International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

Identification of Deleterious Nssnps of Interleukin-2 (IL-2) Gene and Its Structural Stability Using Computational Methods

Tayyaba Sadaf[1] *, Fatima Darakhshan[1]

[1]*Atta-ur-Rahman School of Applied Biosciences, National University of Sciences and Technology. Islamabad, Pakistan.*

*Corresponding author email: taibsadaf@gmail.com

Second author email: fatimadarakhshan@hotmail.com

Abstract

Non-synonymous SNPs (nsSNP) found in the coding regions of a gene cause variations in amino acid residues and are known to influence the function as well as structure of the encoded proteins. Interleukin 2 (IL-2) is described as a proinflammatory cytokine which plays a significant part to govern the response of immune system. It viably contributes in the pathogenesis of several types of cancers and autoimmune diseases. Taking into account the significance of IL-2, a functional examination by various in silico approaches was performed to investigate the probable effect on phenotypic variation due to genetic mutations. In the current study, out of 1071 SNPs, 42 nsSNPs were collected and scrutinized to categorize the deleterious variants. The functional impact of diverse set of mutations was first annotated and protein instability changes were evaluated. Additionally, the protein movement analysis and flexibility of protein at residual level was identified. Secondly, in order to gain insight in protein structural analysis, the IL-2 protein structure was generated; refined modelled structure was then minimized and further assessed by different tools. The conserved regions, solvent accessibility of native protein and structural divergence of mutant models were also estimated. As a result, our study helped to screen the most pathogenic and highly deleterious mutations and suggested that these substitutions may alter the structure–function relationship. It was inferred that the selected nsSNPs can be a vital candidate for diseases caused by the IL-2 gene and other pathological conditions.

Keywords

IL-2; single nucleotide polymorphisms; protein modelling; computational analysis; deleterious variations

1. Introduction

Cytokines represent diverse set of proteins and polypeptides that act as signalling molecules performing various functions in the immune system. Cytokines comprise interleukins, interferons, colony-stimulating factors, and a variety of growth factors. The imbalance in the secretion of these cytokines and their resulting signalling systems is an essential factor of the pathogenesis of autoimmune diseases.

Human Interleukin-2 (IL-2), described as T-cell growth factor, is a pro-inflammatory cytokine produced by T-cells in response to antigenic or mitogenic stimulation which is vital for T-cell proliferation and several other functions important for the regulation of the immune response (Figure 1) (Shen et al., 2012). IL-2 efficiently contributes in the activation of T cells to produce the cytokines interferon gamma and tumour necrosis factor alpha, augments the activity of natural killer (NK) cells and stimulates lymphokine-activated killer cell activity (Lenardo, 1991; Sakaguchi et al., 2008). Activated T cells result in expression of IL-2 receptors (IL-2R). The assembly of IL-2 to these receptors stimulates the cell division. However, when the T cells are no more activated by antigen, the IL-2R availability will ultimately decline, thus terminating the proliferative stage (Figure1).

Moreover, IL-2 effects cell differentiation, cell survival and the development of immune memory cells (Gaffen & Liu, 2004) and helps to control the

negative immune activation (D'Souza & Lefrançois, 2003).

The human IL-2 gene is found on chromosome 4q26-q28 consisting of four exons separated by three introns (Fujita et al., 1983). Several genetic variations in IL-2 are known and biomedical scientists have discussed their association with susceptibility to a range of conditions including breast cancer (Hu et al., 2013), gastric atrophy (Togawa et al., 2005), rheumatoid arthritis (Fedetz et al., 2003), hepatocellular carcinoma (HCC), oesophageal squamous cell carcinoma and multiple sclerosis (Sayad & Movafagh, 2014). The detailed investigation of polymorphisms linked to IL-2 gene expose their function in regulating the rate of inducible expression and secretion of IL-2 (Williams et al., 1988).

IL-2 polymorphisms recognized in the promoter region (-330T/G, rs2069762) and exon 1 region (+114T/G, rs2069763) have been found to be associated with chronic neuroinflammatory and autoimmune disorder i.e. multiple sclerosis in Japanese and Iranian population (Sayad &Movafagh, 2014). Similarly, these two mutations have also been assessed in breast cancer patients suggesting a positive association with increased susceptibility to breast cancer (Hu et al., 2013). In Taiwan, IL-2 gene polymorphism (G/T, rs2069763) has also been related with a multisystemic disorder of autoimmune disease known as systemic lupus erythematosus (SLE) (Lin et al., 2008).

The investigation of these variations i.e. single nucleotide polymorphisms (SNPs) and their association with human diseases provide new insights for the identification of genetic indicators for diagnosis and prognosis and perhaps novel medicinal approaches for therapeutic purpose. In the light of importance of SNPs to pathogenesis of autoimmune diseases mentioned above, this study is designed to evaluate major disease associated genetic variations using computational approach. For this purpose, different computational tools are applied to investigate SNPs in human IL-2 gene and their consequence on the structure, function and stability of its respective protein.

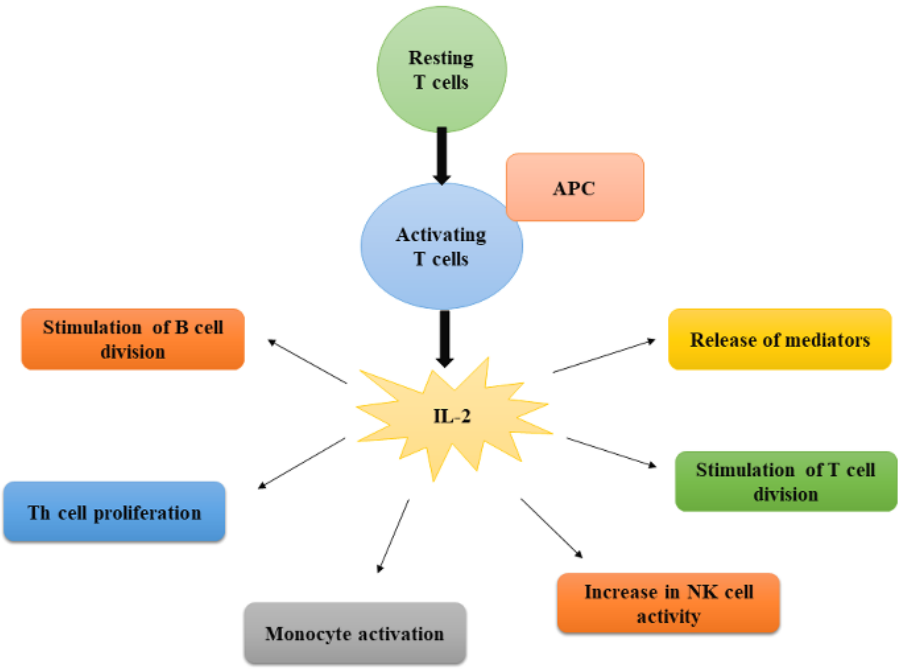

Figure 1. Schematic diagram of Immunoregulatory behaviour of interleukin-2.

2. Material and Methods

2.1. Data collection

Computational analysis was carried out for expatiated sequence and structure based assessment for IL-2 non-synonymous SNPs (nsSNPs). For this purpose, the human IL-2 sequence and SNPs information was retrieved from National Centre for Biotechnology Information (NCBI). Out of human 1071 SNPs, only 42 SNPs were non-synonymous. Hence, nsSNPs corresponding to transcript NCBI reference sequence NM_000586.3 were mined from dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). The amino acid sequence of human IL-2 (NCBI reference sequence NP_000577.2) in FASTA

format was collected from NCBI protein. Ensembl(Aken et al., 2016), Online Mendelian Inheritance in Man (OMIM: 605384) (McKusick, 2016) and RCSB Protein Data Bank (Berman et al., 2006) were used to obtain the associated data for our computational study.

2.2. Softwares used for SNPs annotations

The functional impact of nsSNPs were assessed using different computational tools including: SIFT (http://sift.jcvi.org) (Kumar et al., 2009; Ng & Henikoff, 2006), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2013), SNAP-2 (https://rostlab.org/services/snap2web/) (Bromberg et al., 2008) and PROVEAN (http:// provean.jcvi.org/index.php) (Choi & Chan, 2015).

