Network Biology of Inflammatory Bowel Diseases

By

Tushar Dhyani

Under the Supervision of -

Dr. Ganesh Bagler

Dr. Arjun Ray

Indraprastha Institute of Information Technology Delhi

August, 2021

Network Biology of Inflammatory Bowel Diseases

by

Tushar Dhyani

Submitted

in partial fulfillment of the requirements for the degree of

Master of Technology

to

Indraprastha Institute of Information Technology Delhi

August, 2021

Certificate

This is to certify that the thesis titled “Network Biology of Inflammatory Bowel Diseases” being submitted by Tushar Dhyani to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

August, 2021

Dr. Ganesh Bagler
Department of Computational biology
Indraprastha Institute of Information Technology Delhi
New Delhi 110 020

Dr. Arjun Ray
Department of Computational biology
Indraprastha Institute of Information Technology Delhi
New Delhi 110 020

Acknowledgements

I would first like to thank my thesis Supervisions **Dr. Ganesh Bagler** and **Dr. Arjun Ray** of the computational biology department at **Indraprastha Institute of Information Technology Delhi.** Your wise feedback pushed me to sharpen my thinking and pointed me in the right direction. I want to thank you for your patient and for all of the opportunities I was given.

I would also like to acknowledge **Dr. Sriram K** of the computational biology department at **Indraprastha Institute of Information Technology Delhi** for his wise counsel and sympathetic ear.

I am thankful to my college seniors Gayatri Panda, Sukriti Sacher and Shruti Kaushal without their passionate participation and feedbacks, the molecular modeling and simulations could not have been successfully conducted.

I am grateful to Mrs. Kusum Dhyani, my mother, for listening to my constant babbling about my work and giving continuous encouragement. I am thankful to Mrs. Kanika Dhyani and Miss. Kajol Dhyani, my sisters, for giving me support so that I may pursue my passion i.e. research.

Lastly I want to thank Late Mr. Tapeshwar Prasad Dhyani, my father, for everything.

Author

Tushar Dhyani

MT19221

Abstracts

Inflammatory bowel disease (IBD) is a general term for chronic conditions caused by inflammation of the gastrointestinal tract. IBD is classified mainly into two forms namely; Crohn's disease (CD) and Ulcerative colitis (UC). In 2017 globally total IBD cases count crossed 6.8 million. The mechanism of pathogenesis of IBD is interplay between genetic factors, environmental factors, immune response and intestinal barrier conditions. Previous studies based on isolated therapies show very little effectiveness but applying a system view perspective shows promising results. In this study we tried to find the pathways which are most affected because of mutations in genome of IBD patients. For this mutations identified in IBD patients by GWAS studies were mapped to numerous genes by D Muraro et. al. We used this gene list as candidate gene list to understand how risk polymorphisms in these associated genes play role in IBD disease development. From candidate gene list 150 top ranked genes prioritized by ENDEAVOUR tools were shortlisted. 666 genes found to be closely related to these 150 genes either in the form of complex formation or common pathways were also identified using GeneMANIA. This helped to identify the pathways in which these genes can be functionally classified. IL-12 and IL23 pathways were the top pathways in which these 816 nodes were functionally classified. From literature survey it was found that IL-23 inhibition can lower the inflammation mediated by TH-17 cells. TH-17 cells play a crucial role in inflammation caused in IBD. In the second half of the study, inhibitors for IL-23 signaling axes were found by identifying plant derived molecules found in Indian medicinal plants. The basic idea is to mask the binding site of the IL-23(cytokine) and IL-23R (receptor) and not allowing IL-23 signaling. For this targeted molecular docking experiments were done focusing on the IL-23:IL23R interacting residues. Terflavin B, Punicalin and Punicalagin were the best docked molecule to IL-23 and IL-23R both in separate docking experiments. Molecular dynamic simulation showed that these three were forming stable bond with IL-23 throughout the 10ns simulation. Only Terflavin B and Punicalagin were forming stable bond with IL-23R at the end of 10ns simulation. It is known that Terflavin B showed antibacterial activity against *Bacillus subtilis* and *Pseudomonas fluorescen.* Punicalin and Punicalagin shows strong antioxidant and anti-inflammatory properties. In this study Terflavin B has better binding stability with both IL-23 and IL-23R as compared to all the analyzed ligands. The results from this study suggest that these plant derived molecules can be further explored as potential inhibitor of IL-23 pro-inflammatory signaling axis.

List of figures

List of Table

Contents

Chapter 1

Introduction

Inflammatory bowel disease (IBD) is a general term for chronic conditions caused by inflammation of the gastrointestinal tract. IBD is classified mainly into two forms namely; Crohn's disease (CD) and Ulcerative colitis (UC). CD can cause inflammation in many part of digestive tract whereas UC is mainly limited to colon[1]. There had been a drastic increase in cases of IBD worldwide. In 2017 globally total IBD cases count crossed 6.8 million. IBD cases are consistently greater in developed countries like Australia, Canada, UK and USA. There are many studies linking urbanization with this trend[2]. IBD cases are also rapidly increasing in developing countries in Asia, Eastern Europe, and Africa [3].

The mechanism of pathogenesis of IBD is interplay between genetic factors, environmental factors, immune response and intestinal barrier conditions. A genetic susceptible person when exposed to adverse environmental conditions may develop microbial dysbiosis in the gut; this further through dysfunctional intestinal barrier leads to immune inflammation leading to clinical symptoms. Previous studies based on isolated therapies show very little effectiveness but applying a system view perspective shows promising results[4][5].

Reportedly the risk of IBD is increased by high consumption of processed meat and saturated fatty acid. High-fiber diet is reported to reduce CD risk by 40% [6]. Exposure to pets, breastfeeding, antibiotics use and infection during early stages of life is associated with risk for IBD development, as they may modulate the intestinal microbiota composition. Microbiome plays an important role in intestinal immune cell development. In IBD patient microbiome diversity is decreased, there is increase in inflammation causing bacteria and decrease in bacteria showing anti-inflammatory capacities. Increase in *Escherichia coli, defsulfovibrio and mycolytic* bacteria are linked with different complications in the intestinal barrier, adding up to inflammation in IBD patient's gut[7]. The number of *F praunitzii* and *faecalibacterium prausnitzii* has been observed to be lower in IBD patients [8].

Normally intestinal mucosal layer exists is in equilibrium with lumen but this equilibrium is disturbed in IBD patients. Junctional protein i.e. epithelial cadherin's down regulation is observed in CD patients. Goblet cells are essential for mucosal defense and repair in the intestine. Deletion of MUCIN 2, a secretory mucin derived from goblet cells causes development of colitis in murine models. Paneth, another type of cells found in small intestinal crypts, are important as they maintain intestinal stem cell niche, crypt homeostasis and are responsible to maintain equilibrium between mucosa and microbiota. NOD2 gene is also expressed in the paneth cell, its mutated variants here leads to lower levels of alpha-defensins causing defective antimicrobial function. Patients with IBD also show defective ATG16L1 gene in paneth cells, compromising their autophage function[9][10].

In a healthy individual's gut, macrophages apart from their normal function also produce anti-inflammatory cytokines which restrain TH1 and TH17 cell response and promote regulatory T-cell differentiation. In CD patients these macrophages are attenuated and facilitate microbe passage through mucosa and because of this microbial leakage inflammation causing neutrophil are recruited [11]. IBD patients also have a sub-population of inflammatory macrophages, which produces large amounts of pro-inflammatory cytokines and also express CD14 a dendritic cell marker[12][13]. After the intestinal barrier is compromised the adaptive immune response starts working in a self-sustaining amplifying manner. In CD this self-sustained amplification is caused as the APCs and macrophages produces cytokines which differentiate T naïve into TH17 and TH1 cells, these cells produces pro-inflammatory cytokines which in turn stimulates APCs and macrophages giving positive feedback to inflammation [14][15]. Leukocytes migrates to inflamed intestinal sites. This migration occurs by binding of integrin molecules (present on leukocyte surface) with adhesion molecules (present on endothelial cells). T-cell immune response is also exaggerated in IBD patients. Excessive Th1 and Th2 cell response causes inflammation in CD patients whereas Th2 cell-type like cytokine response causes inflammation seen in UC[16].

Another difference between CD and UC is that UC patients have IL-13 producing CD1d-restricted nonclassic natural killer T cells whereas in CD patients, APC and macrophages produce IL-12, IL-18, TGF-beta and IL-23. Signaled by cytokines Th1 and Th17 cells secrete IL-17, TNF-alpha and interferon gamma which are pro-inflammatory cytokines. These cytokines stimulate APCs, endothelial cells, macrophages and fibroblast to produce IL-1, IL-6, IL-12, IL-8, IL-18 and TNF alpha completing the self-sustaining amplification cycle of inflammation[17]. Normally TGF-beta promotes Treg differentiation in naïve lamina propria CD4+ T cells. Treg contributes to normal homeostasis. Lower number of Treg has a critical role in pathogenesis of IBD. In IBD, TGF-beta because of inflammation caused by cytokines and mircrobial promotes Th17- cell differentiation, which produces pro-inflammatory cytokines. In total increase in cytokines is a major reason of inflammation in IBD. Signaling pathways associated with cytokines are targeted to block their effect [18][19].

IBD patient's first-degree relatives have 5-fold higher risk for developing the disease. Which suggest that IBD have a genetic cause[20][21]. There are certain rare gene mutations identified in X-linked inhibitors of apoptosis (XIAP) and interleukin 10 receptor gene region related to early onset of IBD. The cases caused by these mutations represent only 10-20% of all IBD cases, suggesting that IBD is moreover a polygenic disease[22]. Genome-wide association studies (GWAS) are used to obtain insights into polygenic diseases like IBD. Many loci associated with IBD are identified. Most of these loci are common with other immunodeficiency diseases and autoimmune disease like multiple sclerosis, psoriasis and rheumatoid arthritis[23]. These identified genetic loci helps to understand biological pathways disturbed during IBD pathogenesis [24]. For example multiple loci mutation at IL-23 receptor loci underlines the role on IL-23 signaling in CD. A large number of GWAS identified variations are found in non-coding regions, exerting the importance of modulated gene expression [25]. Polymorphisms in autophagy-related 16-like 1 gene and NOD2 mutations suggest the role of autophagy in IBD pathogenesis.

354 genes associated with Crohn's disease annotated by Jostins et al. are used in our analysis[26]. To further subset the proteins which were most important for our analysis we ranked the gene list and from this ranked gene list the top 150 prioritized candidate genes were selected. These 150 genes were used to build a PPI network along with their closely related genes. Next to elucidate the biological pathways which may be greatly affected by these genes were identified. IL-23 and IL-12 pathways were the top consolidated pathways identified. From further literature survey the pro-inflammatory role of IL-23 was confirmed [27]–[29]. Indian medicinal plants that are already well known in Ayurveda for their

medicinal properties to treat IBD like gut related diseases were identified from the literature [30]. *Terminalia chebula , Cuminum cyminum,* and *Syzygium cumini* were among best Indian medicinal plants used for treating gut related problems [31]–[33]. Structure of Compounds found in these plants were downloaded from IMPPAT database. In the next phase of our this study we tried to identify compounds which can mask the binding site of IL-23:IL-23R interaction so that IL-23 signaling can be blocked. We used molecular docking to identify the compounds which were able to attach with the critical residues of IL-23 and IL-23R in separate experiments. In this way we tried to mask interacting site of both IL-23 (cytokine) and IL-23R (receptor) to completely hinder their binding. This mitigates the pro-inflammatory effect of IL-23 signaling and disrupts the self-sustained inflammation amplification caused as the APCs and macrophages produces cytokines which differentiate T naïve into TH17 and TH1 cells.

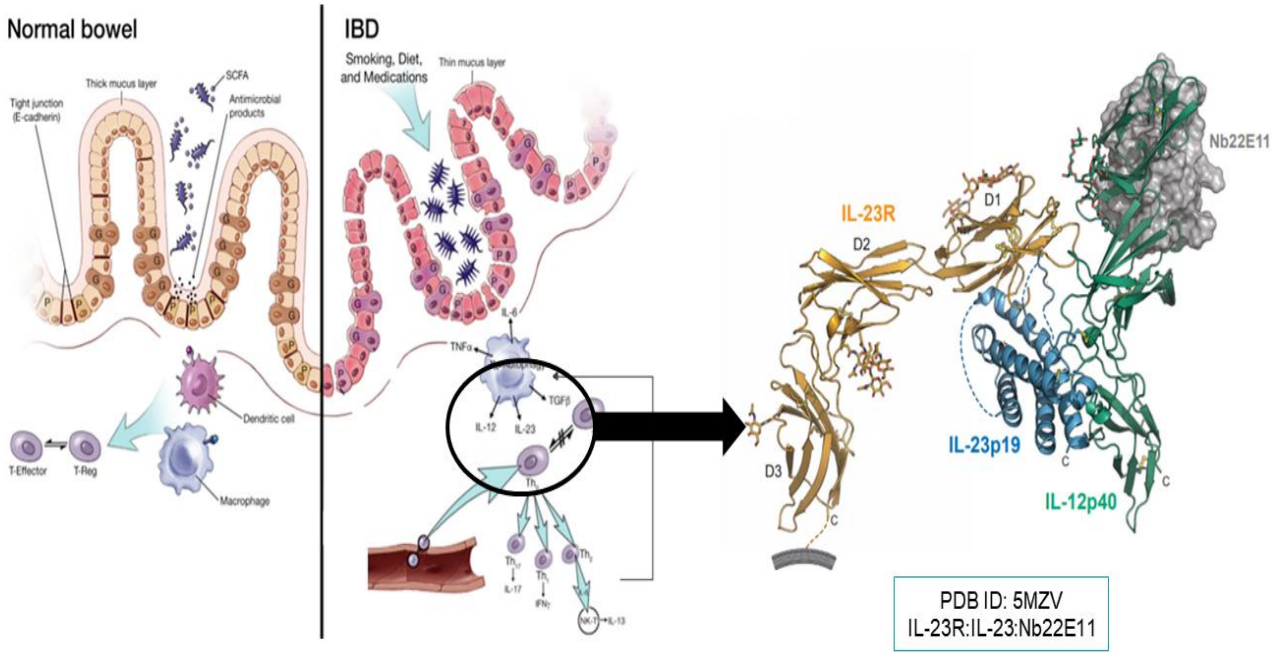


Figure 1: The difference between normal person's and IBD patient's gut lining[27][22]. IBD patient's gut lining is ruptured leading to infiltration of pro-inflammatory cells and cytokine. As depicted in The figure, In CD self-sustained inflammatory amplification is caused as the APCs and macrophages produces cytokines which differentiate T naïve into TH17 and TH1 cells, these cells produces pro-inflammatory cytokines which in turn stimulates APCs and macrophages giving positive feedback to inflammation [14][15]. IL-23 cytokine is depicted on left while binding to its receptor (IL-23R)[23].

