

Journal homepage: http://www.journalijar.com Journal DOI: <u>10.21474/IJAR01</u> INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Prediction of the next mutation in Hemagglutinin protein of Influenza-A virus using the variation pattern analysis.

Simranpal Singh¹, Arun Malik¹, Nishtha Pandey², Ravi Kant Pathak^{2*}.

- 1. Department of Computer Science Engineering. Lovely Professional University, Phagwara, Punjab, 144002, India.
- 2. Department of Biotechnology and Biosciences, Lovely Professional University, Phagwara, Punjab, 144002, India.

.....

Manuscript Info

Abstract

Received: 15 March 2016 Final Accepted: 12 May 2016 Published Online: May 2016

Manuscript History:

Key words: Hemagglutinin, Influenza, MSA, variation, probability score, CD-HIT, clustering.

*Corresponding Author

Ravi Kant Pathak.

Every year certain treatment strategies are developed to combat influenza that too only after the epidemic hits the population. The major setback in designing the treatment strategy is due to the variation in the surface antigenic determinants (Hemagglutinin and Neuraminidase) of the virus. In this work, the position specific contribution of an amino acid in the variation of the Hemagglutinin protein has been derived. Multiple sequence alignment of non-redundant sequences of Hemagglutinin from different strains has been used to derive a position specific weighted probability score matrix. The next-in-line variation in the subtype of the Hemagglutinin protein of the influenza A virus has been predicted using the calculated score matrix. Although the prediction has been accomplished with an average accuracy of 60%, the accuracy can still be improved. This strategy may be proven to be useful to design a drug before the outburst of the disease.

.....

Copy Right, IJAR, 2016,. All rights reserved.

.....

Introduction:-

Influenza A virus has been found to cause the most severe disease in human and have been reported to cause pandemics when crossed the species barrier (Klenk et. al., 2008). Since its first appearance in Spain during 1918-19, it has challenged human population in USA during 1957-58, Hong kong in 1968 (NIAID, NIH, 2011) and over 74 countries in 2009 (WHO, 2013). This virus is empowered with a unique structure and variation in the expression of surface proteins, which imparts a perfect self-defense mechanism. The viral envelop, mainly constituted of 2 types of glycoproteins, Hemagglutinin (HA) and Neuraminidase (NA), protect the core RNA genome which is segmented in nature (Bouvier et. al., 2008). There are 18 different types of HA reported till date (Tong et. al., 2013). HA has been reported to function in recognition of target host cells, and to facilitate the entry of the viral genome into the target cells (Whiteet al.,1997). This makes it the most important and primary target of neutralizing antibodies(Throsby et. al., 2008, Ekiert et. al., 2009, Sui et. al., 2009 and Corti et. al., 2011).

Drift from one strain to another probably depends on point mutation, which might change antigenic determinants, while the region which does not play significant role in triggering the immune response remains highly conserved between different subtypes. These single base substitutions give rise to the diversity in the pathogen (Willy et. al., 1980). Among the various strategies developed till date to solve this problem, tert-butylhydroquinone (TBHQ)(Russell et. al., 2008), neutralizing human antibodies (nABs) (Ekiert et. al., 2009), peptides (Xintian et. al., 2013) and vaccines (Chen et. al., 2011, Anne et. al., 2013) are some. However due to the high level of variability from one season to another season in the strain, these strategies have gained limited success (Ekiert et. al., 2009). The major challenge is to know which strain is going to be prevailing in the coming season, to be prepared with the defense strategy. Current work intends to address this problem and to device a method to predict the next variation in the protein.

Material and Methods:-

Data collection:-

Available protein sequences of all the types of HA were taken from PDB (Berman et. al. 2000) with the keyword Hemagglutinin. Further refinement has been done with experimental method-X-RAY AND taxonomy-VIRUS AND release date between 01-01-2010 up to 31-07-2015. The results were downloaded in the form of fasta files.

Cluster Analysis and Redundancy Check:-

Redundancy in the data obtained from PDB has been removed using CD-HIT (Li et. al., 2006). CD-HIT clustered the input protein sequences based on the identity of the characters. The value for identity percentage has been kept as 100 to cluster the duplicate entries. One representative sequence from each cluster has been derived for further analysis. This has been done to reduce the possibility of biasness of the analysis towards any specific type.

Multiple Sequence Alignment and Block Identification:-

Multiple Sequence Alignment (MSA) (Sievers et. al., 2011) was carried out on the representative sequences derived from CD-HIT. The results were visualized in Jalview (Waterhouse et. al., 2009). The consensus sequence observed from MSA has been stored for further reference. An un-gapped block of all the protein sequences has been identified from the MSA.

Formulation of a Weighted Probability Score Matrix:-

A position specific 2-D weighted probability score matrix was formulated based on a score value which is obtained by the product of probability of occurrence and weight of each amino acid at that position in the un-gapped block derived from MSA. The method for calculation of this matrix is devised based on the concept of sequence logo (Crooks et. al., 2004). A stack is calculated for each of the positions in the alignment data of proteins. The frequency of occurrence of the amino acids in that position is taken into consideration. Procedure:

- 1. In every row I (or the y axis) the number of distinct amino acid, $n_j(distinct\{aa\})$ and their frequencies, $f(aa_i)$, was counted.
- 2. In every column J (or the x axis) there is the increasing order of positions of the block.
- 3. Probability is calculated as the ratio of number of occurrence of a particular amino acid over the total number of amino acids in jth column.

P is the probability of occurrence of amino acid

4. Weight is calculated as inverse of the number of distinct amino acids in the jth column

W is the weight of each position.

5. In every cell (i,j) following formula is applied:

 $Score = P * W \qquad(3)$

Based on the above procedure, a 20 X (number of positions in the identified block) 2-D score matrix was calculated, it was then used a base in the prediction algorithm.

Identification of critical positions for prediction:-

Global pairwise alignment (Needleman and Wunsch, 1970) using EMBOSS NEEDLE (Rice et. al., 2000) was performed on the sequence for which next prediction is to be made and the consensus sequence that has been obtained during the MSA of all the non-redundant protein sequences of HA. The purpose of performing a pairwise alignment is to identify the significant positions having similarity in terms of function, structure or evolution.

Statistically Predicted Output:-

For each of the positions obtained after the pairwise alignment of the input sequence with the consensus sequence, the corresponding amino acid for that position is stored in the database. The values in the score matrix are updated for every position being predicted. Out of the updated values, the amino acid with maximum chance of occurrence is identified based on the calculated score. The amino acid showing a greater chance of occurrence is concluded to the amino acids that will occur next in the chronology.

Result and Discussion:-

Data collection:-

155 protein sequences of HA were taken after refinement as compared to the initial 498 hits obtained through simple keyword search as Hemagglutinin.

Keyword	Hits
Keyword {Hemagglutinin}	498
Keyword {Hemagglutinin} AND Taxonomy {Virus}	423
Keyword {Hemagglutinin} AND Taxonomy {Virus} AND Experimental Method {X-Ray}	412
Keyword {Hemagglutinin} AND Taxonomy {Virus} AND Experimental Method {X-Ray}	155
AND Time Period {2010 to 2015}	

Cluster analysis and redundancy check:-

155 refined sequences obtained from PDB became input to the CD-HIT which was operating on the default parameters. The identity cut off had been changed to 100% in order to eliminate the repeated sequences completely. CD-HIT has returned 79 unique clusters as shown in Figure 1. Each cluster contained one or more than one protein sequence in it. From every cluster a representative sequence has been chosen in such a way that it represents the whole cluster. A total of 79 representative sequences were retrieved.



Figure 1: Cluster Analysis

Multiple sequence alignment:-

MSA of 79 representative sequences was performed. An un-gapped block of 164 positions was observed using Jalview as shown in Figure 2 and Figure 3:

