Original Article

Predicting Rheumatoid Arthritis Associated Significant Amino-Acid Residues Using Residue-Residue Interaction Analysis

Prachi Sao[1], Anupam Singh[2], Sachidanand Singh[3]*

Abstract

Rheumatoid Arthritis (RA) is a prevalent autoimmune and inflammatory disease that requires restructuring. A lot of research information is available, but a clear etiology and drug target information is still unclear. A bottom-up approach can add more information to existing knowledge about RA. One better way of understanding the disease-related mechanism and drug objectives can be a detailed residue-residue interaction of the proteins involved with RA. In the current research work, we have studied the significant proteins reported in the Indian population that are involved in RA progression and have represented each of them as a complex network of amino acid residues to understand the significance of individual residues in the network. We implied the graph theory approach to identity central important residue, based on topological properties of the network. This approach allows us to look at a more precise method to identify potential drug targets. Our result identified leucine, phenylalanine, tyrosine, and tryptophan as essential nodes in the network, their activity was mainly connected with immune system. Understanding the function of these amino acids in CTLA4, CD40, IRF5, IL2RB, and TRAF could lead to a new treatment options in the fight against Rheumatoid Arthritis.

Keywords: Rheumatoid Arthritis; Residue-Residue Interaction; Graph Theory; Drug Target; Topological Properties; Network Biology

Introduction

Rheumatoid arthritis (RA) is an inflammatory condition in which the body's immune system attacks its own tissues and joints, causing inflammation and swelling. If left unattended for a long time, this can lead to severe illnesses, joint deterioration, impairment, which can affect internal organs and premature death. Understanding the etiology of RA is challenging because it involves a series of immune reactions involving T-cell, B-Cell, macrophages, cytokines, interleukins, tumor necrosis factors, and various other mediators which are still unknown, all migrate to synovial tissues and initiate an immune response that ultimately produces cartilage damage.

RA affects approximately 0.5% to 1% of the world's population, with the majority affecting more women than men[1]. In India, ~75% of the population is affected. Although RA has been extensively studied but still effective targets for drugs are still unknown. The current objective of RA treatment is to suppress inflammation, pain, and prevent bone destruction. While the symptoms of RA are similar to those of other types of arthritis, such as Rheumatic Fever, Arthralgia, gouty arthritis, and they all have a high

1. Prachi Sao, Faculty of Biotechnology, Institute of Biosciences and Technology,Shri Ramswaroop Memorial University, Barabanki, 225003, India

2. Anupam Singh, Bansal Institute of Engineering and Technology, NH-24, Sitapur Road, Behind Sewa Hospital, Lucknow, U.P, India-226021

3. Sachidanand Singh[3]Department of Biotechnology, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Vadlamudi, Guntur-522213, Andhra Pradesh, India.

Correspondence: Dr. Sachidanand Singh, Associate Professor, Department of Biotechnology, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Vadlamudi, Guntur-522213, Andhra Pradesh, India.(drsachinbioinfo@gmail.com; drss_bt@vignan.ac.in)

rate of morbidity and disability, it is important to distinguish between them[2,3,4].Both doctors and patients are misled about the disparities in diagnostic capabilities and treatment options available to them[5]. It is critical to comprehend the disease at the molecular level, specifically at the amino acid level, in order to provide correct diagnosis and therapy options. In general, Non-steroidal anti-inflammatory drugs (NSAIDs), Disease-modifying antirheumatic drugs (DMARDs), corticosteroids, and biologic agents are few drugs popularly used for RA. NSAIDs and corticosteroids do not prevent RA progression, while DMARDS may lower the rate of RA progression, however, the mechanism is still not fully understood as a consequence, DMARDS dosage and efficacy are still under development[6]. Looking at the rate of increasing RA cases and the side effects of available drugs, it is important to identify potential drug targets. Since RA involves many signaling pathways and molecules, a bottom-up approach has more potential to explain disease etiology and drug targets[7]. The reason behind RA is a molecule's malfunction, which ultimately leads to a series of over-expressed or under-expressed cells. A thorough understanding of proteins can, however, help describe the drug-target relationship for RA.

Literature studies have also suggested several factors causing RA, and one among them represents *Porphyromonas gingivalis* (*P. gingivalis*) a gram-negative, anaerobic bacterium. *P. gingivalis* is found in the oral cavity and it's a major periodontitis-related bacteria. One of the known causes of RA is the citrullination of the host protein by protein arginine deiminases (PADs)[8]. Peptidyl-arginine is converted to peptidyl-coralline. The citrullinated protein is recognized as a foreign particle by the immune system, which attacks it. *P. Gingivalis* produces proteinases that cleave protein from the residues of arginine and lysine. These proteinases have many adverse reactions on the host such as reducing the bactericidal effect and reducing host immune responses that proves that bacteria may be a RA-causing agent and proteinases may be a good drug target candidate[9,10]. TNFα (Tumor Necrosis Factor Alpha) is one of RA-Periodontitis connecting links. When *P.gingivalis* attacks host, the host releases TNFα, which activates various pathways like cytokine cascade, B-cell, IL-1 activation, NF-μ(nuclear factor-μ), and JNK(c-Jun N-terminal kinases). TNFα over-expression destroys bone and cartilage, so controlling TNFα activity is a significant approach in RA pathogenesis.

TRAF1 (TNF Receptor-Associated Factor 1) has adversely regulated TNFα[11] activity[12]. Citrullination affects TRAF1 activities and may contribute to RA development, which may serve as possible therapeutic targets[13]. The entire immune response changes with a change in amino acid residue and the bottom-up approach are therefore appropriate.

In most of the drug-target interaction, the target protein active site or the binding site leads to conformational and functional modifications. Residue in the protein[14] often influences the properties of these sites. It is well known that a single alteration of the residue can lead to an entirely different protein structure and function. There is little information on how the amino acid sequence forms and retains a 3D structure, so this approach provides insights into such residues that can be a more effective target for a drug. Amino acid seldom works in isolation, rather than the form of a complex network in a protein[15]. The protein's three-dimensional structure can be represented and analyzed using graph theory as a network of amino acid interactions. Where each amino acid is a node and physicochemical bonds between each amino acid are represented as edges. In an amino acid network, essential residues can be defined by analyzing its topological properties, such as centrality, node distribution, shortest path, central closeness, coefficient of clustering, and neighborhood connectivity. Residue network is dense and fewer nodes are connected and act as a communication center with other subnetworks[13,16,17,18,19,20]. Central residues may play an important role in protein function, stability, and transmission of informatio[21,22,23].