Chapter 2

Method

Prioritisation

GWAS identifies many variations that may contribute to complex diseases. 354 genes associated with Crohn's disease annotated by Jostins et al. [24] were downloaded and prioritized using ENDEAVOUR [34][35]. The ENDEAVOUR web based tool requires two inputs one in the form of training dataset and second as the candidate genes list we are aiming prioritizing. The training dataset genes are the ones whose role in the disease under study is already known [26]. ENDEAVOUR provides a large set data sources that can be used to build model using data fusion. This concept of data fusion is more powerful than using any data-source alone. We used the whole array of data sources provided in the tool. Next we uploaded the candidate set and launched the tool. The basic aim here was to rank the candidate genes according to their closeness with the profile derived from training set genes whose role in IBD is already well known.

Microarray study

The microarray dataset with accession number GSE20881 publically available at gene expression omnibus was analyzed. The data contains 172 colon biopsies from healthy controls (73) and Crohn's patients (99). The web tool GEO2R was used to identify the differentially expressed genes in this data set. The Benjamini and Hochberg false discovery rate method was used to adjust the p-values. The genes whose p-value was less than 0.01 were selected as differentially expressed.

Functional classification

We used PANTHER classification system to classify the top 150 prioritized candidate list [36]. This enrichment is done using Statistical over-representation test to find the genes which are observed more than expected based on reference list. Homo sapiens whole genome is selected as reference list. Binomial test is applied to compare the reference list and candidate list genes occurring by chance. Bonferroni correction is selected for False Discovery rate calculations.

Network construction and analysis

We used GeneMANIA to identify 666 genes related to the 150 top prioritized candidate genes. Subsequently various networks of these 816 genes were analyzed using GeneMANIA [37]. Co-expression network identifies the connection between two genes if their levels of expression are similar in gene expression studies. Physical interaction networks are made using the information regarding the protein interaction, edge symbolizes that the gene products are found to interact in studies. Shared protein domain network have gene lined if the gene products have similar protein domain. Gene interaction network have genes connected if it was found that effect of perturbing a gene can be modulated by perturbations to an another gene. Co-localization network have genes linked if they both are expressed

in same tissue location. Pathway network generated using GeneMANIA have nodes linked if the gene products participate in reaction within a same pathway. Predicted network are generated using predicted functional relationship between genes or their product found in other organism. All the above defined network were analyzed using NetworkAnalyzer tool in cytoscape [38].

To analyze how node removal affects the network, open source software NEXCADE was used [39].This monitors how the numbers of interactions vary as nodes are removed randomly, from lowest to highest degree and from highest to lowest degree. The primary extinct are the nodes which are remove in this process and secondary extinct are the nodes which gets totally disconnected from all other nodes because of the primary extinct. We also monitored how the numbers of secondary extinct vary as the nodes are removed from higher to lowest degree, randomly or lowest to higher degree.

Identification of natural molecules against IL-23 and IL-23R by molecular docking and molecular dynamics simulation

In Ayurveda, class of fermented preparations called Asava and Arishta are used for treating digestive disorders. From ancient times the technique used for their preparation consist of extracting water soluble components and then fermenting them[30]. The fermentation medically activates the compounds and help in better absorption by human body [31]. Different medicinal plant's parts are used for these fermented preparations like pericarp of *Terminalia chebula* (rich in hydrolysable tannins) and cumin seeds (rich in Flavonoid glycosides and aglycones). There are many modern day studies supporting the ancient tradition. In modern studies many of the compounds extracted and fermented in Ayurvedic formulation are found to having antioxidant [40]–[42] and anti-inflammatory properties [33]. Here we tried to identify potential therapeutic targets from the plant derived molecules extracted from the medicinal plants used in Asava and Arishta fermented preparations.

Ligand library preparation

For library preparation the compounds present in *Terminalia chebula , Cuminum cyminum,* and *Syzygium cumini* were identified. Zinc database[43] and IMPPAT (Indian Medicinal Plants, Phytochemistry And Therapeutics) a database constructed via manual Curation of knowledge generated for Indian medicinal plants were used to extract the 3D structure of compounds found in plants under study[44]. The final dataset generated contain 127 molecules. Hydrogen were explicitly added to their structures and all 127 molecules were added to a primary MOL2 file using OpenBabel 2.3.1 software [45].

For virtual screening python scripts and UNIX shell commands provided by Autodock suite were used[46]. Different directories for ligand, receptor and docking were used for better data management. We populated the ligands from the primary MOL2 file to the ligand directory. The individual MOL2 files were then converted PDBQT format using the python script provided in AutoDock.

Receptor Preparation

By studying the assembly of the receptor cytokine complex it was shown that the binding of the IL-23 and IL-23R to form a binary complex is a crucial step for the recruitment of the other half the shared receptor i.e. IL-12Rβ1. The IL-12Rβ1 has no measurable affinity to IL-23R and also its affinity to IL-23 was poor, so its recruitment majorly depends on the IL-23:IL:23R binary complex formation. To damper this complex formation the interaction hotspot of IL-23:IL23R binary complex was interrogated[29].

IL-23 is a member of IL-12 family of cytokines. IL-23 is a heterodimer having IL-23p19 and IL-12p40 subunits[23]. IL-23 signaling is mediated by IL-23 receptor (IL-23R) and β1 subunit of IL-12 receptor. IL-23 signaling is found to be essential for inflammation mediated by TH-17 cells. In absence of IL-23 signaling development of TH-17 was obstructed which further decreased IL-17 production and other pro – inflammatory events[47]. For docking study, protein coordinates for unbound human IL-23 (PDB ID: 5mxa) was used. First all unwanted structures from PDB file were manually removed. After this polar hydrogens were added and each atom was assigned gasteiger partial charge using AutoDockTools. The protein was then converted to PDBQT format file and saved in receptor directory.

A second docking experiment targeting IL-23R was also done simultaneously with docking study for IL-23 cytokine. IL-23R has three extracellular domains. By studying the IL-23:IL23R binary complex (PDB ID: 5mzv) it was elucidated that in IL-23R only N-terminal Ig domain was interacting with IL-23 forming the cytokine-binding domain. IL-23R structure from PDB ID: 5mzv was extracted. IL-23R structure was further refined manually by removing unwanted structures like Nano bodies, adding polar hydrogens and assigning gasteiger partial charge using AutoDockTools. This structure was further converted to PDBQT file using the script provided in AutoDock.

GridBox Preparation

We want to identify molecules which can hinder the IL-23 and IL-23R interaction. To find the interaction interface between IL-23 and IL23R, PDB ID: 5mzv was observed. 5mzv contains crystallographic coordinates for human IL-23:IL-23R:Nb22E11 complex. The molecular interaction between the cytokine and the receptor was analyzed using Ligplot+ v.1.4.3 software [48]. The IL23p19 (B) and IL-23R (C) were forming hydrogen bond between residues- Lys164(B):Gly24, Glu58(B):Asn29, Arg57(B):Leu113, Arg57(B):Glu111, Arg57(B):Thr112, Trp156(B):Asp118.

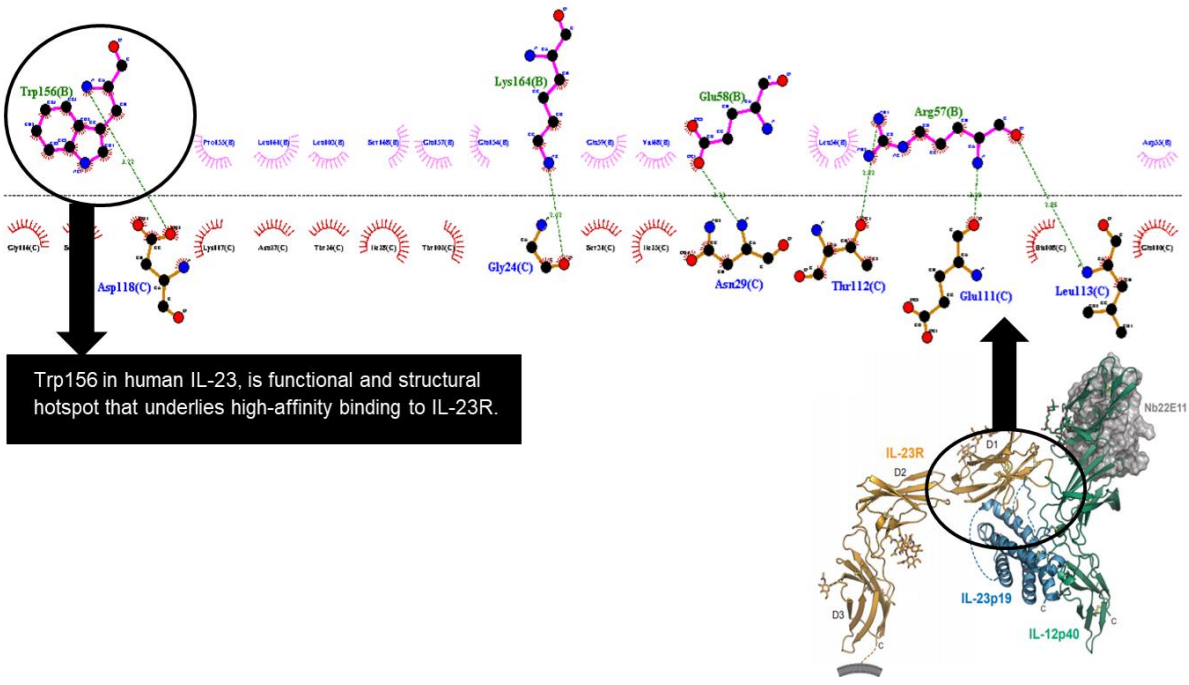


Figure 2: Plot generated by Ligplot+ showing IL-23p19 (here B) subunit and IL-23R (here C) interacting interface for PDB ID: 5mzv.

First GridBox for IL-23 was made. From literature survey it was also pointed that Trp156 in IL-23 is the functional and structural hotspot. This Trp156 is responsible for binding of IL-23 to IL-23R with high affinity and important for signaling purpose [29]. So for molecular Docking studies of IL-23, gridbox covering Trp156 and previously identified IL-23:IL23R interface was generated manually using AutoGrid4 and map types were assigned directly. Similarly for IL-23R, residues important for its interaction with IL-23 were covered during GridBox preparation.

Molecular docking

Autodock4 was used for docking simulation for both IL-23 cytokine and IL-23R receptor. For each ligand 20 feasible binding conformations were obtained. At the end of simulation the ranked docked conformation were sorted according to their binding affinities which is sum total of total internal, intermolecular and torsional free energy minus unbounded system's energy. Interactions between the ligand and the receptor were analyzed using PLIP web tool [49]. Further the docked conformations were visualized using PyMOL software (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

Analysis of structural analogs

ChEMBL is a database of bioactive molecules having drug-like properties. ChEMBL was used to find molecules having more than 95% substructure similarity with the top docked molecule found during library screening (i.e. Terflavin B). We found 8 bioactive molecules of which 4 were unnamed. Other 4 named bioactive molecules were Terflavin A (CHEMBL504094), Punicalin (CHEMBL502440), Punicalagin (CHEMBL506814) and Isoterchebulin (CHEMBL508570). Molecular docking for these molecules was done with the same receptor and conditions as previously used.

Molecular dynamic (MD) simulations

The best molecular leads obtained by Molecular docking were subjected to MD simulations using GROMACS package [50]. Topology files for ligand and the proteins were generated using CHARMM 36 force field. TIP3P water model was used for solvation. After this unit cell was defined and filled with water. To neutralize the charge of the system appropriate ions were added to the system. The system was then subjected to energy minimization using steepest descent minimization algorithm. A total of 50000 steps were performed with minimization step size equal to 0.01. After minimization was successfully completed position restraint were applied during equilibrating step. First NVT equilibration was done for 50000 steps (100 ps) with a time step of 2 fs, at 300 k temperature. A second equilibration under NPT conditions with 1 atm pressure was performed. After two successful equilibration phases, a production run for 10 ns with 2 fs time steps at 300K temperature and 1 bar pressure was simulated. After the completion of 10 ns MD simulation, nonbounded interaction energy between the ligand and the protein , Root mean square fluctuation (RMSF), temperature fluctuation, Radius of Gyration (Rg) ,pressure fluctuation , the Root mean square deviation (RMSD) and density fluctuation were analyzed. The obtained xvg files for these analyses were plotted using MATLAB 2020b.

