Human BLK tyrosine protein kinase is closely related to *Pongo Albelli*: An *in silico* structural and functional analysis

Princy Princy

Shaban Ahmad

Fatima Khan

Sahil Mehta

Sahar Qazi

Khalid Raza

BLK belongs to the family of SRC kinases (SFKs), and are diagnosed with the aid of the presence of an SH3 and SH2 regulatory domains of N-terminal to the catalytic kinase domain. BLK in signalling has a vast position in transmitting alerts via immunoglobulins and ends in pro-B to pre-B conversion, and in signalling for boom arrest and apoptosis downstream of the B-cellular receptor. We have performed a series of computational analysis on various aspects on BLK viz phylogenetic analysis, domain analysis, secondary structure prediction, charge distribution, prediction of the antigenic region and have executed structural analysis by first structure modelling and then its refinement, and active site prediction for better understanding of the human BLK as a drug target. Our study includes a detailed analysis and graphical representation of different domains, charge distribution, prediction of the antigenic region etc with corresponding sequence and its secondary structure for the pharmacological aspect of BLK which observed that *Tulipa suaveolens* is the most outed clade in the BLK and Human BLK is found to be very closed to *Pongo Abelii* through the phylogenetic tree assessment.

**Introduction**

BLK is one of the important members of the SRC family kinases (SFKs). In human, there are eight closely related non-receptor tyrosine kinases of SFKs includes SRC, YES1, FYN, FGR, LCK, LYN, HCK and BLK. These all kinases are related to each other as they all possess quite similar domain structures including N-terminal segment that includes the SH4 domain and 50-70 additional residues that are different to each family member (Oda, Kumar, & Howley, 1999). All are differentiated by SH3 and an SH2 regulatory domains N-terminal to the catalytic kinase domain. The SFKs are linked with the transmembrane receptors and turn on by extracellular receptor stimulation (Petersen et al., 2017). The SFKs are tightly regulated, it exhibits little activity in normal cells in the absence of regulatory signal. This regulation is mediated by phosphorylation/dephosphorylation of a specific tyrosine residue in the carboxy-terminal tail (Tyr 527 in c-Src). Regulatory tyrosine residue is phosphorylated by means of cytoplasmic tyrosine kinase which includes Csk causes a conformational exchange that effects within the closed-form of c-Src due to an intramolecular bonding among the phosphorylated tyrosine and SH2 domains. Due to closed conformation, Src could now not have interaction with different targeted proteins thru SH2 and SH3 domain. After dephosphorylation of Tyr 527, Src embraces an active (open) conformation in which the SH2 and SH3 domains are allowed to have interaction with some other goal. Inactive conformation, the kinase area of Tyr 416 is obtainable for autophosphorylation, for growing the kinase activity (Oda et al., 1999).

It has been shown that BLK, similar to other members of the Src group of tyrosine kinases (such as Lyn), interacts with BANK1, an adaptor/scaffold protein primarily expressed in B cells (Castillejo-López et al., 2012). Recently, various genome-wide association studies (GWAS) identified different single nucleotide polymorphisms (SNPs) in BLK (rs13277113A/G and rs2736340T/C) and BANK1 (rs10516487C/T R61H and rs3733197G/A A383T) to be associated with systemic lupus erythematosus (SLE), firstly in European and Asian-derived populations(Ramírez-Bello et al., 2019). Tumour necrosis factor receptor superfamily 4 (TNFSF4) part of TNF superfamily, and is expressed on dendritic cells, macrophages, cluster of differentiation (CD)4+/CD8+ T cells, activated NK cells and other cells (Murata et al., 2000). Recently research showed that gene‑degree interaction among BLK and TNFSF4 may additionally display a synergistic effect on T cells and B cells thru the nuclear signalling NF‑kB pathway, and this can describe the function in figuring out the immunologic aberration (Shen et al., 2017).

BLK entails in signal transduction downstream of the B‑cellular receptor; subsequently, it might affect the proliferation and differentiation of B cells. (Harley et al., 2008). For example, SRC is an oncogene concerned in several malignancies which include breast, lung, colon and pancreatic most cancers (Summy & Gallick, 2003). Recently it is shown that the active human blk has an oncogenic property to induce tumours in mice (Petersen et al., 2017). Transgenic mice expressed constantly active murine Blk in their lymphoid organ developed lymphomatous B- and T-cell-like disorders confirming a transforming potential of orthologue murine (Malek et al., 1998).

The works, however, have never been reviewed. Here, we will try to relate with the function, and their motif is conserved or not, we have compiled for the first time a detailed analysis of all the relevant information for BLK that may be helpful in the understanding of functional importance of BLK in the body system.

**Materials and Methods**

We have done various deep analysis on human BLK sequence, homology modelling is the development of an atomic version of a target protein primarily based totally at the goal’s amino acid series and the templates; experimentally determined systems of homologous proteins. In recent years structural biology shifted towards structure-based drug discovery. We have modelled the structure using the techniques of homology modelling (Jacobson et al., 2004; Waterhouse et al., 2018). The structure was pre-processed using the protein preparation wizard and have filled the gap using Prime, the further structure was minimized in terms of its energy using the same wizard for the next step of the analysis. The active site is the region of an enzyme in which substrate molecules bind and go through a chemical reaction. The lively site consists of residues that shape brief bonds with the binding site and residues that catalyse a response of the catalytic site, prediction of the active site of the BLKhave bee done using SiteMap (Singh et al., 2011; Wass et al., 2010) with the specific parameters of more restrictive and standard grid of crops site of 4 maps for the further analysis and have got total of 5 active site. We have also extracted the amino acid distribution throughout the sequence using the ProtParam and got the CSV format data after that we have performed various analysis in terms of statistical analysis and have plotted the data for the quick understanding of the BLK. ProtParam lets in the computation of diverse bodily and chemical parameters such as molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity for a given protein saved in Swiss-Prot or TrEMBL or for a user-entered protein sequence(Gasteiger et al., 2005). Secondary structure of the protein has been predicted using emboss (Garnier et al., 1978). GeneCards server was used for the categorisation of mRNA expression level in different human tissues. Antigenic regions on protein defined a lot about inhibitors and the future research scope on that particular protein. In our studies Emboss was used for the extraction of the antigenic region and Geneious prime was used for the plotting of the data. Additionally, we have also used InterProScan for the characterisation of the domains on the surface of BLK for better functional understanding. We have collected BLK from thousands of species and have mined the best, based on the characterisation and have taken the only largest size of BLK from every species and got a total of 324 species. Further data have been aligned using MAFFT v7.450 Alignment (Rozewicki et al., 2019). We have used FFT-NS-1 and 200PAM/k=2 as main principle parameters and have got comparative annotations of every aspect of the various BLK. Geneious Tree has performed for the plotting of the tree with automatic direction determiner. Further, Itol server (Letunic & Bork, 2019) used for editing of a phylogenetic tree for better understanding with its annotation characterization and bootstrapping.

**Result and Discussion**

**Global reports on BLK**

There are many platforms for the scientific reporting about any biological issues, PubMed considered one of the best databases to enlist the manuscript data and other databases of NCBI for the biological issues such as the genome, nucleotide gene etc. We have collected the data of PubMed and plotted the data with few mining techniques for understanding its importance in the scientific community. From the NCBI website, we have downloaded the publications on Schizont egress antigen and classified them based on its publication year. We identified a total number of 1046 of publications on BLK, we have seen scientists are taking their interest on this kinase from the year of 1992 but because of lack of techniques, there were very few reports they have been published before 1992. After the 2000s, the number of publications increased exponentially. Every year the number of publications is growing two-fold. In the year 2017, it was highest with the 85 publications, which has been represented in Fig. 1. From the above observation, we can find interest in this particular kinase.