Strongly connected nodes remain close and lightly or unconnected nodes remain isolated, hence, the highest degree node plays a central role[24,25]. Node degree distribution can clarify locally important residues, yet some central parameters need to be explored to get a globally important residue[26,27]. The residue-residue networks, therefore, involve analysis with the Cytoscape Network Analyser plug-in. Network Analyser computes various network properties like Density, Closeness, Betweenness, Coefficient of Clustering, Neighborhood, Connectivity, and more. In our studies, we focus on degree centrality (DC) parameters, Closeness centrality (CC), and Betweenness centrality (BC)[28]. The degree centrality shows how many interactions a node has. More interaction means the node is important because

it is connected to other nodes or it plays an important role in holding the network together[29]. Closeness centrality emphasizes the importance of a node in communication. A chief value of the closeness centrality parameter of the node (n) strongly suggests that the node (n) can easily transmit information and itsan essential central residue. Betweenness centrality explains how important a node is for communicating with each other, thus a node of high betweenness is a crucial residue for the whole survival of the network. Progressession in RA involves many signaling pathways, these signaling paths should have a high central parameter's value, which explains its key role in cell communication and signal transduction[30].

In this study, five RA-causing proteins that have been significantly involved in immunosuppression, inflammation, and signaling processes in the Indian population are taken from the RA variome database. A network residue-residue interaction (RRI) study is conducted and significant amino acids are identified using the topological properties of graph theory. The RING web server (URL: http://protein.bio.unipd.it/ring) is used to convert protein structure to an amino acid interaction network[31] and Cytoscape is used to visualize and analyze the network. We've studied each protein as an interactive agent. A Network Analyzer plugin[32] was used to study the topology parameters of interactome residues, explaining the importance of each residue over the interactome.

Material and method

Data were retrieved from the RAvariom database[33], which includes information on genetic variants for rheumatoid arthritis based on nationality. A complete list of genetic variants for RA reported in the Indian population has been compiled. Further, through a comprehensive literature mining and GENECARD[34] analysis, only those proteins are been sorted for the further analysis which are reported as proteins coding unique genes with protein data bank (PDB)[35] structures. Hence, for the current study viz., cytotoxic T-lymphocyte-associated protein 4 (CTLA 4), Cluster of differentiation 40 (CD40), Interleukin-2 receptor subunit beta (IL2RB), interferon regulatory factor 5 (IRF5), and Tumor necrosis factor receptor (TNFR) associated factor 1 (TRAF1), are been considered which are also are listed in Table 1.
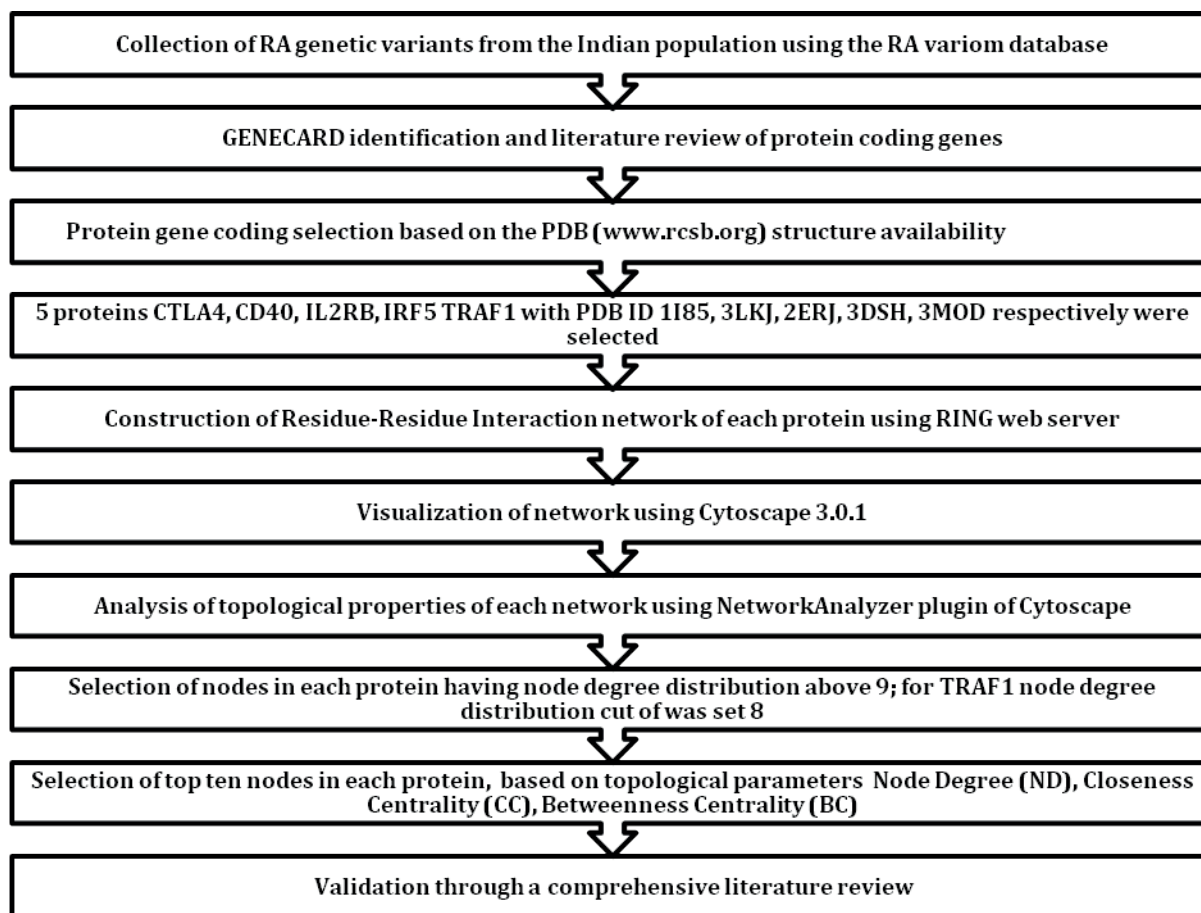
Table 1. Proteins investigated through the present studies

S.No	Protein	Name	PDB-ID	References
1.	CTLA4	Cytotoxic T-lymphocyte – associated protein 4	1I85	57
2.	CD40	CD40 molecule, TNF receptor superfamily member 5	3LKJ	58
3.	IL2RB	Interleukin-2 receptor subunit beta	2ERJ	59
4.	IRF5	Interferon regulatory factor 5	3DSH	60
5.	TRAF1	TNF receptor-associated factor 1	3M0D	61

PDB (www.rcsb.org) Identifiers 1I85, 3LKJ, 2ERJ, 3DSH, and 3M0D for all proteins were downloaded and fed to the RING (URL: http://protein.bio.unipd.it/ring)webserver, a residue interaction network generator[36] for the retrieval of residue-residue interactions (RRIs) for each protein. RING webserver transforms the protein structure into an interaction network, based on minimum distances and the type of link between the residues. The constructed network isstudied in the form of a graph (G) = (N, E) where, each residue is represented as a node (N) and the interaction between nodes as an edge (E). Since each protein is represented as a complex network, the importance of each node is represented by Node Degree Distribution, which calculates the total number of connections in which one node connects to other nodes in the network.

The individual protein network was visualized using Cytoscape version 3.0.1[37] and the centrality parameters such as Node Degree (ND), Closeness Centrality (CC), Betweenness Centrality (BC) were computed using the NetworkAnalyzer[38] plug-in of Cytoscape. Based on Node Degree Distribution from RING and Centrality parameters from NetworkAnalyser, topologically significant residues were chosen to identify important amino acid residues[17,18,39]. The resulting residues were further validated through a systematic approach to the mining of literature. Steps involved in methodology are schematically represented in figure 1.