Chapter 3

Results

Prioritisation using Endeavour

A genome-Wide association study identifies many candidate genes associated with Crohn's disease. We applied Endeavour which uses data fusion to priorities the candidate genes and identifies the most likely genes to be involved in the disease of interest. 354 genes associated with IBD annotated by Jostins et al. [24]were prioritized using Endeavour. The top 150 genes of this prioritized list were selected as a sub-list for the further steps (Workbook list-1).

Microarray study

The web tool GEO2R identifies the genes which are differentially expressed between the normal control and Crohn's patient's microarray data. The data set we analyzed (accession number GSE20881) is publically available at Gene Expression Omnibus. The analysis of this data identifies 4926 genes to be differentially expressed of the total 41616 genes. 41 of the top 150 prioritized genes were identified to be among these 4926 differentially expressed genes.

Functional classification

PANTHER was used to classify the candidate list [36] .The top biological processes in which the candidate genes are functionally classified are immune system processes, Cytokine receptor binding and other immune receptor activity according to molecular function. The Reactome pathways in which the candidates are functionally classified are cytokine signaling, interferon signaling and interleukin signaling. When this classification is done using PANTHER pathways as an annotation data source then the classification is as follows; Inflammation mediated by chemokine and cytokine signaling pathway, Apoptosis signaling pathway, Interleukin signaling pathway, Gonadotropin-releasing hormone receptor pathway ,Toll receptor signaling pathway (Workbook Panther Output).

Network analysis

The network found using GeneMANIA were analyzed using NetworkAnalyzer tool in cytoscape [38]. The density of physical interaction network was higher than NCBI proteome network. This higher density signifies that there is higher interaction between disease proteins than proteins which are unrelated to crohn's disease. The physical interaction network, genetic interaction network and pathway network have a downward sloping power law degree distribution [51]. The topological coefficient distribution of physical interaction network and genetic interaction network shows decreasing topological coefficient with increase in number of neighbors (figure 1,2). This indicates that the nodes with fewer links have more common neighbors than nodes (hubs) with high degree [52].

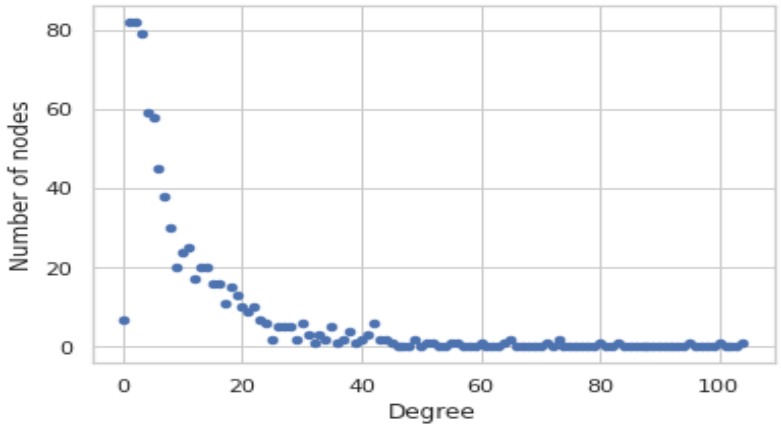


Figure 3: Degree distribution of physical interaction network (800 nodes)

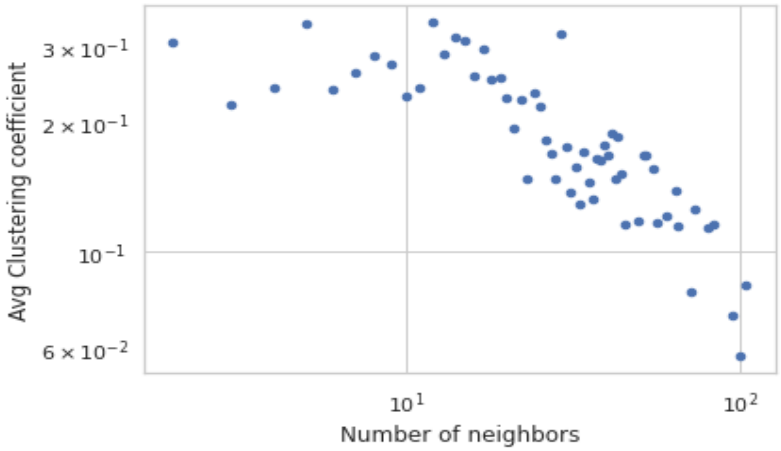


Figure 4: Average clustering coefficient distribution of physical interaction network (800 nodes)

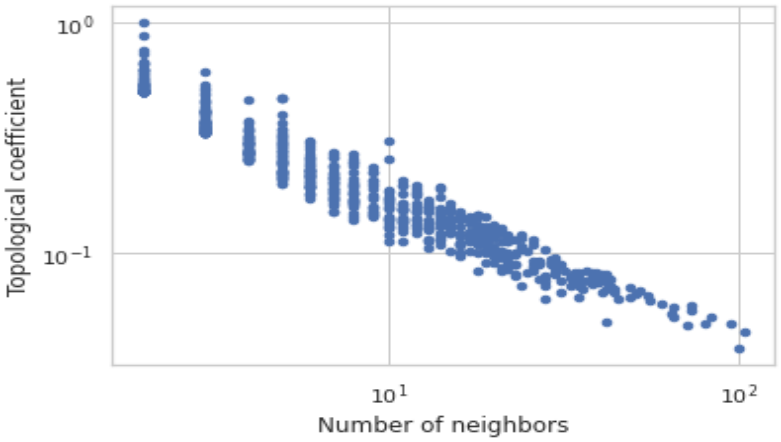


Figure 5: Topological coefficient distribution of physical interaction network (816 nodes)

To analyze the effect of random perturbation (failure) and selected perturbation (attack) on the network structure open source software NEXCADE was used [53]. The total number of interactions decreases more rapidly when high degree nodes were targeted first as compared to removal of lower degree nodes (figure 6.a). When nodes were randomly removed the number interactions in the network decrease linearly. Another analysis related to secondary extinct using NEXCADE shows that number of

secondary extinct increases more rapidly with the removal of high degree nodes as compared to random and lower degree nodes (figure 6.b).

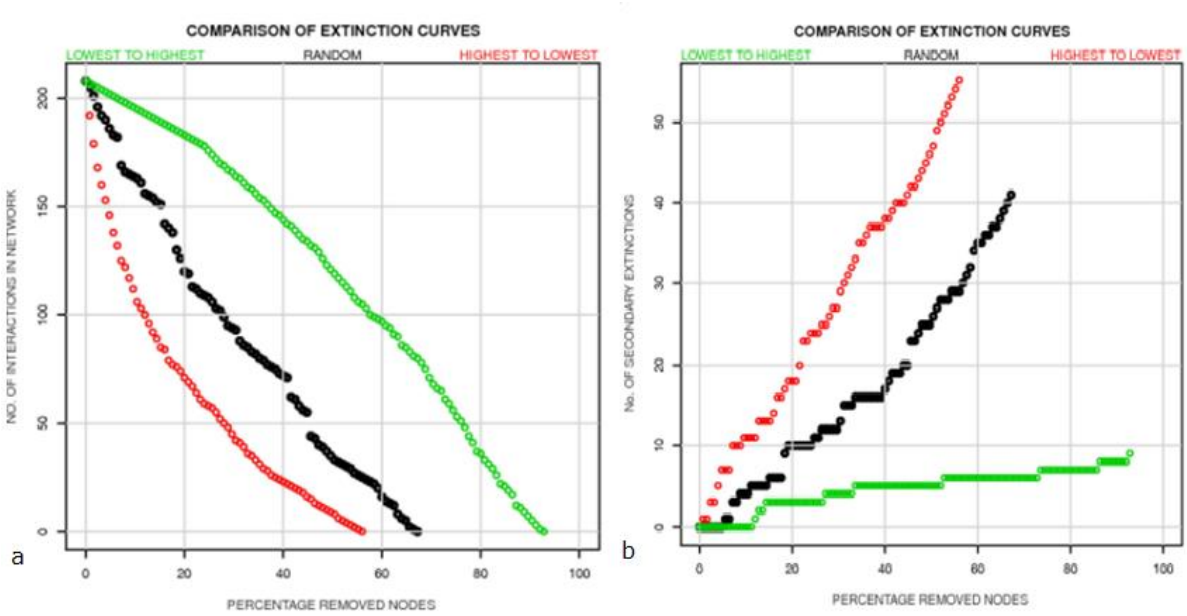


Figure 6: Change in (a) number of interactions and (b) number of secondary extinct after perturbations (Removal of nodes: randomly (black), lower to higher degree nodes (green) and higher to lower degree nodes (red)).

As the physical interaction network shows signatures of scale free network [52] we further focused on identifying the sub-graphs in physical network. We used Mfinder for over-represented sub-graph (Motif) search [53].Here motif of 3 and 4 nodes were searched and their over-representation was evaluated over only 100 random networks because of computational time required. When Mfinder tool was applied to physical interaction network of 100 prioritized candidate proteins with 50 closely related nodes identified by GeneMANIA, a three node motif (id 238) and 2 four node motifs (id 4958,13278) were identified (figure 4). Then motifs for the network having 316 candidate and 500 closely related proteins identified by GeneMANIA were calculated. Here we find found a three node motif (id 238) and 4 four node motifs (id 4958, 13260, 13278, 31710) (figure5).

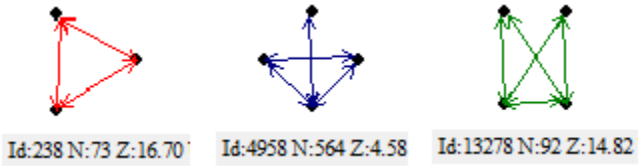


Figure 7: Identified motif of physical interaction network (150 nodes)

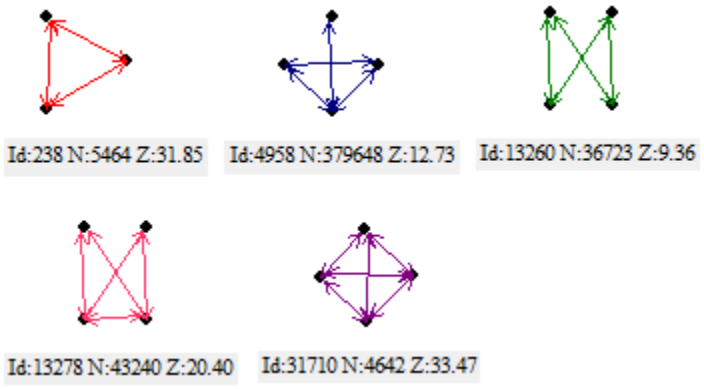


Figure 8: Identified motif of physical interaction network (816 nodes)

GeneMANIA identified the top 10 consolidated pathways in which the 816 node PPI network's proteins were classified (Table 1). These pathways are automatically curated from KEGG, Reactome database and pathway interaction database. This is also like functional classification but here all the 666 proteins related to the 150 prioritized proteins are also being classified adding up to the depth and precision in the classification. The top 2 attributes identified here for this 816 node network are IL-12 and IL-23 signaling events. IL-12 and IL-23 both come under the IL-12 family of cytokines[28]. Role of IL-23 in IBD is already well recognized and explored by various studies [47][29][27].

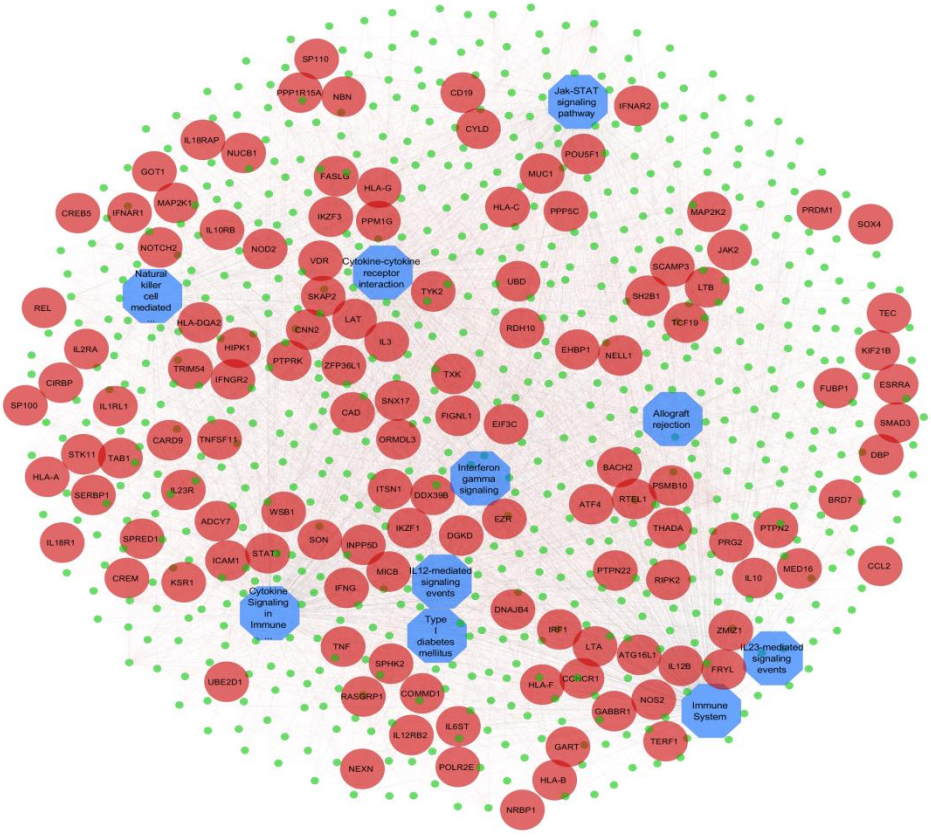


Figure 9: Network of 150 top prioritized candidate proteins (red) plus 650 related proteins (green) to these candidate proteins along with top 10 pathways (blue) in which they can be functionally classified

S.no	Pathways
1.	IL12-mediated signaling events
2.	IL23-mediated signaling events
3.	Allograft rejection
4.	Cytokine Signaling in Immune ...
5.	Type I diabetes mellitus
6.	Interferon gamma signaling
7.	Natural killer cell mediated ...
8.	Jak-STAT signaling pathway
9.	Cytokine-cytokine receptor interaction
10.	Immune System

Table 1 : Top 10 pathways in which the 816 nodes of PPI network were classified using GeneMANIA.

