MQP-BIO-JBD-22

SLiM Analysis in Human SALMs and Linx

A Major Qualifying Project Report

Submitted to the Faculty of

WORCESTER POLYTECHINC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

In

Biology and Biotechnology

By

Daniele Anina

December 15, 2016

APPROVED:

Joseph B. Duffy, Ph.D.

Biology and Biotechnology

WPI Project Advisor

Abstract

Leucine rich repeat and Immunoglobulin domain (LIG) containing proteins play a critical role in protein-protein and intercellular interactions. While the extracellular domains of many of these proteins have been studied in depth, little is known about their intracellular domains, which can add to our knowledge of LIG protein functionality. Here, within the intracellular domains of two subfamilies of LIG molecules, the SALM and Linx subfamilies, sequences conserved over evolutionary time in four species: human, mouse, chicken, and elephant shark and representing putative functionally relevant Short Linear Motifs, SLiMs have been identified.

Acknowledgements

I would like to thank Duff for being my advisor and for being the first to introduce me to the world of research and all it entails. I want to thank Duff for all of the career and life advice that I have obtained along the way, that has helped shape who I am and has helped me grow beyond my comfort zone in the classroom and in my life. This project has added immensely to my knowledge of many LIGs structure- function relationships. I am grateful to have been a part of the continuous process of discovery that is characteristic of Biology and science in general. Thank you Duff, for being there for me from Day 1 and believing in my capabilities from the outset. Your support has been invaluable to me. I am also extremely grateful to all the Biology professors and courses I have taken which have enhanced my undergraduate career and assisted me in identifying my future career aspirations.

Table of Contents

1. Introduction

Transmembrane proteins play an important role in intercellular and protein-protein signaling events that are essential to metazoan survival. Leucine rich repeats (LRR) and Immunoglobulin (Ig) domain containing proteins are thought to be one class of transmembrane proteins involved in these intercellular interactions. 36 human proteins with LRRs and Ig domains, called LIGS, have been identified and many have been shown to be involved in neuronal growth and synapse formation (Homma et al., 2008).

Domain Architecture of LIGs

Leucine Rich Repeats are between 20 to 30 amino acids and their N terminus has the conserved sequence LxxLxLxxN/CxL, with x representing any amino acid. These repeats are organized into loops, forming a horseshoe-like shape region. These LRR play a role as key binding sites for many protein-protein interactions to form. Figure 1 depicts the structure of these LRR (Hilling et al., 1999).

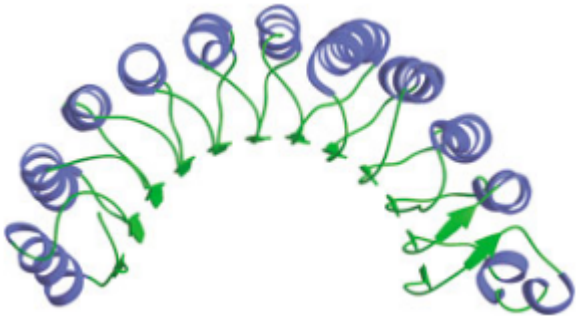


Figure 1. LRR structure found in Rna1p

The immunoglobulin domain is made up of a pair of beta sheets bonded by a disulfide bond where each beta sheet surrounds a hydrophobic core. At the N terminus, there are three loops called hypervariable loops that are present in antibodies and T cell receptors. This domain is one of the most prevalently encoded in the human genome with over 750 genes encoding proteins with at least one Ig domain (Berg et al, 2002) and is believed to play a role in protein- protein interactions. Figure 2 depicts the typical structure for Ig domains (Berg et al., 2002).

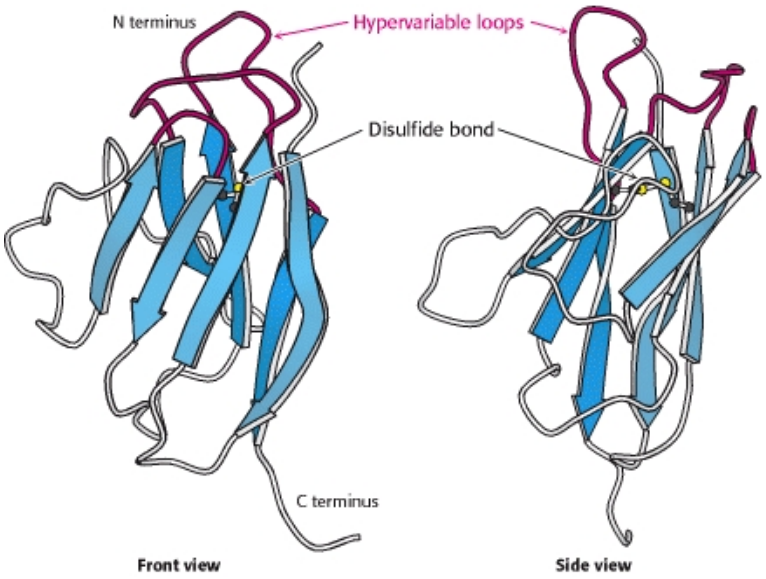


Figure 2. Structure of Ig domain (adapted from Berg et al.,2002)

All LIGs start with LRRs, presumably with their canonical horseshoe shape followed by an Ig domain(s), as can be seen below in Figure 3 (Wit et al., 2011). Different LIGs have variable numbers of each, but must contain both LRR and Ig domains to be considered a LIG. Some LIG proteins may also have additional domains as the SALMs contain a fibronectin domain as well.

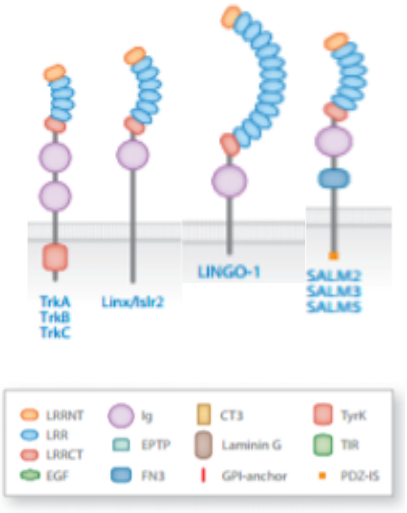


Figure 3. Domain Architecture of some LIG proteins (adapted from Wit et al., 2011)

SALMs: LIG protein family serve as CAMs

Cell adhesion molecules (CAM) play a major role in the formation of synapses and many are involved in the development and maintenance of these synapses (Missler et al., 2012). These transmembrane molecules can also act to hold neuronal membranes together. In addition, they often are necessary to allow for proper communication and contact between axons and dendrites and are involved in intercellular signaling (Missler et al., 2012). One family of LIG proteins that appear to function as adhesion molecules are known as Synaptic Adhesion Like Molecules, or SALMs. This family has 5 members, SALM1, SALM2, SALM3, SALM4, and SALM5 that are structurally similar to each other as they all contain extracellular Leucine rich repeats (LRR), IG C2 type domains, and fibronectin type III domains (Wang et al., 2006). SALMs have been found to have both pre and post synaptic functions and play a role in neurite outgrowth and branching (Choi et al., 2016). These molecules, which are also called Lrfn because they contain the LRR and Fibronectin type III motifs, are transmembrane proteins that also contain a PDZ-domain

binding site (Wang et al., 2006). This site is about 4 amino acids in length with the sequence –X,-T,-X,-V at the carboxyl terminus with x being any amino acid, that binds PDZ domain proteins that play a critical role in protein-protein recognition, protein trafficking, localization, and cell signaling (Lee et al., 2004). These PDZ domain binding sites are found in SALM1-3, but not SALM4 and SALM5, suggesting a variety of functionalities between family members.

CAMs are important for proper neuronal and brain development and mutations in CAMs have been associated with neurological disorders and developmental damage (Seabold et al., 2012). Specifically, SALM1 and SALM5 have been associated with autism disorders and changes in neuronal morphology. Understanding the structural components of these proteins can give further insight into the functional properties of these molecules and how they contribute to the development of neurological disorders. While the extracellular domains of many of these proteins have been studied in detail and have been found to be highly conserved, little is known about the intracellular domains. Studying the intracellular domains of these molecules can add to our knowledge of their function. Specifically, identifying functionally important sequences or motifs in the intracellular domains, such as SLiMs for example, can direct further research efforts in understanding the function of many LIG proteins.

Short linear motifs or SLIMs are approximately 3-10 adjacent amino acid stretches in a protein's primary sequence that are believed to be functionally important for protein activity. SLiMs are thought to mediate between 15-40 % of protein-protein interactions and are therefore critical to understand their mechanistic contributions to cellular signaling events (Edwards et al., 2007). Unfortunately, these sequence elements are very difficult to identify due to their short length sequence. One potential method to identify SLiMs is to take a phylogenetic approach

through the identification of protein orthologs and analysis of conserved sequences among the orthologous molecules.

Identification of orthologs, which represent the same functional protein in different species, is a critical step towards understanding the function of a protein and can be used to identify functionally important sequences that are conserved over time. In this study, the objective was to take a comparative approach to identifying putative SLiMs within the intracellular domains of the SALM and Linx members of the LIG family to better understand their biological activity and mechanism of action. Sequence comparison of SALMs and Linx orthologs across a phylogenetically diverse set of vertebrates, including Homo sapien (human), Mus musculus (mouse), Gallus gallus (chicken), and Callorhinchus milii (elephant shark) species was performed. Orthologs were identified using BLAST, their intracellular domains identified using the online transmembrane prediction tool CCTOP, and their sequence conservation analyzed through protein alignments. Sequence conservation over time revealed intracellular motifs, or putative SLiMs, present within and among the SALMs and Linx orthologs, indicative of sequence units whose study is likely to reveal novel *in vivo* functions linked to the LIG family.

2. Materials and Methods

Identification of protein sequences of LIGS

The NCBI protein database was utilized in order to identify the LIG protein sequences. Using their respective accession numbers, collected from the literature (Homma et al., 2008), the Homo sapien sequences were identified. The NCBI protein BLAST program was then used to identify orthologs of the SALM and Linx proteins in Mus musculus, Gallus gallus, and Callorhinchus milii, mouse (Tax ID10090), chicken (Tax ID 9031), and elephant shark (Tax ID7868), respectively. Putative SALM and Linx orthologs were defined as the match with the highest identity to the human protein query, which in a reciprocal BLAST also identified the initial human LIG query as the highest match.

Identification of the IC domain

The program CCTOP was used to identify the transmembrane region for all orthologs, thereby allowing for the extracellular and intracellular regions of each protein to be defined. CCTOP combines output from 10 different programs to generate a consensus for the prediction of the transmembrane region of proteins. Below is a sample output of a CCTOP consensus and output (Figure 4).

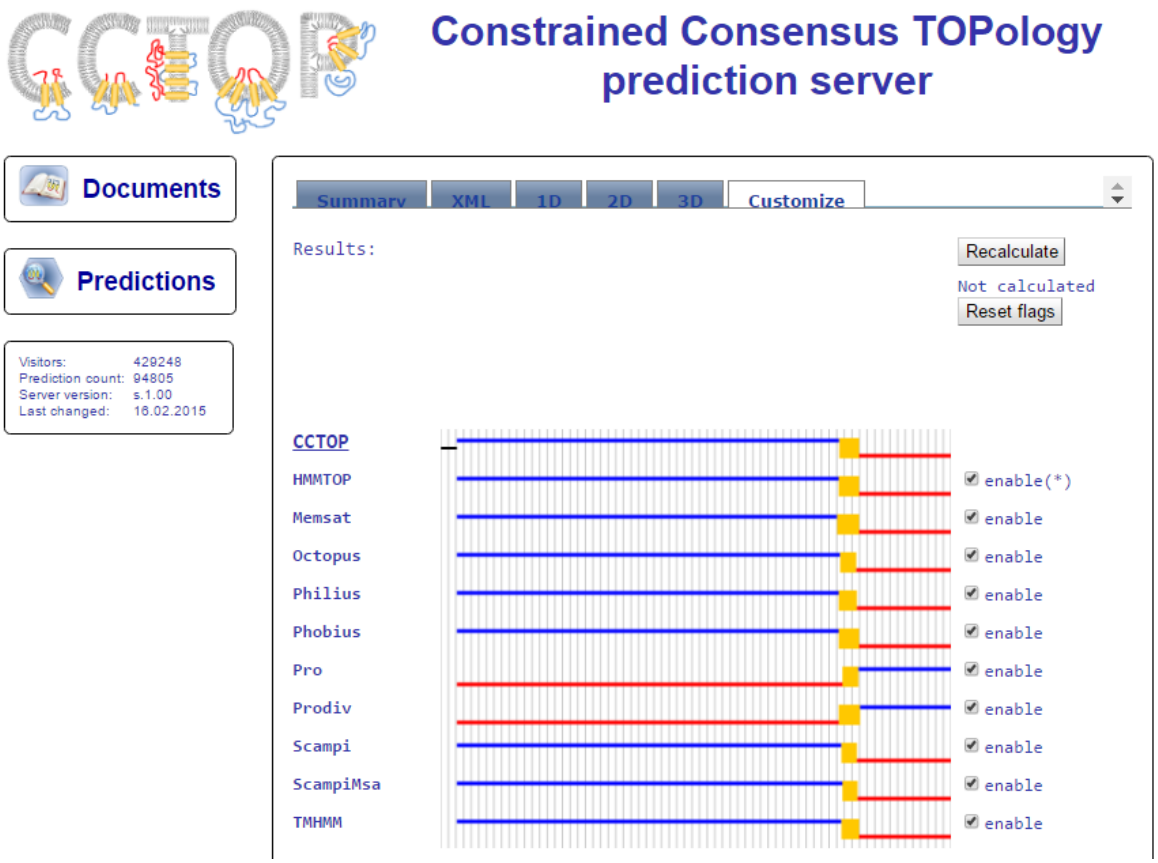


Figure 4. CCTOP based prediction and representation of transmembrane domains

These programs all have the underlying assumption that the transmembrane domain is made up of stretch of hydrophobic amino acids and use amino acid physical properties, as well as structural information to generate a prediction of the position and length of the transmembrane domain and extracellular and intracellular regions by consequence.

Ortholog IC domain Sequence Alignment

After obtaining the predictions of the extracellular, transmembrane, and intracellular regions of the SALM and Linx proteins, the IC domains were collected for each ortholog and sequence alignments built using Clustal Omega. Using only the IC domain, the orthologs were aligned and output in a fasta format. The program Boxshade was then used to create a highlighted protein

alignment of the sequence conservation between orthologs identified in each species. Using the CLUSTAL fasta output, the Boxshade program shades sequences of amino acids based on the level of agreement of the amino acid sequence between species when aligned. For the purpose of this study, conservation was set to shade residues at either 100% or 75 % identity between species as shown in Appendix D. The shading is black when the level of sequence identity is 100% between amino acids at a certain residue, while simple sequence conservation at a residue is shaded in gray. A consensus line is a part of the output that specifies which amino acids are conserved between the 4 species.

Identification of Fingerprints

WebLogo was used to generate a graphical representation of putative IC domain SLiMs based on the overall sequence alignments built with CLUSTAL and Boxshade. The logo is built using a stack of symbols, with the height of the stack indicating sequence conservation at the given position, and with each stack representing a single position in the sequence. A custom color scheme was used in order to also provide information regarding conservation of biochemical properties of amino acids at each position (below).

Chemistry (AA): Color amino acids according to chemical properties.

Polar	G,S,T,Y,C	green
Neutral	Q,N	purple
Basic	K,R,H	blue
Acidic	D,E	red
Hydrophobic	A,V,L,I,P,W,F,M	black

3. Results

To better understand the biological activity and mechanism of action of human LIGs, a comparative approach to identifying putative SLiMs was undertaken. Of the 36 human LIGs, 17 were chosen for initial analyses of IC domain sequences with respect to location and length (see Appendix A)_. Of the 17, six proteins (SALMs 1-5 and Linx) were analyzed for SLIMs. To identify these short linear motifs, the human amino acid sequences for all proteins were found, followed by identifying orthologs in selected vertebrates (mouse, chicken, and elephant shark) that served as a basis for comparison of protein sequences to determine conservation. Conservation in the IC domain was the primary focus since little is known for these regions in LIGs as compared with their extracellular domains with the hope of gaining a deeper understanding of protein functionality. An overview of the experimental approach is presented in Figure 5.

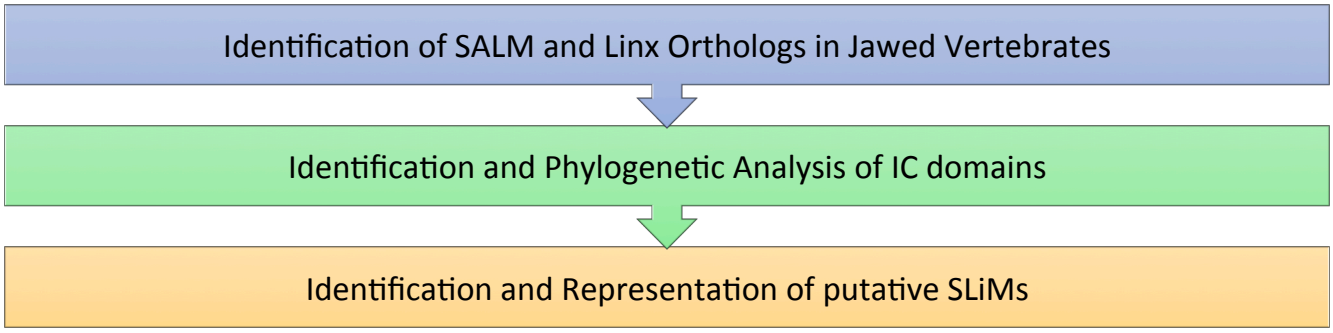


Figure 5. Phylogenetic analyses and SLiM identification in LIGs.

3.1 Identification of LIG family in Jawed Vertebrates

To understand and identify conserved sequences, the human amino acid sequences of 17 LIGs were obtained. Using the accession numbers for each LIG and the NCBI protein database, the

LIG protein sequence in homo sapiens were identified (Homma et al., 2008). Complete sequences for all 17 can be found in Appendix A.

