Chapter 12

Bioinformatics as a Tool to Identify Infectious Disease Pathogen Peptide Sequences as Targets for Antibody Engineering

Lavanya Suneetha, Prasanna Marsakatla,
Rachel Supriya Suneetha and Sujai Suneetha

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71011

Abstract

Bioinformatics is an interdisciplinary field of information technology for understanding biological data from genome to protein. It includes a combination of fields of science, computer science, statistics, mathematics, and engineering to analyze, interpret and derive biological data. This chapter describes how to use Bioinformatics to identify pathogen virulence factor peptide sequence similarities in human nerve tissue proteins and for evaluation as antibody engineering target peptides.

Keywords: bioinformatics, infectious diseases, peptide, recombinant antibody

1. Introduction

Bioinformatics is the application of techniques derived from disciplines such as applied mathematics, computer science, and statistics to analyze and interpret biological data. In this chapter, you will learn how to use bioinformatic techniques to identify pathogen virulence factor (VF) peptide sequence similarities to human nerve tissue proteins and then how to identify target peptides that could form the basis for engineering recombinant antibodies. Also, wet experiments could be conducted on the identified overlapping sequences to help us to single out target antibodies to be tested for tissue culture studies [1, 2]. The most ideal targeted peptide sequences for antibody engineering are those physiologically relevant, easy to access, and comprise amino acid sequence regions which have high specificity in pathogenic steps and reduced amino acid string length.

1.1. Bioinformatics and its role in peptide discovery

The accessibility to the extensive genomic and proteomic databases and the availability of tools to compare and evaluate the information have given rise to a new interdisciplinary field that combines biology and computer science [3]. Bioinformatics conceptualizes physical and chemical biology in terms of macromolecules and then applies "informatics" techniques (derived from disciplines such as applied mathematics, computer science, and statistics) to assimilate and organize the information associated with these molecules, on a large scale [4]. Bioinformatics is an exciting and exploratory method for peptide discovery in antibody engineering and development of antimicrobial therapies and vaccination strategies [5].

There is significantly growing evidence that a number of neurodegenerative diseases are a result of the association of host cell proteins with viral and bacterial infectious agents [6]. When pathogenic micro organisms such as bacteria, viruses, parasites, or fungi cause an infectious disease, there are many molecular interactions between the host-pathogen proteins and host peptides [7] through all the stages of the disease whether incubation, prodromal illness, decline, and convalescence. There is much experimental evidence identifying the virulence factors (VF) of pathogen and host components such as receptors and tissue-specific proteins [8, 9]. Though the pathogenic pathway of the infectious agent in various host tissues is unknown, many of these processes are suspected to be attributable to the yet undiscovered role of molecular mimics identified in pathogenic microorganisms and its corresponding host tissue proteins. The sequence and structural similarities between the pathogenic VF protein and nerve peptides could impact either directly or indirectly the pathogenesis of the infectious disease [10–12]. It could contribute to molecular mimicry, steric hindrance, receptor binding, cell signaling, and autoantibody production events (involved in neuro degeneration) in the host.

Leprosy patients with peripheral nerve damage develop autoimmunity to myelin P0 (nerve protein). The above conclusion was drawn by gathering known scientific evidence that are as follows: (1) labeling and binding studies found that *Mycobacterium leprae* (bacterium causing leprosy) binds to myelin P0 [13]; (2) clinical studies confirmed the production of autoantibodies as a response of the bacterium to interact with myelin P0 [14, 15]; and (3) bioinformatics searches identified sequences and structural similarities between *M. leprae* and the immunoglobulin regions of myelin P0 [16].

Identification of molecular mimics in pathogen-host peptide sequences is one approach to identify target peptides for antibody engineering. There are about 180 extensive biological databases to retrieve information on sequence and functional aspects of biological molecules. The updated list is available in Nucleic Acids Research [17].

1.2. The use of bioinformatics in identifying sequence similarities

This section teaches you how to conduct a search for proteins present in a target host, how to obtain its amino acid sequence/s from the existing databases, how to compare the sequence/s of the host protein to that of the pathogen protein, and finally how to interpret the results based on existing evidential data. In our case study, we identify the virulence factor peptide

sequence similarities of a few selected infectious agents with human nerve tissue proteins for selecting peptides to engineer antipeptide antibodies which recognizes corresponding host/viral proteins.

1.2.1. Selection of nerve proteins

63 proteins were extracted from the Human Protein Atlas Database that were enriched and enhanced in the nervous tissue as observed by immunehistochemistry (**Figure 1**).

To conduct a search for human proteins in the nervous tissue, access the website (www.proteinatlas.org), enter the tissue of study (e.g. nervous tissue) into the search box provided and click on search.

Manual protein selection was carried out based on their tissue expression (enriched and enhanced) and also on immunohistochemistry evidence (**Figure 2**).

The list of selected proteins are as follows: agrin (AGRN_HUMAN, O00468), calbindin (CALB1_HUMAN, P05937), n-chimaerin (CHIN_HUMAN, P15882), secretogranin-2 (SCG2_HUMAN, P13521), neuromodulin (NEUM_HUMAN, P17677), kinesin (KIFC1_HUMAN, Q9BW19), tau (TAU_HUMAN, P10636), 2′,3′-cyclic-nucleotide 3′-phosphodiesterase (CN37_HUMAN, P09543), myelin-associated glycoprotein (MAG_HUMAN, P20916), myelin P0 (MYP0_HUMAN, P25189), myelin P2 (MYP2_HUMAN, P02689), oligodendrocyte-myelin

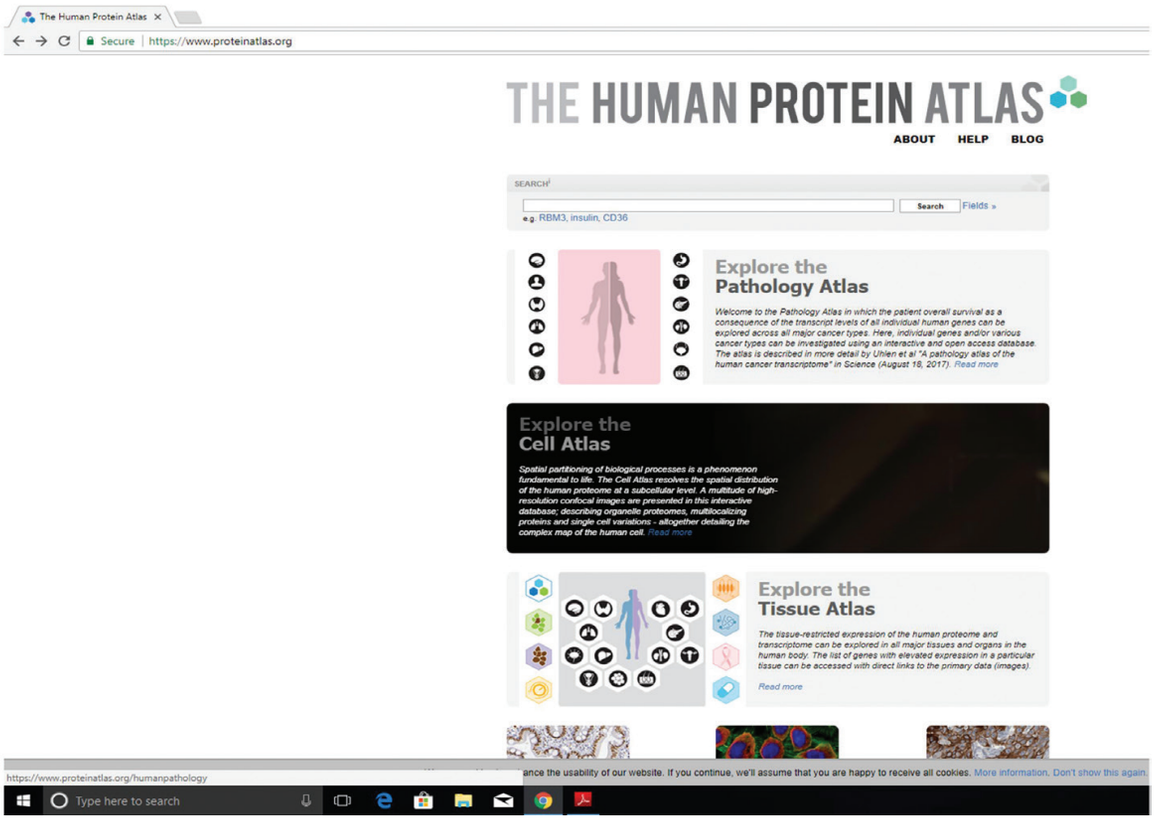


Figure 1. Conducting a search on the Human Protein Atlas Database.

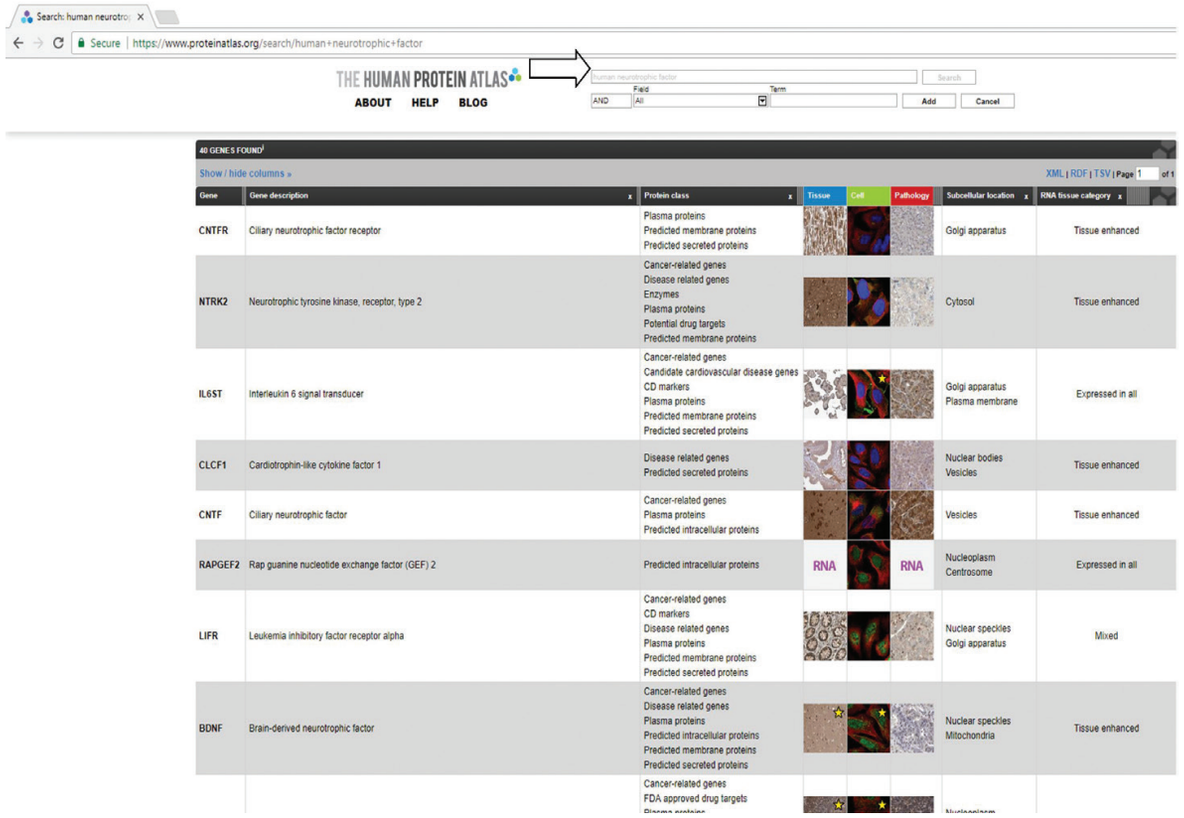


Figure 2. Conducting an advanced search on the Human Protein Atlas Database.

