Biology Direct

RESEARCH

Open Access

A genome-wide search of Toll/Interleukin-1 receptor (TIR) domain-containing adapter molecule (TICAM) and their evolutionary divergence from other TIR domain containing proteins

Shailya Verma[1] and Ramanathan Sowdhamini[1,2,3*]

Abstract

Toll/Interleukin-1 receptor (TIR) domains are cytoplasmic domain that mediates receptor signalling. These domains are present in proteins like Toll-like receptors (TLR), its signaling adaptors and Interleukins, that form a major part of the immune system. These TIR domain containing signaling adaptors binds to the TLRs and interacts with their TIR domains for downstream signaling. We have examined the evolutionary divergence across the tree of life of two of these TIR domain containing adaptor molecules (TICAM) *i.e.,* TIR domain-containing adapter-inducing interferon-β (TRIF/TICAM1) and TIR domain containing adaptor molecule2 (TRAM/TICAM2), by using computational approaches. We studied their orthologs, domain architecture, conserved motifs, and amino acid variations. Our study also adds a timeframe to infer the duplication of TICAM protein from *Leptocardii* and later divergence into TICAM1/TRIF and TICAM2/TRAM. More evidence of TRIF proteins was seen, but the absence of conserved co-existing domains such as TRIF-NTD, TIR, and RHIM domains in distant relatives hints on diversification and adaptation to different biological functions. TRAM was lost in *Actinopteri* and has conserved domain architecture of TIR across species except in *Aves*. An additional isoform of TRAM, TAG (TRAM adaptor with the GOLD domain), could be identified in species in the Mesozoic era. Finally, the Hypothesis based Likelihood ratio test was applied to look for selection pressure amongst orthologues of TRIF and TRAM to search for positively selected sites. These residues were mostly seen in the non-structural region of the proteins. Overall, this study unravels evolutionary information on the adaptors TRAM and TRIF and how well they had duplicated to perform diverse functions by changes in their domain architecture across lineages.

Keywords: GWS, TICAM, TRIF, TRAM, TAG, TLR4, TLR3, Protein evolution

Background

Toll-like receptors play a major role in the innate immune system by recognizing diverse exogenous and endogenous biomolecules (viral RNA, Bacterial or self DNA, LPS, etc.) as their ligands and produces cytokines along with other inflammatory mediators during infections. Structurally they have an extracellular Leucine-rich repeat (LRR) domain, a transmembrane domain, and an intracellular Toll/Interleukin-1 receptor (TIR) domain [1]. The Toll-like receptor proteins identify the ligands by their extracellular LRR domain and then homo or heterodimerize to recruit the adaptor molecules like MyD88, TIRAP, TRIF (specific to TLR4 and TLR3) and

*Correspondence: mini@ncbs.res.in

[1] National Centre for Biological Sciences, GKVK Campus, Bellary Road, Bangalore 560065, India
Full list of author information is available at the end of the article

TRAM (exclusive for TLR4 signaling pathway). The MyD88 & TIRAP dependent pathway ultimately releases pro-inflammatory cytokines, NF-κB, Tumor Necrosis Factor alpha (TNF-α), interleukin (IL-)1β, IL-6, and chemokines. TRIF & TRAM pathway releases both pro-inflammatory cytokines as well as anti-inflammatory mediator like Interferon Regulatory Factor 3 (IRF3), beta interferon (IFN-β), delayed NF-κB activation, type 1 IFN-α/β, IFN-α-inducible protein 10 (IP-10), MCP-5, RANTES, and nitric oxide [2].

Amongst these TLR proteins and its signaling adaptors, TIR domains are found in both. Apart from that, these domains are also present in Interleukins receptors as well as some accessory proteins. Overall there are 25 genes in the human genome that contain TIR domains as per PROSITE database—(Prosite ID: PS50104, Toll-like receptor proteins: TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, TLR signaling adaptor proteins: MyD88, TIRAP, TRIF/TICAM1, TRAM/TICAM2, SARM1, Interleukins receptor, and accessory proteins: SIGIRR, IRPL1, IL1R1, ILRL1, ILRL2, IL18R, I18RA, IRPL2, PIK3AP1 and BCAP/BANK1). A figure depicting the domain architecture of these TIR domain containing proteins are shown in Fig. 1.

Amongst these TIR domain containing proteins, Toll-like receptors, MyD88, SIGIRR, and Interleukin receptor proteins share similar kind of TIR domain structure (Pfam ID: PF01582.22), but the adaptor proteins TIRAP, TRAM, TRIF and SARM1 share a different Pfam TIR_2 domain (PF13676.8). Besides these PIK3AP1 and BCAP have TIR_3 domain (PF18567.3).

These TIR domains are cytoplasmic in nature and consist of approximately 200 amino acids. They in general promote the assembly of signaling complexes via protein–protein homotypic or heterotypic interactions. The TIR structure contains a central five-stranded parallel β sheet surrounded by five helices. Although TIR domains from different proteins have similar structure, their amino acid sequence identity is less than 30% when compared amongst each other. This makes them significantly diverse in sequence and structure amongst Toll-like receptors (TLRs), TLR adaptors, and Interleukin receptors [3]. Sequence analysis has shown three highly conserved regions among the different family of TIR proteins: Box1 (FDAFISY), Box2 (GYKLC-RD-PG), and Box3 (a conserved W surrounded by basic residues) [4].

Multiple evolutionary studies are performed on the evolution of TLR family protein across vertebrates [5–7]. These phylogeny-based studies add to our understanding

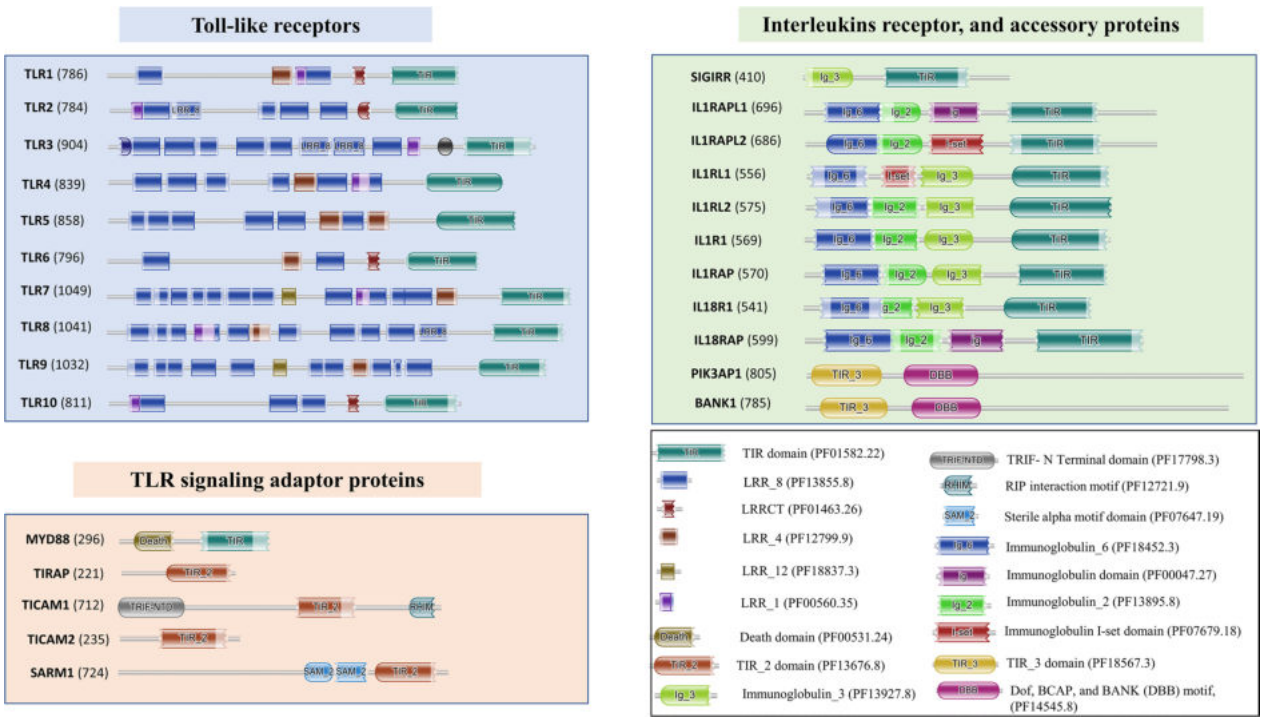


Fig. 1 The domain architecture for TIR domain containing proteins from human are shown. Pfam ID for each of the domains are mentioned in the legend. The length of the protein is mentioned in the bracket. (TIR: Toll/Interleukin-1 receptor, LRR: Leucine rich repeat, LRRCT: C-terminal LRR domain)

of the origin of TLR family proteins, their adaptors and probable signalling pathway. Evidence of **MyD88 in older taxa** explains the origins of MyD88 dependent pathways in invertebrates, although the TRIF and TRAM have resulted from an early duplication event in vertebrate TLR phylogeny [6]. This makes it interesting to mark the emergence of MyD88 independent pathways from the vertebrates. Another study showed the comparative and phylogeny analysis of TLR adaptors (TRIF, TRAM, and **TIRAP**) across 25 representative metazoans. The study aids to add knowledge about these adaptors and the evolution of their functional sites. Also, they have found shark to be the only non-Mammalia group to have TRAM [8].

All the previous evolutionary and phylogeny-based analyses were found to be focused around the TLR family and its divergence from the primitive organisms. Although the presence of the MyD88 adaptor molecules started in early invertebrates, it was interesting to study the origin of the MyD88 independent pathway. Early research suggests the emergence of the TIR domain-containing adapter molecule (TICAM) pathway among the basal chordates (Amphioxus) along with its structural and functional role [9]. Our study aims to look into the TLR MyD88 independent pathway adaptors (TRIF and TRAM) across all lineages. In this paper, we report genome-wide search for orthologs and analysis of domain architecture, sequence conservation, and evolutionary selection amongst those.

Results

Human TIR containing proteins and conserved motif

As described previously, TIR subfamily is represented at least in 25 genes in the human genome. They can be majorly classified into three categories: Toll-like receptor proteins (TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10), TLR signaling adaptor proteins (MyD88, TIRAP, TRIF/TICAM1, TRAM/TICAM2, SARM1), Interleukin receptors and accessory proteins (SIGIRR, IL1AP, IRPL1, IL1R1, ILRL1, ILRL2, IL18R, I18RA, IRPL2), PIK3AP1 and BCAP.

This set of proteins were further employed as queries to search for representative sequences from other taxa (Primates, Odd-toe ungulate, Even-toe ungulate, Carnivore, Placental, Whale and dolphins, Chiropteran, Rodentia, Lagomorpha, and Insectivores) to understand the phylogenetic relationships amongst these proteins. The sequences of the 25 TIR containing genes, along with their representative sequences from different taxa, are provided in Additional File 2.

A maximum likelihood-based tree is as Additional File 3. The respective branch lengths are mentioned and bootstrap values are shown in different sizes of circles. Also, the node ID shows the representative organism from each group. Each TIR domain containing proteins is colored distinctly. The phylogeny clearly shows the varying branch length and clustering of sequences into three major groups as specified earlier. The plasma membrane-based TLR proteins (TLR1, TLR2, TLR4, TLR6, TLR10) and endosomal membrane TLR proteins cluster separately (TLR3, TLR6, TLR7, TLR8, TLR9). TLR5 is found to cluster with TLR3 which may be because of closer identity amongst them. TRIF clusters with other adaptor molecules TRAM and TIRAP as the third cluster. Interestingly, in this phylogeny, the TRIF/TICAM1 was the protein with the highest branch lengths and appears within this cluster. From the motif search, TRAM and TRIF were seen to only have Common-4 motif conserved, unlike conventional TIR domain containing proteins which have Box1, Box2, and Box3 as well. Hence, we decided to perform genome-wide sequence search of TRAM and TRIF across all available taxa.

Search for TRAM and TRIF orthologues

TRAM and TRIF play role in the MyD88 independent pathway and, are involved in IRF3 and IFN β signaling via TLR4 (involves both TRAM and TRIF) and TLR3 (only TRIF) receptors. Activation of TRIF dependent pathway also helps in dendritic cell maturation, thereby acts as a link between innate and adaptive immune responses [2]. Therefore, a query set of TRIF and TRAM proteins from 25 organisms across different orders of Mammalia were considered to search for their orthologs (please see "Methods" section).

TRIF and TRAM share higher sequence identity compared to other TIR containing proteins. So, for comparative analysis, a sequence identity comparison was done for the TIR domain of TRAM and TRIF which is shown in Fig. 2A. This shows TRAM-TIR shares around 40% sequence identity with TRIF-TIR across all Mammalia taxa. The orthologs obtained after CS-BLAST search from TRAM and TRIF queries were used to construct a subfamily-specific sequence similarity network (SSN) using ZEBRA which is shown in Fig. 2B [10]. The SSN show 10 distinct protein subfamily clusters represented by different colors. The nodes between 45 to 100% pairwise sequence identity was connected. Also, we found a pairwise sequence identity of only 31% across the TRAM-TIR (PDB ID: 2M1W) and TRIF-TIR (PDB ID: 2M1X) [11]. Additionally, Subfamily 6 and 10 were seen to be clustered together with TRAM-TIR thereby representing TRAM orthologs. Similarly, Subfamilies 1, 2, and 5 were clustering well with TRIF-TIR representing TRIF orthologs. Apart from these, Subfamily 3 and 4 were found to be scattered and Subfamily 7, 8, and 9 clustered together. Few sequences were considered as

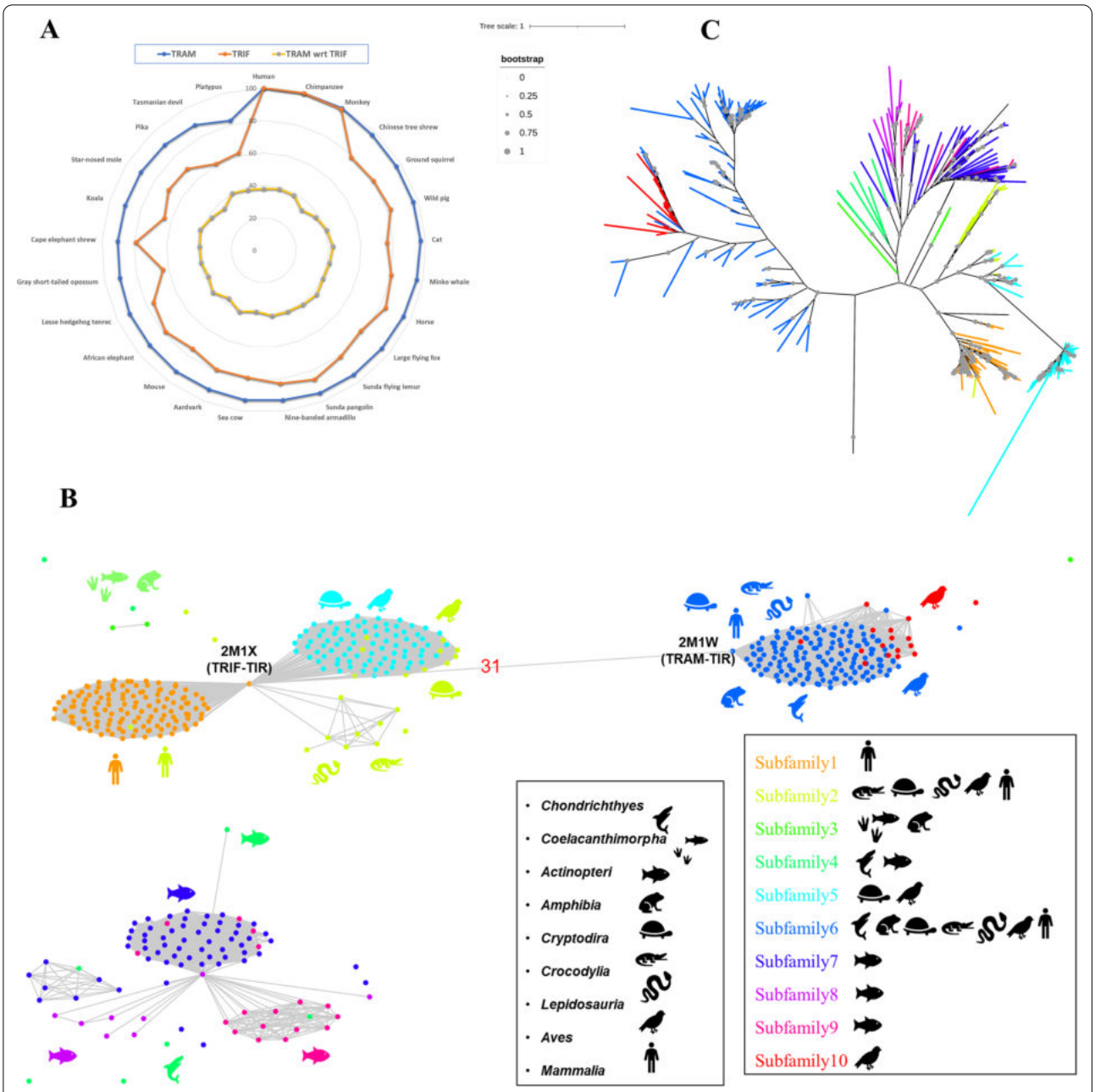


Fig. 2 **A** The radar plot shows the sequence identity between the TIR domain of TRAM (blue) and TRIF (orange) of respective organism wrt human TRAM-TIR and TRIF-TIR. Also, the inner most line (yellow) represent sequence identity percentage of TRAM- TIR and TRIF-TIR within the same organism. **B** Sequence similarity network, with edges between 45 and 100% pairwise sequence identity, and clustering of sequences into subfamilies. Representative organisms from each taxon are represented across each subfamily. In the clustering method, TRAM protein separates distinctively from the other clusters. For the TRIF representatives, Actinopteri and Chondrichthyes separate as a distinct cluster. Mammals represents an individual cluster closer to clusters of Aves, Bifurcata, Crocodylia, Cryptodira, and Amphibia. Interestingly, members from the Amphibia clusters separates notably from the Reptiles clusters (Bifurcata, Crocodylia, Cryptodira). Thereby, explaining the extent of differences amongst these sequence in accordance with evolutionary perspective. **C** An unrooted phylogeny showing different subfamilies. The left cluster represents TRAM and right represents TRIF orthologs

outliers and were not shown in SSN. We constructed an unrooted phylogeny from CSBLAST hits to examine the relationship between these bigger clusters. The unrooted phylogeny is shown in Fig. 2C and it shows well-separated clades diverging at the base of the tree. The basal node represents sequences from Belcher's lancelet