The program SIFT (Sorting Intolerant From Tolerant) was used to characterize intolerant from tolerant amino acid substitutions and to predict the effect of amino acid substitution on the function of protein. The SIFT scores vary from 0 to 1. The scores less than 0.05 indicates amino acid substitutions are 'damaging' or 'not tolerated', while scores higher than 0.05 are thought to be 'tolerated' (Ng & Henikoff, 2003).

PolyPhen-2 was used for further screening of detrimental SNPs from benign. This server uses a naïve Bayesian classifier to evaluate the effect of amino acid substitution on the structure and function based characteristics of protein. The tool estimates the position-specific independent count (PSIC) for each variant and computes the score difference between variants. The PSIC score ranges from 0-1 and categorises the mutations into three groups as benign, possibly damaging or probably damaging. The nsSNPs with score closer to one, is ranked as damaging (Adzhubei et al., 2013).

SNAP-2, a neural network based program, was then implemented to estimate the functional effects of nsSNPs. SNAP-2 employs the information of automatically generated multiple sequence alignment and various structural features of protein i.e. solvent accessibility and predicted secondary structure etc. to make predictions if any variation is expected to change protein function. This tool takes the input as FASTA format of protein sequences. The output provides a score ranging from −100 (strong neutral prediction) to +100 (strong effect prediction) and expected accuracy, thus classifying the nsSNPs as 'Effect' or 'neutral' (Bromberg et al., 2008).

Biological functional changes of a protein due to a variant were then processed by a web based tool PROVEAN (Protein Variant Effect Analyzer) that operates on sequence clustering and alignment based scoring. The score is calculated based on the variation in sequence similarity of a query sequence to a protein sequence homolog between without and with an amino acid variation of the query sequence. The protein variant is categorized as deleterious if the prediction score is not as much as the cutoff value i.e. - 2.5 (Choi et al., 2012).

2.3. Predicting effect of nsSNPs on protein stability

nsSNPs leading to protein instability upon single amino acid substitution was evaluated by a support vector machine (SVM) based I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html)) tool. The free energy change value (DDG) is calculated by subtracting the unfolding Gibbs free energy value of mutated protein from wild protein (kcal/mol).The output indicates the decreased stability if DDG value is less than 0, conversely, DDG value greater than 0 shows an increase in protein stability i.e. a positive DDG value indicates that the mutated protein possesses high stability and vice versa (Capriotti et al., 2005).

In addition, Mupro server (http://mupro. proteomics.ics.uci.edu/) established on SVM and neural network algorithms was used to find out the effect of nsSNPs on protein stability. The server uses the primary amino acid and substituted amino acid and the protein sequences with mutation location as the input. The output includes the change in energy value and a confidence score between -1 and 1 indicating the confidence of the predicted value. The output score less than 0 shows decrease in protein stability due to the mutation; contrariwise, a score greater than 0 refers an increase in protein stability (Cheng et al., 2006).

2.4. Modelling and Validation of protein structure

The amino acid substitutions can remarkably alter protein structure and function. In this context, acquaintance of a protein's 3D structure is crucial for better insight of protein functionality. Therefore, IL-2 structure was modelled using I-TASSER (Yang et al., 2015) available at: https://zhanglab.ccmb.med.umich.edu/I-TASSER/ by specifying the template 1M47. Alignment was generated using MUSTER threading program. The model was refined by Galaxy refine web (http://galaxy.seoklab.org/refine) (Heo et al., 2013; Ko et al., 2012) which is based on an effectively established refinement method i.e. CASP10(Heo et al., 2013). The Swiss-PdbViewer(Guex et al.,

2000) was then applied for the energy minimization of refine modelled structure. Further, the structure was validated through MolProbity (http://molprobity.biochem.duke.edu/) (Chen et al., 2010; Davis et al., 2007; Davis et al., 2004) and ERRAT (https://services.mbi.ucla.edu/SAVES/).
The Swiss-PDBviewer(Guex et al., 2000) was utilized to perform amino acid substitutions, followed by the energy minimization of mutant modelled structures. GROMOS96 43B1 implementation of Swiss-PdbViewer 4.1.0 (Guex et al., 2000) was used to compute the force field energy of wild model including all bonds, torsions, angles, improper, electrostatic and non-bonded atoms (Johansson et al., 2012).

2.5. Exploring the flexibility at residual level

The protein movement analysis, its binding efficiency and several types of interactions can be studied by identifying the flexibility of certain amino acids in protein. Therefore, the potential structural changes based on the B-factor distribution of the residues were estimated by ResQserver (https://zhanglab.ccmb.med.umich.edu/ResQ/).
ResQ is a method for approximating residue specific quality of protein structure prediction and the inherent B-factor profile of all residues along the chain by combining local structure assembly variations with sequence- and structure-based profiling(Yang et al., 2016).

2.6. Detection of functional SNPs in conserved Regions

The evolutionary conservation of amino acid substitution in proteins based on the phylogenetic relations between homologous sequences was analysed by ConSurf (http://consurf.tau.ac.il/2016/) web tool that uses an empirical Bayesian inference. The conserved regions were assessed with the help of conservation scores and colouring grading system by submitting the FASTA sequence of IL-2 gene as an input option.

2.7. Analysis of protein structural divergence

In order to compute changes resulting from the nsSNPs, the wild and mutated predicted structures were then evaluated using a range of structure assessment tools.

The divergence of the mutant structure from the native structure is expected due to the amino acid substitution. This difference between two structures can change the functional activity with respect to binding efficiency of the molecules. Structural divergence due to missense mutations in IL-2 protein was scrutinized by computing root mean square deviation (RMSD) values of native and mutant structures by SuperPose (http://wishart.biology.ualberta.ca/SuperPose/). The output of SuperPose tool gives a diversity of RMSD values for heavy atoms, alpha carbons, backbone atoms and all atoms (average and pairwise) to find individual residue shifts (Maiti et al., 2004).

2.8. Analysis of secondary structure and solvent accessibility

The estimation of secondary structure and solvent accessibility of structure was achieved by RaptorX property prediction (Wang et al., 2016a; Wang et al., 2016b), PredictProtein(Rost et al., 2004)) and artificial neural network based program NetSurfP(Petersen et al., 2009). The output of RaptorX gives information about the tertiary structure, 3-state and 8-state secondary structure, disordered regions and solvent accessibility for a protein (Källberg et al., 2012). Additionally, the output of NetSurP server provides two subclasses i.e. buried and exposed; based on the z-score, defined for solvent accessibility of amino acids for both mutant and wild structures of IL-2 Protein.

2.9. Prediction of phylogeny-based protein function

The molecular functions of IL-2 protein sequence was predicted using a statistical method SIFTER (Statistical Inference of Function Through Evolutionary Relationships) that utilizes a protein family's phylogenetic tree, as the natural assembly for expressing protein relationships (Sahraeian et al., 2015).

2.10. Active site identification

The relationship of structure and function of IL-2 protein can be demonstrated by identifying the ligand binding sites of the protein. FTsite server (http://ftsite.bu.edu.) was applied to predict the ligand binding sites of proteins with high accuracy. The tool gives an estimation if the identified nsSNPs are present in the IL-2 protein binding

region or not. A residue of the protein is considered to be a part of binding site if the residue is less than 4 Å from any atom of the probe from probe clusters (Ngan et al., 2011). Additionally, COACH (http://zhanglab.ccmb.med.umich.edu/COACH/), an algorithm-based web server, was applied for further evaluation of the ligand binding sites and active site residues of IL-2 protein (Yang et al., 2013).

2.11. Analysis of protein-protein Interactions

The functional interaction of IL-2 protein was also analyzed using the online database STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) version 10.5 (Szklarczyk et al., 2017). Both the physical (direct) and functional (indirect) associations were identified depending on the known experimental interactions, predicted genomic interactions, text mining, co-expression and protein homology.