Figure 1. A Schematic Representation of the Workflow of Predicting Rheumatoid Arthritis Associated Significant Amino-Acid Residues Using Residue-Residue Interaction Analysis

Collection of RA genetic variants from the Indian population using the RA variom database

↓

GENECARD identification and literature review of protein coding genes

↓

Protein gene coding selection based on the PDB (www.rcsb.org) structure availability

↓

5 proteins CTLA4, CD40, IL2RB, IRF5 TRAF1 with PDB ID 1I85, 3LKJ, 2ERJ, 3DSH, 3MOD respectively were selected

↓

Construction of Residue-Residue Interaction network of each protein using RING web server

↓

Visualization of network using Cytoscape 3.0.1

↓

Analysis of topological properties of each network using NetworkAnalyzer plugin of Cytoscape

↓

Selection of nodes in each protein having node degree distribution above 9; for TRAF1 node degree distribution cut of was set 8

↓

Selection of top ten nodes in each protein, based on topological parameters Node Degree (ND), Closeness Centrality (CC), Betweenness Centrality (BC)

↓

Validation through a comprehensive literature review

Results and discussion

Altered *CTLA4*, *CD40*, *IL2RB*, *IRF5*, and *TRAF1* genes are associated with RA pathogenesis. A thorough study of the amino acid residues found in each protein provided new insight into the detection of drug targets. In order to find statistically and functionally important amino acid residues, an interactome was created for each protein using the RING (URL: http://protein.bio.unipd.it/ring) web server. Followed by amino acid residue selection, based on RING node degree distribution and NetworkAnalyzer's high centrality values. The node degree distribution cut-off value was taken as nine, i.e. residues with node degree distribution higher than or equal to nine were considered for RING construction . TRAF1 has a node degree of less than nine for all amino acids, therefore residues with node degree distribution 8 have been considered. Details of node-degree distribution residues above nine are shown in Table 2.

CTLA4 (CYTOTOXIC T-LYMPHOCYTE-ASSOCIATED PROTEIN 4)

CTLA4 belongs to the Ig family and plays a significant role in the downregulation of activation of T-cells, macrophages, and osteoclast cells. CTLA4 controls inflammation and structural damage[40]

The underexpression of CTLA4 is therefore associated with RA. CTLA4 is a control point protein expressed at a very small amount on T-cell and translocated on T-cell when required to deactivate T-cell. CTLA4 competes with CD28 to bind to CD80/86 to switch off the immune response. CTLA4 protein represented as a network of 108 nodes with an average of 7.91 neighbors, for the current research, network is shown in figure 2. Top ten residues were selected based on centrality parameters, which are listed in Table 3 and sorted based on a degree because the high node degree represents high connectivity. Analyzing Tables 2 and 3 it is evident that Leucine has a node of seventeen degrees with a high proximity centrality parameter and, as a reported node with a high degree of node and proximity, can be considered as a central residue[41]. A comprehensive literature review was performed to verify the interpretations which Indicated that Leucine and Valine play a significant role as immunomodulatory protein[42]. Leucine also plays a key role in enhancing the activity of CTLA4[43]. Any alteration of leucine residue can,

Table 2. Details of the Amino Acid Having High Node Degree for Each Protein from RING web server. (Three letter code is used for amino acid and number represents the position of the residue in the amino acid sequence.)

NODE DEGREE	CTLA4	CD40	IL2RB	IRF5	TRAF1
17-14	40 Leu	256Phe,258Leu, 239Val,	24Thr	315Leu, 387Phe, 317Leu	
13-10	32Trp, 83Tyr,30Val,13Leu,72Leu, 68Trp, 28Leu	161Leu,175Val, 37Val,126Val, 225Leu, 140Trp, 193Leu, 254Thr 171Ile, 238Phe,191Ala	55Trp,20Tyr,125Gln, 26Leu,129Tyr, 145Ser	416Leu,324Leu, 389Phe, 264Leu, 404Ile, 336Trp, 263Arg, 316Ile, 385Ile, 408Val, 406Val,360Phe, 311Leu, 362Leu, 289Phe, 410Pro,365Phe, 246Tyr, 240Leu,326Ala ,359Leu, 405Thr, 383Phe, 314Gly	
9		169Tyr,231Leu,159Lys, 138Leu,223Ile,124Ala, 241Val,172Tyr,257Gly	127Val,119Tyr, 44Met, 156Trp,46Cys		
8					275Leu, 282Phe, 295Arg, 300Ala,315Glu,271Arg, 322Gln

therefore, affect the CTLA4 production and that may potentially contribute to under-expression of CTLA4 which ultimately results in an over-expressed immune system. Tyrosine is the next residue with high topological parameters and plays an important role in the internalization of CTLA4 from the plasma membrane[44]. Followed by Cystine which is involved in the activation of the T-cell. The residue referred above explains the importance of each residue in CTLA4 and its importance in the reduction of T-cell proliferation, which further regulates various immune response pathways, therefore other residues listed in Table 3 may be targeted as a key drug target for the treatment of RA.

Table 3. Details of CTLA4 protein analysis via NetworkAnalyzer plugin of Cytoscape. (list of top 10 residues based on node degree is been represented. A high value of CC (Closeness Centrality) and BC (Betweenness Centrality) highlights the importance of the residue in the network. Residue-Residue network of CTLA4 protein has 108 nodes. Three letter code is used to represent amino acid and position of each residue in the protein sequence is represented as a number before amino acid, such as 83TRY: tyrosine residue is present at position 83)

| Amino acids	Degree	CC	BC
83Tyr	16	0.41	0.11
72Leu	15	0.38	-----
32Trp	15	0.41	-----
68Trp	14	0.41	0.14
30Val	12	0.39	-----
70Leu	12	0.41	-----
13Leu	12	0.42	0.09
40Leu	12	0.41	------
85Cys	11	0.40	------

CD40 (Cluster of differentiation 40)

CD40 is part of the superfamily TNF. CD40 is very crucial for activating B-Cell. CD40 binds with other
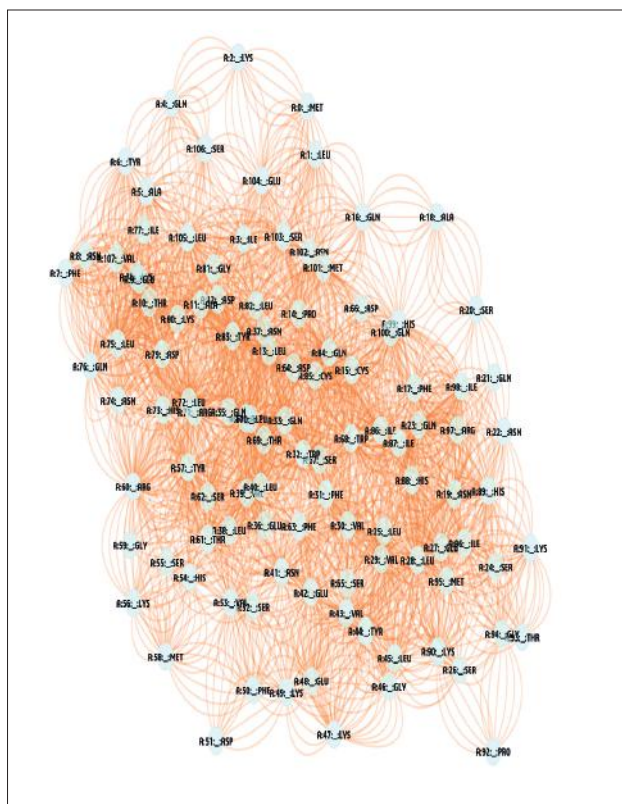
Figure. 2. Amino acid Residue-Residue Network of CTLA4