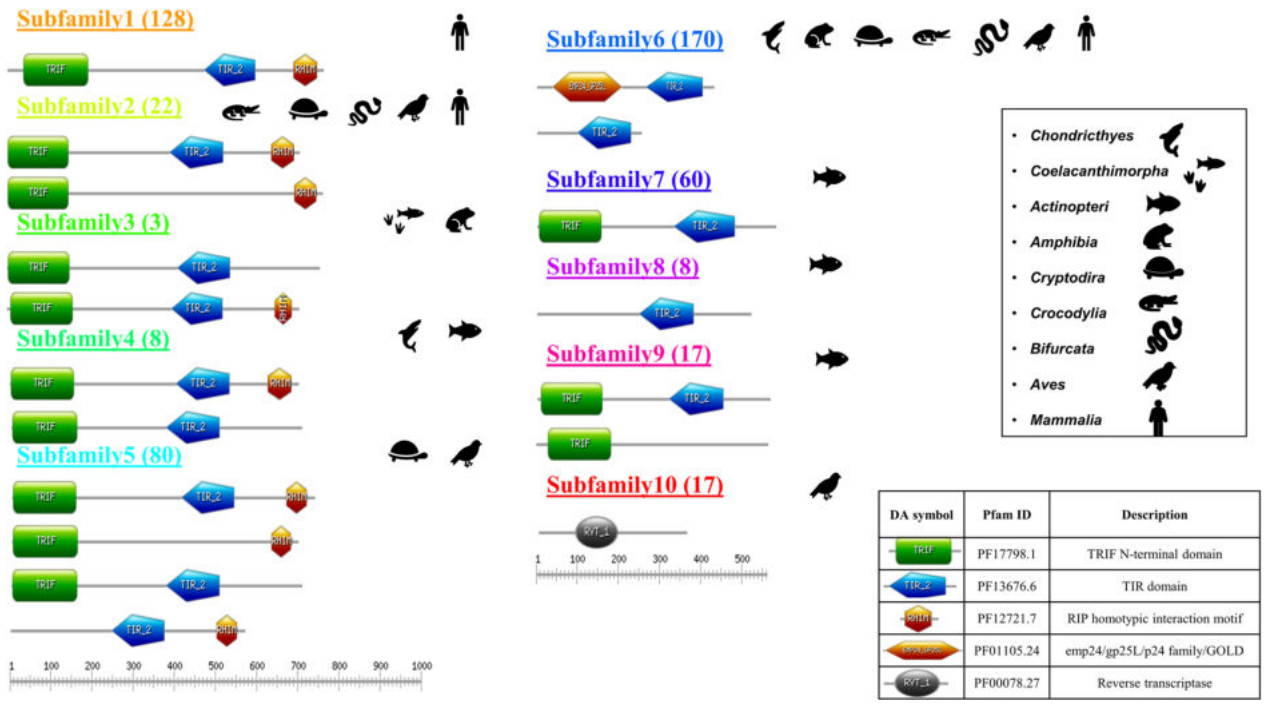


Fig. 3 The picture shows different subfamily and corresponding unique domain architecture present in them. Also, the representative taxa of the cluster are shown as cartoon icons. The PfamID and description for each domain is also mentioned

(*Branchiostoma belcheri*) and Florida lancelet (*Branchiostoma floridae*), commonly referred as Amphioxus and is grouped together as *Leptocardii*. These are known to be the oldest basal chordates with MyD88-independent pathway [9]. A similar pattern of clustering is seen in the phylogeny tree and TRAM separates well and distinctly from TRIF. Further, since subfamilies 7, 8, and 9 cluster with TRIF, a detailed analysis of genes that harbour these domains were performed. Analysis of domain architecture can enable to understand gain and loss of co-existing domains and diversification of function.

Domain architecture among subfamilies

A typical human TRIF protein has TRIF-NTD, TIR_2, and RHIM domains. Here the N-terminal is used for activation of IFNβ promoter activity [12]. TIR_2 domain is involved in homo and heterotypic TIR interaction for signaling and the RHIM domain is important for NF-κB activation [13]. Whereas the TRAM protein contains only the TIR_2 domain. Additionally, TRAM's isoform, TAG contains EMP24_GP25L along with the TIR_2 domain. This EMP24_GP25L domain is implicated in bringing the cargo forward and binding to coat protein [14].

Next, we examined the unique domain architectures and the representative taxa amongst them. A pictorial representation of the same has been shown in Fig. 3. Also, the number of sequences in each category is mentioned. Thereby with respect to Subfamily clusters as obtained in SSN analysis, Subfamily 6 and 10 includes TRAM and TAG which is a splice variant of TRAM (TRAM adaptor with the GOLD domain) [15]. Also, subfamily 10 majorly consists of hits from Aves, and in which no conventional TIR domain was found instead a RVT_1 (Reverse transcriptase domain family) domain was found. Apart from this Subfamily 1, 2, and 5 consists of TRIF sequences with well-annotated domain architecture from *Mammalia, Aves, Bifurcata, Crocodylia,* and *Cryptodira*. Amongst these Subfamily 1, 2, and 5 each of them clusters distinctively, with Subfamily 1 including sequence only from *Mammalia.* Whereas Subfamily 5 and 2 both have sequences from *Aves and Cryptodira*. Moreover Subfamily 2 also have sequences exclusively from *Bifurcata and Crocodylia.* Besides these, Subfamilies 3 and 4 have sequences from *Chondrichthyes, Coelacanthimorpha* (living fossil), and few sequences from *Amphibia* and *Actinopteri*. But most members of *Actinopteri* clusters together under Subfamily 7, 8, and 9. The domain architecture for *Actinopteri consists* of TRIF, TIR_2 domain and lacks the RHIM domain (HMM scan, inclusion e-value = 0.001).

Sequences from *Leptocardii* were considered as an outlier, so we looked at the secondary structure of sequences from the primitive organisms. Sequences

from *Leptocardii, Chondrichthyes,* and *Coelacanthimorpha* were compared with respect to sequences from human TRAM and TRIF. Good conservation along the TIR domains can be seen in the alignment. The image is attached in Additional File 1: Fig. S3. Moreover, good secondary structure conservation hints towards ancestral homology.

Phylogeny of TRAM and TRIF orthologs

We separated the hits into TRAM, TAG, and TRIF subgroups and constructed a phylogeny using Maximum likelihood with 100 bootstraps and branch lengths.

TRIF orthologs were found across *Chondrichthyes, Coelacanthimorpha, Actinopteri, Amphibia, Cryptodira, Crocodylia, Bifurcata, Aves,* and *Mammalia*. Human TRIF has three typical and two atypical TRAF6 binding sites. The generic motif representing this site is denoted by [PxExxD/W/E/F/Y]. The motif sequence and residue positions of human TRIF protein for typical TRAF6 binding sites are PEEPPD (86–91), PEEMSW (250–255) and PVECTE (301–306), whereas the same for atypical sites are PLESSP (491–496) and PPELPS (264–269). Amongst these, the second typical TRAF6 binding site [PEEMSW (250–255)] is one of the most important sites for TRIF interaction with TRAF6 and NF-κb induction [16]. Apart from these, TRIF also has a pLxIS motif ([[NDQEKR]LxIS]) which is a phosphorylation site used for inducing IRF3 activation and RHIM interacting motif ([[IV]Q[ILV]GxxNx[MLI]]) which is part of RHIM domain and is important for inducing apoptosis [17, 18]. Human TRIF has these motifs in following order with their residue position ~ TRAF6(86–91)~**pLxIS(207–210)**~**TRAF6(250–255)**~atypical-TRAF6(264–269)~TRAF6(301–306)~atypical-TRAF6(491–496)~**RHIM motif(687–695)**. Amongst these, the highlighted motifs are shown in the detailed phylogeny in alignment form in the same order as they appear in human TRIF protein (Uniprot ID: Q8IUC6) [12]. Members of *Mammalia, Amphibia, Cryptodira, Coelacanthimorpha* shows conserved pLxIS motif and among *Bifurcata*, *Actinopteri, Aves, Crocodylia* only few organisms have pLxIS motifs. Besides these, the second typical TRAF6 binding sites (PEEMSW) is seen conserved only in *Mammalia* and some organisms of *Aves*, it may be possible that other TRAF6 binding sites may be functional in other taxa. Whereas the RHIM motif along with the RHIM domain are seen to be conserved in all taxa except for *Actinopteri* and *Leptocardii*.

We have represented motifs and domain annotation for each organism in the phylogeny. We observe that none of the members of *Crocodylia* taxa show TIR_2 domains with inclusion e-value (HMM scan, inclusion e-value = 0.001). Even on increasing the value to 0.1 only one of the members *Crocodylus porosus*, shows TIR domain at e-value = 0.048, but that was an insignificant search (Fig S6). The detailed phylogeny with motifs and domain annotations is attached as Additional File 4.

Unlike TRIF, human TRAM (Uniprot ID: Q86XR7) consists of a putative *N-Myristoylation site* (residue position: 2–7) that helps in its localization to the plasma membrane and is critical for TLR4 pathways in response to LPS [12, 19]. The *N-Myristoylation site* has a consensus motif sequence of [G{EDRKHPFYW}xx[STAGCN]{P}], which is important for signal transduction [20]. TRAM protein also has a putative TRAF6 binding motif [PxExxP] (residue position: 181–186) that helps to interact with TRAF6 and mediate activation of the inflammatory responses by TLR4 [21]. Upon observing these in the phylogeny, we found the TIR_2 domain annotation and conserved motifs were missing from *Aves*, suggesting they may not have a functional TRAM protein. Additionally, *Aves* taxa has the highest branch lengths and highest amino acid distances as shown in Additional File 5. This implicates that *Aves* taxa may have undergone a higher divergence and genetic changes over evolutionary time. Previously reported studies with representative sequences across taxa claims TRAM to be lost in fishes, birds, and amphibians. But interestingly by our genome-wide search, we found TRAM orthologues for Aves [22]. None of the TRAM orthologue sequences were retrieved from *Actinopteri* (bony fishes).We found TRAM ortholog sequence in *Callorhinchus mili* from Chondrichthyes (cartilaginous fish), it also shows a concentional TIR_2 domain architecture along with conserved motif. This may be the oldest organism with functional motifs conserved with TIR_2 domain. Also, amongst members of *Amphibia*, all the hits have conserved TIR domain along with conventional *N-Myristoylation* motif except for *Xenopus laevis*. On combining all of our observation we observe that TRAM may have diverged from TRIF and first appeared in *Chondrichthyes* (399 mya), was then lost in *Actinopteri,* and then reappeared in *Amphibia* (323 mya), *Cryptodira, Crocodylia, Bifurcata, Aves*, and *Mammalia.*

TRAM isoform (Uniprot ID: Q86XR7-2) with GOLD domains was also seen across one *Bifurcata*, one *Crocodylia*, a few *Aves,* and *Mammalia* taxa. All of them depicted the well-conserved domain architecture of EMP24_GP25L (primarily involved in the transport of cargo molecules from the endoplasmic reticulum to the Golgi complex [23]) and TIR_2 domain. The difference in sequences of canonical human TRAM and isoform TAG is seen at the position from 1 to 20 (MGIGKSKINSCPL SLSWGKR→MPRPGSAQRW..), this has been shown

in the alignment. A similar pattern can be seen across *Mammalia.* TAG seems to be found mainly in *Mammalia, Aves*, and in a few cases of *Crocodylia*, and *Bifurcata*. Extensive genome sequencing can help recognize TAG across other organisms. An extensive phylogeny showing the presence of TAG in different taxa is shown in Fig. 4.

A representative phylogeny from each category is shown below showing the proportion of organisms. The evolutionary timeline for the divergence of TICAM is also shown below in Fig. 5 [24]. The detailed phylogeny for TRIF and TRAM is attached as Additional Files 4 and 5.

Conservation of Synteny

To focus on the orthologous relationships of TRIF and TRAM, we next examined the syntenies and check the preservation of the order of genes amongst the species and check for the neighbour genes. One representative from each taxon e.g., *Homo sapiens (Mammalia), Falco peregrinus (Aves), Gekko japonicus (Bifurcata), Alligator mississippiensis (Crocodylia), Chelonia mydas (Cryptodira), Xenopus laevis (Amphibia), Callorhinchus milii (Chondrichthyes), Danio rerio (Actinopteri), Latimeria chalumnae (Coelacanthimorpha)* and *Branchiostoma belcheri (Leptocardii)* were selected. The accession ID and domain architecture for these representatives can be seen in Additional Files 4 and 5.

There is overall good correspondence amongst the neighbours of TRAM and TRIF orthologs, except TRIF of *Amphibia* (lacks the gene neighbours), and TRAM of *Aves* (does not have a full one-to-one correspondence) as seen in Fig. 6. Interestingly, FEM1C and CDO1 like genes were also found to be neighbouring TICAM2 genes in *Chondrichthyes* and other TICAM2 orthologs. Due to the event of whole-genome duplication occurring at *Actinopteri*, two forms of TICAMs are generated. Since these duplicates are preserved across the lineages, there may be subfunctionalization or neofunctionalization [25]. To ascertain this, it will be interesting to perform functional characterization of distant orthologs from *Coelacanthimorpha* and *Actinopteri*.

Evolutionary selection pressure in TRIF and TRAM

From the previously obtained TRIF and TRAM phylogeny we observed the variation among branch lengths, and also the varied domain architecture amongst the sequences.

We were further interested to look into the selection pressure of TRIF and TRAM protein. We used the orthologues sequences and used site model from Codon substitution models (codeml) of PAML4.9 package. The implemented site models (M0, M1, M2, M3, M7, and M8) allowed ω ratio ($\omega = d_N/d_S$, ratio of nonsynonymous/synonymous substitution rates) to vary among codon sites in protein. Based on the Likelihood ratio test, using the Bayes empirical Bayes (BEB) method or posterior probabilities of model **M8** for 11 site classes ($k = 11$) along the sequence of the protein was plotted. The values from 11 site classes were grouped into two categories as $\omega > 1$ and $\omega < 1$. Also, another graph depicting the mean probabilities for each site was plotted in Fig. 7.

