INTERNATIONAL RESEARCH JOURNAL OF PHARMACY
www.irjponline.com
ISSN 2230 – 8407

Review Article

PROTEIN MODELLING OF THE GENES OF ARTHRITIS: A REVIEW

Jerine Peter S, Nagesh Kishan Panchal, Bapna I, Kulkarni S, Pratyusha S, Padma Thiagarajan, Sabina Evan Prince *
School of Biosciences and Technology, VIT, Vellore, Tamil Nadu, India
*Corresponding Author Email: eps674@gmail.com

Article Received on: 03/03/20 Approved for publication: 02/04/20

DOI: 10.7897/2230-8407.110435

ABSTRACT

Arthritis is a medical condition which causes painful inflammation and stiffness of the joints. Apart from joints, tendons, ligaments, bones and muscles are also seen to be affected. In this study we are mainly focussing on the genes causing Rheumatoid arthritis (RA) and gout. Rheumatoid arthritis is one of the most common types of autoimmune arthritis affecting the joints of hands and feet. On the other hand, gout is caused by accumulation of crystals in the joints. The list of genes that could cause arthritis was obtained from review of literature. This was followed by acquiring information about the protein structure with the help of RCSB PDB (PDB-101). SNPs for the protein without the structure were recovered through the SNP database (dbSNP) from the NCBI page. For them, the structure was predicted using I-TASSER software which is one of the most common approaches to protein structure and function prediction. Through the literature review 53 genes affecting RA were found. There 19 of them which were further checked for single nucleotide variation and protein structure prediction. Through our study we have modelled the protein for the genes without a structure which could be used to find binding affinity against active compounds of plants. This would help in drug discovery through target inhibition.

Keywords: Arthritis, Rheumatoid arthritis, SNP, I-TASSER, protein modelling, RCSB-PDB

INTRODUCTION

Rheumatoid Arthritis is a chronic, progressive, autoimmune disorder which is caused by inflammation of synovial membrane, which affects joints, causing inflammation, swelling and pain in and around the joints and other body organs[1]. It usually affects hands and feet and may result in complications like low blood count, inflammation around the lungs and heart, cardiovascular disease, osteoporosis, interstitial lung disease, mental disabilities like depression, cancer, trouble working and fatigue. It can also lead to peripheral neuropathy and mononeuritis multiplex. It occurs with a frequency of 0.5-1% of adults in the developed world and the onset is gradual/sudden, but most frequent during middle age[2]. Frequency of occurrence in women is 2.5 times the frequency in men. Symptoms include joint pain, stiffness, redness, swelling, decreased range of motion of affected joints. The cause of Rheumatoid Arthritis involves genetic as well as environmental factors. There are also other types of Arthritis like Gout, Osteoarthritis (degenerative joint disease). Gout is a form of inflammatory arthritis which is caused by accumulation of uric acid crystals in the joints. It affects the metatarsal-phalangeal joint at the base of big toe and also affects heels, knees, wrists and finger joints[3]. It may result in kidney stones or kidney damage. The usual onset is in the older males and post-menopausal women. It occurs with a frequency of 1-2% of adults in the developed world. The aim of this research is to acquire information and to bring to light all the genes which could probably lead to the cause of Rheumatoid Arthritis and Gout and then the protein modelling of the genes without the 3D structure. Of all the genes involved, most significant are the HLA (Human Leukocyte Antigen) gene variations, like the HLA-DRB1 gene. Large genome-wide association studies have identified more than 30 loci involved in the pathogenesis of Rheumatoid Arthritis. Protein modelling is basically the prediction of 3D structure of a target protein, from the amino acid sequence of a homologous protein for which an X-ray or NMR structure is available[4].