3. Results and Discussion

The Single nucleotide polymorphisms (SNPs) are assumed to perform a vital part in various human diseases. Generally, more than 4 million distinctive human SNPs are described by several SNP databanks including; human genome variation database (HGVBase) and the NCBI database dbSNP, providing insights to future research. It is believed that structural modelling of a protein and impact of SNPs on protein is imperative in understanding biological processes of associated diseases at molecular level. Furthermore, nsSNPslead to change in the protein sequence which may modify the protein function. Recently, several in silico studies (Akhoundi et al., 2016; Ali Mohamoud et al., 2014; Chakrapani et al., 2017; D RASAL et al., 2016; Dabhi & Mistry, 2014; Goswami, 2015; Kamaraj & Purohit, 2013; Nailwal & Chauhan, 2017; Naveed et al., 2016; Raghav et al., 2013; Rajasekaran et al., 2008)predicted the importance of nsSNPs associated with genes and provided a platform to ascertain the molecular mechanisms of diseases and influence of variations on the structure and function of protein.

In our study, nonsynonymous SNPs were chosen from coding region for the computational analysis. A total of 42 SNPs of IL-2 gene from NCBI dbSNPwere first functionally investigated using different computational programs, namely, SIFT, PolyPhen-2, PROVEAN and SNAP-2 as summarized in Table 1. These insilico tools estimated whether the SNPs have detrimental or neutral effect on protein function. The computational tools used as a part of this investigation have already been tried and endorsed by several other biomedical scientists as well (Masoodi et al., 2013; Kumar et al., 2015).

SIFT predicted 22 (52.3%) out of 42 nsSNPs as 'damaging' based on the tolerance index score. Among these nsSNPs predicted as damaging, 16 (72.7%) were with score 0.00, 5 (22.7%) were with score 0.01 and only one with 0.03 score. The remaining 20 (47.6%) were categorized as 'tolerated' with score ranging 0.05-0.57 (Table 1 and Figure 2A). The PSIC score calculated by PolyPhen-2 filtered a total of 26 (61.9%) variants as damaging. Among them, 15 (57.6%) nsSNPs were classified as probably damaging and 11 (42.3%) were predicted as possibly damaging. The rest of 16 (38.1%) variants were ranked benign according to PolyPhen-2 as described in Table 1 and Figure 2A. However, 37 (88.1%) nsSNPs were predicted as effective and 05 (11.9%) nsSNPs as neutral based on the score calculated by SNAP-2 as summarized in Table 1 and Figure 2A. The effect of nsSNPs on protein function was further validated by PROVEAN program which classified 26 (61.9%) nsSNPs as deleterious with score < -2.5 whereas 16 (38.1%) nsSNPs were predicted as neutral (Table 1 and Figure 2A).

Table 1. Screening and prediction of deleterious nsSNPs of IL-2 gene by different computational tools

rsID	ChrPos	Variant position		PolyPhen-2		SIFT		PROVEAN		SNAP2	
		Nt	AA	Pred	Score	Effect	Score	Pred	Score	Pred	Score
rs778990655	*122456439*	*2T>C*	*M1T*	*pr dmg*	*0.999*	*dmg*	*0*	*dele*	*-3.524*	*eff*	*69*
rs757619199	122456430	11T>C	M4T	benign	0.005	dmg	0	neu	-2.214	eff	42
rs533458551	122456379	62C>A	A21E	pr dmg	1	dmg	0	dele	-4.026	eff	73
		62C>T	A21V	pr dmg	0.996	dmg	0	dele	-3.261	eff	51

rs752534234	122456380	61G>A	A21T	pr dmg	1.000	dmg	0	dele	-3.203	eff	55
rs754421965	122456376	65C>T	P22L	pr dmg	1	dmg	0	dele	-8.096	eff	75
rs758928770	122456367	74G>T	S25I	pos dmg	0.956	dmg	0	dele	-4.25	eff	42
rs77806995	122456361	80C>A	T27K	pr dmg	1	Tol	0.22	dele	-2.994	eff	30
rs765882548	*122456346*	*95T>C*	*L32P*	*pos dmg*	*0.9*	*dmg*	*0.01*	*neu*	*-1.431*	*eff*	*23*
rs762536978	*122456344*	*97C>A*	*Q33K*	*pos dmg*	*0.82*	*tol*	*0.05*	*dele*	*-2.62*	*eff*	*25*
rs3087209	*122456328*	*113T>G*	*L38R*	*pr dmg*	*0.998*	*dmg*	*0*	*dele*	*-4.882*	*eff*	*68*
rs146566026	122456312	129G>T	M43I	benign	0.011	tol	0.17	neu	-1.445	eff	12
rs988966007	122456308	133T>G	L45V	pos dmg	0.891	dmg	0.01	neu	-2.291	eff	49
rs751468439	122456202	149A>G	N50S	pos dmg	0.903	dmg	0.01	dele	-3.718	eff	27
rs765829053	122456196	155A>G	K52R	pos dmg	0.527	tol	0.22	neu	-1.406	neu	-2
rs202027273	122456191	160C>T	P54S	benign	0.002	tol	0.34	dele	-3.208	eff	2
rs760934382	122456178	173G>T	R58M	benign	0.114	tol	0.24	dele	-3.845	neu	-10
rs767776092	*122456158*	*193T>C*	*Y65H*	*pos dmg*	*0.955*	*tol*	*0.07*	*dele*	*-2.908*	*eff*	*70*
rs994708825	*122456152*	*199C>G*	*P67A*	*pr dmg*	*1*	*dmg*	*0*	*dele*	*-7.188*	*eff*	*67*
rs771093163	122456146	205A>G	K69E	benign	0.000	dmg	0.03	neu	-2.164	eff	37
rs753445824	*122453853*	*208G>A*	*A70T*	*pos dmg*	*0.818*	*dmg*	*0.01*	*dele*	*-2.668*	*eff*	*29*
rs767872192	122453852	209C>T	A70V	benign	0.025	tol	0.14	neu	-2.484	neu	0
rs774269792	*122453849*	*212C>T*	*T71I*	*pr dmg*	*0.96*	*dmg*	*0*	*dele*	*-4.843*	*eff*	*52*
rs759849798	*122453843*	*218T>C*	*L73P*	*pr dmg*	*1.000*	*dmg*	*0*	*dele*	*-6.237*	*eff*	*79*
rs752080371	122453841	220A>C	K74Q	pr dmg	0.999	tol	0.08	neu	-2.46	neu	-3
rs146270985	122453835	226C>T	L76F	pr dmg	1.000	dmg	0	dele	-3.738	eff	70
rs866077447	122453799	262G>A	E88K	pos dmg	0.526	tol	0.06	dele	-2.751	eff	32
rs773490003	*122453786*	*275T>C*	*L92S*	*pr dmg*	*0.994*	*dmg*	*0*	*dele*	*-4.245*	*eff*	*67*
rs997605621	122453765	296A>C	H99P	benign	0.356	tol	0.23	dele	-4.003	eff	10
rs749017354	122453746	315A>C	L105F	benign	0.069	tol	0.42	neu	-0.607	eff	19
		315A>T	L105F	benign	0.069	tol	0.42	neu	-0.607	eff	19
rs746832307	122453745	316A>G	I106V	benign	0.028	tol	0.23	neu	-0.679	eff	22
rs772064424	*122453738*	*323A>G*	*N108S*	*benign*	*0.149*	*dmg*	*0*	*dele*	*-4.752*	*eff*	*42*
rs756737575	122453730	331G>A	V111I	benign	0.001	tol	0.57	neu	-0.703	eff	29
		331G>T	V111L	benign	0	tol	0.32	neu	-2.199	eff	47
rs778851120	122451852	362C>T	T121I	pr dmg	1	dmg	0	dele	-4.469	eff	52
rs1000418138	122451843	371T>C	M124T	benign	0	tol	1	neu	-0.688	neu	-2
rs377374894	*122451829*	*385G>C*	*D129H*	*pos dmg*	*0.89*	*dmg*	*0*	*dele*	*-5.239*	*eff*	*12*

rs865810716	122451819	395C>A	A132E	pr dmg	0.984	tol	0.08	neu	-2.468	eff	30
rs777464920	122451810	404T>C	V135A	benign	0.007	tol	0.08	neu	-2.159	eff	23
rs543296992	122451796	418A>G	R140G	benign	0.000	dmg	0.01	dele	-3.142	eff	56
rs374465594	*122451789*	*425T>C*	*I142T*	*pos dmg*	*0.709*	*tol*	*0.06*	*dele*	*-4.055*	*eff*	*69*

Chrpos= chromosome position; Nt= nucleotide; Pred= prediction; AA= amino acid; Pr dmg= probably damaging; pos dmg= possibly damaging; tol= tolerated; neu= neutral; dele= deleterious; eff= effect.