In order to identify SLiMs in the human protein sequence, the idea of evolutionary conservation was utilized as a technique for motif discovery. Because SLiMs are often very short (4-10 amino acid residues) in length, they are difficult to identify (Edwards et. al., 2007). Therefore, a phylogenetic approach was undertaken to identify short sequences conserved in LIGs from species with different evolutionary relationships. With this methodology, short but highly identical amino acid sequences between species would suggest an important functionality of the protein that needed to be conserved over varying evolutionary distances/time.

Three species were identified on the basis of their evolutionary relationships to Homo sapiens, and each other, for this analysis, Mus musculus, Gallus gallus, and Callorhinchus milii or mouse, chicken, and elephant shark, respectively. Together the four species represent divergence times up to ~450 Mya and two branchpoints - divergences between cartilaginous fishes and bony vertebrates, and reptiles and mammals (Smith and Keinath, 2015; Schmutz and Grimwood, 2004; Venkatesh et. al., 2014; Waterston et. al., 2002). Mus musculus is the closest in genetic makeup to Homo sapiens as it has had the least amount of time to evolve as compared to the chicken and elephant shark genomes. The elephant shark has had the longest time to evolve and is the least similar to Homo sapien of the three species, while chicken falls in the middle of evolutionary distance relative to humans and sharks.

Figure 6 below illustrates the evolutionary relationships between these species in a phylogenetic tree.

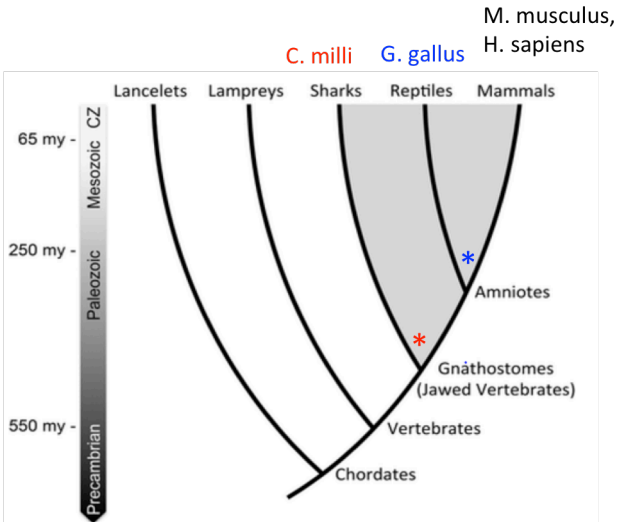


Figure 6. Phylogenetic tree for Homo sapien, Mus musculus, Gallus gallus, and Callorhinchus milii (modified from Smith and Keinath, 2015). This phylogeneic tree depicts a timeline for evolution between chordates and vertebrates, including sharks, reptiles, and mammals. The timeline shows key events in the evolution of these major family groups.

After the LIG sequences were identified in Homo sapiens, NCBI BLAST and protein database was used to find the ortholog sequences of the SALM subfamily and LINX in the selected species - Mus musculus, Gallus gallus, and Callorhinchus milii. Using Protein Blast, the sequences of each SALM (1-5) and Linx were used to find their respective orthologs in the other species (see Materials and Methods for details). Percent similarity, high score and reciprocal BLASTing back to the h. sapiens database were used to determine the correct orthology. Table 1 below summarizes the presence of SALM and Linx orthologs in these species. While Linx was found in all species analyzed here, SALM3 was not found in chicken, and SALM4 was not found in either chicken or elephant shark.

Name	Human	Mouse	Chicken	Elephant Shark
LINX	Yes	Yes	Yes	Yes
SALM1	Yes	Yes	Yes	Yes
SALM2	Yes	Yes	Yes	Yes
SALM3	Yes	Yes	No	No
SALM4	Yes	Yes	No	Yes
SALM5	Yes	Yes	Yes	Yes

Table 1. Presence of Linx or SALMs in the human, mouse, chicken, and elephant shark genomes.

The sequence data for each LIG and the accession numbers can be found in Appendix B.

3.2 Identification of IC domains using CCTOP

After orthologs were identified, the protein sequences were analyzed using the prediction program CCTOP, which predicts the location of transmembrane domains, as well as the extracellular (EX) and IC domains. Green was used to highlight the EX domain, blue for the transmembrane domain, and red for the IC domain. A complete set of topology predictions for all SALM1-5 and Linx orthologs can be found in Appendix A and B. A compilation of the predicted IC domain length for each protein can be found in Table 2 below.

Name	Human	Mouse	Chicken	Elephant Shark
LINX	132	132	134	100
SALM1	234	233	79	226
SALM2	214	209	217	215
SALM3	96	97	NA	NA
SALM4	69	66	NA	121
SALM5	169	196	187	225

Table 2. Length of IC domains (amino acids) in SALM and Linx orthologs.

Within the SALM family, SALM1's IC domain was the largest across human, mouse, and elephant shark species, but not in chicken. As expected based on the relatively short evolutionary distance between human and mouse the length of the IC domains across orthologs was most similar between human and mouse species.

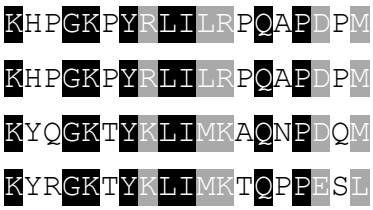
3.3 Alignment of ortholog sequences in Clustal Omega

With the orthologs in hand and their IC domains defined, sequence alignments were performed to identify regions of conservation. The IC domains of each LIG and their respective orthologs were aligned using Clustal Omega to look for sequence conservation between species. The IC domain sequences for each LIG in each species that were used for the alignments can be found in Appendix C.

3.4 Identification of Motifs using Box shade

To identify putative SLiMs, or conserved regions, the CLUSTAL alignments were analyzed with the program Boxshade. Parameters were varied to display regions of either 100% and 75% conservation and the consensus amino acid residue conserved between all represented sequences. The box shade outputs can be found in Appendix D. These alignments were then analyzed for putative SLiMs, conserved sequences across species, as well as between other members of the SALM family. Three motifs were identified in Linx based on visual analysis of sequences. Putative SLiMs were selected based on 100% conservation and were extended to include areas when only some conservation was found. Figure 7 displays the motifs for Linx. SALM1 was found to have six motifs, while SALM2 had four, SALM3 had one motif, SALM4 had two motifs, and SALM5 contained four motifs. For a list of all motifs identified and the sequence alignments refer to appendix E. In total, twenty motifs were identified.

DA18



DA19

DA20

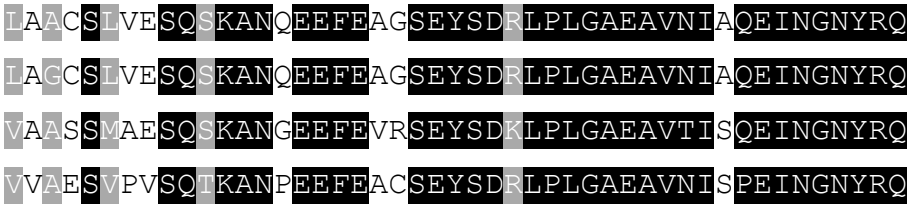
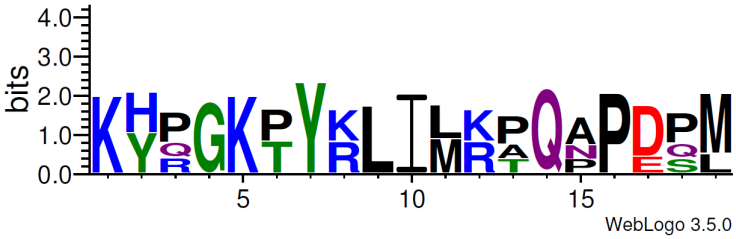


Figure 7. Putative Linx motifs.

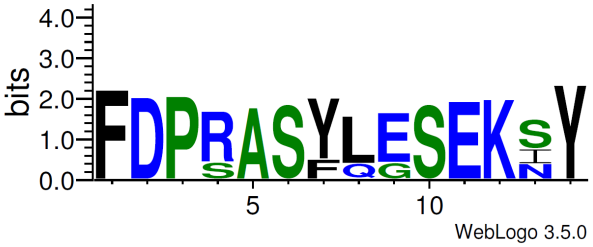
3.5 Identification and Representation of putative SLiMs

To provide a graphical representation of the degree and possible biochemical characteristics associated with regions conservation, the bioinformatics tool WebLogo was used to visualize all motifs. Figure 8 represents the graphical output for the three Linx motifs - DA18, DA19, and DA20.

DA18



DA19



DA20

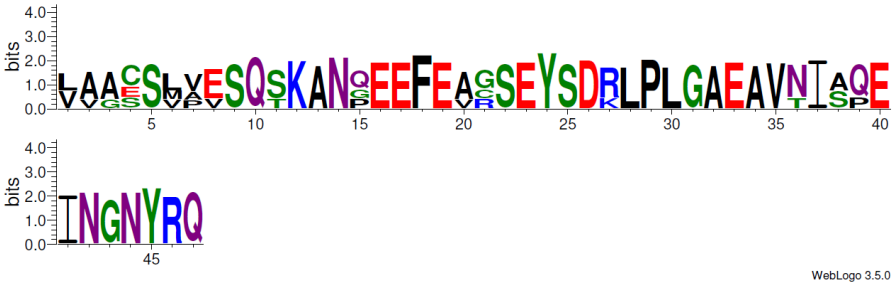


Figure 8. Putative SLiMs for Linx represented in WebLogo.

The graphic shows the representation of specific amino acids found at every position in a given SLiM. Letter height represents the frequencies of specific residues at a certain position, while different colors are representative of specific chemical properties of the amino acids.

Conserved sequences were analyzed within motifs across the family and among different family members, specifically for the SALM family. In some cases submotifs present in a SLiM were defined as fingerprints. Two such fingerprints were identified from analysis of motifs in the SALM family and are shown in Figure 9.

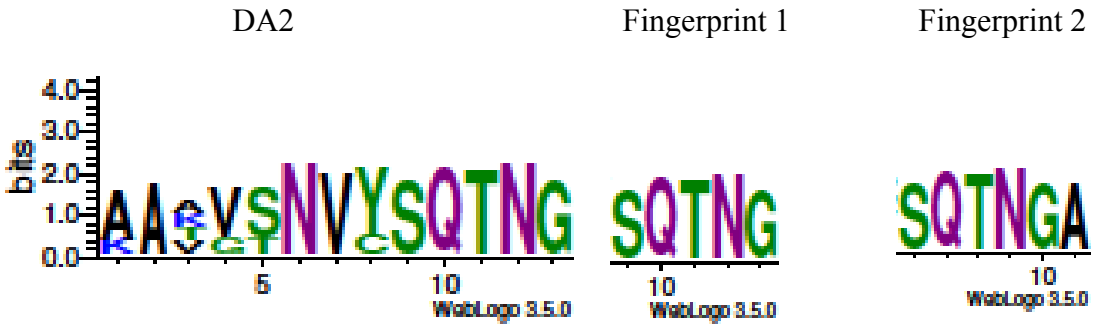


Figure 9. Identification of DA2 Fingerprints from SALM family

The shortened amino acid sequence SQTNG in DA2 is found in all SALM homologs and was therefore identified as a fingerprint because its conservation across species, as well as across different SALM family members. Four additional fingerprints were identified through this analysis, including ESVV, ESTV, RYKV, and SFD. Motifs were identified that were also unique to their specific SALM and were not conserved between all family members. The sequence ESTV at the terminus of the IC domain was conserved in SALM1-3, but not SALM4-5. The sequence RYKV was found in SALMs1, 2, 4, and 5 but not SALM3. The fingerprints for each LIG can be found in Appendix E. Table 3-7 below summarize the fingerprints found in the SALM family in the 4 species.

Name	Human	Mouse	Chicken	Elephant Shark
SALM1	+	+	+	+
SALM2	+	+	+	+
SALM3	+	+	+	+
SALM4	+	+	+	+
SALM5	+	+	+	+

Table 3. Prevalence of SQTNG Fingerprint in SALM1-5

Name	Human	Mouse	Chicken	Elephant Shark
SALM1	+	+	+	+
SALM2	+	+	+	+
SALM3	-	-	-	-
SALM4	-	-	-	-
SALM5	-	-	-	+

Table 4. Prevalence of ESTV Fingerprint in SALM1-5

Name	Human	Mouse	Chicken	Elephant Shark
SALM1	-	-	-	-
SALM2	-	-	-	-
SALM3	+	+	-	-
SALM4	-	-	-	-
SALM5	-	-	-	-

Table 5. Prevalence of ESVV Fingerprint in SALM1-5

Name	Human	Mouse	Chicken	Elephant Shark
SALM1	+	+	+	+
SALM2	+	+	+	+
SALM3	-	-	-	-
SALM4	+	+	-	+
SALM5	+	+	+	+

Table 6. Prevalence of RYKV Fingerprint in SALM1-5

Name	Human	Mouse	Chicken	Elephant Shark
SALM1	+	+	+	+
SALM2	+	+	+	+
SALM3	-	-	-	-
SALM4	-	-	-	-
SALM5	-	-	-	-

Table 7. Prevalence of SFD Fingerprint in SALM1-5

As Table 3 portrays, the fingerprint SQTNG, is found in all SALM family members. The fingerprint ESTV is found in all species for SALM1 and SALM2, but not for the other family members. The fingerprint ESVV is found only in SALM3. The fingerprint RYKV is found in SALMs 1,2,4 (except for Gallus gallus), and 5 in all four species. The fingerprint SFD is only found in SALM1 and SALM2 for all species. The presence and distribution of unique fingerprints suggests both diversification and redundancy in functionality of the proteins over evolutionary time.

Discussion

Leucine rich repeat and immunoglobulin (LIG) containing proteins are thought to have a significant role in protein-protein and intercellular interactions. Like the SALM proteins, many are involved with cellular interactions during stages of neural development and function. While the extracellular domains of these proteins have been studied with some detail, little is known about the functionalities of their intracellular domains. Through the study of the IC domains of these LIG proteins, we can add to our knowledge of these protein's roles *in vivo* and the molecular mechanisms by which they act.

In this study, I identified evolutionarily conserved sequences that may represent functional SLiMs in the intracellular domains of two families of LIG molecules, SALM and Linx, that are conserved in four species: human, mouse, chicken, and elephant shark. These putative SLiMs are likely conserved between species because they provide specific essential or vital functions that an organism or species needs to survive. The sequences were found using the idea of evolutionary conservation because of the short length that would not be easily recognizable using standard search algorithms.

The fingerprints found were SQTNG, ESVV, ESTV, RYKV, and SFD. The SQTNG fingerprint's functionality can be emphasized in that it appears in every species and is conserved among all SALM family members. This prevalence suggests a functional importance and a clear target for further studies. The ESVV and ESTV fingerprints end the IC domain sequence in SALM3 and SALMs 1 and 2, respectively, across all species. They do not appear as motifs in

SALM4 and 5. These fingerprints are consensus PDZ-domain binding sites and suggest some degree of functional diversification between SALMs 1, 2, and 3 with SALMs 4 and 5.

The SLiMs and fingerprints founds can be used for further analysis and to further our understanding of SALM and Linx proteins *in vivo*. The SALM subfamily has been associated with autism disorders and changes in neuronal morphology. Understanding the relationship between these sequence elements and their contributions to protein function can give further insight into the role of these molecules and how they contribute to the development of neurological disorders.

References

Berg, J. et al. (2002). Biochemistry: Section 33.2 The Immunoglobulin Fold Consists of a Beta-Sandwich Framework with Hypervariable Loops. New York: W H Freeman https://www.ncbi.nlm.nih.gov/books/NBK22461/

Choi, Y. et al (2016). SALM5 trans-synaptically interacts with LAR- RPTPs in a splicing dependent manner to regulate synapse development . *Scientific Reports*, 1-12.

Edwards, R. J., Norman, E. D., & Shields, D. C. (2007). SLiMFinder: A probabilistic Method for Identifying Over- Represented, Convergently Evolved, Short Lnear Motifs in Proteins. PLOS ONE, 2 (10), e967.

Hillig, R.C. et al. (1999). LRR Protein Domain. Mol. Cell 3(6), 781–791.

Homma, S. et al. (2008). Expression Pattern of LRR and Ig domain- containing (LRRIG) protein in the early mouse embryo. *Gene Expression Patterns,* 1-26

Lee, C. et al. (2004). Role of the PDZ Domain-Binding Motif of the Oncoprotein E6 in the Pathogenesis of Human Papillomavirus Type 31. *Journal of Virology*, 12366-12377.

Lie, E. et al. (2015). SALM4 suppresses excitatory synapse development by cis-inhibiting trans-synaptic SALM3-LAR adhesion. *Nature Communications* , 1-15.

Missler, M, Sudhof, T.C., & Biederer, T. (2012). Synaptic Cell Adhesion. Cold Spring Harb Perspect Biol., 4: 1-18.

Seabold, Gali K. et al (2012). Dileucine and PDZ binding motifs Mediate Synaptic Adhesion like Molecules 1 (SALM1) Trafficking in Hippocampal Neurons. *The Journal of Biological Chemistry*, 4470-4484.

Schmutz, J. and Grimwood, J. (2004). Fowl Sequence. *Nature, Vol. 432,* 679-680.

Smith, J. J. & Keinath, M. C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrae genome duplications. *Genome Research*, 25, 1081- 1090

Venkatesh, B., Lee, A., Ravi, V., Maurya, A., …Warren, W. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature, Vol. 505,* 174-179.

Wang, Chag-Yu et al. (2006). A Novel Family of Adhesion- Like Molecules that Interacts with the NMDA Receptor. *The Journal of Neuroscience*, 2174-2183.

Waterston, R., Lindblad-Toh, K., Birney, E., Rogers, J., Brent, M., …Lander, E. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, Vol. 420,* 520-562.