From both these sequences, one can notice that the posterior probability of ω was moreover $< = 1$ for sequences in domain regions. Although some positively selected sites were detected. The propensity of these

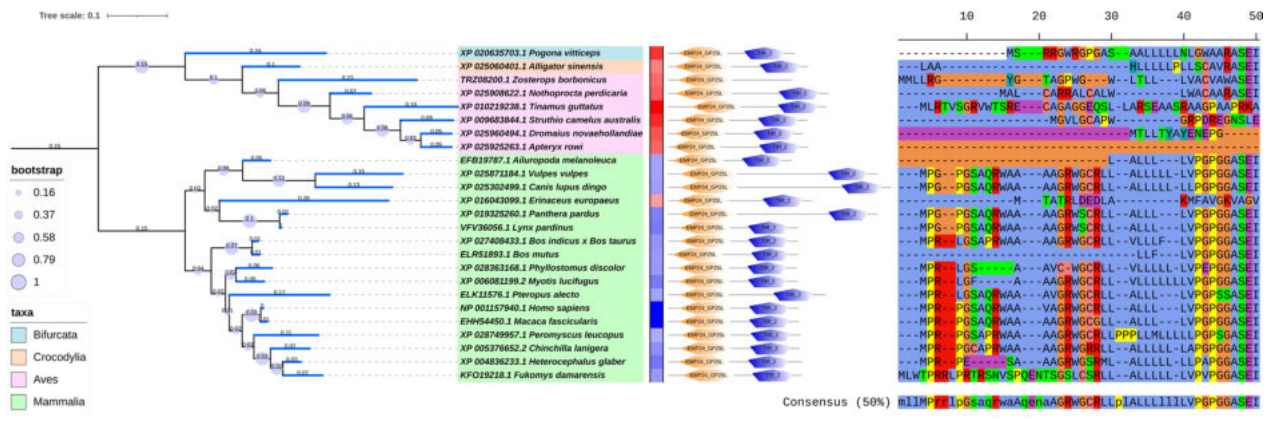


Fig. 4 A detailed phylogeny of TAG from different organisms, along with amino acid distance wrt *Homo sapiens*, domain architecture, and alignments across initial positions 1–20. The branches were colored to display the subfamily it belonged to and node IDs were colored based on the taxa group. A relative amino acid distance was shown with respect to Homo sapiens sequences by color gradient boxes (blue representing the minimum distance, white being the midpoint gradient, and red being the maximum distance). Also, the domain architecture was shown for each sequence. Besides these conserved motifs were shown in the order of their occurrence in an alignment form next to each node of the phylogeny

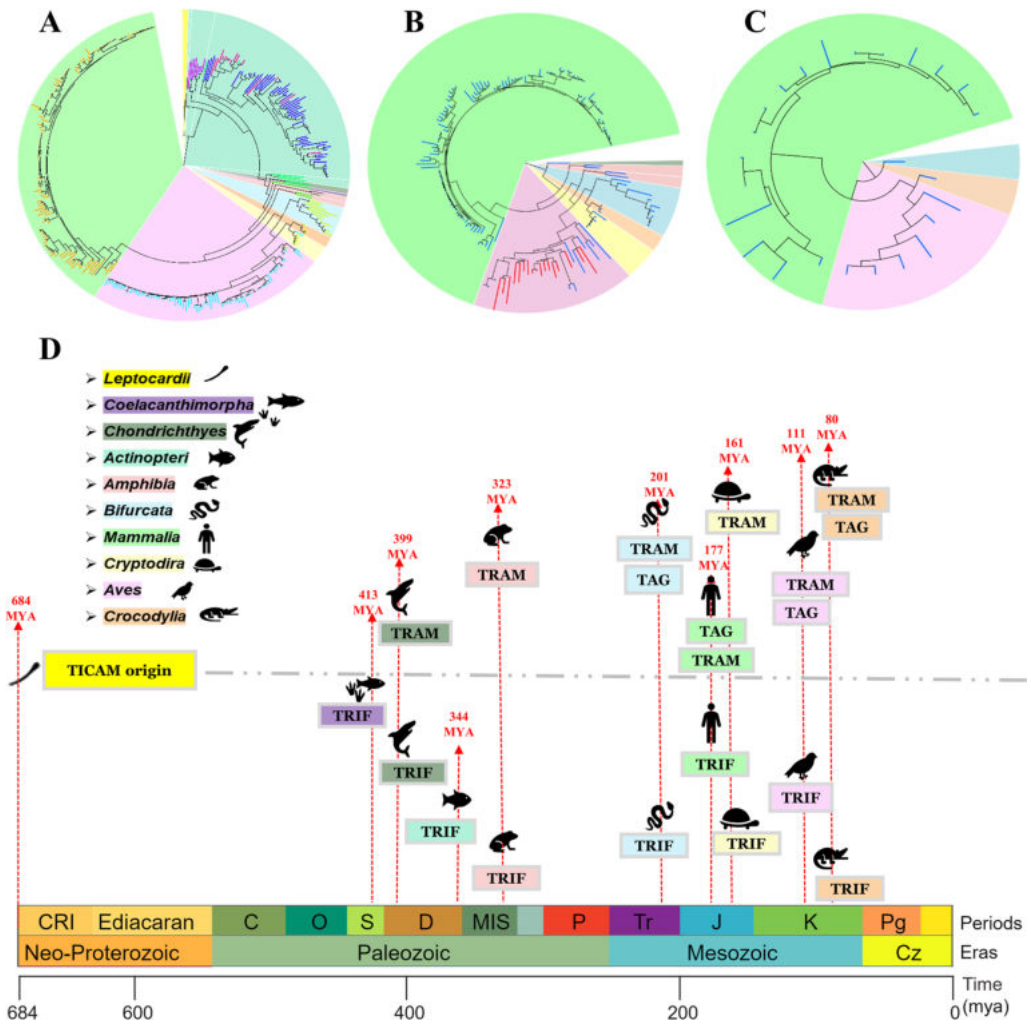


Fig. 5 The figure in the top panel shows a representative phylogeny for TRIF (**A**), TRAM (**B**) and TAG (**C**) respectively. The colored region represents organisms from different taxa and the node colour represents previously categorised ZEBRA subfamily. **D** This panel represents the evolutionary timeline from TICAM origin from Leptocardii and divergence into TRIF, TRAM and TAG. Here, TRAM was seen to be lost in Coelacanthimorpha and Actniopteri. The time period for the organisms were taken from Time Tree considering divergence between the hits. The order of the Eras with their Periods are Neo-Proterozoic Era (Cryogenian and Ediacaran), Paleozoic Era (Cambrian, Ordovician, Silurian, Devonian, Mississippian, Pennsylvanian, Carboniferous, and Permian), Mesozoic Era (Triassic, Jurassic and Cretaceous) and Cenozoic Era (Paleogene, Neogene and Quaternary)

positively selected sites was high and mostly amongst the non-structural regions. A list of positively selected sites with a probability above 0.95 has been provided in the Additional File 1: Figs. S4 and S5.

Discussion

The role of Toll-like receptors is important for innate immunity and these adaptors, apart from the conventional (MyD88 and TIRAP) ones, provide an alternate pathway for the production of inflammatory mediators. Previous studies have concentrated on the evolution of TIRs and adaptors involved in the MyD88-dependent pathway. However, the evolutionary lineage of adaptors that are known to operate in MyD88-independent pathway have not been studied in detail. The prime objective of our study has been to look into the divergence of TICAM to TRIF and TRAM. Unlike the MyD88 pathway, which seems to have emerged in early invertebrates, TRIF and TRAM related pathway trace back and emerge with the duplication event in vertebrates. A previously reported study has shown these adaptors in 25 metazoans [8], but details about the evolution of these proteins is lacking.

We performed our sequence searches in whole genomes, by employing all the TIR domains in the human genome as our queries. We later extended the searches to include TRIF and TRAM adaptors and accumulated their orthologues in our study. As expected, the

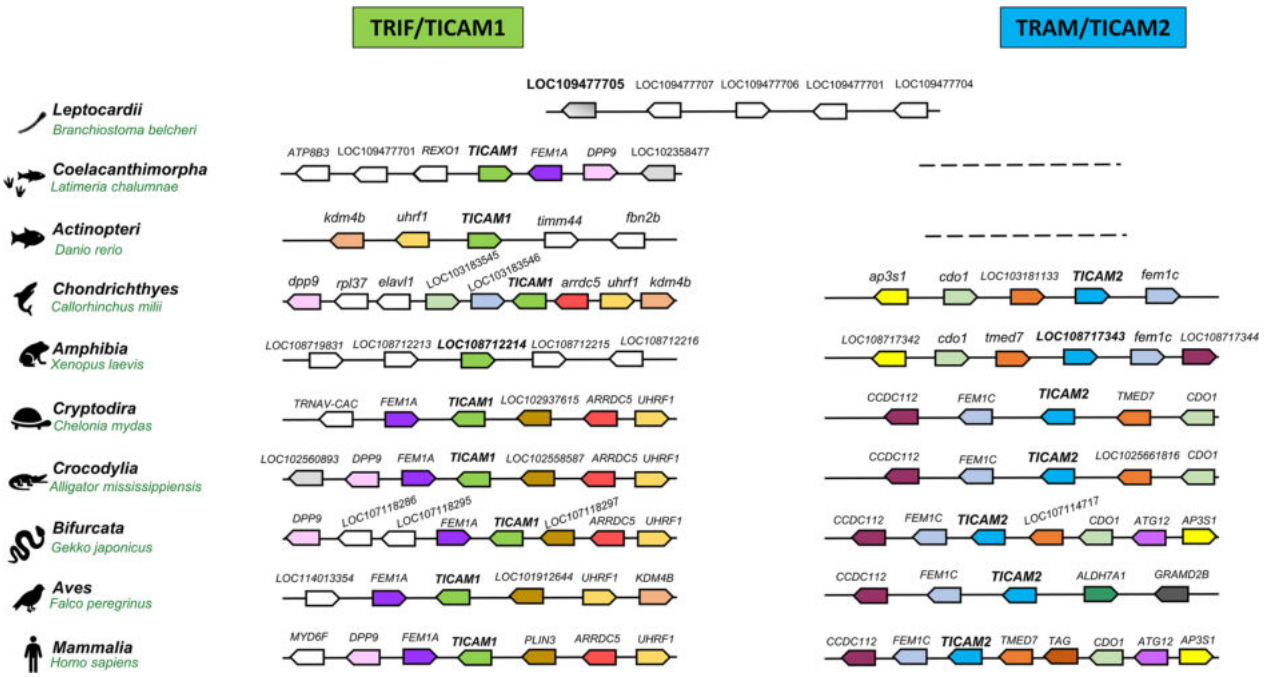


Fig. 6 Gene synteny for TRIF and TRAM protein from different taxon

conservation of Box1, 2, and 3 were found in these proteins, apart from that an additional motif was seen to be conserved. This was named Common4 and was the only motif present in TRIF and TRAM. This makes these adaptors peculiar to be studied. From the sequence-based phylogeny of the orthologs, *Leptocardii* appears as the oldest ancestor for these proteins. This was referred as TICAM in basal chordates for the MyD88 independent pathway [9]. Although they do not seem to have conserved domain structures or motifs responsible for signaling like *Homo sapiens.* Further, with our sequence search approaches, we found TAG (an isoform of TRAM with GOLD domain) which is seen in some species of *Bifurcata, Crocodylia, Aves,* and *Mammalia.* We also observed, through synteny analysis, that TRAM lineages and their immediate gene neighbours to be more highly conserved, as compared to TRIF where some ambiguity was seen for *Actinopteri* and *Amphibia.* TIR domains within TRAM are more conserved than in TRIF (Fig. 2a) and the variety of domain architectures are more in TRIF (Fig. 3).

Additionally, to extend the residue-based study on the full-length sequence of the protein, we performed a site model-based analysis to detect positively selected residues. Amino acid pertaining to non-domain regions were found to be positively selected in both the protein family. Overall, this study helps us in understanding a little more about these adaptors and their evolution (Additional files 6, 7, 8, 9, 10, 11).

Based on the presence of these adaptors amongst different taxa, we can explain the signaling of TLR3 and TLR4 based immune pathways. The presence of TRIF aids to boost the endosomal TLR3 pathway by recognizing double-stranded RNA, a major form of genetic information carried by viruses. Whereas the TRAM's presence along with TRIF can tell us about the functioning of the endosomal TLR4 pathway. The evidence of TRAM from *Chondrichthyes* makes us wonder about the need for the endosomal pathway of TLR4. This study can be further extended by a comparative analysis of orthologs among interacting partners of the signaling pathways. It will be interesting to track the presence or function of TRIF related inflammatory mediators, like IRF3 and IFN, in the older taxa. A structural determination for TICAM from ancestral taxa will help us know better about its function.

A potential limitation of the study would be the absence of high-quality data for ancestral taxa. Also, among groups like *Crocodylia, Amphibia,* and *Cryptodira,* whole genome sequence information is available only for a few species. To elucidate a taxa-specific evolutionary pattern and comment on group-specific evolution, we need to accumulate more data from multiple organisms. This will become better and more feasible with increasing numbers of whole genome sequencing of non-model organisms.

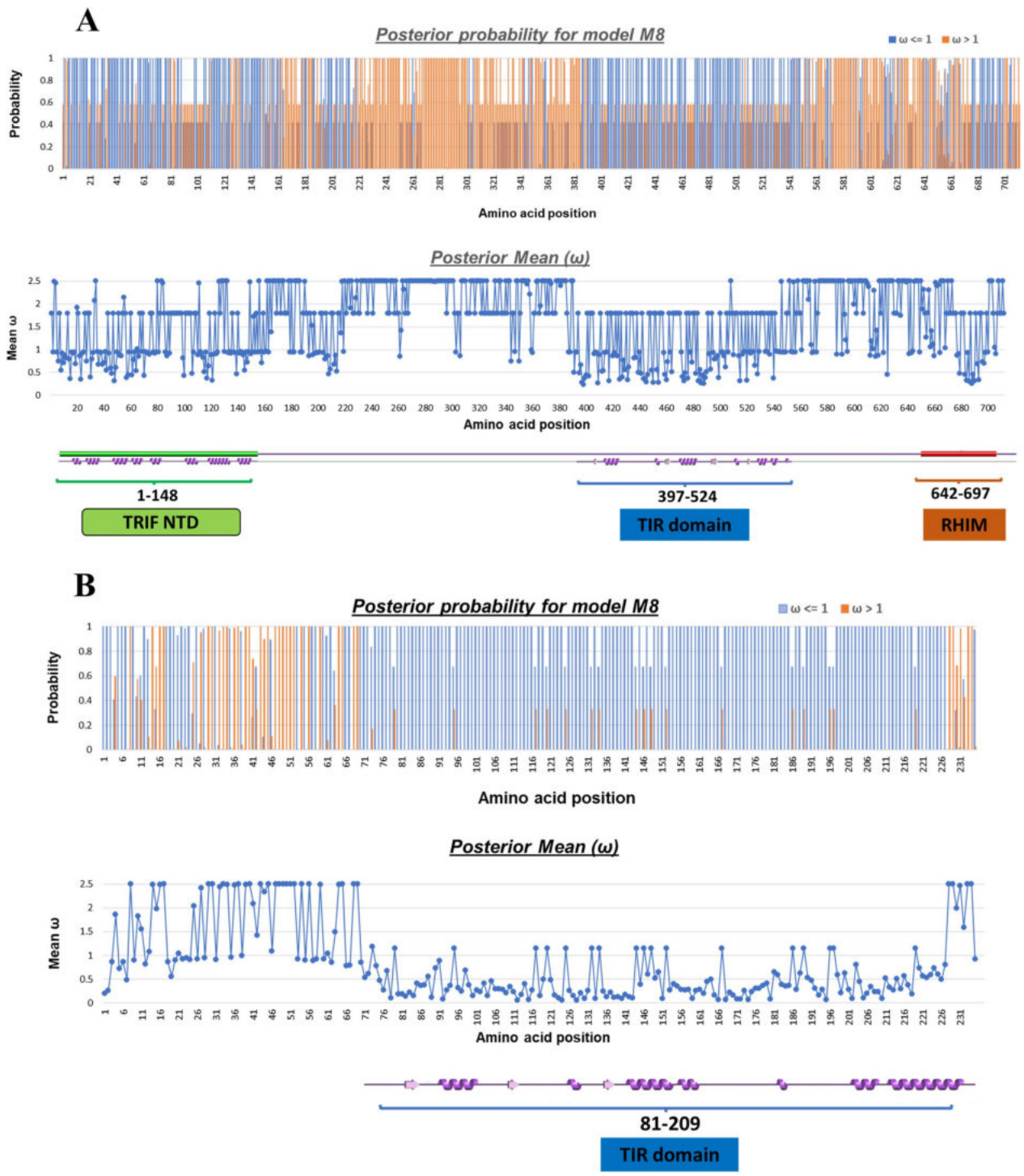


Fig. 7 The plots show two graph (i) posterior probability obtained from model M8 along the amino acid of the protein. Here ω < = 1 is shown in blue and ω > 1 is shown in orange. (ii) Shows the mean ω along the amino acid of the protein. **A** TRIF protein **B** TRAM protein

Conclusion

The current study is aimed at a systematic search and survey of TICAM orthologs in all the available genomes. We examined the domain architecture of genes that bear these domains and map the TICAM divergence to TRIF and TRAM across timescales. We also found evidence of the isoform of TRAM, TAG, and its presence is dated around 201 mya. Analysis of conserved, co-evolving residues and codon-based analysis was performed to identify positively selected sites amongst orthologs.