**Fig. 1** Year-wise publication of BLK in PubMed

**Sequence analysis of the BLK**

We have retrieved the data of Human BLK from uniport (https://www.uniprot.org/uniprot/P51451) and analysed the sequence of BLK using Geneious Prime (https://www.geneious.com/prime/) for the basic quality assessment. Our sequence is having charge of 2.21 at pH 7 and isoelectric point is 7.61. Frequency of Acidic residues is 66 which is 13.1% of the sequence, frequency of basic residues is 76 which is 15%, Polarly uncharged and charged are 121 and 142 which are 24% and 28.1% respectively. Hydrophobic residues are 252 in number which is making 49.9% of the total sequence.ProtParam (Gasteiger et al., 2005), EMBOSS plugin (Rice, Longden, & Bleasby, 2000)and have arranged the data using excel and plotted using tableau. Computer-based clustering analysis of kinase is picking up in significance as an expository apparatus, it is getting conventional for the investigation because of the accessibility of quick increment in the succession.In a complete sequence of analysis of BLK, we have found, Lysine is having the highest frequency of total 68 times and Arginine is second highest of 66residues and both are negatively charged. Cysteine (sulphur-containing amino acid) and histidine which considered to be an inflammatory agent in immune response whichare the lowest occurring amino acid andare eight in numbers (Fig. 2).

**Fig. 2** Amino acid distribution throughout the sequence of BLK

**Secondary structure prediction**

Structural analysis is one of the major fields and it helped in the identification of many ligand molecules for the proper binding with the receptor (Sliwoski et al., 2014; Kaul et al., 2019). We have got the 23 alpha-helix which considered energetically favourable for hydrogen-bonding, 29 beta-strand which are generically more rigid than other components, 36 coils and 36 turns which considered to contain unusual conformational abilities. We have plotted sequence and secondary structure parallelly using Geneious prime (Geneious 2020.0.5 (https://www.geneious.com)) (Fig. 3). We have got the length ranges from 2-27 and the maximum length of the alpha helix is from 357-383. Beta strand is showing the length complexity from 1-14. While the turn and coil are showing the length ranges from 1-7 residues (Fig. 4) Aloha helix are made more hydrogen bonds while the beta-strand are giving high stability to the structure. From Ramachandran plot, we have found there is Cis proline is present which is 2/24 (A306 GLU-A307 PRO), (A504 GLN-A505 PRO) and Ramachandran outliers which is 1.13% A501 TYR, A500 GLN, A496 ALA, A386 ASP, A142 ILE.

**Fig. 3** Count of Secondary structure components

**Fig. 4** Secondary structure of Human BLK

**Homology modelling**

The Crystal structure of human BLK was not available on RCSB PDB databases, hence its secondary structure was predicted from its primary sequence. We have done the homology modelling of Human BLK, for its further accuracy confirmation it was necessary to analyze its various structural properties. So, we have used the Schrodinger suite (Schrödinger, 2016) for the modelling and minimized the structure in protein preparation wizard and predicted active site on BLK.

The length of blk is 505 amino acids and the molecular weight is 57706 Da. We modelled three- dimensional structure of blk by using Schrodinger’s Prime (Jacobson et al., 2004) (Fig. 5A) and have got the sequence identity of 69.98. The Ramachandran Z-score value of the model is -1.182, the model is found to be related to HAEMATOPOETIC CELL KINASE (HCK). GMQE of the model is 0.77. Favoured by Ramachandran plot which is 95.25% (Fig. 7). The active site has been predicted using Schrodinger’s SiteMap (Halgren, 2009). We have got a total of 5 active sites on Human BLK. The best binding site for the ligand has shown in Fig. 5B.

**Fig. 5** (A) Homology model of BLK, (B) Predicted active site

**Molecular Dynamic Simulation**

Protein structure refinement is an essential component and is the process of improving the qualities of modelled protein structures to approximate them r to their native states. Thus, to improve our modelled structure of BLK, we employed molecular dynamics (MD) simulations using two well-known software namely – Visual Molecular Dynamics (VMD) (Humphrey et al, 1996), which is a molecular graphics tool designed for the display and analysis of molecular structures, and i3DRefine software (Bhattacharya & Cheng, 2013) for the fulfilment of stabilization. Our analysis has shown to refine the originally modelled structure of BLK to approximately ~90%. The refined structure has a MolProbity score of 9.7, clash score of 0.5, Ramachandran favoured regions have significantly improved to be 98.2 with no poor rotamers. Fig. 6 depicts the two refined structures of human BLK protein a) using VMD and b) using i3DRefine software respectively.

**Fig 6.** Refined BLK structures a) VMD refined structure and b) i3DRefine refined structure.

**Table 1.** Results of refinement software(s) for refined BLK protein structure.

**Fig. 7** Ramachandran Plot for BLK

**Site of Expression**

B-cellular receptor (BCR) signalling requires tight law of numerous protein tyrosine kinases and phosphatases and associated coreceptors. B-mobile antigen receptor binds to the antigen triggers signalling that causes B-cell activation. BLK in signalling has a numerous position like transmitting indicators via floor immunoglobulins and leads pro-B to pre-B conversion, and signalling for growth arresting and apoptotic downstream of the B-cell receptor. It binds specifically and phosphorylates the CD79A at ’Tyr-188’ and ’Tyr-199’, also CD79B at ’Tyr-196’ and ’Tyr-207’ and Signalling additionally phosphorylates the IgG receptors FCGR2A, FCGR2B and FCGR2C (Bewarder et al., 1996). This additionally results in BTK activation by not directly stimulating BTK intramolecular autophosphorylation. In islets of the pancreas, it acts as a regulator of beta-cells characteristic via the up-regulation of PDX1 and NKX6-1 and leads to stimulation of insulin secretion within the presence of glucose(Fishilevich et al., 2016). Phosphorylates CGAS, and promote retention of CGAS inside the cytosol (Liu et al., 2018). The expression of signalling is under the manipulate of NF-kappa-B and the B-cell particular transcription factors PAX5 and EBF1(Akerblad & Sigvardsson, 1999). The steady-state state of BLK expression showed no obvious correlation with the mode of B cell transformation. Expression in these cell lines is therefore not an aberrancy related to a particular process of transformation, nor is it likely to completely allow the transformed phenotype. Expression of BLK turned into discovered in B cell precursors as well as in cellular lines consultant of mature B cells, suggesting that during B cellular ontogeny, BLK is expressed before the advent of surface immunoglobulin (Fig. 8). The B lymphoid cell specificity of blk expression distinguishes it from the three other individuals of the Src own family which might be preferentially expressed in hematopoietic cells: LCK, FGR, and HCK.B lymphocytes contain a tyrosine kinase hobby with a substrate specificity and phosphorylation kinetics which are awesome from the principal pastime in T cells; a main endogenous substrate for tyrosine phosphorylation in B cells has an obvious molecular mass of 55 kD. B cellular progenitors can gather in secondary lymphoid tissues below situations of polyclonal activation of stress in addition to in heavy chain-poor mice expressing activated Ras within the B lymphoid compartment (Tretter et al., 2003).

**Fig. 8** Level of mRNA expression in normal human tissue

**Tissue specificity**

BLK showed its expression in various tissue of human beings such as lymphatic organs, pancreatic islets, Leydig cells, striate ducts of salivary glands and hair follicles etc. GeneCards (Stelzer et al., 2016) used to extract the data of mRNA expression in normal human tissues and found the expression ranges from 2-3. It has been noticed higher expression in blood, retina, spinal cord, liver, lung, Thyroid, Prostrate, Lymph Node.

**BLK and Cancer**

Recently, confirmed the novel evidence that human BLK in its energetic form show oncogene belongings with the ability to guide the growth of lymphoid cells in vitro and to promote tumour increase in vivo. BLK may be a potential novel therapeutic goal in Cutaneous T-cell lymphoma (CTCL) (Petersen et al., 2017). The constitutively active shape of human BLK became capable of remodelling the lymphocytic Ba/F3 cells into increase aspect independence through conferring resistance to apoptosis brought about with the aid of cytokine withdrawal in vitro (Petersen et al., 2017).