glycoprotein (OMGP_HUMAN, P23515), brain-derived neurotrophic factor (BDNF_HUMAN, P23560), ciliary neurotrophic factor (CNTF_HUMAN, P26441), neurotrophin-3 (NTF3_HUMAN, P20783), beta-nerve growth factor (NGF_HUMAN, P01138), nestin (NEST_HUMAN, P48681), neurofilament heavy polypeptide (NFH_HUMAN, P12036), neurogranin (NEUG_HUMAN, Q92686), voltage-dependent T-type calcium channel subunit alpha-1G (CAC1G_HUMAN, O43497), hippocalcin (HPCL1_HUMAN, P37235), neurocalcin-delta (NCALD_HUMAN, P61601), recoverin (RECO_HUMAN, P35243), bombesin receptor subtype-3 (BRS3_HUMAN, P32247), kininogen-1/bradykinin (KNG1_HUMAN, P01042), calcitonin (CALC_HUMAN, P01258), cholecystokinin (CCKN_HUMAN, P06307), galanin peptides (GALA_HUMAN, P22466), pro-neuropeptide Y (NPY_HUMAN, P01303), neurotensin/neuromedin N (NEUT_HUMAN, P30990), protein S100-B (S100B_HUMAN, P04271), synapsin-1 (SYN1_HUMAN, P17600), probable tubulin polyglutamylase (TTLL1_HUMAN, O95922), myelin basic protein (MBP_HUMAN, P02686), protein phosphatase 1 regulatory subunit 1B (PPR1B_HUMAN, Q9UD71), Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2 (AGAP2_HUMAN, Q99490), cathepsin L2 (CATL2_HUMAN, O60911), D(1A) dopamine receptor (DRD1_HUMAN, P21728), BDNF/NT-3 growth factors receptor (NTRK2_HUMAN, Q16620), melanoma-associated antigen E1 (MAGE1_HUMAN, Q9HCI5), microtubule-associated protein 6 (MAP6_HUMAN, Q96JE9), protocadherin alpha-12 (PCDAC_HUMAN, Q9UN75), carboxypeptidase E (CBPE_HUMAN, P16870), Down syndrome cell adhesion molecule (DSCAM_HUMAN, O60469), dyslexia-associated

protein KIAA0319 (K0319_HUMAN, Q5VV43), uncharacterized protein KIAA1211-like (K121L_HUMAN, Q6NV74), microtubule-associated protein 1B (MAP1B_HUMAN, P46821), neuronal calcium sensor 1 (NCS1_HUMAN, P62166), neurofilament light polypeptide (NFL_HUMAN, P07196), receptor expression-enhancing protein 2 (REEP2_HUMAN, Q9BRK0), secretogranin-3 (SCG3_HUMAN, Q8WXD2), ubiquitin carboxyl-terminal hydrolase isozyme L1 (UCHL_HUMAN, P09936), galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1 (B3GA1_HUMAN, Q9P2W7), beta-1,4 N-acetylgalactosaminyltransferase 1 (B4GN1_HUMAN, Q00973), caprin-2 (CAPR2_HUMAN, Q6IMN6), dopamine beta-hydroxylase (DOPO_HUMAN, P09172), FAM81A (FA81A_HUMAN, Q8TBF8), mitogen-activated protein kinase 10 (MK10_HUMAN, P53779), N-terminal EF-hand calcium-binding protein 1 (NECA1_HUMAN, Q8N987), neuroligin-3 (NLGN3_HUMAN, Q9NZ94), protein kinase C and casein kinase substrate in neurons protein 1 (PACN1_HUMAN, Q9BY11), sodium channel protein type 7 subunit alpha (SCN7A_HUMAN, Q01118), and clathrin coat assembly AP180 (AP180_HUMAN, O60641). The biological accepts of the proteins have been derived from the information presented in UniProt database for each protein [18–20].

1.2.2. Retrieving FASTA formats

FASTA formats for each of the above proteins were retrieved from NCBI PubMed. The FASTA format is a text-based format obtained from the PubMed search and represents either nucleotide sequences or peptide sequences (**Figure 3**).

Upon accessing the website, select the database in which the search is to be conducted (e.g. Protein). Type the name of the protein and its species in brackets into the search text box provided (e.g. Agrin (*Homo sapiens*)) and click on the search button.

The protein with the highest number of amino acids is chosen. Click on the hyperlinked protein to access its gene bank. Upon reaching the gene bank of the selected protein, click on the hyperlinked *FASTA* (**Figures 4**, **5** and **6**).

Obtain the FASTA format by copying all the information (Starting from the > symbol).

1.2.3. Arranging the FASTA formats

All the FASTA formats of the human proteins are saved in a sequence on Microsoft Notepad (**Figure 7**).

1.2.4. Running the BLAST

Pathogen-protein mimics, nerve protein sequences were BLAST (Basic Local Alignment Search Tool; Version 2.7.1; e-value ≤0.01) [21] against a pathogen genome (**Figure 8**).

Access the BLAST website at https://blast.ncbi.nlm.nih.gov/Blast.cgi and click on Protein (Protein~Protein) BLAST. The FASTA formats of 63 nerve proteins were copied and pasted from the notepad into the text box provided. Enter the species of the organism against which the blast has to be performed/the sequence comparison has to be carried out specifying its Tax ID (**Figure 8**).

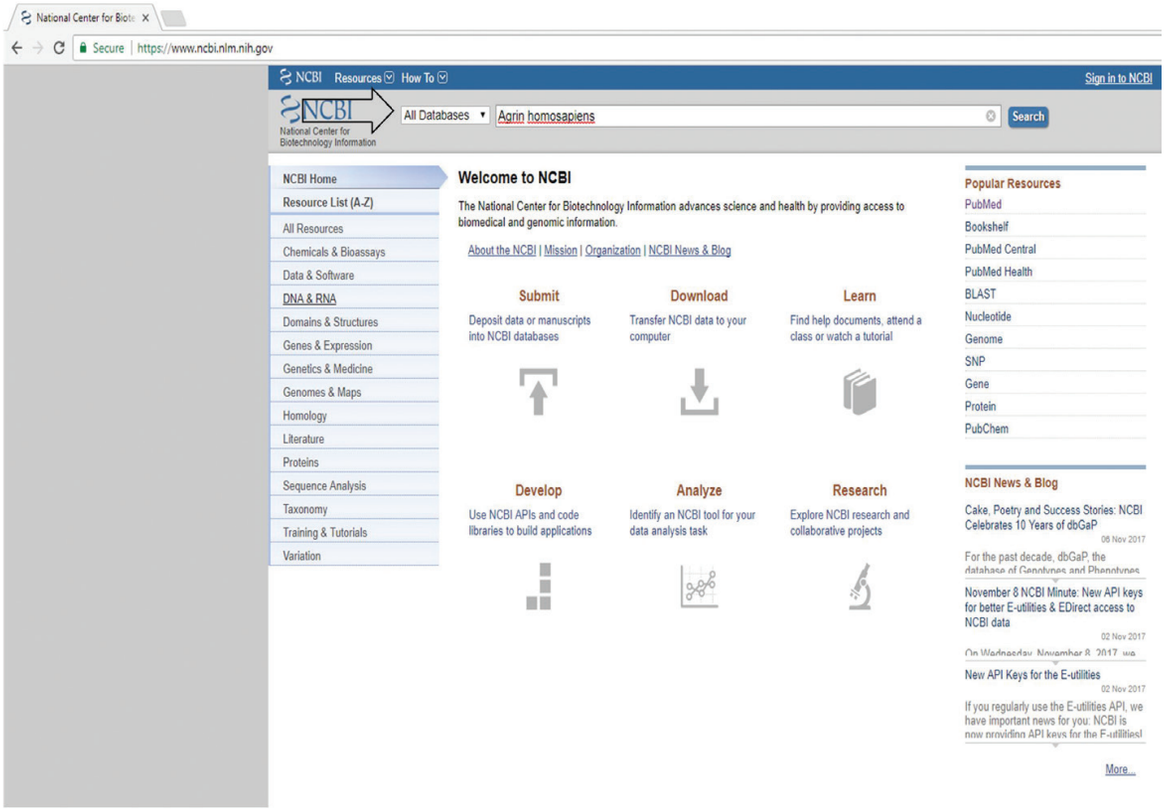


Figure 3. Conducting a search on the PubMed database.

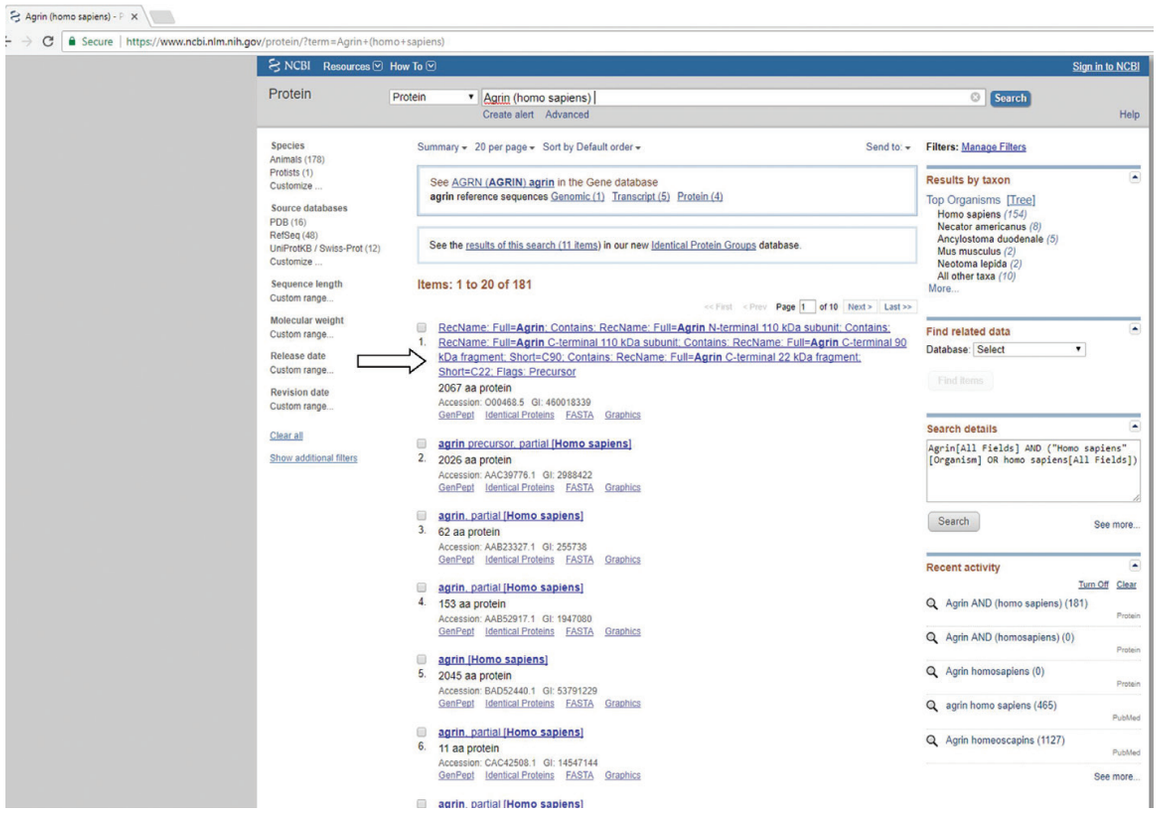


Figure 4. List of available sequenced protein information.