Williams, A. & Barclay, A. (1988). The Immunoglobulin Superfamily- Domains for Cell Surface Recognition. *Annual Reviews of Immunology*, 6, 381-405 http://www.annualreviews.org/doi/10.1146/annurev.iy.06.040188.002121

Wit, J., Hong, W., Luo, L., & Ghosh, A. (2011). Role of Leucine- Rich Proteins in the Development and Function of Neural Circuits. *Annual Review of Cell and Developmental Biology,* 27, 697- 729.

Appendix A: Protein Sequence for Human LIGS

SALM1

Genbank:Q9P244.2

771aa

Leucine rich repeat and fibronectin type III domain containing 1 [Homo sapiens]

>gi|189028858|sp|Q9P244.2|LRFN1_HUMAN RecName: Full=Leucine-rich repeat and fibronectin type III domain-containing protein 1; AltName: Full=Synaptic adhesion-like molecule 2; Flags: Precursor

MAPGPFSSALLSPPPAALPFLLLLWAGASRGQPCPGRCICQNVAPTLTMLCAKTGLLFVPPAIDRRVVEL
RLTDNFIAAVRRRDFANMTSLVHLTLSRNTIGQVAAGAFADLRALRALHLDSNRLAEVRGDQLRGLGNLR
HLILGNNQIRRVESAAFDAFLSTVEDLDLSYNNLEALPWEAVGQMVNLNTLTLDHNLIDHIAEGTFVQLH
KLVRLDMTSNRLHKLPPDGLFLRSQGTGPKPPTPLTVSFGGNPLHCNCELLWLRRLTREDDLETCATPEH
LTDRYFWSIPEEEFLCEPPLITRQAGGRALVVEGQAVSLRCRAVGDPEPVVHWVAPDGRLLGNSSRTRVR
GDGTLDVTITTLRDSGTFTCIASNAAGEATAPVEVCVVPLPLMAPPPAAPPPLTEPGSSDIATPGRPGAN
DSAAERRLVAAELTSNSVLIRWPAQRPVPGIRMYQVQYNSSVDDSLVYRMIPSTSQTFLVNDLAAGRAYD
LCVLAVYDDGATALPATRVVGCVQFTTAGDPAPCRPLRAHFLG

GTMIIAIGGVIVASVLVFIVLLMI

RYKVYGDGDSRRVKGSRSLPRVSHVCSQTNGAGTGAAQAPALPAQDHYEALREVESQAAPAVAVEAKA
MEAETASAEPEVVLGRSLGGSATSLCLLPSEETSGEESRAAVGPRRSRSGALEPPTSAPPTLALVPGGAAAR
PRPQQRYSFDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARACLAFTSTEWM
LESTV

SALM2

Leucine rich repeat and fibronectin type III domain containing 2 [Homo sapiens]

GenBank: AAI42617.1 (NP_065788)

789 aa

GenPept Identical Proteins Graphics

>gi|148745628|gb|AAI42617.1| Leucine rich repeat and fibronectin type III domain containing 2 [Homo sapiens]

METLLGGLLAFGMAFAVVDACPKYCVCQNLSESLGTLCPSKGLLFVPPDIDRRTVELRLGGNFIIHISRQ
DFANMTGLVDLTLSRNTISHIQPFSFLDLESLRSLHLDSNRLPSLGEDTLRGLVNLQHLIVNNNQLGGIA
DEAFEDFLLTLEDLDLSYNNLHGLPWDSVRRMVNLHQLSLDHNLLDHIAEGTFADLQKLARLDLTSNRLQ
KLPPDPIFARSQASALTATPFAPPLSFSFGGNPLHCNCELLWLRRLERDDDLETCGSPGGLKGRYFWHVR
EEEFVCEPPLITQHTHKLLVLEGQAATLKCKAIGDPSPLIHWVAPDDRLVGNSSRTAVYDNGTLDIFITT
SQDSGAFTCIAANAAGEATAMVEVSIVQLPHLSNSTSRTAPPKSRLSDITGSSKTSRGGGGSGGGEPPKS
PPERAVLVSEVTTTSALVKWSVSKSAPRVKMYQLQYNCSDDEVLIYRMIPASNKAFVVNNLVSGTGYDLC
VLAMWDDTATTLTATNIVGCAQFFTKADYPQCQSMHSQI

LGGTMILVIGGIIVATLLVFIVILMV

RYKVCNHEAPSKMAAAVSNVYSQTNGAQPPPPSSAPAGAPPQGPPKVVVRNELLDFTASLARASDSSSSSS
LGSGEAAGLGRAPWRIPPSAPRPKPSLDRLMGAFASLDLKSQRKEELLDSRTPAGRGAGTSARGHHSDREP
LLGPPAARARSLLPLPLEGKAKRSHSFDMGDFAAAAAGGVVPGGYSPPRKVSNIWTKRSLSVNGMLLPFEE
SDLVGARGTFGSSEWVMESTV

SALM3

Leucine rich repeat and fibronectin type III domain containing 3 [Homo sapiens]

GenBank: AAH03578.1 (NP_078785.1)

628 aa

>gi|13097762|gb|AAH03578.1| Leucine rich repeat and fibronectin type III domain containing 3 [Homo sapiens]
MAILPLLLCLLPLAPASSPPQSATPSPCPRRCRCQTQSLPLSVLCPGAGLLFVPPSLDRRAAELRLADNF
IASVRRRDLANMTGLLHLSLSRNTIRHVAAGAFADLRALRALHLDGNRLTSLGEGQLRGLVNLRHLILSN
NQLAALAAGALDDCAETLEDLDLSYNNLEQLPWEALGRLGNVNTLGLDHNLLASVPAGAFSRLHKLARL
D
MTSNRLTTIPPDPLFSRLPLLARPRGSPASALVLAFGGNPLHCNCELVWLRRLAREDDLEACASPPALGG
RYFWAVGEEEFVCEPPVVTHRSPPLAVPAGRPAALRCRAVGDPEPRVRWVSPQGRLLGNSSRARAFPNGT
LELLVTEPGDGGIFTCIAANAAGEATAAVELTVGPPPPPQLANSTSCDPPRDGDPDALTPPSAASASAKV
ADTGPPTDRGVQVTEHGATAALVQWPDQRPIPGIRMYQIQYNSSADDILVYRMIPAESRSFLLTDLASGR
TYDLCVLAVYEDSATGLTATRPVGCARFSTEPALRPCGAPHAPF
LGGTMIIALGGVIVASVLVFIFVLL
MRYKVHGGQPPGKAKIPAPVSSVCSQTNGALGPTPTPAPPAPEPAALRAHTVVQLDCEPWGPGHEPVGP

SALM4

Leucine rich repeat and fibronectin type III domain containing 4 [Homo sapiens]

GenBank: AAH15581.2 (NP_076941)

635 aa

GenPept Identical Proteins Graphics

>gi|22800525|gb|AAH15581.2| Leucine rich repeat and fibronectin type III domain containing 4 [Homo sapiens]
MAPPLLLLLLASGAAACPLPCVCQNLSESLSTLCAHRGLLFVPPNVDRRTVELRLADNFIQALGPPDFRN
MTGLVDLTLSRNAITRIGARAFGDLESLRSLHLDGNRLVELGTGSLRGPVNLQHLILSGNQLGRIAPGAF
DDFLESLEDLDLSYNNLRQVPWAGIGAMPALHTLNLDHNLIDALPPGAFAQLGQLSRLDLTSNRLATLAP
DPLFSRGRDAEASPAPLVLSFSGNPLHCNCELLWLRRLARPDDLETCASPPGLAGRYFWAVPEGEFSCEP
PLIARHTQRLWVLEGQRATLRCRALGDPAPTMHWVGPDDRLVGNSSRARAFPNGTLEIGVTGAGDAGGY
T
CIATNPAGEATARVELRVLALPHGGNSSAEGGRPGPSDIAASARTAAEGEGTLESEPAVQVTEVTATSGL
VSWGPGRPADPVWMFQIQYNSSEDETLIYRIVPASSHHFLLKHLVPGADYDLCLLALSPAAGPSDLTATR
LLGCAHFSTLPASPLCHALQAHV
LGGTLTVAVGGVLVAALLVFTVALLV

RGRGAGNGRLPLKLSHVQSQTNGGPSPTPKAHPPRSPPPRPQRSCSLDLGDAGCYGYARRLGGAWARRSH
SVHGGLLGAGCRGVGGSAERLEESVV

Linx

Immunoglobulin superfamily containing leucine-rich repeat 2 [Homo sapiens]

GenBank: (NP_065902.1)AAI52430.1

745 aa

GenPept Identical Proteins Graphics

>gi|156230954|gb|AAI52430.1| Immunoglobulin superfamily containing leucine-rich repeat 2 [Homo sapiens]

MFPLRALWLVWALLGVAGSCPEPCACVDKYAHQFADCAYKELREVPEGLPANVTTLSLSANKITVLRRGA
FADVTQVTSLWLAHNEVRTVEPGALAVLSQLKNLDLSHNFISSFPWSDLRNLSALQLLKMNHNRLGSLPR
DALGALPDLRSLRINNNRLRTLAPGTFDALSALSHLQLYHNPFHCGCGLVWLQAWAASTRVSLPEPDSIA
CASPPALQGVPVYRLPALPCAPPSVHLSAEPPLEAPGTPLRAGLAFVLHCIADGHPTPRLQWQLQIPGGT
VVLEPPVLSGEDDGVGAEEGEGEGDGDLLTQTQAQTPTPAPAWPAPPATPRFLALANGSLLVPLLSAKEA
GVYTCRAHNELGANSTSIRVAVAATGPPKHAPGAGGEPDGQAPTSERKSTAKGRGNSVLPSKPEGKIKGQ
GLAKVSILGETETEPEEDTSEGEEAEDQILADPAEEQRCGNGDPSRYVSNHAFNQSAELKPHVFELGVIA
LDVAEREARVQLTPLAARWGPGPGGAGGAPRPGRRPLRLLYLCPAGGGAAVQWSRVEEGVNAYWFRGL
RP

GTNYSVCLALAGEACHVQVVFSTKKELPSL

LVIVAVSVFLLVLATVPLLGAAC

CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEKSYPAGGEAGGEEPEDVQGEGLDEDAEQG
DPSGDLQREESLAACSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEINGNYRQTAG

LRIT2 protein [Homo sapiens]

GenBank: AAI44476.1

560 aa

GenPept Identical Proteins Graphics

>gi|219518288|gb|AAI44476.1| LRIT2 protein [Homo sapiens]

MASVFHYFLLVLVFLDTHAAQPFCLPGCTCSEESFGRTLQCTSVSLGKIPGNLSEEFKQVRIENSPLFEM
PQGSFINMSTLEYLWLNFNNISVIHLGALEHLPELRELRLEGNKLCSVPWTAFRATPLLRVLDLKRNKID
ALPELALQFLVSLTYLDLSSNRLTVVSKSVFLNWPAYQKCRQPDCGAEILSSLVVALHDNPWVCDCRLRG
LVQFVKSITLPVILVNSYLICQGPLSKAGQLFHETELSACMKPQISTPSANITIRAGQNVTLRCLAQASP
SPSIAWTYPLSMWREFDGLLGGKHLTPVLTSSTGEDTALSELAIPAAHLVDSGNYTCMASNSIGKSNLVI
SLHVQPAQALHAPDSLSIPSEGNAYIDLRVVKQTVHGILLEWLAVADTSKEEWFTLYIASDEAFRKEVVH

IGPGINTYAVDDLLPGTKYEACLSLEGQPPHQGQCVAFVTGRDAGGLEAREH
LLHVTVVLCVVLLAVPVGAYAWAAQGPC
SCSKWVLRGCLHRRKAPSCTPAAPQSKDGSFREHPAVCDDGEGHIDTEGDKEKGGTEDNS

LRIT3 protein [Homo sapiens]

GenBank: AAI04038.1

552 aa

GenPept Identical Proteins Graphics

>gi|74355215|gb|AAI04038.1| LRIT3 protein [Homo sapiens]

MPLLRTLDLHNNKITSVPNEALRYLKNLAYLDLSSNRLTTLTPDFLENWTHLVSTPSGVLDLSPSRIILG
LQDNPWFCDCHISKMIELSKVVDPAIVLLDPLMTCSEPERLTGILFQRAELEHCLKPSVMTSATKIMSAL
GSNVLLRCDATGFPTPQITWTRSDSSPVNYTVIQESPEEGVRWSIMSLTGISSKDAGDYKCKAKNLAGMS
EAVVTVTVLGITTTPIPPDTSERTGDHPEWDVQPGSGRSTSVSSASSYLWSSSFSPTSSFSASTLSPPST
ASFSLSPFSSSTVSSTTTLSTSISASTTMANKRSFQLHQGGKRNLKVAKNGSKLPPASTSKKEELALLDQ
TMLTETNATIENLRVVSETKESVTLMWNMINTTHNSAVTVLYSKYGGKDLLLLNADSSKNQVTIDGLEPG
GQYMACVCPKGVPPQKDQCITFSTERVEGDDSQWS
LLLVVTSTACVVILPLICFLL
YKVCKLQCKSEPFW
EDDLAKETYIQFETLFPRSQSVGELWTRSHRDDSEKLLLCSRSSVESQVTFKSEGSRPEYYC

Leucine rich repeat containing 24 [Homo sapiens]

GenBank: AAI11068.1

513 aa

GenPept Identical Proteins Graphics

>gi|83405784|gb|AAI11068.1| Leucine rich repeat containing 24 [Homo sapiens]

MALRAPALLPLLLLLLPLRAAGCPAACRCYSATVECGALRLRVVPLGIPPGTQTLFLQDNNIARLEPGAL
APLAALRRLYLHNNSLRALEAGAFRAQPRLLELALTSNRLRGLRSGAFVGLAQLRVLYLAGNQLARLLDF
TFLHLPRLQELHLQENSIELLEDQALAGLSSLALLDLSRNQLGTISREALQPLASLQVLRLTENPWRCDC
ALHWLGAWIKEGGQRLLTSRDRKIMCAEPPRLALQSLLDVSHSSLICIPPSVHVQPLELTANLGEDLRVA
CQASGYPQPLVTWRKVPQPREGRPRAQAQLEGGLLGLGGHSASDTGSGMLFLSNITLAHAGKYECEASNA
GGAARVPFRLLVNASRQQPQQPAQPPPPAARPAGSEPRPEAGSMAFRALGVATQ
TAIAAAIALLALTALLLVAMI
CRRRRRRKKARGPPGEGALFVNDYLDGPCTFAQLEELRDERGHEMFVINRSKPLFAEGPAEAPAD
CGPAQGAGPGLRVPPPVAYEIHC

LRRC4C protein [Homo sapiens]

GenBank: AAH41374.3

640 aa

GenPept Identical Proteins Graphics

>gi|73909151|gb|AAH41374.3| LRRC4C protein [Homo sapiens]

MLNKMTLHPQQIMIGPRFNRALFDPLLVVLLALQLLVVAGLVRAQTCPSVCSCSNQFSKVICVRKNLREV
PDGISTNTRLLNLHENQIQIIKVNSFKHLRHLEILQLSRNHIRTIEIGAFNGLANLNTLELFDNRLTTIP
NGAFVYLSKLKELWLRNNPIESIPSYAFNRIPSLRRLDLGELKRLSYISEGAFEGLSNLRYLNLAMCNLR
EIPNLTPLIKLDELDLSGNHLSAIRPGSFQGLMHLQKLWMIQSQIQVIERNAFDNLQSLVEINLAHNNLT
LLPHDLFTPLHHLERIHLHHNPWNCNCDILWLSWWIKDMAPSNTACCARCNTPPNLKGRYIGELDQNYFT
CYAPVIVEPPADLNVTEGMAAELKCRASTSLTSVSWITPNGTVMTHGAYKVRIAVLSDGTLNFTNVTVQD
TGMYTCMVSNSVGNTTASATLNVTAATTTPFSYFSTVTVETMEPSQDEARTTDNNVGPTPVVDWETTNVT
TSLTPQSTRSTEKTFTIPVTDINSGIPGIDEVMKTTK
IIIGCFVAITLMAAVMLVIF
YKMRKQHHRQNHH
APTRTVEIINVDDEITGDTPMESHLPMPAIEHEHLNHYNSYKSPFNHTTTVNTINSIHSSVHEPLLIRMN
SKDNVQETQI

Leucine rich repeat containing 4 [Homo sapiens]

GenBank: AAI11562.1

653 aa

GenPept Identical Proteins Graphics

>gi|109730363|gb|AAI11562.1| Leucine rich repeat containing 4 [Homo sapiens]

MKLLWQVTVHHHTWNAILLPFVYLTAQVWILCAAIAAAASAGPQNCPSVCSCSNQFSKVVCTRRGLSEVP
QGIPSNTRYLNLMENNIQMIQADTFRHLHHLEVLQLGRNSIRQIEVGAFNGLASLNTLELFDNWLTVIPS
GAFEYLSKLRELWLRNNPIESIPSYAFNRVPSLMRLDLGELKKLEYISEGAFEGLFNLKYLNLGMCNIKD
MPNLTPLVGLEELEMSGNHFPEIRPGSFHGLSSLKKLWVMNSQVSLIERNAFDGLASLVELNLAHNNLSS
LPHDLFTPLRYLVELHLHHNPWNCDCDILWLAWWLREYIPTNSTCCGRCHAPMHMRGRYLVEVDQASFQ
C
SAPFIMDAPRDLNISEGRMAELKCRTPPMSSVKWLLPNGTVLSHASRHPRISVLNDGTLNFSHVLLSDTG
VYTCMVTNVAGNSNASAYLNVSTAELNTSNYSFFTTVTVETTEISPEDTTRKYKPVPTTSTGYQPAYTTS
TTVLIQTTRVPKQVAVPATDTTDKMQTSLDEVMKTTK
IIIGCFVAVTLLAAAMLIVF
YKLRKRHQQRSTV
TAARTVEIIQVDEDIPAATSAAATAAPSGVSGEGAVVLPTIHDHINYNTYKPAHGAHWTENSLGNSLHPT
VTTISEPYIIQTHTKDKVQETQI