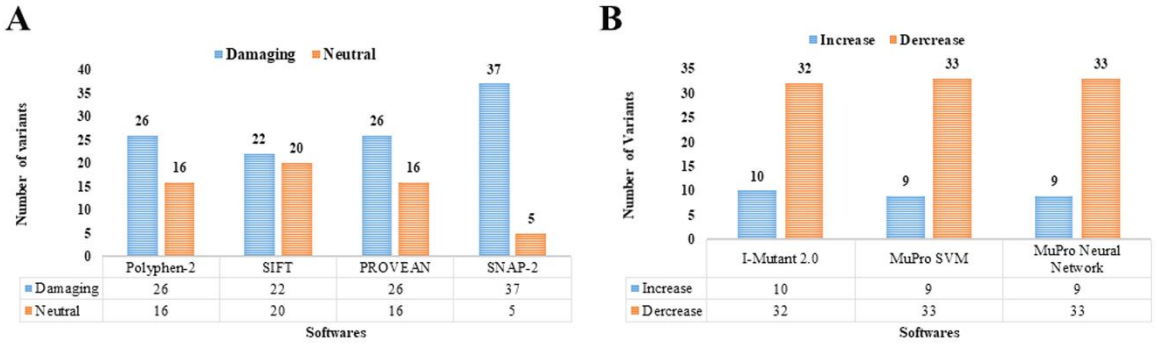

Figure 2. Bar chart representation of annotated IL-2 SNPs. A) Number of deleterious and tolerated nsSNPs predicted by various softwares. B) Number of screened nsSNPs based on stability of protein predicted through I- mutant and Mupro.

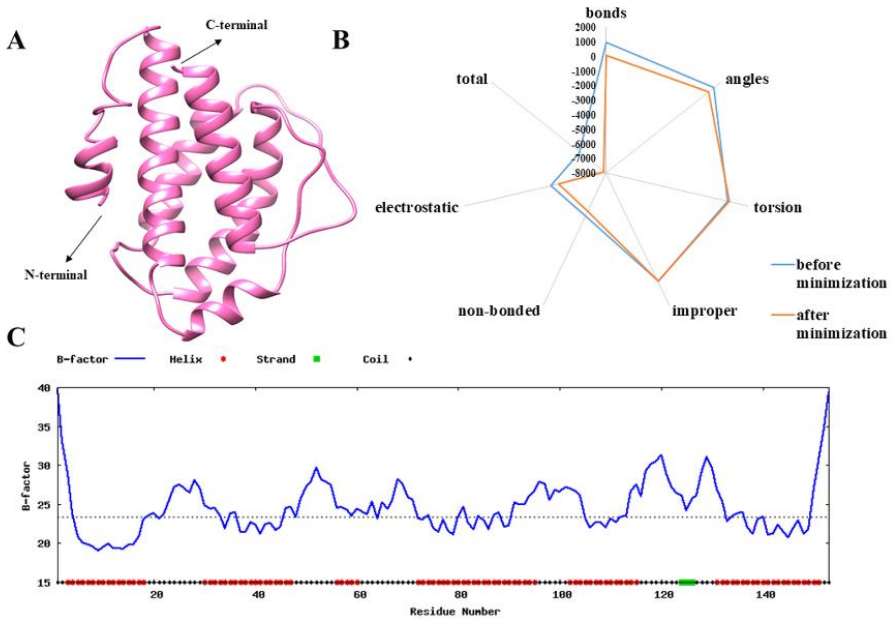

Figure 3. A) The 3D modelled structure of IL-2 protein predicted by I-TASSER. B) Energy parametric profile of IL-2 protein before and after minimization. C) B-factor profile for IL-2 protein sequence predicted by ResQ.

Table 2: Prediction of effect of nsSNPs on IL-2 protein stability through I-Mutant 2 and MuPro

Variants	I-Mutant 2	MuPro

			SVM		Neural Network	
	Effect	**DDG**	**Effect**	**Confidence Score**	**Effect**	**Confidence Score**
Met1Thr	*decrease*	*-0.85*	*decrease*	*-1*	*decrease*	*-0.9*
Met4Thr	decrease	-1.08	decrease	-0.7	decrease	-0.8
Ala21Glu	decrease	-0.22	decrease	-1	decrease	-0.9
Ala21Val	increase	0.26	decrease	-0.8	decrease	-0.9
Ala21Thr	decrease	-0.52	decrease	-1	decrease	-0.9
Pro22Leu	increase	-1.07	decrease	-0.1	increase	0.6
Ser25Ile	increase	0.68	increase	0.1	increase	0.7
Thr27Lys	decrease	-1.5	decrease	-1	decrease	-0.9
Leu32Pro	*increase*	*0.79*	*decrease*	*-0.9*	*decrease*	*-0.9*
Gln33Lys	*increase*	*0.17*	*increase*	*0*	*decrease*	*-0.6*
Leu38Arg	*decrease*	*-0.59*	*decrease*	*-0.1*	*decrease*	*-0.7*
Met43Ile	decrease	-0.37	decrease	-0.1	increase	0.5
Leu45Val	decrease	-0.8	decrease	-1	decrease	-0.9
Asn50Ser	decrease	-0.08	decrease	-1	decrease	-0.9
Lys52Arg	decrease	-0.09	decrease	-0.3	decrease	-0.6
Pro54Ser	decrease	-1.17	decrease	-1	decrease	-0.9
Arg58Met	decrease	-1.31	increase	0.4	increase	0.8
Tyr65His	*decrease*	*-0.98*	*decrease*	*-1*	*decrease*	*-0.9*
Pro67Ala	*decrease*	*-0.147*	*decrease*	*-1*	*decrease*	*-0.9*
Lys69Glu	increase	0.42	decrease	-0.2	decrease	-0.6
Ala70Thr	*decrease*	*-0.59*	*increase*	*0.7*	*decrease*	*-0.5*
Ala70Val	decrease	-0.2	increase	1	increase	0.7
Thr71Ile	*decrease*	*-0.6*	*decrease*	*-0.3*	*increase*	*0.5*
Leu73Pro	*decrease*	*-1.2*	*decrease*	*-1*	*decrease*	*-0.9*

Lys74Gln	decrease	-0.38	decrease	-0.8	decrease	-0.9
Leu76Phe	increase	0.63	decrease	-0.9	decrease	-0.9
Glu88Lys	decrease	-0.49	decrease	-0.4	increase	0.5
Leu92Ser	*decrease*	*-2.33*	*decrease*	*-1*	*decrease*	*-0.9*
His99Pro	decrease	-0.07	increase	0.03	increase	0.5
Leu105Phe	decrease	-0.29	decrease	-1	decrease	-0.8
Ile106Val	decrease	-0.54	decrease	-0.9	decrease	-0.8
Asn108Ser	*increase*	*0.35*	*increase*	*0.09*	*decrease*	*-0.6*
Val111Ile	decrease	-0.52	decrease	-0.7	decrease	-0.8
Val111Leu	decrease	-1.4	decrease	-0.4	decrease	-0.8
Thr121Ile	increase	0.46	increase	0.1	decrease	-0.5
Met124Thr	decrease	-0.59	decrease	-0.2	decrease	-0.7
Asp129His	*increase*	*1.27*	*decrease*	*-0.4*	*decrease*	*-0.5*
Ala132Glu	decrease	-1.25	increase	0.5	increase	0.5
Val135Ala	decrease	-1.29	decrease	-1	decrease	-0.9
Arg140Gly	decrease	-0.62	decrease	-1	decrease	-0.8
Ile142Thr	*decrease*	*-2.42*	*decrease*	*-1*	*decrease*	*-0.7*

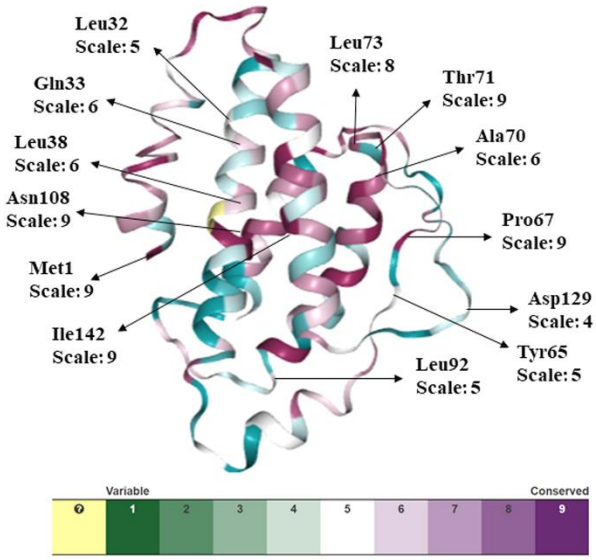

Figure 4. Conservation analysis of IL-2 protein by ConSurf. The color coding bar shows conservation score.