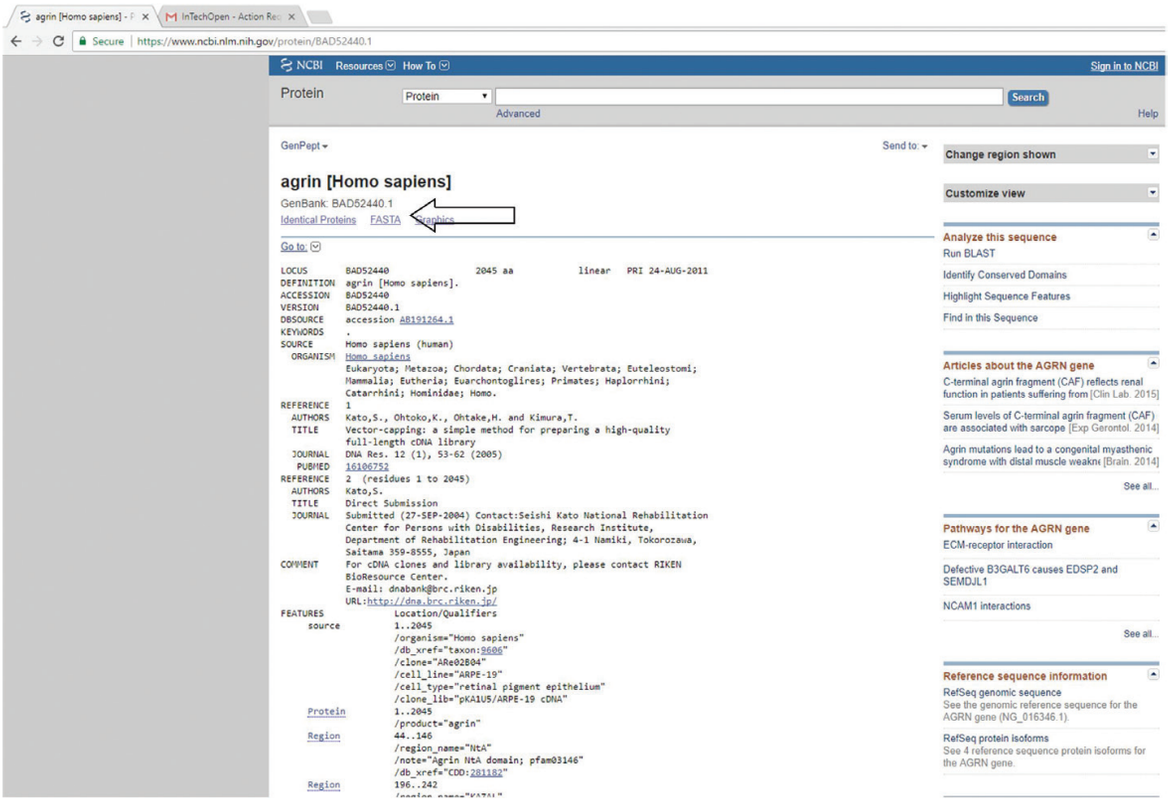


Figure 5. Gene information of agrin.

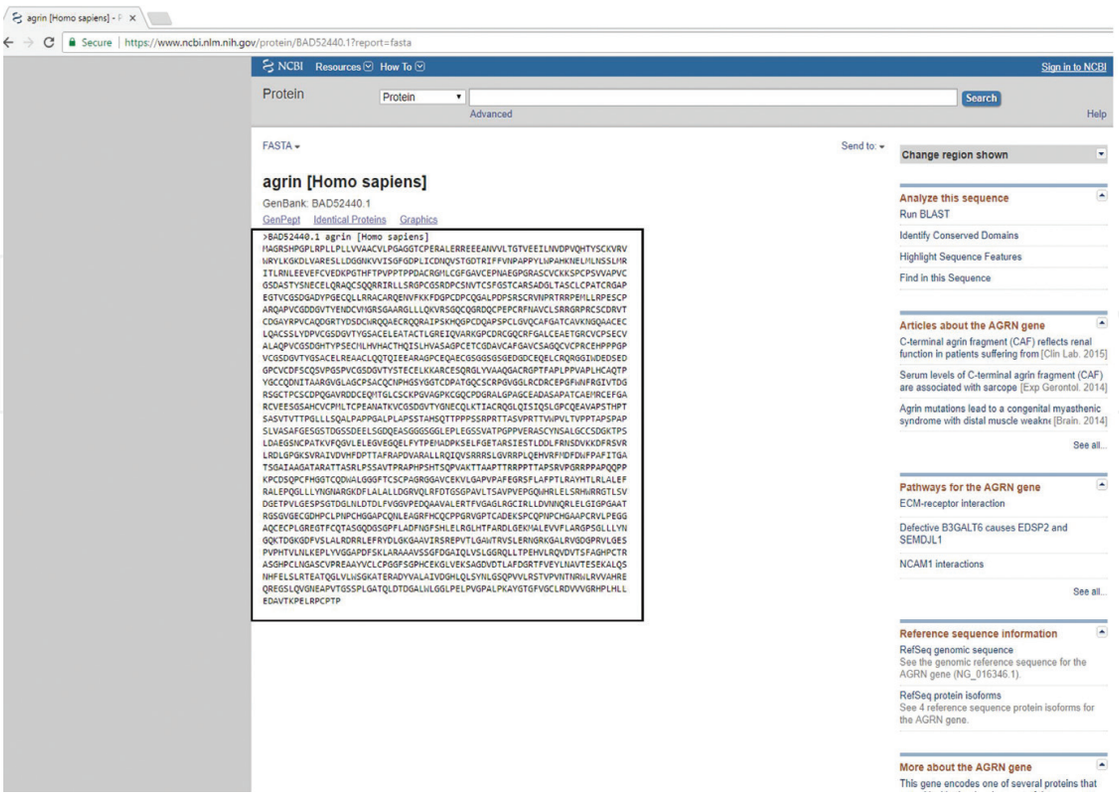


Figure 6. FASTA format of agrin.

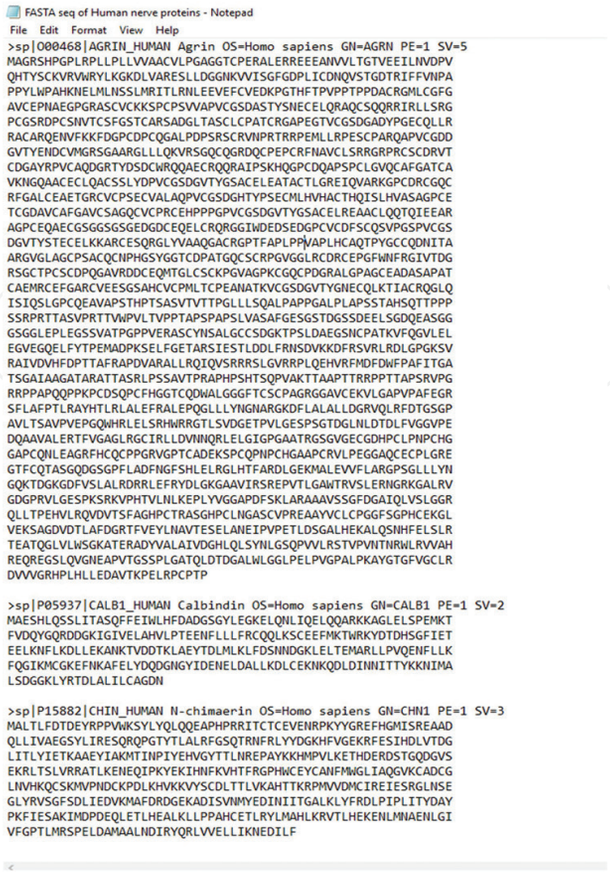


Figure 7. FASTA formats of the 63 proteins in sequence.

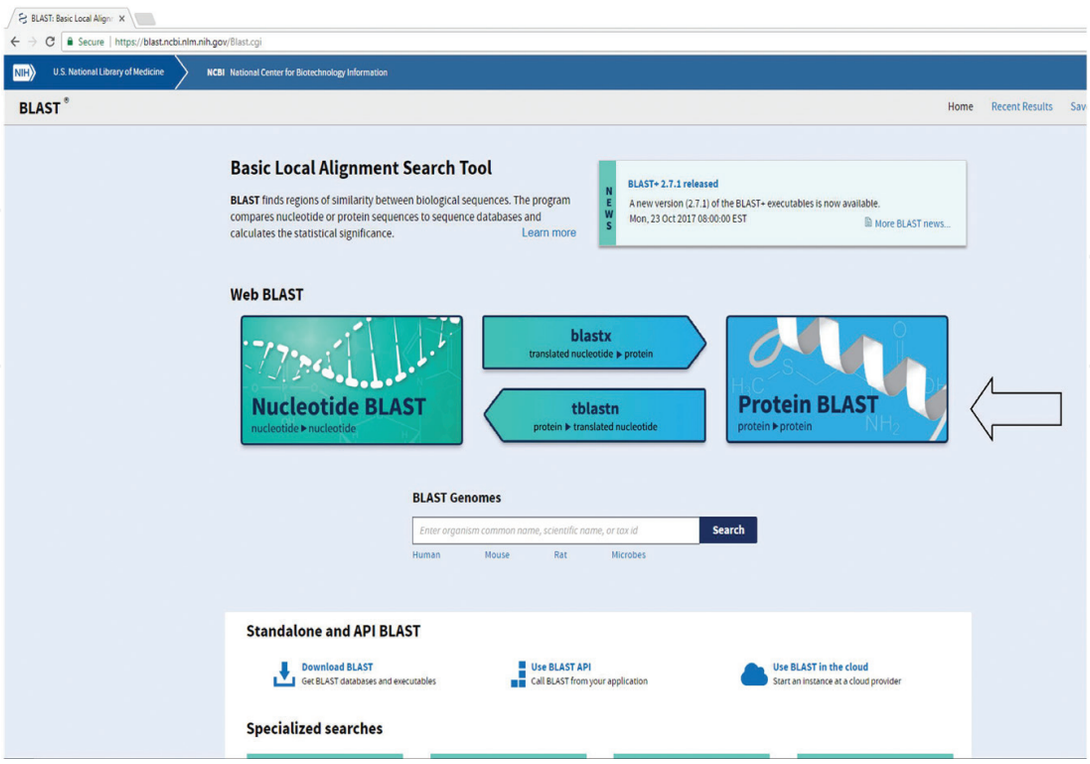


Figure 8. BLAST home page.

The pathogen genome sequences that were compared with the human nerve proteins are as follows: HIV (Tax ID: 11,676), Polio (Tax ID: 138,950), Japanese Encephalitis (Tax ID:64,320), *M. leprae* (Tax ID: 1769), Human herpes virus 1 (Tax ID: 10,298), Human herpes virus 2 (Tax ID: 10,310), Rabies virus (Tax ID: 11,292), Zika virus (Tax ID: 64,320), Corona virus (Tax ID: 11,118), Varicella zoster virus (Tax ID: 10,335).

Select program PSI BLAST as the BLAST algorithm for a more position-sensitive search. It looks deeper into the database to best match to your query. Click on the BLAST button and wait for the results. Take screen shots of your result and also download the provided excel format (**Figure 9**).

The output of the BLAST identified the significant peptide sequence similarities between the human protein and its pathogenic counterpart **Figure 10**. These peptide sequence similarities are identified by amino acid positions, in which amino acids exist in single-letter codes. The BLAST provides us with the number of sequence similarities between the pathogenic genomic sequence and its host proteins. It also identifies viral counterpart peptides and the region of similarity on the host proteins.

Further interpretations of the results can be made by referring to the Uniprot database to obtain the biological and functional aspects of the host and the pathogen proteins (**Figures 11** and **12**).