Protein modelling is necessary as the functional properties of the proteins depend on their 3D structures and deciphering the 3D structure of a protein from its amino acid sequence helps in understanding the properties of the protein. Protein modelling is done by first looking at the single nucleotide variations in the genes. Single Nucleotide Polymorphism or SNPs are the most common type of genetic variation which occurs due to substitution of a single nucleotide that occurs at a specific position in the genome. Single Nucleotide Variations can occur within the coding regions of the gene, non-coding regions and also the intergenic regions[5]. The SNPs which occur within the coding regions may be of synonymous or non-synonymous types. The non-synonymous SNPs include ‘mis-sense’ or ‘nonsense’ variations, which change the amino acid sequence of the protein. The mis-sense SNP variations for the proteins without the 3D structure can be obtained through the SNP database. The SNP database (db SNP) is a free online resource and a public archive for genetic variations within and across several species, which is developed by National Centre for Biotechnology Information (NCBI). It helps for the research in association of genetic variations with phenotypic traits. The structure of these genes with SNP variations can now be predicted using I-TASSER software, which is an on-line platform which allows to automatically generating high-quality model predictions of 3D structures and functions of protein molecules. I-TASSER is a bioinformatics method used to predict the 3D structure model of protein molecules from amino acid sequences. It detects the structure templates from the Protein Data Bank (PDB) by a technique called fold recognition[6].

Gene identification

The very first step in this experiment is being the search for genes which cause Rheumatoid and gout arthritis. It is an important step so as to proceed further with research. Genes responsible for this were found out through literature survey and referring to previously done research on this topic. There were total 54 genes found to play a role in the causation of these conditions. Out of which 53 are pertaining to rheumatoid arthritis and 2 for gout arthritis. Here is the list of genes responsible for the rheumatoid arthritis HLA-DRB major histocompatibility complex, class 2, DR beta 1; AFF3 AF4/FMR2 family member 3; ARID5B AT-rich interaction domain 5B; BLK proto oncogene SRC family tyrosine kinase; complement C5; CCL21c-c motif chemokine ligand 21; CCR6 c-c motif chemokine receptor 6; CD2 molecules; CD5 molecules; CD28 molecule; CD40 molecule; CD58 molecule; CTLA4 cytotoxic T-lymphocyte associated protein 4; FCGR2A Fc fragment of IgG receptor 2a; FCGR2B Fc fragment IgG receptor 2b; GATA3 binding protein 3; HLA-B major histocompatibility complex class 1B; HLADP-B1 major histocompatibility complex class 2 DP beta1; HLA-DR-B1 major histocompatibility complex class 2 DR beta 1; IKZF3 IKAROS family zinc finger 3; IL2 interleukin; IL2RA interleukin 2 receptor subunit alpha; IL2RB interleukin 2 receptor subunit beta; IL6R interleukin 6 receptor; IL6ST interleukin 6 signal transducer; interleukin 21; IRAK1 interleukin 1 receptor associated kinase 1; IRF5 interferon regulatory factor 5; IRF8 interferon regulatory factor 5A; KIF5A kinesin family member 5A; NFKB inhibitor like 1; PADI4 peptidyl arginine deiminase 4; PIPK2, POU3F1 POU class 3 homeobox 1; PRDM1 PR/SET domain 1; PRKCQ protein kinase C theta; PTPN22 protein tyrosine phosphatase non-receptor type 22; PTPRC protein tyrosine phosphatase receptor type C; PXK PX domain containing serine/threonine kinase like; RASGRP1 RAS guanyl releasing protein 1; RBPJ recombination signal binding protein for immunoglobulin kappa J region; RCAN1 regulator of calcineurin; REL proto oncogene NF-kB subunit; RUNX1 family transcription factor 1; SPRED2 sprout related EVH1 domain containing 2; STAT4 signal transducer and activator of transcription 4; TAGAP T-cell activation RhoGTPase activating protein; TLE3 TLE family member 3 transcriptional corepressor; TNFAIP3 TNF alpha induced protein 3; TNFRSF14 TNF receptor superfamily member 14; TRAF1 TNF receptor associated factor 1; TRAF6 TNF receptor associated factor 6; TYK2 tyrosine kinase. Those related with gout arthritis are ABCG2 ATP binding cassette subfamily G member 2 and SLC2A9 solute carrier family 2 members 9. This information was obtained from nih.ncbi site. The next is classification of the above genes as the ones having a protein structure and others which do not had a structure in the databases. Uni-prot which is protein data bank wasused to look for genes having their protein structures and those without any were noted. CCR6 did not have any structure.