Library preparation

Ayurveda formulation used for treating intestinal disorders are made by using *Terminalia chebula , Cuminum cyminum,* and *Syzygium cumini*. Library preparation for plant derived molecules identified from these plants was done. Total of 127 ligands were identified and their 3D structure were downloaded from Zinc database and IMPPAT database. Some of the metals identified to be found in these medicinal plants were ignored in this study. 3D SDF structure files were converted to 3D MOL2 files and all the structures were saved in a primary MOL2 file using OpenBabel 2.3.1 software [45]. .

Molecular docking

IL-23 docking

Docking of the plant derived molecules was done by focusing on the site of IL-23 and IL-23R interaction. IL-23 docking results identify Urolithin B, Urolithin A and Terflavin B as the best docked molecules. Terflavin B with lowest binding energy was the top docked molecule.

For further analysis bioactive molecules having more than 95% substructure similarity with Terflavin B were identified. Molecular docking for all of them with the IL-23 was done. Here the top docked molecules with their binding energy are: Terflavin A (-12.88), Terflavin B (-12.63), Punicalin (-10.90), Punicalagin (-10.65) and CHEMBL416615 (-10.8500).

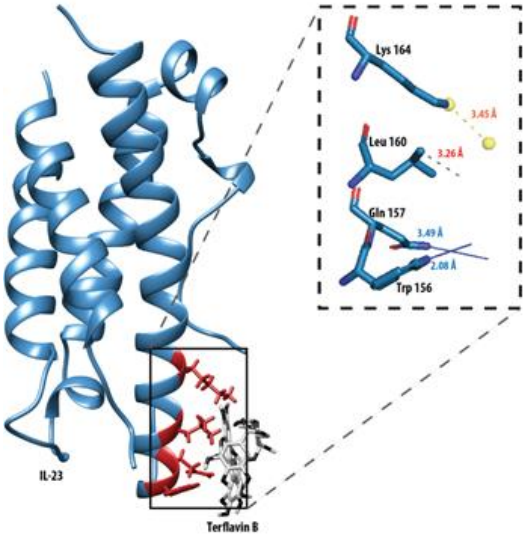
Punicalin and Punicalagin found in pomegranate are known for their antioxidant and anti-inflammatory properties [41][40]. Due to their well-known properties Punicalin and Punicalagin along with Terflavin B were further screened to identify their binding sites with the IL-23 using PLIP web tool [49]. All of them were observed to be interacting with crucial residues for IL-23 and IL-23R interaction. Terflavin B was found to forming hydrogen bonds with the Trp156 and Glu157 along with hydrophobic interactions with Leu160 and a salt bridge with Lys164. Punicalin was found to be forming hydrogen bonds with Asn66, Gln157 and Lys156 along with other hydrophobic interactions at Gln157 and Leu160 and salt bridge formation with Lys164. Punicalagin was also interacting with residues at the active site by forming hydrogen bond with Thr64, Lys164 and Arg167 along with hydrophobic interactions with Leu160 and

Phe163. The chemical interaction of these 3 molecules with the critical IL-23 residues shows their possibility as potential drug intervention.

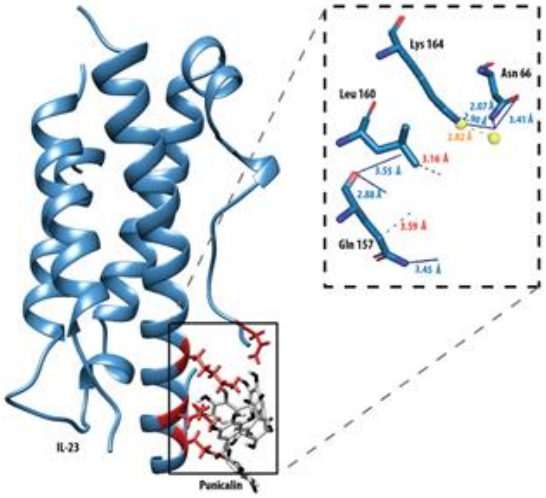
Protein name	PDB ID	Ligand name	ChEMBL ID	Binding energy
IL-23	5mxa	TERFLAVIN B	CHEMBL507965	-12.63
IL-23	5mxa	PUNICALIN	CHEMBL502440	-10.9
IL-23	5mxa	PUNICALAGIN	CHEMBL506814	-10.65

Table 2 : Top Binding energy of ligands binding with IL-23

i)



ii)



iii)

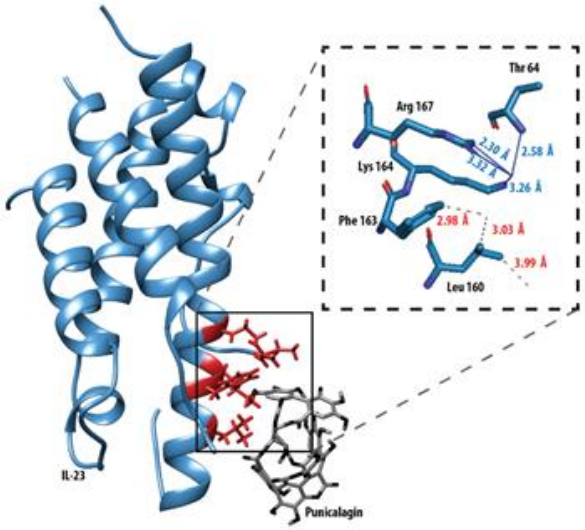


Figure 10: IL-23 residues interacting with (a) Terflavin B (b) Punicalin (c) Punicalagin. (Blue line = hydrogen bond, grey dotted line = hydrophobic interaction, yellow dotted line = salt bridge, yellow sphere = charge center)

IL-23R docking

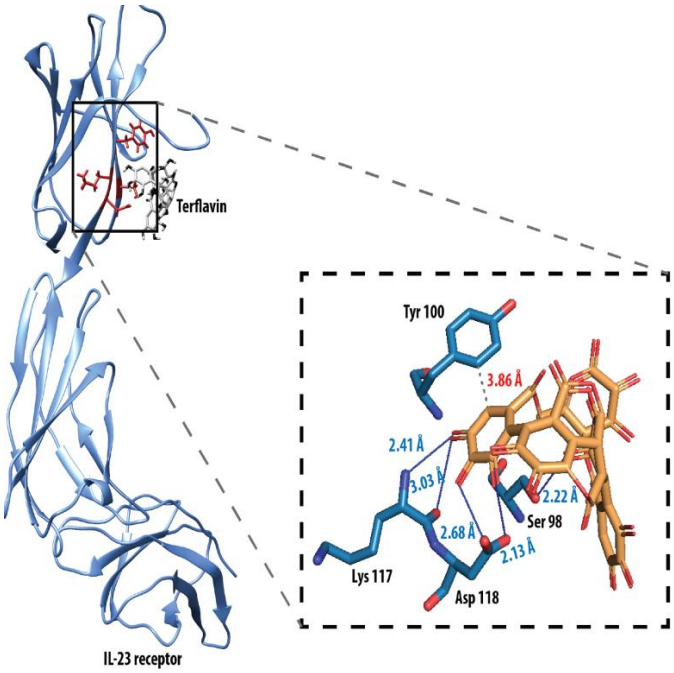
Docking of IL-23R with library having 125 ligands identified Terflavin B, Delphinidin 3-gentiobioside, Urolithin A 3,8-O-Diglucuronide and Petunidin 3-gentiobioside as the best docked ligands. Terflavin B having -14.8700 binding energy score was the best docked ligand for IL-23R as well.

For further analysis, bioactive molecules having more than 95% substructure similarity with Terflavin B were identified form CHEMBL database. All of the Terflavin B analogs had good binding affinity with IL-23R. Because of their known anti-inflammatory properties Punicalin and Punicalagin were selected along with Terflavin B for further analysis [40], [41]. The top docked models of these three were analyzed using PLIP web tool [49]. All of them showed interactions with the critical residues for IL-23:IL-23R interaction. Punicalagin having binding energy equal to -13.96 was forming hydrogen bonds with Thr26, Asn27, Ile28, Ser98, Gly116, Lys117 and Asp118. Terflavin B having -14.47 as binding energy was forming hydrogen bond with Ile28 and two hydrogen bonds with Asp118. Punicalin having binding energy of -11.81 was forming hydrogen bond with Ile28, Gly32, Ile34, Ser98, Gly116, Lys117 and Asp118. Apart from this Punicalin was also forming a salt bridge with Lys117.

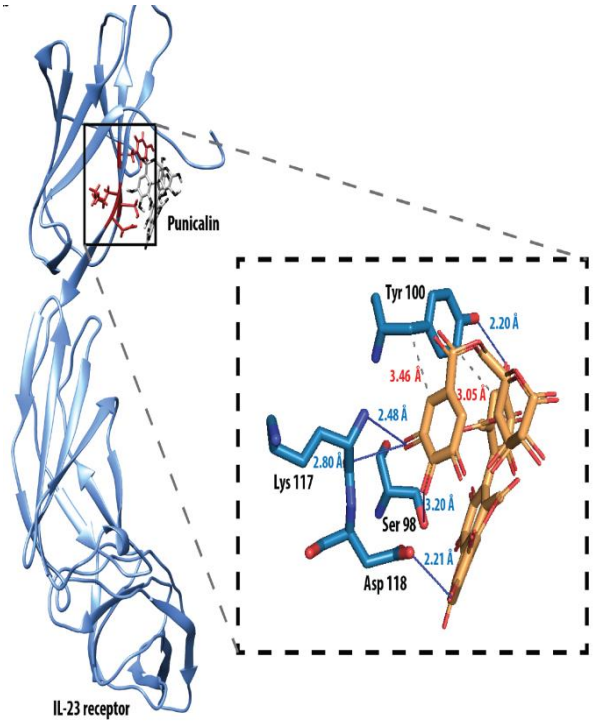
Protein name	PDB ID	Ligand name	ChEMBL ID	Binding energy
IL-23R	5mzv	TERFLAVIN B	CHEMBL507965	-14.47
IL-23R	5mzv	PUNICALIN	CHEMBL502440	-11.81
IL-23R	5mzv	PUNICALAGIN	CHEMBL506814	-13.96

Table 3: Top Binding energy of ligands binding with IL-23R

i)


Terflavin
IL-23 receptor
Tyr 100
3.86 Å
2.41 Å
3.03 Å
2.22 Å
Ser 98
2.68 Å
2.13 Å
Lys 117
Asp 118

ii)


Punicalin
IL-23 receptor
Tyr 100
2.20 Å
3.46 Å
3.05 Å
2.48 Å
2.80 Å
Lys 117
3.20 Å
Ser 98
2.21 Å
Asp 118

iii)

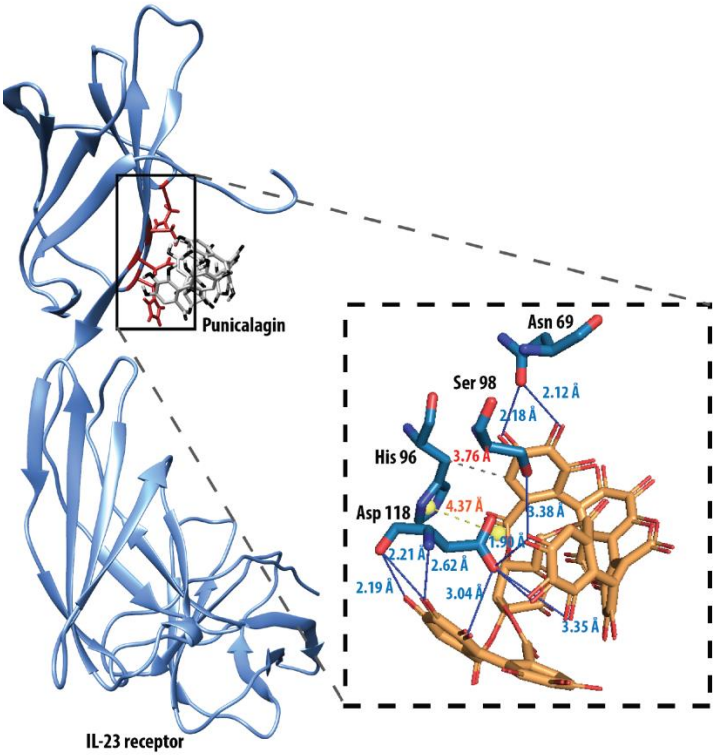


Figure 11: IL-23R residues interacting with (i) Terflavin B, (ii) Punicalin (iii) Punicalagin. (Blue chain/residues = IL-23R residues, yellow chain = Ligand, Blue line = hydrogen bond, grey dotted line = hydrophobic interaction, yellow dotted line = salt bridge, yellow sphere = charge center).

Molecular dynamic simulation

Molecular docking only represents some frames of the protein ligand interaction. To analyses various other states that protein- ligands can acquire because of their dynamic behavior in solvent molecular, dynamic simulations are used. In this study simulation was done for 10ns using TIP3P water model. Various dynamic behaviors of the systems are as follows.