1.2.5. Ascribing a biological role and application

The results show a number of sequence similarities existing between host proteins and various pathogen proteins. The maximum number of peptide sequence similarities were found between host protein caprin-2 which had 495 similarities with polio; neurogranin had 230

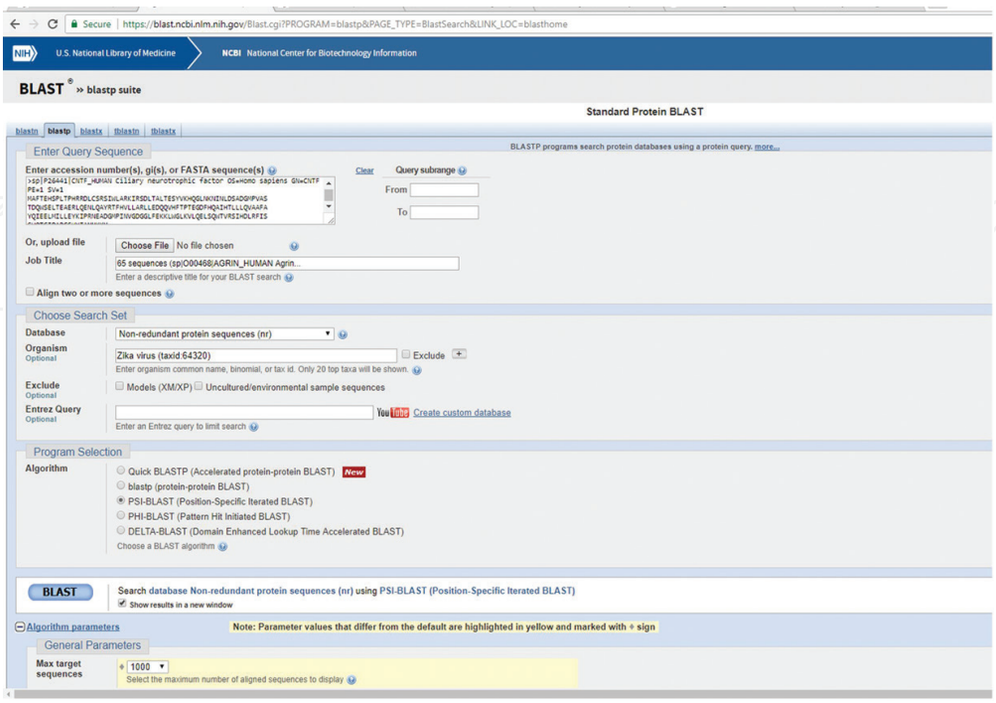


Figure 9. BLAST search.

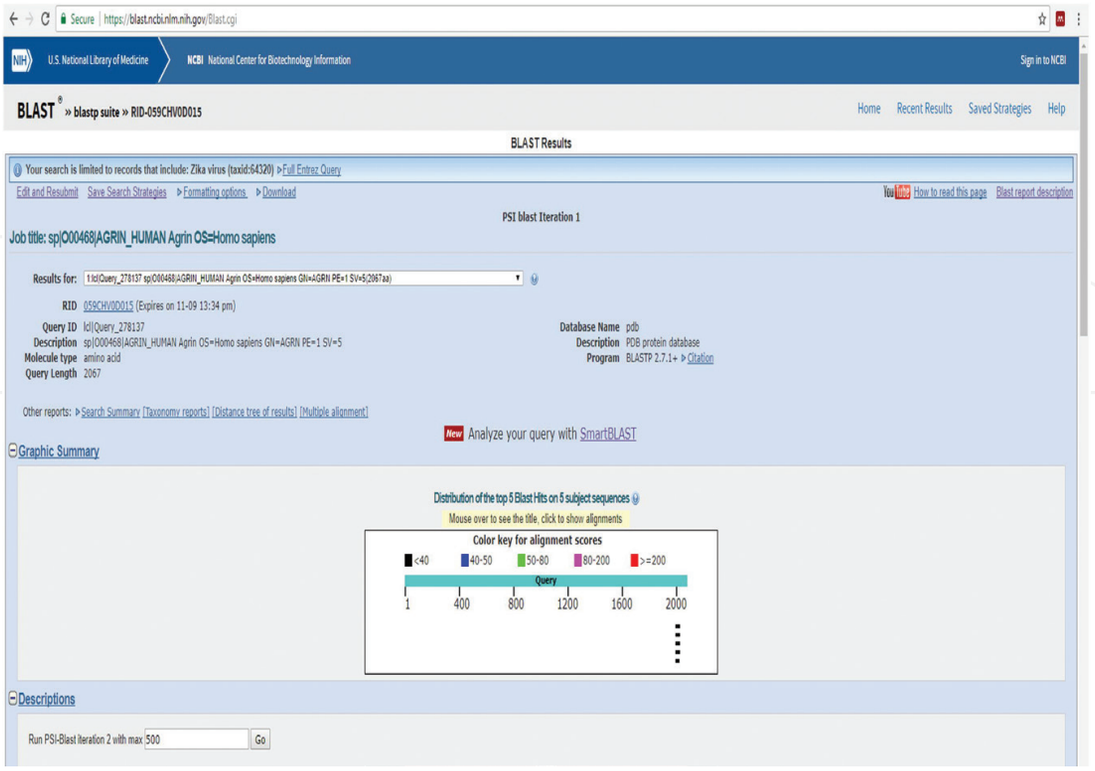


Figure 10. BLAST results of nerve proteins showing similarity to pathogen proteins.

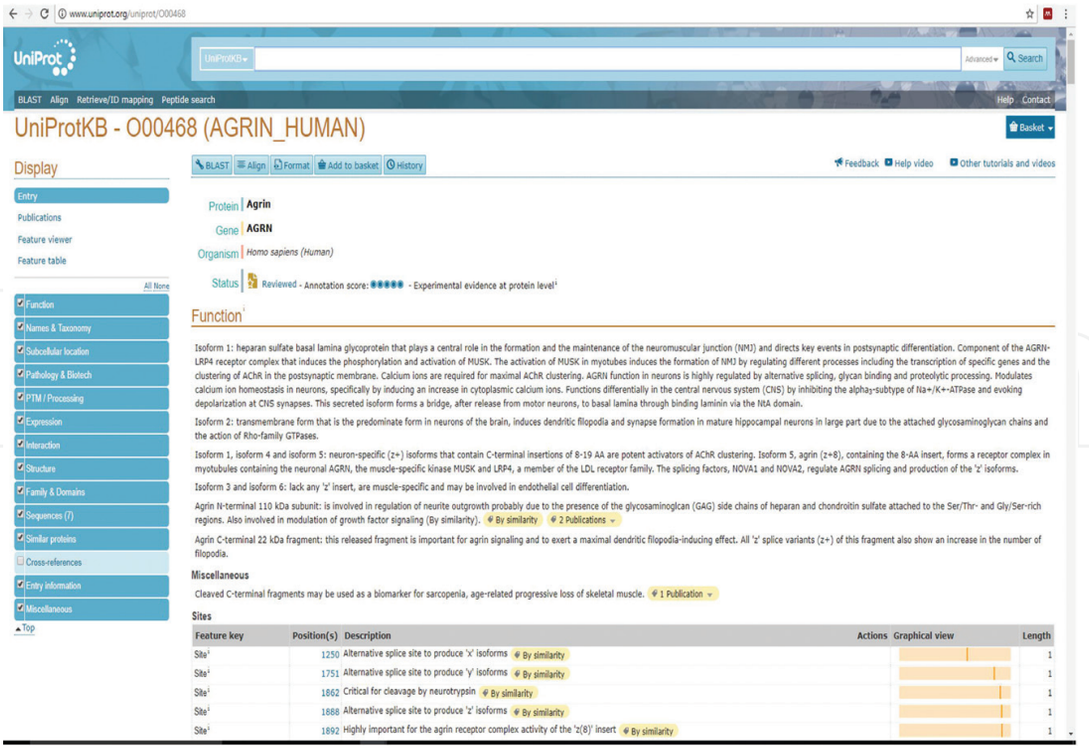


Figure 11. UniProtKB screenshot showing the biological and functional data of the human protein.

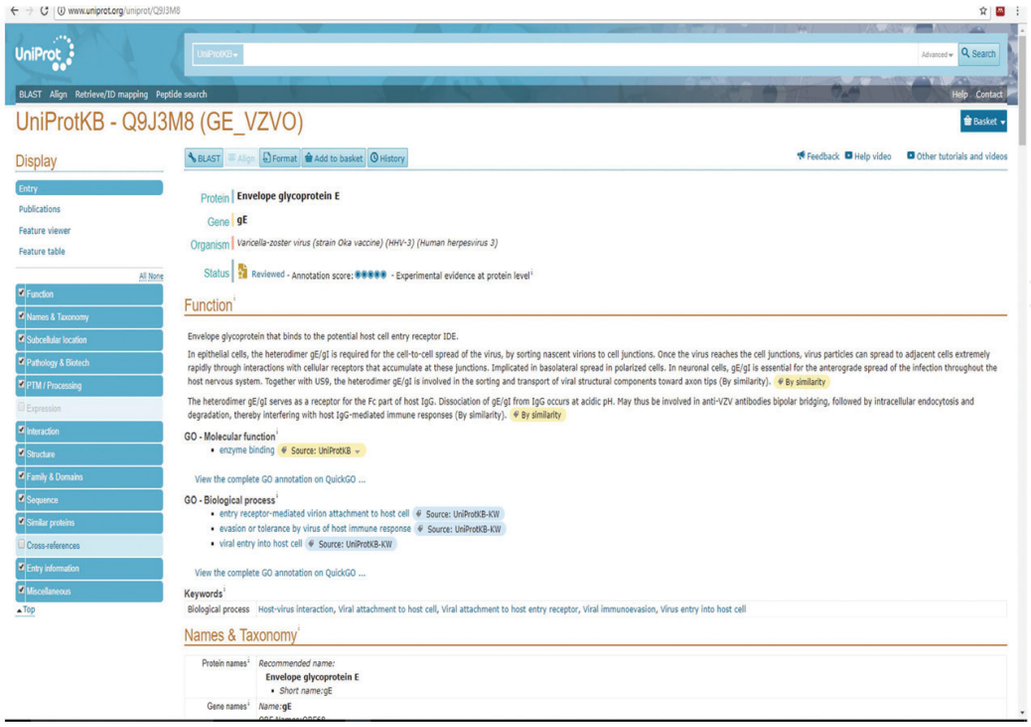


Figure 12. Uniprot screenshot showing the biological and functional data of the viral protein.

similarities with HHV2; secretogranin-3 had 221 similarities with Japanese encephalitis; agrin had 212 similarities with varicella; caprin-2 had 198 similarities with rabies virus; galanin peptides had 87 similarities with Zika virus; kinesin had 54 similarities with HIV; neurofilament heavy polypeptide had 46 similarities with corona virus; neurogranin had 39 similarities with HHV1; and 2′,3′-cyclic-nucleotide 3′-phosphodiesterase had 21 similarities with *M. leprae*.

This method identifies significant virulent factors which have sequence similarities to human nerve tissue proteins. The nerve proteins that exhibited sequence similarities with four or more pathogenic virulent factors are displayed in **Table 1**. All 63 proteins are found to have sequence similarities with *M. leprae* proteins.

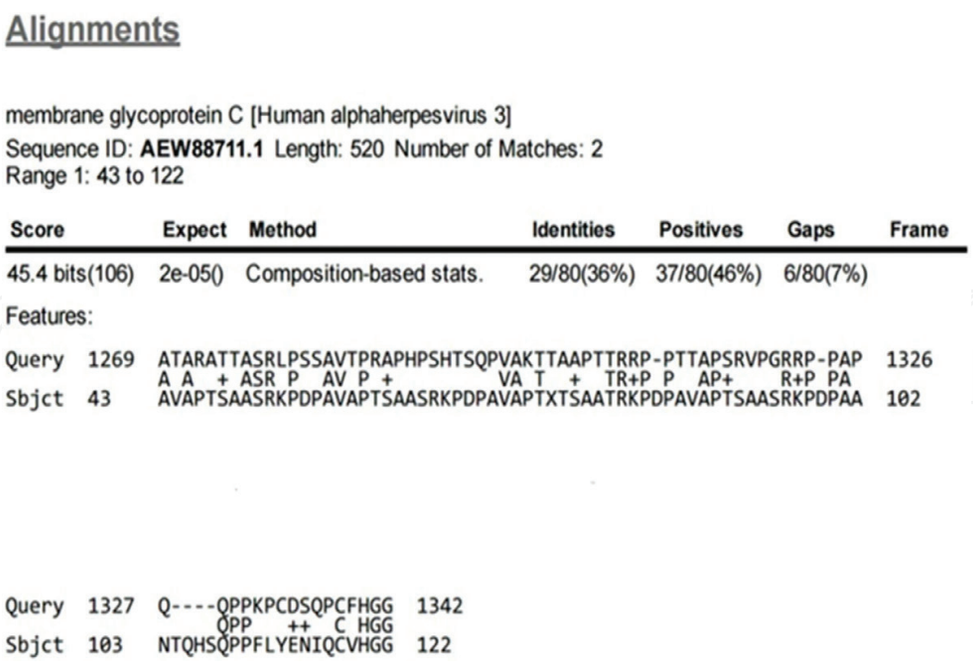
Agrin is a heparin sulphate basal lamina glycoprotein with a molecular mass of 217,232 Da. It plays a central role in the formation and maintenance of the neuromuscular junction. It is known to direct events in postsynaptic differentiation. Agrin also induces the phosphorylation and activation of muscle-specific kinase (MUSK), the clustering of Acetyl choline esterase receptor (AChR) in the postsynaptic membrane, regulates calcium ion homeostasis in neurons, and is involved in regulation of neuritis outgrowth [22, 23].