SNP analysis of genes

The single nucleotide polymorphism (SNPs) was to be analysed for the genes. For the gene CCR6 five isoforms were obtained having a single nucleotide change at the 369th base. The clinical significance of genes showed no particular role in the rheumatoid or gout arthritis. Go to NCBI, apply the filter for SNPdb, and search for the particular gene in the search tab, in functions column choose for missense and click the first result in homo-sapiens. All the gene related information will be shown, check for clinical significance if any found in Clin Var, copy the SNP forms and make an excel sheet in columns as molecule type, change in sequence, amino acid and clinical significance.

Protein modelling

Those genes which do not have a structure need to model. This was done using iTASSER. CCR6 having 5 different forms were to be submitted on it Asser to obtain a modelled protein. Go to uniprot, enter the gene, from the given information the FASTA format was downloaded using BLAST, copied to Microsoft word, check for the change in sequence by referring to the excel sheet made, the missense variant taken, made the same change in the copied sequence in Microsoft word; registration in iTASSER, copy paste the sequence with missense and submit for protein modelling; results to be obtained within a month.

Quality Check of the protein model

The quality check of protein structures is done either using RAMPAGE by Ramchandran plot analysis or by PDB sum. Ramchandran plot is mainly to determine the role of amino acids in the secondary structure of a protein. It can also be used for protein crystal verification. The rotations of the main chain in a protein are given by torsion angles phi and psi. These angles correspond to conformations which are sterically disallowed of the polypeptide backbone. The plot is an important tool to analyse the protein structure. PDB sum is a database provides a wide illustrated summary of the operative information about every macromolecular deposited in the Protein Data Bank. The protein model obtained using I-TASSER, is to be uploaded in the PDB sum and checked the quality. The model having the highest percentage has to bechosen as the best model. If the Ramchandran plot shows percentage above 85% for a given model then it is considered to be a good quality model, which can be further used for docking.

Protein model analysis of the isoform NP_113597.2:p.Ala369Glu (A (Ala) > E (Glu)) [isoform 1] of the gene, C-C Chemokine Receptor type 6 (CCR6)

Five protein models of the isoform 1 of the gene, CCR6, were obtained through the I-TASSER software, for which quality check was carried out using PDB sum database and RAMPAGE software. Quality of protein (Figure 1) found for Model 1 was 82.8%, Model 2 was 82.3%, Model 3 was 80.9%, Model 4 was 82.0%, Model 5 was 81.7%. Model 1 with 82.8% is the highest and can be considered as the best of all models to be used fordocking. The C-Score of Model 1 is -0.59, Model 2 is -1.03, Model 3 is -0.71, Model 4 is -3.45, Model 5 is - 4.00. PDB structure and ligand binding site of the best model was represented in Figure 2. Table 1 represent the ligand binding residues.