TRAM domains play important role in TLR4-mediated endosomal MyD88 independent pathway, and TRIF is the sole adaptor domain for the TLR3 signaling pathway involved in ds RNA recognition. Therefore, this study will help us know more about the immune systems in older taxa and how they evolved during evolution.

Methods

Query dataset and sequence search

We initially curated the human TIR domain sequences from Uniprot and filtered the hits to include Swiss-Prot reviewed sequences. Further, amongst these, those which followed PROSITE-ProRule annotation: PRU00204, were only selected. This PROSITE profile is specific to the pattern of TIR domains' scaffold that promotes assembly of signaling complexes via protein–protein interactions. Motif search was performed using MEME suite by classic search method and gapped local alignment was performed by the GLAM2 module using protein seqeuences as input structure [26–28]. Such human TIR sequences were used as a query to search for best homologs using BLAST [29] with an e-value of 10^{-10}. Best representatives of Primates, Odd-toe ungulate, Even-toe ungulate, Carnivore, Placental, Whale and dolphins, Chiropteran, Rodentia, Lagomorpha, and Insectivores were selected for each query. The motif search was extended to this set of proteins (275 proteins).

In order to search the homologues of TRAM and TRIF, we enriched our dataset to include organisms from 22 orders across the Mammalian class. TRAM and TRIF proteins of different organisms from the following orders were included. Monotremata (*Ornithorhynchus anatinus*), Didelphimorphia (*Monodelphis domestica*), Dasyuromorphia (*Sarcophilus harrisii*), Diprotodontia (*Phascolarctos cinereus*), Cingulata (*Dasypus novemcinctus*), Proboscidea (*Loxodonta africana*), Afrosoricida (*Echinops telfairi*), Tubulidentata (*Orycteropus afer afer*), Rodentia (*Marmota flaviventris*), Primates (*Pan troglodytes*), Eulipotyphla (*Condylura cristata*), Chiroptera (*Pteropus vampyrus*), Artiodactyla (*Sus scrofa*), Cetacea (*Balaenoptera acutorostrata scammoni*), Perissodactyla (*Equus caballus*), Carnivora (*Felis catus*), Lagomorpha (*Ochotona princeps*), Macroscelidea (*Elephantulus edwardii*), Scandentia (*Tupaia chinensis*), Dermoptera (*Galeopterus variegatus*), Sirenia (*Trichechus manatus latirostris*), Pholidota (*Manis javanica*) orders along with two model organisms (*Mus musculus, Macaca mulatta*) protein and *Homo sapiens*.

Later, the TRAM and TRIF protein from 25 organisms were used as a query to perform CS-BLAST [30] against the NR_Sept2019 database. The search was done using a python script to include all sequences individually with a very stringent e-value of 10^{-10} and up to 5 iterations after which it got saturated. The results from all CS-BLAST searches were combined and using an in-house script the output of CS-BLAST was converted into a tabular format. These results were further filtered using a query coverage filter of more than or equal to 50% and a sequence identity filter of more than or equal to 30% (keeping in mind the Twilight zone of protein sequence alignment) [31]. The list of hits obtained from multiple query search after considering the query coverage and sequence identity cut off for TRIF and TRAM orthologues are shown in Additional Files 14 and 15 respectively. The sequence of reference ID from these hits was retrieved using blastdbcmd module of BLAST version 2.9.0+. To remove the redundancy amongst sequences, CD-HIT was used to cluster the hits with 100% identity cutoff [32]. These hits were further divided into subfamilies based on clustering pattern by constructing sequence similarity network (SSN). These classifications were done using ZEBRA2, based on the CD-HIT clustering approach with a sequence identity threshold of 30% [10]. Based on subfamily clusters and phylogeny, proteins homologous were categorized into TRAM and TRIF family. Here we also used the protein sequence from PDB entries 2M1X and 2M1W, that corresponds to TIR domain region from TRIF and TRAM human protein respectively. We used these TIR sequence along with the full-length TRIF and TRAM orthologs sequence to see where does the TIR sequence clusters in the SSN.

Domain annotation

Domain architectures were searched for these sequences using Hmmscan modules from the HMMER suite package (version 3.1b2) against Pfam database (version 31.0) [14] with an e-value of 0.01 and inclusion threshold of 0.001. The output for domain architecture was generated using a python script. Sequences, which connected to the Pfam entry, TIR_2, were alone considered as true positives. Those sequences which connected to Pfam domains, TRIF-NTD, TIR_2, and RHIM domains were pooled as TRIF homologues. The domain architecture of each ortholog with the domain boundary and e-value are shown in Additional File 13. The domain architecture

diagram was made using My Domains from Expasy [33]. For Fig. 1, the domain architecture were taken from the HMMER webserver, by using hmmscan with e-value cut-off of 0.01 [34]. Sequences apart from these were looked at manually concerning their secondary structure and alignment profile. The secondary structure predictions were performed using PSIPRED [35]. Also, Ali2d scan was used for multiple secondary structure prediction and its results were visualized in 2dSS alignment viewer [36, 37].

Phylogeny tree construction

The fasta sequences were aligned using MUSCLE v3.8.31 [38] and a phylogenetic tree was constructed using the maximum likelihood method with 100 bootstraps in MEGAX [39]. The evolutionary timelineof the taxa were calibrated using TimeTree [24] based timeline for divergence of nodes. Amino acid distances were also calculated in MEGA using the p- distance method with uniform rate among sites, the homogenous rate among lineage, and pairwise deletion of gaps. Relative rate test was performed for sequences with higher branch lengths using Tajima relative rate test in MEGAX. This was calculated using 'Homo Sapiens' as a reference and the oldest descendant species as an outgroup to confirm if organisms with higher branch length evolve at the same rate or reject the null hypothesis. These data were added to the phylogeny and visualized using iTOL [40].

Conservation of Synteny

Genome assembly for one representative organism from each taxon was used to look into the gene neighbours. The sequence ID corresponding to TRAM and TRIF protein were searched in NCBI and the neighbours were found using the Genome Data Viewer [41].

Strength of evolutionary selection

Nucleotide codon sequences were retrieved for each protein ID using the Batch Entrez mode of NCBI [42]. They were further aligned using MUSCLE and then converted to codon alignment using PAL2NALv14 in PAML format [43]. These sequences were used along with a phylogeny tree in the CODEML module of the PAML4.9j package [44]. Site models (**M0**-one ratio, **M1**-nearly neutral, **M2a**-positive selection, **M3**-discrete, **M7**-beta, **M8**-beta, and $\omega > 1$, **M8a**-beta and $\omega = 1$) were used for these sequences and individual dN/dS ratio was also calculated for each branch. The hypothesis of different rates of evolution amongst different groups (*Leptocardii, Condricthyes, Coelacanthimorpha, Actinopteri, Amphibia, Cryptodira, Crocodylia, Bifurcata, Aves, Mammals*) were also tested. The phylogeny used considered branch lengths obtained from the Maximum Likelihood. Also, species from the *Leptocardii* group was included in both TRIF and TRAM phylogeny as an outgroup to root the phylogeny tree.

The parameters used for codeml run of site model were as shown in the table in Additional File 12.

The site model was performed for both protein family (TRAM and TRIF) using three models; **M3** versus **M0**, **M1** versus **M2a,** and **M7** versus **M8** for comparing positive selection. For identifying the potential amino acid residues that would have been under selection, we performed a Likelihood ratio test calculation for pairwise comparison of codon models using the Bayes empirical Bayes (BEB) method. The residues with probability above 0.95 and above were documented. Additionally, values for BEB or posterior probabilities of model **M8** for 11 site classes ($k = 11$) along the sequence of protein was plotted. The values from 11 site classes were grouped into two categories as $\omega > 1$ and $\omega < 1$. Also, the mean probabilities for each site were examined.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13062-022-00335-9.

Additional file 1: Figs. S1–S6.

Additional file 2: Amino acid sequence of TLR family proteins from 10 different orthologues.

Additional file 3: A maximum likelihood tree for TLR family proteins from 10 different orthologues.

Additional file 4: A detailed phylogeny with amino acid distance, domain architecture and motif conservation of TRIF orthologues.

Additional file 5: A detailed phylogeny with amino acid distance, domain architecture and motif conservation of TRAM orthologues.

Additional file 6: Protein Sequence ID's and respective organism name of TRIF orthologues.

Additional file 7: Multiple Sequence Alignment of TRIF orthologues in FASTA format.

Additional file 8: Protein Sequence ID's and respective organism name of TRAM orthologues.

Additional file 9: Multiple Sequence Alignment of TRAM orthologues in FASTA format.

Additional file 10: Protein Sequence ID's and respective organism name of TAG orthologues.

Additional file 11: Multiple Sequence Alignment of TAG orthologues in FASTA format.

Additional file 12: The parameters and values used for codeml run of site model.

Additional file 13: Table showing domain architecture of TRIF and TRAM orthologue along with domain boundaries and e-value.

Additional file 14: List of hits obtained from genome wide search of TRIF orthologues using CS BLAST after filtering.

Additional file 15: List of hits obtained from genome wide search of TRAM and TAG orthologues using CS BLAST after filtering.

Acknowledgements
The authors would like to thank NCBS (TIFR) for infrastructural facilities. They would like to thank Mr Adwait Joshi, Ms Teerna Bhattacharyya and Mr Vikas Tiwari for helpful discussions.

Author contributions
RS conceived the idea and provided further ideas for analysis. VS carried out the entire work of data curation and analysis. VS wrote first draft of the manuscript and RS improved the manuscript. All authors read and approved the final manuscript.

Funding
RS acknowledges funding and support provided by JC Bose Fellowship (SB/S2/JC-071/2015) from Science and Engineering Research Board, India and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021). RS would also like to thank Institute of Bioinformatics and Applied Biotechnology for the funding through her Mazumdar-Shaw Chair in Computational Biology (IBAB/MSCB/182/2022).

Availability of data and materials
The authors declare that [the/all other] data supporting the findings of this study are available within the article [and its supplementary information files].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]National Centre for Biological Sciences, GKVK Campus, Bellary Road, Bangalore 560065, India. [2]Institute of Bioinformatics and Applied Biotechnology, Bangalore 560100, India. [3]Molecular Biophysics Unit, Indian Institute of Science, CV Raman Road, Karnataka 560012 Bangalore, India.

Received: 20 April 2022 Accepted: 16 August 2022
Published online: 02 September 2022

References

1. Botos I, Segal DM, Davies DR. The structural biology of Toll-like receptors. Structure. 2011. https://doi.org/10.1016/j.str.2011.02.004.
2. Zughaier SM, Zimmer SM, Datta A, Carlson RW, Stephens DS. Differential induction of the toll-like receptor 4-MyD88-dependent and -independent signaling pathways by endotoxins. Infect Immun. 2005;73(5):2940–50. https://doi.org/10.1128/IAI.73.5.2940-2950.2005.
3. Xu Y, et al. Structural basis for signal transduction by the toll/interleukin-1 receptor domains. Nature. 2000. https://doi.org/10.1038/35040600.
4. Slack JL, et al. Identification of two major sites in the type I interleukin-1 receptor cytoplasmic region responsible for coupling to pro-inflammatory signaling pathways. J Biol Chem. 2000. https://doi.org/10.1074/jbc.275.7.4670.
5. Roach JC, et al. The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci USA. 2005. https://doi.org/10.1073/pnas.0502272102.
6. Roach JM, Racioppi L, Jones CD, Masci AM. Phylogeny of toll-like receptor signaling: adapting the innate response. PLoS ONE. 2013. https://doi.org/10.1371/journal.pone.0054156.
7. Liu G, Zhang H, Zhao C, Zhang H. Evolutionary history of the toll-like receptor gene family across vertebrates. Genome Biol Evol. 2019. https://doi.org/10.1093/gbe/evz266.
8. Wu B, Xin B, Jin M, Wei T, Bai Z. Comparative and phylogenetic analyses of three TIR domain-containing adaptors in metazoans: implications for evolution of TLR signaling pathways. Dev Comp Immunol. 2011. https://doi.org/10.1016/j.dci.2011.02.009.
9. Yang M, et al. Characterization of bbtTICAM from amphioxus suggests the emergence of a MyD88-independent pathway in basal chordates. Cell Res. 2011. https://doi.org/10.1038/cr.2011.156.
10. Suplatov D, Sharapova Y, Geraseva E, Švedas V. Zebra2: advanced and easy-to-use web-server for bioinformatic analysis of subfamily-specific and conserved positions in diverse protein superfamilies. Nucleic Acids Res. 2020. https://doi.org/10.1093/nar/gkaa276.
11. Enokizono Y, et al. Structures and interface mapping of the TIR domain-containing adaptor molecules involved in interferon signaling. Proc Natl Acad Sci U S A. 2013. https://doi.org/10.1073/pnas.1222811110.
12. Bateman A. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019. https://doi.org/10.1093/nar/gky1049.
13. Kaiser WJ, Upton JW, Mocarski ES. Receptor-interacting protein homotypic interaction motif-dependent control of NF-κB activation via the DNA-dependent activator of IFN regulatory factors. J Immunol. 2008. https://doi.org/10.4049/jimmunol.181.9.6427.
14. Finn RD, et al. Pfam: the protein families database. Nucleic Acids Res. 2014. https://doi.org/10.1093/nar/gkt1223.
15. Palsson-McDermott EM, et al. TAG, a splice variant of the adaptor TRAM, negatively regulates the adaptor MyD88-independent TLR4 pathway. Nat Immunol. 2009. https://doi.org/10.1038/ni.1727.
16. Jiang Z, Mak TW, Sen G, Li X. Toll-like receptor 3-mediated activation of NF-κB and IRF3 diverges at Toll-IL-1 receptor domain-containing adapter inducing IFN-β. Proc Natl Acad Sci U S A. 2004. https://doi.org/10.1073/pnas.0308496101.
17. Liu S, et al. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. Science. 2015. https://doi.org/10.1126/science.aaa2630.
18. Kaiser WJ, Offermann MK. Apoptosis induced by the toll-like receptor adaptor TRIF is dependent on its receptor interacting protein homotypic interaction motif. J Immunol. 2005. https://doi.org/10.4049/jimmunol.174.8.4942.
19. Rowe DC, et al. The myristoylation of TRIF-related adaptor molecule is essential for Toll-like receptor 4 signal transduction. Proc Natl Acad Sci U S A. 2006. https://doi.org/10.1073/pnas.0510041103.
20. Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-myristoylation of proteins: Refinement of the sequence motif and its taxon-specific differences. J Mol Biol. 2002. https://doi.org/10.1006/jmbi.2002.5425.
21. Verstak B, et al. The TLR signaling adaptor TRAM interacts with TRAF6 to mediate activation of the inflammatory response by TLR4. J Leukoc Biol. 2014. https://doi.org/10.1189/jlb.2a0913-487r.
22. Sullivan C, Postlethwait JH, Lage CR, Millard PJ, Kim CH. Evidence for evolving Toll-IL-1 receptor-containing adaptor molecule function in vertebrates. J Immunol. 2007. https://doi.org/10.4049/jimmunol.178.7.4517.
23. Anantharaman V, Aravind L. The GOLD domain, a novel protein module involved in Golgi function and secretion. Genome Biol. 2002. https://doi.org/10.1186/gb-2002-3-5-research0023.
24. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006. https://doi.org/10.1093/bioinformatics/btl505.
25. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 2005. https://doi.org/10.1371/journal.pbio.0030314.
26. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015. https://doi.org/10.1093/nar/gkv416.
27. Bailey TL, Elkan C, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol., 1994.
28. Frith MC, Saunders NFW, Kobe B, Bailey TL. Discovering sequence motifs with arbitrary insertions and deletions. PLoS Comput Biol. 2008. https://doi.org/10.1371/journal.pcbi.1000071.
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990. https://doi.org/10.1016/S0022-2836(05)80360-2.
30. Biegert A, Soding J. Sequence context-specific profiles for homology searching. Proc Natl Acad Sci. 2009;106(10):3770–5. https://doi.org/10.1073/pnas.0810767106.
31. Rost B. Twilight zone of protein sequence alignments. Protein Eng. 1999. https://doi.org/10.1093/protein/12.2.85.

32. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006. https://doi.org/10.1093/bioinformatics/btl158.
33. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003. https://doi.org/10.1093/nar/gkg563.
34. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46(W1):W200–4. https://doi.org/10.1093/nar/gky448.
35. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999. https://doi.org/10.1006/jmbi.1999.3091.
36. Zimmermann L, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol. 2018. https://doi.org/10.1016/j.jmb.2017.12.007.
37. Lotun DP, Cochard C, Vieira FRJ, Bernardes JS, 2dSS: a web server for protein secondary structure visualization. *bioRxiv*, p. 649426, https://doi.org/10.1101/649426. (2019).
38. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004. https://doi.org/10.1093/nar/gkh340.
39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018. https://doi.org/10.1093/molbev/msy096.
40. Letunic I, Bork P. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019. https://doi.org/10.1093/nar/gkz239.
41. Wolfsberg TG. Using the NCBI map viewer to browse genomic sequence data. Curr Protoc Bioinform. 2010. https://doi.org/10.1002/0471250953.bi0105s16.
42. "Batch Entrez," in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, 2006.
43. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006. https://doi.org/10.1093/nar/gkl315.
44. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007. https://doi.org/10.1093/molbev/msm088.

Publisher's Note

Computational and Structural Biotechnology Journal 21 (2023) 3680–3689

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Integrated approaches for the recognition of small molecule inhibitors for Toll-like receptor 4

Shailya Verma [a], Purushotham Reddy [a,b], R. Sowdhamini [a,c,d,*]

[a] National Centre for Biological Sciences (TIFR), GKVK campus, Bangalore 560065, India
[b] NMR-Analytical research and development, Aurobindo Pharma, Research center-II, Hyderabad, Telangana 502307, India
[c] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India
[d] Institute of Bioinformatics and Applied Biotechnology, Electronic City, 560100, India

ARTICLE INFO

Keywords:
Toll-like receptor
TIR domain
Autoimmune
Virtual screening
MD simulations
Reporter assay
NMR

ABSTRACT

Toll-like receptors (TLRs) are pattern recognition receptors present on the surface of cells playing a crucial role in innate immunity. One of the TLRs, TLR4, recognizes LPS (Lipopolysaccharide) as its ligand leading to the release of anti-inflammatory mediators as well as pro-inflammatory cytokines through signal transduction and domain recruitment. TLR4 homodimerizes at its intracellular TIR (Toll/interleukin-1 receptor) domain that helps in the recruitment of the TRAM/TICAM2 (TIR domain-containing adaptor molecule 2) molecule. TRAM also contains TIR domain which in turn, dimerizes and functions as an adapter protein to further recruit TRIF/TICAM1 (TIR domain-containing adaptor molecule 1) protein for mediating downstream signaling. Apart from LPS, TLR4 also recognizes endogenous ligands like fibrinogen, HMGB1, and hyaluronan in autoimmune conditions and sepsis. We employed computational approaches to target TRAM and recognize small molecule inhibitors from small molecules of natural origin, as contained in the Super Natural II database. Finally, cell reporter assays and NMR studies enabled the identification of promising lead compounds. Hence, this study aims to attenuate the signaling of the TLR4-TRAM-TRIF cascade in these auto-inflammatory conditions.

1. Introduction

Toll-like receptors (TLR) play a crucial role in the innate immune response by recognizing PAMPs (Pathogen Associated Molecular Patterns) and DAMPs (Danger Associated Molecular Patterns) as their ligands. The exogenous PAMP are derived from bacterial or viral sources, whereas the DAMP includes endogenous proteins [1]. There are 10 known TLRs in humans and each of them identifies a different ligand.

Some of these TLRs are present on the plasma membrane (1, 2, 4, 5, 6, 10), whereas others (3, 4, 7, 8, 9) are located on the endosomes. Unlike others, TLR4 is present on the plasma membrane as well as the endosome. Structurally, TLRs are type I transmembrane receptors with an extracellular domain that has Leucine Rich Repeat (LRR), followed by a transmembrane region and an intracellular TIR domain (Toll/interleukin-1 receptor) [2]. The cytoplasmic TIR domain is approximately 200 amino acids. It promotes the assembly of signaling complexes via protein-protein interactions. The primordial function of TIR domain containing proteins a per the InterPro Toll/interleukin-1 receptor homology (TIR) domain entry (IPR000157) is of a self-association-dependent nicotinamide dinucleotide (NAD(+)) cleaving enzyme (NADase) activity that cleaves NAD(+) into nicotinamide (Nam) and ADP-ribose (ADPR), cyclic ADPR (cADPR) or variant cADPR (v-cADPR), with catalytic cleavage executed by a conserved Glutamic acid [3–5].

On identification of the ligands, the TLR forms homo or hetero dimers and recruits the adaptor molecules like MyD88, TIRAP, TRIF (also called TICAM1 and is specific to TLR4 and TLR3), TRAM (also called TICAM2 and is exclusive for TLR4 signaling pathway). The MyD88 & TIRAP dependent pathway ultimately releases pro-inflammatory cytokines, NF-κB, Tumor Necrosis Factor alpha (TNF-α), interleukin (IL-)1β, IL-6, and chemokines like Monocyte Chemoattractant Protein 1 (MCP-1), Macrophage Inflammatory Protein 3α (MIP-3α), and IL-8. Whereas TRIF & TRAM pathway releases both pro-inflammatory cytokines as well as anti-inflammatory mediators like Interferon Regulatory Factor 3 (IRF3), beta interferon (IFN-β), delayed NF-κB activation, type 1 IFN-α/β, IFN-α-inducible protein 10 (IP-10), MCP-5, RANTES, and nitric oxide release [1,6]. A schematic representing the TLR signaling pathway with adaptors and mediators is shown in Fig. 1.

* Corresponding author at: National Centre for Biological Sciences (TIFR), GKVK campus, Bangalore, 560065, India.
E-mail address: mini@ncbs.res.in (R. Sowdhamini).

https://doi.org/10.1016/j.csbj.2023.07.026
Received 21 March 2023; Received in revised form 8 July 2023; Accepted 19 July 2023
Available online 22 July 2023

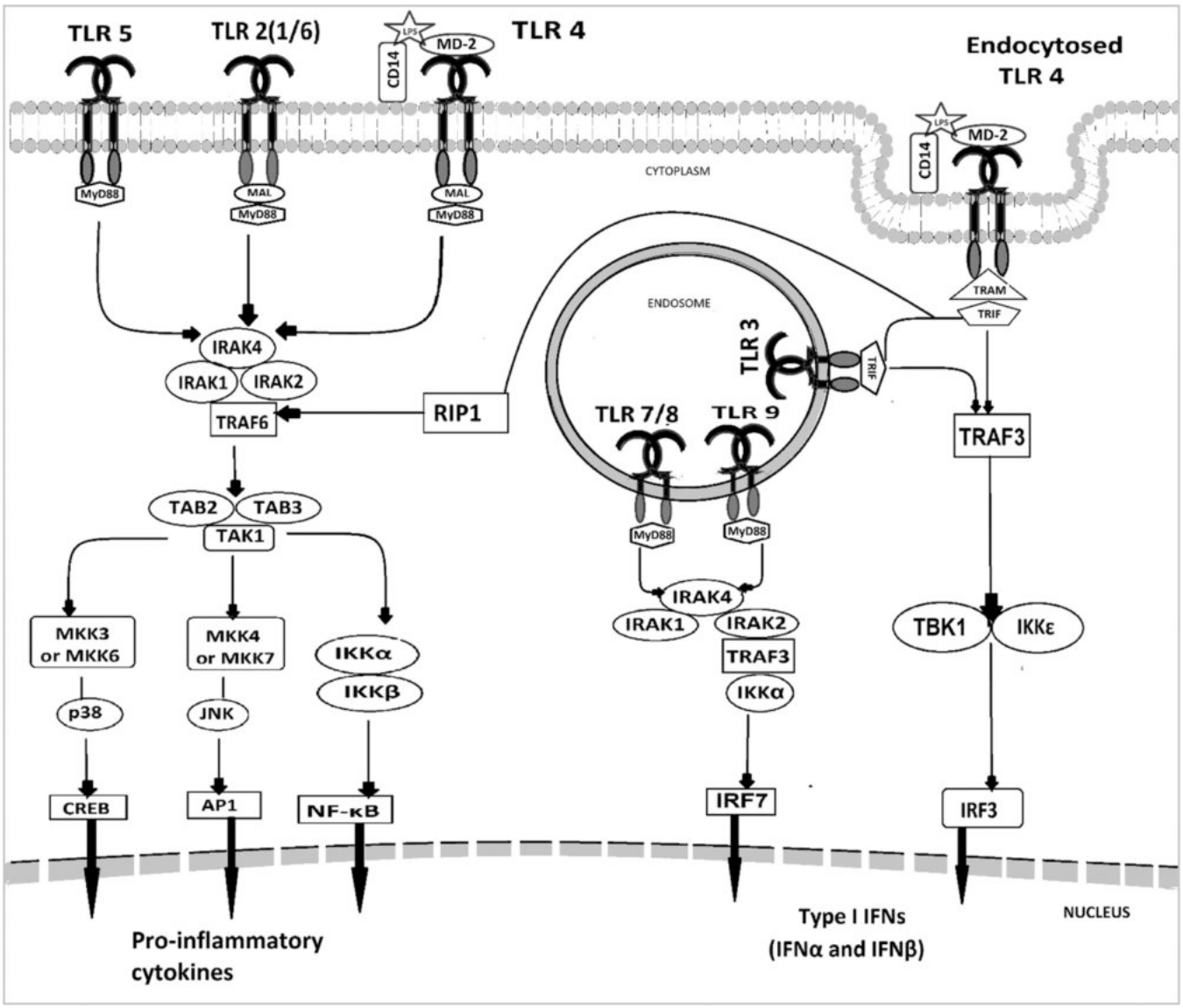


Fig. 1. Schematic of Toll-like receptor downstream signaling pathway.

TLR4 exhibits both MyD88-dependent as well as TRAM dependent pathway. This TLR recognizes Lipopolysaccharides (LPS) as ligands and upon binding to hexa-acylated LPS, CD-14 transfers it to MD-2 causing the dimerization and then recruits the adaptor molecules to control the downstream processes [7]. This study aims to target TLR4-TRAM signaling. Upon recognition of ligand, TLR4 forms a homodimer of the TIR domain with the help of its BB-loop interaction and associates in a twofold symmetry [8]. These BB-loops form the interface which helps in the recruitment of the TRAM adaptor molecule, which also forms dimers symmetrically in a twofold axis along its BB-loop. The TRAM dimer recruits the TRIF through interaction between its EDD & TS site and RK & QI site of TRIF. These crucial residues have been found through mutational studies [9]. These residues are very important for interactions, so they were used in selecting the docking site.

The role of TLR4 is seen evident in sepsis, inflammatory conditions as well as cancer. Also, if the LPS is not controlled properly it can lead to septic shock syndrome which is a major cause of death in patients in the ICU [10]. Besides, these there are several literature reports that support the instances, where sepsis occurs as the worst outcome of host-pathogen interactions, that is a leading cause of death [10,11]. Sepsis is developed equally by the gram-positive and gram-negative bacteria where exaggerated immune trigger leads to multi-organ failure and septic shock. These gram-negative bacteria contain the LPS that contributes to sepsis development, also the septic shock is due to the amplified body immune response rather than the infection [12]. Additionally, there has been various TLR inhibitors clinically evaluated to manage the sepsis conditions [13]. However, all the currently investigated inhibitors are targeting at the level of TLRs. Our objective through this study is to target such overamplified signaling in case of TLR4, with the help of the TRAM adaptor protein.

The aim is to search for small molecule therapeutic that attenuates or enhances TLR4-mediated signaling. The approach we have taken here is to search for a novel compound that targets TRAM and hinders its role as an adaptor.

A previous study from the lab had used TRAM as a target [14] since A46-VIPER motif from vaccinia virus producing protein is known to disrupt TRAM-TLR interactions. The VIPER peptide representing the VIPER motif is now used in the field as TLR4 inhibitor [15]. It binds to the TIR domain of the adaptor proteins thereby masking the binding sites on TRAM and MAL and inhibiting the downstream signaling. Also, another study suggests these VIPER motif being particularly important for TRAM antagonism, as seen in case of Mal-deficient immortalized mouse bone marrow-derived macrophage (iBMDMs) [16]. These peptide do not interacts with MAL (MyD88 mediated pathway) in vitro [17].

Although the exact binding site is not fully elucidated, in our previous attempt we used homology modeling and molecular docking approach to identify potential binding sites on TRAM (focusing on BB-loop and C^{α} helix). In the previous study ligands were screened from the ZINC database and top 12 compounds were selected as potential hits [14]. In this study, we present the findings of virtual screening using small molecules of natural origin from Supernatural Database [18]. We have also validated the small molecules using cell-based assay and NMR titrations and report two promising candidate small molecules which bind better than VIPER motif.

2. Materials and methods

2.1. Virtual screening pipeline

A virtual screening pipeline was followed using Glide [19] from Schrodinger Suite. The protein coordinates were retrieved from PDB (PDB ID: 2M1W, H117C) and the ligand library was prepared from the Supernatural II database which contains all compounds of natural origin (3,25,287 small molecules) [18]. All small molecules and proteins were prepared and optimized at pH = 7.4 using Ligprep and the Protein preparation module of Glide. Ligprep also considers the tautomers and all possible conformations while preparing the ligand library, thereby 4, 31,685 ligands were generated. Further Site Map was used to predict the potential binding site on the protein [20]. Docking was performed through a series of hierarchical filters i.e. HTVS mode (high-throughput virtual screening) for efficiently enriching million compound libraries, to the SP mode (standard precision) for reliably docking tens to hundreds of thousands of ligands with high accuracy, to the XP mode (extra precision) where further elimination of false positives is accomplished by more extensive sampling and advanced scoring, resulting in even higher enrichment. Each step proceeded with the top 10 % from the previous one. The HTVS filtered the library to around 3,40,000, SP filtered to around 34,000 and, XP filtered to around 3000 small molecules respectively. Latter QikProp descriptors were used to select compounds with 95 % drug-like properties using ADME criteria (absorption, distribution, metabolism, and excretion) [21]. Binding energy was calculated using MM-GBSA (molecular mechanics energies combined with generalized Born and surface area continuum solvation) from the Prime module [22]. The formula for ΔG calculation by MMGBSA is as follows: ΔG (bind) = E_complex (minimized) - (E_ligand (minimized) + E_receptor (minimized). After this step top 3000 compounds, in each case, were clustered using CANVAS hierarchical fingerprinting and Schrodinger's interaction-based fingerprint [23,24]. MOLPRINT2D fingerprint method was used and clustering was tried using all available options (Average, Ward, Single, Centroid, Mc Quitty, Complete, Weighted Centroid, Flexible β and Schrodinger). Average and Flexible β methods were chosen as these two methods gave the best clusters. List of clusters centroids from these methods along with top 10 MMGBSA ΔG (bind) score compounds were combined and the top score compounds were then sorted and list of purchasable ones were made.