1.2.6. HHV3 peptide similarity to human protein agrin

Agrin UniProtKB-O00468 (AGRIN_HUMAN) (AA position 1269–1326) (**Figure 13**) has a similarity to membrane glycoprotein C (Sequence ID: AEW88711.1 AA Position 43–122) of the varicella zoster virus UniProtKB-Q9J3M8 (GE_VZVO) which by its similarity has the

S. No	Query No.	Proteins	HIV	Polio	JE	HHV 1	HHV 2	*M. leprae*	Corona	Zika	Rabies	Vericella
1	O00468	Agrin	0	0	4	0	0	6	1	0	1	212
2	P17677	Neuromodulin	1	1	0	1	3	6	28	1	0	75
3	Q9BW19	Kinesin	54	1	0	0	0	9	0	0	0	0
4	P10636	Tau protein	0	1	0	0	14	5	9	0	0	19
5	P25189	Myelin protein P0	2	0	0	1	22	7	1	0	0	0
6	P23515	Oligodendrocyte-myelin glycoprotein	0	2	0	1	0	9	0	0	1	23
7	P48681	Nestin	0	0	3	2	0	8	2	30	0	22
8	P04271	Protein S100-B	0	26	0	2	11	7	0	0	0	12
9	P17600	Synapsin-1	0	1	7	11	2	13	0	0	5	0
10	P02686	Myelin basic protein	0	0	0	0	2	9	4	3	0	5
11	Q16620	BDNF/NT-3 growth factors receptor	0	0	0	23	8	11	0	0	1	15
12	Q5VV43	Dyslexia-associated protein KIAA0319	0	0	0	37	21	5	5	0	2	1
13	P07196	Neurofilament light polypeptide	0	0	0	1	1	2	4	0	0	77
14	Q8WXD2	Secretogranin-3	3	5	221	0	10	8	0	0	9	0
15	Q00973	Beta-1,4 N-acetylgalactosaminyltransferase 1	1	29	0	1	2	8	0	0	0	0

Table 1. Sequence similarities of human nerve tissue proteins with human virulent factors. Multiple alignments obtained in a single BLAST search could result in identities of the amino acids or substitutions of the amino acids in the same peptide region.



Alignments

membrane glycoprotein C [Human alphaherpesvirus 3]
Sequence ID: **AEW88711.1** Length: 520 Number of Matches: 2
Range 1: 43 to 122

Score	Expect	Method	Identities	Positives	Gaps	Frame
45.4 bits(106)	2e-05()	Composition-based stats.	29/80(36%)	37/80(46%)	6/80(7%)	

Features:

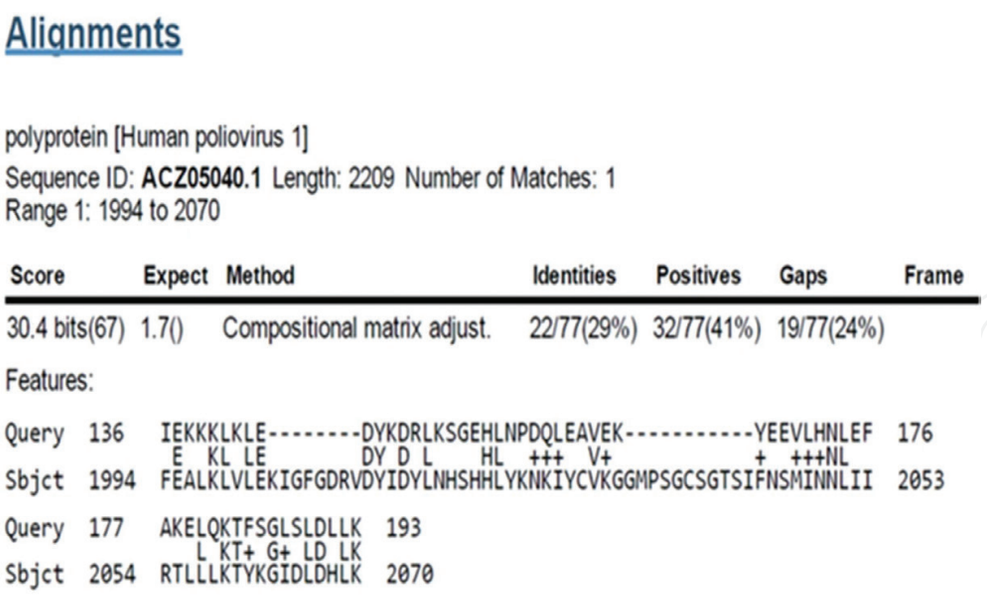
```
Query  1269  ATARATTASRLPSSAVTPRAPHPSHTSQPVAKTTAAPTTRRP-PTTAPSRVPGRRP-PAP  1326
             A A  + ASR P  AV P +          VA T  +  TR+P P  AP+    R+P PA
Sbjct  43    AVAPTSAASRKPDPAVAPTSAASRKPDPAVAPTXTSAATRKPDPAVAPTSAASRKPDPAA  102

Query  1327  Q----QPPKPCDSQPCFHGG  1342
              QPP   ++  C HGG
Sbjct  103   NTQHSQPPFLYENIQCVHGG  122
```

Figure 13. BLAST output of membrane glycoprotein of HHV3 showing similarity to human protein agrin.

potential to bind to the tissue cell receptor. Experimental evidence in epithelial cells shows that the hetero demonization of viral receptors could spread the virus by sorting nascent virion to nerve tissue cell junctions. The virus particles can spread to adjacent cells through interactions with cellular receptors at these cell junctions. The virus at cell junctions spreads extremely rapidly into the tissues [24, 25]. Sequence mimics of agrin to the varicella membrane glycoprotein could have an effect on either viral entry into host cell, evasion or on tolerance of host immune response to the virus and virion attachment to the host cell. These similarities in peptide regions warrant further exploration to understand pathogenesis and to identify target peptides for antibody engineering [26].

1.2.7. Poliovirus and rabies virus peptide similarities to human protein caprin-2

Caprin-2UniProtKB-Q6IMN6 (CAPR2_HUMAN) is a protein of molecular mass 68,429 Da. The structure of caprin-2 was found to be similar to the polio and rabies viruses. Caprin-2 (AA position: 136–176) has a similarity to the polyprotein of polio virus UniProtKB–E0WCG5 (E0WCG5_9ENTO) (polyprotein sequence ID: ACZ05040.1 AA position: 1994–2070) (**Figures 14** and **15**). Caprin-2 (AA position: 13–54) also has a similarity to the phosphoprotein of rabies virus UniProtKB-Q80JL8 (Q80JL8_9RHAB) (phosphoprotein sequence ID: AAO60615.1 AA position 76–110) (**Figure 15**). Caprin-2 has a significant role in influencing phosphorylation of the Wnt-signaling pathways (PubMed:18,762,581) [27]. Caprin-2 also facilitates LRP6 phosphorylation by CDK14/CCNY during G2/M stage of the cell cycle, which may potentiate cells for transport or translation of mRNAs, modulate the expression of neuronal proteins involved in synaptic plasticity [28], while simultaneously influencing cell cycle signaling and regulation of viral transcription and replication [29, 30].



Alignments

polyprotein [Human poliovirus 1]
Sequence ID: **ACZ05040.1** Length: 2209 Number of Matches: 1
Range 1: 1994 to 2070

Score	Expect	Method	Identities	Positives	Gaps	Frame
30.4 bits(67)	1.7()	Compositional matrix adjust.	22/77(29%)	32/77(41%)	19/77(24%)	

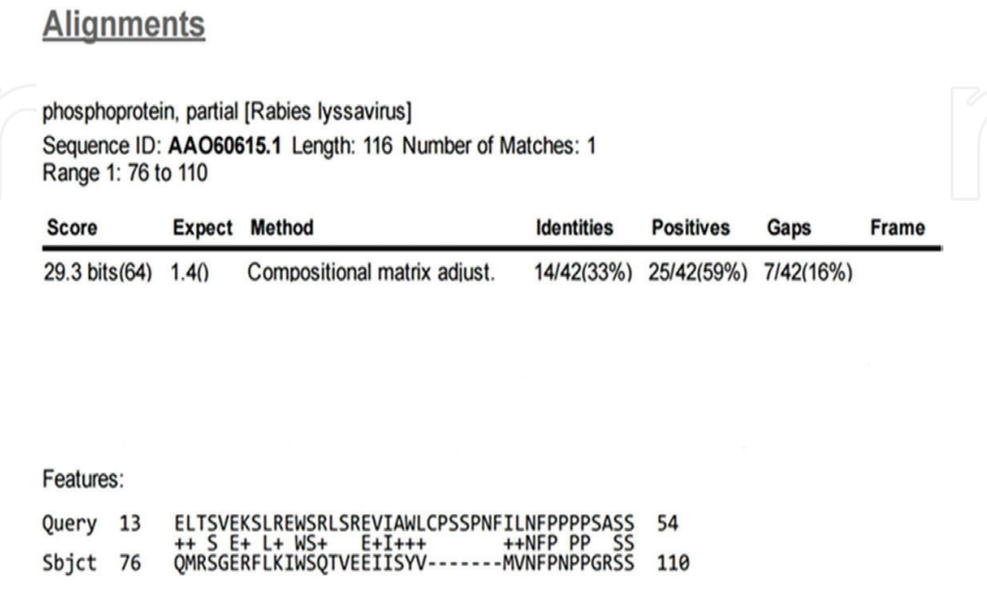
Features:

```
Query  136   IEKKKLKLE--------DYKDRLKSGEHLNPDQLEAVEK-----------YEEVLHNLEF  176
              E  KL LE        DY D L    HL  +++  V+           +  +++NL
Sbjct  1994  FEALKLVLEKIGFGDRVDYIDYLNHSHHLYKNKIYCVKGGMPSGCSGTSIFNSMINNLII  2053

Query  177   AKELQKTFSGLSLDLLK  193
               L KT+ G+ LD LK
Sbjct  2054  RTLLLKTYKGIDLDHLK  2070
```

Figure 14. BLAST output of polyprotein of poliovirus showing similarity to human protein caprin-2.

1.2.8. Mycobacterium leprae peptide similarity to 2′, 3′-cyclic-nucleotide 3′-phosphodiesterase

2′, 3′-cyclic-nucleotide 3′-phosphodiesterase UniProtKB-P09543 (CN37_HUMAN) is a protein of molecular mass 47,579 Da. 2′, 3′-cyclic-nucleotide 3′-phosphodiesterase (sequence ID: WP_010908292.1 AA position 191–261) has a similarity to thiamin pyrophosphokinase of *M. leprae* UniProtKB A0A197SEI9 (A0A197SEI9_MYCLR) (AA position: 170–2166) (**Figure 16**) 2′, 3′-cyclic-nucleotide 3′-phosphodiesterase is involved in RNA metabolism of the myelinating cell, CN37 (2′, 3′-cyclic-nucleotide 3′-phosphodiesterase) is the one of the most abundant myelin protein in nervous system. The sequence similarities identified could impact cell signaling and also regulate energy metabolism [31].