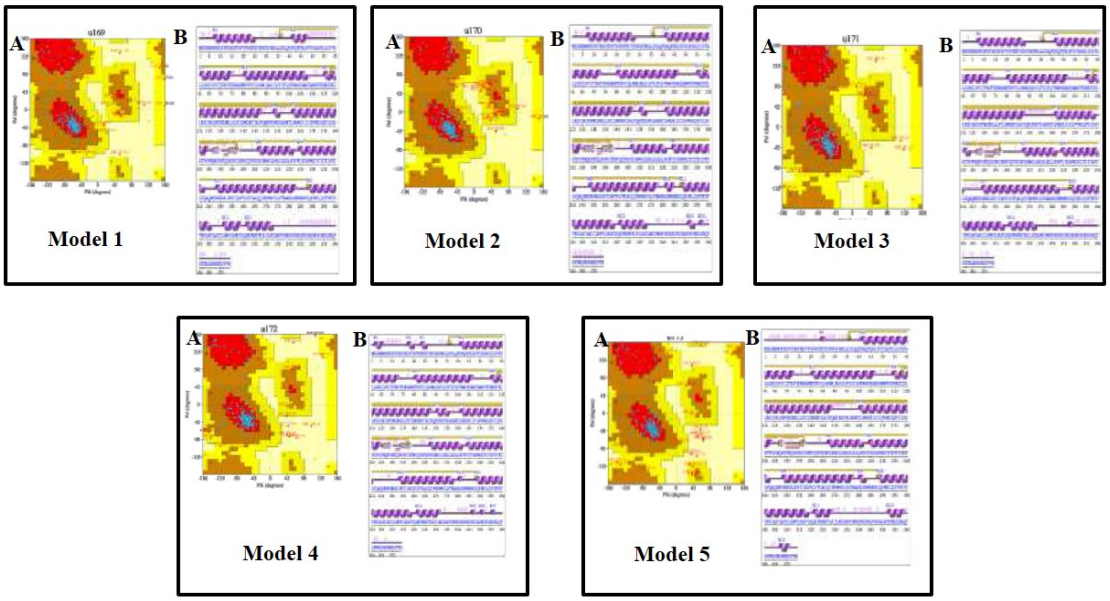

Figure 1: A- represents the Ramachandran Plot images of all the protein models of the Isoform NP_113597.2:p.Ala369Glu of the CCR6 gene, respectively. B- Represents the secondary residue images of all the 5 protein models of the Isoform NP_113597.2:p.Ala369Glu of the CCR6 gene

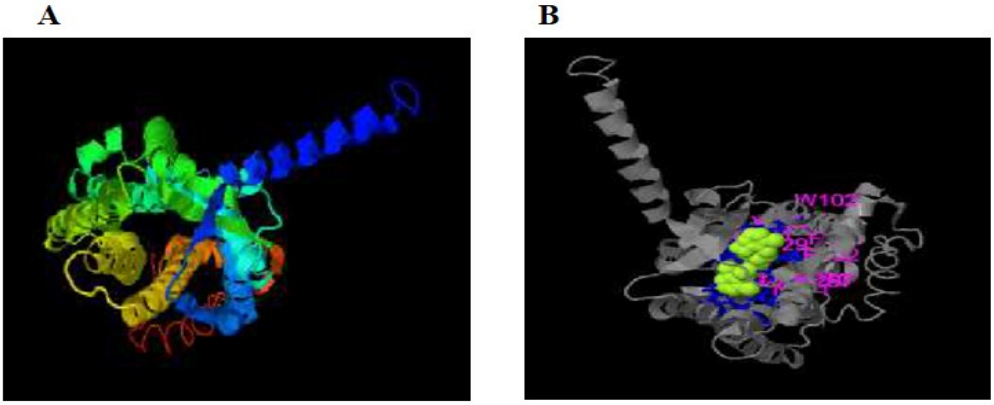

Figure 2:- A- represents the PDB image of the highest quality modelled protein. B- Represent the Ligand Binding Site image of the isoform NP_113597.2:p.Ala369Glu which shows highest binding site.