**Antigenic Regions on BLK**

BLK is having good antigenic property because it has shown the total of 19 antigenic regions. We have used emboss (Garnier et al., 1978) for the prediction of the antigenic region on BLK. We have allotted the parameter as the minimum length of the antigenic region as 6 residues and have got the 19 antigenic region which is ranges from 7-33 regions. We have also plotted the respective data using Geneious prime (Fig. 9) along with its potential in the cytoplasm.

**Fig. 9** Predicted antigenic region on BLK

**Gene structure and its regulation**

The gene for BLK, assigned to the chromosome 8p23.1, exon remember is 15. It is suggested that the 100-kb place entails the BLK gene, which codes for a nonreceptor tyrosine-kinase of the Src family of proto-oncogenes concerned in cell proliferation and differentiation. On the gene transcription, web page of 20kb five’ mutation turned into positioned, one in exon 4 where it decided an Ala to Thr substitution at role 71, 1 at the end of the three’ UTR, 1 right away three’ of the polyadenylation sign, and 18Kb from the gene at the 3’ facet. Whereas BLK was reported within the literature to be expressed best in B lymphocytes (Dymecki et al., 1990), analysis of current expression statistics revealed that this gene turned into additionally expressed in human pancreatic islets-a finding that turned into showed via RT-PCR. Both BLK probes inside the array gave a more potent hybridization sign with RNA isolated from microdissected Beta-cells instead of entire islets Staining of a human tissue array with an anti-BLK antibody showed the microarray findings (Borowiec et al., 2009).

It is stated that during murine the BLK gene involves 13 exons less than 30kb in size of DNA on mouse chromosome 14. In the primary 3 exons, which offers the five’- untranslated place and the amino acid of N-terminal series unique to p55 Blk, the blk gene is different from other individuals of the Src family however final 10 exons, display specific organisation of gene to that of the other Src genes (Dymecki, Zwollo, Zeller, Kuhajda, & Desiderio, 1992). It is pronounced that the Expression of blk regulated throughout B-mobile development and BLK RNA became found in all pro-B-, pre-B, and mature B-cell lines examined however were absent from plasma cellular traces. Nucleotide sequences of blk exons have been decided by means of the dideoxynucleotide termination method (Holbrook et al., 1984).The ordinary company of blk is much like the murine and human lck, murine hck, bird c-src, and human c-fgr. In its first three exons, BLK is not homologous in nucleotide sequence to another src circle of relatives individuals. Nonetheless, the configuration of these three exons resembles that of the T-lymphocyte kinase gene LCK. In each BLK and LCK, exon 1 is untranslated and is separated from exon 1’ with the aid of >10 kb of DNA; exons 1’ and a couple of encoding N-terminal amino acid residues that show little similarity amongst Src kinases. A more general similarity is clear in exons three-12; in BLK, LCK, HCK, C-SRC, and C-FGR, these exonsencode the conserved SRC homology regions SH3 and SH2, the catalytic area, and a C-terminal regulatory place, are homologous with admire to nucleotide sequence and organization.

By sequence analysis, BLK is maximum intently related to the SRC circle of relatives contributors ZCK and HCK; this similarity is pondered in the genomic organisation of the 3 genes. The homology found amongst SRC own family participants shows that they rise through the duplication of an ancestral gene. A phylogenetic map of tyrosine kinases, primarily based on the sequences in their catalytic domain names, places ZCK on a department awesome from that which includes C-SRC and C-FGR. BLK is expressed especially in cells of the B-lineage, it may feature in a sign transduction pathway specific to B-lymphocytes. Several transmembrane proteins expressed on B-cells seem to signal through tyrosine phosphorylation however they do now not show intrinsic kinase hobby. The pleasant characterised consist of floor immunoglobulin and class II important histocompatibility complicated (MHC)1 antigen. Cross-linking of floor Ig seems to stimulate the enzymatic activities of Blk, Fyn, and Lyn. Ig and MHC class1 antigens each have a special pattern of expression whilst B-cellular development take area, but each of them is expressed on mature B-cells and are absent from plasma cells. The proof implicating p55blk in signalling pathways regarding one or both of these proteins led us to study the Structural, Biochemical and developmental law of the blk gene.

**Domain Analysis**

BLK kinase has 3 domains named SH3 domain, SH2 domain, Protein Tyrosine Kinase was located in a region of 64 – 110, 124 – 205, 241 – 490 respectively. We have used Pfam server (El-Gebali et al., 2019) for the domain analysis. We have also checked for glycosylation site on blk using NetNGlyc 1.0 Server (Gupta, Jung, & Brubak, 2004) which shows there is not glycosylation site on BLK. We have also plotted all 3 domains of the BLK along with the secondary structure and sequence representation for better understanding (Fig. 10).

**Fig. 10** Predicted domains on BLK

**Analysis of Isoelectric point and charges of the residues**

We have analysed isoelectric point and charges of every residue on blk using the Geneious prime. This plot can clearly define the Hydrophobicity of the residues along with its isoelectric point, both play an important role in drug designing. We have also categorised the Transmembrane permeability and plotted its affinity at outside, inside and on the plasma membrane of the cell. An amino acid charge is basically defining the charge per residues which can be added for the cumulative analysis of the receptor. The cumulative of hydrophobicity at each residue could reveal a lot about the drug design for BLK to upregulate and downregulate in the human body. This graph is depicting the charge, TM along with the secondary structure for the better understanding of the system. For the detailed analysis of the domains of the Human BLK, we have plotted all the annotations along with the domain region on the sequence (Fig. 11).

**Fig. 11** Graphical representation of annotations of the BLK

**Role of BLK in the human body**

Signalling via BLK performs a critical position in transmitting indicators through floor immunoglobulins and provide the supports to seasoned-B to pre-B transition, as well as the signalling for increase arrest and apoptosis downstream of the B-cell receptor. Phosphorylation of CD79A at ’Tyr-188’and ’Tyr-199’, as well as CD79B at ’Tyr-196’ and ’Tyr-207’ is done by specific binding of BLK and itphosphorylates in addition to the immunoglobulin G receptors FCGR2A, FCGR2B and FCGR2C. Alongwith FYN and LYN, it performs an essential position in the pre-B-cell receptor (pre-BCR)-mediated NF-kappa-B activation (Clark, Mandal, Ochiai, & Singh, 2014). BLKcontributes to BTK activation by using not directly stimulating BTK intramolecular autophosphorylation. In islets of the pancreas, acts as a modulator of beta-cells characteristic through the up-law of PDX1 and NKX6-1 and consequent stimulation the secretion of insulin in reaction to glucose. Phosphorylates CGAS, promoting retention of CGAS inside the cytosol(Kurosaki, 2002).The biological function of BLK consist of B cell receptor signalling pathway, cell differentiation, intracellular signal transduction, peptidyl-tyrosine autophosphorylation, peptidyl-tyrosine phosphorylation, positive regulation of insulin secretion, regulation of B cell receptor signalling pathway, regulation of B cell receptor signalling pathway, regulation of cell population proliferation receptor of transmembrane protein tyrosine kinase signalling pathway (JADWIN, 2017).