Interestingly, it was noted that 18 nsSNPs estimated as damaging by SIFT, were also ranked as deleterious by PolyPhen-2. This observation provides an obvious sign of a strong relation between evolutionary based and the structural based approaches i.e. SIFT and PolyPhen-2 respectively. Furthermore, in this analysis, we have identified and predicted 13 nsSNPs that were found to be common and functionally significant (deleterious) by all four computational tools.

3.1. Predicting effect of nsSNPs on protein stability

The influence of single amino acid substitution on the stability of protein structure and function was evaluated by I-Mutant 2.0 and MuPro servers. The results of stability changes of 42 nsSNPs are demonstrated in Table 2 and Figure 2B which are projected to be either amplify or reduce the free energy change upon amino acid substitutions. I-Mutant 2.0 revealed that 32 (76.2%) out of 42 variants causes decrease in protein stability with DDG value less than zero. Similarly, MuPro predicted protein stability through two algorithms: SVM and Neural network. Both algorithms predicted 33 (78.5%) variants with the confidence score less than zero, thus revealing decrease in the IL-2 protein stability (Table 2 and Figure 2B). This destabilizing effect in majority of the deleterious mutations gives an indication about the disturbance in the structure and function of IL-2 protein because of nsSNPs.

3.2. Modelling and Validation of protein Structure

The 3D structure of protein is extremely convenient to assess the impact of nsSNPs on the whole protein structure (Kumar et al., 2009; Ng & Henikoff, 2006). Therefore, the 3D structures of IL-2 protein was generated through I-TASSER. I-TASSER predicted 5 models of IL-2. The generated models were compared providing three discrete values: C-Score (confidence score), Root-mean-square deviation (RMSD) and TM-score (Roy et al., 2010). C-score normally varies between -5 and 2. The higher C-score value reflects the high quality i.e. more reliability of the model (Roy et al., 2012). Accurate models depict RMSD values less than 2.0Å, and TM-scores tending to 1. We selected the best model having C-Score equals to -0.18, estimated TM-score was 0.69±0.12 and estimated RMSD was 5.2±3.3Å (Figure 3A).
Galaxy Refine server was then used to explore the 3D structure further. The assessment of the 3D models created by I-TASSER and Galaxy refine tool is given in Table 3. The output indicates the improved values of RMSD, Clash score, Poor rotamers and Ramachandran favoured, enlightening the highly refined 3D structure of IL-2.

Thepromising energy minimization techniques can be employed in the domain of crystallography, structure prediction and molecular dynamics simulation for the refinement of the protein structure (Moult et al., 2014). In this context, we performed energy minimization for the selected model of IL-2 protein. Minimized model of IL-2 had lower optimized energy as indicating the higher stability (Figure 3A).
The model was then validated after energy minimization by submitting to the verification tools ERRAT and Molprobity. The Ramachandran analysis using MolProbity justified that most of the residues (84-93%) exist in favourable regions (Table 4). Subsequent outlier removal and poor rotamer correction improved the predicted models. The output values obtained from ERRAT for I-TASSER, Galaxy refined and minimized model were 93.7%, 80.7% and 82%, respectively, revealing the overall good quality of these models.

3.3. Exploring the flexibility at residual level

The flexibility in IL-2 protein model can also be evaluated through the analysis of the B-factor. B-factor demonstrates the inherent thermal mobility of amino acids or atoms in proteins. Therefore, we estimated B-factor by ResQ server. The output (Figure 3C) showed that the areas at the N- and C-terminals and majority of the loop regions have higher B-factor values demonstrating that these regions are structurally more flexible than other regions. Besides, the B-factors calculated for alpha regions are lower, proposing these regions are structurally more stable.

3.4. Detection of functional SNPs in conserved Regions

The selected single nucleotide polymorphisms were examined for conservation analysis to predict the conservation rate. Figure 4. represents the ConSurf results that comprises the color scale based on the conservation scores (9 - conserved, 1 - variable), demonstrating the evolutionary relationships among sequence homologs. The output of the ConSurf tool also gives the score that is the normalized conservation scores. It is noticed from data that among the selected residual positions, the positions 1, 67, 71, 73, 108 and 142 were found to be highly conserved and functional with conservation scale value ranging 8-9 (Figure 4). From these results, we can infer that the mutations at the 1, 67, 71, 73, 108 and 142 positions can have great functional and structural impact on the IL-2 protein due to their high conservation frequency.

3.5. Analysis of protein structural divergence

The accurate RMSD score for all the variants were determined by SuperPose tool as mentioned in Figure 5. The greater the RMSD value is, the higher the variation between the two structures is, which thus modifies their functional performance (Reva et al., 1998). It is noted that the overall RMSD values of alpha carbons, backbone and heavy atoms of modelled structures are all resembling suggesting that they are structurally very similar and might cause a slight change in the mutant structures.

3.6. Analysis of secondary structure and solvent accessibility

The biophysical analysis including secondary structure and solvent accessibility of IL-2 protein was first predicted using RaptorX tool as shown in Figure 6A-B. The 8-class secondary structure of IL-2 precited by RaptorX consist of: 57% helix, 3% extended strand and 39% coil region. The solvent accessed is expected to contain 24% Buried, 51% Medium and 23% Exposed regions.
The solvent accessibility of IL-2 protein when evaluated by PredictProteinFigure 6C, the results revealed the presence of alpha-helices predominantly which is in consonance with the results predicted by RaptorX.
Moreover, the biophysical analysis including surface and solvent accessibility of IL-2 protein achieved through NetSurfP suggested that the stability of the protein can be disturbed due to the position and the nature of a mutated residue. It has been reported that as the solvent accessibility of a residue lessens, it destabilizes the protein. The output generated by NetSurP (Table 5) demonstrates that there is no change in the class assignment for any of 13 SNPs due to mutation. 07 out of 13 of mutant and native residues were found in buried regions that infer their low accessibility to surface and solvent. The mutant residue exhibited the changes in relative surface accessibility (RSA),absolute surface accessibility and Z-fit score for RSA prediction (Table 5).

Table 3. Comparison of models generated by I-TASSER and Galaxy refine tools.

Models	GDT-HA	RMSD	MolProbity	Clash Score	Poor rotamers	Ramachandran favoured
I-TASSER	1.0000	0.000	3.159	13.3	13.2	84.1
Galaxy Refine	0.9542	0.425	2.342	26.5	0.7	93.4

RMSD= root mean square deviation

Table 4. IL-2 protein structure validation through Molprobity and ERRAT

Models	Rotamers		Ramachandran		Cβ deviations	ERRAT
	Poor	Favored	Outliers	Favored		
I-TASSER	22 (15.28%)	115 (79.86%)	10 (6.62%)	127 (84.11%)	11	93.793
Galaxy Refine model	2 (1.39%)	140 (97.22%)	2 (1.32%)	141 (93.38%)	15	80.690
Minimized	2 (1.49%)	132 (98.51%)	1 (0.66%)	137 (90.73%)	5	82.069