Molecular dynamic simulation of IL-23:ligand complexes

The configuration changes of three protein-ligand complexes were analyzed using RMSD (Root mean square deviation) for 10n simulations. For all the three complexes RMSD was observed to be increasing for the starting few picoseconds. This increased RMSD indicates substantial conformation change in protein:ligand complex structure in the starting few picoseconds. After this a plateau can be observed in the three complex systems. Within the 10ns simulation the average RMSD for IL-23 complex with Punicalin was 0.632, with Punicalagin was 0.554 and with Terflavin B was 0.43.

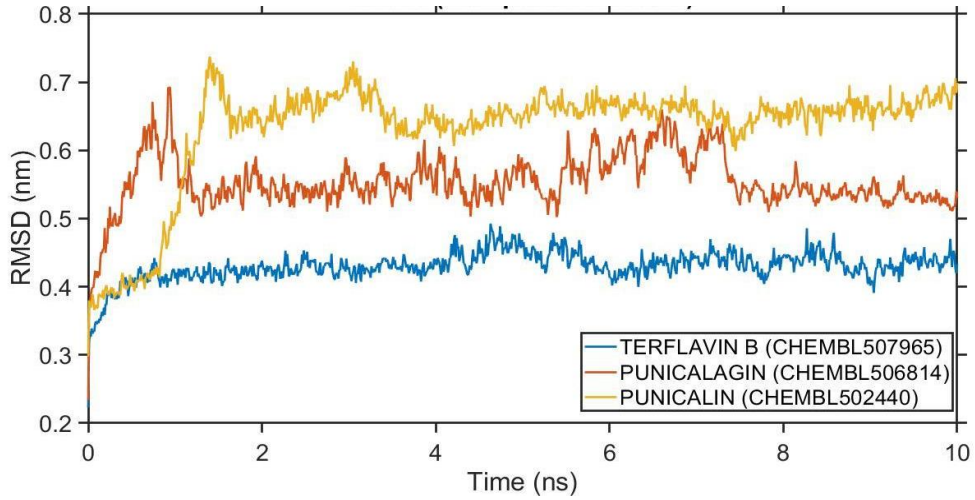


Figure 12: RMSD plot of protein (IL-23) forming complex with ligand.

To calculate the fluctuation of each residue of the protein, RMSF calculations were done. The average residue fluctuations for the three complexes here were below 0.2nm. All the complexes had less fluctuation in backbone, indicating adequate stability.

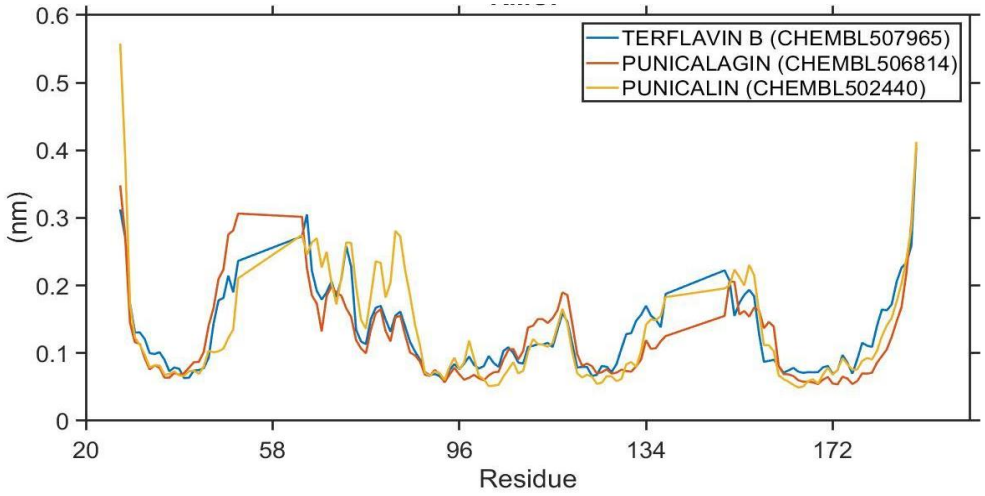


Figure 13: RMSF plots of protein (IL-23)-ligand complexes for 10ns simulation.

Radius of gyration (Rg) is used to measure folding and unfolding stability of protein structure when bonded with ligand. If the Rg value of protein is stable throughout the simulation then it depicts stably folded protein. When Rg value changes abruptly during simulation, protein can be regarded as unstably unfolded during simulation. Here we observed that IL-23:Terflavin B was forming a relatively steady folded complex structure as compared to the other two. These results were also verified by visualizing the last frame of the 10ns simulation. After structural analysis it was verified that Terflavin B was least displaced from its initial docked position.

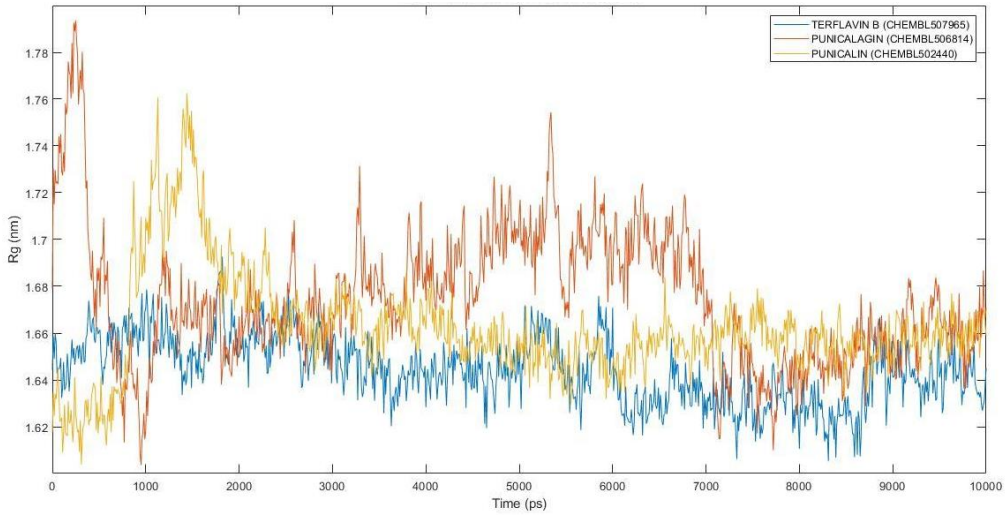


Figure 14: Plot of radius of gyration observed for the protein (IL-23)-ligand complexes.

The most common interaction between protein and a ligand is a hydrogen bond. To further evaluate the stability of the complexes, time series plots for hydrogen bonds were analyzed. The complexes being observed were showing hydrogen bonds between protein and ligand throughout the 10ns simulation. The number of hydrogen bonds between IL-23 and Punicalagin were greatly increased towards the end of the simulation.

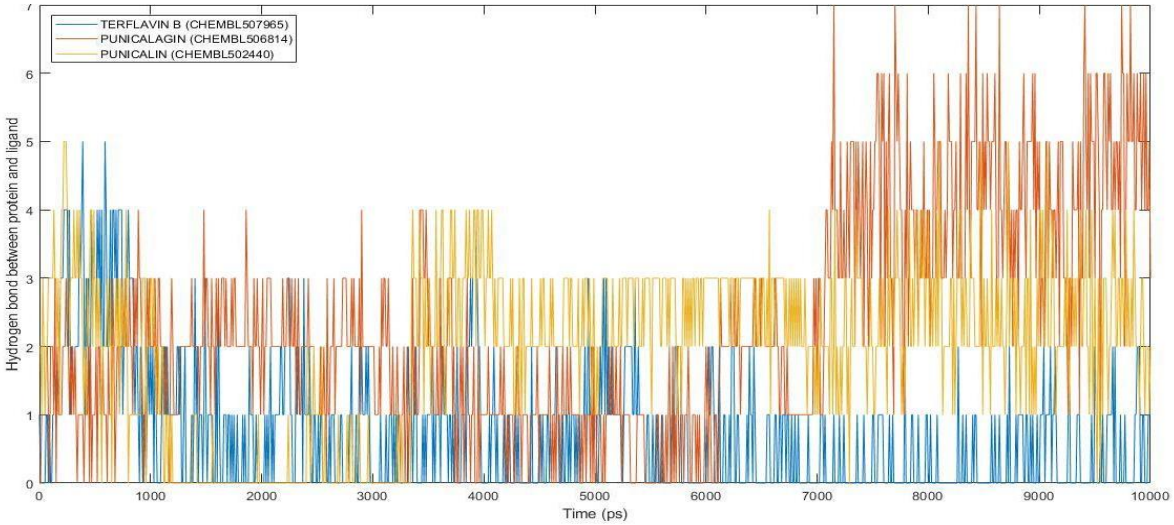


Figure 15: Plot showing number of Hydrogen bonds between protein (IL-23) and the ligands (legends in the box on the northwest side of plot).

The strength of receptor ligand complex is calculated by finding their interaction energy during molecular dynamic simulation. The average interaction energy for IL-23:Terflavin B complex was -

105.89, which was the best among the three observed IL-23:ligand complexes. IL-23:Punicalagin and IL-23:Punicalin complexes had -74.2418, -90.156 as respective interaction energy. The interaction energy result of molecular dynamic simulation complements the molecular docking results.

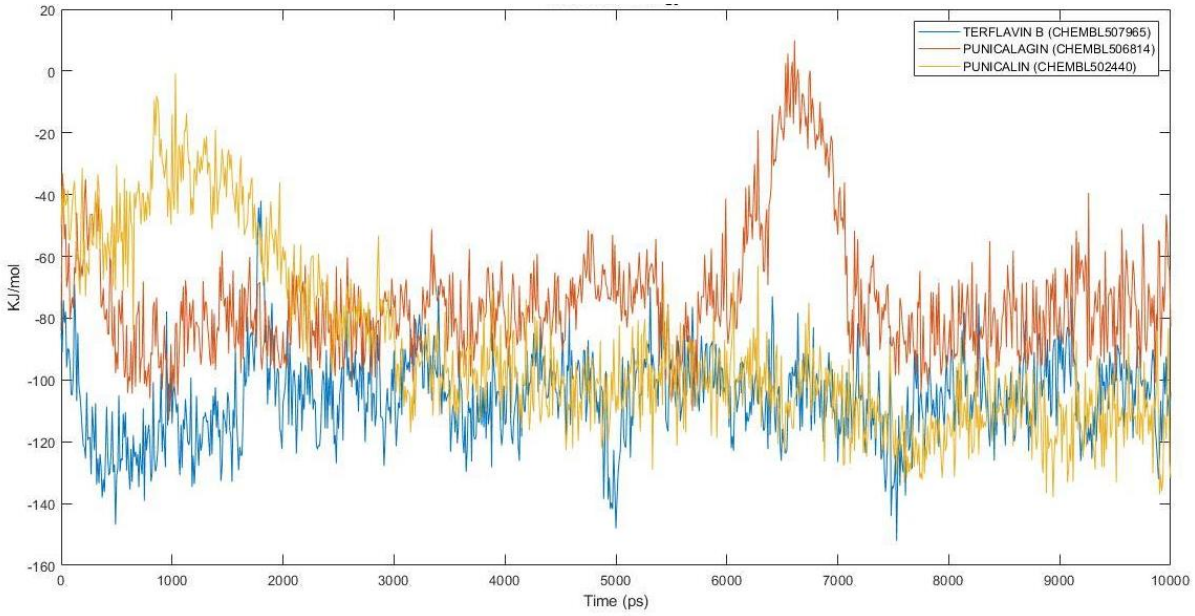


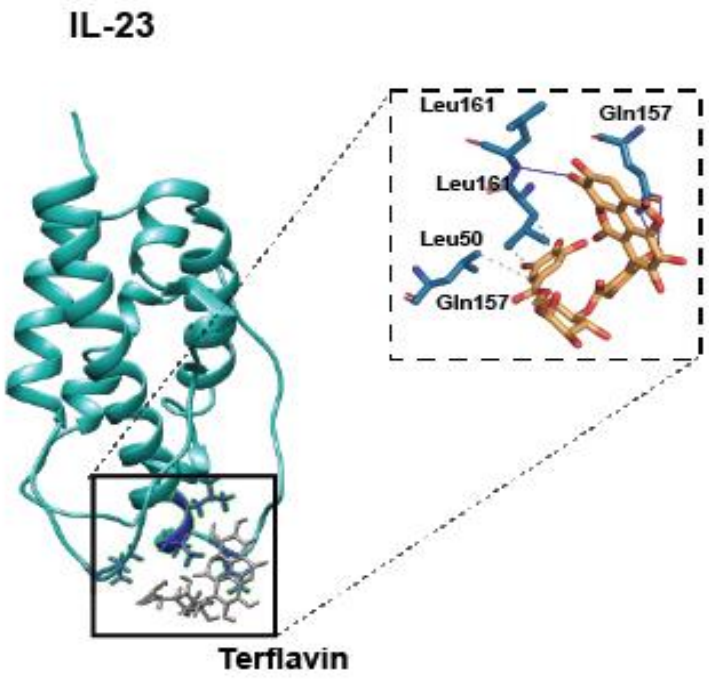
Figure 16: Plot for interaction energy of protein IL-23 and ligands complexes (Ligands are color coded according to the legends provided in upper right corner of the plot).

At the end of the simulation the interactions interface between IL-23 and the ligands were again analyzed to evaluate if the ligands were bound to the desired position or not. At the end of the simulation Terflavin B was making hydrogen bonds with Leu161 and two hydrogen bonds with Gln157. Apart from this it showed hydrophobic interactions with Leu50 and Leu160 (Figure 17(i)). After 10ns simulation Punicalagin was bound to IL-23 by forming hydrogen bonds with Thr64, Thr65, Asn66, Asp67 and Val68. Punicalagin was also a salt bridge with Lys164 (Figure 17(ii)). Punicalin was forming only one hydrogen bond with Gln157 at the end of the 10ns simulation. Punicalin was also forming a salt bridge with Arg158 ,pi-stacking with Trp156 and hydrophobic interactions with Leu159. Punicalagin was greatly displaced from its desired docked position after 10ns simulation.