**BLK and Disease**

The mutation associated with BCL causes MODY(Maturity-onset diabetes of the younger) is a type of Diabetes this is an autosomal ailment this happens on a completely young age, earlier than the age of 25 (Craig et al., 2014). It is pronounced that a hundred-kb vicinity belongs to the BLK gene, this encoded bySrc’s own family non-receptor tyrosine kinase of the which features for mobile proliferation and differentiation(Dymecki et al., 1992). It is suggested that the BLK has an effect on insulin secretion and synthesis, and BLK became both knocked-down or overexpressed by using retrovirus (Borowiec et al., 2009). In low glucose, neither BLK overexpression nor its downregulation had widespread results on insulin secretion. But in a high concentration of glucose decorate the insulin secretion while the antonymic effect was cited in cells wherein BLK were downregulated. The boom in insulin secretion precipitated by BLK overexpression changed into followed by using a 70% increase in insulin content compared with control cells. Thus, BLK may beautify insulin reaction to glucose at the least in part by using growing the amount of insulin available for secretion(M. Borowiec et al., 2009).This impact is blunted by means of the Ala71Thr mutation. In the settlement with the insulin secretion and content material facts, it’s been determined a 40% increase in insulin mark abundance in MIN6 beta-cells overexpressing BLK and a 15% lower in cells in which BLK was knocked down(Dai et al., 2015; Sharma et al., 2015). These findings denoted that the changes in insulin content material regulated with the aid of BLK passed off at the extent of transcription. It is also pronounced that the up-law of the transcription aspect Nkx6.1, this is concerned within the manager of glucose-inspired insulin secretion in pancreatic beta-cells and it’s far viable that the BLK-induced increase in protein stages of Pdx-1 without delay leads the expression of Nkx6.1 and the 2 transcription elements collectively enhance beta-cellular function and mass (Schisler et al., 2005).

**Phylogeny**

The origin and evolutionary aspects of BLK were analysed through the phylogenetic tree. We have chosen a total of 324 number of BLK, only one sequence from one species, from animal and Plantae kingdom through various databases. Species wise categorisation on the longest read of the sequence has provided enough and more accurate amount of data for the prospective analysis. We have got the tree using Geneious tree and have plotted its tree using Itol server (Letunic & Bork, 2019) for better comprehension. We have plotted the name of species to refer BLK from that particular species and have recoloured human BLK for better comparison in terms of origin and occurrence of mutation. Distances are represented at nodes of every species to get the actual benefit of comprehension (Fig. 12). We have characterized the whole data into 4 distinct class excluding human BLK. Our study shows Tulipa suaveolens as outer species from all of the 324 species data with the highest value of 0.77. Human BLK comes in between which is having huge numbers of mutations if compare with the *T suaveolens* . *Eudyptes filholi* is one of the more mutated species with a wide-angle of the same category. The ring belongs to *T chinensis* is further extended with the wider species count and range of development in the last category. Fundulus heteroclitus have shown a wide distinction in the third ring of the tree. Our studies and plot for this BLK have categorically defined the evolutionary aspect of the BLK’s mutational event.

**Fig. 12** Phylogenetic tree of the BLK from every species reported yet

**Conclusion**

In this study, a common putative feature of BLK has been analysed extensively using bioinformatics tools. BLK is majorly chargeable for the formation of B-Lymphocytes. It has been observed that human BLK in its lively shape is an oncogene with the potential aid in terms of growth of lymphoid cells in vitro and promotes tumour boom in vivo, thus BLK can be considered to be a capacity novel therapeutic target for Cutaneous T-cellular lymphoma. The antibody-mediated surface interaction of the B-cell antigen receptor (BCR) leads to phosphorylation of BLK on tyrosine amino acids, shows the enzymatic activity. We have modelled the structure and did the various analysis of BLK structures which will add the knowledge for understanding the BLK’s functionality in our human body and its regular development. This study would not only help pharmacologists to develop new drugs but also to bioinformaticians in developing new inhibitor to regulate its expression. Till date, the crystallised structure of BLK is not available hence our modelled structure followed by structure refinement using MD simulation gives a stable structure that can be used in pharmacology.

Competing interest

The authors declare that there is no conflict of interest in the publication of this manuscript.

References

Akerblad, P., & Sigvardsson, M. (1999). Early B cell factor is an activator of the B lymphoid kinase promoter in early B cell development. *J Immunol, 163* (10), 5453-5461.

Bewarder, N., Weinrich, V., Budde, P., Hartmann, D., Flaswinkel, H., Reth, M., & Frey, J. (1996). In vivo and in vitro specificity of protein tyrosine kinases for immunoglobulin G receptor (FcgammaRII) phosphorylation. *Mol Cell Biol, 16* (9), 4735-4743.

Bhattacharya, D., & Cheng, J. (2013). i3Drefine software for protein 3D structure refinement and its assessment in CASP10. *PloS one* , *8* (7).

Borowiec, M., Liew, C. W., Thompson, R., Boonyasrisawat, W., Hu, J., Mlynarski, W. M., . . . Doria, A. (2009). Mutations at the BLK locus linked to maturity onset diabetes of the young and beta-cell dysfunction. *Proc Natl Acad Sci U S A, 106* (34), 14460-14465.

Castillejo-López, C., Delgado-Vega, A. M., Wojcik, J., Kozyrev, S. V., Thavathiru, E., Wu, Y.-Y., . . . Fineschi, S. (2012). Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK.*Annals of the rheumatic diseases, 71* (1), 136-142.

Clark, M. R., Mandal, M., Ochiai, K., & Singh, H. (2014). Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling.*Nature reviews Immunology,* 14(2), 69-80.

Craig, M. E., Hattersley, A., & Donaghue, K. C. (2009). Definition, epidemiology and classification of diabetes in children and adolescents. *Pediatric diabetes* , *10* , 3-12.

Dai, F. F., Bhattacharjee, A., Liu, Y., Batchuluun, B., Zhang, M., Wang, X. S., … & Wheeler, M. B. (2015). A novel GLP1 receptor interacting protein ATP6ap2 regulates insulin secretion in pancreatic beta cells. *Journal of Biological Chemistry* , *290* (41), 25045-25061.

Dymecki, S. M., Niederhuber, J. E., & Desiderio, S. V. (1990). Specific expression of a tyrosine kinase gene, blk, in B lymphoid cells. *Science* , *247* (4940), 332-336.

Dymecki, S. M., Zwollo, P., Zeller, K., Kuhajda, F. P., & Desiderio, S. V. (1992). Structure and developmental regulation of the B-lymphoid tyrosine kinase gene blk. *Journal of Biological Chemistry* , *267* (7), 4815-4823.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., … & Sonnhammer, E. L. L. (2019). The Pfam protein families database in 2019. *Nucleic acids research* , *47* (D1), D427-D432.

Fishilevich, S., Zimmerman, S., Kohn, A., Iny Stein, T., Olender, T., Kolker, E., . . . Lancet, D. (2016). Genic insights from integrated human proteomics in GeneCards.*Database, 2016* .

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of molecular biology* , *120* (1), 97-120.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In*The proteomics protocols handbook* (pp. 571-607): Springer.

Gupta, R., Jung, E., & Brubak, S. (2004). NetNGlyc: prediction of N-glycosylation sites in human proteins. In: Prep.

Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of chemical information and modeling* , *49* (2), 377-389.

Harley, J. B., Alarcón-Riquelme, M. E., Criswell, L. A., Jacob, C. O., Kimberly, R. P., Moser, K. L., . . . Genetics, I. C. f. S. L. E. (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nature genetics, 40* (2), 204.

Holbrook, N. J., Smith, K. A., Fornace, A. J., Comeau, C. M., Wiskocil, R. L., & Crabtree, G. R. (1984). T-cell growth factor: complete nucleotide sequence and organization of the gene in normal and malignant cells. *Proceedings of the National Academy of Sciences* , *81* (6), 1634-1638.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics* , *14* (1), 33-38.

Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., & Friesner, R. A. (2004). A hierarchical approach to all‐atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics* , *55* (2), 351-367.

JADWIN, J. A. (2017). Quantitative Analysis of EGFR Phosphorylation and SH2 Domain Binding in vivo.*Doctoral Dissertations* , 1437, University of Connecticut Graduate School.