2.2. Molecular dynamics simulations

Further Molecular Dynamics Simulations were performed using Desmond for protein-ligand complex initially for 20 ns. We performed an initial all-atom Molecular Dynamics Simulations (MD) using Desmond (Schrodinger suite) [25]. The Protein-ligand complex with the best scores was selected after XP docking. The system builder was used with the TIP4P solvent model and default boundary conditions [26]. Using the OPLS3e force field and adding ions for neutralization the set-up file was prepared. This was further taken forward for simulation of 20 ns at a temperature of 300 K and Pressure of 1.013 bar. Further, the whole system was relaxed through energy minimization before simulation and detailed interaction analysis was performed. For ligands that showed good results in reporter assay, the MD simulations were extended to 100 ns.

2.3. Small molecule ligands for experimental validation

Computationally predicted top compounds were searched for their availability and ease of purchase. In total 10 compounds were ordered from Molport (https://www.molport.com). The details of compound number with their Molport ID are as follows: Compound 1(MolPort-001–740–483), Compound 2 (MolPort-021–804–591), Compound 3 (MolPort-001–740–229), Compound 4 (MolPort-001–741–384), Compound 5 (MolPort-005–945–341), Compound 6 (MolPort-002–515–588), Compound 7 (MolPort-000–779–136), Compound 8 (MolPort-002–132–868), Compound 9 (MolPort-001–740–491), Compound 10 (MolPort-046–509–064).

We also used VIPER and CP7 peptides as a positive and negative control, for cell-based reporter assay. These peptides were commercially available and were purchased from Novus Biologicals (https://www.novusbio.com/products/tlr4-inhibitor_nbp2–26244) [27]. We used VIPER peptide (KYSFKLILAEYRRRRRRRRR**)** as a known positive control [16,14] that is known to inhibit the downstream signaling thereby, lower the SEAP production. And along with it the CP7 peptide (RNTISGNIYSARRRRRRRRR) is a known negative control that does not affect the downstream signaling. These peptides are well established in the field and is commercially used as TLR4 inhibitor peptide set. The VIPER peptide for the NMR experiment was customized and ordered from Lifetime (https://www.lifetein.com/). The sequence from the VIPER motif used for the peptide was KYSFKLILAEY [27].

2.4. Cell-based SEAP reporter assay

HEK-Blue™ TLR4 cells (Invivo Gen- https://www.invivogen.com/hek-blue-htlr4) were used, to study the stimulation of TLR4 by monitoring the activation of NF-κB and AP-1. These cells have been also used for several similar studies for finding small molecule drug molecules [28–30].

These cells are already co-transfected by human TLR4, MD-2, and CD14 co-receptor genes, and an inducible SEAP reporter gene. The stable cell lines were grown in media containing DMEM, Fetal bovine serum (10 X FBS), Penicillin-Streptomycin-Glutamine (1 X PSG), and 1 X HEK-Blue™ Selection antibiotics. Cells were grown to 80 % confluency and then detached and counted. Cells were added to freshly prepared detection media with ~140,000 cells per ml. 180 μl of cell suspension solution (~25,000 cells) were added to each well of a 96-well plate. Further, we added the water/ peptide/ compounds (effective concentration of 100uM per well) and incubated it for one hour at 37 °C in 5 % CO2.

Lipopolysaccharide purified from Escherichia coli K12 (LPD-EK) was purchased from Invivogen (cat code: tlrl-peklps) for using as PAMP in TLR4 pathway activation [31]. This is an ultrapure form of LPS extracted by successive enzymatic hydrolysis step and purifies using phenol-TEA-DOC extraction protocol [32]. These do not contain lipoprotein, therby it only activated TLR4 pathway in the cells. LPS dose-dependent assay was performed initially to check for optimum LPS needed for TLR4 activation and SEAP production. Further, 10 ng/ml was found to be the optimum amount of LPS, and it was used as a ligand in screening compounds. For screening the compounds, we prepared a 100 mM stock of each compound in DMSO, and then prepared a working solution of 10 mM in water. We first used an effective concentration of 100 μM of each compound and checked for % LPS-induced SEAP production ((Absorbance LPS) / (Absorbance Control) * 100). This was further incubated for 24 h and then the SEAP produced was calculated. For this assay, we have also used positive (VIPER peptide), and negative controls (CP7 peptide) [27]. Compounds that performed better than positive control in the % LPS-induced SEAP production screening were further studied by performing the dose-dependent assay.

Here, HEK-Blue™ TLR4 cells were treated with various

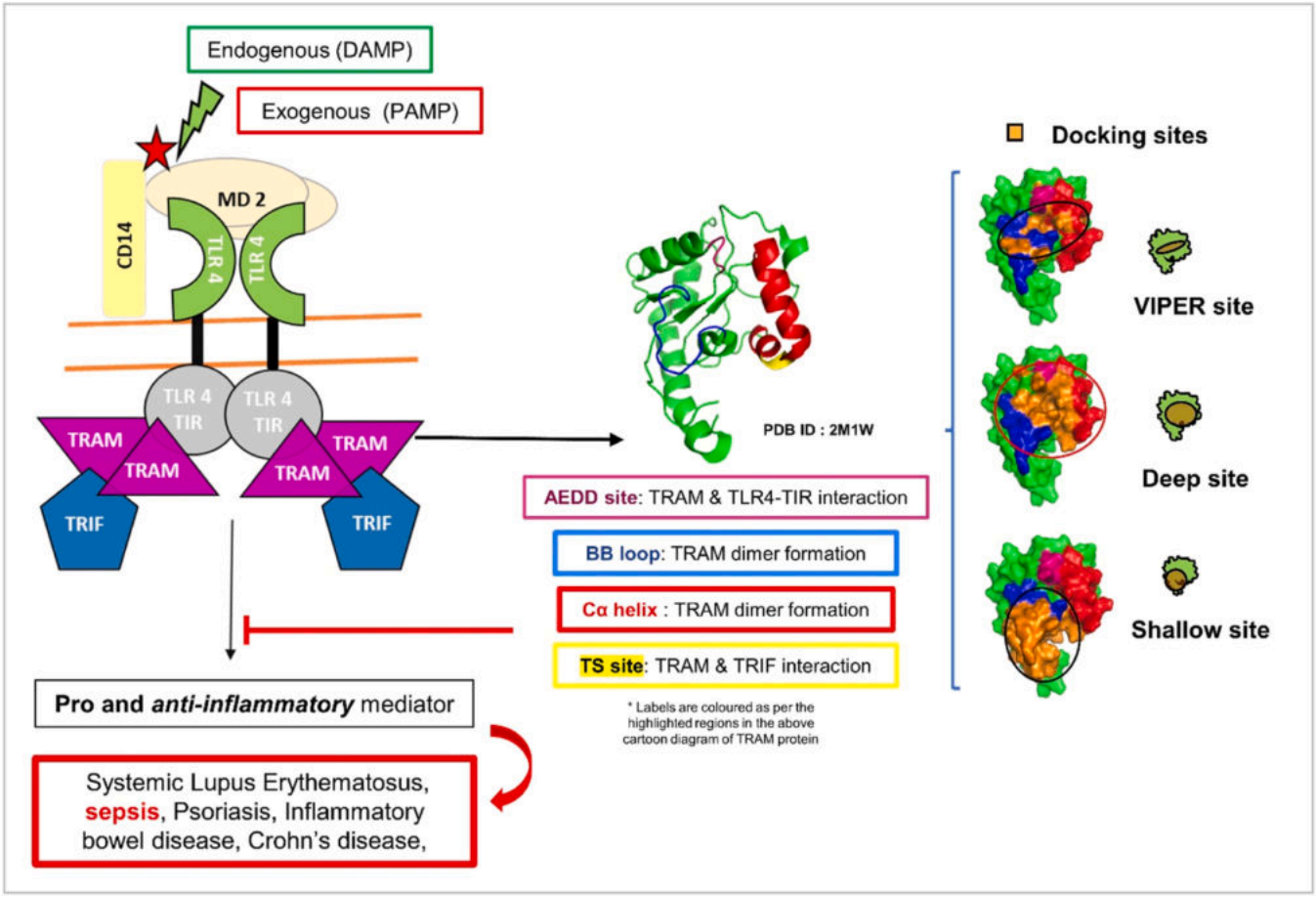


Fig. 2. A schematic of TRAM mediated TLR4 pathway leading to overproduction of inflammatory mediators in presence of excess PAMP and DAMPS causing autoimmune conditions. Highlighted residues on the TRAM cartoon structure depict AEDD site, BB loop, Cα helix, and TS site. The docking sites (VIPER, Deep, and Shallow) are shown on the surface cartoon diagram with important residues highlighted in different colors.

concentrations of compounds in serially diluted manner. Concentrations used were 100 μM, 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.125 μM, 1.5625 μM, 0.78125 μM. Dose-response curves were made for this concentration. The experiment was performed independently three times and % LPS induced SEAP production was calculated with respect to control. IC_{50} values (half maximal inhibitory concentration) were calculated for these by using Graphpad Prism 9 software (Y=Bottom + (TopBottom)/(1 +10^((LogIC$_{50}$-X) *HillSlope)).

We also performed the MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) tetrazolium reduction assay to check the viability of the cells. HEK-Blue™ TLR4 cells were initially seeded into 96 well plates, it was incubated for 48 h for cells to become adhesive and gain confluency for the experiment (40–60 %). After 48 h cells were treated with compounds in serially diluted concentrations (100 μM, 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.125 μM). Cells with treatment was incubated for another 24 h. Subsequently, MTT (0.5 mg/ml) solution was added to each well and incubated for 4 h. Next, incubation formazan crystal was dissolved using isopropanol and then absorbance was taken at 580 nm. The experiment was repeated in triplicates and cell viability was measured in presence of compounds.

2.5. Protein purification

The human TRAM gene (pcDNA3-TRAM-CFP, Plasmid #13027) was purchased from addgene (https://www.addgene.org/13027/#?). The region encoding the TIR domain (amino acid residues 70–235) was cloned into the pET28a (+) vector with restriction enzymes (*Nde*I and Hind lll) and expressed as a His6-tagged fusion protein at the N terminus. pET28a (+)-TRAM-TIR cloned vector was then transformed into Escherichia coli BL21(DE3). The transformed E. coli BL21(DE3) cells were cultured in N^{15} labeled ($N^{15}H_4Cl$) minimal media at 37 °C until the OD_{600} reached ~0.6–0.7. Composition of M9 Minimal media is shown in Supplementary Figure 11. The cells were induced with 1.0 mM IPTG and then cultured for 5 h at 37 °C. The harvested cells were resuspended in lysis buffer [50 mM Tris, 300 mM NaCl,1 mM PMSF, DNase 0.1 mg/ml, 5 mM MgCl2, pH 8.0] and lysed using sonication. TRAM-TIR fused with His6-tag was first purified by Ni2 + -NTA affinity chromatography (HisTrap H- 5 ml) with step elution of TRAM-TIR at 200 mM Imidazole concentration. The purified protein was concentrated and buffer exchange was done to 50 mM Tris, 150 mM NaCl, 1.5 mM CaCl2, pH 8.0. The concentrated protein was then treated with thrombin (SRP6556–1KU) for cleavage of His tag and after 4 h of cleavage at RT, 1 mM PMSF was added to stop the reaction and finally, size exclusion chromatography (Superdex 75 16/600) was done in 50 mM HEPES, 5 mM DTT, 0.1 mM EDTA, pH= 7.4. The labeled protein was then used for conducting NMR titration experiments. The protein purification profile with steps of Affinity and size exclusion chromatography is shown in Supplementary Figure 12.

2.6. NMR titration experiment

The compounds were diluted to a concentration of 5 mM in protein buffer and titrated with protein in the following ratios (1:0, 1:1, 1:2, 1:4). HSQC data were collected for all the runs for compounds 2, 3, 4, 9, 10 and VIPER peptide.

Chemical shift perturbations were calculated using the formula: CSP

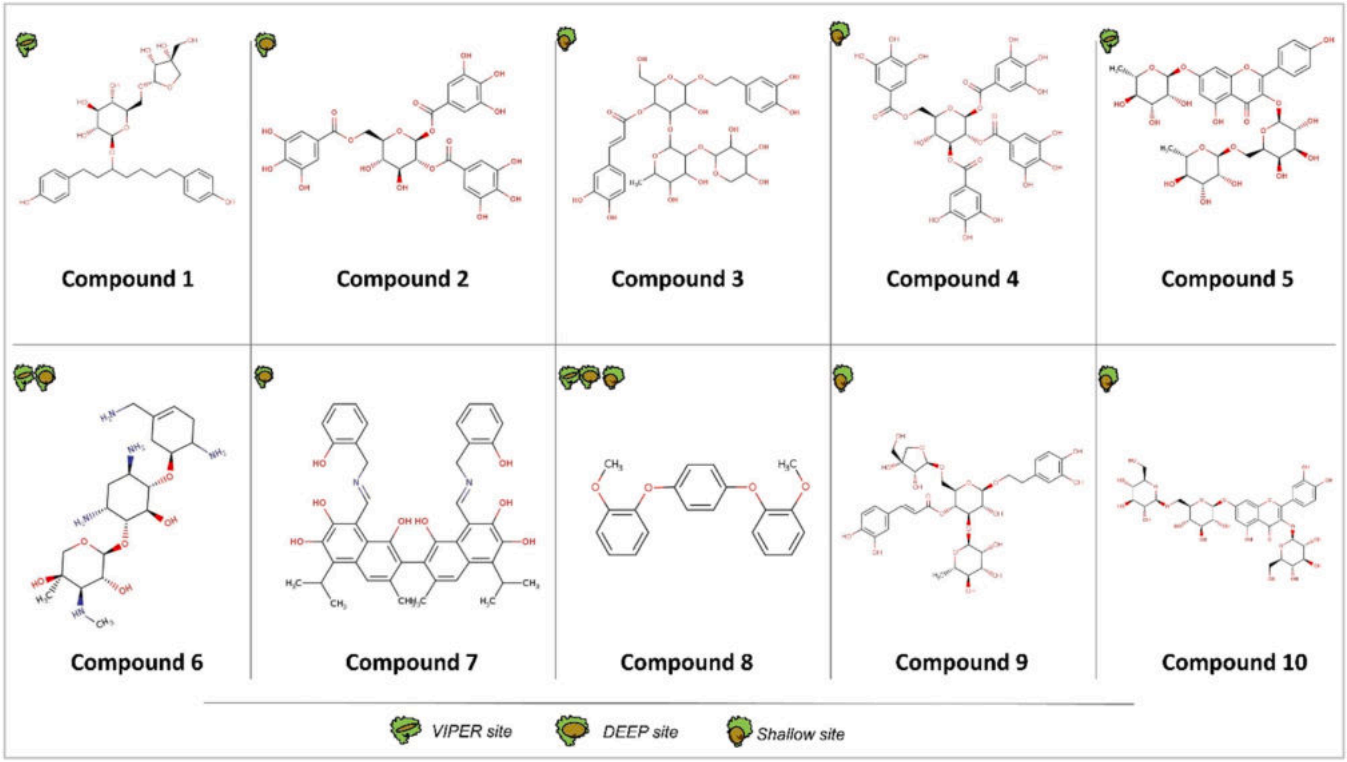


Fig. 3. Chemical structure of top ten ligands selected from Virtual screening along with their binding sites.

$= [(\Delta\delta HN)^2 + (\Delta\delta N)^2/25]^{1/2}$ [33]. All NMR spectra were acquired at 25 °C on 800 MHz/600 MHz Bruker Avance III spectrometers, processed using NMRPipe, and analyzed using Spar [34,35].

3. Results and discussions

3.1. Putative binding sites on the TRAM protein

Potential binding sites on TRAM were chosen such that it interferes with the binding of TRAM to TLR4-TIR or TRIF and thereby abrogating the downstream signaling. The residues of TRAM that were part of binding pockets for the VIPER peptide were selected as one of the putative binding sites – we refer this as "VIPER site" (Fig. 2). We also used SITEMAP within Schrodinger package to predict other potential binding sites. Two other sites were selected using SITEMAP which are named "Deep" and "Shallow" sites (as marked in Fig. 2) on the basis of ligand binding score, docking score, size and volume of the pocket (listed in Supplementary Figure 1).

These pockets also contain structurally important residues like AEDD and TS site, that are involved in upstream TLR4-TIR and downstream TRIF interaction with TRAM respectively. Along with these sites, residues from the BB loop and Cα helix that are important for dimer formation are also included in the docking sites.