Leucine rich repeat neuronal 1 [Homo sapiens]

GenBank: AAH34947.1

716 aa

GenPept Identical Proteins Graphics

>gi|23273823|gb|AAH34947.1| Leucine rich repeat neuronal 1 [Homo sapiens]

MARMSFVIAACQLVLGLLMTSLTESSIQNSECPQLCVCEIRPWFTPQSTYREATTVDCNDLRLTRIPSNL
SSDTQVLLLQSNNIAKTVDELQQLFNLTELDFSQNNFTNIKEVGLANLTQLTTLHLEENQITEMTDYCLQ
DLSNLQELYINHNQISTISAHAFAGLKNLLRLHLNSNKLKVIDSRWFDSTPNLEILMIGENPVIGILDMN
FKPLANLRSLVLAGMYLTDIPGNALVGLDSLESLSFYDNKLVKVPQLALQKVPSLKFLDLNKNPIHKIQE
GDFKNMLRLKELGINNMGELVSVDRYALDNLPELTKLEATNNPKLSYIHRLAFRSVPALESLMLNNNALN
AIYQKTVESLPNLREISIHSNPLRCDCVIHWINSNKTNIRFMEPLSMFCAMPPEYKGHQVKEVLIQDSSE
QCLPMISHDSFPNRLNVDIGTTVFLDCRAMAEPEPEIYWVTPIGNKITVETLSDKYKLSSEGTLEISNIQ
IEDSGRYTCVAQNVQGADTRVATIKVNGTLLDGTQVLKIYVKQTESHSILVSWKVNSNVMTSNLKWSSAT
MKIDNPHITYTARVPVDVHEYNLTHLQPSTDYEVCLTVSNIHQQTQKSCVNVTTKNAAFAVDISDQETST
A

LAAVMGSMFAVISLASIAV

YFAKRFKRKNYHHSLKKYMQKTSSIPLNELYPPLINLWEGDSEKDKDGSA
DTKPTQVDTSRSYYMW

Leucine rich repeat neuronal 2 [Homo sapiens]

GenBank: AAH68541.1

713 aa

GenPept Identical Proteins Graphics

>gi|46249796|gb|AAH68541.1| Leucine rich repeat neuronal 2 [Homo sapiens]

MRLLVAPLLLAWVAGATAAVPVVPWHVPCPPQCACQIRPWYTPRSSYREATTVDCNDLFLTAVPPALPAG
TQTLLLQSNSIVRVDQSELGYLANLTELDLSQNSFSDARDCDFHALPQLLSLHLEENQLTRLEDHSFAGL
ASLQELYLNHNQLYRIAPRAFSGLSNLLRLHLNSNLLRAIDSRWFEMLPNLEILMIGGNKVDAILDMNFR
PLANLRSLVLAGMNLREISDYALEGLQSLESLSFYDNQLARVPRRALEQVPGLKFLDLNKNPLQRVGPGD
FANMLHLKELGLNNMEELVSIDKFALVNLPELTKLDITNNPRLSFIHPRAFHHLPQMETLMLNNNALSAL
HQQTAESLPNLQEVGLHGNPIRCDCVIRWANATGTRVRFIEPQSTLCAEPPDLQRLPVREVPFREMTDHC
LPLISPRSFPPSLQVASGESMVLHCRALAEPEPEIYWVTPAGLRLTPAHAGRRYRVYPEGTLELRRVTAE
EAGLYTCVAQNLVGADTKTVSVVVGRALLQPGRDEGQGLELRVQETHPYHILLSWVTPPNTVSTNLTWSS
ASSLRGQGATALARLPRGTHSYNITRLLQATEYWACLQVAFADAHTQLACVWARTKEATSCHRALGDRP

GLIAILALAVLLLAAGLAAHLG

TGQPRKGVGGRRPLPPAWAFWGWSPPSVRVVSAPLVLPWNPGRKLPRSS

EGETLLPPLSQNS

Leucine rich repeat neuronal 3 [Homo sapiens]

GenBank: AAH35133.1

708 aa

GenPept Identical Proteins Graphics

>gi|23242678|gb|AAH35133.1| Leucine rich repeat neuronal 3 [Homo sapiens]

MKDMPLRIHVLLGLAITTLVQAVDKKVDCPRLCTCEIRPWFTPRSIYMEASTVDCNDLGLLTFPARLPAN
TQILLLQTNNIAKIEYSTDFPVNLTSLDLSQNNLSSVTNINVKKMPQLLSVYLEENKLTELPEKCLSELS
NLQELYINHNLLSTISPGAFIGLHNLLRLHLNSNRLQMINSKWFDALPNLEILMIGENPIIRIKDMNFKP
LINLRSLVIAGINLTEIPDNALVGLENLESISFYDNRLIKVPHVALQKVVNLKFLDLNKNPINRIRRGDF
SNMLHLKELGINNMPELISIDSLAVDNLPDLRKIEATNNPRLSYIHPNAFFRLPKLESLMLNSNALSALY
HGTIESLPNLKEISIHSNPIRCDCVIRWMNMNKTNIRFMEPDSLFCVDPPEFQGQNVRQVHFRDMMEICL
PLIAPESFPSNLNVEAGSYVSFHCRATAEPQPEIYWITPSGQKLLPNTLTDKFYVHSEGTLDINGVTPKE
GGLYTCIATNLVGADLKSVMIKVDGSFPQDNNGSLNIKIRDIHANSVLVSWKASSKILKSSVKWTAFVKT
ENSHAAQSARIPSDVKVYNLTHLNPSTEYKICIDIPTIYQKNRKKCVNVTTKGLHPDQKEYEKNNTTT
LMACLGGLLGIIGVICLISCLS
PEMNCDGGHSYVRNYLQKPTFALGELYPPLINLWEAGKEKSTSLKVKATV
IGLPTNMS

Neurotrophic tyrosine kinase, receptor, type 1 [Homo sapiens]

GenBank: AAI44240.1

790 aa

GenPept Identical Proteins Graphics

>gi|219841840|gb|AAI44240.1| Neurotrophic tyrosine kinase, receptor, type 1 [Homo sapiens]

MLRGGRRGQLGWHSWAAGPGSLLAWLILASAGAAPCPDACCPHGSSGLRCTRDGALDSLHHLPGAENLT
E
LYIENQQHLQHLELRDLRGLGELRNLTIVKSGLRFVAPDAFHFTPRLSRLNLSFNALESLSWKTVQGLSL
QELVLSGNPLHCSCALRWLQRWEEEGLGGVPEQKLQCHGQGPLAHMPNASCGVPTLKVQVPNASVDVGD
D
VLLRCQVEGRGLEQAGWILTELEQSATVMKSGGLPSLGLTLANVTSDLNRKNVTCWAENDVGRAEVSVQ
V
NVSFPASVQLHTAVEMHHWCIPFSVDGQPAPSLRWLFNGSVLNETSFIFTEFLEPAANETVRHGCLRLNQ
PTHVNNGNYTLLAANPFGQASASIMAAFMDNPFEFNPEDPIPDTNSTSGDPVEKKDETPFG
VSVAVGLAVFACLFLSTLLLVL
NKCGRRNKFGINRPAVLAPEDGLAMSLHFMTLGGSSLSPTEGKGSGLQGHIIENPQY

FSDACVHHIKRRDIVLKWELGEGAFGKVFLAECHNLLPEQDKMLVAVKALKEASESARQDFQREAELLTM
LQHQHIVRFFGVCTEGRPLLMVFEYMRHGDLNRFLRSHGPDAKLLAGGEDVAPGPLGLGQLLAVASQVA
A
GMVYLAGLHFVHRDLATRNCLVGQGLVVKIGDFGMSRDIYSTDYYRVGGRTMLPIRWMPPESILYRKFTT
ESDVWSFGVVLWEIFTYGKQPWYQLSNTEAIDCITQGRELERPRACPPEVYAIMRGCWQREPQQRHSIKD
VHARLQALAQAPPVYLDVLG

Neurotrophic tyrosine kinase, receptor, type 2 [Homo sapiens]

GenBank: AAH31835.1

477 aa

GenPept Identical Proteins Graphics

>gi|21594337|gb|AAH31835.1| Neurotrophic tyrosine kinase, receptor, type 2 [Homo sapiens]

MSSWIRWHGPAMARLWGFCWLVVGFWRAAFACPTSCKCSASRIWCSDPSPGIVAFPRLEPNSVDPENITE
IFIANQKRLEIINEDDVEAYVGLRNLTIVDSGLKFVAHKAFLKNSNLQHINFTRNKLTSLSRKHFRHLDL
SELILVGNPFTCSCDIMWIKTLQEAKSSPDTQDLYCLNESSKNIPLANLQIPNCGLPSANLAAPNLTVEE
GKSITLSCSVAGDPVPNMYWDVGNLVSKHMNETSHTQGSLRITNISSDDSGKQISCVAENLVGEDQDSVN
LTVHFAPTITFLESPTSDHHWCIPFTVKGNPKPALQWFYNGAILNESKYICTKIHVTNHTEYHGCLQLDN
PTHMNNGDYTLIAKNEYGKDEKQISAHFMGWPGIDDGANPNYPDVIYEDYGTAANDIGDTTNRSNEIPST
DVTDKTGREHLS
VYAVVVIASVVGFCLLVMLFLL
KLARHSKFGMKGFVLFHKIPLDG

Neurotrophic tyrosine kinase, receptor, type 3 [Homo sapiens]

GenBank: AAH13693.1

612 aa

GenPept Identical Proteins Graphics

>gi|15489168|gb|AAH13693.1| Neurotrophic tyrosine kinase, receptor, type 3 [Homo sapiens]

MDVSLCPAKCSFWRIFLLGSVWLDYVGSVLACPANCVCSKTEINCRRPDDGNLFPLLEGQDSGNSNGNAS
INITDISRNITSIHIENWRSLHTLNAVDMELYTGLQKLTIKNSGLRSIQPRAFAKNPHLRYINLSSNRLT
TLSWQLFQTLSLRELQLEQNFFNCSCDIRWMQLWQEQGEAKLNSQNLYCINADGSQLPLFRMNISQCDLP
EISVSHVNLTVREGDNAVITCNGSGSPLPDVDWIVTGLQSINTHQTNLNWTNVHAINLTLVNVTSEDNGF
TLTCIAENVVGMSNASVALTVYYPPRVVSLEEPELRLEHCIEFVVRGNPPPTLHWLHNGQPLRESKIIHV
EYYQEGEISEGCLLFNKPTHYNNGNYTLIAKNPLGTANQTINGHFLKEPFPESTDNFILFDEVSPTPPIT
VTHKPEEDTFG
VSIAVGLAAFACVLLVVLFVMI

NKYGRRSKFGMKGPVAVISGEEDSASPLHHINHGITT
PSSLDAGPDTVVIGMTRIPVIENPQYFRQGHNCHKPDTWVFSNIDNHGILNLKDNRDHLVPSTHYIYEEP
EVQSGEVSYPRSHGFREIMLNPISLPGHSKPLNHGIYVEDVNVYFSKGRHGF

Appendix B: Protein Sequence for other Species LIGS

Linx

immunoglobulin superfamily containing leucine-rich repeat protein 2 isoform a [Mus musculus]

NCBI Reference Sequence: NP_001155007.1

GenPept Identical Proteins Graphics

protein	1	fasta

>gi|238859603|ref|NP_001155007.1| immunoglobulin superfamily containing leucine-rich repeat protein 2 isoform a [Mus musculus]

MHSPFLPTATATDARSSLRLSPESGDRLAAPQHHTASQRAAGVTMGPFGALCLAWALLGVVRACPEPCAC
VDKYAHQFADCAYKELREVPEGLPANVTTLSLSANKITVLRRGAFVNVTQVTSLWLAHSEVRTVESGALA
VLSQLKNLDLSHNLISNFPWSDLRNLSALQLLKMNHNRLGSLPRDALGALPDLRSLRINNNRLRTLEPGT
FDALSALSHLQLYHNPFHCSCGLVWLQAWAASTRVSLPEPDSIACASPPELQGVPVHRLPALPCAPPSVR
LSAEPPPEAPGTPLRAGLAFMLHCVAEGHPTPRLQWQLQIPGGTVVLVPPVLSKEEDGGDKVEDGEGDGD
EDLPTQTEAPTPTPAPAWPAPPATPRFLALANGSLLVPLLSAKEAGIYTCRAHNELGTNSTSLRVTVAAA
GPPKHAPGTGEEPDAQVPTSERKATTKGRSNSVLPFKPEGKTKGQGLARVSVLGEIEAELEETDEGEQME
GQIPADPMGEKHCGHGDPSRYVSNHAFNQSSDLKPHVFELGVIALDVAEREARVQLTPLAARWGPGPDGA
SGARRPGRRPLRLLYLCPAGGGTAVQWSRVEEGVNAYWFRGLRPGTNYSVCLALAGEACHVQVVFSTKK
E
LPSL
LVIVTVSVFLLVLATVPLLGAAC
CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEK
SYPARGEAGGEEPEEVPEEGLDEDVEQGDPSGDLQREESLAGCSLVESQSKANQEEFEAGSEYSDRLPLG
AEAVNIAQEINGNYRQTAG

immunoglobulin superfamily containing leucine-rich repeat protein 2 precursor [Gallus gallus]

NCBI Reference Sequence: NP_001038132.1

GenPept Identical Proteins Graphics

>gi|113206126|ref|NP_001038132.1| immunoglobulin superfamily containing leucine-rich repeat protein 2 precursor [Gallus gallus]

MAPALWLWLAALLGSARACPEPCACVDKYAHQFADCAYKDLQVVPTGLPSNVTTLSLSANKITALQRRSF
VEVTQVTSLWLAHNEIRAIEPGAFAILVQLKNLDISHNQIVDFPWQDLYNLSALQLLKMNNNHMAVVPQG
AFHTLKDLRSLRINNNKFTTLAEGIFDSLSSLSHLQIYNNPFECSCKLQWLKKWMDSTLISIPEKESITC
SLPEQLRGVEVGKIPDTQCTSPSVQLTYYPNLDTTELFDGFTLTLHCAVTGAPPPEVSWKIRTSSQTLEL
SGSPSESAGKDPPRQDPERFLVFKNGTLVIPHLSKREEGTYTCLATNEMGSNQTSVNVAVAGSQKYPLQP
GRDPTGGKAQPGDKKPGAKGAKNSVLTPDERSKPLSPTRQSQPPSAAGMEPTGDGKVPFQLPPFEKKCGS
MPTSRYISNHAFNQSGDFKQHTFDLGVIALDVSERDARVQLTPTYVQPDKVHLRMLYLCQESSRGHALVQ
WSKIEEGVNSYWFQGLKPGTNYSVCLTYLGEDCQVQVVFTTKKEIPS

LIIIVVVSIFLLLLATLPLMGATWCHLL

SKYQGKTYKLIMKAQNPDQMEKHMAADFDPRASYLESEKNYNPSEVGEGEAEEEDEDEEDDDEGG
RRRRRREAEETTELEREESVAASSMAESQSKANGEEFEVRSEYSDKLPLGAEAVTISQEINGNYRQRPR

PREDICTED: immunoglobulin superfamily containing leucine-rich repeat protein 2-like [Callorhinchus milii]

NCBI Reference Sequence: XP_007906282.1

GenPept Identical Proteins Graphics

>gi|632938762|ref|XP_007906282.1| PREDICTED: immunoglobulin superfamily containing leucine-rich repeat protein 2-like [Callorhinchus milii]

MLEKLLCVISVGYVFCPWGVRGCPEPCVCQDKYFNQFADCAYKNFQAVPVGLPSNVTTLSLSANKIKSLL
RADFAEVTQVTSLWLAHNEIRKIEKGSLTVLLQLKNLDISHNQIVDFPWEDLYNLTALQLLKMNNNYMVH
LSRDAFSTLKELRSLRINSNKFHTIWEGTFDSLSSLSHLQIYSNPFSCTCNLQWLKGWIDQALISIPEQK
DIVCSAPEEFKGTPVVELPDMQCIAPLVHLTYQASNEKGELYEGYALTMHCNATGSPVPVIRWKIQTANK
EIELNDANVEPERNELLLENRKEVRDRFVVLKNGTLVIPHLTKYEEGAYTCLATNEIGSNRSTLNVAVTA
SPKREPTYIQERIPSQPGERKPGLKLPKNNAISWAKPGQKGQRISPATARSFPGQGTERNAVFLPPVAKN
CSKSQGSHYITNHAFNRSSEMKQHTFDYGIIALEVTETDAKVQLTPFQTAPDKISLEMLYLCAEQGGKAA
TVVQWSMIESGVNSYRFQGLNPGSNYTLCLTYTGQDCQVQVVFSTR

RKIPSLLIMIIVSSFLLGLATIPLVAATCCHLM

YKYRGKTYKLIMKTQPPESLHQNAPCTFDPSASFQGSEKIYNPSEVGEESVVAESVPVSQT
KANPEEFEACSEYSDRLPLGAEAVNISPEINGNYRQPVR

SALM1

leucine-rich repeat and fibronectin type III domain-containing protein 1 isoform 1 precursor [Mus musculus]

NCBI Reference Sequence: NP_001135393.1

GenPept Identical Proteins Graphics

>gi|213972562|ref|NP_001135393.1| leucine-rich repeat and fibronectin type III domain-containing protein 1 isoform 1 precursor [Mus musculus]