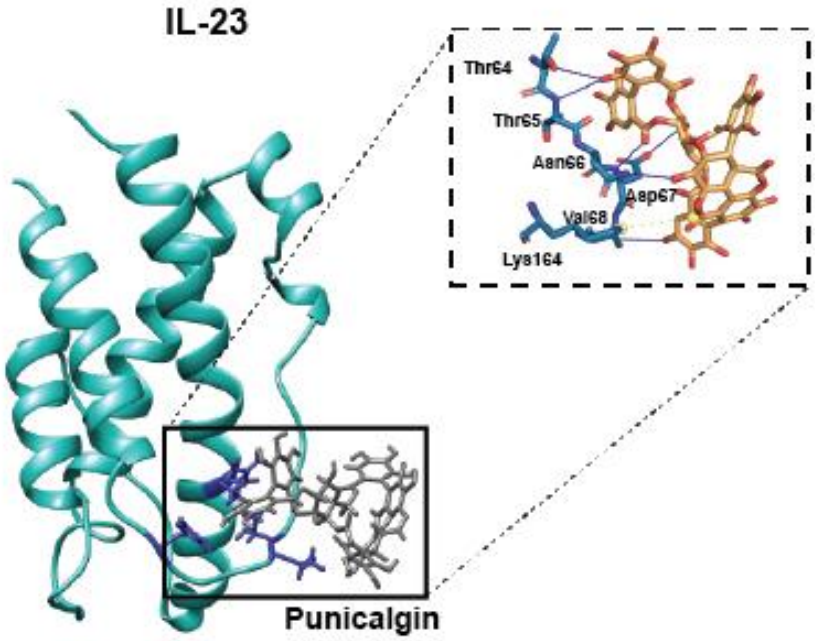
Protein name	PDB ID	Ligand name	ChEMBL ID	Average interaction energy (KJ/mol)
IL-23	5mxa	TERFLAVIN B	CHEMBL507965	-105.89
IL-23	5mxa	PUNICALIN	CHEMBL502440	-90.156
IL-23	5mxa	PUNICALAGIN	CHEMBL506814	-74.2418

Table 4: Average interaction energy between IL-23 and ligands at site of interest

i)


IL-23
Leu161
Gln157
Leu161
Leu50
Gln157
Terflavin

ii)


IL-23
Thr64
Thr65
Asn66
Asp67
Val68
Lys164
Punicalgin

iii)

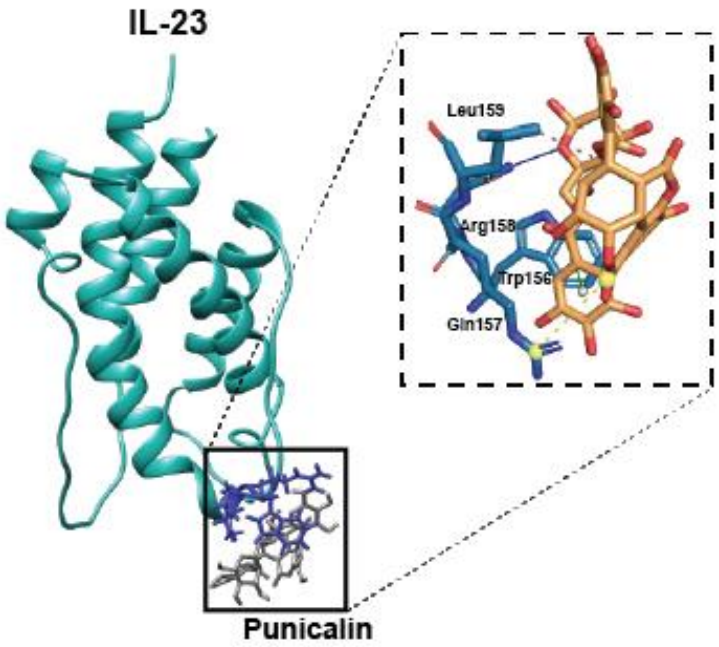


Figure 17: IL-23 residues interacting with (i) Terflavin B (ii) Punicalagin (iii) Punicalin at the end of 10ns simulation. (Blue line = hydrogen bond, grey dotted line = hydrophobic interaction, yellow dotted line = salt bridge, yellow sphere = charge center, green dotted line = pi-stacking)

Molecular dynamic simulation of IL-23R:ligand complexes

During MD simulation, the complex formed between IL-23R and Punicalin got broken, depicting that their bonding was not stable. Configuration change for the rest two protein-ligand complexes were analyzed by calculating RMSD during 10ns simulation. The average RMSD value for Terflavin B-IL23R complex was 0.29941, for Punicalagin-IL23R complex was 0.37359. RMSD fluctuations were in acceptable range for the studied period.

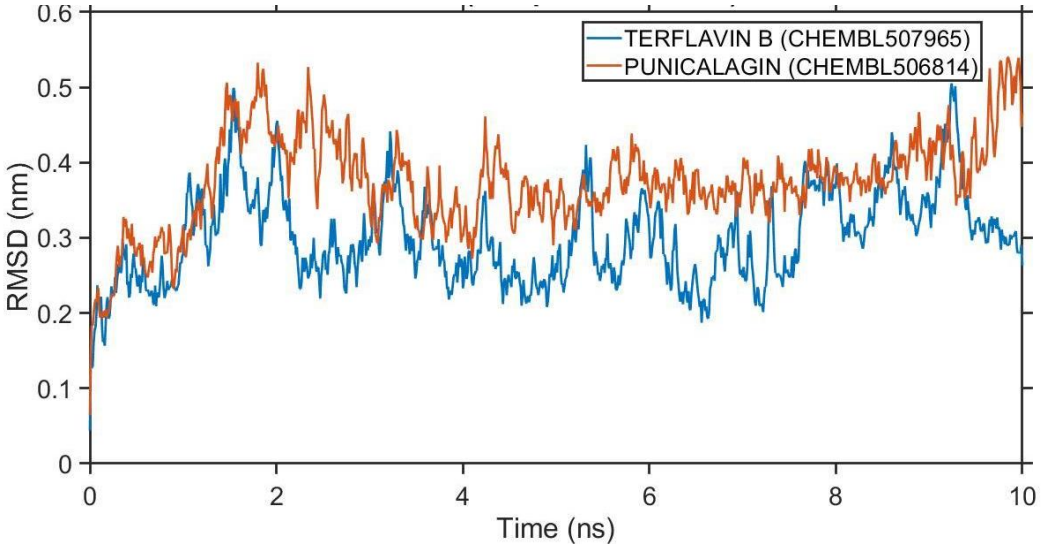


Figure 18: RMSD plot of protein (IL-23R) forming complex with ligand.

RMSF measures local change in protein structure. Backbone atoms of IL-23R in both the complexes were showing below 0.25ns average RMSF fluctuation. This indicates that in both cases stability of the protein (IL-23R) was maintained.

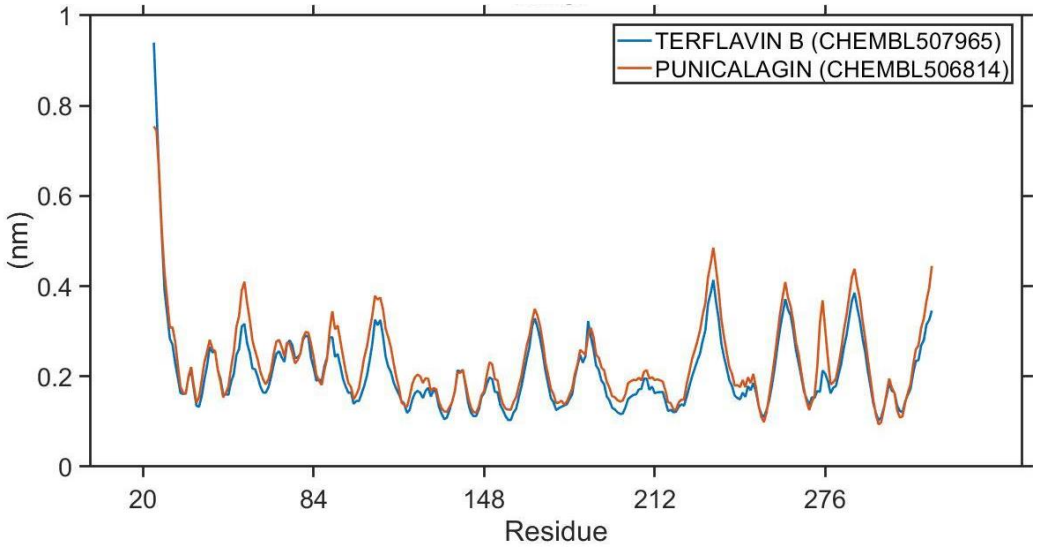


Figure 19: RMSF plots of protein (IL-23R) backbone for 10ns simulation.

Radius of gyration was used to evaluate if the complexes were stably folded or unfolded. The average Rg value for Punicalagin was 3.028 ± 0.05 and for Terflavin B was 3.09 ± 0.046. This relatively small

deviation in radius of gyration depicts that the complexes were stably folded during the 10ns MD simulation.

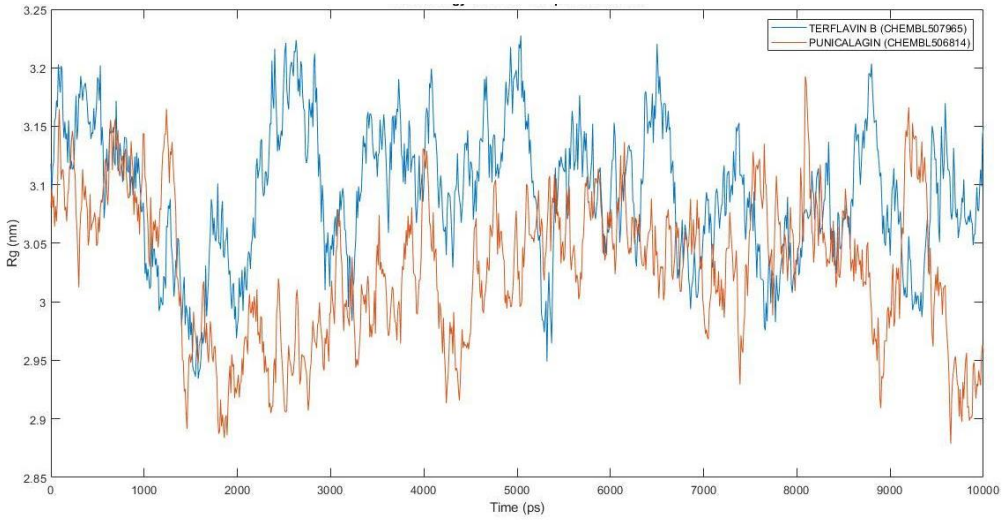


Figure 20: Plot of radius of gyration observed for the proteins (IL-23R) -ligand complexes.

Hydrogen bonds between protein and ligand are critical for formation of stable complex between them. Hydrogen bonds in both the complexes were counted. It was observed that in both the cases ligands were forming hydrogen bonds with the protein till the very last frame of 10ns simulation.

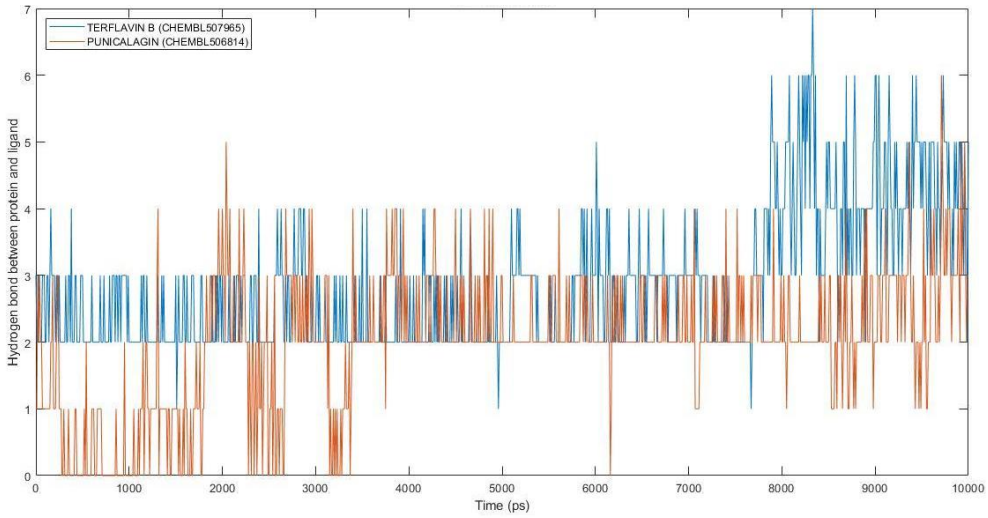


Figure 21: Plot showing number of Hydrogen bonds between protein (IL-23R) and the ligand.

Interaction energy between the receptor and the ligand depicts the strength of the complex formed. The average interaction energy for IL-23R:Terflavin B complex was -100.99 kJ/mol and for IL-23:Punicalagin was -39.39 kJ/mol. Interaction energy for IL-23R:Terflavin B complex was in the acceptable range.

