components adopt a barrel-like architecture, where the helices of CD3ζζ interact with those of CD3γε and CD3δε. Integration of the TCRαβ transmembrane helices into this barrel-like structure occurs through both hydrophobic and ionic interactions, thereby enabling the assembly of the TCR-CD3 complex across the membrane (22).

CALGRANULIN-A (S100-A8)

S100A8 and S100A9 are calcium-binding proteins serving as key markers in various inflammatory conditions such as rheumatoid arthritis and inflammatory bowel disease. These proteins form a heterodimeric complex whose structure has been resolved via X-ray crystallography. The dimerization mechanism in calprotectin is more akin to the homodimerization observed in S100A8 and provides structural stabilization to S100A9. This stabilization is evident in the extended C-terminal alpha-helix of S100A9. Tetramer formation is mediated by the canonical calcium-binding helices of the heterodimer. Two distinct zinc-binding sites are present: one is characterized by a high-affinity arrangement involving three histidine residues and one aspartic acid side chain, a feature unique to the heterotetrameric form.

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Overexpression of this Zn^{2+}-binding capability may result in disrupted zinc homeostasis, leading to clinically significant symptoms (23).

SERUM AMYLOID A1 (SAA1)

SAA1 is a member of the acute-phase protein family involved in inflammatory responses. In humans, SAA1 exists predominantly as a hexamer, with each subunit adopting a four-helix bundle conformation stabilized by an extended C-terminal tail. Two positively charged regions, one near the central core and the other at the apex of the hexamer, mediate interactions with heparin. The apex domain also promotes high-affinity binding to lipoproteins, a process that is competitively inhibited by heparin. Amyloidogenic peptide sequences are found at the N-termini of helices 1 and 3. While hexamer dissociation alone does not trigger amyloidogenesis, proteolytic cleavage or truncation of the C-terminal tail facilitates the formation of a variety of structural aggregates, which include ~5-nm protofibril-like repeating units (23).

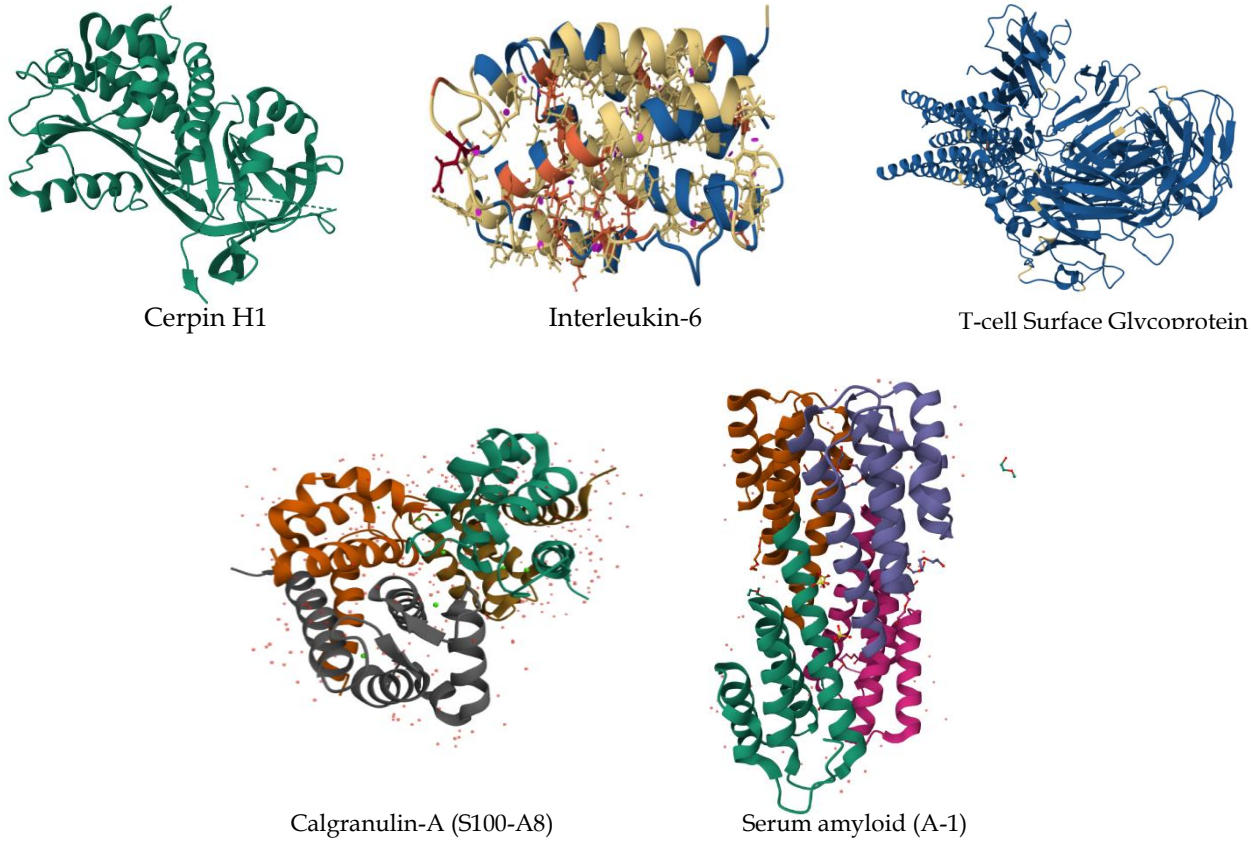

Fig. 3. Protein 3D structures of i. Cerpin _ H1, ii. Interleukin-6 IL6, iii. T-cell Surface Glycoprotein (D3), iv. and v. Serum amyloid (A-1)

CONCLUSION

The investigation explored the physicochemical characteristics, transmembrane domains, secondary structure, and 3D conformations of detected proteins. The findings highlight the growing utility of bioinformatics in identifying potential diagnostic biomarkers and therapeutic targets, offering deeper insights into the molecular mechanisms of rheumatoid arthritis. Continued advancement in proteomic and genomic bioinformatics tools will further enhance our ability to decode complex disease pathways.