MAPGPFSSGLFSPPPAALPFLLLLWAGASRGQPCPGRCICQNVAPTLTMLCAKTGLLFVPPAIDRRVVEL
RLTDNFIAAVRRRDFANMTSLVHLTLSRNTIGQVAAGAFADLRALRALHLDSNRLAEVRGDQLRGLGNLR
HLILGNNQIRKVESAAFDAFLSTVEDLDLSYNNLEALPWEAVGQMVNLNTLTLDHNLIDHIAEGTFVQLH
KLVRLDMTSNRLHKLPPDGLFLRSQGGGPKPPTPLTVSFGGNPLHCNCELLWLRRLTREDDLETCATPEH
LTDRYFWSIPEEEFLCEPPLITRQAGGRALVVEGQAVSLRCRAVGDPEPVVHWVAPDGRLLGNSSRTRVR
GDGTLDVTITTLRDSGTFTCIASNAAGEATAPVEVCVVPLPLMAPPPAAPPPLTEPGSSDIATPGRPGAN
DSTSERRLVAAELTSSSVLIRWPAQRPVPGIRMYQVQYNSSADDSLVYYRMIPSTSQTFLVNDLAAGRAYD
LCVLAVYDDGATALPATRVVGCVQFTTAGDPAPCRPLRAHFLG
GTMIIAIGGVIVASVLVFIVLLMI
RYK
VYGDGDSRRIKGTSRTPPRVSHVCSQTNGAGAQQASAPPAPDRYEALREVAVPAAIEAKAMEAEATSTEL
EVVLGRSLGGSATSLCLLPSEETSGEESRAMTGPRRSRSGALGPPTSAPPTLALVPGGAPARPRPQQRYS
FDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARARLAFTSTEWMLESTV

PREDICTED: leucine-rich repeat and fibronectin type III domain-containing protein 1-like protein isoform X1 [Gallus gallus]

NCBI Reference Sequence: XP_423347.4

GenPept Identical Proteins Graphics

>gi|513240280|ref|XP_423347.4| PREDICTED: leucine-rich repeat and fibronectin type III domain-containing protein 1-like protein isoform X1 [Gallus gallus]

MMTVCPSPTMDRLLVCLLVVSAAVKAMLCPKRCMCQNLSPSFTILCTKTGLLFVPPSIDRRTAELRLMDN
FITTLRRKDFANMTNLIHLTLSRNTISQIMPYAFFDLKGLHALHLDSNRLTYINEDHFKGLINLRHLILS
NNQLSYISPGSLDDFIETIEDLDLSYNNLVNVPWETVAKLSNVNTVSLDHNLIEFVPEGIFSNLHKLARL
DMTSNKLKKIPPDPLFSRIPVYAKSKGSPLTSLVLSFGGNPLHCNCELVWLRRLTREDDLETCASPPELM
GKYFWSIKEEEFVCEPPMITHRTPKVAVSEGQSVSLKCKAVGDPDPYVRWIAPDGKLVSNTSRTTSYENG
TLDIAGTSLGDKGTFTCIASNAAGESTAPVELVVTPYPNLANSTNCEKEAENGPSDILISAKSSFPNETK
GPQERAVVVGELTSSSALIQWPSQQHLPGIRMFQIQYNSSSDEILVYRMIPAASKSFFLTDLVAGREYDL
CVLAVYDDGLTSLTATRVIGCVQFTTQEEYKQCRSLHAQF
LGGTMIIIIGGIIVASVLVFIFILLM
KYKVYNNHHKNKAAKVSNVCSQTNGSHGGSMARSTSKLTEGSHQECSASSSKGKAVLDSDGDKVTPTTH
TTFLT

TDPLS

PREDICTED: leucine-rich repeat and fibronectin type III domain-containing protein 1 [Callorhinchus milii]

NCBI Reference Sequence: XP_007907747.1

GenPept Identical Proteins Graphics

>gi|632981721|ref|XP_007907747.1| PREDICTED: leucine-rich repeat and fibronectin type III domain-containing protein 1 [Callorhinchus milii]

MESLLLCALVLVLGVTVTAQLCPKRCVCQNLSPSIAILCAKTGLLFVPPFIDRRTVELRLTDNFITSVRK
RDFANMTSLVHLTLSRNTISQIMPHSFGDLRGLRALHLDSNRLTKLVDAHLRGLVNLRHLILNNNQLNAI
SDGSFDDFLGSLEDLDMSYNNLETFPWEAISKMVNLNTLSLDHNLIDHIEEGTFSVLHKLSRLDMTSNRL
HKLPPDPLFLRTQLLVNTRGSHSFSLVLSFGGNPLHCNCELLWLRRLMREDDLETCASPPHLMGKYFWSI
AEEEFICEPPLITRLQATKTFVMEGQGVTLKCKAVGDPDPSILWSLPEGKLVSNTSRTIIYDNGTLDILI
TTLKDNGRFACIASNAAGESATNITIGIIPLPHFVNLTQHIKVPDPGSSDISTSSKPGAPSNSSDTKSTQ
DKKVTASELTTTSALVRWPSQRSIPGIRMYQIQYNSSSDNTLVYRMIPSTSQLFLVNDLAPGRDYELCVL
AVYDDGMTTLTATRAVGCVRFTTEQEYTQCHSVHTQF
LGGTMIIIIGGIIVASVLVFIIILMI
RYKVYSS
GLGDSKAVGTNVYSQTNGNGSHNGALDRSCSKPEGPGESVPEALVELPDQSQTVVLSVMCEKAGGAHTT
A
SATASASASVTVPTEGALPQAQRRRVQPGATGQHQHQQQLEPQTSSEEGHTEASTTDSSMSVCLISSSRG
TLPGRGKPAKLSNISLLPREISRTQHRHSFDGDYSLFQSHSYPRRARTKRSLTGSGQQLHCEDRRGTFSS
TEWMLESTV

SALM2

leucine-rich repeat and fibronectin type-III domain-containing protein 2 precursor [Mus musculus]

NCBI Reference Sequence: NP_081728.2

GenPept Identical Proteins Graphics

protein	1	fasta

>gi|226246673|ref|NP_081728.2| leucine-rich repeat and fibronectin type-III domain-containing protein 2 precursor [Mus musculus]

METLLGGLLAFGMAFAVVDACPKYCVCQNLSESLGTLCPSKGLLFVPPDIDRRTVELRLGGNFIIHIGRQ
DFANMTGLVDLTLSRNTISHIQPFSFLDLESLRSLHLDSNRLPSLGEDTLRGLVNLQHLIVNNNQLGGIA
DDAFEDFLLTLEDLDLSYNNLHGLPWDSVRRMVNLHQLSLDHNLLDHIAEGTFADLQKLARLDLTSNRLQ

KLPPDPIFARSQASLLTATPFAPPLSFSFGGNPLHCNCELLWLRRLERDDDLETCGSPGSLKGRYFWHIR
EEEFVCEPPLITQHTHKLLVLEGQAATLKCKAIGDPSPLIHWVAPDDRLVGNSSRTAVYDNGTLDILITT
SQDSGPFTCIAANAAGEATATVEVSIVQLPHLSNSTSRMAPPKSRLSDITGSSKTSRGGGGSGAGEPPKS
TPERAVLVSDVTTTSALVKWSVSKSAPRVKMYQLQYNCSDDEVLIYRMIPASNKAFVVNNLVSGTGYDLC
VLAMWDDTATTLTATNIVGCAQFFTKADYPQCQSMHSQI
LGGTMILVIGGIIVATLLVFIVILMV
RYKVC
NHDTPGKMAAATVSNVYSQTNGSQPPPLGGIPVGQLPQAPPKVVVRNELMDFSTSLARACDSSSSSSLGS
GEAAGLGRGPWRLPPPAPRPKPSLDRLMGAFASLDLKSQRKEELLDSRTPAGRGAGTSSRGHHSDREPLL
GPPATRARSLLPLPLEGKAKRSHSFDMGDFAAAAAAVPGGYSPPRRVSNIWTKRSLSVNGMLLPFEESDL
VGARGTFGSSEWVMESTV

PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 2 [Gallus gallus]

NCBI Reference Sequence: XP_004935432.1

GenPept Identical Proteins Graphics

>gi|513175233|ref|XP_004935432.1| PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 2 [Gallus gallus]

MEKLLCGILVFGMAVMVNACPKYCVCQNLSESLGTLCPSKGLLFVPLDIDRRTVELRLGGNFIINISRQD
FANMSGLVDLTLSRNTISYIQPYSFTDLESLRSLHLDSNRLPDIGEDILRGLINLQHLILNNNQLTSISD
EAFEDFLLTLEDLDLSYNNLRSIPWESIRKMINLHQLSLDHNLIDYITEGTFADLQKLARLDLTSNRLQK
LPPDPIFARSQVIPLAVTPFSPPLSLSFGGNPLHCNCELLWLRRLDRDDDMETCASPPGLKGRYFWYVRE
EEFVCEPPLITQHTHKLLVLEGQTATLKCKAIGDPTPIIHWVAPDDRLIGNSSRTSVYDNGTLDILITTS
KDYGTFTCIAANAAGESTATIELSIVQLPHLSNGTGRAAPPKSRLSDITSSSKSNRGETKGPPERAVLVS
EVTTTSALVKWTVSKSAPRVKMYQLQYNCSDDEVLIYRMIPATNKAFVVNNLVSGTGYDLCVLAMWDDT
A
TTLTATNIVGCAQFFTKEDYPQCQSMHSHFLGGT
MILIIGGIIVATLLVFIVILMV
RYKVCNNSQGKMSS
VSNVYSQTNGAQPVQNGVLPQVNPKVVVRNELMEFNSGSVRSSISSSSSSMNSRDCDNYSLQSEQGTLSS
KWRPPSRSKHNIDRLMGAFASLELKCQKKEETTDSRTSTAARHSDKEPLLGQPESKFRSLLMLPLEGKTK
RSHSFDMGDFATSQCCTYPKKITNIWTKRSLSVNGMLLQYDDNDLTGAKGTYGSSEWVMESTV

PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 2 [Callorhinchus milii]

NCBI Reference Sequence: XP_007908260.1

GenPept Identical Proteins Graphics

>gi|632982664|ref|XP_007908260.1| PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 2 [Callorhinchus milii]

MEKLLCNLLVIGMAVTVYACPKYCVCQNLSESLGTLCPSKGLLFVPPNIDRRTVELRLGGNFILSINRQD
FGNMTGLVDLTLSRNTIDYIQPYSFADLESLRSLHLDSNRLTRIGSNDFRGLLNLQHLILNNNQLNSILD
EAFDDFLLTLEDLDLSYNNLVSLPWEALGKMINLHTLSLDHNLIDYIPEGTFTDLLKLARLDLVSNRLQK
LPPDPIFARSETFVLSTTPYFAPLSLSIGGNPLHCNCELLWLRRLSREDDMETCASPSHLKGRYFWYVPE
EEFVCEQPLITQHSHKVLVLEGQTATLRCKAIGDPKPVIHWVAPDDRILGNSSRTVIYDNGTLDILITTS
KDYGTFTCIAANAAGESTASIELSIVQLPHLSNGTGRAVQPGSRLSDITSSSKTYRGETMSKPEKVVKVY
DVTASTALVKWSVGRSAPKVKMYQFQYNSSTDEVLVYRMIPASNKAFVVKNLVPSSNYDLCVLAIWDDT
L

TTLTATNVVGCVRFTTSEDYTQCKSFHSQ

FLGGTMILIIGGIIVASLLVFIIILTI

KYKLCNGQEKLPDV

NNVCSQTNGGQPVLNGILPQLNPKVVGRDEMLEFNCGSIHSSMSSSTGSSQDCEDCYSLNSNASTLSKKW
RHRSKSRHNIDRLMGAFASLDLRCQRKEDNCESRASTLAHYSDKEPLLGHSESRLNKLLTLPMEVKTKRS
HSFDMSDFATTPCYNYPRRITNIWTRRSLSVNGTLLQYDEEDLESTKGMYCSSEWVMESTV

SALM3

leucine-rich repeat and fibronectin type-III domain-containing protein 3 precursor [Mus musculus]

NCBI Reference Sequence: NP_780687.1

GenPept Identical Proteins Graphics

>gi|30425224|ref|NP_780687.1| leucine-rich repeat and fibronectin type-III domain-containing protein 3 precursor [Mus musculus]

MAVLPLLLCLLPLAPASSPPQPAISSPCPRRCRCQTQSMPLSVLCPGAGLLFVPPSLDRRAAELRLADNF
IAAVRRRDLANMTGLLHLSLSRNTIRHVAAGAFADLRALRALHLDGNRLTSLGEGQLRGLVNLRHLILSN
NQLAALAAGALDDCAETLEDLDLSYNNLEQLPWEALGRLGNVNTLGLDHNLLASVPAGAFSRLHKLARL
D

MTSNRLTTIPPDPLFSR

LPLLARPRGSPASALVLAFGGNPLHCNCELVWLRRLAREDDLEACASPPALGG

RYFWAVGEEEFVCEPPVVTHRSPPLAVPAGRPAALRCRAVGDPEPRVRWVSPQGRLLGNSSRARAFPNGT
LELLVTEPEDGGTFTCIAANAAGEATAAVELTVGPPPPPQLANSTSCDPPRDGEPDALTPPSAASASAKV
ADTVAPTDRGVQVTEHGATAALVQWPDQRPVPGIRMYQIQYNSSADDILVYRMIPADSRSFLLTDLASGR
TYDLCVLAVYEDSATGLTATRPVGCARFSTEPALRPCAAPHAPF

LGGTMIIALGGVIVASVLVFIFVLLL

RYKVHGGQPPGKAKATAPVSSVCSQTNGALGPVPSAPAPEPAAPRAHTVVQLDCEPWGPSHEPAGP

SALM4

leucine-rich repeat and fibronectin type-III domain-containing protein 4 precursor [Mus musculus]

NCBI Reference Sequence: NP_700437.2

GenPept Identical Proteins Graphics

protein	1	fasta

>gi|31559904|ref|NP_700437.2| leucine-rich repeat and fibronectin type-III domain-containing protein 4 precursor [Mus musculus]

MAPPLLLLLLASGAAACPLPCVCQNLSESLSTLCAHRGLLFVPPNVDRRTVELRLADNFIQALGPPDFRN
MTGLVDLTLSRNAITRIGARSFGDLESLRSLHLDGNRLVELGSSSLRGPVNLQHLILSGNQLGRIAPGAF
DDFLDSLEDLDVSYNNLRQVPWAGIGSMPALHTLNLDHNLIDALPPGVFAQLSQLSRLDLTSNRLATLAP
DPLFSRGRDAEASPSPLVLSFSGNPLHCNCELLWLRRLARPDDLETCASPPTLAGRYFWAVPEGEFSCEP
PLIARHTQRLWVLEGQRATLRCRALGDPVPTMHWVGPDDRLVGNSSRAWAFPNGTLEIGVTGAGDAGAY
T
CIATNPAGEATARVELRVLALPHGGNTSAEGGRPGPSDIAASARTAAEGEGTLESEPAVQVTEVTATSGL
VSWGLGRPADPVWMFQIQYNSSEDETLIYRIVPASSHHFLLKHLVPGADYDLCLLALSPAAGPSDLTATR
LLGCAHFSTLPATPLCHALQAHVLG
GTLTVAVGGVLVAALLVFTVALLV
RGRGAGNGRLPLKLSHVQSQT
NGGTSPMPKSHPPRSPPPRPQRSCSLDLGDTGGCYGYARRLGGAWARRSHSVHGGLLGAGCRGVGGSAER
LEESVV

PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 4-like [Callorhinchus milii]

NCBI Reference Sequence: XP_007909247.1

GenPept Identical Proteins Graphics

>gi|632984655|ref|XP_007909247.1| PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 4-like [Callorhinchus milii]

MEKFTFAFLLVGSLAAGSEACPFHCTCQNLSESLSTLCANKGLLFIPINIDRRTVELRLADNFLRVIAQP

DFLNMSGLVDLTLSRNTIISLEPFAFGDLESLRSLHLDSNRLIRIHEDSLRGLINLQHLIINNNQLINIA
LSAFDDFVVTLEDLDLSFNNLQRVPWEAIQSMVNLHMLNLDHNLIDYIMADTFAELFKLARLDMTSNRLQ
TLPPDSLFSRSQTGVINPTPYTSIIILNFGGNPLHCNCELLWLRRLVREDDMETCASPAHLAGRYFWSIP
EEEFICEPPLITRHTHKVWILEGQRATLKCRAIGDPEPIIHWVSPEDKIVSNSSRIVSYRNGTLDILVTT
MREDGVYTCFATNAAGESTALADLKIIPLPHRGNGTLQILHHDPGSSDISTSTKPVTNSTGRSRPRDKTV
SVTDVTGTTALIRWAQSKSPHIVWMYQIQYNCSIDETLVYRIISSKSKAFILKNLISGVDYDLCILAIYD
DSVTQLAATKVVGCIQFSTHEEYPHCHLLHAHF
LGGTLTVIVGGIIVVTLLVFTVIMMV
KYKVCGSARCE
VPKLTDVYSQTNGSQTTVPNGMVSAQRITVLNTRGQPTGGVPVPDLSSANLPRQESRKAPPYSAKTQRKR
YKCKQRGEGDGELATLGCQGGEGPGERTALAKQPCPQSSE

SALM5

leucine-rich repeat and fibronectin type-III domain-containing protein 5 isoform 1 precursor [Mus musculus]

NCBI Reference Sequence: NP_848829.2

GenPept Identical Proteins Graphics

>gi|31559842|ref|NP_848829.2| leucine-rich repeat and fibronectin type-III domain-containing protein 5 isoform 1 precursor [Mus musculus]

MEKFLFYLFLIGIAVRAQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRLADNFVTNIKRKD
FANMTSLVDLTLSRNTISFITPHAFADLRNLRALHLNSNRLTKITNDMFSGLSNLHHLILNNNQLTLISS
TAFDDVFALEELDLSYNNLETIPWDAVEKMVSLHTLSLDHNMIDNIPKGTFSHLHKMTRLDVTSNKLQKL
PPDPLFQRAQVLATSGIISPSTFALSFGGNPLHCNCELLWLRRLSREDDLETCASPALLTGRYFWSIPEE
EFLCEPPLITRHTHEMRVLEGQRATLRCKARGDPEPAIHWISPEGKLISNATRSLVYDNGTLDILITTVK
DTGAFTCIASNPAGEATQTVDLHIIKLPHLLNSTNHIHEPDPGSSDISTSTKSGSNASSSNGDTKMSQDK
IVVAEATSSTALLKFNFQRNIPGIRMFQIQYNGTYDDTLVYRMIPPTSKTFLVNNLASGTMYDLCVLAIY
DDGITSLTATRVVGCIQFTTEQDYVRCHFMQSQFL
GGTMIIIIGGIIVASVLVFIIILMI
RYKVCNNNGQ
HKVTKVSNVYSQTNGAQMQGCSVTLPQSMSKQAMGHEENAQCCKVASDNAIQSSETCSSQDSSTTTSALP
PTWTSSAPVSQKQKRKTGTKPSAEPQSEAVTNVESQNTNRNNSTALQLASCPPDSVTEGPTSQRAHTKPS
KFLTVPAEGSRARHRASLSGGLKDSFHYGNSQLSLKRSMSMNAMWT

PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 5 [Gallus gallus]

NCBI Reference Sequence: XP_421485.2

GenPept Identical Proteins Graphics

>gi|118092246|ref|XP_421485.2| PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 5 [Gallus gallus]

MEKLLLFLLFIGIAVRAQICPKRCVCQILSPNLATLCAKKGLLFVPPNIDRRTVELRLADNFVTNIKRKD
FANMTSLVDLTLSRNTISFITPHAFADLRNLRALHLNSNRLTKITNDMFSGLSNLHHLILNNNQLTLISS
TAFDDVLALEELDLSYNNLETIPWDAVEKMVSLHTLSLDHNMIDHIPKGTFSHLHKMTRLDVTSNKLQKL
PPDPLFQRAQVLATSGIISPSTFALSFGGNPLHCNCELLWLRRLSREDDLETCASPQLLSGRYFWSIPEE
EFLCEPPLITRHTHELRVLEGQRAALRCKARGDPEPAIHWISPEGKLISNATRSVVYDNGTLDILITTVK
DTGSFTCIASNPAGEATQTVDLHIIKLPHLLNSTNHIHEPDPGSSDISTSTKSGSNASSSNGDTKVSQDK
KVVVAEATSSTALLKFNFQRNIPGIRMFQIQYNGTYDDSLVYRMIPPTSKTFLVNNLAAGTMYDLCVLAI
YDDGITSLTATRVVGCTQFTTEQDYVRCHFMQSQFL
GGTMIIIIGGIIVASVLVFIIILMI
RYKVCNNNG
QHKATKVSNVYSQTNGAQVQACGGALSQSASKQAVGHEEAAQCCRAASDGAGPSPEPSPGPEATAATTTS
PSPHAWAAGTSAAQKPKRKPGPKPSSEPQSEAAMSIESQNTNRNNSTALQLASRPPDSDKGVPTYKRAQS
KPKAGADLKDTHTAPLLESSCPNLATRQKTKRSQRTKD

PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 5 [Callorhinchus milii]

NCBI Reference Sequence: XP_007891516.1

GenPept Identical Proteins Graphics

>gi|632951833|ref|XP_007891516.1| PREDICTED: leucine-rich repeat and fibronectin type-III domain-containing protein 5 [Callorhinchus milii]

MEKLLFYLLLIGMAVKAQVCPKRCVCQNLSPNLATLCAKKGLLFVPPNIDRRTVELRLGDNFITSIKRKD
FANMTGLVDLTLSRNTINHIAPQAFSDLCNLRALHLNSNRLTQITNEMFSRLSKLHHLIVNNNQLIEISS
GAFSDILLSLEELDVSYNNLKTIPWEAVEKMVNLHTLSLDHNMLEHIDEGTFSHLHKLIRLDMTSNKLRK
LPPDPLFTRVQVLANLGIMNPTGFVLSFGGNPLHCNCELLWLRRLSREDDLETCASPTHLTGRYFWSIPE
EEFICDQPLITRHTHELRVLEGQRATLKCKAIGDPDPSIHWSSPEGKLISNMSRTVLYANGTLDILITTV
KDTGTFTCIASNAAGETTAMVELHIIKLPHLINSTNHIHEPDPGSSDISTSTKSGSNTSNSVSDTKVKPE
RRVAVAETTSSSALIKFNLQHNIPGIRMFQIQYNGSYDDSLVYRMIPSTSKTFLVTNLAAGTLYDLCVLA
IYDDGITSLTATRVVGCVEFTTDQDYVRCHFMPSQFL
GGTMIIIIGGIIVASVLVFIIILMI
RYKVCNNN
DQHKMTKVSNVYSQTNGAHLQMCGSVLSHSNSKVAMGHDDNITRCNKDPSESKTQLSESTLSQDCSTTTS
TLPHDWTASVSPSQKLKRKAGLNPSVESPMEAFTNVESLKKKENTAILQKSTCAQISLKDTPTFRRAHSK
SIKFLTLPTEISRAKRRYSLDAEVSEYHCYTHSQSINSLWSKRSMSMNGMLLQLANSDVDGGKAVFSSSE

WIMESTV

Appendix C: IC domain

LINX

>HsLinx

CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEKSYPAGGEAGGEEPEDVQGEGLDEDAEQG
DPSGDLQREESLAACSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEINGNYRQTAG

>MmLinx

CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEK
SYPARGEAGGEEPEEVPEEGLDEDVEQGDPSGDLQREESLAGCSLVESQSKANQEEFEAGSEYSDRLPLG
AEAVNIAQEINGNYRQTAG

>GgLinx

SKYQGKTYKLIMKAQNPDQMEKHMAADFDPRASYLESEKNYNPSEVGEGEAEEEDEDEEDDDEGG
RRRRRREAEETTELEREESVAASSMAESQSKANGEEFEVRSEYSDKLPLGAEAVTISQEINGNYRQRPR

>CmLinx

YKYRGKTYKLIMKTQPPESLHQNAPCTFDPSASFQGSEKIYNPSEVGEESVVAESVPVSQT
KANPEEFEACSEYSDRLPLGAEAVNISPEINGNYRQPVR

SALM1

>HsSALM1

RYKVCNHEAPSKMAAAVSNVYSQTNGAQPPPPSSAPAGAPPQGPPKVVVRNELLDFTASLARASDSSSSSS
LGSGEAAGLGRAPWRIPPSAPRPKPSLDRLMGAFASLDLKSQRKEELLDSRTPAGRGAGTSARGHHSDREP
LLGPPAARARSLLPLPLEGKAKRSHSFDMGDFAAAAAGGVVPGGYSPPRKVSNIWTKRSLSVNGMLLPFEE
SDLVGARGTFGSSEWVMESTV

>MmSALM1

RYKVC
NHDTPGKMAAATVSNVYSQTNGSQPPPLGGIPVGQLPQAPPKVVVRNELMDFSTSLARACDSSSSSSLGS
GEAAGLGRGPWRLPPPAPRPKPSLDRLMGAFASLDLKSQRKEELLDSRTPAGRGAGTSSRGHHSDREPLL

GPPATRARSLLPLPLEGKAKRSHSFDMGDFAAAAAAVPGGYSPPRRVSNIWTKRSLSVNGMLLPFEESDL
VGARGTFGSSEWVMESTV

>GgSALM1
KYKVYNNHHKNKAAKVSNVCSQTNGSHGGSMARSTSKLTEGSHQECSASSSKGKAVLDSDGDKVTPTTH
TTFLT
TDPLS

>CmSALM1
RYKVYSS
GLGDSKAVGTNVYSQTNGNGSHNGALDRSCSKPEGPGESVPEALVELPDQSQTVVLSVMCEKAGGAHTT
A
SATASASASVTVPTEGALPQAQRRRVQPGATGQHQHQQQLEPQTSSEEGHTEASTTDSSMSVCLISSSRG
TLPGRGKPAKLSNISLLPREISRTQHRHSFDGDYSLFQSHSYPRRARTKRSLTGSGQQLHCEDRRGTFSS
TEWMLESTV

SALM2
>HsSALM2
RYKVYGDGDSRRVKGSRSLPRVSHVCSQTNGAGTGAAQAPALPAQDHYEALREVESQAAPAVAVEAKA
MEAETASAEPEVVLGRSLGGSATSLCLLPSEETSGEESRAAVGPRRSRSGALEPPTSAPPTLALVPGGAAAR
PRPQQRYSFDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARACLAFTSTEWM
LESTV

>MmSALM2
RYK
VYGDGDSRRIKGTSRTPPRVSHVCSQTNGAGAQQASAPPAPDRYEALREVAVPAAIEAKAMEAEATSTEL
EVVLGRSLGGSATSLCLLPSEETSGEESRAMTGPRRSRSGALGPPTSAPPTLALVPGGAPARPRPQQRYS
FDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARARLAFTSTEWMLESTV

>GgSALM2
RYKVCNNSQGKMSS
VSNVYSQTNGAQPVQNGVLPQVNPKVVVRNELMEFNSGSVRSSISSSSSSMNSRDCDNYSLQSEQGTLSS
KWRPPSRSKHNIDRLMGAFASLELKCQKKEETTDSRTSTAARHSDKEPLLGQPESKFRSLLMLPLEGKTK
RSHSFDMGDFATSQCCTYPKKITNIWTKRSLSVNGMLLQYDDNDLTGAKGTYGSSEWVMESTV

>CmSALM2

KYKLCNGQEKLPDV

NNVCSQTNGGQPVLNGILPQLNPKVVGRDEMLEFNCGSIHSSMSSSTGSSQDCEDCYSLNSNASTLSKKW

RHRSKSRHNIDRLMGAFASLDLRCQRKEDNCESRASTLAHYSDKEPLLGHSESRLNKLLTLPMEVKTKRS

HSFDMSDFATTPCYNYPRRITNIWTRRSLSVNGTLLQYDEEDLESTKGMYCSSEWVMESTV

SALM3

>HsSALM3

RGRGAGNGRLPLKLSHVQSQTNGGPSPTPKAHPPRSPPPRPQRSCSLDLGDAGCYGYARRLGGAWARRSH

SVHGGLLGAGCRGVGGSAERLEESVV

>MmSALM3

RGRGAGNGRLPLKLSHVQSQT

NGGTSPMPKSHPPRSPPPRPQRSCSLDLGDTGGCYGYARRLGGAWARRSHSVHGGLLGAGCRGVGGSAER

LEESVV

>GgSALM3

>CmSALM3

SALM4

>GgSALM4

>HsSALM4

MRYKVHGGQPPGKAKIPAPVSSVCSQTNGALGPTPTPAPPAPEPAALRAHTVVQLDCEPWGPGHEPVGP

>MmSALM4

RYKVHGGQPPGKAKATAPVSSVCSQTNGALGPVPSAPAPEPAAPRAHTVVQLDCEPWGPSHEPAGP

>CmSALM4

KYKVCGSARCE

VPKLTDVYSQTNGSQTTVPNGMVSAQRITVLNTRGQPTGGVPVPDLSSANLPRQESRKAPPYSAKTQRKR

YKCKQRGEGDGELATLGCQGGEGPGERTALAKQPCPQSSE

SALM5

>HsSALM5

RYKVCNNNGQHKVTKVSNVYSQTNGAQIQGCSVTLPQSVSKQAVGHEENAQCCKATSDNVIQSSETCSSQ
DSSTTTSALPPSWTSSTSVSQKQKRKTGTKPSTEPQNEAVTNVESQNTNRNNSTALQLASRPPDSVTEGPTS
KRAHIKPNALLTNVDQIVQETQRLELI

>MmSALM5

RYKVCNNNGQ
HKVTKVSNVYSQTNGAQMQGCSVTLPQSMSKQAMGHEENAQCCKVASDNAIQSSETCSSQDSSTTTSALP
PTWTSSAPVSQKQKRKTGTKPSAEPQSEAVTNVESQNTNRNNSTALQLASCPPDSVTEGPTSQRAHTKPS
KFLTVPAEGSRARHRASLSGGLKDSFHYGNSQLSLKRSMSMNAMWT

>GgSALM5

RYKVCNNNG
QHKATKVSNVYSQTNGAQVQACGGALSQSASKQAVGHEEAAQCCRAASDGAGPSPEPSPGPEATAATTTS
PSPHAWAAGTSAAQKPKRKPGPKPSSEPQSEAAMSIESQNTNRNNSTALQLASRPPDSDKGVPTYKRAQS
KPKAGADLKDTHTAPLLESSCPNLATRQKTKRSQRTKD

>CmSALM5

RYKVCNNN
DQHKMTKVSNVYSQTNGAHLQMCGSVLSHSNSKVAMGHDDNITRCNKDPSESKTQLSESTLSQDCSTTTS
TLPHDWTASVSPSQKLKRKAGLNPSVESPMEAFTNVESLKKKENTAILQKSTCAQISLKDTPTFRRAHSK
SIKFLTLPTEISRAKRRYSLDAEVSEYHCYTHSQSINSLWSKRSMSMNGMLLQLANSDVDGGKAVFSSSE
WIMESTV

Appendix D: Boxshade Outputs of LIGS

```
Boxshade for 1.0 match
HsLinx       1 CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEKSYPAGGEAGGEEPEDVQ
MmLinx       1 CHLLAKHPGKPYRLILRPQAPDPMEKRIAADFDPRASYLESEKSYPARGEAGGEEPEEVP
GgLinx       1 ----SKYQGKTYKLIMKAQNPDQMEKHMAADFDPRASYLESEKNYNPSEVGEGEAEEEDE
CmLinx       1 ----YKYRGKTYKLIMKTQPPESLHQNAPCTFDPSASFQGSEKIYNPSEVGEES------
consensus    1      K  GK YrLIlr Q Pd m       FDP ASy  SEK Y     a
```

```
HsLinx       61 ------GEGLDEDAEQGDPSGDLQREESLAACSLVESQSKANQEEFEAGSEYSDRLPLGA
MmLinx       61 ------EEGLDEDVEQGDPSGDLQREESLACCSLVESQSKANQEEFEAGSEYSDRLPLGA
GgLinx       57 DEEDDDEGGRRRRRREAEETTELEREESVAASSMAESQSKANGEEFEVRSEYSDKLPLGA
CmLinx       51 ----------------------------VVAESVPVSQTKANPEEFEACSEYSDRLPLGA
consensus    61                       L REESl a Sl  SQsKAN EEFE  SEYSDrLPLGA

HsLinx      115 EAVNIAQEINGNYRQTAG
MmLinx      115 EAVNIAQEINGNYRQTAG
GgLinx      117 EAVTISQEINGNYRQRPR
CmLinx       83 EAVNISPEINGNYRQPVR
consensus   121 EAVNI QEINGNYRQ
```