File Edit Select View Annotations Format Colour Calculate Web Senice 2022/24/260(24/m)SEQUENCE/1403 201 20 20 201 20 201
2YP2 AJPDBDJCHAMJSEQUENCE/1-503 DKLPGNDN STATLCLGHHAVPNGT IVKTITNDQ I EVTNATELVQSSTGG I CDSPHQ I LDGENCTL I DALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNES ZYP5 AJPDBDJCHAMJSEQUENCE/1-503 DKLPGNDN STATLCLGHHAVPNGT IVKTITNDQ I EVTNATELVQSSTGG I CDSPHQ I LDGENCTL I DALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFNNES ZYP5 AJPDBDJCHAMJSEQUENCE/1-523 QLPGNDN STATLCLGHHAVPNGT IVKTITNDQ I EVTNATELVQSSTGG I CDSPHQ I LDGENCTL I DALLGDPQCDFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLVASSGTLEFTNES ZYP5 AJPDBDJCHAMJSEQUENCE/1-523 QLPGNDN STATLCLGHHAVPNGT IVKTITDQ I EVTNATELVQSSTGG I CDSPHQ I DAGLKYDG TVL JULSNYNHL LUXLANDATI I DTSENNKLIFEKTR ZYP5 AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVPEWSYI MEKENPROGLCYPGST NDYEELKHLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVPEWSYI MEKENPROGLCYPGST NDYEELKHLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVPEWSYI MEKENPROGLCYPGST NDYEELKHLLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVPEWSYI MEKENPROGLCYPGST NDYEELKHLLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVEWSYI MEKENPROGLCYPGST NDYEELKHLLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEKVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVEWSYI MEKENPROGLCYPGST NDYEELKHLLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PGDD I CIGYHANNSTEVVDT I LERNTVTHAKD I LEKTNHGKLCKLNG I PPELGADCS I AGWLLGNPECDRLLSVEWSYI MEKENPROGLCYPGST NDYEELKHLLSSVKHF EKVKI I PKDRW ZWC3AJPDBDJCHAMJSEQUENCE/1-527 PDFD I CIGYHANNSTEVVTI T DYN I LENTYTVTI I KWI T TOTVI I KWNTOTFD EXVEKF SSUG ZWC3APDBDJCHAMJSEQUENCE/1-527 PDFD I CIGYHANNSTEVVTI T I SWNDAGAPGADAVYCVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV
2172 AIPDBD(CHAM)SEQUENCE/1-03 2172 AIPDBD(CHAM)SEQUENCE/1-03
2/P7-A/PDBD(CHAM)SEQUENCE/-332 DLEPONDSTATLCLGHHAVPNGTIVKTITNDDIEVTNATELVDSSTDGICDSPHDILDGENCTLIDALLGOPDCDGFDNKKNDLFVERSKAYSNCYPDVPDYASLRSLVASSTLEFNNES 2/PG-B/DBD(CHAM)SEQUENCE/-327 OLFONDSTATLCLGHHAVPNGTLVKTITDDDIEVTNATELVDSSTDGICDSPHDILDGENCTUDIDLEKSVEGRIDDLEKYVEDTKIDUWSYNAELLVALENDTILDTDSEMNKLFEKTR 2/PG-B/DBD(CHAM)SEQUENCE/-327 OLFONDSTATLCLGHHAVPNGTLVKTITDDDIEVTNAKDILEKTNNGKLGKNOSIFVEGSTDGICDSTDIEVESKAKSTOPTDVDYVESKSTDEFTDGICDYPDSSTDKICLGHAV 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEKVDTILERNTVTHAKDILEKTNNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEKVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEKVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEKVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLELGDCSIAGWLLGNPECDRLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEKVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLEGDCSIAGWLLGNPECDRLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEVVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLEGDCSIAGWLLGNPECDRLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLSSVKHFEKVKILPKDRW 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEVVDTILERNTVTHAKDILEKTHNGKLCKLNGIPPLGDCSIAGWLLGNPECDELIENTENPRGUENCE/-327 OFDOICIGYHANNSTEVVDTIERGITUNTHAKDILEKTNIGKLOKUNSTENDI 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEVVDTIERGITUNTHAKDILEKTNIGKLOXGAGAGAZVLVVCYDSDRSGIPERFSGSNSONTATLTISREADGALVIKVDUSSSDHVIFGGGTKLTVLGGFKAAPSVLFPSG 2/UG-B/DBD(CHAM)SEQUENCE/-327 OFDOICIGYHANNSTEVVDTIVEKNIVYAADKSTAAADGAGAZVLVVCYDSDRSGIPERFSGSNSONTATLTISREADGALVXCDDGFLUVWEYGRHAAPSVLFPSS 2/US-B/DBD(CHAM)SEQUENCE/-327 OFDICIGYHANNSTEVVDTUSKASSYTHDYHINNURDAGAZVLVVCYDSDRSGIPERFSGSNSONTATLTISREADGALVXCDDGFLUVWEYGRHAAPSVLFPSS 2/US-B/DBD(CHAM)SEQUENCE/-327 OFDICIGYHANNSTEVVDTVUEKNVVTHAQDILEKKNEVKLVS
2/PG / PDBD/CHAM/SEQUENCE/127 % OLF GAIAGFIENGWEGMIDGWYGFRHONSEGTGOAADLKSTDAAIDD IN KLIRVIEKTNEK HDIEKEFSEVEGRIDDLEKYVEDTVIDLWSYNAELLVALENDHTIDLTD SEMNKLEFTRR 2/PG / PDBD/CHAM/SEQUENCE/127 % OLF GAIAGFIENGWEGMIDGWYGFHRONSEGTAAFDE IN KKNED GAIS STGKICNNPHRILDG ID CIIDALGOPHCDVFQNETWDLFVERSKAFSNCYPYDVPD VASLRSLVASG GAIS 2/US3/PDBC/CHAM/SEQUENCE/137 % OF GAIAGFIENGWEGMIDGWYGFHRONSEGTAAFDE IT NYNVSVIEKMITDT EAVGKEFSNLERRE ENLNKKMED FLDYWTYNAELLUWENERTLDFH DSNVKNL YDKVR 2/US3/PDBC/CHAM/SEQUENCE/137 % OF GOICIG YHANNSTEKVDTILERNYTYHAKDILEKTHNGKLCKLNG IPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVK ILPKDRW 2/US3/PDBC/CHAM/SEQUENCE/1327 % OD ICIG YHANNSTEKVDTILERNYTYHAKDILEKTHNGKLCKLNG IPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVK ILPKDRW 2/US3/PDBC/CHAM/SEQUENCE/1327 % OD ICIG YHANNSTEKVDTILERNYTYHAKDILEKTHNGKLCKLNG IPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEKVK ILPKDRW 2/US3/PDBC/CHAM/SEQUENCE/137 % OF GAIAGFIENGWEGY INERVOTTHERG IEVVNATETVETTYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
2/PGE/DBCD(CHAM)SEQUENCE/1-22 SUG3/
30/03/POBC/CHAINSEQUENCE/137 OF DAIAGF IEGGWOGMVOGWYGYHHSNDOGSGYAADKESTQAKPODITNKVNSVIEKMNTOF EAVGKEFSNLERRLENN KKMEDGFLDVWTYNAELLVLWENERTLOFHDSNVKNLYDKVRM 30/03/POBC/CHAINSEQUENCE/137 PODICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGOCSIAGWLLGNPECDRLLSVPEWSYIMEKENPROGLCYPOSFNDYEELKHLLSSVKHFEKVKILPKORW 30/03/POBC/CHAINSEQUENCE/137 PODICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGOCSIAGWLLGNPECDRLLSVPEWSYIMEKENPROGLCYPOSFNDYEELKHLLSSVKHFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 PODICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGOCSIAGWLLGNPECDRLLSVPEWSYIMEKENPROGLCYPOSFNDYEELKHLLSSVKHFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNGKLCKLNGIPPLELGOCSIAGWLLGNPECDRLSVFEWSYIMEKENPROGLCYPOSFNDYEELKHLLSSVKHFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGYHANNSTEKVDTILERNSVIKFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGYHANNSTEKVDTILERNSVIKFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGYHANNSTEKVDTILERNSVIKFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGYHANNSTEKVDTILERNSVIKFEKVKILPKORW 30/06/JPDBC/CHAINSEQUENCE/137 OFODICIGHHAXANGTKVNTTTERDIEVVNATETVETTNIKKICTOGKRPTDLGGCLGLGTIGFADVKTGAUGVTVTVNAELLVLWENEHTLDFHDSNVKNLYDKVRM 30/05/JPDBC/CHAINSEQUENCE/137 OFODICIGHHAXANGTKVNTTTERDIEVVNATETVETTNIKKICTOGKRPTDLGGCGLGLGTIGFADVKTGVUVTNTSVDLEFSSDIJUGCUMDSSDHVITYNAELLVLWENEHTLDFHDSNVKNLYDKVRM 30/05/JPDBC/CHAINSEQUENCE/137 OFTGAIGGTIKUVSKIKSKIKSKIKSKIKAGUTTTTNIKKICTOGKRPTDEFSS 30/06/JPDBC/CHAINSEQUENCE/137 OFTGAIGGTIKOVGVVTTNAEDVDTVGCATVTV 35/05/JPDBC/CHAINSEQUENCE/132 OFTGAIGGTIKGVVXKIKSKSSKIKSKIKSKIKSKIKSKIKSKIKSKIKSKIKS