Table 1: Ligand binding residues of modelled protein

C-Score	PDB Hits	Ligand Binding Residues
0.15	4ea3A	102,122,125,126,129,130,198,218,222,267,271,274,302,306
0.07	5cxvA	71,78,83,86,90,127,172,176
0.04	3oduB	102,105,111,121,125,194,197,198,302
0.03	1gzmA	49,52,269,300,301,304
0.03	5dhgA	102,105,111,122,125,126,129,197,218,222,270,271,274,306

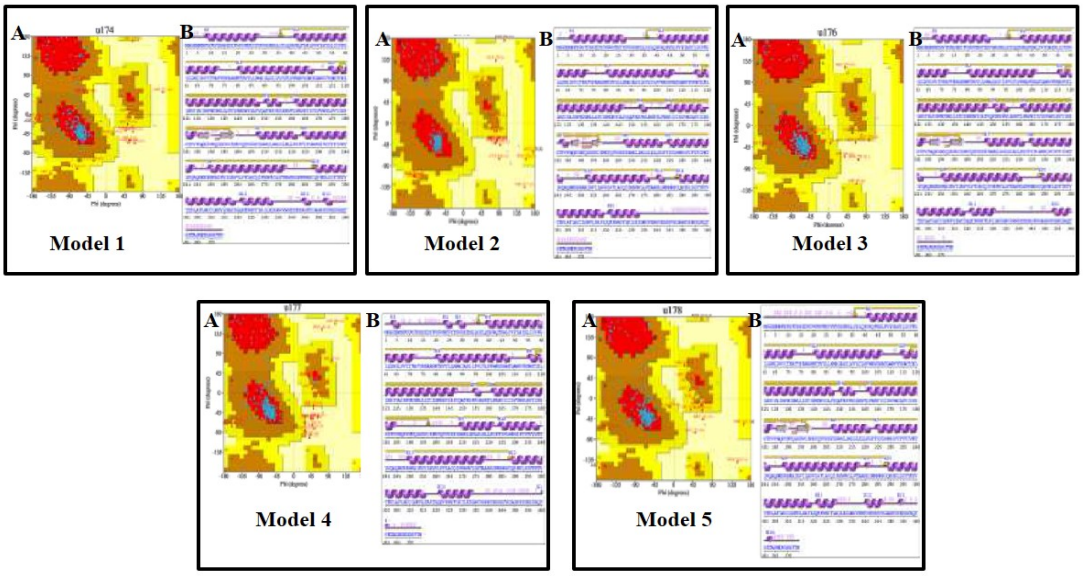

Figure 3: A-represents the Ramachandran Plot images of all the protein models of the Isoform NP_113597.2: p.Ala369 Gly of the CCR6 gene, respectively. B-represents the secondary residue images of all the 5 protein models of the Isoform NP_113597.2: p.Ala369 Gly of the CCR6 gene

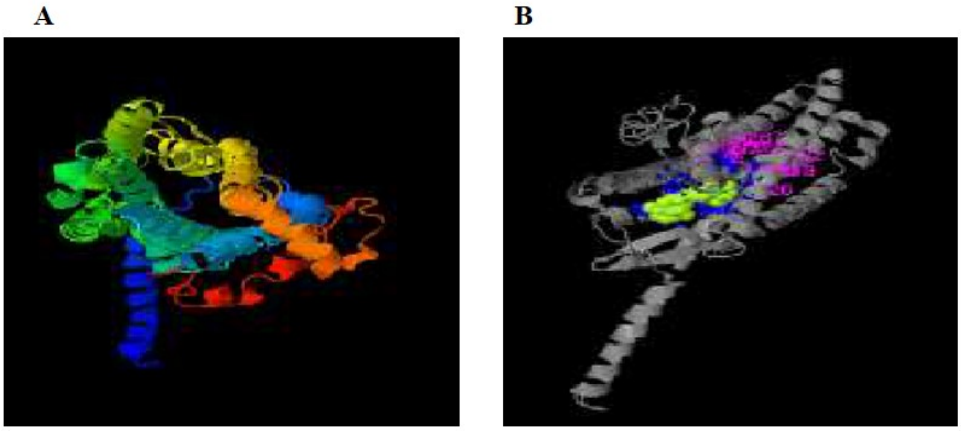

Figure 4: A-Represents the PDB image of the highest quality modelled protein. B- Represents the Ligand Binding Site image of the isoform NP_113597.2: p.Ala369 Gly which shows highest binding site.