TNF family proteins and initiates various immune pathways such as Nuclear Factor Kappa B Subunit 1 (NFkb), P38, AKT, c-Jun N-terminal kinases (JNK), Signal Transducer and Activator of Transcription A (STAT), and also leads to the production of pro-inflammatory cytokines and controls the expression of co-stimulating molecules[45], which is why overexpression of CD40 leads to RA. In this study, CD40 is represented as a Residue-Residue interaction network with 128 nodes with an average of 791 neighbors, network is shown in figure 3. Top ten residues have been selected based on the centrality parameters, which are listed in Table 4 and sorted by degree parameter because the higher number of node degree represents high connectivity. According to Table 2 and 4, phenylalanine has a high value for node degree, closeness centrality, and betweenness centrality, which indicates that phenylalanine is a significant residue for CD40 and phenylalanine along with uncharged residues helps transmembrane domain CD40 to produce signals in hydrophobic and lipid-rich environments[46]. Another residue with a high centrality parameter was threonine, it plays an important role in binding CD40 to intracellular proteins, which in turn activates other signaling

cascades[47]. Also, threonine is important for M12 signaling.
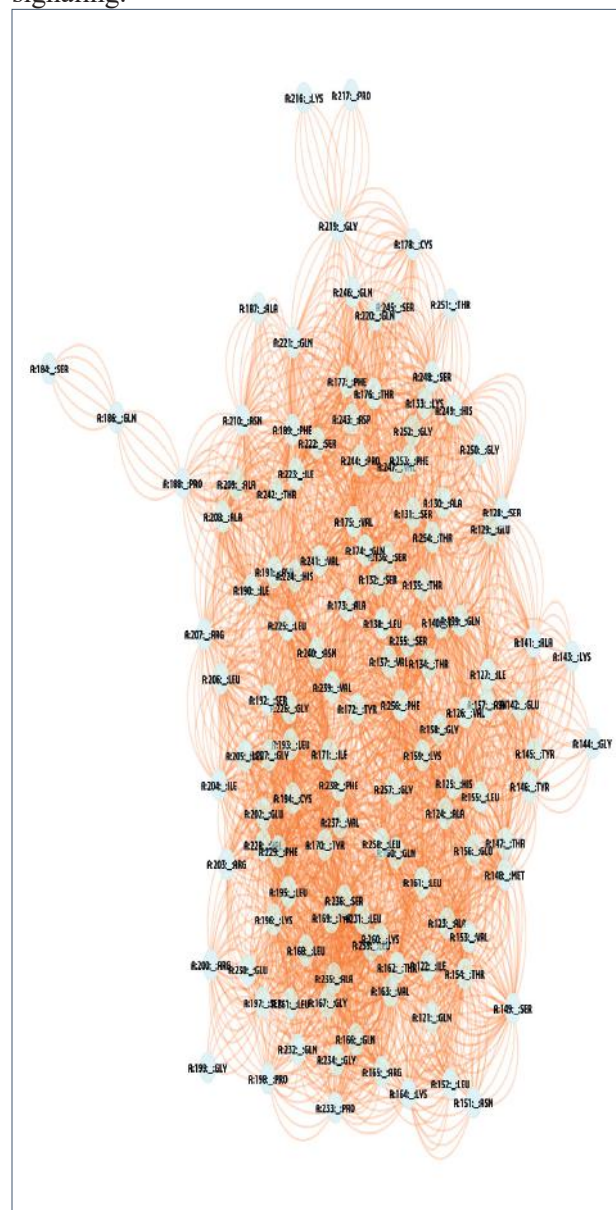
Figure 3. Amino acid Residue-Residue Network of CD40

Table 4. Details of CD40 protein analysis via NetworkAnalyzer plugin of Cytoscape. (list of top 10 residues are listed here, selected and sorted based on node degree. A high value of CC (Closeness Centrality) and BC (Betweenness Centrality highlights the importance of the residue in the network. Residue-Residue network of CD40 protein has 128 nodes. Three letter code is used to represent amino acid and position of each residue in the protein sequence is represented as a number before amino acid, such as 256Phe: phenylalanine residue is present at position 256)

Amino acids	Degree	CC	BC	NC
256Phe	17	0.42	0.11	12.29
239Val	15	0.40	0.07	11.60
175Val	14	0.37	0.06	10.79
161Leu	14	0.37	0.06	11.14
258Leu	14	0.37	0.05	11.43
254Thr	13	0.38	0.08	10.54
159Lys	13	0.38	0.06	10.77
237Val	13	0.39	0.06	-------
238Phe	12	0.37	0.05	10.33

IL2RB (INTERLEUKIN-2 RECEPTOR SUBUNIT BETA)

IL2R regulates adaptive immunity by controlling the proliferation of T_{reg} cells. The IL2R complex is associated with the activation of CD40, T-Cell, B-Cell, Natural Killer cells (NK cell) fibroblast, mono leukocytes, cytoplasmic tyrosine kinase, and other cytokines. Interleukin-2 Receptor (IL2R) is complex and has three subunits IL-2Rα, IL2-Rβ, and IL-2RΥ. Out of all three subunits, IL2-Rβ plays a key role in the activation of the IL2R complex and also triggers various signaling pathways such as mitogen-activated protein kinase (MAPK),Phosphoinositide 3-kinase (PI-3K), and Signal Transducer and Activator of Transcription 5 (STAT5)[48]. A network of 128 nodes with an average number of 7.71 neighborhoodwas constructed, network is shown in figure 4. The top eight residue-based centrality parameters listed in Table 5 were selected and sorted based on the node degree distribution. Tables 2 and 4 shows that Tyrosine has high values of node degree and proximity centrality, and the activation of different pathways of Tyrosine phosphorylation is very important [49,50]. The results of the RING web server shown in Table 2, indicates that Leucine is an important node, which is essential for the association of all three IL-2R long subunits with Tyrosine, Threonine, and Proline residues[51]. Residues listed in Table 5 can, therefore, be reported as a significant player in these target proteins.