3WC3AP20BC/CHAINSEQUENCE/-327 PODICIGYHANNSTEKVDTILERNUTVTHAKDILEKTHNOKLCKLNGIPPLELGOCSIAGWLLGNPECDRLLSVPEWSYIMEKENPROGLCYPOSTNOYEELKHLLSSVKHFEKVKILPKORW 3WC5AP20BC/CHAINSEQUENCE/-327 PODICIGYHANNSTEKVDTILERNUTVTHAKDILEKTHNOKLCKLNGIPPLELGOCSIAGWLLGNPECDRLLSVPEWSYIMEKENPROGLCYPOSTNOYEELKHLLSSVKHFEKVKILPKORW 3WC5AP20BC/CHAINSEQUENCE/-327 GODICIGYHANNSTEKVDTILERNUTVTHAKDILEKTHNOKLCKLNGIPPLELGOCSIAGWLLGNPECDELLGNPEGNUT MEKENPROGLCYPOSTNOYEELKHLLSSVKHFEKVKILPKORW 3WC5AP20BC/CHAINSEQUENCE/-327 GODICIGYHANNSTEVUTTERGIEVVNATETVETTNIKKICTOGKRPTDLGOCGLLGTLIGPPCDDDLEFSDLIERREDTDICYPORTNEESLRDILTBRSGIGKESMAGY 3WC5AP20BC/CHAINSEQUENCE/-327 GOTICIGYHANNSTEVUTTERGIEVVNATETVETTNIKKICTOGKRPTDLGOCGLLGTLIGPPCDDDLEFSSDLIIERREDTDICYPORTNEESLRDILTBRSGIGKESMAGY 3WC5AP20BC/CHAINSEQUENCE/-327 GOTICIGYHANNSTEVVTTERGIEVVNATETVETTNIKKICTGGKRPTDLGOCGLLGTLIGPPCDDDIEFSSDLIIERREDTDICYPORTNEESLRDILTBRSGIGKESMAFYSGI 3QQC3B/PDBC/CHAINSEQUENCE/-327 GUTICIGYHANNSTEVVTTERGIEVVNATETVETTNIKKICTGGKRPTDLGOCGLLGTLIGPPCDDDIEFSSDLIIERREDTDICYPORTNEESLRDILTUSGRKABASVKVSCK 3WC5AP20BC/CHAINSEQUENCE/-327 EVDICIGYHANNSTEVVTTERGIEVVNATERVETVETTNIKKICTGGKRPTDLGOCGLLGTLIGPKPDLGOCGLLGTLOFYGVUNTYAELLVLMENEHTLDFHDSNVKNLYVKVM 3WC5AP20BC/CHAINSEQUENCE/-327 EVDICIGYHANNSTEVVTYUSGRSGYFADAKSTQAAFDGITNKVNSVIEKMNTDFEAVGKEFSNLERRENLNKKVDDGFTDIWTYNAELLVLMENEHTLDFHDSNVKNLYVKVM 3WC5AP20BC/CHAINSEQUENCE/-327 EVDICIGYHANNSTEVVTVUEKNVTVTNSVIEKNMNGGATVXVVXVXVKVVXVXVXVXVXVXVXVXVXVXVXVXVXVXV
30/03/POBD(CHAIN)SEQUENCE/1327 PODDICIOYHANDSTEKVDTILEENTVTHAKDILEKTHNOKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPOSTNOYEELKHLLSSVKHFEKVKILPKDRW 30/03/PDBD(CHAIN)SEQUENCE/1327 PODDICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNOKLCKLNGIPPLELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPOSTNOYEELKHLLSSVKHFEKVKILPKDRW 30/03/PDBD(CHAIN)SEQUENCE/1377 DFDDKICLGHHAVANGTKVNTLTERGIEVVNATETVETTNIKKICTQGKRPTDLGQCGLLGTLIGPPQCDQFLEFSDLIIERREGTDICYPORFTNEESLRQILRRSGGIGKESMGFTYSGI 30/03/PDBD(CHAIN)SEQUENCE/1377 DFDDKICLGHHAVANGTKVNTLTERGIEVVNATETVETTNIKKICTQGKRPTDLGQCGLLGTLIGPPQCDQFLEFSDLIIERREGTDICYPORFTNEESLRQILRRSGGIGKESMGFTYSGI 30/03/PDBD(CHAIN)SEQUENCE/1370 SVTAPOTABITCGGNOIGNEGYAADVESTQAAIDGESTAAVKESTQAFIDOITNKVNSVIEKNMTQFEAVGKEFSNLERRLENLNKKMEDGFLDVWTVNAELLVLMEMEHTLDFHDSNVKNLYDKVM 30/03/PDBD(CHAIN)SEQUENCE/1371 OSVITOPSVSVAPGGATSUVANGTGAKSVMINQQKFGQAPVLVVCYOSDNSGNFATLTISRVEADGEADVYCYOUDDSSDNVFIGGTTVTV 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKSVKNQCKASGYTFTDYHINWVRQAPGQAPVLVVCYOSDNSGNTATLTISRVEADGEADVYCYOUDDSSDNVFIGGTTVTV 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATSUVXCKASGYTFTDYHINWVRQAPGQAPLUVCYOSDNSWSNTATTISRVEADGEADVYCYOUDDSSDNVFIGGTTVTV 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKVSCKASGYTFTDYHINWVRQAPGQAPLUVCYOSDNSWIKTVTNAELLVLEKS 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKVSCKASGYTFTDYHINWVRQAPGQAPLUVCYOSDNSWIKTVTNAELLVLEKS 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKVSCKASGYTFTDYHINWVRQAPGQAPLUVCYOSDNSWIKTVTNAELUKKNDDGFIOVYCHGASDVYCYCHABDSVXKLYEKVKS 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKVSSCKASGYTFTDYHINWVRUAPGQALUKVSVILVEKNSVIEKKMTQTTAVGKFNKLERRMENLNKKVDDGFIOVYTARGULVLEKENTVVHSVILVEKVKS 30/03/PDBD(CHAIN)SEQUENCE/1322 OSVIAPGGATASUKVSSCKASGYTFTDYHINWVRUAPGGALUKVSVIEKKMTQTTAVGKFNKLERRMENLNKKVDDGFIOVYTABELVULUENERTLDPHDSNVKNLYVKVSS 30/03/PDBD(CHAIN)SEQUENCE/1323 OSVIAPGATASUKSVISSGKAADUKSTQUALUSHNKUSVIEKKMTQTTAVGKFNKLERRMENLNKKVDDGFIOVYTABELVLUENERTLDPHDSNVKNLYVKVSS 30/03/PDBD(CHAIN)SEQUENCE/1323 OSVIAPGATASUKSSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKS
30/06/JDDDC/CHAIN_SEQUENCE/1-327 PODDICIGYHANNSTEKVDTILEENTHVTYTHAKDILEKTHNOKLCKNGIPPELEGDCSIAGWULENPECDRLESVPENTYIMEKENPROBLCYPOSTNOYEELKHLLSSVKHFEKVK KLPYERVKK 30/06/JDDDC/CHAIN_SEQUENCE/1-377 ÅDFODKICLGHHAVANGTKVNTITERGIEVVNATETVETTNIKKICTOGKENTDLGODELGTESDLIEFRSDLIERREDTDICYPOSTNOYEELKHLLSSVKHFEKVK KLPYERVKK 30/02/JDDDC/CHAIN_SEQUENCE/1-377 ÅDFODKICLGHHAVANGTKVNTITERGIEVVNATETVETTNIKKICTOGKENTDLGODELGTESDLIEFRSDLIERREDTDICYPOSTNOYEELKHLLSSVKHFEKVK KLPYERVKK 30/02/JDDDC/CHAIN_SEQUENCE/1-377 ÅDFODKICLGHHAVANGTKVNTITERGIEVVNATETVETTNIKKICTOGKENTUSVIEKMNTOFEAVGKEFSNLERRLEN.NKKMEDGFLDVITYNAELLVLMENEHTLDFHDSVKNLVDKVRM 30/02/JDDDC/CHAIN_SEQUENCE/1-377 ÅDFODKICLGHHAVANGTKVKNTITERGIEVVNATETVETTNIKKICTOGKENTUSVIEKMNTOFEAVGKEFSNLERRLEN.NKKMEDGFLDVITYNAELLVLMENEHTLDFHDSVKNLVDKVRM 30/02/JDDDC/CHAIN_SEQUENCE/1-377 ÅDFIGAIAGFIEGGMTANTDOGSGYAADKESTDAKFDGOD TINKVNSVIEKMNTOFEAVGKEFSNLERRLEN.NKKMEDGFLDVITYNAELLVLMENEHTLDFHDSVKNLVDKVRM 30/05/JDDDC/CHAIN_SEQUENCE/1-327 ÅDTICIGYHANNSTDVDTVUSKASYTTDYIN 100/05/05/4ADGKSTQNAADGEGUENMGWI IHFNSGDTNYAKKPGGWTNTHTNAMEVNGLKSDDTAVYYCARGGUERSDVYYYGWNWGGGTTVT 30/05/JDDDC/CHAIN_SEQUENCE/1-327 ÅDTICIGYHANNSTDVDTVUEKVNTVTHSVNLLEDSHNGKLCKLRGVARINFDKJOWTNATENUNGLKSDDTAVYYCARGGUERSDVYYYGWNWCXGYTHV 30/05/JDDDC/CHAIN_SEQUENCE/1-327 ÅDTICIGYHANNSTDVDTVUEKVNTVTHSVNLLEDSHNGKLCKLRGVAPLHLGKSVISVIEKMNTOFTAVGKEFNKLERRMENLNKKVDDGFID IWTYNAELLVLLENERTLDFHDSSVSSTERFEIFFKSSS 30/05/JDDDC/CHAIN_SEQUENCE/1-327 ÅDTICIGYHANNSTDVDTVUEKVNTVTHADDILEKTHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVEKNPENDTCYPGHFADYEELREDLSSVSSTERFEIFFKSSS 30/05/JDDDC/CHAIN_SEQUENCE/1-329 ÅDGTLCIGYHANNSTDVDTVUEKNVTVTHADDILEKTHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVEKAPENDTCYPGHFADYEELKHLLSRINHFEKIDIIPKSSWS 1862/JDDDC/CHAIN_SEQUENCE/1-327 ÅDTICIGYHANNSTEGVDTIMEKNVTVTHADDILEKTHNGKLCCLDGVCVFLILRDCSVAGWULGNPMCDEFINVPEWSYIVEKANPVNDLCYPGOFNDYEELKHLLSRINHFEKIDIIPKSSWS 1862/JDDDC/CHAIN_SEQUENCE/1-329 ÅDGTLCIGYHANNSTEGVDTIMEKNVTVTHADDILEKTHNGKLCDLDGVKVLLIERUSSVS 1862/JDDDC/CHAIN_SEQUENCE/1-327 ÅDICIGYHANNSTEGVDTIMEKNVTVTHADDILEKKHNGKLCDLDGVCVF
3MG63/pDBD(pCMM)SEQUENCE/-327 ADPOKICLGHAAVANGTKVNTLTENGIEVVNATETVETTNIKKICTQGKRPTALGQCGLGFTIGPQGCDGFLEFSSDLIIERREGTDICYPORFTNEESLRQILVAMENDHTIDLADSEMSKLYERVKK 3MG63/pDBD(pCMM)SEQUENCE/-327 ADPOKICLGHAAVANGTKVNTLTENGIEVVNATETVETTNIKKICTQGKRPTALGQCGLGFTIGPQGCDGFLEFSSDLIIERREGTDICYPORFTNEESLRQILVAMENDHTIDLADSEMSKLYERVKK 3MG63/pDBD(pCMM)SEQUENCE/-327 ADPOKICLGHAAVANGTKVNTLTENGIEVVNATETVETTNIKKICTQGKRPTALGQCGLGFTIGPQGCDGFLEFSSDLIIERREGTDICYPORFTNEESLRQILRRSGGGKESMGFTYSG 3MG63/pDBD(pCMM)SEQUENCE/-327 EVQLVQSGAEVKLYPGASVKVSCKASGYTFTDYHINWVRQAPGQAEVVLVCYDSDRPSGIPERFSGSNSGNTATLTISRVEAGDEADYYCQVWDSSDHVIFGGETKLTVLQQPKAAPSVTLFPPSS 3MG63/pDBD(pCMM)SEQUENCE/-327 EVQLVQSGAEVKLYPGASVKVSCKASGYTFTDYHINWVRQAPGGGLENMOWIHPNSGDTINYQKFGGWTNTRDTAISTAYMEVNGLKSDDTAVYYCARGGLEPRSVDYYYYGMDVWGOTTUTV 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLVQSGAEVKLYPGASVKVSCKASGYTFTDYHINWVRQAPGGGLENMOWIHPNSGDTINYQKFGGWTNTRTDTAISTAYMEVNGLKSDDTAVYYCARGGLEPRSVDYYYYGMDVWGOTTUTV 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLVQSGAEVKLYPGASVKVSCKASGYTFTDYHINWVRQAPGGLENMOWIHPNSGDTINYQKFGGWTNTRTDTAISTAYMEVNGLKSDDTAVYYCARGGLEPRSVDYYYYGMDVWGOTTUTV 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLVQSGALGAGAEVKLYPGASVKVSCKASGYTATDYHINWVRQAPGGLENGXXGNWIGNTGHTNYGKFFNKLERRMENLNKKVDDGFIDIWTYNAELLVLLENERTLDFHDSNVKLYEKVKS 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLUQSGALGAEVKLYDVTVTVUTVTVVVVTHAQULEKVVSVLUEKNINTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWTYNAELLVLENERTLDFHDSNVKLYEKVS 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLUGSGAADLKSTQNAIDEUSKISTERFEIFFKSS 3SMG53/pDBD(pCMM)SEQUENCE/-327 EVQLUGSGALGAEVKLYVTHAQULEEDKHNGKLCLKGVAPLHLGKCNIAGWILGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFDIDYEELKHLSRINHFEKQIIPSS 3SMG53/pDBD(pCMM)SEQUENCE/-337 EVGLUGSGAADLKSTQNAIDEUSKSSFXADKESTQAIDGVTNKVSVLEKNVVSVSVKVVSVKVSVKVVSVSVSVSVSVSVSVSVSVSV