Table 2: Ligand binding residues of modelled protein

C- Score	PDB Hits	Ligand Binding Site Residues
0.11	5DHHB	102,125,126,129,218,222,267,270,271,302,306
0.05	4MBSA	53,102,105,125,126,129,201,214,215,218,267,270,274,278,298,302,303
0.04	3ODUB	102,105,111,121,125,194,196,197,198,302
0.04	2Y01A	72,83,90,124,127,172
0.03	5CXVA	71,78,83,86,90,127,172,176

Protein model analysis of the isoform NP_113597.2: p.Ala369Gly (A (Ala) > G (Gly) [isoform 2] of the gene, C-C chemokine receptor type 6 (CCR6)

Five protein models of the isoform 2 of the gene, CCR6, were obtained through the I-TASSER software, for which quality check was carried out using PDB sum database and RAMPAGE software. Quality of protein (Figure 3) found for Model 1 was 83.6%, Model 2 was 87.4%, Model 3 was 86.6%, Model 4 was 80.0%, Model 5 was 82.5%. Model 2 with 87.4% is the highest and can be considered as the best of all models to be used for docking. And the C-Score of model is -1.64. PDB structure and ligand binding site of the best model was represented in Figure 4. Table 2 represent the ligand binding residues.

The purpose of this study was to model protein for the genes involved in Rheumatoid Arthritis. Through the course of the entire study we came across several genes that are related to Rheumatoid Arthritis in some or the other way. The list of the genes collected by literature has already been mentioned in the previous sections[7]. The autoimmune conditions we are focussing on are triggered due to various genes. These genes tend to perform various functions which directly or indirectly lead to manifestation of the symptoms and hence lead to a disease. AFF3 belongs to AF4/FMR2 family which encodes for a nuclear transcriptional activator that is generally expressed in lymphoid tissue. The transcription activator may function in lymphoid tissue development and oncogenesis. In vitro it is observed to bind to double-stranded DNA. Through studies it has been