3.2. Virtual screening against TLR4 adaptor molecule TRAM-TIR

Virtual screening was performed at pH = 7.4 using the three docking pockets of TRAM protein against the Supernatural II database that consists of 3,25,287 ligands [18]. With a rigorous screening of the naturally derived small molecule library using steps of virtual screening (HTVS, SP, and XP) we filtered the dataset to a smaller number, around 3500 small molecules. Further with the energy scores, and ADMET properties, we clustered top hits using various clustering approaches to get the best representative of the small molecules. Finally, the top ten hits were selected and extensively searched in literature for their known properties. The curated list of the top ten compounds is provided in Fig. 3. Other details like SMILES format, docking scores, binding energy values, and availability of these compounds are mentioned in Supplementary Figure 2.

Most of the small molecules are associated with anti-inflammatory, antipyretic and antioxidant properties. Some are of Asian origin and have been documented as regularly used in traditional Chinese medicines. In particular, Compound 1 is found in plants like *Betula platyphylla* (common name: Asian white birch) and *Acer maximowiczinum* (common name: Nikko maple, native to China and Japan). It is an HDAC6 inhibitor that synergistically enhances anticancer activity [36]. Compound 2 obtained from the cell culture of *Cornus Officinalis* (common name: Japanese cornel dogwood) which has antiviral properties and works as an inhibitor of HCV NS3 Protease and Vertebrate Squalene Epoxide [37]. Compound 3 is found in the stem bark of *Magnolia officinalis* (common name: Chinese Magnolia) has antioxidant properties and is used in traditional Chinese medicines [38]. Compound 4 is found in fruits like *Punica granatum* (pomegranate) and *Mangifera indica* (mango). It shows anti-microbial, anti-inflammatory, antioxidant, and anti-carcinogenic properties [39]. Compound 5 found in *Vinca erecta* (European periwinkle) and Robinia pseudoacaia (White Locust Tree) mentioned in traditional Chinese medicine has anti-bacterial, anti-inflammatory, and diuretic properties [40]. Compound 6 has a scaffold that was an integral part of many other top hits natural compounds. Compounds 7 and 8, albeit recognised from the Supernatural II database, do not have evidence of their use in literature. Compound 9 is found in leaves of *Lamioplomis rotate (Benth.) Kudo* (known as Dabuba in Tibet and China) and is used as a Chinese folk medicine plant for an anti-inflammatory condition like sepsis [41]. It is known to inhibit TNF-alpha, IL-6, and IκB and modulate NF-κb. Compound 10 is found in the seed of *Desuranina Sophia* (flixweed or tansy mustard) and has anti-inflammatory, antipyretic, analgesic, antioxidant, and anthelmintic properties [42].

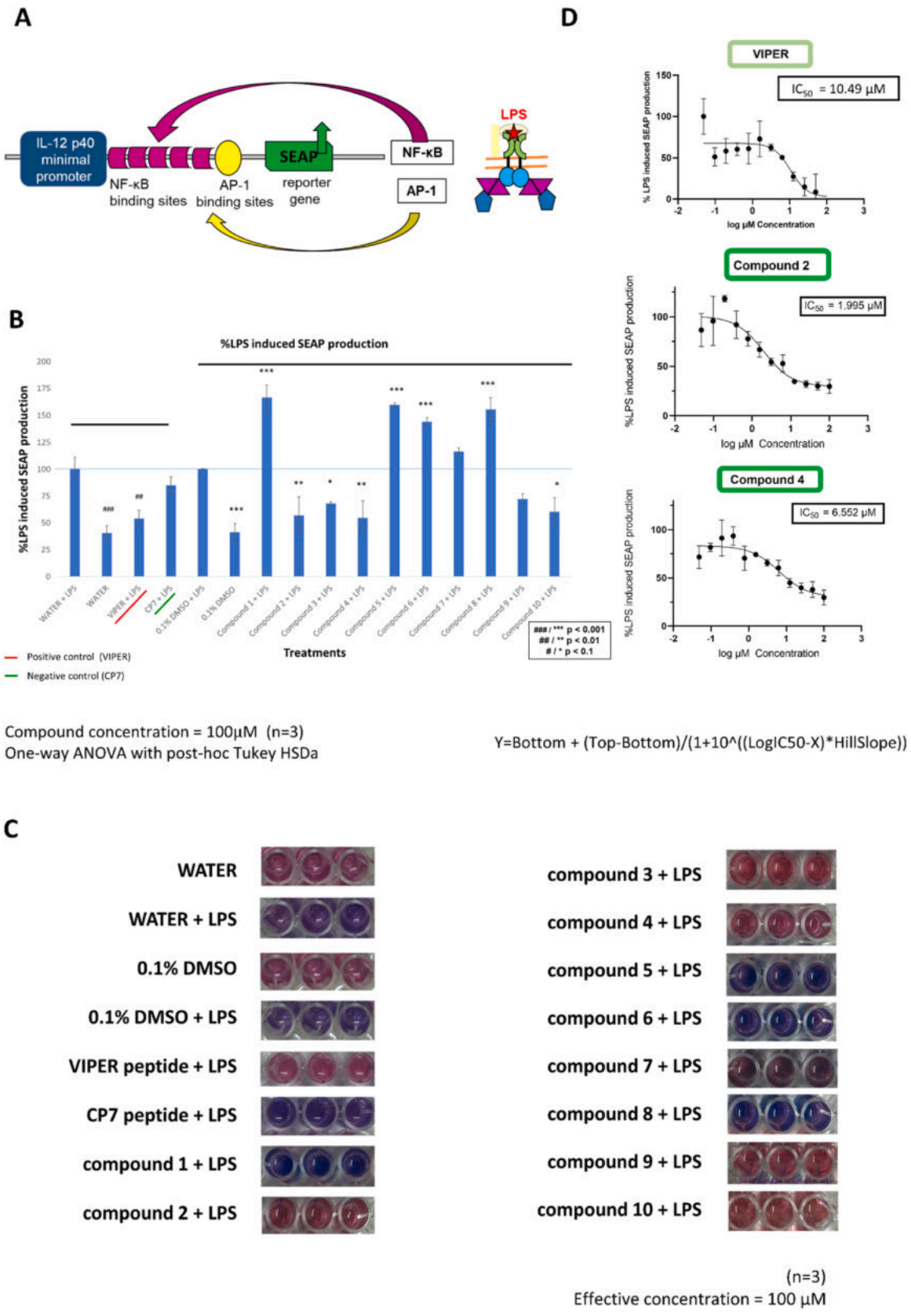


Fig. 4. A) Schematic depicting the activation of TLR4 complex in presence of LPS and thereby producing NF-kB and AP-1. These transcription factor bind to their respective binding sites downstream of the SEAP reporter gene. These lead to SEAP production by the HEK Blue TLR4 cells. B) Graph shows the % LPS induced SEAP production in presence of compounds at 100 µM concentration. C) Observed colour change while screening of compounds using cell-based reporter assay to validate the SEAP production. D) Dose-dependent assay for VIPER peptide, Compounds 2, and 4 to calculate the IC_{50} values.

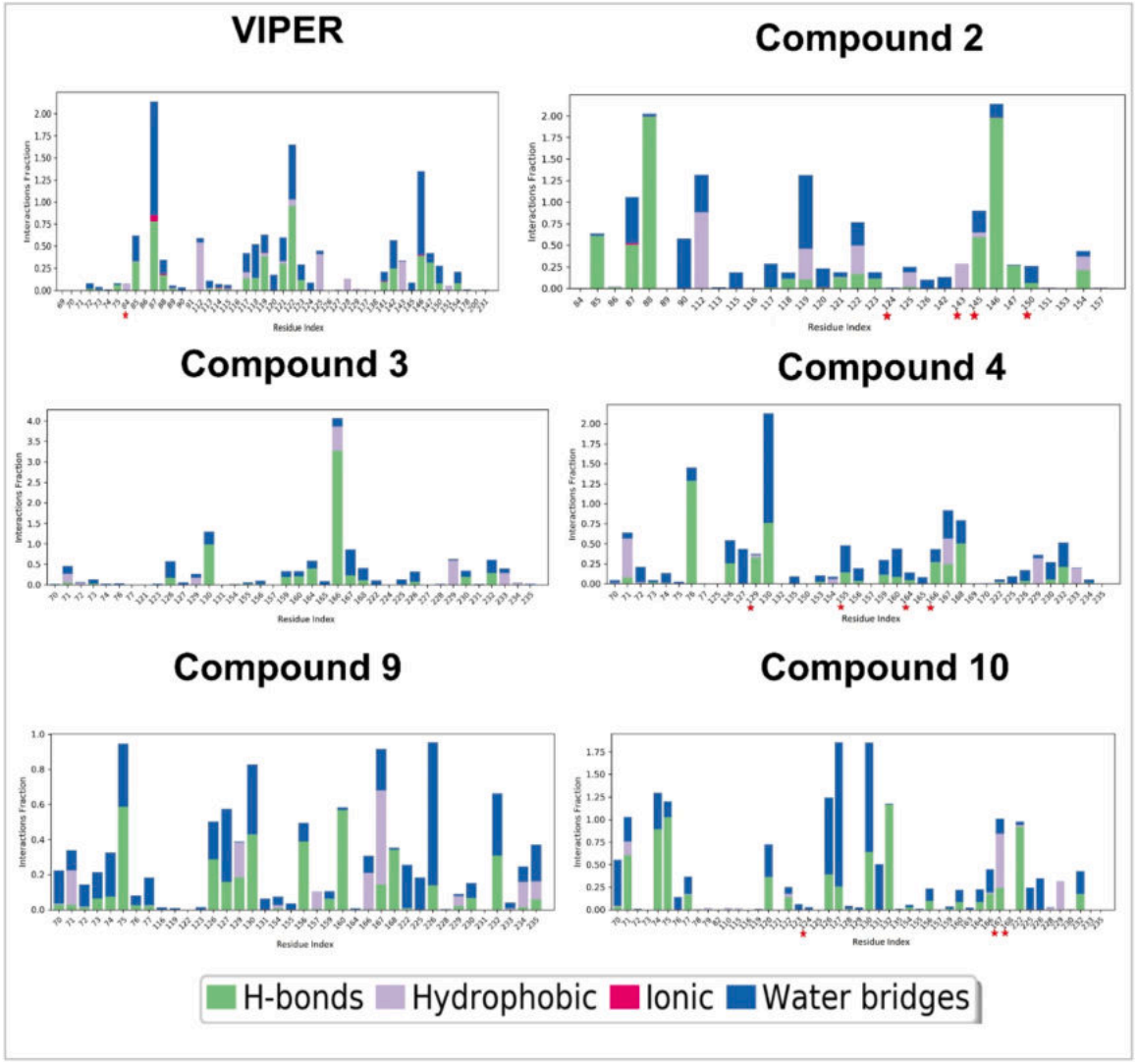


Fig. 5. Stacked bar blot showing residues involved in H-bonds, Hydrophobic, Ionic, and Water bridges. The results are plotted as per the 100 ns MD simulation trajectories for VIPER peptide, Compounds 2, 3, 4, 9, and 10 docked complexes. Residues highlighted with a red star are also found significantly interacting in the NMR titration experiment (**VIPER**: LEU_84, **Compound 2**: ASN_124, PHE_143, ARG_145, ASN_150, **Compound 4**: VAL_129, THR_155, GLN_164, LYS_166, **Compound 10**: ASN_124, TYR_167, ASN_168).

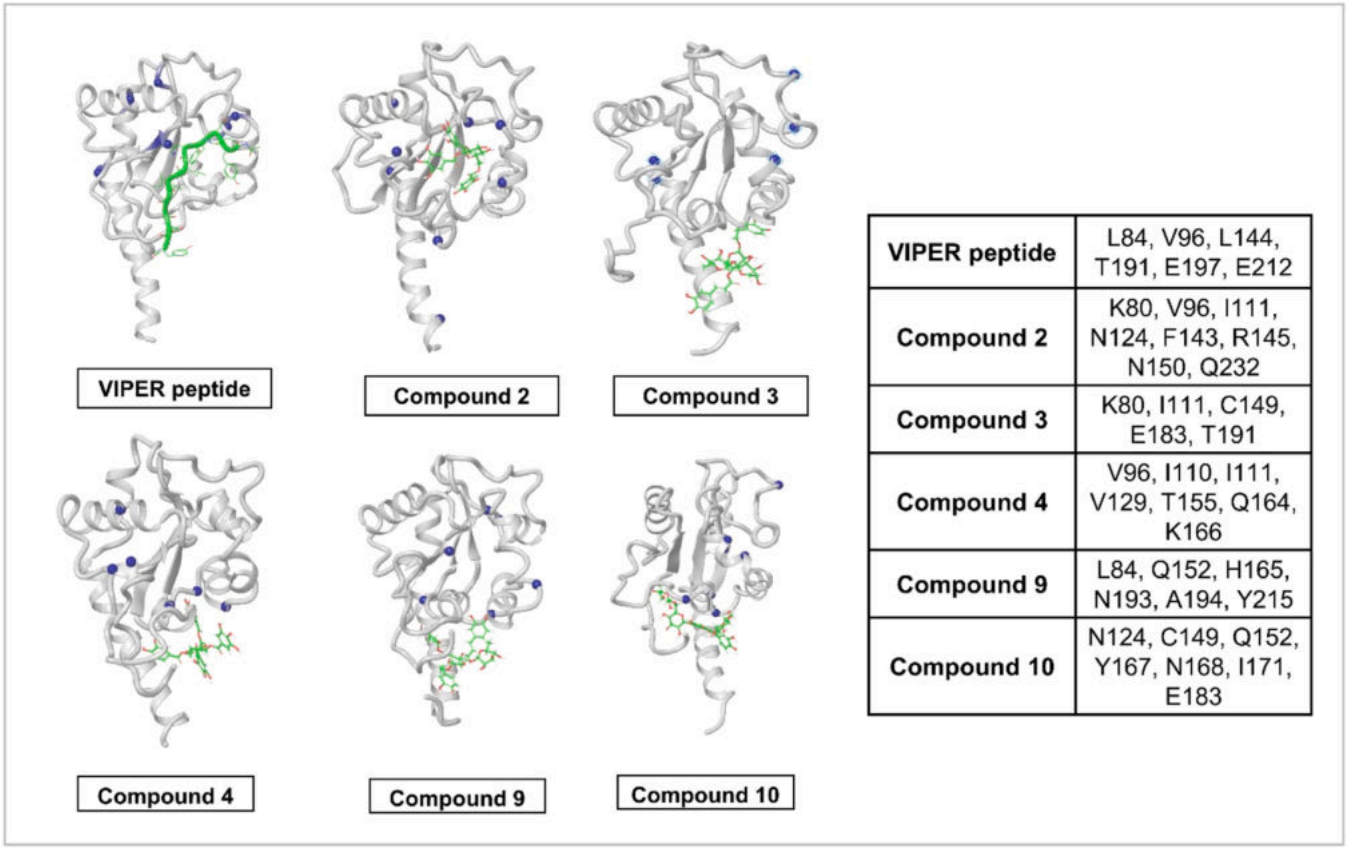
3.3. Reporter gene assay using HEK-Blue hTLR4 cells

We screened the top compounds using cell-based reporter assay and found Compounds 2, 3, 4, 9, and 10 were inhibiting TLR4 activation in presence of ligand (LPS). The inhibition by these compounds were comparable to the known positive control i.e. VIPER peptide. The LPS optimization values and SEAP assay for compounds are shown in Supplementary Figure 3. We also performed dose-dependent assay to calculate the IC_{50} value of the compounds and found Compounds 2 and 4 inhibits the activity at IC_{50} of 1.995 μM and 6.552 μM respectively. Based on the IC_{50}, both the compounds have higher efficacy against TLR4 signaling as compared to know inhibition by VIPER peptide (IC_{50} of 10.49 μM). Results from dose dependent assay of other compounds are shown in Supplementary Figure 3. A figure depicting the schematic of the functioning of HEK-Blue hTLR4 cells, LPS-induced cell-based reporter screening assay, and dose-dependent assay for compounds (VIPER, Compound 2 and 4) that are inhibiting the downstream signaling is shown in Fig. 4 and for Compound 3, 9 and 10 is shown in Supplementary Figure 4. We also performed cell viability assay to check the toxicity of these compounds. This cell viability assay was done in the absence of LPS and using range of concentration of small molecule compound (100 μM, 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.125 μM). As these are small molecule compounds of natural origin, they are expected and found to cause no toxicity to cells. The results are shown in Supplementary Figure 5.