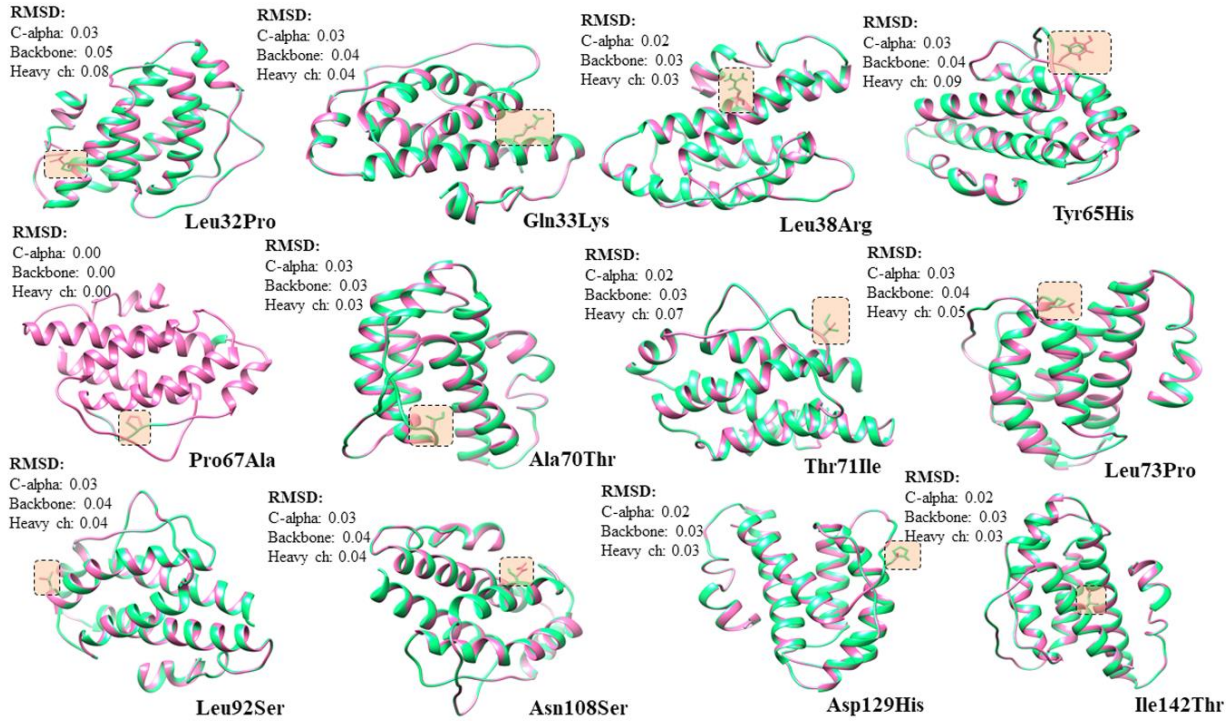

Figure 5. RMSD values of IL-2 mutant model protein calculated by SuperPose.

Table 5. Surface accessibility of wild type and mutants of IL-2 protein.

Variants	Class assignment	NetSurfP

		Relative Surface Accessibility (RSA)	Absolute Surface Accessibility (ASA)	Z-fit score for RSA prediction
Met1Thr	Exposed	0.784	156.958	-0.108
	Exposed	0.860	119.227	0.693
Leu32Pro	Exposed	0.548	100.339	1.141
	Exposed	0.576	81.777	1.199
Gln33Lys	Exposed	0.311	55.455	0.887
	Exposed	0.335	68.930	1.019
Leu38Arg	Buried	0.120	22.027	0.282
	Buried	0.150	34.396	0.527
Tyr65His	Buried	0.267	57.143	-0.377
	Buried	0.273	49.677	-0.364
Pro67Ala	Buried	0.286	40.555	-0.519
	Buried	0.295	32.465	-0.504
Ala70Thr	Buried	0.166	18.271	-0.933
	Buried	0.182	25.243	-0.785
Thr71Ile	Exposed	0.428	59.294	-0.424
	Exposed	0.399	73.852	-0.442
Leu73Pro	Buried	0.136	24.938	0.667
	Buried	0.162	23.045	0.139
Leu92Ser	Buried	0.114	20.928	0.463
	Buried	0.123	14.439	0.266
Asn108Ser	Exposed	0.321	46.965	-0.460
	Exposed	0.318	37.328	-0.418
Asp129His	Exposed	0.600	86.417	-0.199
	Exposed	0.600	109.213	-0.255
Ile142Thr	Buried	0.085	15.762	0.715
	Buried	0.090	12.469	0.718

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 04
February 2018

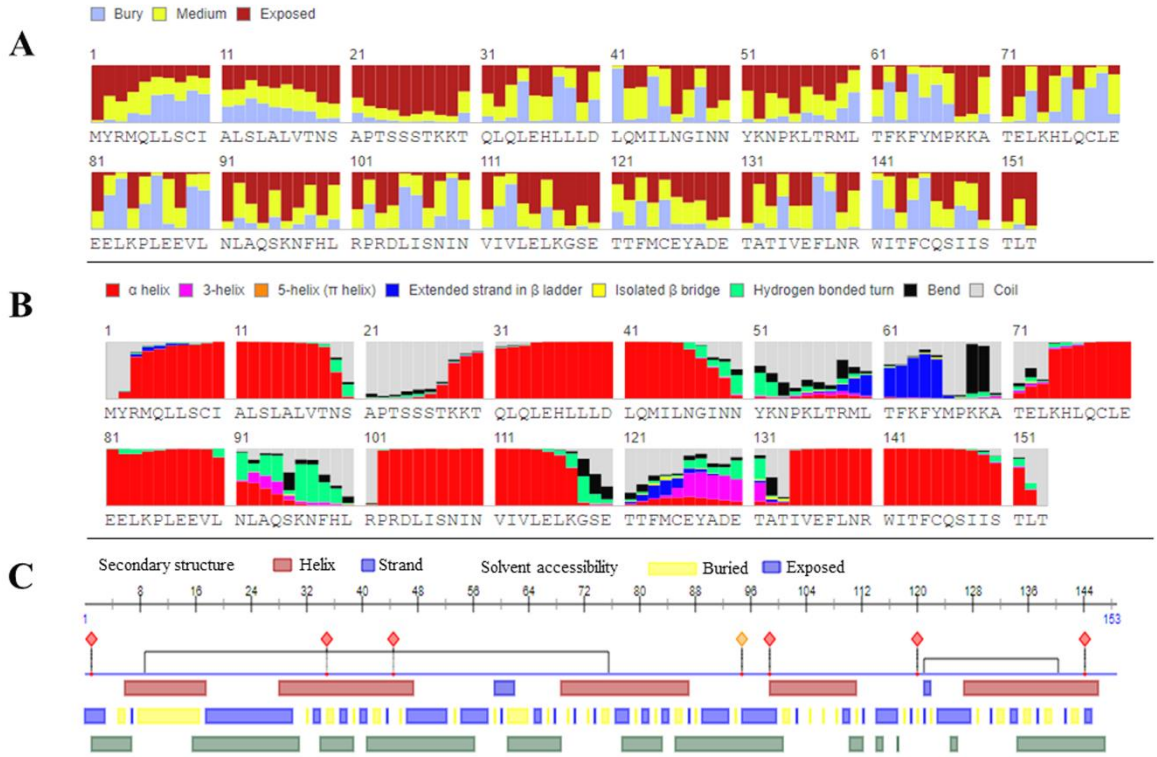

Figure 6. Secondary structure and solvent accessibility of native IL-2 protein: A) Solvent accessibility prediction of IL-2 protein by RaptorX B) 8-class secondary structure prediction by RaptorX; C) Secondary structure prediction by PredictProtein

Table 6. Molecular function predictions for the entire human IL-2 protein by SIFTER.

GO term	Protein Prediction	Confidence score
GO:0005125	Cytokine activity	0.81
GO: 0005134	Interleukin-2 receptor binding	0.81
GO: 0019209	Kinase activator activity	0.75
GO: 0008083	Growth factor activity	0.75

Table 7. Residues at ligand binding sites of IL-2 protein predicted by FTSite.

Binding Sites	**AminoAcid Residues**
FTSite 1:	Met66, **Pro67**, Lys68, Lys69, **Ala70**, Arg140, Trp141, Thr143, Phe144, Ser147
FTSite 2:	Lys29, **Leu32, Gln33,** His36, Leu37, **Asn108**, Ile109, Val111, Ile112, Glu115

FTSite 3:	Leu6, Cys9, Ile10, Val17, Glu35, **Leu38**, Gln42, Gln146, Ile149, Ser150, Thr153

3.7. Prediction of phylogeny-based protein function

The IL-2 protein sequence subjected for the phylogeny-based protein function prediction reveals the presence of different domains in IL-2 protein structure, highlighting the multiple functions of IL-2 protein that includes Cytokine activity, binding to Interleukin-2 receptor, Kinase activator activity and Growth factor activity as stated in Table 6. All together, we can say that human IL-2 act as a growth factor and help to regulate different signalling systems that are essential to understand the pathogenesis of associated diseases.