speculated that variation in AFF3 may lead to an increased inflammatory response by lymphocytes, resulting more proinflammatory molecules in circulation, leading to a reduced response to TNF antagonists[8]. Another gene studied in relation to Rheumatoid Arthritis is CCL21 is also referred to as Ckb9, ECL, SCYA21 etc. The protein encoded by this gene inhibits hemopoiesis and stimulates chemotaxis. Under *in vitro* conditions it acts as a chemotactic protein for thymocytes and activated T cells. It acts as a ligand for chemokine receptor 7. When activated it binds to ACKR and mediates the recruitment of beta-arrestin (ARR1/2) to ACKR4.CCR6 CKRL3, CMKBR6, GPR29, STRL22 is used synonymously for CCR6 gene. The gene is located on the q armof chromosome 6.C-C chemokine receptor type 6 acts as the ligand for CCL20 also referred to as MIP-3α. It is also a member of chemokine family that attract leukocytes to the rheumatoid arthritis (RA) joint and also mediate angiogenesis. Through studies elevated amounts of MIP-3α has been observed in synovial fluid in patients with Rheumatoid arthritis as compared to patients with osteoarthritis. Activation of CCR6 transduces a signal by increasing levels of Calcium ions intracellularly. Although CCL20 acts as the major type of ligand but it is also observed to act as receptor for non-chemokine ligand like DEFB1 which in turn also affects calcium ion as well as cAMP levels. The ligand receptor pair is responsible for chemotaxis dendritic cells (DC), effector/ memory T-cells and B-cells and plays an important role at skin and mucosal surfaces under homeostatic and inflammatory conditions, as well as in pathology, including cancer and various autoimmune diseases[9]. The receptor is essential for recruitment of both IL-17 producing Th-17 and T-Reg cells. Also essential for B- cell localisation as well as class switching to IgA. While doing literature survey we came across that CCR6 is one of those genes whose encoded protein structure is still not discovered[10]. Therefore we used it as our gene of interest for protein modelling. Positive stimulation of the gene leads to sperm motility and chemotaxis. The HLA group of genes binds to peptides presented by antigen presenting cells.HLA-DPB1, HLA-DRB1, HLA-B are the protein coding genes involved in rheumatoid arthritis. Although it is still unknown as how HLA genes contribute to the condition susceptibility but they have been found to be in close association with the specific histocompatibility markers like HLA-DR4. According to some studies only two of the subtypes of HLA-DR4 have been found to be prevalent in patients with Rheumatoid Arthritis which are Dw4 and Dw14[6]. But there are subtypes which are not at all associated with the condition. Association studies between HLA-DR4 and RA is one the most thoroughly researched cases of genetic susceptibility to an autoimmune condition. There are studies done which have suggested that PTPN22 are expressed in hematopoietic tissues primarily. Thymus, spleen, PBMCs, bone marrow is amongst those tissues which express PTPN22 extensively but there seems to some hierarchy. NK cells seem to express it more on the contrary B cells and CD4+ T cells express it the least. This observation raise a possibility between the gene and Rheumatoid Arthritis could lead to functional cell changes in other cell types. It acts as a negative regulator of TCR. The gene has been observed to regulate lot of mechanism including autophagy, antiviral responses etc. However the finding that the minor allele of the PTPN22 SNP is also related to T1D supports the hypothesis that there are common genetic variants that contribute to general immune dysregulation and susceptibility to autoimmunity. Another gene peptidyl arginine deiminase 4 abbreviated as PADI-4 is observed to catalyse the conversion of arginine residues to citrulline residues. This gene may play a role in granulocyte and macrophage development which could lead to inflammatory and immune responses. A single nucleotide polymorphism in RUNX1 is also associated with Rheumatoid Arthritis. Through studies it has been observed that expression of SLC22A4 (organic cation transporter gene) is specific to haematological and immunological tissues and the transporter gene is also highly expressed in the inflammatory joints of mice with collagen-induced arthritis[11]. This indicates that the regulation of SLC22A4 expression by RUNX1 is related to susceptibility to rheumatoid arthritis and may act as an example of an epistatic effect of two genes on this disorder. According to a study it has been hypothesised that the immune-related genes complement component 5 (C5) and/or TNF receptor-associated factor 1 (TRAF1) would show some correlation with Rheumatoid Arthritis. The protein encoded by TRAF1 is a member of the TNF receptor-associated factor (TRAF) protein family, which associates with and mediates the signal transduction from various receptors of the TNF receptor superfamily, including the receptor for TNFα and in the activation and proliferation of T cells. Therefore TRAF1 could play a role in Rheumatoid Arthritis by aiding the maintenance of the proinflammatory environment[12]. Also endurance of inflammation coincides with increased levels of the anaphylatoxin C5a in the synovial fluid of Rheumatoid Arthritis patients has been observed. There are many other genes observed to trigger Rheumatoid Arthritis. But there are various non genetic factors which are claimed to affect onset of the disorder. Potential triggers include occupational exposure to certain kinds of dust or fibres, viral or bacterial infections and changes in sex hormones in females. Rheumatoid arthritis is a very common autoimmune illness that affects around 1% of people globally. It is caused by an abnormal immune reaction thereby affecting joints and causing an inflammation. Currently treatments aim to give symptomatic relief with the use of simple analgesics, or anti-inflammatory drugs. In addition, most patients are also treated with what are known as disease-modifying agents, which aim to prevent joint damage. The onset of the illness is observed to be due to a lot of genetic factors, therefore greater study in this direction could help managing the disease and devising a better treatment for the same[1].