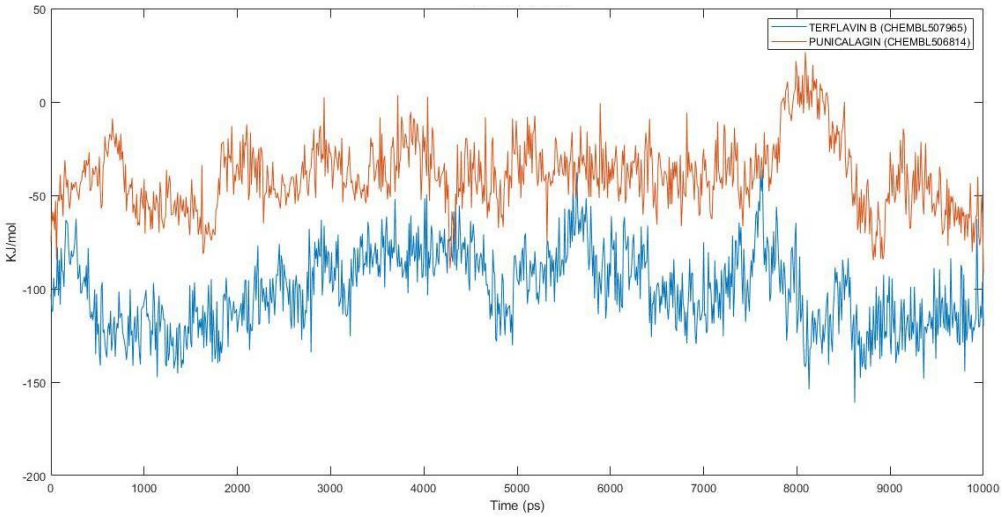


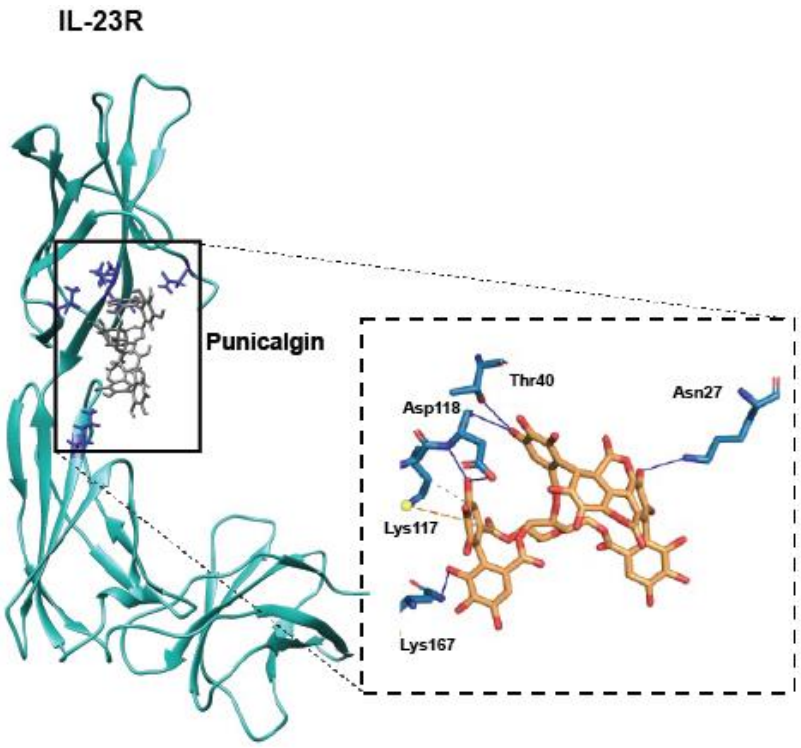
Figure 22: Plot for interaction energy of IL-23R - ligands complexes (Ligands are color coded according to the legends provided in upper right corner of the plot).

At the end of 10ns simulation Terflavin B was forming hydrogen bonds with Asn29, Ser98, Met99, Tyr100, Ile114 and Asp118. Punicalagin was forming hydrogen bonds with Asn27, Thr40, Asp118 and Lys157. In both the cases the ligand was still bound to Asp118, the residue important for high affinity binding of IL-23 to IL-23R.

Protein name	PDB ID	Ligand name	ChEMBL ID	Average interaction energy kJ/mol
IL-23R	5mzv	TERFLAVIN B	CHEMBL507965	-100.99
IL-23R	5mzv	PUNICALAGIN	CHEMBL506814	-39.39

Table 5: Average interaction energy between IL-23R and ligands at site of interest

i)



ii)

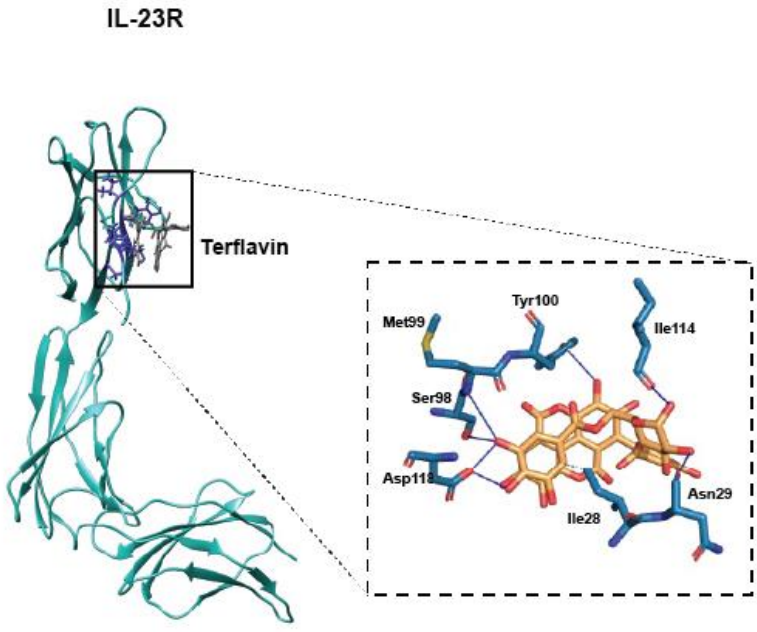


Figure 23: IL-23R residues interacting with (i) Punicalagin (ii) Terflavin B at the end of 10ns simulation. (Blue line = hydrogen bond, grey dotted line = hydrophobic interaction)

Chapter 4

Discussion

IBD (Inflammatory bowel disease) is a complex disease which causes inflammation in gastrointestinal tract[22]. In absence of an adequate treatment this inflammation can lead to weight loss, abdominal pain, bloody stool and severe damage of the intestine can also cause ulceration[54]. A genetically susceptible person when exposed to adverse environmental conditions may develop microbial dysbiosis in gut which impair the intestinal epithelial barrier function[4]. This causes invasion of microbes which stimulate invasion of macrophages and dendritic cells (DC) [55]. These DC and macrophages produce various cytokines (IL-4, IL-6, IL-12, IL-23, IL-27, TGFβ). These cytokines causes differentiation of t-naïve cells into Th1, Th2, Th17 cell subtypes. These T helper cell subtypes produce pro inflammatory cytokines (like IL-17, IFNγ). In this study , 354 genes found to be associated with Crohn's disease (CD) [56] were downloaded and prioritized according to their relatedness to already known CD susceptible genes . 41 of these 150 top prioritized genes were found to be differentially expressed in CD patient's colon biopsies (accession number GSE20881). To look deeply into the processes in which these 150 genes' products are involved, an additional set of 650 genes related to them were identified using GeneMANIA[37]. PPI network of these 800 gene product showed scale free nature. GeneMANIA also identified the consolidated pathways in which these 800 genes can be classified. Top 10 pathways were all related to immune system. In this list top two pathways were IL-12 and IL-23 mediated signaling events. In IBD these cytokines are responsible for differentiation of t-naïve cells into Th1, Th2, Th17 cell subtypes. Cytokines like IL-17 produced by these cells are pro-inflammatory in nature[55]. In mouse model IL-17 inhibition was not very effective. Interestingly IL-17 inhibition worsened the inflammation and also intestinal epithelial barrier was weakened. In mouse models inhibition of IL-23 increase T regulatory cell accumulation, which are essential for intestinal homeostasis [27]. Targeting IL-23 or IL-23 receptor (IL-23R) shows promising positive results for various autoimmune diseases [23].Extracts from plants like *Terminalia chebula and Cuminum cyminum* are used Ayurveda to treat gut related issues[30]. In mice model, Dried fruit extracts from *Terminalia chebula* was found to be modulating TH1 response[32]. Molecules identified in these medicinal plants were screened to identify inhibitors against IL-23. Our docking studies identified that Terflavin B, Punicalin and Punicalagin were binding to critical residues of both IL-23 and IL-23R with good binding energy. MD Simulations of the protein ligand complexes were conducted to further analyses the insight behavior of their bond formed. After MD simulation it was analyzed that Terflavin B was stably bound to the critical residues of IL-23R with interaction energy of -100.99 kJ/mol. Whereas Punicalin (-90.156 kJ/mol) and Terflavin B (-105.89 kJ/mol) were bound to critical residues of IL-23 effectively during the entire course of 10ns MD simulation. By different bioassays it is known that Terflavin B showed antibacterial activity against *Bacillus subtilis* and *Pseudomonas fluorescens* [57]. Punicalin and Punicalagin found in pomegranate shows strong antioxidant and anti-inflammatory properties [40][41]. In this study Terflavin B has better stability with both IL-23 and IL-23R as compared to all the analyzed ligands. The results from this study suggest that these plant derived molecules can be further explored as potential inhibitor of IL-23 signaling axis. Inhibiting IL-23 can in turn help to minimize inflammation caused in IBD and in other autoimmune diseases like multiple sclerosis, psoriasis and rheumatoid arthritis[23].

Bibliography

[1] K. L. Hammond, "Inflammatory Bowel Disease," in *The Clinician's Handbook of Natural Medicine*, H. J.-B. Joseph E. Pizzorno, Michael T. Murray, Ed. Churchill Livingstone, 2016, pp. 547–564.

[2] G. 2017 I. B. D. Collaborators*, "The global , regional , and national burden of inflammatory bowel disease in 195 countries and territories , 1990 – 2017 : a systematic analysis for the Global Burden of Disease Study," *Lancet Gastroenterol Hepatol*, vol. 5, pp. 17–30, 2020, doi: 10.1016/S2468-1253(19)30333-4.

[3] S. Aniwan, S. H. Park, and E. V. Loftus, "Epidemiology, Natural History, and Risk Stratification of Crohn's Disease," *Gastroenterol. Clin. North Am.*, vol. 46, no. 3, pp. 463–480, 2017, doi: 10.1016/j.gtc.2017.05.003.

[4] M. Zaeem Cader and A. Kaser, "Recent advances in inflammatory bowel disease: Mucosal immune cells in intestinal inflammation," *Gut*, vol. 62, no. 11, pp. 1653–1664, 2013, doi: 10.1136/gutjnl-2012-303955.

[5] T. Zuo and S. C. Ng, "The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory bowel disease," *Front. Microbiol.*, vol. 9, no. SEP, pp. 1–13, 2018, doi: 10.3389/fmicb.2018.02247.

[6] T. C. Liu and T. S. Stappenbeck, "Genetics and Pathogenesis of Inflammatory Bowel Disease," *Annu. Rev. Pathol. Mech. Dis.*, vol. 11, no. 7351, pp. 127–148, 2016, doi: 10.1146/annurev-pathol-012615-044152.

[7] R. Balfour Sartor and G. D. Wu, "Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches," *Gastroenterology*, vol. 152, no. 2, pp. 327–339, 2017, doi: 10.1053/j.gastro.2016.10.012.

[8] D. N. Frank, A. L. St. Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace, "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 34, pp. 13780–13785, 2007, doi: 10.1073/pnas.0706625104.

[9] S. Buhner *et al.*, "Genetic basis for increased intestinal permeability in families with Crohn's disease: Role of CARD15 3020insC mutation?," *Gut*, vol. 55, no. 3, pp. 342–347, 2006, doi: 10.1136/gut.2005.065557.

[10] A. Kaser *et al.*, "XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease," *Cell*, vol. 134, no. 5, pp. 743–756, 2008, doi: 10.1016/j.cell.2008.07.021.

[11] Y. R. Na, M. Stakenborg, S. H. Seok, and G. Matteoli, "Macrophages in intestinal inflammation and resolution: a potential therapeutic target in IBD," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 16, no. 9, pp. 531–543, 2019, doi: 10.1038/s41575-019-0172-4.

[12] A. A. Kühl, U. Erben, L. I. Kredel, and B. Siegmund, "Diversity of intestinal macrophages in inflammatory bowel diseases," *Front. Immunol.*, vol. 6, no. DEC, pp. 1–7, 2015, doi: 10.3389/fimmu.2015.00613.

[13] A. W. Segal, "Studies on patients establish Crohn's disease as a manifestation of impaired innate immunity," *J. Intern. Med.*, vol. 286, no. 4, pp. 373–388, 2019, doi: 10.1111/joim.12945.

[14] T. L. Denning, Y. C. Wang, S. R. Patel, I. R. Williams, and B. Pulendran, "Lamina propria macrophages and dendritic cells differentially induce regulatory and interleukin 17-producing T cell responses," *Nat. Immunol.*, vol. 8, no. 10, pp. 1086–1094, 2007, doi: 10.1038/ni1511.

[15] P. S. Steinbach EC, "The role of macrophages and dendritic cells in the initiation of inflammation in IBD," *Inflamm Bowel Dis*, vol. 20, no. 1, pp. 166–175, 2014, doi: 10.1097/MIB.0b013e3182a69dca.

[16] M. G. Neuman, "Immune dysfunction in inflammatory bowel disease," *Transl. Res.*, vol. 149, no. 4, pp. 173–186, 2007, doi: 10.1016/j.trsl.2006.11.009.

[17] B. Ahluwalia, L. Moraes, M. K. Magnusson, and L. Öhman, "Immunopathogenesis of inflammatory bowel disease and mechanisms of biological therapies," *Scand. J. Gastroenterol.*, vol. 53, no. 4, pp. 379–389, 2018, doi: 10.1080/00365521.2018.1447597.