Additionally, compounds 1, 5, 6, and 8 were increasing the SEAP production. It may be binding to TRAM in such a way that strengthen the downstream signaling. Thereby, these may act as potential agonists for the TRAM-mediated TLR4 signaling pathway.

3.4. Molecular dynamics simulations for best ligand-protein complexes

The strength of small molecule ligand interaction with TRAM protein was verified using MD simulations. We observed the Root Mean Square Deviation (RMSD) of protein and ligand from the reference frame to measure the changes in the position of atoms. Protein appears to be stable by the end of the simulation in each case. We also looked at Root Mean Square Fluctuations of the protein to determine the protein region involved in major interactions and see if the secondary structures were maintained across the trajectory. The trend of these values across trajectory for TRAM protein is in Supplementary Figure 6, 7 and 8 respectively.



VIPER peptide	L84, V96, L144, T191, E197, E212
Compound 2	K80, V96, I111, N124, F143, R145, N150, Q232
Compound 3	K80, I111, C149, E183, T191
Compound 4	V96, I110, I111, V129, T155, Q164, K166
Compound 9	L84, Q152, H165, N193, A194, Y215
Compound 10	N124, C149, Q152, Y167, N168, I171, E183

Fig. 6. Cartoon diagram of TRAM protein in docked pose from final frame of molecular dynamics run (100th ns frame). Residues highlighted in blue show CSP higher than average CSP + 2 SD. The table on the right lists the position and name of the residues identified as significantly interacting with each of the compounds through NMR titration experiments.

Apart from these, we also monitored the protein-ligand contacts throughout the simulation. These interactions were categorized mainly into four major categories: Hydrogen Bonds, Hydrophobic, Ionic, and Water Bridges. Interactions that were consistent for 70 % of the simulation are shown in the stacked bar plot in Fig. 5. A schematic for the simulation interaction diagram for ligands docked pose in protein pocket is shown in Supplementary Figure 9.

3.5. NMR protein-ligand titration studies

Some of the residues identified through the simulation interaction diagram are significantly found to be interacting in the NMR titration experiment as shown in Fig. 5. While performing the NMR titration with each protein-ligand complex, the residues which show Chemical shift perturbations (CSP) more than the average value + 2 SD (standard deviation) are highlighted in the images below. For the pictorial representation final frame snapshots (100th ns) from each protein-ligand complex were taken and interacting residues as observed in NMR are highlighted in blue color (CSP > average CSP + 2 SD) (Fig. 6). A detailed plot with CSP values for each complex is shown in Supplementary Figure 10.

Through both the computational and experimental NMR studies, following list of residues were found to be consistent at the interaction site of small molecule compounds and TRAM. Residues N124, F143, R145, N150 for Compound 2; V129, T155, Q164, K166 for Compound 4, and N124, Y167, N168 for Compound 10. Among these residues, N124 and V129 are part of the second alpha helix that is structurally after the BB loop region. Residues F143, R145, and N150 are part of the third alpha helix (Cα helix) that is important for TRAM dimer interface. Residue T155 is part of the fourth alpha helix and it forms the TS site that is important for the interaction of TRAM-TIR with TRIF for downstream signaling. Apart from these, other residues i.e., Q164, K166, Y167, and N168 are part of the loop region connecting the fourth and fifth helix. Amongst these, Y167 is a crucial residue since its mutation leads to complete loss of phosphorylation in response to LPS [43]. Interestingly, in the case of Compounds 3 and 4, we noticed several residues that are far from the predicted binding site. This may imply presence of allosteric binding site or some structural changes occurring at distant residue because of ligand binding. This can be because of long distance interaction occurring in protein structures. A network-based approach can highlight these better, which is currently out of the scope of the paper. Some literature which supports such long range allostery due to ligand binding has been cited [44,45].

4. Conclusion

Overall, the study aims to target the TLR4 pathway in case of overproduction of pro and anti-inflammatory mediators leading to autoimmune conditions. We selected the TRAM adaptor protein as the potential candidate for our study to target TRAM mediated TLR4 pathway. We screened the docking sites against ligands from natural origin followed by extensive steps of filtering, binding energy scoring, clustering, and selecting the top hits. We also extended our computational study to perform the cell-based assay for validating the effect of compounds in HEK-Blue™ TLR4 cells. We performed dose-dependent assay and calculated the IC_{50} value. Further, we purified the TRAM-TIR domain and performed NMR titration experiments to verify the interacting residues in vitro. The involvement of these residues in the significant interaction of small molecules and TRAM protein can interfere with the signaling of TRAM with its downstream TRIF and while forming dimers.

The current approach targets the structurally important regions like the BB loop, Cα helix, TS site, and phosphorylation site.

Altogether using our integrated approach, we identified Compound 2 and Compound 4 as promising hits and probable antagonists for TRAM protein. Compound 2 binds 10 times better than the control VIPER peptide as measured by the reporter assay, whereas Compound 4 retains 60 % lower IC_{50} value in comparison with this peptide. These two compounds also affect the structural positions of multiple residues through NMR titrations. Much of these interactions, especially hydrogen bonds, are retained throughout the MD simulations. These naturally derived small molecule compounds can be useful for abrogating the downstream signaling in case of autoimmune disorders.

A potential limitation of the study will be false negatives, due to our stringent filtering. In addition, the initial docking considers target protein structure as a rigid entity, although multiple docking poses were taken into consideration for selecting the best possible pose. We also used the clustering method to choose the best representative of similar structures. The cluster centroids were, therefore, considered as top hits. Further, cluster members were next selected as secondary hits and the experiments were performed. However, none of the secondary hits were observed to bind better than the initial set of hits (data not shown).

Apart from this, the experiments were conducted using HEK-Blue™ TLR4 cells, replicating the same on other cell lines will give more insight into the behavior and efficacy of these compounds. But as the major objective was to validate these compounds, our LPS dose-dependent assay provides confidence for computationally predicted compounds. Further, the top hits can be used as parent structure and chemically modified to promote stronger binding and be developed as a stronger antagonist against TRAM-mediated TLR4 signaling pathway.

Ethics approval and consent to participate

Not applicable.

Credit authorship contribution statement

SV carried out all the experiments and analyses. SV and PR carried out NMR analyses. RS and SV conceptualized the study. SV wrote the first draft of the manuscript. All authors improved the manuscript.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

The authors would like to thank NCBS (TIFR) for infrastructural facilities. They would like to thank Dr Vinoth Kumar, Dr Praveen Vemula and Dr Ranabir Das for useful discussions. The authors would also thank Shaileshanand Jha, Tripti Kharbanda, Radhika Rao and Rajamohammed Khader for helping with experiments. RS acknowledges funding and support provided by JC Bose Fellowship (JBR/2021/000006) from Science and Engineering Research Board, India and Bioinformatics Centre Grant funded by Department of Biotechnology, India (BT/PR40187/BTIS/137/9/2021). RS would also like to thank Institute of Bioinformatics and Applied Biotechnology for the funding through her Mazumdar-Shaw Chair in Computational Biology (IBAB/MSCB/182/2022).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.07.026.

References

[1] Vidya MK, Kumar VG, Sejian V, Bagath M, Krishnan G, Bhatta R. Toll-like receptors: significance, ligands, signaling pathways, and functions in mammals. Int Rev Immunol 2018. https://doi.org/10.1080/08830185.2017.1380200.
[2] Botos I, Segal DM, Davies DR. The structural biology of toll-like receptors. Structure 2011. https://doi.org/10.1016/j.str.2011.02.004.
[3] Slack JL, et al. Identification of two major sites in the type I interleukin-1 receptor cytoplasmic region responsible for coupling to pro-inflammatory signaling pathways. J Biol Chem 2000. https://doi.org/10.1074/jbc.275.7.4670.
[4] Essuman K, et al. TIR domain proteins are an ancient family of NAD+-consuming enzymes. e4 Curr Biol 2018;28(3):421–30. https://doi.org/10.1016/j.cub.2017.12.024.
[5] Loring HS, et al. A phase transition enhances the catalytic activity of sarm1, an nad + glycohydrolase involved in neurodegeneration. Elife 2021;10:1–33. https://doi.org/10.7554/eLife.66694.
[6] Zughaier SM, Zimmer SM, Datta A, Carlson RW, Stephens DS. Differential induction of the toll-like receptor 4-MyD88-dependent and -independent signaling pathways by endotoxins. Infect Immun 2005;73(5):2940–50. https://doi.org/10.1128/IAI.73.5.2940-2950.2005.
[7] Tan Y, Zanoni I, Cullen TW, Goodman AL, Kagan JC. Mechanisms of toll-like receptor 4 endocytosis reveal a common immune-evasion strategy used by pathogenic and commensal bacteria. Immunity 2015. https://doi.org/10.1016/j.immuni.2015.10.008.
[8] Miguel RN, et al. A dimer of the toll-like receptor 4 cytoplasmic domain provides a specific scaffold for the recruitment of signalling adaptor proteins. PLoS One 2007. https://doi.org/10.1371/journal.pone.0000788.
[9] Y. Enokizono et al., Structures and interface mapping of the TIR domaincontaining adaptor molecules involved in interferon signaling, Proc Natl Acad Sci. U S A, 2013, doi: 10.1073/pnas.1222811110.
[10] Anwar MA, Shah M, Kim J, Choi S. Recent clinical trends in Toll-like receptor targeting therapeutics. Med Res Rev 2019. https://doi.org/10.1002/med.21553.
[11] Martin GS. Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes. Expert Rev Anti Infect Ther 2012;10(6):701–6. https://doi.org/10.1586/eri.12.50.
[12] Torrecillas FCalbo. Germs. Esp Pedia 1989;31(Suppl 3):165–72. https://doi.org/10.1177/000992288702600906.
[13] Savva A, Roger T. Targeting Toll-like receptors: promising therapeutic strategies for the management of sepsis-associated pathology and infectious diseases. Front Immunol 2013;4(NOV):1–16. https://doi.org/10.3389/fimmu.2013.00387.
[14] Mahita J, Harini K, Pichika MR, Sowdhaminia R. An in silico approach towards the identification of novel inhibitors of the TLR-4 signaling pathway. J Biomol Struct Dyn 2015. https://doi.org/10.1080/07391102.2015.1079243.
[15] Lysakova-Devine T, et al. Viral inhibitory peptide of TLR4, a peptide derived from vaccinia protein A46, specifically inhibits TLR4 by directly targeting MyD88 adaptor-like and TRIF-related adaptor molecule. J Immunol 2010;185(7):4261–71. https://doi.org/10.4049/jimmunol.1002013.
[16] Stack J, Bowie AG. Poxviral protein A46 antagonizes toll-like receptor 4 signaling by targeting BB loop motifs in Toll-IL-1 receptor adaptor proteins to disrupt receptor: adaptor interactions. J Biol Chem 2012;vol. 287(27):22672–82. https://doi.org/10.1074/jbc.M112.349225.
[17] ichiro Oda S, Franklin E, Khan AR. Poxvirus A46 protein binds to TIR domain-containing Mal/TIRAP via an α-helical sub-domain. Mol Immunol 2011;48(15–16):2144–50. https://doi.org/10.1016/j.molimm.2011.07.014.
[18] Banerjee P, Erehman J, Gohlke BO, Wilhelm T, Preissner R, Dunkel M. Super natural II-a database of natural products. Nucleic Acids Res 2015. https://doi.org/10.1093/nar/gku886.
[19] Friesner RA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 2006. https://doi.org/10.1021/jm051256o.
[20] Halgren TA. Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model 2009. https://doi.org/10.1021/ci800324m.
[21] W.L. Jorgensen, QikProp, Schrödinger LLC, New York. 2016.
[22] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 2015. https://doi.org/10.1517/17460441.2015.1032936.
[23] Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. J Mol Graph Model 2010. https://doi.org/10.1016/j.jmgm.2010.05.008.
[24] Sastry M, Lowrie JF, Dixon SL, Sherman W. Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. J Chem Inf Model 2010. https://doi.org/10.1021/ci100062n.
[25] Schrödinger Release, Desmond molecular dynamics system Schrödinger LLC, 2019.
[26] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79(2):926–35. https://doi.org/10.1063/1.445869.
[27] Kim Y, Lee H, Heo L, Seok C, Choe J. Structure of vaccinia virus A46, an inhibitor of TLR4 signaling pathway shows the conformation of VIPER motif. Protein Sci 2014;23(7):906–14. https://doi.org/10.1002/pro.2472.
[28] Hu Z, et al. Small-Molecule TLR8 antagonists via structure-based rational design. e3 Cell Chem Biol 2018;25(10):1286–91. https://doi.org/10.1016/j.chembiol.2018.07.004.
[29] Salyer ACD, Caruso G, Khetani KK, Fox LM, Malladi SS, David SA. Identification of adjuvantic activity of amphotericin B in a novel, multiplexed, poly-TLR/NLR high-throughput screen. PLoS One 2016;11(2):1–17. https://doi.org/10.1371/journal.pone.0149848.

[30] Pérez-Regidor L, et al. Small molecules as toll-like receptor 4 modulators drug and in-house computational repurposing. Biomedicines 2022;10(9). https://doi.org/10.3390/biomedicines10092326.
[31] Kuhnert P, Nicolet J, Frey J. Rapid and accurate identification of Escherichia coli K-12 strains. Appl Environ Microbiol 1995;61(11):4135–9. https://doi.org/10.1128/aem.61.11.4135-4139.1995.
[32] Hirschfeld M, Ma Y, Weis JH, Vogel SN, Weis JJ. Cutting edge: repurification of lipopolysaccharide eliminates signaling through both human and murine toll-like receptor 2. J Immunol 2000;165(2):618–22. https://doi.org/10.4049/jimmunol.165.2.618.
[33] Grzesiek S, Stahl SJ, Wingfield PT, Bax A. The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. Biochemistry 1996;35(32):10256–61. https://doi.org/10.1021/bi9611164.
[34] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6(3):277–93. https://doi.org/10.1007/BF00197809.
[35] Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 2015;31(8):1325–7. https://doi.org/10.1093/bioinformatics/btu830.
[36] Ryu HW, Lee DH, Shin DH, Hyun Kim S, Kwon SH. Aceroside VIII is a new natural selective hdac6 inhibitor that synergistically enhances the anticancer activity of HDAC inhibitor in HT29 cells. Planta Med 2015. https://doi.org/10.1055/s-0034-1396149.
[37] Duan D, Li Z, Luo H, Zhang W, Chen L, Xu X. Antiviral compounds from traditional Chinese medicines Galla Chinese as inhibitors of HCV NS3 protease. Bioorg Med Chem Lett 2004. https://doi.org/10.1016/j.bmcl.2004.09.067.
[38] Yu SX, Yan RY, Liang RX, Wang W, Yang B. Bioactive polar compounds from stem bark of Magnolia officinalis. Fitoterapia 2012. https://doi.org/10.1016/j.fitote.2011.11.020.
[39] Torres-León C, Ventura-Sobrevilla J, Serna-Cock L, Ascacio-Valdés JA, Contreras-Esquivel J, Aguilar CN. Pentagalloylglucose (PGG): a valuable phenolic compound with functional properties. J Funct Foods 2017. https://doi.org/10.1016/j.jff.2017.07.045.
[40] J. Zhou, G. Xie, and X. Yan, Encyclopedia of Traditional Chinese Medicines - molecular structures, pharmacological activities, natural sources and applications. 2011.
[41] Jiang WL, Yong-Xu SP, Zhang HB, Zhu, Jian-Hou. Forsythoside B protects against experimental sepsis by modulating inflammatory factors. Phytother Res 2012. https://doi.org/10.1002/ptr.3668.
[42] Nimrouzi M, Zarshenas MM. Phytochemical and pharmacological aspects of Descurainia sophia Webb ex Prantl: modern and traditional applications. Avicenna J Phytomed 2016;6(3):266–72. https://doi.org/10.22038/ajp.2016.4469.
[43] Huai W, et al. Phosphatase PTPN4 preferentially inhibits TRIF-dependent TLR4 pathway by dephosphorylating TRAM. J Immunol 2015. https://doi.org/10.4049/jimmunol.1402183.
[44] Strotz D, et al. Protein allostery at atomic resolution. Angew Chem - Int Ed 2020;59(49):22132–9. https://doi.org/10.1002/anie.202008734.
[45] Ragona L, et al. NMR dynamic studies suggest that allosteric activation regulates ligand binding in chicken liver bile acid-binding protein. J Biol Chem 2006;281(14):9697–709. https://doi.org/10.1074/jbc.M513003200.