3MG6/2002/DAM/SEQUENCE/-317 ADPONNING LEGENDA VANDIKVNTLTERGIEVVNATETVETTNIKKINGENDEKDELDEGELGELLGEREGEDLIERREGIDLCYGARFINEESLRDILTERREGIDLCYGARFINEESLRDILERREGIDCYGARFINEESLRDILTERGIGUND WYGYRHNSNDQOSGYAADKESTQAAFDGITNKVNSVIEKMNTQFEAVGKEFSNLERRLENLNKKMEDGFLDVWTYNAELLVLMENEHTLDFHDSNVKNLYGYG 3QQG/201/2020/CAM/SEQUENCE/-327 EVQLVQSGAEVKKPGASVKVSCKASGYTFTDYHINWVRQAPGQAEUNVGYDSDRPSGIPERFSSNSGNTATLTISTVEADELNKKMEDGFLDVWTYNAELLVLMENEHTLDFHDSNVKNLYEKVK 3SMG5/2020/CAM/SEQUENCE/-327 EVQLVQSGAEVKKPGASVKVSCKASGYTFTDYHINWVRQAPGQAEUNVGYDSDRPSGIPERFSSNSGNTATLTISTVEADEADYYCQUWDSSSDHVIFGGGTKLTVLGGFNADVWGQGTVTV 3SMG5/2020/CAM/SEQUENCE/-327 EVQLVQSGAEVKKPGASVKVSCKASGYTFTDYHINWVRQAPGQAEUNVGYDSDRPSGIPERFSSNSGNTATLTISTVEADEADYYCQUWDSSSDHVIFGGGTKLTVLGGFNKVKNLYEKVKS 3SMG5/2020/CAM/SEQUENCE/-327 EVQLVQSGAEVKKPGASVKVSCKASGYTFTDYHINWVRQAPGQAEUNVGYDSDRPSGIPERFSSNSGNTALTLTISTVEADEADYYCQUWDSSSDHVIFGGETLUULENERTLDFHDSWVKNLYEKVKS 3SMG5/2020/CAM/SEQUENCE/-327 EVTIGGTAINSTDTVDTVEKNVTVTHSVNLLEDKHOKLGKGAADUSTTAALIDEITNKVNSVIEKMNTQTTAVGKERNLERRMENLNKKVDDGFIDIWTYNAELLVLLENERTLDFHDSNVKNLYEKVKS 3SMG5/2020/CAM/SEQUENCE/-327 FDTIGGTAIGFIEGGWTGNNDGWYGYHHONEQGSGYAADDKSTQUALISTUALIDEITNKVNSVIEKMNTQTAVGKERNLERRIENLNKKVDDGFIDIWTYNAELLVLLENERTLDFHDSNVKNLYEKVKS 3SMG5/2020/CAM/SEQUENCE/-327 FDTIGGTAINSTDTVDTVLEKNVTVTHSVNLLEDKHNGKUCKGVAPLHLGKCHIAGWILGNPCDESISTASSMSYIVETPSSNDTCYPGDFIDYEELREDLSSVSSFERFEIFPKTSS 2020/E2020/CAM/SEQUENCE/-327 FDTIGGTAINSTDTVDTVLEKNVTVTHSVNLLEDKHNGKUCKUGVAPLHLGKCHIAGWILGNPCDEFINVPEWSYIVEKANPVNDLCYPGDFIDYEELKHLLSRINHFEKIQIIPKSSNS 3805/2020/CAM/SEQUENCE/-327 FDDICIGYANNSTDTVDTVLEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILGKCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFIDYEELKHLLSRINHFEKIQIIPKSSNS 4882/2020/CAM/SEQUENCE/-327 FDDICIGYANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILGCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSNS 4882/2020/CAM/SEQUENCE/-327 FDDICIGYANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILGCSVAGWLLGNPMCDEFINVPEWSYIVEKNPVNDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSNS 4882/2020/CAM/SEQUENCE/-327 FDDICIGYANNSTEQVDTIMEKNVTVTHAQDIL
3QQDB/DCM/MISEQUEMCE/1/17% GLF 6A IAGF IEG6WOGMVG6WYGYHHSNDOGSGYAADKESTQAAFDG IT NKVNSVIEKMNTDF EAVGKEFSNLERRLEN.NKKMEDGFLDVWTYNAELLVLWEHEHTLDFHDSNVKNLYVKVM 3SW5L/DCB/C/AM/SEQUEMCE/1/27 EVQLVQSGAEVKKPGASVKVSCAASGYTF DYHINWRQAPGQGLEWMGWIHSO IT NKVNSVIEKMNTDF AISTAYWEVNGLKSDDTAVYYCAWDSSDNVJFGGGTKLTVLGGT KAVKVFM 3SW5J/DCB/C/AM/SEQUEMCE/1/27 EVQLVQSGAEVKKPGASVKVSCAASGYTF DYHINWRQAPGQGLEWMGWIHSDDT NYAKFPKUGWVTHTDTAISTAYWEVNGLKSDDTAVYYCARGGLEPS 3SW5J/DCB/C/AM/SEQUEMCE/1/27 EVQLVQSGAEVKKPGASVKVSCKASGYTF DYHINWRQAPGQGLEWMGWIHSDDT NYAKFPKUGWVTHTDTAISTAYWEVNGLKSDDTAVYYCARGGLEPS 3SW5J/DCB/C/AM/SEQUEMCE/1/27 EVQLVQSGAEVKKPGASVKVSCKASGYTF DYHINWRQAPGQGLEWMGWIHSDDT NYAKFPKUKERMEN.NKKVDDGFID IWTYNAELLVLLENERTLDFDFDSNVXKLVYEKVK 3SW5J/DCB/C/AM/SEQUEMCE/1/27 EVGLVQSGAEVKKPGASVKVSCKASGYTF DYHINWRQAPGQGLEWMGWIHSDT NYAKFPKUKERMKEN.NKKVDDGFID IWTYNAELLVLLENERTLDFDFDSNVXKLVYEKVK 3SW5J/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNDGWYGYHHONEDGSGYAADUKSTQNAID EI TNKVNSVIEKKMTGT TAVGKEFNKLERRMEN.NKKVDDGFLD IWTYNAELLVLENERTLDFDFDSNVXKLYYEKVKS 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNDGWYGYHHONEDGSGYAADUKSTQNAID EI TNKVNSVIEKMTGT TAVGKEFNKLERRIEN.NKKVDDGFLD IWTYNAELLVLENERTLDFNSS 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNDGWYGYHHONEDGSGYAADUKSTQNAID EI TNKVNSVIEKMTGT TAVGKEFNKLERRIEN.NKKVDDGFLD IWTYNAELLVLENERTLDFNSS 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNDGWYGYHHONEDGSGYAADUKSTQNAID EI TNKVSVIEKMTGT EAVGKFNKLERRIEN.NKKVDDGFLD IWTYNAELLVLESTIN HTEKTOFT 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNGWYGHHSDESGYAADUKSTQNAIVKLYSVEKL 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 GLFGAIAGFIEG6WTGWNGWYGHHSDESGYAADUKSTQNAID EI TNKVSVIEKNTGT EAVGKFNKLERRIEN.NKKVDDGFLD IWTYNAELLVLESTIN HTEKTOFT 3SBEJ/DCB/C/AM/SEQUEMCE/1/27 DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVTKVLISVENKLYSVEKLEBRENEN KKWEDGFD INVFEWSYIVEKANPVNDLCYFGDFNDYEELKHLLSRINHFEKTQIIPKSSW 3BBCJ/DCB/C/AM/SEQUEMCE/1/27 DQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKELILBCSVAAGWLLONPMCDEFINVFEWSYIVEKANPVNDLCYFGDFNDYEELKHLLSRINHFEKTQIIPKSSW 3BBCJB/DCB/C/AM/SEQUEMCE/1/27 DDICIGYHANNSTEQ
330/6_JDBD(CHAM)SEQUENCE/-211 QSUTDPPSVSVAP00TARITC060NDIG KSVHWINQD KPG0APVLVVCYDSD RPS0 PERFS0SVISOH ATTITISRVEA 40 EAD YYCO WDDSSSD HVIFG0G TKLTVL00PKAAPSVTLFPPS0 330/6 JPDBD(CHAM)SEQUENCE/-227 EVQLVQS0AEVKKPGASVKVSCKASGYFT DYHI INWVRQAP60QEUWVGVDSD RPS0 PERFS0SVISOH ATTTISRVEA 40 EAD YYCO WDDSSD HVIFG0G TKLTVL00PKAAPSVTLFPPS0 330/6 JPDBD(CHAM)SEQUENCE/-327 EVQLVQS0AEVKKPGASVKVSCKASGYFT DYHI INWVRQAP60QEUWMGWI HPNS0D TNAQKF0GWVTMTRD AISTAYMEVNGLKSD DTAVYYCAR06U EPRSVD YYYG BNVVKLVYEKVTV 330/6 JPDBD(CHAM)SEQUENCE/-327 EVDTICIG YHANNSTD TVD TVLEKNVTVT HSVNLLED SHNGKLCLLK0 I APLQL0 NCSVAGWI LG NPECELLISRESWSYI VEKP NPENGTCYPOHFAD YEEL WLENETTD FHDSVKKI VYEKVK 330/6 JPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTD TVD TVLEKNVTVT HSVNLLED SHNGKLCLLK0 I APLQL0 NCSVAGWI LG NPECELLISRESWSYI VEKP NPENGTCYPOHFAD YEEL VLENETTD FHDSVKKI VYEKVR 330/6 JPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTD TVD TVLEKNVTVT HSVNLLED SHNGKLCLLK0 GVAPLHLGKCNI AGWI LG NPECELLISRESWSYI VEKP NPENGTCYPOHFAD YEEL VLENETTD FHDSVKKI VYEKVR 330/6 JPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTD TVD TVT HSVNLLED SHNGKEK SKILVEK NG GVAPLHLGKCNI AGWI LG NPECES LSTAS SWSYI VETPS SDNOT CYPOHF DY DESTINE SKILVEK NVT Y HADD LEKTHNGKLCD LG VYEKVI VYEKVR VEKVR VYEKVR VYEKVR VYEKVR 320/6 ZJPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTD TVD TVEKVNT THADD I LEKTHNGKLCD LG VYEKVI LIRD CSVAGWLLG NPMCDEF INVPEWSYI VEKNP VND LCYPOF NVEELKHLLSRI NHFEK I DI I PKSSWS 3876 ZJPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTED VDTIME KNVTVTHADD I LEKTHNGKLCD LG VYEVIL I RD CSVAGWLLG NPMCDEF INVPEWSYI VEKNP VND LCYPOF NDYEELKHLLSRI NHFEK I DI I PKSSWS 3876 ZJPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTED VDTIME KNVTVTHADD I LEKTHNGKLCD LG GVKPLI LRC CSVAGWLLG NPMCDEF INVFEWSYI VEKNP VND LCYPOF NDYEELKHLLSRI NHFEK I DI I PKSSWS 3876 ZJPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTED VDTIME KNVTVTHADD I LEKTHNGKLCD LG GVKPLI LRC CSVAGWLLG NPMCDEF INVFEWSYI VEKNP VND LCYPOF NDYEELKHLLSRI NHFEK I DI I PKSSWS 3876 ZJPDBD(CHAM)SEQUENCE/-327 FOTICIG YHANNSTED VDTIME KNVTVTHAD I LEKTHNGKLCD LG GVKPLI LRC SSVAGWLG N