Alignments

phosphoprotein, partial [Rabies lyssavirus]
Sequence ID: **AAO60615.1** Length: 116 Number of Matches: 1
Range 1: 76 to 110

Score	Expect	Method	Identities	Positives	Gaps	Frame
29.3 bits(64)	1.4()	Compositional matrix adjust.	14/42(33%)	25/42(59%)	7/42(16%)	

Features:

```
Query  13   ELTSVEKSLREWSRLSREVIAWLCPSSPNFILNFPPPPSASS  54
            ++ S E+ L+ WS+   E+I+++       ++NFP PP  SS
Sbjct  76   QMRSGERFLKIWSQTVEEIISYV-------MVNFPNPPGRSS  110
```

Figure 15. BLAST output of phosphoprotein of rabies virus showing similarity to human protein caprin-2.

http://dx.doi.org/10.5772/intechopen.71011

thiamin pyrophosphokinase [Mycobacterium leprae]

Sequence ID: **WP_010908292.1** Length: 393 Number of Matches: 1

See 5 more title(s)

Range 1: 191 to 261

Score	Expect	Method	Identities	Positives	Gaps	Frame
28.5 bits(62)	0.70()	Compositional matrix adjust.	24/75(32%)	33/75(44%)	17/75(22%)	

Features:

```
Query  170  SADDLKKLKPGLEKDFLPLYFGWFLTKKSSETLRKAG----------QVFLEEL--GNH  216
            +ADDLK LKP + K++ P+  G       ++ LRKAG          Q+  E L  G H
Sbjct  191  AADDLKSLKPFI-KEYQPVLVG---VSGGADVLRKAGYRPQLIVGDPEQISTEALRCGAH  246

Query  217  KAFKKELRQFVPGDE  231
            +     PG E
Sbjct  247  VVLPADADGHAPGLE  261
```

Figure 16. BLAST output of thiamin pyrophosphate of *Mycobacterium leprae* showing similarity to human protein 2′, 3′-cyclic-nucleotide 3′-phosphodiesterase.

1.2.9. Zika virus peptide similarity to human protein galanin

Galanin peptide UniProtKB-P22466 (GALA_HUMAN) is a protein of molecular mass 13,302 Da. Galanin (AA position 53–99 position) has a similarity to polyprotein envelope protein E of Zika virus UniProtKB-Q73880 (Q73880_9HIV1) sequence ID: ARB07952.1 (AA position: 729–765) (**Figure 17**). Galanin is involved in the smooth muscle contraction of the gastrointestinal and genitourinary tract, regulation of growth hormone release, modulation of insulin release, and might also be involved in the control of adrenal secretion [32]. The envelope protein E of the Zika virus is responsible for binding to host cell surface receptors and mediating fusion between viral and cellular membranes. It is synthesized in the endoplasmic

Alignments

polyprotein, partial [Zika virus]

Sequence ID: **ARB07952.1** Length: 1939 Number of Matches: 1

Range 1: 729 to 765

Score	Expect	Method	Identities	Positives	Gaps	Frame
24.3 bits(51)	2.2()	Compositional matrix adjust.	14/47(30%)	23/47(48%)	10/47(21%)	

Features:

```
Query  53   SFSDKNGLTSKRELRPEDDMKPGSFDRSIPENNIMRTIIEFLSFLHL  99
            SF  K+G     E+RP          R  PE+N++R+++   S  H+
Sbjct  729  SFRAKDGCWYGMEIRP----------RKEPESNLVRSVVTAGSTDHM  765
```

Figure 17. BLAST output of polyprotein of Zika virus showing similarity to human Galanin peptide.

reticulum with protein prM and forms a heterodimer. Galanin's similarity with the ZIKA polypeptide could subsequently affect neural regulation of muscle function and play a role in immune evasion pathogenesis and viral replication [33].

1.2.10. HIV 1 peptide similarity to human kinesin-like protein

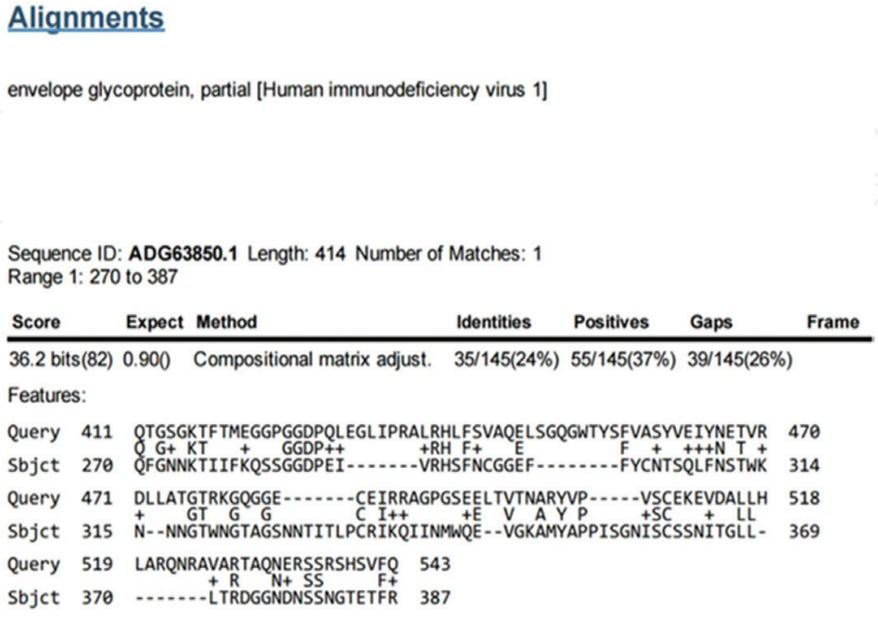
Kinesin-like protein KIFC1 UniProtKB-Q9BW19 (KIFC1_HUMAN) is a protein of molecular mass 73,748 Da. Kinesin-like protein (AA position: 411–470) has a similarity to HIV virus envelope glycoprotein UniProtKB-D6QPK9 (D6QPK9_9HIV1) sequence ID:ADG63850.1 (AA position:270–387)(**Figure 18**). KIFC1 along with microtubules contributes to movement of endocytic vesicles. These similarities could affect viral attachment to the host cell, membrane fusion, and entry into the cell and the nucleus [34, 35].

1.2.11. Corona virus peptide similarity to human neurofilament heavy polypeptide

Neurofilament heavy polypeptide UniProtKB-P12036 (NFH_HUMAN) is a protein of molecular mass 112,479 Da. Neurofilament heavy polypeptide (AA position: 819–872) has a similarity to ORF1a UniProtKB-A0A0F6SKM6 (A0A0F6SKM6_9GAMC) of Corona virus sequence ID: AKF17723.1 (AA positions: 890 –1031) (**Figure 19**) neurofilament of the nerve tissue usually contain three intermediate filament proteins: L, M, and H (NFH-human) which is involved in the maintenance of neuronal caliber. NFH-H has an important function in axon maturation. These similarities could affect viral replication, protein processing, and could generate autoantibody production [36, 37].

1.2.12. HHV 1 and HHV 2 peptide similarity to human protein neurogranin

Neurogranin UniProtKB-Q92686 (NEUG_HUMAN) is a protein of molecular mass 7618 Da. The structure of neurogranin at identical regions has a similarity to envelope glycoprotein M of



Alignments

envelope glycoprotein, partial [Human immunodeficiency virus 1]

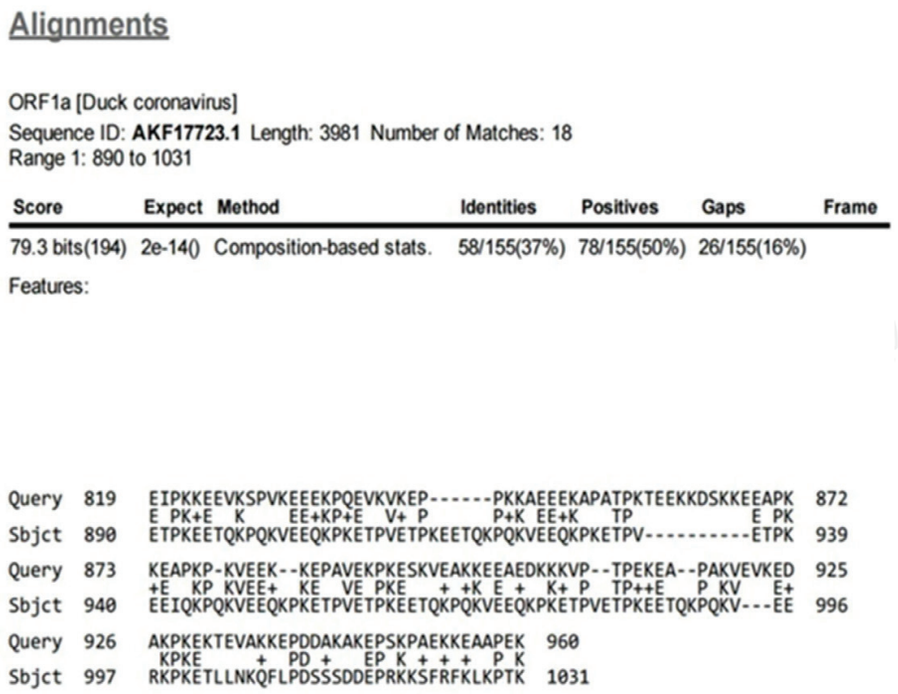
Sequence ID: **ADG63850.1** Length: 414 Number of Matches: 1
Range 1: 270 to 387

Score	Expect	Method	Identities	Positives	Gaps	Frame
36.2 bits(82)	0.90()	Compositional matrix adjust.	35/145(24%)	55/145(37%)	39/145(26%)	

Features:

```
Query  411  QTGSGKTFTMEGGPGGDPQLEGLIPRALRHLFSVAQELSGQGWTYSFVASYVEIYNETVR  470
            Q G+ KT    +   GGDP++        +RH F+    E          F   +  +++N T +
Sbjct  270  QFGNNKTIIFKQSSGGDPEI-------VRHSFNCGGEF--------FYCNTSQLFNSTWK  314

Query  471  DLLATGTRKGQGGE-------CEIRRAGPGSEELTVTNARYVP-----VSCEKEVDALLH  518
            +     GT  G  G         C I++      +E  V  A Y P     +SC    +  LL
Sbjct  315  N--NNGTWNGTAGSNNTITLPCRIKQIINMWQE--VGKAMYAPPISGNISCSSNITGLL-  369

Query  519  LARQNRAVARTAQNERSSRSHSVFQ  543
                    + R    N+ SS     F+
Sbjct  370  -------LTRDGGNDNSSNGTETFR  387
```

Figure 18. BLAST output of envelope glycoprotein of HIV 1 showing similarity to human kinesin-like protein.