Figure 4. Amino acid Residue-Residue Network of IL2RB

Table 5. Details of IL2RB protein analysis via NetworkAnalyzer plugin of Cytoscape. (list of top 10 residues are represented, and sorted based on node degree. A high value of CC (Closeness Centrality) and BC (Betweenness Centrality) highlights the

importance of the residue in the network. Residue-Residue network of IL2RB protein has 128 nodes. Three letter code is used to represent amino acid and position of each residue in the protein sequence is represented as a number before amino acid, such as 121Phe: phenylalanine residue is present at position 121)

Amino acids	Degree	CC	BC	NC
20Tyr	16	0.34	0.11	-----
44Met	14	0.31	0.09	9.85
24Thr	14	0.34	0.11	------
121Phe	13	0.35	0.19	------
127Val	13	0.31	0.08	10.0
117Arg	11	0.29	0.08	------
17Ala	11	------	0.10	9.56
107Pro	10	0.29	-------	10.35

IRF5 (INTERFERON REGULATORY FACTOR 5)

IRF5 belongs to a family of regulatory interferon factors, it regulates pathogen-induced acquired and innate immunity and it is involved in toll-like receptor signaling, B-cell receptor signaling. It plays an essential function in the DNA repair process. Ubiquitination and phosphorylation activate IRF5, which transcribes pro-inflammatory cytokines and type1 INF. IRF5 also determines the ultimate fate of macrophage phenotype[52]. Hence overexpression of IRF5 is associated with RA. A 234-node network was constructed with an average of 7.70 neighbors, network is shown in figure 5. Top nine centrality-based residue have been selected and are specified in Table 6 which are sorted based on node degree. The result is that Serine and Threonine have good value for the parameter of centrality. Serine and Threonine phosphorylation is essential for the dimerization of IRF5 and nuclear translocation. Leucine, Lysine, Serine, and Valine are necessary to establish a hydrophobic environment[53] as shown in Table 6.

Table 6. Details of IRF5 protein analysis via NetworkAnalyzer plugin of Cytoscape. (list of top 10 residues are represented, select, and sorted based on node degree. A high value of CC (Closeness Centrality) and BC (Betweenness Centrality highlights the importance of the residue in the network. Residue-Residue network of IRF5 protein has 234 nodes. Three letter code is used to represent
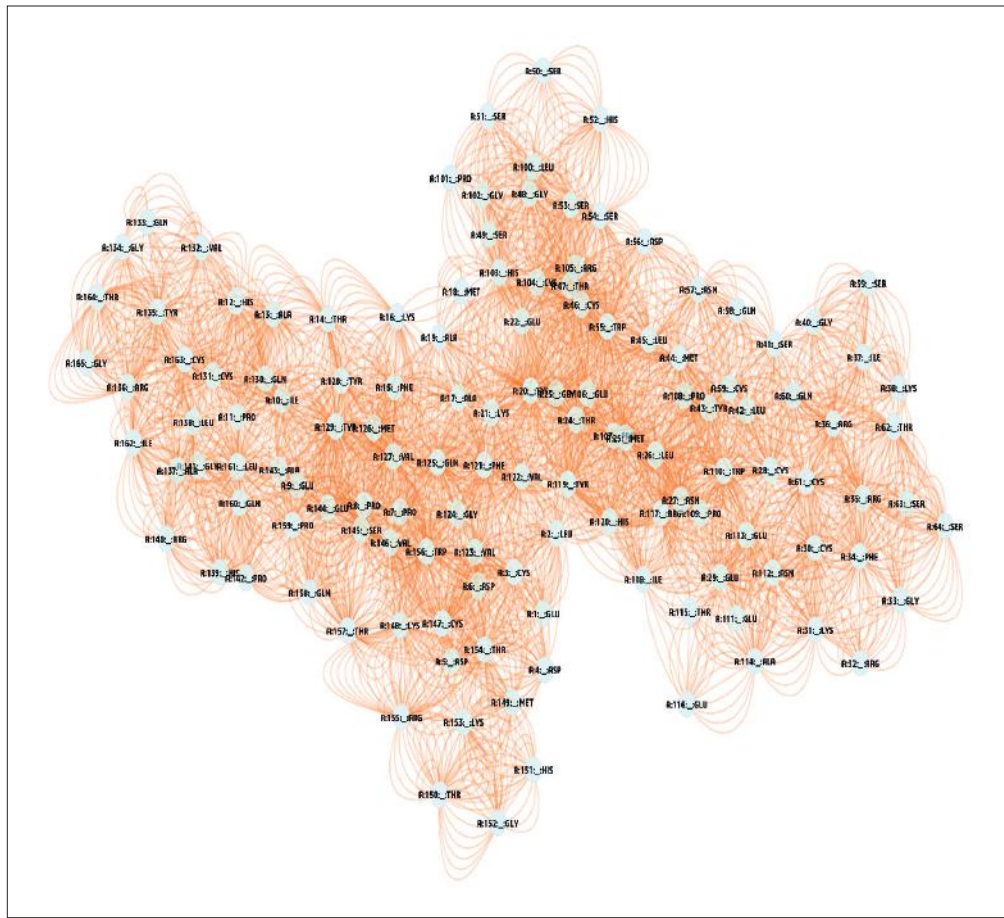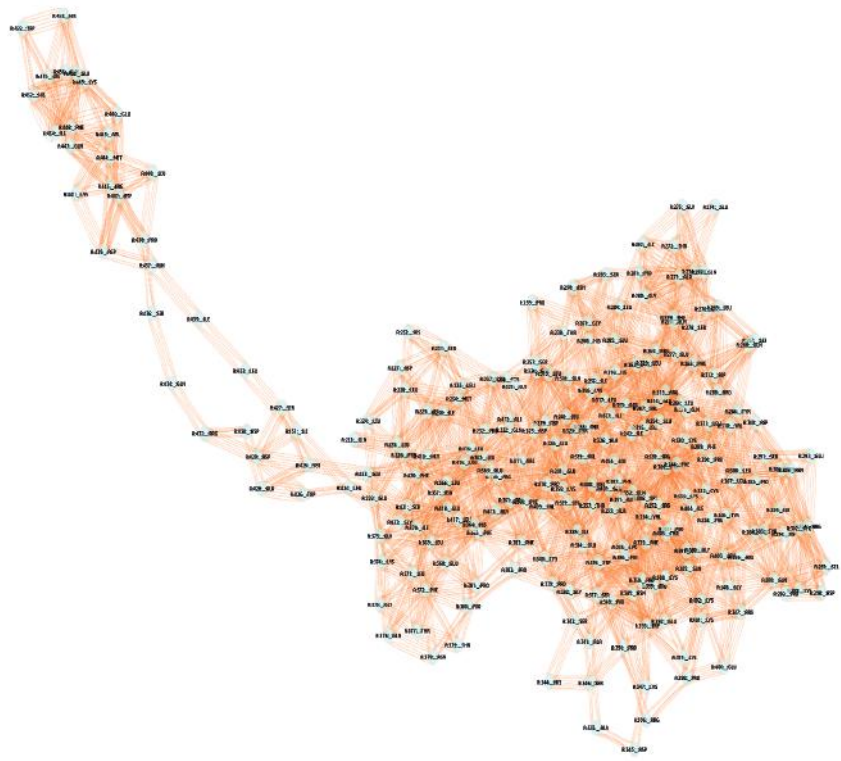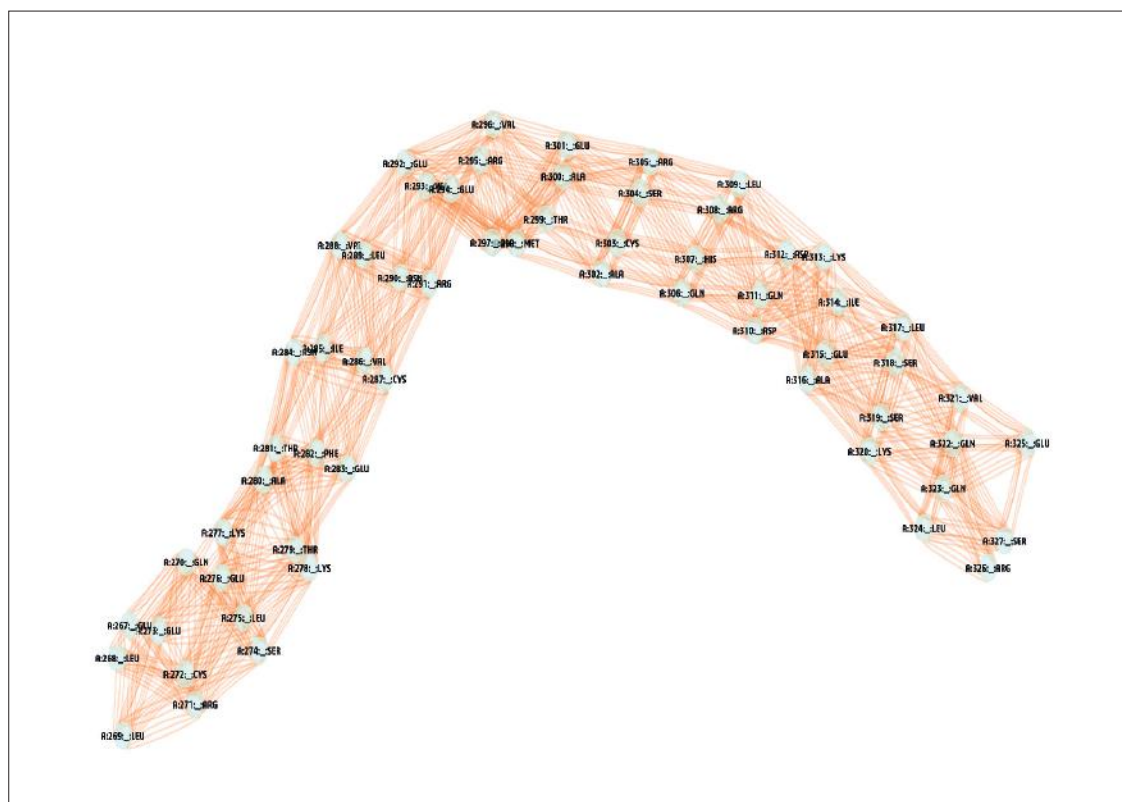
Figure 5. Amino acid Residue-Residue Network of IRF5