Kaul, T., Eswaran, M., Ahmad, S., Thangaraj, A., Jain, R., Kaul, R., … & Bharti, J. (2019). Probing the effect of a plus 1bp frameshift mutation in protein-DNA interface of domestication gene, NAMB1, in wheat. *Journal of Biomolecular Structure and Dynamics* , 1-15.

Kurosaki, T. (2002). Regulation of B-cell signal transduction by adaptor proteins. *Nature Reviews Immunology* , *2* (5), 354-363.

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* , *47* (W1), W256-W259.

Liu, H., Zhang, H., Wu, X., Ma, D., Wu, J., Wang, L., . . . Ge, B. (2018). Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature, 563* (7729), 131-136.

Malek, S. N., Dordai, D. I., Reim, J., Dintzis, H., & Desiderio, S. (1998). Malignant transformation of early lymphoid progenitors in mice expressing an activated Blk tyrosine kinase. *Proceedings of the National Academy of Sciences, 95* (13), 7351-7356.

Murata, K., Ishii, N., Takano, H., Miura, S., Ndhlovu, L. C., Nose, M., . . . Sugamura, K. (2000). Impairment of antigen-presenting cell function in mice lacking expression of OX40 ligand. *The Journal of experimental medicine, 191* (2), 365-374.

Oda, H., Kumar, S., & Howley, P. M. (1999). Regulation of the Src family tyrosine kinase Blk through E6AP-mediated ubiquitination. *Proc Natl Acad Sci U S A, 96* (17), 9557-9562.

Petersen, D. L., Berthelsen, J., Willerslev-Olsen, A., Fredholm, S., Dabelsteen, S., Bonefeld, C. M., . . . Woetmann, A. (2017). A novel BLK-induced tumor model. *Tumour Biol, 39* (7), 1010428317714196.

Ramírez-Bello, J., Jiménez-Morales, S., Montufar-Robles, I., Fragoso, J. M., Barbosa-Cobos, R. E., Saavedra, M. A., & Sánchez-Muñoz, F. (2019). BLK and BANK1 polymorphisms and interactions are associated in Mexican patients with systemic lupus erythematosus. *Inflammation Research, 68* (8), 705-713.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics: TIG* , *16* (6), 276-277.

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research* , *47* (W1), W5-W10.

Schisler, J. C., Jensen, P. B., Taylor, D. G., Becker, T. C., Knop, F. K., Takekawa, S., … & Newgard, C. B. (2005). The Nkx6. 1 homeodomain transcription factor suppresses glucagon expression and regulates glucose-stimulated insulin secretion in islet beta cells. *Proceedings of the National Academy of Sciences* , *102* (20), 7297-7302.

Schrödinger, L. (2016). Schrödinger Suite. *Schrödinger, LLC, New York, NY* .

Sharma, R. B., O’Donnell, A. C., Stamateris, R. E., Ha, B., McCloskey, K. M., Reynolds, P. R., … & Alonso, L. C. (2015). Insulin demand regulates β cell number via the unfolded protein response. *The Journal of clinical investigation* , *125* (10), 3831-3846.

Shen, Y., Liu, Y., Wang, X. Q., Ke, X., Kang, H. Y., & Hong, S. L. (2017). Association between TNFSF4 and BLK gene polymorphisms and susceptibility to allergic rhinitis.*Molecular medicine reports, 16* (3), 3224-3232.

Singh, T., Biswas, D., & Jayaram, B. (2011). AADS-An automated active site identification, docking, and scoring protocol for protein targets based on physicochemical descriptors. *Journal of chemical information and modeling* , *51* (10), 2515-2527.

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological reviews* , *66* (1), 334-395.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., … & Kaplan, S. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics* , *54* (1), 1-30.

Summy, J. M., & Gallick, G. E. (2003). Src family kinases in tumor progression and metastasis.*Cancer and metastasis reviews, 22* (4), 337-358.

Tretter, T., Ross, A. E., Dordai, D. I., & Desiderio, S. (2003). Mimicry of pre-B cell receptor signaling by activation of the tyrosine kinase Blk. *J Exp Med, 198* (12), 1863-1873.

Wass, M. N., Kelley, L. A., & Sternberg, M. J. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic acids research* , *38* (suppl\_2), W469-W473.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … & Lepore, R. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* , *46* (W1), W296-W303.

**Figure legends:**

**Fig. 1** Year-wise publication of BLK in PubMed.

**Fig. 2** Amino acid distribution throughout the sequence of BLK.

**Fig. 3** Count of Secondary structure components.

**Fig. 4** Secondary structure of Human BLK.

**Fig. 5** (A) Homology model of BLK, (B) Predicted active site.

**Fig 6.** Refined BLK structures a) VMD refined structure and b) i3DRefine refined structure.

**Fig. 7** Ramachandran Plot for BLK.

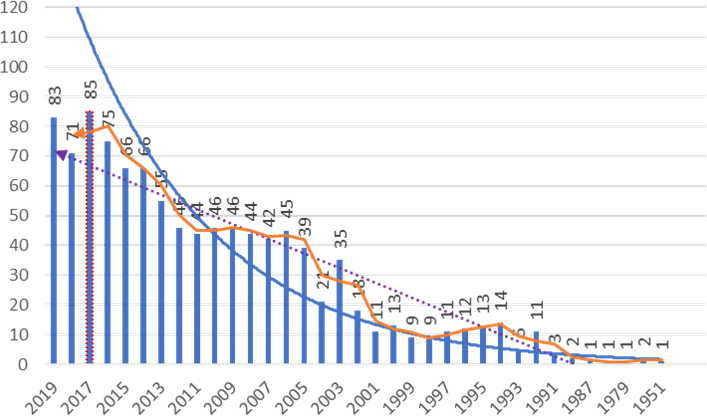
**Fig. 8** Level of mRNA expression in normal human tissue.

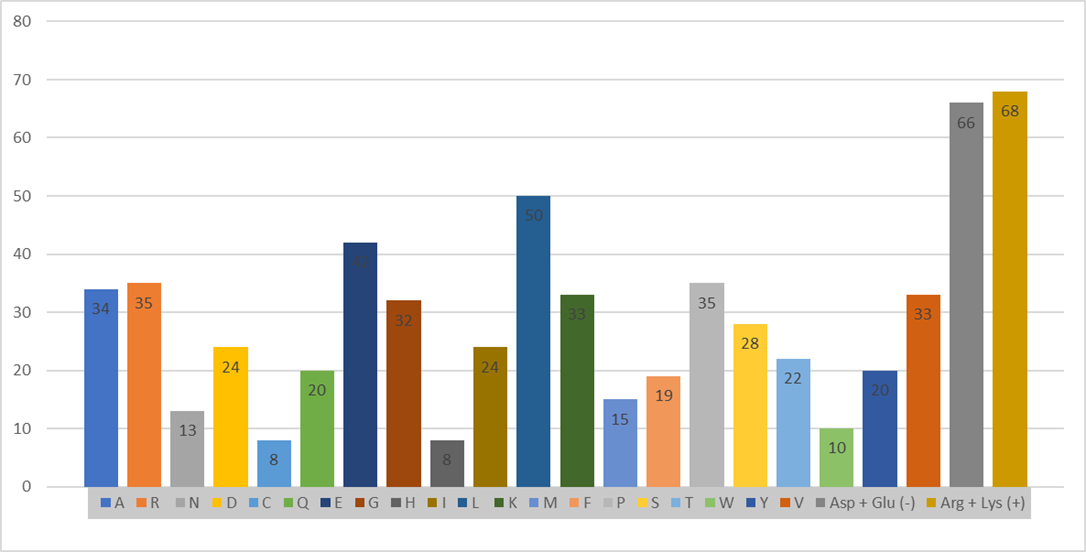
**Fig. 9** Predicted antigenic region on BLK.