Alignments

ORF1a [Duck coronavirus]
Sequence ID: **AKF17723.1** Length: 3981 Number of Matches: 18
Range 1: 890 to 1031

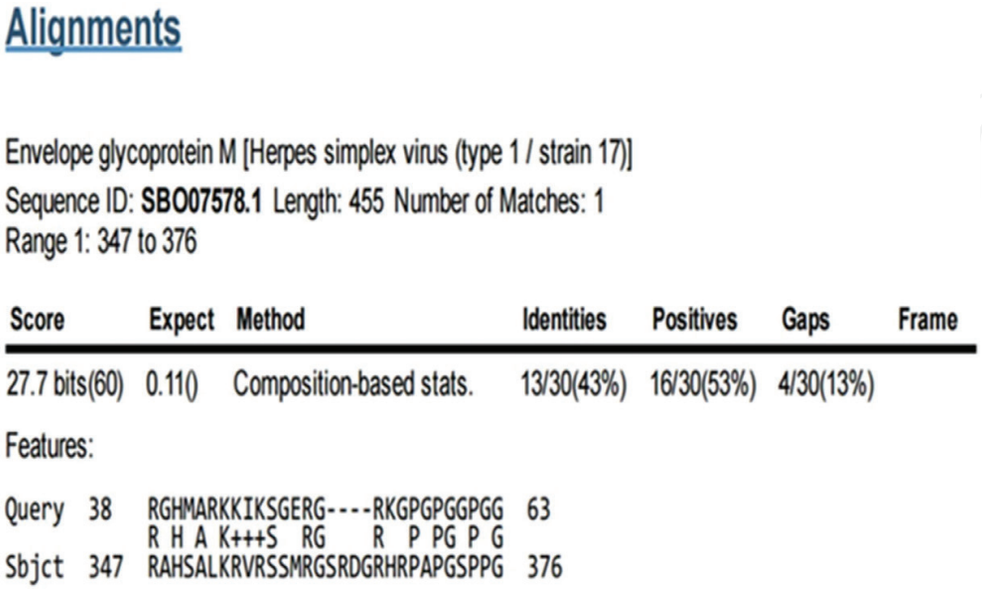
Score	Expect	Method	Identities	Positives	Gaps	Frame
79.3 bits(194)	2e-14()	Composition-based stats.	58/155(37%)	78/155(50%)	26/155(16%)	

Features:

```
Query  819   EIPKKEEVKSPVKEEEKPQEVKVKEP------PKKAEEEKAPATPKTEEKKDSKKEEAPK  872
             E PK+E  K    EE+KP+E  V+ P      P+K EE+K   TP          E PK
Sbjct  890   ETPKEETQKPQKVEEQKPKETPVETPKEETQKPQKVEEQKPKETPV----------ETPK  939

Query  873   KEAPKP-KVEEK--KEPAVEKPKESKVEAKKEEAEDKKKVP--TPEKEA--PAKVEVKED  925
             +E  KP KVEE+  KE  VE PKE    + +K E +   K+ P  TP++E   P KV    E+
Sbjct  940   EEIQKPQKVEEQKPKETPVETPKEETQKPQKVEEQKPKETPVETPKEETQKPQKV---EE  996

Query  926   AKPKEKTEVAKKEPDDAKAKEPSKPAEKKEAAPEK  960
              KPKE      +  PD +   EP K + + +  P K
Sbjct  997   RKPKETLLNKQFLPDSSSDDEPRKKSFRFKLKPTK  1031
```

Figure 19. BLAST output of ORF1 of corona virus showing similarity to human neurofilament heavy polypeptide.

HHV1 and envelope glycoprotein M of HHV2 at partially overlapping positions. Neurogranin (AA position: 38–63) has a similarity to the envelope glycoprotein M of HHV1(UniProtKB-A0A181ZHE7 (A0A181ZHE7_HHV11) (sequence ID: SBO07578.1 AA position: 347–376) (**Figure 20**). Neurogranin (AA position: 38–64) also has a similarity to the envelope glycoprotein M of HHV2 (UniProtKB-A0A0Y0R357 (A0A0Y0R357_HHV2)) (sequence ID: AMB66044.1 AA position 389–416) (**Figure 21**). Neurogranin functions as a signaling messenger, a substrate for protein kinase C and has affinity to calmodulin in the absence of calcium. These similarities



Alignments

Envelope glycoprotein M [Herpes simplex virus (type 1 / strain 17)]
Sequence ID: **SBO07578.1** Length: 455 Number of Matches: 1
Range 1: 347 to 376

Score	Expect	Method	Identities	Positives	Gaps	Frame
27.7 bits(60)	0.11()	Composition-based stats.	13/30(43%)	16/30(53%)	4/30(13%)	

Features:

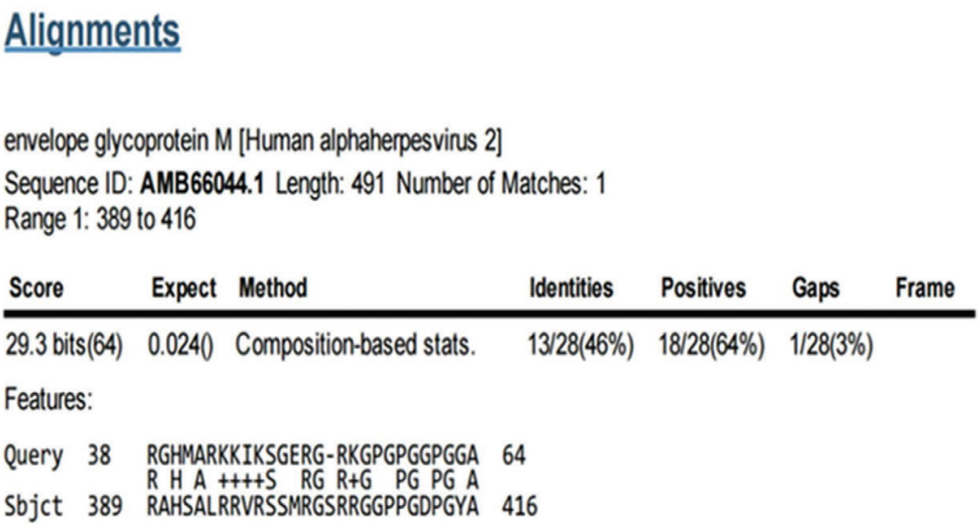
```
Query  38   RGHMARKKIKSGERG----RKGPGPGGPGG  63
            R H A K+++S  RG    R  P PG P G
Sbjct  347  RAHSALKRVRSSMRGSRDGRHRPAPGSPPG  376
```

Figure 20. BLAST output of envelope glycoprotein of HHV 1 showing similarity to human protein neurogranin.

of HHV1 & 2 with neurogranin could have an interaction with viral transport into the host cell Golgi network and subsequently to the host nucleus [38].

1.2.13. JE 2 peptide similarity to human protein secretogranin-3

Secretogranin-3 UniProtKB-Q8WXD2 (SCG3_HUMAN) is a protein of molecular mass 53,005 Da. Secretogranin-3 (AA position: 139–190) has a similarity to the polyprotein of Japanese encephalitis virus (UniProtKB-G3LHD8 (G3LHD8_9FLAV) (sequence ID: SBO07578.1 AA position: 2744 to (**Figure 22**). Secretogranin-3 is a member of the chromogranin/secretogranin



Alignments

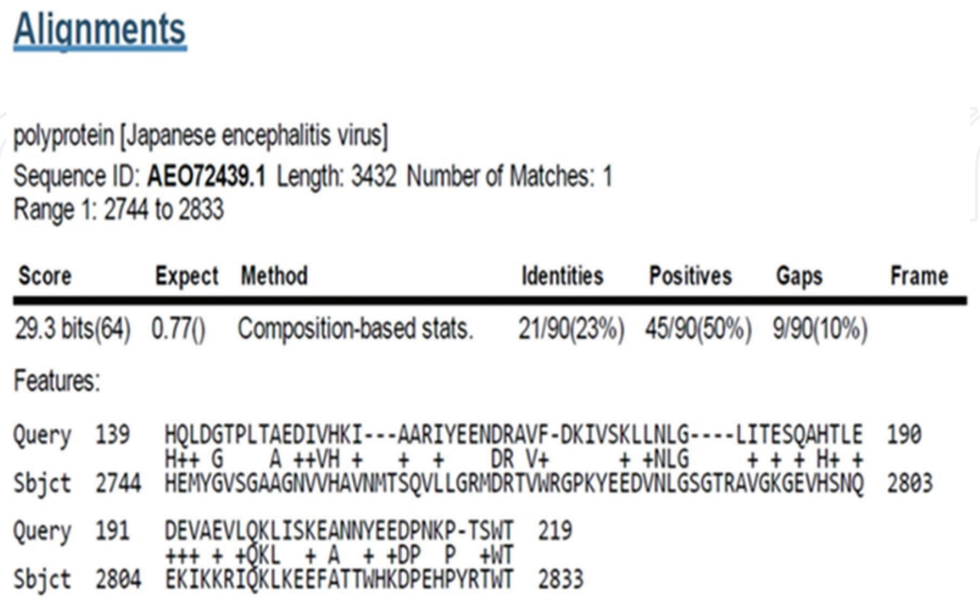
envelope glycoprotein M [Human alphaherpesvirus 2]
Sequence ID: **AMB66044.1** Length: 491 Number of Matches: 1
Range 1: 389 to 416

Score	Expect	Method	Identities	Positives	Gaps	Frame
29.3 bits(64)	0.024()	Composition-based stats.	13/28(46%)	18/28(64%)	1/28(3%)	

Features:

```
Query  38   RGHMARKKIKSGERG-RKGPGPGGPGGA  64
            R H A ++++S  RG R+G  PG PG A
Sbjct  389  RAHSALRRVRSSMRGSRRGGPPGDPGYA  416
```

Figure 21. BLAST output of envelope glycoprotein of human alpha herpes virus 2 showing similarity to human protein neurogranin.



Alignments

polyprotein [Japanese encephalitis virus]
Sequence ID: **AEO72439.1** Length: 3432 Number of Matches: 1
Range 1: 2744 to 2833

Score	Expect	Method	Identities	Positives	Gaps	Frame
29.3 bits(64)	0.77()	Composition-based stats.	21/90(23%)	45/90(50%)	9/90(10%)	

Features:

```
Query  139   HQLDGTPLTAEDIVHKI---AARIYEENDRAVF-DKIVSKLLNLG----LITESQAHTLE  190
              H++ G    A ++VH +   +  +    DR V+      + +NLG     + + + H+ +
Sbjct  2744  HEMYGVSGAAGNVVHAVNMTSQVLLGRMDRTVWRGPKYEEDVNLGSGTRAVGKGEVHSNQ  2803

Query  191   DEVAEVLQKLISKEANNYEEDPNKP-TSWT  219
              +++ + +QKL  + A  + +DP  P  +WT
Sbjct  2804  EKIKKRIQKLKEEFATTWHKDPEHPYRTWT  2833
```

Figure 22. BLAST output of polyprotein of JE 2 showing similarity to human protein secretogranin-3.

family of neuroendocrine secretory proteins comprising a number of significant cellular functions. In an experimental mouse model, autoimmunity with secretogranin was associated with encephalitis [39]. These similarities identified in the host-pathogen could affect neuro endocrine secretory protein release and autoimmunity.

2. Creating a schematic model

The sequence similarities in agrin,caprin-2,2′,3′-cyclic-nucleotide 3′-phosphodiesterase, galanin peptide, kinesin-like protein, neurofilament heavy polypeptide, neurogranin and secretogranin-3 with its corresponding pathogenic peptide/s could have a number of cellular-level implications which include alternations in receptor binding, signaling/synaptic transmission, metabolic alteration, inflammation, resulting in autoimmunity and consequently neuropathy (**Figure 23**) [11, 40].

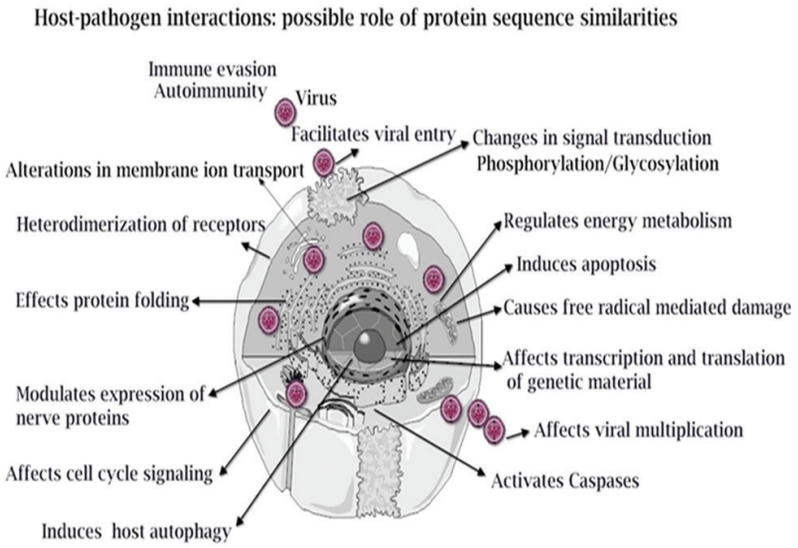


Figure 23. A model for the modes of host-pathogen interaction and possible intracellular regulation of metabolic activities.