3016#70EDICHAMISEQUENCE/-327 EVILVQ36AEVKK96ASVKVSCKASGYTF TDYHINWVRQAFGGGUPMOWIHFNISOTTNYQACFGGWUYTMTRDTAISTAYMEVNOLKSDDTAVYYCAR60LEPRSVDYYYYOMDVWGQATVTY SANGJPD20LCHAMISEQUENCE/-1327 EVILIGVHANNSTDTVDTVLEKWTVTHVNLLDSHNGKLCLLKGIAFUGLGKSVAGWILGNPEGLEINSKSVSYUEKPNPENGTCYPGHFADYEELERGLSSVSSFERFEIFPKESSW 308E-JPD20LCHAMISEQUENCE/-1377 OLFGAIAGFIEGGWTGMVDGWYGYHHQNEQOSGYAADLKSTQNAIDGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWTYNAELLVLLENERTLDSVNSFERFEIFPKESSW 308E-JPD20LCHAMISEQUENCE/-1377 OLFGAIAGFIEGGWTGMVDGWYGYHHQNEQOSGYAADLKSTQNAIDGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRS 308E-JPD20LCHAMISEQUENCE/-1372 OLFGAIAGFIEGGWTGMVDGWYGYHHQNEQOSGYAADLKSTQNAIDGITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRS 308E-JPD20LCHAMISEQUENCE/-1372 OLFGAIAGFIEGGWTGMVDGWYGYHHQNEQOSGYAADLKSTQNAIDGITNKVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLVLENERTLDYHDSNVKNLYEKVRS 308E-JPD20LCHAMISEQUENCE/-1327 OLFGAIAGFIEGGWTGMUNGWYGYHHSNEQOSGYAADLKSTQNAIDIEKKNNGVEKINGVAENLGKSVGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFIDYEELKHLLSRINHFEKIQIIPKSSS 408C-3/PD20LCHAMISEQUENCE/-1327 PD0ICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVFKLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSS 487C-3/PD20LCHAMISEQUENCE/-1327 PD0ICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKNPANDLCYPGNFNDYEELKHLLSRINHFEKIGIIPKSSS 487C-3/PD20LCHAMISEQUENCE/-1327 D0ICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKNPANDLCYPGNFNDYEELKHLLSRINHFEKIGIIPKSSS 487C-3/PD20LCHAMISEQUENCE/-1327 DVICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKNPANDLCYPGNFNDYEELKHLLSRINHFEKIGIIPKSSS 487C-3/PD20LCHAMISEQUENCE/-1327 DVICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKCCDLGVKFLIRCSSNAGWLLGNPMCDEFINVFEWSYIVEKNPADLCYPGNFNDYEELKHLLSRINHFEKIGIIPKSSS
330/63/2020/CHAINSEQUENCE/-322 OF CAIAGFIEGGWITGWVDGWYGYHHODEOSGYAADOXSTDNAING TNVKNSVIEKMITOFTAVGKEPINLERRENLNKKVDDGFIDIWTYNAELLVLLENERTLDFHDSWVKNLYKKKS 350/6-A/2020/CHAINSEQUENCE/-322 OF CAIAGFIEGGWITGWVDGWYGYHHODEOSSGYAADOXSTDNAING TNVKNSVIEKMITOFTAVGKEPINLERRIENLNKKVDDGFIDIWTYNAELLVLLENERTLDFHDSWVKNLYKKKS 350/6-A/2020/CHAINSEQUENCE/-329 OF CIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREALSSVSSFERFEIFPKTSS 350/6-A/2020/CHAINSEQUENCE/-329 OF CIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREALSSVSSFERFEIFPKTSS 350/6-A/2020/CHAINSEQUENCE/-329 OF CIGYHANNSTDTVDTVLEKNVTVTHAODILEKTHNGKLCCNLOVENUKVLVYEKVRS 350/6-A/2020/CHAINSEQUENCE/-329 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCCNLOVENUKVILVEKNKVSIDKINTGENENCE/-329 350/6-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILADCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKIOIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILADCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLSRINHFEKIOIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILADCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKIOIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILADCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKICIIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILADCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKICIIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLOGVKPLILENCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKICIIPKSSWS 362-A/2020/CHAINSEQUENCE/-327 OF CIGYHANNSTEOVDTIMEKNVTVTHAODILEKKNOKLCDLOGUKPLILENCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNOYEELKHLLSRINHFEKICIIPKSSWS 363-A/2020/CHAINSEQUENCE/-327 OKIGLGHALSGTKANTTERGSOVCYPGKFISGSINGFTSGINGFTSGINGFTSGINFTYGGTFNGFTSGINGFTSGINGFTSGINGFTSGINGFTSGINGFTSGINGFTSGINGF
33/05/a/POBD(CHAM)SEQUENCE/-327 EDTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCLLKGIAPLDLGNCSVAGWILGNPECELLISRESWSYIVEKPNPENGTCYPGHFADYEELREDLSSVSFERFEIFPKESSW 30/05/a/POBD(CHAM)SEQUENCE/-327 GLFA IAGFIEGGWTGNVDGWYGYHHDNEOGSGYAADLKSTDAAIDET TNVKNSVIEKNNTDFTAVGKEFNHLEKRIEN NKKVDDGFLDIWTTNAELLVLLENERTLDYHDSNVKNLYKVRS 30/05/a/POBD(CHAM)SEQUENCE/-326 GDTLCIGYHANNSTDTVDTVUEKNVTVHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVEKPSD010 CYPGDFNDYEELKHLLSRINHFEKIDIIPKSSS 30/05/a/POBD(CHAM)SEQUENCE/-326 DDICIGYHANNSTDTVDTVDEKNVTVHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-326 DDICIGYHANNSTEDVDTIMEKNVTVHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-326 DDICIGYHANNSTEDVDTIMEKNVTVHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKNPDADLCYPGNFNDYEELKHLLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-327 PDDICIGYHANNSTEDVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKINPADLCYPGNFNDLEKHLLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-327 PDDICIGYHANNSTEDVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKINPADLCYPGNFNDLCYPGNFNNEELKHLLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-327 PDDICIGYHANNSTEDVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKINPADLCYPGNFNDLCYPGNFNDLCYPGNFNNEELKILLSRINHFEKIDIIPKSSS 38/05/a/POBD(CHAM)SEQUENCE/-327 PDDICIGYHANNSTEDVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFNNEELKHLLSRINHFEKIDIIPKSSS 38/05/05/04/04/05/05/04/05/05/05/05/05/04/04/05/05/04/05/05/04/05/05/05/05/05/05/05/05/05/05/05/05/05/