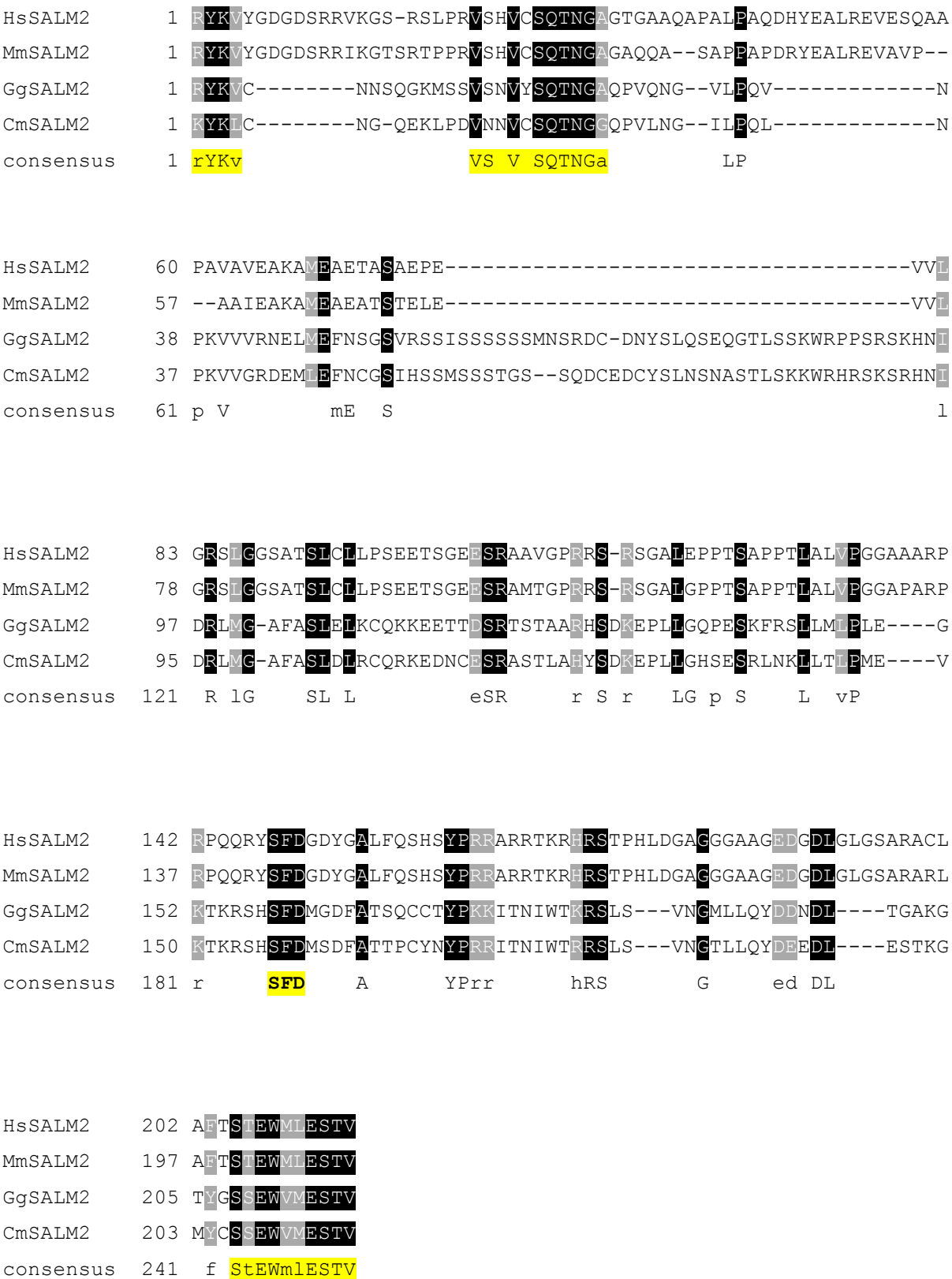
SALM1

0.7 match

```
CmSALM1       1 RYKVYSSGLGD--SKAVGTNVYSQTNGNGSHNGALDRSCSKPEGPGESV-PEALVELPDQ
GgSALM1       1 KYKVYNNHHK--NKAAKVSNVCSQTNGSHGGSMARSTS--------K---------LTE-
HsSALM1       1 RYKVCNHEAPSKMAA-AVSNVYSQTNGAQPPPPSSAPAGAPPQGPPKVVVRNELLDFTA-
MmSALM1       1 RYKVCNHDTPGKMAAATVSNVYSQTNGSQPPPLGGIPVGQLPQAPPKVVVRNELMDFST-
consensus     1 rYKV n        aa vsNVySQTNG      a       p gp k v    lve t

CmSALM1      58 SQTVVLSVMCEKAGGAHTTA-SATASASASVTVPTEGALPQAQRRRVQPGATGQ----HQ
GgSALM1      41 ----GSHQECSASSSK--GKAV---LDSDGDKVTPTTHTTFLTTDPLS-----------
HsSALM1      59 ----SLARASDSSSSSLGSGEAACLGRAPWRIPPSAPRPKPSLDRLMGAFASLDLKSQR
MmSALM1      60 ----SLARACDSSSSSLGSGEAACLGRGPWRLPPPAPRPKPSLDRLMGAFASLDLKSQR
consensus    61      l   ce sss   g a a ala a  kvpp g  p  t drl  g

CmSALM1     113 HQQQLEPQTSSEEGHTEASTTDSSMSVCLISSSRGTLPGRGKPAKLSNISLLPREIS-RT
GgSALM1         ------------------------------------------------------------
HsSALM1     115 KEELLDSRTPAGRGAGTSARGH--------HSDRE--PLLGPPAA-RARSLLPLPLEGKA
MmSALM1     116 KEELLDSRTPAGRGAGTSSRGH--------HSDRE--PLLGPPAT-RARSLLPLPLEGKA
```

```
consensus  121 h   le  t    g               s r   p  g pa     sllp  i  r


CmSALM1    172 QHRHSFD-GDYSLFQ-------SHSYPR---RARTKRSLTGSGQQL-----HCEDRRGTF
GgSALM1        ------------------------------------------------------------
HsSALM1    164 KRSHSFDMGDFAAAAAGGVVPGGYSPPRKVSNIWTKRSLSVNGMLLPFEESDLVGARGTF
MmSALM1    165 KRSHSFDMGDFAAAA--AAVPGGYSPPRRVSNIWTKRSLSVNGMLLPFEESDLVGARGTF
consensus  181  h hsfd gdy             s pr      tkrslt  g  l          rgtf


CmSALM1    216 SSTEWMLESTV
GgSALM1        -----------
HsSALM1    224 GSSEWVMESTV
MmSALM1    223 GSSEWVMESTV
consensus  241  stewmlestv
```

SALM2


HsSALM2 1 RYKVYGDGDSRRVKGS-RSLPRVSHVCSQTNGAGTGAAQAPALPAQDHYEALREVESQAA
MmSALM2 1 RYKVYGDGDSRRIKGTSRTPPRVSHVCSQTNGAGAQQA--SAPPAPDRYEALREVAVP--
GgSALM2 1 RYKVC--------NNSQGKMSSVSNVYSQTNGAQPVQNG--VLPQV----------N
CmSALM2 1 KYKLC--------NG-QEKLPDVNNVCSQTNGCQPVLNG--ILPQL----------N
consensus 1 rYKv VS V SQTNGa LP
HsSALM2 60 PAVAVEAKAMEAETASAEPE----------VVL
MmSALM2 57 --AAIEAKAMEAEATSTELE----------VVL
GgSALM2 38 PKVVVRNELMEFNSGSVRSSISSSSSSMNSRDC-DNYSLQSEQGTLSSKWRPPSRSKHNI
CmSALM2 37 PKVVGRDEMLEFNCGSIHSSMSSSTGS--SQDCEDCYSLNSNASTLSKKWRHRSKSRHNI
consensus 61 p V mE S l
HsSALM2 83 GRSLGGSATSLCLLPSEETSGEESRAAVGPRRS-RSGALEPPTSAPPTLALVPGGAAARP
MmSALM2 78 GRSLGGSATSLCLLPSEETSGEESRAMTGPRRS-RSGALGPPTSAPPTLALVPGGAPARP
GgSALM2 97 DRLMG-AFASLELKCQKKEETTDSRTSTAARHSDKEPLLGQPESKFRSLLMLPLE----G
CmSALM2 95 DRLMG-AFASLDLRCQRKEDNCESRASTLAHYSDKEPLLGHSESRLNKLLTLPME----V
consensus 121 R lG SL L eSR r S r LG p S L vP
HsSALM2 142 RPQQRYSFDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARACL
MmSALM2 137 RPQQRYSFDGDYGALFQSHSYPRRARRTKRHRSTPHLDGAGGGAAGEDGDLGLGSARARL
GgSALM2 152 KTKRSHSFDMGDFATSQCCTYPKKITNIWTKRSLS---VNGMLLQYDDNDL----TGAKG
CmSALM2 150 KTKRSHSFDMSDFATTPCYNYPRRITNIWTRRSLS---VNGTLLQYDEEDL----ESTKG
consensus 181 r SFD A YPrr hRS G ed DL
HsSALM2 202 AFTSTEWMLESTV
MmSALM2 197 AFTSTEWMLESTV
GgSALM2 205 TYGSSEWVMESTV
CmSALM2 203 MYCSSEWVMESTV
consensus 241 f StEWmlESTV

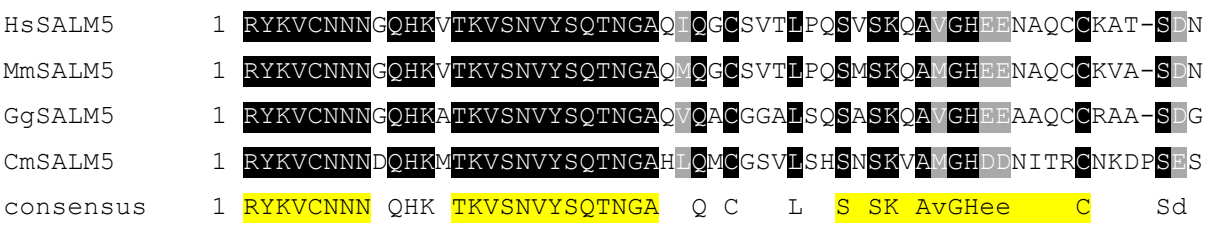
SALM3

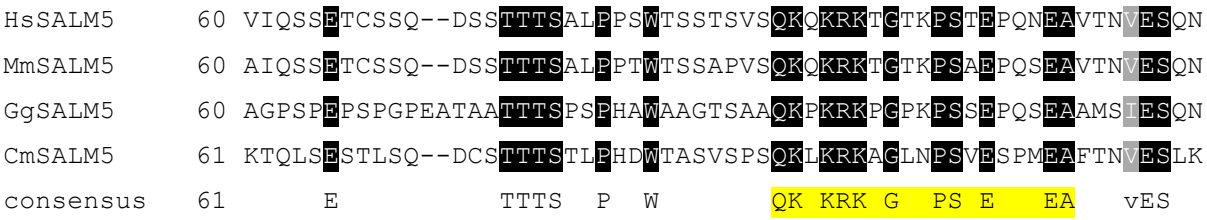
```
HsSALM3       1 RGRGAGNGRLPLKLSHVQSQTNGGPSPTPKAHPPRSPPPRPQRSCSLDLGDA-GCYGYAR
MmSALM3       1 RGRGAGNGRLPLKLSHVQSQTNGGTSPMPKSHPPRSPPPRPQRSCSLDLGDTGGCYGYAR
consensus     1 RGRGAGNGRLPLKLSHVQSQTNGG SP PK HPPRSPPPRPQRSCSLDLGD  GCYGYAR

HsSALM3      60 RLGGAWARRSHSVHGGLLGAGCRGVGGSAERLEESVV
MmSALM3      61 RLGGAWARRSHSVHGGLLGAGCRGVGGSAERLEESVV
consensus    61 RLGGAWARRSHSVHGGLLGAGCRGVGGSAERLEESVV
```

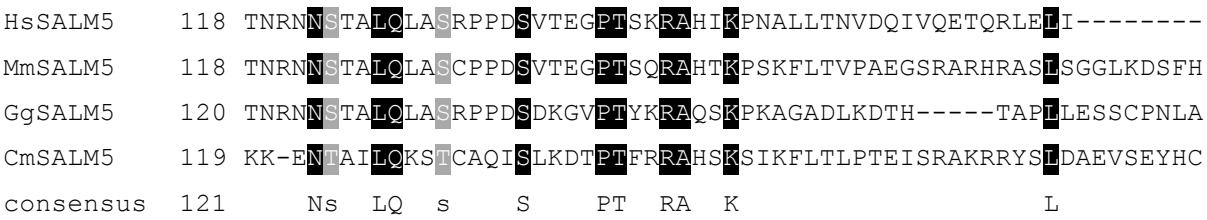
SALM4

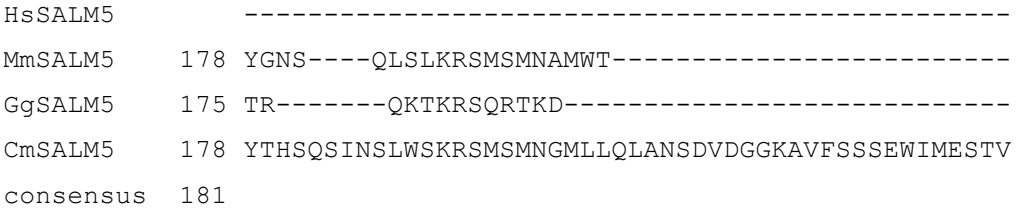
```
HsSALM4       1 MRYKVHGGQPPGKAKIPAPVSSVCSQTNGALGPTPTPA------------------PPAP
MmSALM4       1 -RYKVHGGQPPGKAKATAPVSSVCSQTNGALGPVP--S------------------APAP
CmSALM4       1 -KYKVCGSARCEVPK----LTDVYSQTNGSQTTVPNGMVSAQRITVLNTRGQPTGGVPVP
consensus     1  rYKV G      K    vs V SQTNG     P                     P P

HsSALM4      43 EP--------------AALRAHTVVQLDCEPWGPGHEPVGP------------------
MmSALM4      40 EP--------------AAPRAHTVVQLDCEPWGPSHEPAGP------------------
CmSALM4      56 DLSSANLPRQESRKAPPYSAKTQRKRYKCKQRGEGDGELATLGCQGGEGPGERTALAKQP
consensus    61 e                           C   G      g

HsSALM4         ------
MmSALM4         ------
CmSALM4     116 CPQSSE
consensus   121
```

SALM5


HsSALM5 1 RYKVCNNNGQHKVTKVSNVYSQTNGAQIQGCSVTLPQSVSKQAVGHEENAQCCKAT-SDN
MmSALM5 1 RYKVCNNNGQHKVTKVSNVYSQTNGAQMQGCSVTLPQSMSKQAMGHEENAQCCKVA-SDN
GgSALM5 1 RYKVCNNNGQHKATKVSNVYSQTNGAQVQACGGALSQSASKQAVGHEEAAQCCRAA-SDG
CmSALM5 1 RYKVCNNNDQHKMTKVSNVYSQTNGAHLQMCGSVLSHSNSKVAMGHDDNITRCNKDPSESS
consensus 1 RYKVCNNN QHK TKVSNVYSQTNGA Q C L S SK AvGHee C Sd


HsSALM5 60 VIQSSETCSSQ--DSSTTTSALPPSWTSSTSVSQKQKRKTGTKPSTEPQNEAVTNVESQN
MmSALM5 60 AIQSSETCSSQ--DSSTTTSALPPTWTSSAPVSQKQKRKTGTKPSAEPQSEAVTNVESQN
GgSALM5 60 AGPSPEPSPGPEATAATTTSPSPHAWAAGTSAAQKPKRKPGPKPSSEPQSEAAMSIESQN
CmSALM5 61 KTQLSESTLSQ--DCSTTTSTLPHDWTASVSPSQKLKRKAGLNPSVESPMEAFTNVESLK
consensus 61 E TTTS P W QK KRK G PS E EA vES

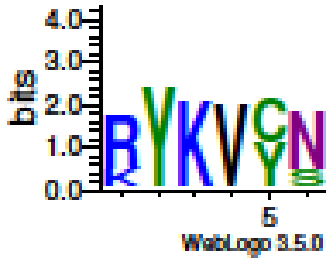

HsSALM5 118 TNRNNSTALQLASRPPDSVTEGPTSKRAHIKPNALLTNVDQIVQETQRLELI--------
MmSALM5 118 TNRNNSTALQLASCPPDSVTEGPTSQRAHTKPSKFLTVPAEGSRARHRASLSGGLKDSFH
GgSALM5 120 TNRNNSTALQLASRPPDSDKGVPTYKRAQSKPKAGADLKDTH-----TAPLLESSCPNLA
CmSALM5 119 KK-ENTAILQKSTCAQISLKDTPTFRRAHSKSIKFLTLPTEISRAKRRYSLDAEVSEYHC
consensus 121 Ns LQ s S PT RA K L


HsSALM5 ----------
MmSALM5 178 YGNS----QLSLKRSMSMNAMWT----------
GgSALM5 175 TR-------QKTKRSQRTKD----------
CmSALM5 178 YTHSQSINSLWSKRSMSMNGMLLQLANSDVDGGKAVFSSSEWIMESTV
consensus 181