Figure 6. Amino acid Residue-Residue Network of TRAF1

amino acid and potion of each residue in the protein sequence is represented as a number before amino acid, such as 383Phe: phenylalanine residue is present at position 383)

Amino acids	Degree	CC	BC	NC
387Phe	**14**	0.26	-------	13.0
417Leu	13	0.23	0.14	------
406Val	12	0.25	--------	12.41
383Phe	11	0.25	0.15	10.91
410Pro	11	0.25	------	11.27
337Ser	10	0.22	------	10.44
413Ala	9	0.23	------	12.22
339Pro	9	0.24	------	10.67
358Lys	9	0.24	-------	11.11

TRAF1 (TNF RECEPTOR-ASSOCIATED FACTOR 1)

TRAF1 is a receptor-associated tumor necrosis factor that regulates cell survival, proliferation, and death. TRAF1 is an adjuster molecule instead of an enzyme. This regulatory molecule is used for different signaling pathways, including mitogen-activated protein kinase (MAPK), c-Jun N-terminal kinases (JNK), Nuclear factor kappa B (NF-kB) as well as for signaling anticonception[54]. The 60 node network with an average neighboring number of 6.33 has been built and the result is that serine has good closeness centrality value because serine phosphorylation plays an important role in activating the NF-kappaB and JNK pathway[55], network is shown in figure 6. whereas Arginine and Glutamic acid have a high value for the centrality parameters and it is involved in the formation of the salt bridge. Leucine and Valine have a good value for the centrality parameters and involved in the establishment of hydrophobic patches. Both salt bridges and hydrophobic patches are important for the stable structure of TRAF1[56]. Bacteria generates a PAD enzyme which can be a cause for aggressive RA by citrullinating arginine residues of TRAF1. This could be one cause of clearly reflecting arginine as a residue of high value in Table 2 and Table 7 of all central parameters with

a high node degree.

Table 7. Details of TRAF1 protein analysis via NetworkAnalyzer plugin of Cytoscape. (list of top 10 residues are represented, and sorted based on node degree. A high value of CC (Closeness Centrality) and BC (Betweenness Centrality) highlights the importance of the residue in the network. Residue-Residue network of TRAF1 protein has 60 nodes. Three letter code is used to represent amino acid and position of each residue in the protein sequence is represented as a number before amino acid, such as 300Ala: alanine residue is present at position 30)

Amino acids	Degree	CC	BC	NC
295Arg	8	0.26	0.27	7.00
300Ala	8	------	0.27	6.63
291Arg	7	0.24	0.27	------
305Arg	7	0.24	0.21	------
298Met	7	0.25	------	6.86
287Cys	7	0.23	------	6.71
304Ser	6	0.24		6.83
294Glu	6	0.24		7.00
289leu	6	0.23		7.00

Conclusion

This work illuminates the importance of the amino acid residues interaction network for any diseases associated target proteins. The methodical analysis supports the fact that residue with high centrality parameters plays a central role in protein function and structure, therefore, residues identified by the Residue-Residue Interaction network may be represented as potential drug targets for RA. Each protein has been interactively studied and combiningly the result reports that Leucine has a high node degree in all five proteins, hence it is very crucial for the overall function of all five proteins and therefore it may be playing a significant role

in binding affinities and protein folding of these proteins which are represented as potential drug targets. The node degree distribution and centrality values were also very promising for Valine, Tyrosine, Phenylalanine, and tryptophan and this confirmed that they are also significant amino acid residues. 32 branched-chain amino acids such as Valine and Leucine have been already reported of having a highly immunomodulatory effect explaining the significance of this experiment. Tyrosine phosphorylation plays a crucial role in activating and surviving various immune molecules and their associated pathways, while tryptophan is extremely important concerning antibody-mediated immunity. The current study can be made more significant by examining already reported RA proteins, to understand the importance and the relationship between specific amino acids of each protein on their respective pathways.

Authors Contribution:

Data gathering and idea owner : Prachi Sao, Dr. Anupam Singh, Dr. Sachidanand Singh

Writing, editing final draft and Submitting of the manuscript: Prachi Sao

Approval of final draft: Dr. Sachidanand singh

Conflicts of interest

None of the authors have any conflict of interest. We also would like to declare that we do not have any competing interests.