308E_UPDBD(HAM)SEQUENCE/-177 GLFGALAGFIEGGWTGMVDGWYGYHHONEDGSGYAADLKSTDNAIDE ITNKVNSVIEKMNTDFTAVGKER NHLEKRIEN. NKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRS 308E_WPDBD(HAM)SEQUENCE/-329 PGDTLCIGYHANNSTDTVDTVEKNUTHSVNLEDKHNGKLGKEGVAPLHLGKCNLAGWILGNPECESLSTASSWSYIVETPSSDNGTCPGOFIDYEELKRLDSRIVSFERFEIFSS 308E_WPDBD(HAM)SEQUENCE/-329 PGDTLCIGYHANNSTDTVDTVHDDILEKTHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLISRINHFEKIDIIPKSSWS 48G2#PDBD(HAM)SEQUENCE/-327 PGDTLCIGYHANNSTEGVDTIMEKNVTVTHADDILEKTHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLISRINHFEKIDIIPKSSWS 48G2#PDBD(HAM)SEQUENCE/-327 PGDTLCIGYHANNSTEGVDTIMEKNVTVTHADDILEKTHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLISRINHFEKIDIIPKSSW 48G2#PDBD(HAM)SEQUENCE/-327 PGDTLCIGYHANNSTEGVDTIMEKNVTVTHADDILEKTHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLLSRINHFEKIDIIPKSSW 48G2#PDBD(HAM)SEQUENCE/-327 PDDTLCIGYHANNSTEGVDTIMEKNVTVTHADDILEKKHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLLSRINHFEKIDIIPKSSW 48G2#PDBD(HAM)SEQUENCE/-327 PDDTCIGYHANNSTEGVDTIMEKNVTVTHADDILEKKHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLLSRINHFEKIDIIPKSSW 48G3#PDBD(HAM)SEQUENCE/-327 PDDTCIGYHANNSTEGVDTIMEKNVTVTHADDILEKKHNGKLCDLDGVKPLILRDCSVAGWLGNPMCDEFINVPEWSYIVEKNPVNDLCYPOFNDYEELKHLLSRINHFEKIDIIPKSSW 48G3#PDBD(HAM)SEQUENCE/-327 PDTCIGHALAGFIAG9FIADAADGEGTAADYKSTDAAIDDITGKLNRIEIFNT
SUBE-BYDBOLOCHAMISEQUENCE/-329 POOTLCIGYHANN STOTVDTVLEKNUTVTHSVNLLEDKHNOKLCKLRGVAPLHLGKCHIAGWILGNPHCDESLSTASSWSYIVETPSSNOSTCYPODFIDYEELREDLSSVSSFERFEIPFSS SZMX-EJCBELOCHAMISEQUENCE/-329 POOTLCIGYHANN STOTVDTVLEKNUTVTHAQDILEKTHNGKLCDLDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFIDYEELKHLLSRINHFEKIQIIPKSWS GROZEJPCOELONAMISEQUENCE/-327 PODICIGYHANN STEQVDTIMEKNUTVTHAQDILEKTHNGKLCDLDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFIDYEELKHLLSRINHFEKIQIIPKSSW GROZEJPCOELONAMISEQUENCE/-327 PODICIGYHANN STEQVDTIMEKNUTVTHAQDILEKTHNGKLCDLDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW GROZEJCOCHAMISEQUENCE/-327 PODICIGYHANN STEQVDTIMEKNUTVTHAQDILEKKHNOKLCDLDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW GROZEJCOCHAMISEQUENCE/-327 PODICIGYHANN STEQVDTIMEKNUTVTHAQDILEKKHNOKLCDLDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW GROZEJCOCHAMISEQUENCE/-327 PLGICIGYHANN STEQVDTIMEKNUTVTHAQDILEKKHNOKLCDLGDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW GROZEJCOCHAMISEQUENCE/-327 PLGICIGYHANN STEQVDTIMEKNUTVTHAQDILEKKHNOKLCDLGDGVKPLILDCSVAGWLLGNPHCDEFINVPEWSYIVEKANPVNDLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW GROZEJCOCHAMISEQUENCE/-327 PLGICIGYHANSTEQVDTIMEKNUTVTHAQDILEKKHNOKLCDLGGGULGTITGFDODFLEFSADLILENGSVVYVKIVTBOLCYPODFNDYEELKHLLSRINHFEKIQIIPKSSW
32WEEPOBD(CHAM)SEQUENCE/4326 DOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWSS 48G22APOBD(CHAM)SEQUENCE/4326 DOIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSWS 48G2APOBD(CHAM)SEQUENCE/4327 PDOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWS 48G2APOBD(CHAM)SEQUENCE/4327 PDOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWS 48G2APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGFNDYEELKHLLSRINHFEKIQIIPKSSW 48G2APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGFNDYEELKHLLSRINHFEKIQIIPKSSW 48G2APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGFNDKOVTVTHAODILEKKHNGKLCVHAGU 48G2APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMEKNVTVTHAODILEKKHNGKLCNLDGVKPLIKNCK 48G3APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMENNTVTHAODILEKKHNGKLCNLDGVKPLIKNCK 48G3APODB(CHAM)SEQUENCE/4327 DOICIGYHANNSTEOVDTIMENNTVTTHAODINTEKTORT
486228/2080/CHAMISEQUEMCE/-466 % LFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSNVKNLYDKVRL 4862-24/2080/CHAMISEQUEMCE/-327 (PDQICIGYHANNSTEQVDTIMEKNVTVTHADDILEKTHNGKLCDLODGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEK INPANDLCYPGNFHDVSELKHLLSRINHFEKIQIIPKSSWS 1842-24/2080/CHAMISEQUEMCE/-327 (PDQICIGYHANNSTEQVDTIMEKNVTVTHADDILEKTHNGKLCDLDDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFHDVSELKHLLSRINHFEKIQIIPKSSWS 1845-24/2080/CHAMISEQUEMCE/-327 (DICIGYHANNSTEQVDTIMEKNVTVTHADDILEKKHNGKLCDLDDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFHDVSELKHLLSRINHFEKIQIIPKSSWS 1855-24/2080/CHAMISEQUEMCE/-327 (DICIGYHANNSTEQVDTIMEKNVTVTHADDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGFNDYELKHLLSRINHFEKIQIIPKSWS 1855-24/2080/CHAMISEQUEMCE/-327 (DICIGYHANSTEQVDTIMEKNVTKHADDILEKKHNGKLCDLDGVKPLIRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGFNDHDTIDLADSED 1855-24/2080/CHAMISEQUEMCE/-327 (DICIGYHANSTEQVDTIMEKNVTVTHADDICYPSIDEX 1855-24/2080/CHAMISEQUEMCE/-327 (DICIGYHANSTURG) TEVWSYNAELUXANENDAGEGTAADYKSTQSAIDQITGKLMRL
4862-AIPOBOLOHAMISEQUENCE/-327 PODICIGYHANNSTEDVDTIMERKUTUTHADDILEKTHNGKLCOLDGVKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFHDYEELKHLLSRINHFEKIDIIPKSWG 4882-AIPOBOLOHAMISEQUENCE/-328 OPDDICIGYHANNSTEDVDTIMERKUTUTHADDILEKKHNGKLCHOGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGNFHDYEELKHLLSRINHFEKIDIIPKSWG 4853-AIPOBOLOHAMISEQUENCE/-1-77 ÖLFGAIAGFIAGWEGUIDGWYGFRHDANDGEGTAADYKSTDSAIDDITGKLNRIEKTNDOFELIDNEFNEVEKDIGNIN
48H2/AIPOBICICHAMISEQUEMO2/438 DPDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVKPLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFEKIQIIPKSSW 18SA BIPOBICICHAMISEQUEMO2/477 OLFGAIAGFIENGWEGLIDGWYGFRHQNAQGEGTAADYKSTQSAIDQITGKLNRLIEKTNQQFELIDNEFNEVEKQIGNVINWTRDSITEVWSYNAELLVAMENQHTIDLADSEMDKLYERVKR 18SA AIPOBICICHAMISEQUEMO2/4321 DKICLGHHALSNGTKVNTLTERGVEVVNATETVERTNIPRICSKGKRTVDLGQCGLLGTITGPPQCDQFLEFSADLIIERREGSDVCYPGKFVNEEALRQILRESGGIDKEAMGFTYSGIRTNG
185A BIPOBICICHAMISEQUENCE/1-177 GLEGA IAGE I ENGWEG LI DOWYGERHON NAO GEGTA A DYKSTOSA I DO I TOKLNRLI EKTNODE EL IDNEFNEVEKO I GNVI NWTRDSI TEVWSYN A ELLVAMENO HTI DLAD SEMDKLYERVKR 185A AIPOBICICHAMISEQUENCE/1-321 DKI CLGHHALSNG TKVNTLTERGVEVVNA TETVERTNI PRI CSKGKRTVDLGOCGLLGTI TOPPOCODE LEFSADLI I ERREGSD VCYPOKEVNEEA LROLI LESGO I DKEAMIGETYSGI RTNO!
188A: AIPOBIOJCHAMISEQUENCE/1-321 DKICLGHHALSNGTKVNTLTERGVEVVNATETVERTNIPRICSKGKRTVDLGQCGLLGTITGPPQCDQFLEFSADLIIERREGSDVCYPGKFVNEEALRQILRESGGIDKEAMGFTYSGIRTNG
400W EJPOBDICHAMISEQUENCE/1327 OPDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLD6VKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSW
40YW BIPOBIO CHAMISEQUENCE/1473 GLFGA I AGFIENGWEGMVD GWYGFRHQN AQG TGQA AD YKSTQA A I DQITGKLNRLIEKTN TEFESIESEFSEIEHQIGN VINWTKD SITDIWTYQA ELLVAMENQHTIDMAD SEMLNLYERVRK
40YW:APDBLD[CHAM]SEQUENCE/1-324_LDKICLGHHAVANGTIVKTLTNEDEEVTNATETVESTSLDRLCMKGRSHKDLGNCHPIGMLIGTPACDLHLTGTWDTLIERENAIAYCYPGATVNEEALRQKIMESGGISKISTGFTYGSSINS
120/68/POBC/CHAMISEQUENCE/1-177 §LFGAIAGFIENGWEGLIDGWYGFRHQNAQGEGTAADYKSTQSAIDQITGKLNRLIEKTNQQFELIDNEFTEVERQIGNVINWTRDSMTEVWSYNAELLVAMENQHTIDLADSEMNKLYERVKR
12016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/11/2016/1
1EDA BIPOBLICHAM SEQUENCEM-182 OLF GA I AG FIEG GWTGMVDGWYG YHHAN EQG SG YAADLKSTANA I DE I TNKVNSVI EKMNTAFTAVG KEFNHLEKRIENLNKKVDD GFLD I WTYNAELLVLENERTLD YHDSNVKNLYEKVRS 🧋