References:

1. Scutellari PN, Orzincolo C. Rheumatoid arthritis: sequences. European journal of radiology. 1998;27:S31-S8.
2. Singh SP, Wal P, Wal A, Srivastava V, Tiwari R, Sharma RD. Understanding autoimmune disease: an update review. IJPTB. 2016;3:51-65.

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

3. Chatzidionysiou K, Emamikia S, Nam J, Ramiro S, Smolen J, van der Heijde D, et al. Efficacy of glucocorticoids, conventional and targeted synthetic disease-modifying antirheumatic drugs: a systematic literature review informing the 2016 update of the EULAR recommendations for the management of rheumatoid arthritis. Annals of the rheumatic diseases. 2017;76(6):1102-7.
4. Sharif K, Sharif A, Jumah F, Oskouian R, Tubbs RS. Rheumatoid arthritis in review: Clinical, anatomical, cellular and molecular points of view. Clinical Anatomy. 2018;31(2):216-23.
5. Zandman-Goddard G, Peeva E, Rozman Z, Ben-Zvi I, Langevitz P, Shvartser Y, et al. Sex and gender differences in autoimmune diseases. Sex and gender aspects in clinical medicine: Springer; 2011. p. 101-24.
6. Shlotzhauer TL. Living with rheumatoid arthritis: JHU Press; 2014.
7. Walsh DA, McWilliams DF. Mechanisms, impact and management of pain in rheumatoid arthritis. Nature Reviews Rheumatology. 2014;10(10):581-92.
8. Liao H, Wu J, Kuhn E, Chin W, Chang B, Jones MD, et al. Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. Arthritis & Rheumatism. 2004;50(12):3792-803.
9. Cohen J. Bioinformatics—an introduction for computer scientists. ACM Computing Surveys (CSUR). 2004;36(2):122-58.
10. Califf RM. Biomarker definitions and their applications. Experimental biology and medicine. 2018;243(3):213-21.
11. Iqbal N, Kumar P. From data science to bioscience: emerging era of bioinformatics applications, tools and challenges. Procedia Computer Science. 2023;218:1516-28.
12. Sulehri SA, Zunera T, Sumaira A, Rimsha RM, Syed HS, Mariam S, Madiha S, Mariam K, Khawar I. Analysis of serological proteins of breast cancer patients using bioinformatics tools. Pak-Euro J Med Life Sci. 2020;3(4):187–194..
13. Talat R, Mustafa MZ, Tanveer Z, Ahmad I, Farooq MS, Asmat R, Ali HM. Bioinformatics analysis of serologic proteins of prostate cancer patients separated by SDS-PAGE. Pak-Euro J Med Life Sci. 2018;1(1):5–11.
14. Khan AN, Hyder N, Gillani SM, Ahmad S, Asghar MA, Ali FM. Establishing neuroprotective insights: in-vivo exploration of quercetin's therapeutic potential in autism induced rats model. Pak-Euro J Med Life Sci. 2024;7(4):599–604.
15. Mendoza Ospina JE, Martínez Chavarro PA. Computational physicochemical analysis and production of OmpA's mutants with functional potential in drug delivery. 2020.
16. Schwede T, Peitsch MC. Protein Structure Modeling and Docking at the Swiss Institute of Bioinformatics. Bioinformatics: A Swiss Perspective. 2009:219.
17. Vidhya V. An in silico analysis of physicochemical characterization and protein-protein interaction network analysis of human anti-apoptotic proteins. Asian Journal of Pharmaceutics (AJP). 2018;12(04).
18. Lemmon MA. Membrane recognition by phospholipid-binding domains. Nature reviews Molecular cell biology. 2008;9(2):99-111.
19. Widmer C, Gebauer JM, Brunstein E, Rosenbaum S, Zaucke F, Drögemüller C, et al. Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. Proceedings of the National Academy of Sciences. 2012;109(33):13243-7.
20. Xu G-Y, Yu H-A, Hong J, Stahl M, McDonagh T, Kay LE, et al. Solution structure of recombinant human interleukin-6. Journal of molecular biology. 1997;268(2):468-81.
21. Dong D, Zheng L, Lin J, Zhang B, Zhu Y, Li N, et al. Structural basis of assembly of the human T cell receptor–CD3 complex. Nature. 2019;573(7775):546-52.
22. Korndörfer IP, Brueckner F, Skerra A. The crystal structure of the human (S100A8/S100A9) 2 heterotetramer, calprotectin, illustrates how conformational changes of interacting α-helices can determine specific association of two EF-hand proteins. Journal of molecular biology. 2007;370(5):887-98.
23. Sun L, Richard DY. Serum amyloid A1: structure, function and gene polymorphism. Gene. 2016;583(1):48-57.

Copyright © 2024 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.