CONCLUSION

Protein modelling of the genes affecting Rheumatoid Arthritis was performed. This review basically focusses on the genetic factors affecting the autoimmune disease. Through the literature study it was found that CCR6 gene encoding protein doesn't have a structure. Therefore modelling of the same was done. Throughout the work various types of genes and their underlying functions were studied which helped us understand the mechanism of the disease and further research could also help to devise new treatments for curing Rheumatoid Arthritis. In future we would like to use these results for docking which could help find a potential compound that could inhibit the triggering of the factors that could lead to Rheumatoid Arthritis.

REFERENCES

1. Hyndman D, Liu S, Miner JN. Urate Handling in the Human Body. Current Rheumatology Reports 2016; 18(6): 34.
2. Gao W, McGarry T, Orr C, *et al*. Tofacitinib regulates synovial inflammation in psoriatic arthritis, inhibiting STAT activation and induction of negative feedback inhibitors. Annals of the Rheumatic Diseases 2016; 75(1): 311–315.
3. Zhao B, Lehr R, Smallwood AM, *et al*. Crystal structure of the kinase domain of serum and glucocorticoid-regulated kinase 1 in complex with AMP–PNP. Protein Science: A Publication of the Protein Society 2007; 16(12): 2761–2769.

4. Tao J-H, Zhang Y, Li X-P. P2X7R: a potential key regulator of acute gouty arthritis. Seminars in Arthritis and Rheumatism 2013; 43(3): 376–380.
5. Abdelwahab NS, Hassan HM, Magd AMA. Rapid microwave-assisted hydrolytic degradation of colchicine: *In silico* ADME/Tox profile, molecular docking, and development of innovative RP-Chromatographic methods. Micro chemical Journal 2019; 104 (4): 1-9.
6. Yar M, Bajda M, Shahzadi L, *et al*. Novel synthesis of dihydropyrimidines for α-glucosidase inhibition to treat type 2 diabetes: *In vitro* biological evaluation and *in silico* docking. Bioorganic Chemistry 2014; 54: 96–104.
7. Bjordal JM, Ljunggren AE, Klovning A, *et al*. Non-steroidal anti-inflammatory drugs, including cyclo-oxygenase-2 inhibitors, in osteoarthritic knee pain: meta-analysis of randomised placebo controlled trials. BMJ (Clinical research ed) 2004; 329(7478): 1317.
8. Xiong H, Xu Y, Tan G, *et al*. Glycyrrhizin Ameliorates Imiquimod-Induced Psoriasis-like Skin Lesions in BALB/c Mice and Inhibits TNF-a-Induced ICAM-1 Expression via NF-κB/MAPK in HaCaT Cells. Cellular Physiology and Biochemistry 2015; 35(4): 1335–1346.
9. Wu B, Song H-P, Zhou X, *et al*. screening of minor bioactive compounds from herbal medicines by *in silico* docking and the trace peak exposure methods. Journal of Chromatography A 2016; 143(6): 91–99.
10. Singh H, Samani D, Nain S, *et al*. Interleukin-10 polymorphisms and susceptibility to ARV associated hepatotoxicity. Microbial Pathogenesis 2019; 133(10): 35-44.
11. Bridges CC, Krasnikov BF, Joshee L, *et al*. New insights into the metabolism of organo mercury compounds: Mercury-containing cysteine S-conjugates are substrates of human glutamine transaminase K and potent in activators of cystathionine γ-lyase. Archives of Biochemistry and Biophysics 2012; 517(1): 20–29.
12. Morelli M, Scarponi C, Mercurio L, *et al*. Selective Immunomodulation of Inflammatory Pathways in Keratinocytes by the Janus Kinase (JAK) Inhibitor Tofacitinib: Implications for the Employment of JAK-Targeting Drugs in Psoriasis. Journal of Immunology Research 2018; 5: 1-18.

Cite this article as:

Jerine Peter S *et al*. Protein modelling of the genes of Arthritis: A Review. Int. Res. J. Pharm. 2020;11(4):13-18 http://dx.doi.org/10.7897/2230-8407.110435

Source of support: Nil, Conflict of interest: None Declared