Figure 2: Ungapped block of positions from 1-124

🖳 F:\Mtech\data collection\vepresentative sequences\vepresentative_seq.fasta 🔤 🗗
File Edit Select View Annotations Format Colour Calculate Web Service
120 136 140 150 250 170 180 190 200 210 220 230 240
2/P2/APOBOCICHAMISEQUENCE/-803 EFNNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHLKFKYPALINVTMPNNEKFDKLYIWGVHHPGTDNDQISLYAQASGRITVSTKRSQQTVIPNIGSRPRVRDIPSRISIYWTIVKPGDIA
21/77.AI/2080/CHAMISEQUENCE/-503 EFNNESFNWTGVTQNGTSSACKRKSNNSFFSRLNWLTHLKFKYPALINVTMPNNEKFDKLYIWGVHHPGTDNDQIFLYAQASGRITVSTKRSQQTVIPNIGSRPRVRNIPSRISIYWTIVKPGDI
2VPGFIPDBDICHAMISEQUENCE/1-175 FEKTRRQLRENAEEM6N6CFKIYHKCDNACIESIRN6TYDHDVYRD[EALNNRFQIK6
2VPGEPDBDCCHAMISEQUENCE/4328 EFITEGFTWTGVTQNGGSNACKRGPGSGFFSRLNWLTKSGSTYPVLINVTMPNNDNFDKLYIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRPWVRGLSSRISIYWTIVKPGDV
3WU35JPDBDCjCHAMISEQUENCE/1-174 YDKVRMQLRDNVKELGNGCFEFYHKCDDECMNSVKNGTYDYPKYEEjESKLNRNEIK
3XU3:AP20BD(CHAM)SEQUENCE/1-327 LPKDRWTQHTTT66SRACAVS6NPSFFRNMVWLTEK6SNYPVAK6S(YNNTS6EQMLIIW6VHHPNDETEQRTLYQNV6TYVSV6TSTLNKRSTPEIATRPKVN6Q66RMEFSWTLLDMWDTINF
3XU5XHPCBD(CHAM)SEQUENCE/1-327 LPKDRWTQHTTT66SRACAVS6NPSFFRNNVWLTEK6SNYPVAK6SYNNTS6EQMLIIWGVHHPNDETEQRTLYQNV6TYVSV6TSTLNKRSTPEIATRPKVN6L6SRMEFSWTLLDMWDTINF
3X/U&/a/POBLD(CHAM)SEQUENCE/3-327 LPKDRWTQHTTTG6SRACAVS6NPSFFRNMVWLTEK6SNYPVAK6SYNNTS6EQMLIIWGVHHPNDETEQRTLYQNVGTYVSVGTSTLNKRSTPEIATRPKVN6L66RMEFSWTLLDMWDTINF
3W665B/20BLD(CHAM)SEQVENCE/1-182 YERVKKQLRENAEED6T6CFEIFHKCDDQCMESIRNNTYDHTQYRT(ESLQNRIQIDSGRLVPR6
30/65/a/2013/c/am/sequence/4317 FTYS6 IRTNGATSACTRS6SSFYAEMKWLLSNSDNAAFPQMTKAYR/NPRNKPAL I WGVHHSESVSEQTKLY6SGNKL ITVRSSKYQQSFTPNPGARR I DFHWLLLDPNDTVTFTFNGAF I APD
3QQOB/PBBD(CHAM)SEQUENCE/1-174 YDKVRMQLRDNVKELGNGCFEFYHKCDDECMNSVKNGTYDYPKYEE/ESKLNRNEIK
33MBJPDBD(CHAM)SEQUENCE/1211 LFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSHRSYSCQVTHEGSTVEKTVAPT
33MB3HP2BDQCHAMJSEQUENCE/4227 GTTVTVSSASTK6PSVFPLAPSSKSTS66TAAL6CLVKDYFPEPVTUVSWNS6ALTS6VHTFPAVLQSS6LYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPK
33MB5B/DBDICHAMJ3EQUENCE/1-182 YEKVKSQLKNNAKEIGNGCFEFYHKCNDECMESVKNGTYDYPKYSE[ESKLNREKIDGVRSLVPR
33MB/JPDB/JCHAM/3EQUENCE/1-327 PKESSWPNHTTTGVSASCSHNGESSFYKNLLWLTGKNGLYPNLSKS/YANNKEKEVLVLWGVHHPPNIGDDRALYHTENAYVSVVSSHYSRKFTPEIAKRPKVRDREGRINYYWTLLEPGDTIIF
3UBELIPOBDICHAMISEQUENCE/1-177 YEKVRSQLKNNAKEIGNGCFEFYHKCDNTCMESVKNGTYDYPKYSEjEAKLNREEIDSGR
3UBEXIPDBDICHAMISEQUENCE/1329 FPKTSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADTYVFVCSSRYSKKFKPEIAICPKVRDQEGRMNYYWLLVEPGDKI
32NW:EJPOBIDJCHAMISEQUENCE/1326 KSSWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVNGQSGRMEFFWTILKPNDAINFE
48628/POBD/CHAM/SEQUENCE/1-166 YDKVRLQLRDNAKELGNGCFEFYHRCDNECMESVRNGTYDYPQYSE/EA
4862.4P20BD(CHAM)SEQUENCE/4.327 PKSSWSDHEASAGVSSACPYQGRSSFFRNVVWLIKKDNAYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKPNDAINF
48H2:AP20BD(cHAM)SEQUENCE/1-328 IPKSSWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKDSTYPTIKR ⁱ SYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVKGLSGRMEFFWTILKPNDAIN
485AB/POBD/CHAM/SEQUENCE/1-177 YERVKRQLRENAEED0TGCFEIFHKCDDDCMASIRNNTYDHSKYRE/EAMQNRIQIDPVK
485A AIPOBOLICHAMISEQUENCE/1-321 GIRTNGTTSACRRSGSSFYAEMKWLLSNTDNAAFPOMTKSYKNTRKSPALIVWGIHHSVSTAEOTKLYGSGNKLVTVGSSNYOOSFVPSPGARPOVNGLSGRIDFHWLMLNPNDTVTFSFNGAF
40CW EPOBDICHAM SEQUENCE/1327 I PKSSWSDHEAS6VSSACPYQGRSSFRNVVWLTKKDNAYPTIKRSYNNTNQEDLLVLWGIHHPNDAAEQTRLYQNPTTYISV6TSTLNQRLVPKIATRSKVN6QS6RMEFFWTILKPNDAINF
4CWWBJPDBDJCHAMJSEQUENCE/1-173 YERVRKQLRQNAEEDGKGCFEIYHACDDSCMESIRINITYDHSQYREJEALLNRLNI
40YWAPDBDCHAMSEQUENCE/1-324 GSSINSAGTTKACMRNGGNSFYAELKWLVSKSKGQNFPQTTNTYRNTDTAEHLIMWGIHHPSSTQEKNDLYGTQSLSISVGSSTYQSNFVPVVGARPQVNGQSGRIDFHWTLVQPGDNITFSHN
4DV68JPDBDCJCHAMISEQUENCE/1-177 YERVKRQLRENAEEDGTGCFEIFHKCDDDCMASIRNNTYDHSKYREJEAIQNRIQIDSGR
400%:AIPOBBCICHAMISEQUENCE/1-327 FTYSGIRTNGTTSACRRSGSSFYAEMKWLLSNTDNAAFPOMTKSYK/NTRKDPALIIWGIHHSGSTTEQTKLYGSGNKLITVGSSNYQOSFVPSPGARPOVNGOSGRIDFHWLILNPNDTVTFSF
<pre>#EDABJPEBD(CHAMISEQUENCE/F-182 YEKVRSQLKNNAKE)GNGCFEFYHKCDNTCMESVKNGTYDYPKYSE[EAKLNREE]DGVRSLVPR.</pre>

Figure 3: Ungapped block of positions from 125 - 164

The consensus sequence retrieved from MSA was:-

GDFGAIAGFIENGWEGMVDGWYGYEHQNEQGSGTAADKKSTQGAIDGITGKLNSLIEKTNTQFELIGNEFN ELEKRIENLIKKVEDGFDDVWSYNAELLVLLENELTLDSHDSEVKKLYEKVRSQLRENAKESGNGCFEFYH KCDNFCMESIRNGTYDYTKYREEANLNREEIDGLRGIHHPHDEAEQTTLYQNYTTYSSVGTSSTSQRNQPEI PTRSKVNGVRGRMEFFWTILTILDPIDFESNGNNIAPEAPYKIKKKGSSGIMKSEGSEGNCGTKCQTPTGAIN SSNPFHNIHPLTIGECPKYVKSKKLVLATGLRNLPNIEKRERRIFGRIAGFIEAGWEEGGDGWYGFGHQNSE GIGEAADGIATQKAINQIAGKENRLIGKNNEEFHNGEKEFIEGEFRIQDLEINVEDDKIDDWSYNAELLVALE NQHTEDDRDLDLNDLFERNKHQLIENAEDMGNGCFKIGHKCDNACCGDICNGTYDHDTYRDEALKEEFQI KRQEIEGIRLVPR

Since the range of positions of the observed un-gapped block is from 1-164, same range of positions have been marked significant in the consensus sequence as well.