3.8. Active site identification

The identification of active sites is very important, any mutation in these regions may disrupt the binding capacity of the ligand to its protein. Thus, change in ligand binding site affects the function of protein by changing the stability of protein (Ng & Henikoff, 2006; Yang et al., 2012).
In our analysis, FTSite detected 3 ligand binding sites on IL-2 Protein. The amino acid residues observed in these 3 sites of IL-2 protein are presented in Table 7. It is observed that 6 amino acid residues i.e. Leu32, Gln33, Leu38, Pro67, Ala70 and Asn108 out of 13 selected deleterious nsSNPs are identified among these sites.

Moreover, ligand binding residues were also identified by COACH algorithm that employs various ligands to examine their binding sites. The ligands with greater C-score (confidence score) signify an authentic result. Table 8 indicates that the ligand FRH has better C-score with 0.46 value. Only two of these sites i.e. Tyr65His and Leu92Ser matches the deleterious SNPs selected for IL-2 protein. However, three sites Leu32Pro, Gln33Lys and Asn108Ser that also act as the deleterious SNPs, were found having the binding affinity for 2JY ligand but with low C-score.

3.9. Analysis of protein-protein Interactions

The analysis of protein–protein interaction is crucial to reveal all functional interactions among cellular proteins. The STRING database exhibited 10 functional partners of IL-2 (Figure 7) with highest interaction confidence score > 0.90 indicating strong association with IL-2 protein. Predicted interaction network with different parameters and filters observed with IL-2 protein has demonstrated IL2R as the strongest interaction partner (Figure 7).
All the study reflects that the selected nsSNPs of IL-2 gene might influence the function of protein and might be potent drug target for the treatment of diseases associated with IL-2 gene polymorphisms.

Table 8. Function prediction of IL-2 modelled protein using COACH server.

Rank	C-score	Cluster size	PDB hit	Lig name	Binding residues
1	0.46	16	1py2A	FRH	54,55,58,59,61,62,63,64,**65**,82,85,89,**92**,93,131
2	0.07	3	4nejA	2K1	62,63,64,85
3	0.07	2	1qvnC	FRI	49,56,60,133,136
4	0.06	2	4nemA	2JY	**32**,**33**,36,**108**,111,112
5	0.06	2	1nbpA	MHC	47,51,55,90,93,94,96,102,105

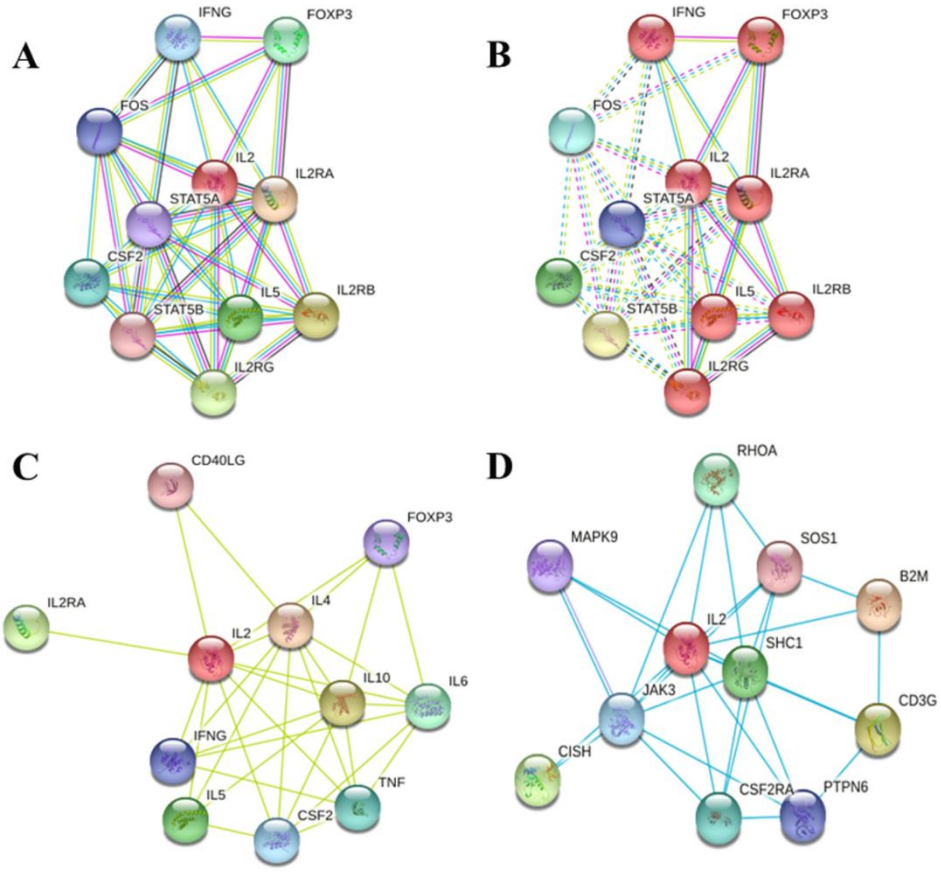

Figure 7. Protein–protein interaction network of IL-2 using STRING 10.5 server: A) top 10 interactor with highest confidence score without clustering, B) K means clustering with 5 number of clusters, C) active interaction source is text mining, D) active interaction sources is databases.

4. Conclusion

The study was focused to thoroughly investigate the plausible effects of variations on the structure and function of IL-2 protein. Therefore, the complete 3D structure of IL-2 was primarily modelled and validated. Furthermore, both sequence based and structure based methods were implemented for evaluation of non-synonymous single nucleotide polymorphisms (nsSNPs). Out of 42 nsSNPs, 12 nsSNPs namely p.Leu32Pro, p.Gln33Lys, p.Leu38Arg, p.Tyr65His, p.Pro67Ala, p.Ala70Thr, p.Thr71Ile, p.Leu73Pro, p.Leu92Ser, p.Asn108Ser, p.Asp129His and p.Ile142Thr were found pathogenic and highly deleterious and might induce alterations in protein structure, stability, solvent accessibility and binding affinity. The computational analysis of free energy change indicates the decrease in IL-2 protein stability due to mutations. Consequently, our study recommends that the selected nsSNPs of IL-2 gene can be viewed as significant candidates in developing diseases related to IL-2 gene polymorphisms.

Conflict of interest None.

References

[1] Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 7-20.

[2] Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., et al. (2016). Ensembl 2017. *Nucleic acids research*, *45*(D1), D635-D642.

[3] Akhoundi, F., Parvaneh, N., & Modjtaba, E.-B. (2016). In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene. *Meta gene*, *9*, 142-150.

[4] Ali Mohamoud, H. S., Manwar Hussain, M. R., El-Harouni, A. A., Shaik, N. A., Qasmi, Z. U., Merican, A. F., et al. (2014). First comprehensive in silico analysis of the functional and structural consequences of SNPs in human GalNAc-T1 gene. *Computational and mathematical methods in medicine*, *2014*.

[5] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2006). The Protein Data Bank, 1999–*International Tables for Crystallography Volume F: Crystallography of biological macromolecules* (pp. 675-684): Springer.

[6] Bromberg, Y., Yachdav, G., & Rost, B. (2008). SNAP predicts effect of mutations on protein function. *Bioinformatics*, *24*(20), 2397-2398.

[7] Capriotti, E., Fariselli, P., Calabrese, R., & Casadio, R. (2005). Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, *21*(suppl_2), ii54-ii58.

[8] Chakrapani, V., Rasal, K. D., Kumar, S., Mohapatra, S. D., Sundaray, J. K., Jayasankar, P., et al. (2017). In Silico Analysis of nsSNPs of Carp TLR22 Gene Affecting its Binding Ability with Poly I: C. *Interdisciplinary Sciences: Computational Life Sciences*, 1-12.

[9] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12-21.

[10] Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*, *62*(4), 1125-1132.