3. Conclusion

In conclusion, it is important to conduct bioinformatic searches and design wet experiments with the objective of identifying a vast number of functionally significant peptides for further comparison and study. Bioinformatic search tools and various available databases are to be extensively explored to rapidly develop possible neuroprotective or pathogenic peptide sequences. These peptides can be further explored as targets to generate recombinant antibodies. This exercise can also be used to develop an efficacious and safe vaccine against pathogens that demonstrate no autoimmune cross-reactions. It can also contribute to design peptide/drug molecules to neutralize the effects of neurotoxins. Bioinformatics is the key to open the door of understanding medical and biological processes in the future.

Acknowledgements

We acknowledge short-term project works of Do Eon Lee of York University, 700 Keele St, Toronto, ON M3 J 1P3, Canada and Logeshwaran Vasudevan of Bharathidasan University, Palkalaiperur, Tiruchirappalli, Tamil Nadu 620024, and Dr Sharon Bushi of Morristown Med CtrIntnlMedcn, 100 Madison Ave, Morristown, NJ 07960 on the preliminary work of nerve protein pathogen similarity searches.

Author details

Lavanya Suneetha*, Prasanna Marsakatla, Rachel Supriya Suneetha and Sujai Suneetha

*Address all correspondence to: drlavanyasuneetha@gmail.com

Infectious Disease Research Laboratory, Nireekshana ACET, Hyderabad, India

References

[1] Pardy RD, Rajah MM, Condotta SA, Taylor NG, Sagan SM, Richer MJ. Analysis of the T cell response to Zika virus and identification of a novel CD8+ T cell epitope in immunocompetent mice. PLoS Pathogens. Feb 2017;**13**(2):e1006184

[2] Iversen T-G, Skotland T, Sandvig K. Endocytosis and intracellular transport of nanoparticles: Present knowledge and need for future studies. Nano Today. 2011;**6**(2):176-185

[3] Bullman S, Meyerson M, Kostic AD. Emerging concepts and Technologies for the discovery of microorganisms involved in human disease. Annual Review of Pathology: Mechanisms of Disease. Jan 2017;**12**(1):217-244

[4] Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods of Information in Medicine. 2001;**40**(4):346-358

[5] Wilson MR, Zha L, Balskus EP. Natural product discovery from the human microbiome, Journal of Biological Chemistry. May 26 2017;**292**(21):8546-8552. DOI: 10.1074/jbc.R116.762906. Epub 2017 Apr 7

[6] McManus RM, Heneka MT. Role of neuroinflammation in neurodegeneration: New insights. Alzheimer's Research & Therapy. Mar 2017;**9**(1):14

[7] Martinez-Martin N. Technologies for proteome-wide discovery of extracellular host-pathogen interactions. Journal of Immunology Research. 2017;**2017**:1-18

[8] Stanfield BA, Luftig MA. Recent advances in understanding Epstein-Barr virus. F1000 Research. Mar 2017;**6**:386

[9] Briken V. Molecular mechanisms of host-pathogen interactions and their potential for the discovery of new drug targets. Current Drug Targets. Feb 2008;**9**(2):150-157

[10] Barel M, Charbit A. Role of glycosylation/deglycolysation processes in *Francisella tularensis* pathogenesis. Frontiers in Cellular and Infection Microbiology. Mar 2017;**7**:71

[11] Yu X, Decker KB, Barker K, Neunuebel MR, Saul J, Graves M, Westcott N, Hang H, LaBaer J, Qiu J, Machner MP. Host–pathogen interaction profiling using self-assembling human protein arrays. Journal of Proteome Research. Apr. 2015;**14**(4):1920-1936

[12] Gutlapalli VR, Sykam A, Nayarisseri A, Suneetha S, Suneetha LM. Insights from the predicted epitope similarity between mycobacterium tuberculosis virulent factors and its human homologs. Bioinformation. Dec 2015;**11**(12):517-524

[13] Suneetha LM, Singh SS, Vani M, Vardhini D, Scollard D, Archelos JJ, Srinivasulu M, Suneetha S. *Mycobacterium leprae* binds to a major human peripheral nerve glycoprotein myelin P zero (P0). Neurochemical Research. Sep 2003;**28**(9):1393-1399

[14] Su MA, Davini D, Cheng P, Giang K, Fan U, DeVoss JJ, Johannes KPA, Taylor L, Shum AK, Valenzise M, Meloni A, Bour-Jordan H, Anderson MS. Defective autoimmune regulator-dependent central tolerance to myelin protein zero is linked to autoimmune peripheral neuropathy. Journal of Immunology. May 2012;**188**(10):4906-4912

[15] Raju R, Devi SK, Mehervani C, Kumar AS, Meena AK, Reddy PP, Pranay P, Jain S, Archelos-Gracia JJ, Suneetha S, Suneetha LM. Antibodies to myelin P0 and ceramide perpetuate neuropathy in long standing treated leprosy patients. Neurochemical Research. May 2011;**36**(5):766-773

[16] Vardhini D, Suneetha S, Ahmed N, Joshi DSM, Karuna S, Magee X, Vijayalakshmi DSR, Sridhar V, Karunakar KV, Archelos JJ, Suneetha LM. Comparative proteomics of the *Mycobacterium leprae* binding protein myelin P0: Its implication in leprosy and other neurodegenerative diseases. Infection, Genetics and Evolution. Mar 2004;**4**(1):21-28

[17] Rigden DJ, Fer Andez-Sú Arez XM, Galperin MY. The 2016 database issue of nucleic acids research and an updated molecular biology database collection, Nucleic Acids Research, 2016;**44**

[18] Alpi E, Griss J, da Silva AWS, Bely B, Antunes R, Zellner H, Ríos D, O'Donovan C, Vizcaíno JA, Martin MJ. Analysis of the tryptic search space in UniProt databases. Proteomics, Jan 2015;**15**(1):48-57

[19] Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. Methods in molecular biology (Clifton, N.J.). 2016;**1374**:23-54

[20] Foulger RE, Osumi-Sutherland D, McIntosh BK, Hulo C, Masson P, Poux S, Le Mercier P, Lomax J. Representing virus-host interactions and other multi-organism processes in the gene ontology. BMC Microbiology. Jul. 2015;**15**:146

[21] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. Sep 1997;**25**(17):3389-3402

[22] Zhang W, Coldefy A-S, Hubbard SR, Burden SJ. Agrin binds to the N-terminal region of Lrp4 protein and stimulates association between Lrp4 and the first immunoglobulin-like domain in muscle-specific kinase (MuSK), The Journal of Biological Chemistry. Nov 2011;**286**(47):40624-40630

[23] Kröger S, Pfister H. Agrin in the nervous system: Synaptogenesis and beyond. Future Neurology. Jan 2009;**4**(1):67-86

[24] Alfsen A, Yu H, Magérus-Chatinet A, Schmitt A, Bomsel M. HIV-1-infected blood mononuclear cells form an integrin- and agrin-dependent viral synapse to induce efficient HIV-1 transcytosis across epithelial cell monolayer. Molecular Biology of the Cell. Sep 2005;**16**(9):4267-4279

[25] Kenyon TK, Cohen JI, Grose C. Phosphorylation by the varicella-zoster virus ORF47 protein serine kinase determines whether endocytosed viral gE traffics to the trans-Golgi network or recycles to the cell membrane. Journal of Virology. Nov 2002;**76**(21):10980-10993

[26] Zerboni L, Sen N, Oliver SL, Arvin AM. Molecular mechanisms of varicella zoster virus pathogenesis. Nature Reviews Microbiology. Mar 2014;**12**(3):197-210

[27] Ding Y, Xi Y, Chen T, Wang J, Tao D, Wu Z-L, Li Y, Li C, Zeng R, Li L. Caprin-2 enhances canonical Wnt signaling through regulating LRP5/6 phosphorylation. The Journal of Cell Biology. Sep 2008;**182**(5):865-872

[28] Wang X, Jia Y, Fei C, Song X, Li L. Activation/proliferation-associated protein 2 (Caprin-2) positively regulates CDK14/Cyclin Y-mediated lipoprotein receptor-related protein 5 and 6 (LRP5/6) constitutive phosphorylation., The Journal of Biological Chemistry. Dec 2016;**291**(51):26427-26434

[29] Ono EAD, Taniwaki SA, Brandão P. Short interfering RNAs targeting a vampire-bat related rabies virus phosphoprotein mRNA. Brazilian Journal of Microbiology. Jul - Sep 2017;**48**(3):566-569. DOI: 10.1016/j.bjm.2016.11.007. Epub 2017 Feb 5

[30] Kammouni W, Wood H, Jackson AC. Serine residues at positions 162 and 166 of the rabies virus phosphoprotein are critical for the induction of oxidative stress in rabies virus infection. Journal of Neurovirology. Jun 2017;**23**(3):358-368. DOI: 10.1007/s13365-016-0506-8. Epub 2016 Dec 19

[31] Teles RMB, Krutzik SR, Ochoa MT, Oliveira RB, Sarno EN, Modlin RL. Interleukin-4 regulates the expression of CD209 and subsequent uptake of *Mycobacterium leprae* by Schwann cells in human leprosy. Infection and Immunity. Nov 2010;**78**(11):4634-4643

[32] Bersani M, Johnsen AH, Højrup P, Dunning BE, Andreasen JJ, Holst JJ. Human galanin: Primary structure and identification of two molecular forms. FEBS Letters. Jun 1991;**283**(2): 189-194

[33] Routhu NK, Byrareddy SN. Host-virus interaction of ZIKA virus in modulating disease pathogenesis. Journal of Neuroimmune Pharmacology. Jun 2017;**12**(2):219-232. DOI: 10.1007/s11481-017-9736-7. Epub 2017 Mar 27

[34] Gaudin R, Cunha de Alencar B, Jouve M, Bèrre S, Le Bouder E, Schindler M, Varthaman A, Gobert F-X, Benaroch P. Critical role for the kinesin KIF3A in the HIV life cycle in primary human macrophages. The Journal of Cell Biology. Oct 2012;**199**(3):467-479

[35] Malikov V, da Silva ES, Jovasevic V, Bennett G, de Souza Aranha Vieira DA, Schulte B, Diaz-Griffero F, Walsh D, Naghavi MH. HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus. Nature Communications. Mar 2015;**6**:6660

[36] Liu Q, Xie F, Siedlak SL, Nunomura A, Honda K, Moreira PI, Zhua X, Smith MA, Perry G. Neurofilament proteins in neurodegenerative diseases. Cellular and Molecular Life Sciences. Dec 2004;**61**(24):3057-3075

[37] Elliott C, Lindner M, Arthur A, Brennan K, Jarius S, Hussey J, Chan A, Stroet A, Olsson T, Willison H, Barnett SC, Meinl E, Linington C. Functional identification of pathogenic autoantibody responses in patients with multiple sclerosis. Brain. Jun. 2012;**135**(6): 1819-1833

[38] Yamamoto T, Nakamura Y, Carlson J, Tonon M, Zen M, Bassi N. A single tube PCR assay for simultaneous amplification of HSV-1/−2, VZV, CMV, HHV-6A/−6B, and EBV DNAs in cerebrospinal fluid from patients with virus-related neurological diseases. Journal of Neurovirology. Jan 2000;**6**(5):410-417

[39] Suri A, Walters JJ, Rohrs HW, Gross ML, Unanue ER. First signature of islet β-cell-derived naturally processed peptides selected by Diabetogenic class II MHC molecules. Journal of Immunology. Mar 2008;**180**(6):3849-56

[40] Dirlikov E, Major CG, Mayshack M, Medina N, Matos D, Ryff KR, Torres-Aponte J, Alkis R, Munoz-Jordan J, Colon-Sanchez C, Salinas JL, Pastula DM, Garcia M, Segarra MO, Malave G, Thomas DL, Rodríguez-Vega GM, Luciano CA, Sejvar J, Sharp TM, Rivera-Garcia B. Guillain-Barré syndrome during ongoing Zika virus transmission - Puerto Rico, January 1-31-07-2016. MMWR. Morbidity and Mortality Weekly Report. Sep 2016;**65**(34):910-914

IntechOpen