Weighted probability score matrix:-

For every position of the un-gapped block a weighted probability score has been calculated with respect to every amino acid. The same procedure is followed for each of the 164 positions and the complete 20 X 164 score matrix is obtained. Example of the score matrix for 1st position of the conserved block has been shown in the following table:

Amino acid	Weighted Probability score at Position 1	Amino acid	Weighted Probability score at Position 1
А	0	М	0
С	0.005	Ν	0
D	0	Р	0.005
E	0	Q	0.006
F	0	R	0.001
G	0.057	S	0.001
Н	0.003	Т	0.001
Ι	0.013	V	0
K	0	W	0
L	0.008	Y	0

Table 2: Weighted Probability Score for position 1

Prediction:-

Chronologically two latest outbreaks occurred as H7N9 in China, April 2016 and H5N6 again in china in March 2016 (WHO, Disease Outbreak News (DONs), 2016). Since the H5 subtype has been expressed earlier than H7, sequence of H5 has been chosen to be the input to the methodology and the predicted output is expected to be of strain H7.

Predicted output:-

Pairwise alignment of input sequence and consensus sequence was carried out using EMBOSS-NEEDLE with default parameters. The aligned positions were then processed through the prediction algorithm and the predicted sequence was obtained.

Input sequence:-

H5 (PDB ID-4KWM)

ADPGDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVKPLILRDCSVAGWLLGNPM CDEFINVPEWSYIVEKANPANDLCYPGNFNDYEELKHLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQG TPSFFRNVVWLIKKNNTYPTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRLVP KIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIVKSEVEYGNCNTKCQTPIG AINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRER

Predicted sequence:-

ADPGAQICFIYHAWNSTEQGWWIMEHNVEVTHAQDALEKTHNGAICDIDGVKPLILRDCSVAGWLLGNP MCEEEINIPELIEIVEKANPAVDSCYPGNFNDYEEEKHLLSRINHFKKIQIIPKSSWSDHEASSGVSSACPYQ

KTPS**F**FR**E**VVWLIKK**D**N**TK**PTIKRSYNNTNQEDLLILWGIHHSNDAAEQTKLYQNPTTYISVGTSTLNQRL VPKIATRSKVNGQSGRMDFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSAIVKSEVEYGNCNTKCQTP IGAINSSMPFHNIHPLTIGECPKYVKSNKLVLATGLRNSPLRER

A pairwise alignment of the predicted sequence and the expected H7 sequence having PDB ID-3M5G was carried out and a similarity of 51.5% was obtained.

Accuracy:-

The method to check the accuracy of the prediction has been chosen such that a global pairwise alignment of the predicted sequence is performed with the actual sequence P' in the phylogeny of the representative sequences. P' represents that sequence in the phylogeny that stands next to the input sequence. The prediction algorithm works with the accuracy as shown in the table below:

Table 3: Validation results								
S.	Input	Next	Number of predicted	Identity	Similarity			
No.:	Sequence	Sequence	positions	percentage	percentage			
1	4BGZ:A	4CYW:A	29	33.9	52			
2	3ZNK:E	2YP2:A	25	21.3	34			
3	2YP2:A	2YP7:A	33	93.6	95			
4	3UBE:L	4BGZ:B	129	72.9	86.4			
5	4N60:D	4NRJ:F	115	38.5	52.2			
6	4M40:E	4NRJ:E	32	81.2	86.4			
7	4F23:C	4FIU:C	33	95.1	95.7			
8	4HKX:A	4FQR:X	36	24.2	41.5			
9	3KU3:A	3KU5:A	37	33.6	49.6			
10	4LKI:A	4M4Y:A	28	30.7	49.7			
			AVERAGE	53	60			

Using this method, upon calculation of an average of 10 completely random protein sequences a similarity of 60% and an identity of 53% percent was observed.

Conclusion:-

79 non redundant representative sequences have been used to perform MSA and based on the position specific weighted probability score, which represents the variation effect in the due course of evolution, a methodology has been designed to predict the next in line subtype. Although the accuracy of the method has been calculated as 60% (based on similarity), scope of improvement still lies open. The accuracy if could be increased further, it can be implemented to other viral diseases also, in which the viral pathogen adopts the same strategy of variation to bypass the immune system without getting identified. This includes HIV-AIDS, SIV and other diseases. Future endeavour would be to increase the accuracy and develop a prediction tool based on the developed methodology. The output of the prediction tool shall be helpful in designing drugs/vaccines which can be effective against any subtype.

References:-

- 1. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, and G. J. Barton, (2009), Jalview Version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, vol. 25, pp. 1189-1191.
- Anne-Cécile V. Bayne, David Boltz, Carole Owen, Yelena Betz, Goncalo Maia, ParastooAzadi, Stephanie Archer-Hartmann, Ross Zirkle, J. Casey Lippmeier, (2013), Vaccination against Influenza with Recombinant Hemagglutinin Expressed by Schizochytrium sp. Confers Protective Immunity. PLOS ONE, 8(4): e61790, doi:10.1371/journal.pone.0061790
- 3. Bouvier NM, Palese P (September 2008). The biology of influenza viruses. Vaccine. 26 Suppl 4: D49–53. doi:10.1016/j.vaccine.2008.07.039. PMC 3074182.PMID 19230160.
- 4. **Chen JR, Ma Č, Wong CH.,**(2011),Vaccine design of hemagglutinin glycoprotein against influenza. Trends in Biotechnology 29 (9): 426–434doi:10.1016/j.tibtech.2011.04.007
- Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, Vachieri SG, Pinna D, Minola A, Vanzetta F, Silacci C, Fernandez-Rodriguez BM, Agatic G, Bianchi S, Giacchetto-Sasselli I, Calder L, Sallusto F, Collins P, Haire LF, Temperton N, Langedijk JP, Skehel JJ, Lanzavecchia A (August 2011). A

neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. Science 333 (6044): 850–6. doi:10.1126/science.1205669. PMID 21798894

- Ekiert DC, Bhabha G, Elsliger MA, Friesen RH, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA (April 2009). Antibody recognition of a highly conserved influenza virus epitope. Science 324 (5924): 246–51. doi:10.1126/science.1171491. PMC 2758658. PMID 19251591
- 7. G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, (2004), WebLogo: a sequence logo generator *Genome research*, vol. 14, pp. 1188-1190, .
- 8. **H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne** (2000) The Protein Data BankNucleic Acids Research, 28: 235-242.
- 9. Klenk, Hans-Dieter; Matrosovich, Mikhail; Stech, Jürgen (2008). Avian Influenza: Molecular Mechanisms of Pathogenesis and Host Range. Animal Viruses: Molecular Biology.Caister Academic Press. ISBN 978-1-904455-22-6.
- Needleman, Saul B.; and Wunsch, Christian D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48 (3): 443–53. doi:10.1016/0022-2836(70)90057-4. PMID 5420325
- 11. Rice P , Longden I , Bleasby A,(2000), EMBOSS: the European Molecular Biology Open Software Suite. Trends in genetics : TIG 16 (6) :276-7PMID: 10827456
- Russell RJ, Kerry PS, Stevens DJ, Steinhauer DA, Martin SR, Gamblin SJ, Skehel JJ. (2008). Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. Proc Natl AcadSci 105(46):17736-41
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG,(2011), Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology 7 :539 doi: 10.1038/msb.2011.75, PMID: 21988835
- 14. Sui J, Hwang WC, Perez S, Wei G, Aird D, Chen LM, Santelli E, Stec B, Cadwell G, Ali M, Wan H, Murakami A, Yammanuru A, Han T, Cox NJ, Bankston LA, Donis RO, Liddington RC, Marasco WA (March 2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat. Struct. Mol. Biol. 16 (3): 265–73. doi:10.1038/nsmb.1566. PMC 2692245. PMID 19234466
- 15. Throsby M, van den Brink E, Jongeneelen M, Poon LL, Alard P, Cornelissen L, Bakker A, Cox F, van Deventer E, Guan Y, Cinatl J, terMeulen J, Lasters I, Carsetti R, Peiris M, de Kruif J, Goudsmit J (2008). Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. PLoS ONE 3 (12): e3942. doi:10.1371/journal.pone.0003942. PMC 2596486. PMID 19079604
- 16. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, Yang H, Chen X, Recuenco S, Gomez J, Chen LM, Johnson A, Tao Y, Dreyfus C, Yu W, McBride R, Carney PJ, Gilbert AT, Chang J, Guo Z, Davis CT, Paulson JC, Stevens J, Rupprecht CE, Holmes EC, Wilson IA, Donis RO (October 2013). New World Bats Harbor Diverse Influenza A Viruses. .PLoS Pathogens 9 (10): e1003657. doi:10.1371/journal.ppat.1003657. PMC 3794996.PMID 24130481.
- 17. W. Li and A. Godzik, (2006), Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics*, vol. 22, pp. 1658-1659.
- White JM, Hoffman LR, Arevalo JH, et al. (1997). Attachment and entry of influenza virus into host cells. Pivotal roles of hemagglutinin. In Chiu W, Burnett RM, Garcea RL. Structural Biology of Viruses. Oxford University Press. pp. 80–104
- Willy Min Jou, Martine Verhoeyen, René Devos, Eric Saman, Rongxiang Fang, Danny Huylebroeck, Walter Fiers (1980). Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. Cell 19 (3): 683-96. doi: 10.1016/S0092-8674(80)80045-6. PMID: 6153930