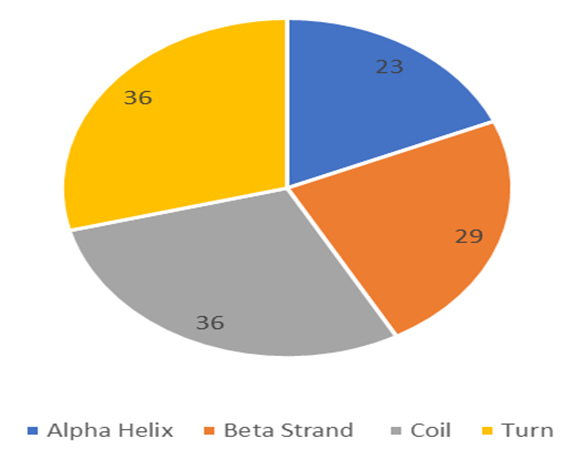
**Fig. 10** Predicted domains on BLK.

**Fig. 11** Graphical representation of annotations of the BLK.

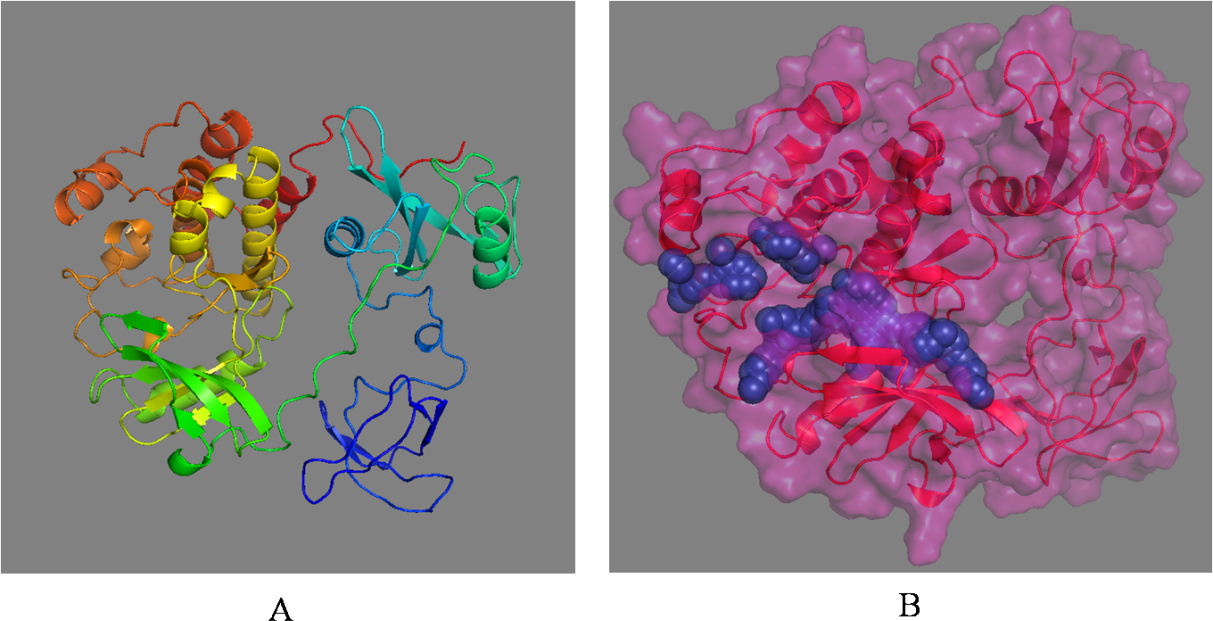
**Fig. 12** Phylogenetic tree of the BLK from every species reported yet.