References

1. Handa, R. *et al.* Literature review of rheumatoid arthritis in India. *Int. J. Rheum. Dis.* **19**, 440–451 (2016).

2. DAWN, Indranil; BISWAS, Gouranga. Assessment of serum Ischemia Modified Albumin (IMA) Levels in Acute Rheumatic Fever. International Journal of Human and Health Sciences (IJHHS), [S.l.], v. 5, n. 2, p. 222-225, oct. 2020. ISSN 2523-692X. Available at: <https://ijhhsfimaweb.info/index.php/IJHHS/article/view/263>. Date accessed: 14 oct. 2021. doi:http://dx.doi.org/10.31344/ijhhs.v5i2.263.

3. Rhahman, A. F. A., Mohamad, N. I., & Teoh, S. H. (). An uncommon case of gouty arthritis in a teenage boy. *Bangladesh Journal of Medical Science*, 2021;**20**(1): 200–203. https://doi.org/10.3329/bjms.v20i1.50372

4. Sao, P. et al. Putative Target Identification for Gout; A NetworkBiologyApproach.https://www.ingentaconnect.com/content/asp/jobn/2013/00000007/00000006/art00006;jsessionid=67m9m8k0ps5j8.x-ic-live-03 (2013) doi:info:doi/10.1166/jbns.2013.1166

5. Soomro, M. H., Magsi, M., Soomro, M. A., Akram, M., & Lahmar, O. (). Patients' Knowledge on Rheumatoid Arthritis presenting with Arthralgia in a Tertiary Care Teaching Hospital, Pakistan. *Bangladesh Journal of Medical Science*, 2019;**18**(4): 808–813. https://doi.org/10.3329/bjms.v18i4.42909

6. Di, Y. M., Zhou, Z.-W. & Zhou, C. G. L. and S.-F. Current and Future Therapeutic Targets of Rheumatoid Arthritis. *Anti-Inflammatory & Anti-Allergy Agents in Medicinal Chemistry* 2011;**10**: 92–120 http://www.eurekaselect.com/68675/article.

7. Singh, S., Vennila, J. J. & Kant, R. Gene interaction map: a paradigm for identifying significant pathways responsible for rheumatoid arthritis. *Netw. Model. Anal. Health Inform. Bioinforma* 2014;.**1**: 1–7.

8. Maresz, K. J. *et al.Porphyromonas gingivalis* Facilitates the Development and Progression of Destructive Arthritis through Its Unique Bacterial Peptidylarginine Deiminase (PAD). *PLoSPathog.* **9**, e1003627 (2013).

9. Travis, J., Pike, R., Imamura, T. & Potempa, J. Porphyromonas gingivalis proteinases as virulence factors in the development of periodontitis. *J. Periodontal Res.* 1997;**32**:120–125.

10. Sao, P., Chand, Y., Kumar, A., & Singh, S. (. Potential Drug Target Identification in Porphyromonas gingivalis using In-silico Subtractive Metabolic Pathway Analysis. *Bangladesh Journal of Medical Science*, 2021;**20**(4), 887–896. https://doi.org/10.3329/bjms.v20i4.54149

11. Tsitsikov, E. N. *et al.* TRAF1 Is a Negative Regulator of TNF Signaling: Enhanced TNF Signaling in TRAF1-Deficient Mice. *Immunity* **1**2001;**5**: 647–657 ().

12. Persson, G. R. Rheumatoid arthritis and periodontitis – inflammatory and infectious connections. Review of the literature. *J. Oral Microbiol.* 2012;**4**: 11829 ().

13. Quirke, A.-M., Fisher, B. A. C., Kinloch, A. J. & Venables, P. J. Citrullination of autoantigens: Upstream of TNFα in the pathogenesis of rheumatoid arthritis. *FEBS Lett.* 2011;**585**: 3681–3688

14. Singh, S., Vennila, J. J., Snijesh, V. P., George, G. & Sunny, C. Implying Analytic Measures for Unravelling Rheumatoid Arthritis Significant Proteins Through Drug–Target Interaction. *Interdiscip. Sci. Comput. Life Sci.* 2016;**8**: 122–131 (

15. Li, Y. *et al.* Predicting disease-associated substitution of a single amino acid by analyzing residue interactions. *BMC Bioinformatics* 2011;**12**: 14.

16. Singh, S. Network Biology Approach for Identifying Significant Drug Targets and Pathways for Rheumatoid Arthritis. *University* (2015).

17. Snijesh, V. P., Matchado, M. S. & Singh, S. Classifying Rheumatoid Arthritis gene network signatures for identifying key regulatory molecules and their altered pathways by adopting network biology approach. *Gene Rep.* 2018;**13**: 199–211

18. V.P., Snijesh& Singh, Sachidanand. (2014). Network Based Approach for Understanding the Chondrocyte Biology for Rheumatoid Arthritis. *International Journal of Innovative Research in Computer and Communication Engineering.*

19. Blessia, T. F., Singh, S., & Vennila, J. nwinding the Novel Genes Involved in the Differentiation of Embryonic Stem Cells into Insulin-Producing Cells: A Network-Based Approach. Interdisciplinary Sciences: Computational Life Sciences,2017; **9**(1): 88–95. https://doi.org/10.1007/s12539-016-0148-9.

20. Rheumatoid Arthritis Candidate Genes Identification by Investigating Core and Periphery Interaction Structures. (n.d.). Springerprofessional.De. Retrieved October 17, 2021, from https://www.springerprofessional.de/en/rheumatoid-arthritis-candidate-genes-identification-by-investiga/4862424.

21. Miyazawa, S. & Jernigan, R. L. Residue – Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading. *J. Mol. Biol.* 1996;**256**:623–644 ().

22. Snijesh, V. P., & Singh, S. Molecular Modeling and Network Based Approach in Explaining the Medicinal Properties of Nyctanthesarbortristis, Lippianodiflora for Rheumatoid Arthritis. Journal of Bioinformatics and Intelligent Control,2014;**3**(1),: 31–38. https://doi.

org/10.1166/jbic.2014.1072

23. Singh, S. (2015). Network Biology Approach for Identifying Significant Drug Targets and Pathways for Rheumatoid Arthritis. University. http://shodhganga.inflibnet.ac.in:8080/jspui/handle/10603/106794

24. Xie, J., Chen, M. & Szymanski, B. K. LabelRankT: incremental community detection in dynamic networks via label propagation. in *Proceedings of the Workshop on Dynamic Networks Management and Mining - DyNetMM '13* 25–32 (ACM Press, 2013). doi:10.1145/2489247.2489249.