Appendix E: Motifs and Fingerprints of LIGS

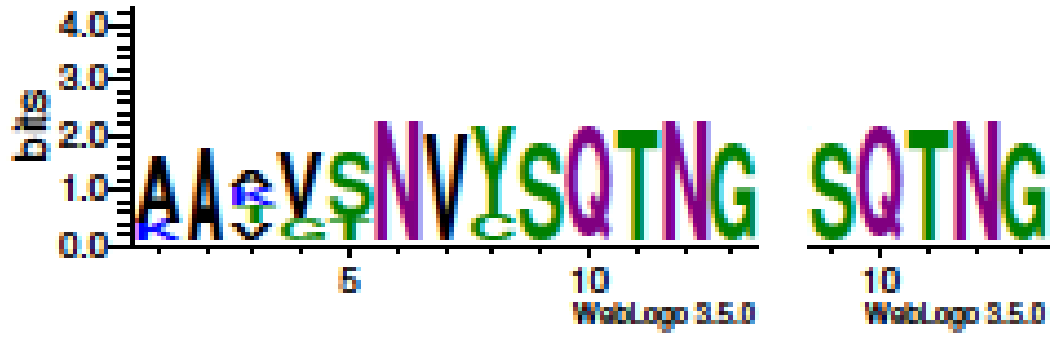
DA1

RYKVYS
KYKVYN
RYKVCN
RYKVCN



DA2

KAVGTNVYSQTNG
AAKVSNVCSQTNG
A-AVSNVYSQTNG
AATVSNVYSQTNG



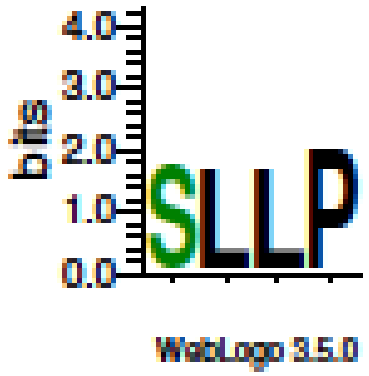
fingerprint

DA3

SLLP

SLLP

SLLP

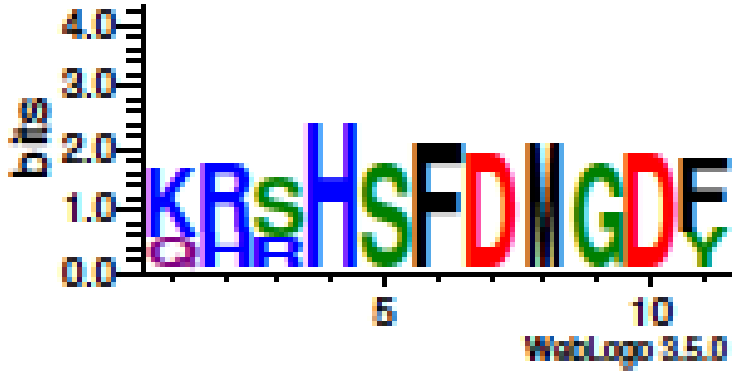


DA4

QHRHSFD-GDY

KRSHSFDMGDF

KRSHSFDMGDF

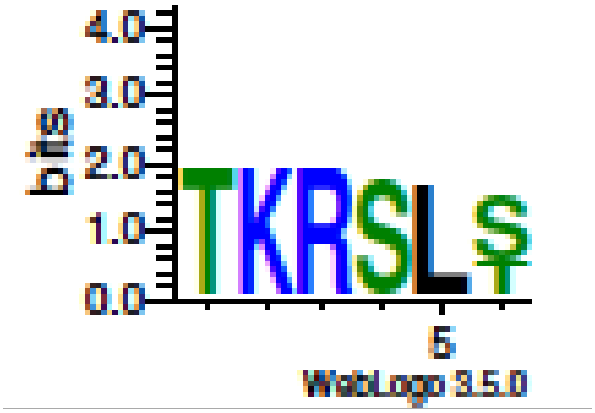


DA5

TKRSLT

TKRSLS

TKRSLS

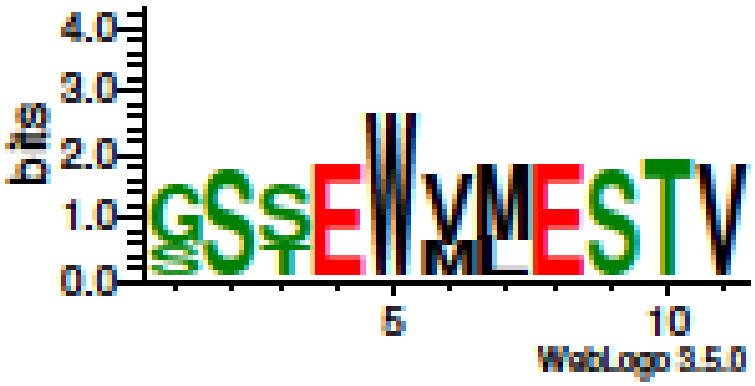


DA6

SSTEWMLESTV

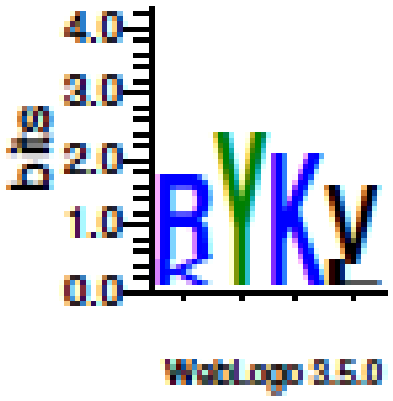
GSSEWVMESTV

GSSEWVMESTV



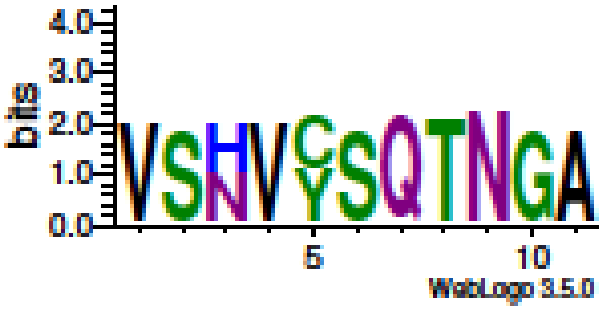
DA7

RYKV
RYKV
RYKV
KYKL



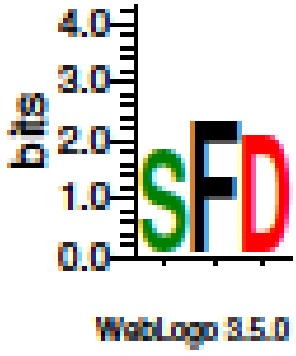
DA8

VSHVCSQTNGA
VSHVCSQTNGA
VSNVYSQTNGA
VSNVYSQTNGA



DA9

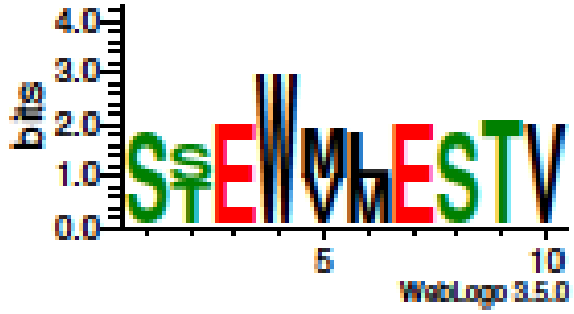
SFD
SFD
SFD
SFD



DA10

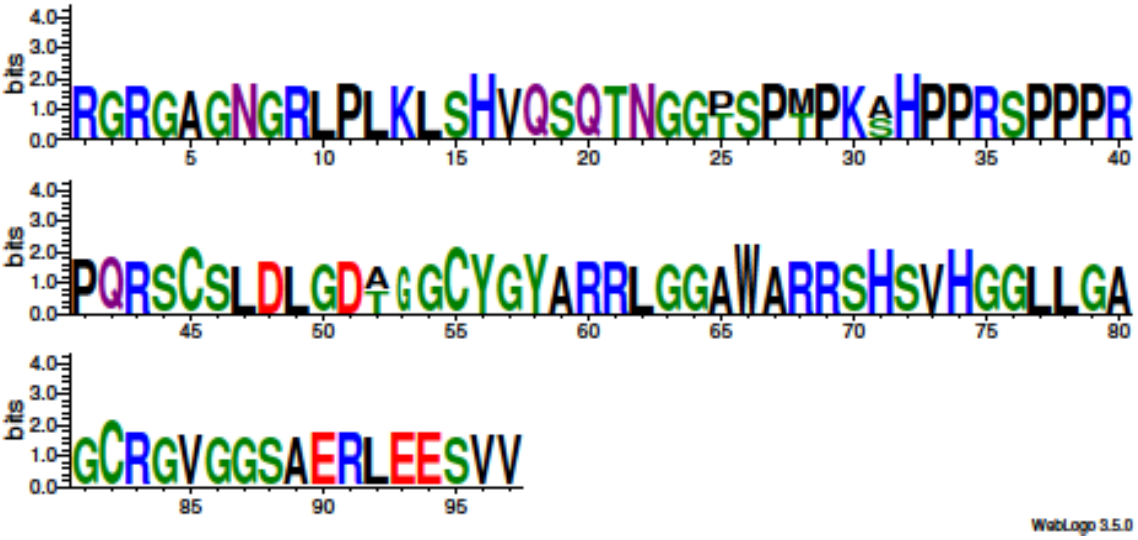
STEWMLESTV
STEWMLESTV
SSEWVMESTV

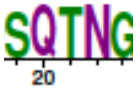
SSEWVMESTV

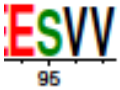


DA11

RGRGAGNGRLPLKLSHVQSQTNGGPSPTPKAHPPRSPPPRPQRSCSLDLGDA-
GCYGYARRLGGAWARRSHSVHGGLLGAGCRGVGGSAERLEESVV
RGRGAGNGRLPLKLSHVQSQTNGGTSPMPKSHPPRSPPPRPQRSCSLDLGDTGGCYGYARRLGGAWARRSHSVHGGLLGAGCRGVGG
SAERLEESVV

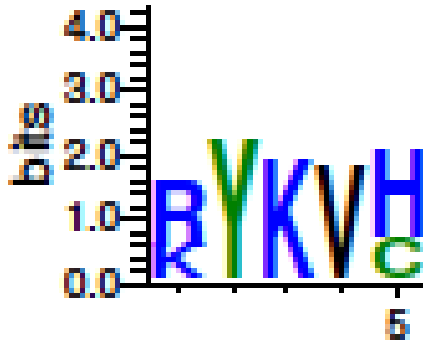


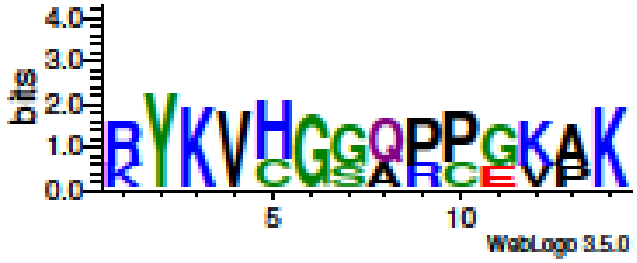




DA12

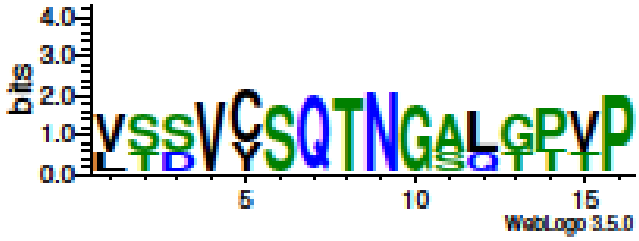
RYKVHGGQPPGKAK
RYKVHGGQPPGKAK
KYKVCGSARCEVPK





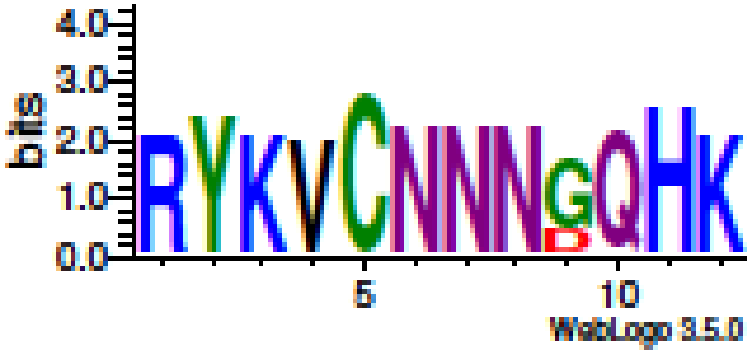
DA13

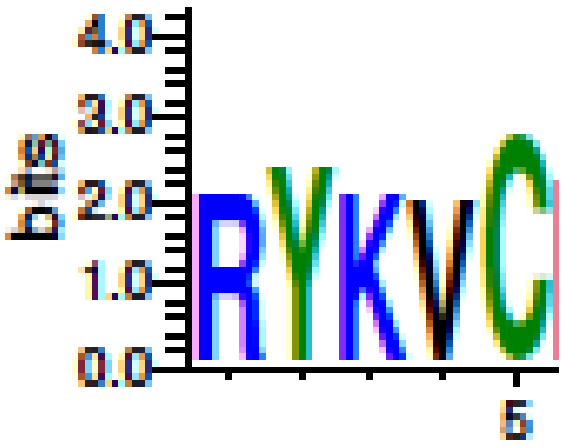
VSSVCSQTNGALGPTP
VSSVCSQTNGALGPVP
LTDVYSQTNGSQTTVP



DA14

RYKVCNNNGQHK
RYKVCNNNGQHK
RYKVCNNNGQHK
RYKVCNNNDQHK





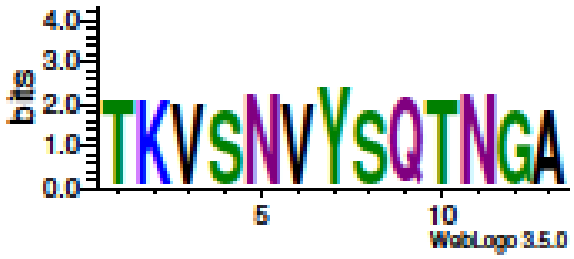
DA15

TKVSNVYSQTNGA

TKVSNVYSQTNGA

TKVSNVYSQTNGA

TKVSNVYSQTNGA



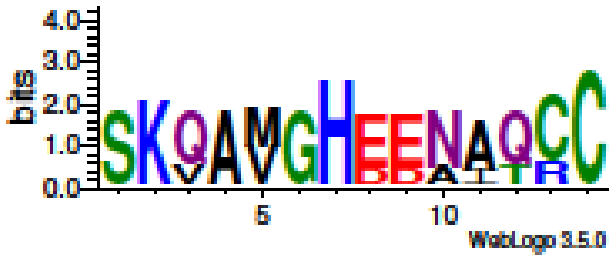
DA16

SKQAVGHEENAQCC

SKQAMGHEENAQCC

SKQAVGHEEAAQCC

SKVAMGHDDNITRC



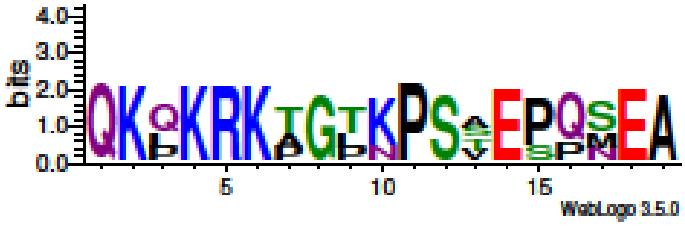
DA17

QKQKRKTGTKPSTEPQNEA

QKQKRKTGTKPSAEPQSEA

QKPKRKPGPKPSSEPQSEA

QKLKRKAGLNPSVESPMEA



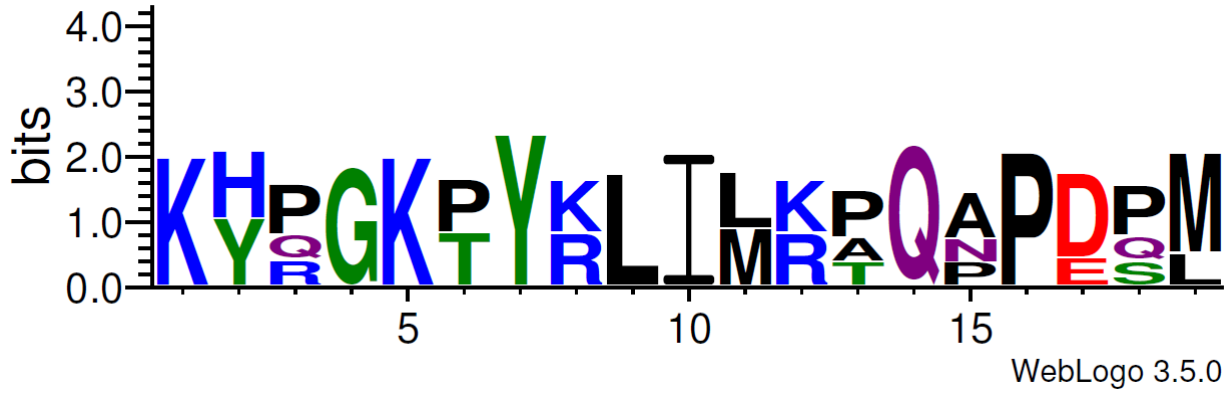
DA18

KHPGKPYRLILRPQAPDPM

KHPGKPYRLILRPQAPDPM

KYQGKTYKLIMKAQNPDQM

KYRGKTYKLIMKTQPPESL



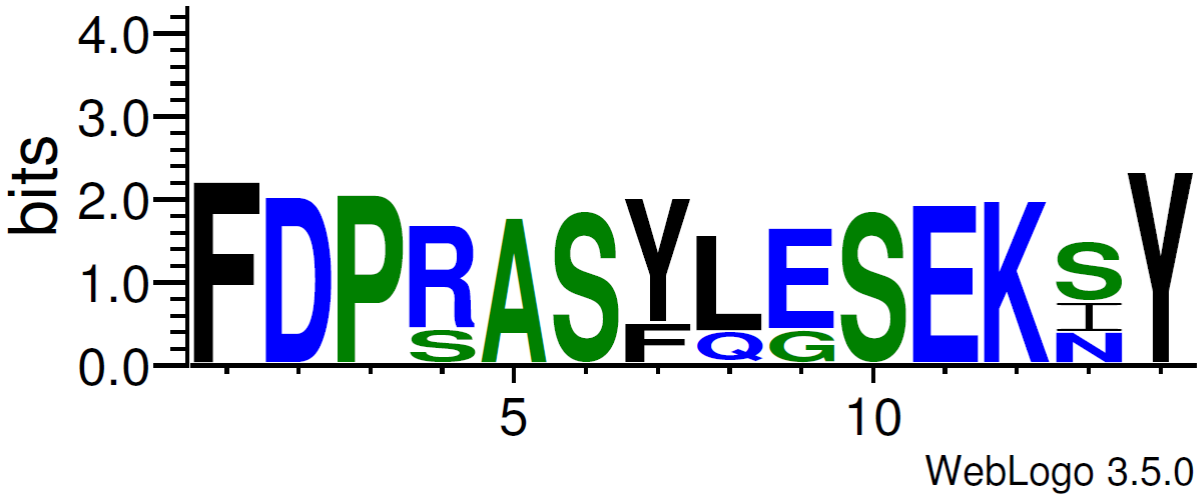
DA19

FDPRASYLESEKSY

FDPRASYLESEKSY

FDPRASYLESEKNY

FDPSASFQGSEKIY



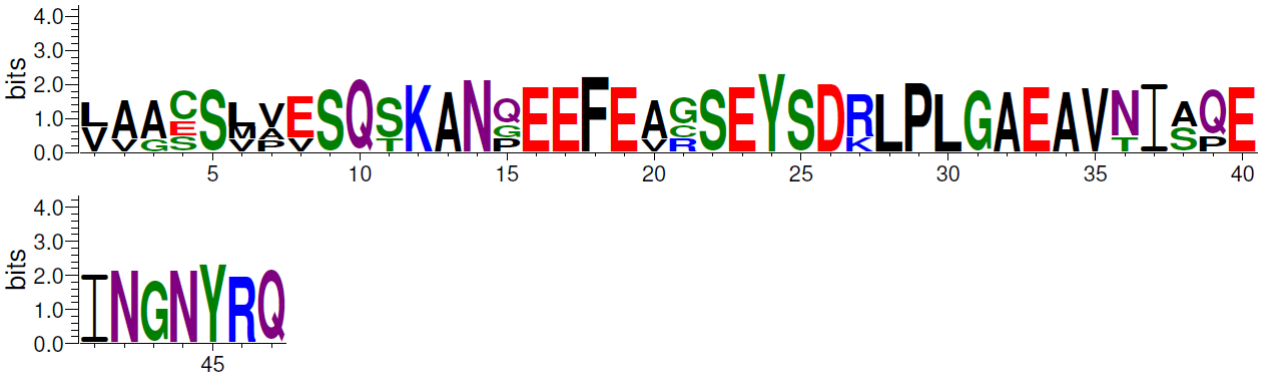
DA20

LAACSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEINGNYRQ

LAGCSLVESQSKANQEEFEAGSEYSDRLPLGAEAVNIAQEINGNYRQ

VAASSMAESQSKANGEEFEVRSEYSDKLPLGAEAVTISQEINGNYRQ

VVAESVPVSQTKANPEEFEACSEYSDRLPLGAEAVNISPEINGNYRQ


bits
4.0
3.0
2.0
1.0
0.0
5
10
15
20
25
30
35
40
bits
4.0
3.0
2.0
1.0
0.0
45
WebLogo 3.5.0