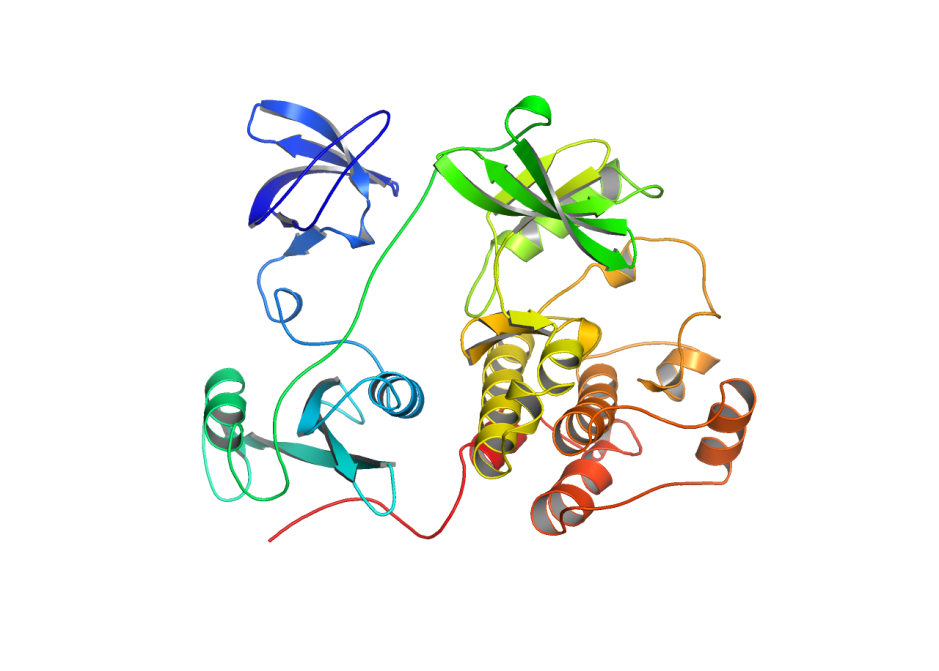


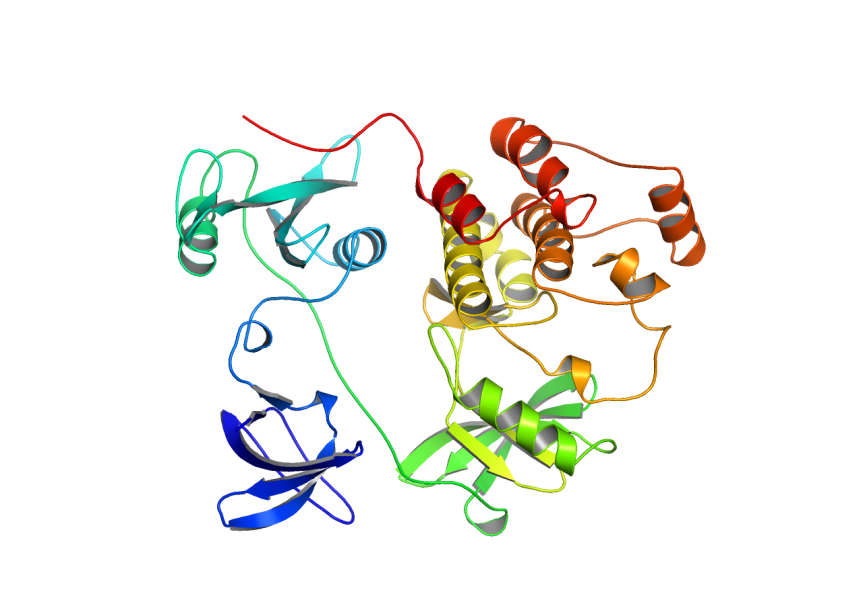




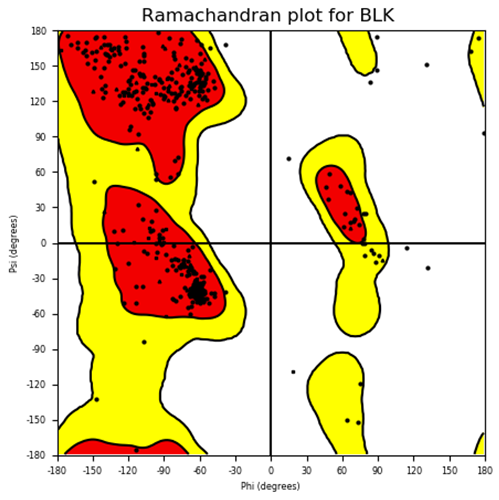


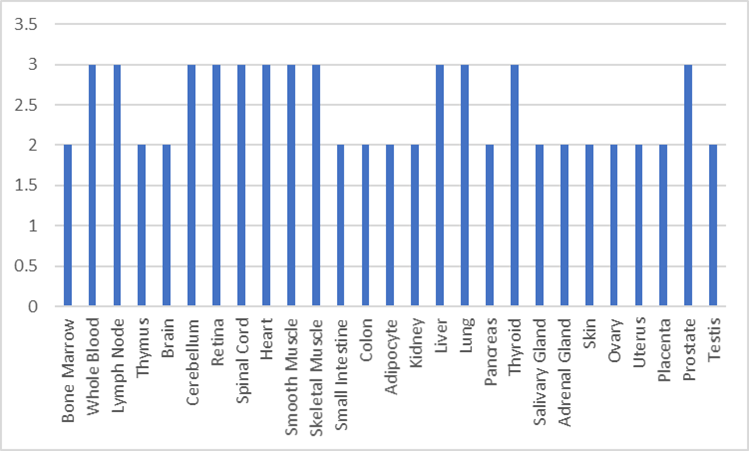


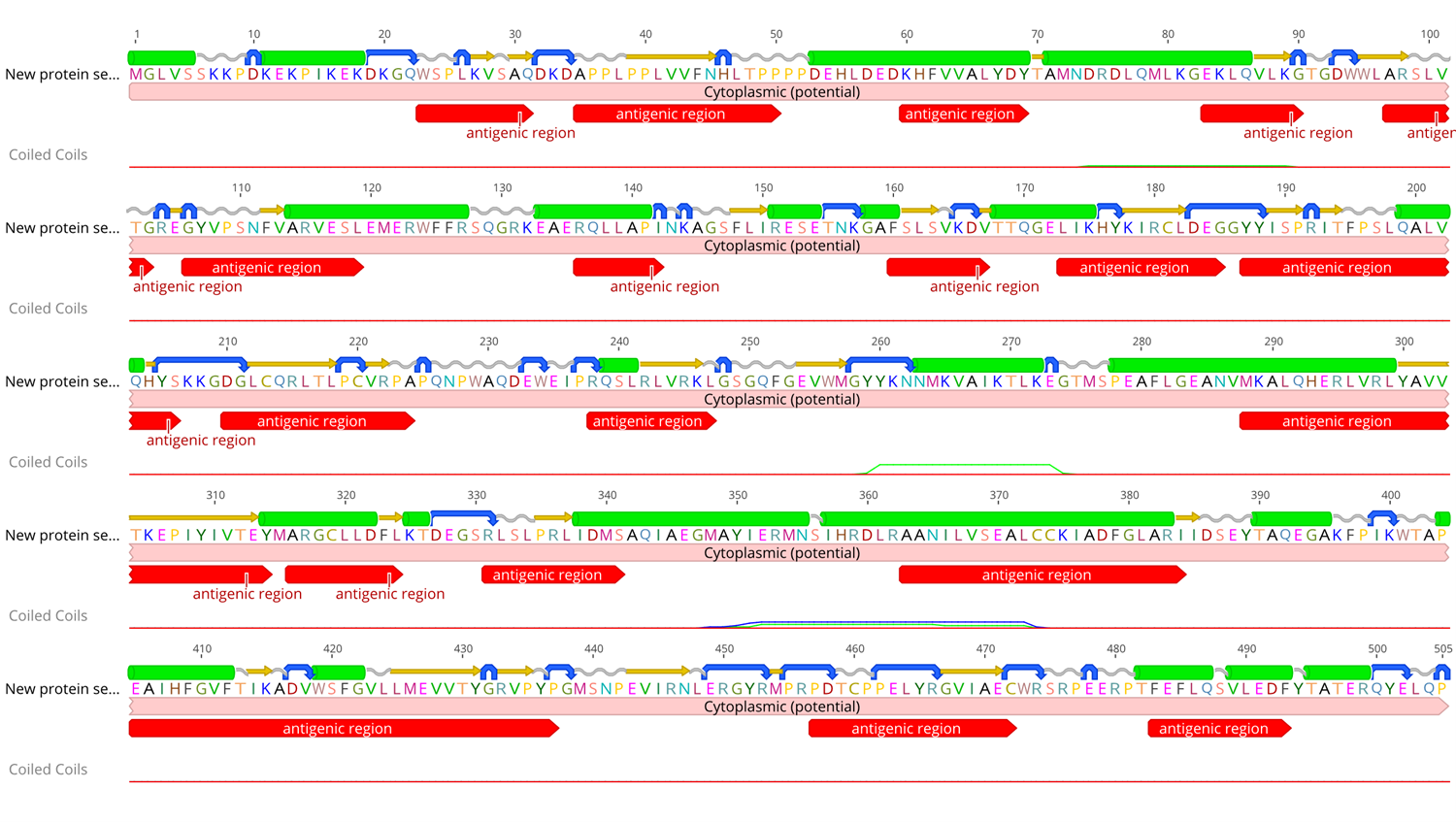


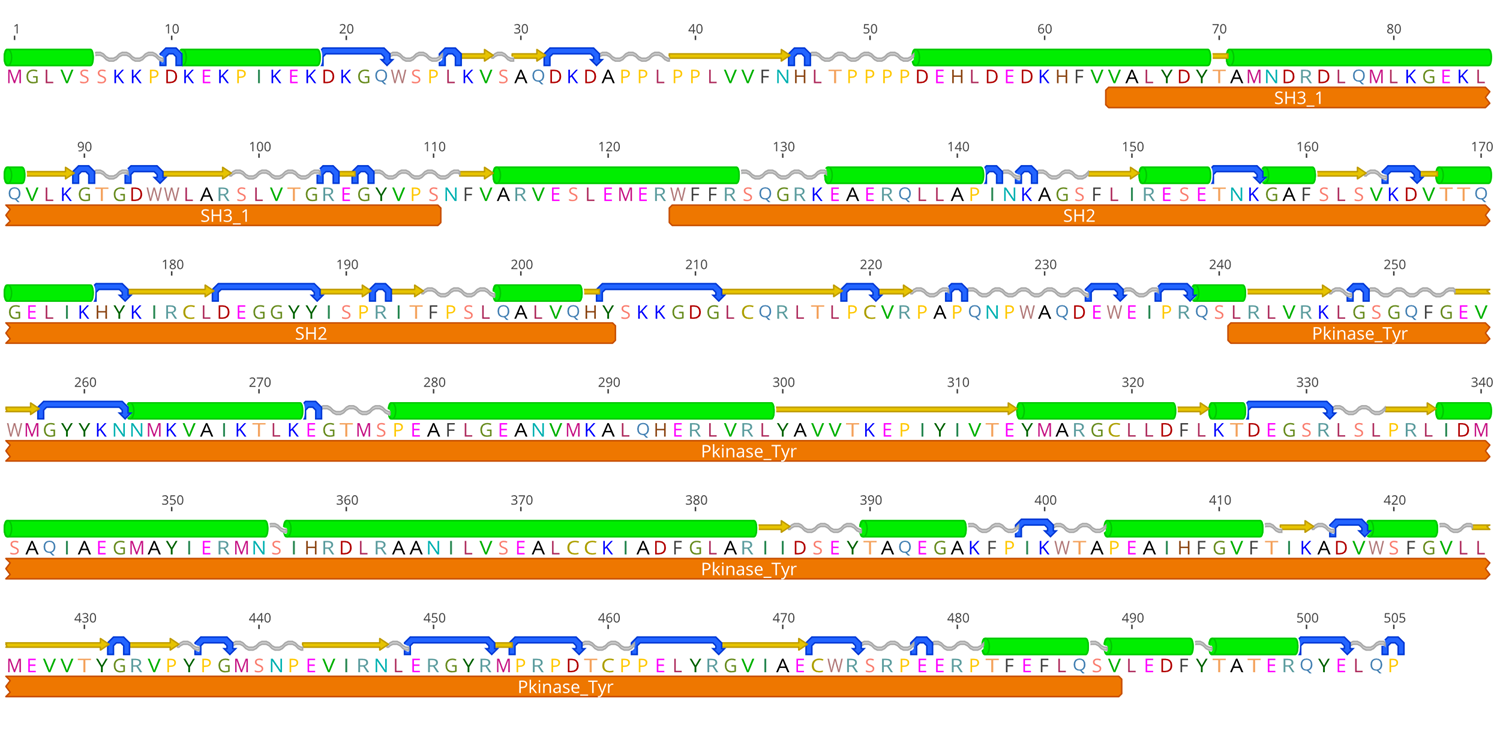


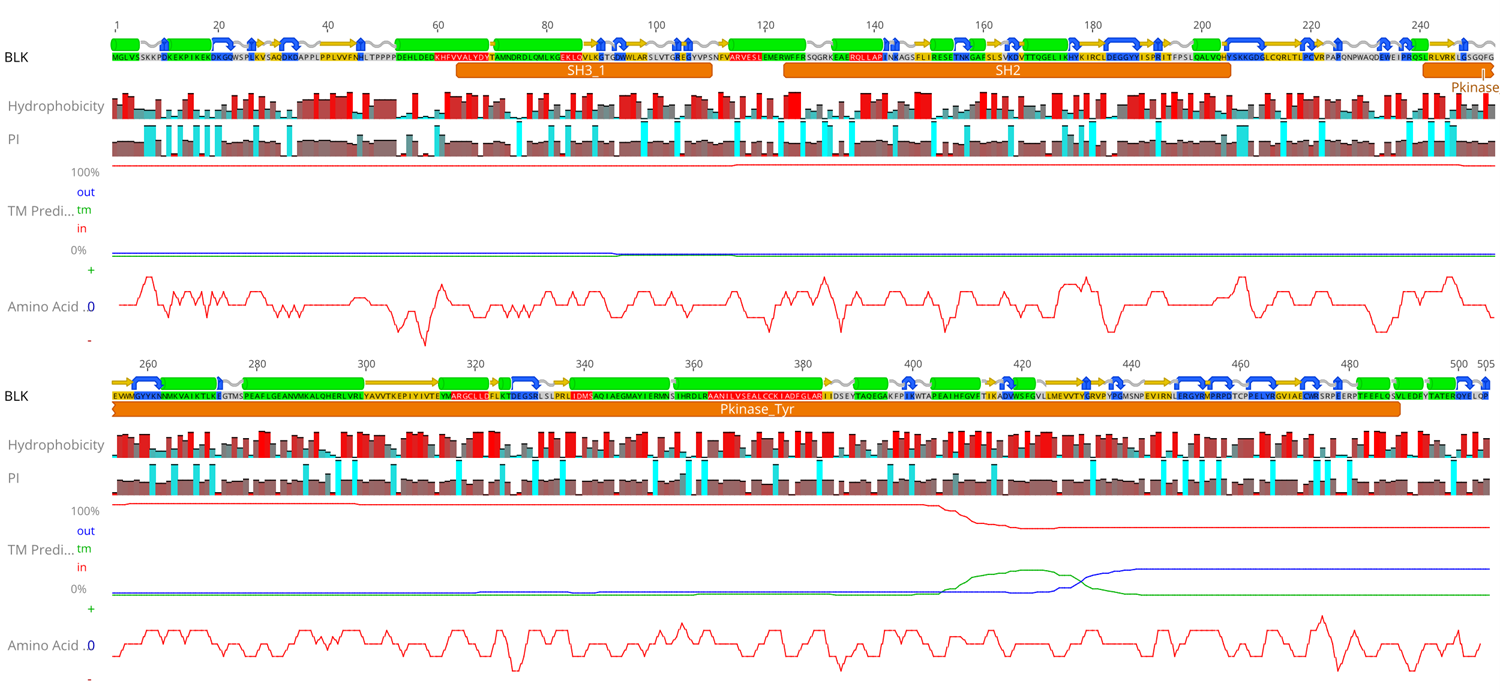
a) b)

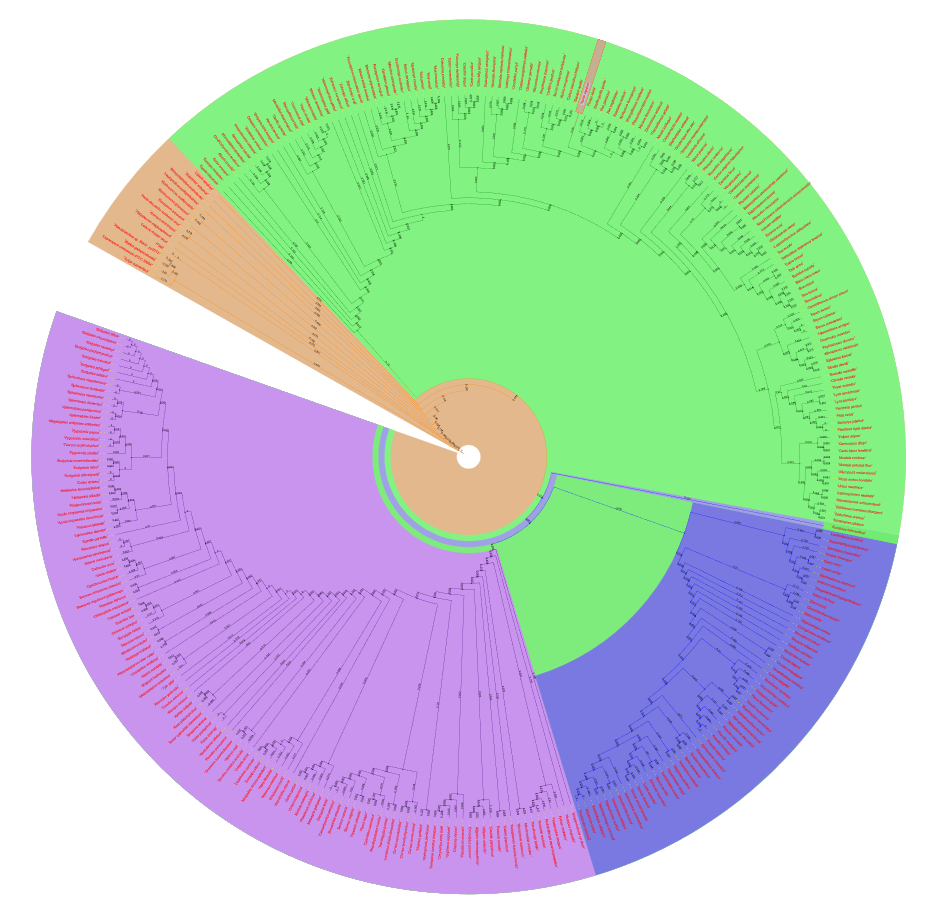












**Table 1.** Results of refinement software(s) for refined BLK protein structure.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Refinement** | **Clash Score** | **GDT-HA** | **RMSD** | **MolProbity** | **Poor Rotamers** | **RWPlus** | **Rama Favoured** |
| VMD | 0.5 | 0.9792 | 0.325 | 9.7 | 0.0 | -98320.0905 | 98.2 |
| i3DRefine | 0.56 | 0.9977 | 1.213 | 8.6 | 0.3 | -98193.9601 | 97.7 |