25. Singh, S., Jannet Vennila, J., & Kant, R. Gene interaction map: A paradigm for identifying significant pathways responsible for rheumatoid arthritis. *Network Modeling Analysis in Health Informatics and Bioinformatics*,2014;**3**(1), 62. https://doi.org/10.1007/s13721-014-0062-1

26. George, G., Singh, S., Lokappa, S. B. & Varkey, J. Gene co-expression network analysis for identifying genetic markers in Parkinson's disease - a three-way comparative approach. *Genomics2019;***111**: 819–830

27. Singh, S., Vennila, J. J., Snijesh, V. P., George, G., & Sunny, C. (2016). Implying Analytic Measures for Unravelling Rheumatoid Arthritis Significant Proteins Through Drug–Target Interaction. Interdisciplinary Sciences: *Computational LifeSciences*,2016;**8**(2):122–131. https://doi.org/10.1007/s12539-015-0108-9

28. Fokas, A. S., Cole, D. J., Ahnert, S. E., & Chin, A. W. Residue Geometry Networks: A Rigidity-Based Approach to the Amino Acid Network and Evolutionary Rate Analysis. *Scientific Reports*,2016;**6**(1): 33213. https://doi.org/10.1038/srep33213

29. del Sol, A., Fujihashi, H., Amoros, D. &Nussinov, R. Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families. *Protein Sci. Publ. Protein Soc.2006;***15**: 2120–2128 (2006).

30. Pavlopoulos, G. A. *et al.* Using graph theory to analyze biological networks. *BioData Min.***4**, 10 (2011).

31. Piovesan, D., Minervini, G. &Tosatto, S. C. E. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.2016;***44**: W367–W374 (2016).

32. Csermely, P. Creative elements: network-based predictions of active centres in proteins, cellular and social networks. 11.

33. Nagai, Y. & Imanishi, T. RAvariome: a genetic risk variants database for rheumatoid arthritis based on assessment of reproducibility between or within human populations. *Database***2013**, (2013).

34. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses: The GeneCards Suite. in *Current Protocols in Bioinformatics* (eds. Bateman, A., Pearson, W. R., Stein, L. D., Stormo, G. D. & Yates, J. R.) 1.30.1-1.30.33 (John Wiley & Sons, Inc., 2016). doi:10.1002/cpbi.5.

35. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.2016;***28**: 235–242

36. Martin, A. J. M. *et al.* RING: networking interacting residues, evolutionary information and energetics in protein structures. *Bioinforma. Oxf. Engl.*1011**;7**: 2003–2005

37. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.*2003**;3**: 2498–2504

38. Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinforma. Oxf. Engl 2008;***24**: 282–284

39. Chand, Y. *et al.* Prioritizing potential diagnostic biomarkers of Alzheimer's disease by investigating gene expression data: A network-based approach. *Alzheimers Dement.2020;***16**, e044322

40. Cutolo, M., Sulli, A., Paolino, S. &Pizzorni, C. CTLA-4 blockade in the treatment of rheumatoid arthritis: an update. *Expert Rev. Clin. Immunol.2016;***12**:417–425 (2016).

41. Amitai, G. *et al.* Network analysis of protein structures identifies functional residues. *J. Mol. Biol2004;.***44**: 1135–1146 (2004).

42. Hou, H., Fan, Y., Li, B., Xue, C. & Yu, G. Preparation of immunomodulatory hydrolysates from Alaska pollock frame. *J. Sci. Food Agric2012;.***92**: 3029–3038

43. El-Far, M. *et al.* Down-Regulation of CTLA-4 by HIV-1 Nef Protein. *PLoS ONE***8**, e54295 (2013).

44. Shiratori, T. *et al.* Tyrosine Phosphorylation Controls Internalization of CTLA-4 by Regulating Its Interaction with Clathrin-Associated Adaptor Complex AP-2. *Immunity 1997;***6**:583–589 (1997).

45. Elgueta, R. *et al.* Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol. Rev 2009;.***229**, 152–172 (2009).

46. Bock, J. &Gulbins, E. The transmembranous domain of CD40 determines CD40 partitioning into lipid rafts. *FEBS Lett2003;.***534**: 169–174

47. Hostager, B. S., Hsing, Y., Harms, D. E. & Bishop, G. A. Different CD40-mediated signaling events require distinct CD40 structural features. 8.

48. Gruss, H. J., Scott, C., Rollins, B. J., Brach, M. A. & Herrmann, F. Human fibroblasts express functional IL-2

receptors formed by the IL-2R alpha- and beta-chain subunits: association of IL-2 binding with secretion of the monocyte chemoattractant protein-1. 8.

49. Park, S. J. *et al.* Impact of IL2 and IL2RB genetic polymorphisms in kidney transplantation. *Transplant. Proc2011;.***43**:2383–2387

50. Asao, H., Kumaki, S., Takeshita, T., Nakamura, M. &Sugamura, K. IL-2-dependent in vivo and in vitro tyrosine phosphorylation of IL-2 receptor γ chain. *FEBS Lett.*1992;**304**:141–145 ().

51. Stauber, D. J., Debler, E. W., Horton, P. A., Smith, K. A. & Wilson, I. A. Crystal structure of the IL-2 signaling complex: Paradigm for a heterotrimeric cytokine receptor. *Proc. Natl. Acad. Sci.2006;***103**: 2788–2793

52. Platanitis, E. & Decker, T. Regulatory Networks Involving STATs, IRFs, and NFκB in Inflammation. *Front. Immunol. 1018"***9**:

53. Chen, W. *et al.* Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat. Struct. Mol. Biol.2008;***15**: 1213–1220

54. Qiao, Y. Q. *et al.* Gene expression of tumor necrosis factor receptor associated-factor (TRAF)-1 and TRAF-2 in inflammatory bowel disease: TRAF-1 and TRAF-2 in IBD. *J. Dig. Dis.1013;***14**:244–250

55. Kato, T., Gotoh, Y., Hoffmann, A. & Ono, Y. Negative regulation of constitutive NF-kappaB and JNK signaling by PKN1-mediated phosphorylation of TRAF1. *Genes Cells Devoted Mol. Cell. Mech.2008;***13**:509–520

56. Kim, C. M. *et al.* Crystal structure of TRAF1 TRAF domain and its implications in the TRAF1-mediated intracellular signaling pathway. *Sci. Rep.***6**, 25526 (2016).

57. Flores-Borja, F., Jury, E. C., Mauri, C. & Ehrenstein, M. R. Defects in CTLA-4 are associated with abnormal regulatory T cell function in rheumatoid arthritis. *Proc. Natl. Acad. Sci2008;***105**:19396–19401 (2008).

58. van der Linden, M. P. M. *et al.* Association of a single-nucleotide polymorphism in *CD40* with the rate of joint destruction in rheumatoid arthritis. *Arthritis Rheum2009;.***60**: 2242–2247 (2009).

59. Akhmed, N. & Lichtenstein, J. R. MRSA infections in patients treated with tumour necrosis factor inhibitors. *Ann. Rheum. Dis2009;.***68**: 1788–1789

60. Rueda, B. *et al.* Analysis of IRF5 gene functional polymorphisms in rheumatoid arthritis. *Arthritis Rheum.2006;***54**: 3815–3819

61. Plenge, R. M. *et al.* TRAF1-C5 as a Risk Locus for Rheumatoid Arthritis-A Genomewide Study. doi:10.1056/NEJMoa073491.

―――――――