[18] Q. Guan and J. Zhang, "Recent Advances: The Imbalance of Cytokines in the Pathogenesis of Inflammatory Bowel Disease," *Mediators Inflamm.*, vol. 2017, no. Dc, 2017, doi: 10.1155/2017/4810258.

[19] M. Mahapatro, L. Erkert, and C. Becker, "Cytokine-Mediated Crosstalk between Immune Cells and

Epithelial Cells in the Gut," *Cells*, vol. 10, no. 1, 2021, doi: 10.3390/cells10010111.

[20] et al Marianne Orholm, Pia Munkholm, Ebbe Langholz, "Familial occurrence of Inflammatory Bowel Disease," *N. Engl. J. Med.*, vol. 329, no. 14, pp. 977–986, 1991, [Online]. Available: http://content.nejm.org/cgi/content/abstract/329/14/977%5Cnhttp://www.nejm.org/doi/abs/10.1056/NEJM199309303291401.

[21] C. Tysk, E. Lindberg, G. Jarnerot, and B. Floderus-Myrhed, "Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking," *Gut*, vol. 29, no. 7, pp. 990–996, 1988, doi: 10.1136/gut.29.7.990.

[22] G. P. Ramos and K. A. Papadakis, "Mechanisms of Disease: Inflammatory Bowel Diseases," *Mayo Clin. Proc.*, vol. 94, no. 1, pp. 155–165, 2019, doi: 10.1016/j.mayocp.2018.09.013.

[23] C. Tang, S. Chen, H. Qian, and W. Huang, "Interleukin-23: As a drug target for autoimmune inflammatory diseases," *Immunology*, vol. 135, no. 2, pp. 112–124, 2012, doi: 10.1111/j.1365-2567.2011.03522.x.

[24] G. et al. Xie, "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease," *Physiol. Behav.*, vol. 176, no. 12, pp. 139–148, 2013, doi: 10.1038/nature11582.Host-microbe.

[25] C. J. Cardinale and H. Hakonarson, "Genetics of inflammatory bowel diseases," *Gastroenterology*, vol. 149, no. 5, pp. 1163–1176, 2015, doi: 10.1053/j.gastro.2015.08.001.

[26] D. Muraro, D. A. Lauffenburger, and A. Simmons, "Prioritisation and network analysis of Crohn's disease susceptibility genes," *PLoS One*, vol. 9, no. 9, 2014, doi: 10.1371/journal.pone.0108624.

[27] J. R. Maxwell *et al.*, "Differential Roles for Interleukin-23 and Interleukin-17 in Intestinal Immunoregulation," *Immunity*, vol. 43, no. 4, pp. 739–750, 2015, doi: 10.1016/j.immuni.2015.08.019.

[28] D. A. A. Vignali and V. K. Kuchroo, "IL-12 family cytokines: immunological playmakers," *Nat. Immunol.*, vol. 13, no. 8, pp. 722–728, 2012, doi: 10.1038/ni.2366.

[29] Y. Bloch *et al.*, "Structural Activation of Pro-inflammatory Human Cytokine IL-23 by Cognate IL-23 Receptor Enables Recruitment of the Shared Receptor IL-12Rβ1," *Immunity*, vol. 48, no. 1, pp. 45-58.e6, 2018, doi: 10.1016/j.immuni.2017.12.008.

[30] I. P. S. Uma Ranjan Lal, "Review and Implications of Traditional Indian Medicine for Inflammatory Bowel Disease," *Intech*, vol. 32, no. tourism, pp. 137–144, 1989, doi: http://dx.doi.org/10.5772/intechopen.89465.

[31] U. R. Lal, S. M. Tripathi, S. M. Jachak, K. K. Bhutani, and I. P. Singh, "Chemical changes during fermentation of Abhayarishta and its standardization by HPLC-DAD," *Nat. Prod. Commun.*, vol. 5, no. 4, pp. 575–579, 2010, doi: 10.1177/1934578x1000500417.

[32] I. Rubab and S. Ali, "Dried fruit extract of Terminalia chebula modulates the immune response in mice," *Food Agric. Immunol.*, vol. 27, no. 1, pp. 1–22, 2016, doi: 10.1080/09540105.2015.1055554.

[33] L. V. Brumatti, A. Marcuzzi, P. M. Tricarico, V. Zanin, M. Girardelli, and A. M. Bianco, "Curcumin and inflammatory bowel disease: Potential andlimits of innovative treatments," *Molecules*, vol. 19, no. 12, pp. 21127–21153, 2014, doi: 10.3390/molecules191221127.

[34] S. Aerts *et al.*, "Gene prioritization through genomic data fusion," *Nat. Biotechnol.*, vol. 24, no. 5, pp. 537–544, 2006, doi: 10.1038/nbt1203.

[35] L. C. Tranchevent *et al.*, "Candidate gene prioritization with Endeavour," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W117–W121, 2016, doi: 10.1093/nar/gkw365.

[36] H. Mi, A. Muruganujan, and P. D. Thomas, "PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees," *Nucleic Acids Res.*, vol. 41, no. D1, pp. 377–386, 2013, doi: 10.1093/nar/gks1118.

[37] D. Warde-Farley *et al.*, "The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function," *Nucleic Acids Res.*, vol. 38, no. SUPPL. 2, pp. 214–220, 2010, doi: 10.1093/nar/gkq537.

[38] Y. Assenov, F. Ramírez, S. E. S. E. Schelhorn, T. Lengauer, and M. Albrecht, "Computing topological

parameters of biological networks," *Bioinformatics*, vol. 24, no. 2, pp. 282–284, 2008, doi: 10.1093/bioinformatics/btm554.

[39] G. Yadav and S. Babu, "Nexcade: Perturbation analysis for complex networks," *PLoS One*, vol. 7, no. 8, 2012, doi: 10.1371/journal.pone.0041827.

[40] C. J. Lee, L. G. Chen, W. L. Liang, and C. C. Wang, "Anti-inflammatory effects of Punica granatum Linne in vitro and in vivo," *Food Chem.*, vol. 118, no. 2, pp. 315–322, 2010, doi: 10.1016/j.foodchem.2009.04.123.

[41] Y. Wang, H. Zhang, H. Liang, and Q. Yuan, "Purification, antioxidant activity and protein-precipitating capacity of punicalin from pomegranate husk," *Food Chem.*, vol. 138, no. 1, pp. 437–443, 2013, doi: 10.1016/j.foodchem.2012.10.092.

[42] K. Š. · R. Š. Relja Suručić1 · Biljana Tubić · Miloš P. Stojiljković · Dragan M. Djuric · Maja Travar · Milkica Grabež7, "Computational study of pomegranate peel extract polyphenols as potential inhibitors of SARS-CoV-2 virus internalization." Springer Science+Business Media, LLC, part of Springer Nature 2020 Abstract, doi: https://doi.org/10.1007/s11010-020-03981-7.

[43] T. Sterling and J. J. Irwin, "ZINC 15 - Ligand Discovery for Everyone," *J. Chem. Inf. Model.*, vol. 55, no. 11, pp. 2324–2337, 2015, doi: 10.1021/acs.jcim.5b00559.

[44] K. Mohanraj *et al.*, "IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry and Therapeutics," *Sci. Rep.*, vol. 8, no. 1, pp. 1–17, 2018, doi: 10.1038/s41598-018-22631-z.

[45] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," *J. Cheminform.*, vol. 3, no. 33, pp. 1–14, 2011, doi: doi:10.1186/1758-2946-3-33.

[46] G. M. Morris *et al.*, "AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility," *J. Comput. Chem.*, vol. 30, no. 16, pp. 2785–2791, 2009, doi: 10.1002/jcc.21256.

[47] M. J. Mcgeachy *et al.*, "Interleukin 23 receptor is essential for terminal differentiation of effector T helper type 17 cells in vivo," *Nature*, vol. 10, no. 3, pp. 314–324, 2009, doi: 10.1038/ni.1698.Interleukin.

[48] L. RA and S. MB, "LigPlot+: multiple ligand-protein interaction diagrams for drug discovery," *J. Chem. Inf. Model.*, vol. 51, pp. 2778–2786, 2011.

[49] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, "PLIP: Fully automated protein-ligand interaction profiler," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W443–W447, 2015, doi: 10.1093/nar/gkv315.

[50] S. Pronk *et al.*, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, 2013, doi: 10.1093/bioinformatics/btt055.

[51] A. L. Barabási and Z. N. Oltvai, "Network biology: Understanding the cell's functional organization," *Nat. Rev. Genet.*, vol. 5, no. 2, pp. 101–113, 2004, doi: 10.1038/nrg1272.

[52] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science (80-.).*, vol. 297, no. 5586, pp. 1551–1555, 2002, doi: 10.1126/science.1073374.

[53] N. Kashtan, S. Itzkovitz, R. Milo, and U. Alon, "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs," *Bioinformatics*, vol. 20, no. 11, pp. 1746–1758, 2004, doi: 10.1093/bioinformatics/bth163.

[54] E. Szigethy, L. McLafferty, and A. Goyal, "Inflammatory bowel disease," *Child Adolesc. Psychiatr. Clin. N. Am.*, vol. 19, no. 2, pp. 301–318, 2010, doi: 10.1016/j.chc.2010.01.007.

[55] M. Kumar, M. Garand, and S. Al Khodor, "Integrating omics for a better understanding of Inflammatory Bowel Disease: A step towards personalized medicine," *J. Transl. Med.*, vol. 17, no. 1, pp. 1–13, 2019, doi: 10.1186/s12967-019-02174-1.

[56] L. et al Jostins, "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease," *Nature*, vol. 491, pp. 119–124, 2012, doi: 10.1038/nature11582.

[57] J. Conrad *et al.*, "Isoterchebulin and 4,6-O-isoterchebuloyl-D-glucose, novel hydrolyzable tannins from Terminalia macroptera," *J. Nat. Prod.*, vol. 64, no. 3, pp. 294–299, 2001, doi: 10.1021/np000506v.

Curriculum Vitae (CV)

Tushar Dhyani

Email: tushar19221@iiitd.ac.in
Phone: 9711532625
DOB: 06, 30, 1997
Address: E-151/A Brij vihar, Ghaziabad, U.P, India

Education

Indraprastha Institute of Information Technology, Delhi CGPA: 8
M.Tech (Computational biology) (3nd Semester)
2019 – Present

Jaypee Institute of Information Technology, Noida CGPA: 7.7
B.Tech (Biotechnology)
2015 – 2019

Bhai Parmanand Vidya Mandir School, Delhi CGPA: 7.16
C.B.S.E (Science - PCMB)
2012 – 2014

Vidya Bharati School, Ghaziabad CGPA: 10
C.B.S.E
2011 – 2012

Skills

Expertise Area	Computational biology, Biotechnology
Programming Language	Python, R
Tools and Technologies	Gromacs, AutoDock, Cytoscape, MEGA software, RNA-seq, Docker, PCR, AWS Machine Learning (Basics) , Immunological assays, Nanotechnology
Technical Electives	Data science, Network biology , Computer aided drug design, Genetics, Clinical trial management, Bio-pharmaceutics, Bio-economics, Bioinformatics for infectious diseases, Market research

Internship

Modern Diagnostic and Research Centre (Industrial) (JUNE–JULY)(2018)
Team Size- 6

Trained in Microbiology department
Worked on staining techniques, biochemical tests and microbial cultures

Sun Pharma Industrial Limited (Industrial) (MAY–JULY)(2017)
Team Size- 4

Trained in Clinical Pharmacology and Pharmacokinetics department

Projects

Network Analysis of Inflammatory Bowel Disease (JULY 2020-MAY 2021)
Guide : Dr. Ganesh Bagler

- Prioritization of the genes associated with IBD
- PPI Network construction of the top prioritized genes
- Analysis of the pathways associated with those genes and correlating with IBD pathogenesis
- Targeting the leads obtained by network analysis:
    - Molecular docking
    - Molecular dynamic simulation

Reverse Vaccinology (AUG–NOV)(2019)
Guide: Dr. Vibhor Kumar

- Epitope prediction for subunit vaccine development

Analysis of SARS CoV-2 Main Protease (JAN-MAY)(2020)
Guide: Dr. Vibhor Kumar

- Analyzed evolutionary profile of SARS CoV-2
- Compared SARS and SARS-2's main protease
- Coevolution study using CoeViz web-based tool
- Residue interaction network analysis using structural and coevolution profile

Course Coordination System Designing (Python) (AUG–DEC) (2019)
Team Size- 2

- Designed and coded the system using object oriented programming principles

Fabrication of Bio-inspired Solar Cell (JUN 2018-MAY 2019)
Guide: Prof. Sudha Srivastava Team Size- 5

- Developed solar cell using chlorophyll, TiO_2 and Gold Nanoparticles

In Silico Exploration of *Brucella* Species using Network Reconstruction (JAN–MAY)(2018)
Team Size- 5
Guide: Dr. Chakresh Kumar Jain

- Literature survey for *Brucella* Species
- Analyzed PPI network of *Brucella* Species

Survey for Usage of Blue-Green Dustbin (JAN–MAY)(2019)
Guide: Dr. Ashwani Mathur Team Size- 2

- Conducted primary survey
- Data analysis

Awards and Achievements

- GATE 2019 Qualified *(AIR 1236)*
- JNUEE-CEEB 2019 Qualified *(AIR 33)*
- NPTEL certification course "Bioengineering: An Interface with Biology and Medicine"

(*AIR 1)*
- Winner of Talent hunt competition for Dance (2015, Jiit Noida)
- Meritorious Performance Award in Secondary School Examination

Interests and Hobbies

- Dance
- Poetry
- Taekwondo

Declaration: The above information is correct to the best of my knowledge.
Tushar Dhyani
Date: 05,19,2021