[11] Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745-2747.

[12] Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PloS one*, *7*(10), e46688.

[13] D RASAL, K., Chakrapani, V., PATRA, S. K., Jena, S., D MOHAPATRA, S., Nayak, S., et al. (2016). Identification and prediction of the consequences of nonsynonymous SNPs in glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene of zebrafish Danio rerio. *Turkish Journal of Biology*, *40*(1), 43-54.

[14] D'Souza, W. N., & Lefrançois, L. (2003). IL-2 is not required for the initiation of CD8 T cell cycling but sustains expansion. *The Journal of Immunology*, *171*(11), 5727-5735.

[15] Dabhi, B., & Mistry, K. N. (2014). In silico analysis of single nucleotide polymorphism (SNP) in human TNF-α gene. *Meta gene*, *2*, 586-595.

[16] Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic

acids. *Nucleic acids research*, *35*(suppl_2), W375-W383.
[17] Davis, I. W., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic acids research*, *32*(suppl_2), W615-W619.
[18] Fedetz, M., Matesanz, F., Cáliz, R., Ferrer, M. A., Collado, M. D., Alcina, A., et al. (2003). Lack of association between-384 and 114 IL-2 gene polymorphisms and rheumatoid arthritis. *The Journal of rheumatology*, *30*(3), 435-437.
[19] Fujita, T., Takaoka, C., Matsui, H., & Taniguchi, T. (1983). Structure of the human interleukin 2 gene. *Proceedings of the National Academy of Sciences*, *80*(24), 7437-7441.
[20] Gaffen, S. L., & Liu, K. D. (2004). Overview of interleukin-2 function, production and clinical applications. *Cytokine*, *28*(3), 109-123.
[21] Goswami, A. M. (2015). Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2. *Meta gene*, *5*, 162-172.
[22] Guex, N., Diemand, A., Peitsch, M., & Schwede, T. (2000). SwissPDBViewer program. *Glaxo Smith Kline R. &D*.
[23] Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic acids research*, *41*(W1), W384-W388.
[24] Hu, X.-B., Ouyang, L.-Z., & Tang, L.-L. (2013). Interleukin-2 gene polymorphisms and prognosis of breast cancer. *Genetic testing and molecular biomarkers*, *17*(6), 453-457.
[25] Johansson, M. U., Zoete, V., Michielin, O., & Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC bioinformatics*, *13*(1), 173.
[26] Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., et al. (2012). Template-based protein structure modeling using the RaptorX web server. *Nature protocols*, *7*(8), 1511-1522.
[27] Kamaraj, B., & Purohit, R. (2013). In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *BioMed research international*, *2013*.
[28] Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research*, *40*(W1), W294-W297.
[29] Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*, *4*(7), 1073-1081.
[30] Kumar, P., Singh, R. K., & Mahalingam, K. (2015). In silico analysis of fat mass obesity associated (FTO) gene using computational algorithms. *Int. J. Pharm. Bio Sci*, *6*, 589-599.
[31] Lenardo, M. J. (1991). lnterleukin-2 programs mouse αβ T lymphocytes for apoptosis. *Nature*, *353*(6347), 858-861.
[32] Lin, Y.-J., Wan, L., Sheu, J. J.-C., Huang, C.-M., Lin, C.-W., Lan, Y.-C., et al. (2008). G/T polymorphism in the interleukin-2 exon 1 region among Han Chinese systemic lupus erythematosus patients in Taiwan. *Clinical Immunology*, *129*(1), 36-39.
[33] Maiti, R., Van Domselaar, G. H., Zhang, H., & Wishart, D. S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic acids research*, *32*(suppl_2), W590-W594.
[34] Masoodi, T. A., Al Shammari, S. A., Al-Muammar, M. N., Alhamdan, A. A., & Talluri, V. R. (2013). Exploration of deleterious single nucleotide polymorphisms in late-onset Alzheimer disease susceptibility genes. *Gene*, *512*(2), 429-437.
[35] McKusick, V. (2016). Online Mendelian Inheritance in Man, OMIM™. Johns Hopkins University: Baltimore. MIM Number:# 193700.
[36] Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., & Tramontano, A. (2014). Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins: Structure, Function, and Bioinformatics*, *82*(S2), 1-6.
[37] Nailwal, M., & Chauhan, J. B. (2017). Analysis of consequences of non-synonymous SNPs of USP9Y gene in human using bioinformatics tools. *Meta Gene*, *12*, 13-17.
[38] Naveed, M., Tehreem, S., Mubeen, S., Nadeem, F., Zafar, F., & Irshad, M. (2016). In-silico analysis of non-synonymous-SNPs of STEAP2: To provoke the progression of prostate cancer. *Open Life Sciences*, *11*(1), 402-416.
[39] Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.

[40] Ng, P. C., & Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.*, *7*, 61-80.

[41] Ngan, C.-H., Hall, D. R., Zerbe, B., Grove, L. E., Kozakov, D., & Vajda, S. (2011). FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics*, *28*(2), 286-287.

[42] Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., & Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC structural biology*, *9*(1), 51.

[43] Raghav, D., Sharma, V., & Agarwal, S. M. (2013). Structural investigation of deleterious non-synonymous SNPs of EGFR gene. *Interdisciplinary sciences, computational life sciences*, *5*(1), 60.

[44] Rajasekaran, R., Doss, C. G. P., Sudandiradoss, C., Ramanathan, K., Rituraj, P., & Rao, S. (2008). Computational and structural investigation of deleterious functional SNPs in breast cancer BRCA2 gene. *Chinese Journal of Biotechnology*, *24*(5), 851-856.

[45] Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Folding and Design*, *3*(2), 141-147.

[46] Rost, B., Yachdav, G., & Liu, J. (2004). The predictprotein server. *Nucleic acids research*, *32*(suppl_2), W321-W326.

[47] Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, *5*(4), 725-738.

[48] Roy, A., Yang, J., & Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic acids research*, *40*(W1), W471-W477.

[49] Sahraeian, S. M., Luo, K. R., & Brenner, S. E. (2015). SIFTER search: a web server for accurate phylogeny-based protein function prediction. *Nucleic acids research*, *43*(W1), W141-W147.

[50] Sakaguchi, S., Yamaguchi, T., Nomura, T., & Ono, M. (2008). Regulatory T cells and immune tolerance. *Cell*, *133*(5), 775-787.

[51] Sayad, A., & Movafagh, A. (2014). The Association of− 330 Interleukin-2 Gene Polymorphism with Its Plasma Concentration in Iranian Multiple Sclerosis Patients. *Scientifica*, *2014*.

[52] Shen, Y., Liu, Y., Liu, S., & Zhang, A. (2012). The association between-330T/G polymorphism of interleukin 2 gene and bladder cancer. *DNA and cell biology*, *31*(6), 983-987.

[53] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*, *45*(D1), D362-D368.

[54] Togawa, S., Joh, T., Itoh, M., Katsuda, N., Ito, H., Matsuo, K., et al. (2005). Interleukin-2 gene polymorphisms associated with increased risk of gastric atrophy from Helicobacter pylori infection. *Helicobacter*, *10*(3), 172-178.

[55] Wang, S., Li, W., Liu, S., & Xu, J. (2016). RaptorX-Property: a web server for protein structure property prediction. *Nucleic acids research*, *44*(W1), W430-W435.

[56] Wang, S., Peng, J., Ma, J., & Xu, J. (2016). Protein secondary structure prediction using deep convolutional neural fields. *Scientific reports*, *6*.

[57] Williams, T., Eisenberg, L., Burlein, J., Norris, C., Pancer, S., Yao, D., et al. (1988). Two regions within the human IL-2 gene promoter are important for inducible IL-2 expression. *The Journal of Immunology*, *141*(2), 662-666.

[58] Yang, J., Roy, A., & Zhang, Y. (2012). BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic acids research*, *41*(D1), D1096-D1103.

[59] Yang, J., Roy, A., & Zhang, Y. (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, *29*(20), 2588-2595.

[60] Yang, J., Wang, Y., & Zhang, Y. (2016). ResQ: an approach to unified estimation of B-factor and residue-specific error in protein structure prediction. *Journal of molecular biology*, *428*(4), 693-701.

[61] Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, *12*(1), 7-8.