ISSN 0974-1453
Research Article

MODELING OF CANINE DISTEMPER VIRUS PROTEINS USING SWISS-MODEL

Manjusha Tyagi[1], Chakra Pal Singh[2] and Anant Rai[2]*

[1]Sri Guru Ram Rai University, Dehradun, Uttarakhand, India.

[2]Institute of Biotechnology & IT, Mudia Ahmadnagar, Bareilly-243122, UP, India.

***Corresponding author: raia48@gmail.com**

Received: 12-11-2019, Accepted: 30-11-2019

ABSTRACT

Proteins of canine distemper virus-nucleocapsid (N), phosphoprotein (P), matrix protein (M), fusion protein (F), hemagglutinin (H), large polymerase protein (L), were modelled using SWISS-MODEL available at https://swissmodel.expasy.org. It yielded valuable structural details of the viral proteins.

Keywords: CDV, modelling, protein, swiss-model, virus.

INTRODUCTION

Canine distemper virus (CDV) is an enveloped, single-stranded RNA virus within the genus Morbillivirus (family Paramyxoviridae) and is closely related to the viruses of measles and rinderpest. The genome consists of a single linear molecule of negative-sense, single-stranded RNA, ~16 kb in size (Fig.1). The structural details of the viral proteins are very important to understand the functions of these proteins. Modeling is the best way to achieve it.

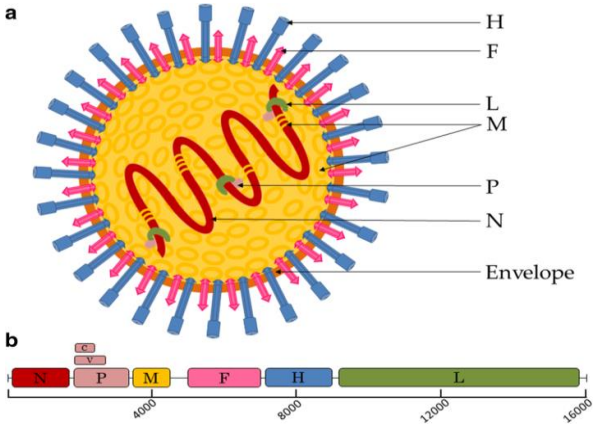

Fig 1. CDV virion and genome organization. N: nucleocapsid, P: phosphoprotein, M: matrix protein, F: fusion protein, H: hemagglutinin, L: large polymerase protein. b Map of genomic RNA 3′ to 5′ of CDV. (Source:

https://upload.wikimedia.org/wikipedia/commons/thumb/5/5e/12985_2019_1136_Fig1_HTML.webp/1200px-12985_2019_1136_Fig1_HTML.webp.png).

MATERIALS AND METHODS

Protein sequences

These were downloaded from PubMed and are shown in tables 1-7.

Table 1: Primary amino acid sequence of CDV H protein for which templates were searched and models were built.

MLSYQDKVGAFYKDTARANSSKLSL
VTEEQGGRRPPYLLFVLLILLIGILALL
AITGVRFHQVSTSNMEFSRLLKEDME
KSEAVHHQVIDVLTPLFKIIG
DEIGLRLPQKLNEIKQFILQKTNFFNP
NREFDFRDLHWCINPPSKIKVNFTNY
CDTVGVKNSIASAANPIILSALSGARG
DIFPPYRCSGATTSVGRVFP
LSVSLSMSLISRTSEIINMLTAISDGVY
GKTYLLVPDYIEGEFDSQKIRVFEIGFI
KRWLNNMPLLQTTNYMVLPETSKAK
VCTIAVGELTLASLCVGES
TVLLYHDSNGSQNGILVVTLGIFGATP
MDQVEEVIPIAHPSVERIHITNHRGFIK
DSVVTWMVPVLVSEKQREQKNCLES
ACQRKSYPMCNQTSWEPFGG
GQLPSYGRLTLPLDPSVDLQLNISFTY
GPVILNGDGMDYYESPLLESGWLTIP
PKNGTVLGLINKASRGDQFTVTPHVL
TFAPRESSGNCYLPIQTSQIM
DKDVLTESNLVVLPTQNFRYVIATYD
ISRGDHAIVYYVYDPIRTIFYTYPFRLT
TKGRPDFLRIECFVWDDDLWCHQFY
RFEANITNSTTSVENLVRIRSPCNGPN
P

Table 2: Primary amino acid sequence of CDV P protein for which templates were searched and models were built.

MAEEQAYHVSKGLECLKALRENPPDI
EEIQEVSSLRDQTCNPGQENGTTGMQ
EEEDSQNLDESHEPTKGSNYVGHVPQ
NNPGCGERNTALVEAERPPRED
IQPGPGIRCDHVYDHSGEEVKGIEDA
DSLVVPAGTVGNRGFERGEGSLDDST
EDSGEDYSEGNASSNWGYSFGLKPD
RAADVSMLMEEELSALLRTSRNV
GIQKRDGKTLQFPHNPEVRQGIRSVD
PLKRGTEERSVSHGMGIVAGSTSGAT
QSALKSTGGSSEPSVSAGNVRQPAMN
AKMTQKCKLESGTQLPPRTSNE
AESDSEYDDELFSEIQEIRSAITKLTED
NQAILTKLDTLLLLKGETDSIKKQISK
QNIAISTIEGHLSSIMIAIPGFGKDTGD
PTANVDINPELRPIIGR
DSGRPLAEVLKQPASSRGNRKDSGIT
LGSKGQLLRDLQLKPIDKESSSAIGYK
PKDTAPSKAVLASLIRSSRVDQSHKH
NMLALLKNIKGDDNLNEFYQMVKSI
THA

Table 3: Primary amino acid of CDV C protein sequence for which templates were searched and models were built.

MSAKGWNASKPSERILLTLRRFKRSA
ASETKPATQAKRMEPQACRKRRTLRI
SMNHTSQQKDQTMSAMYLKIIRDVE
NAILRLWRRSGPLERTSNQDLEY
DVIMFMITAVKRLRESKMLTVSWYL
QALSVIEDSREEKEALMIALRILAKIIP
KEMLHLTGDILSALNRTEQLM

Table 4: Primary amino acid sequence of CDV M protein for which templates were searched and models were built.

MTEVYDFDQSSWYTKGSLAPILPTTY
PDGRLIPQVRVIDPGLGDRKDECFMYI
FLMGIIEDNDGLGPPIGRTFGSLPLGV
GRTTARPEELLKEATLLDIM
VRRTAGVKEQLVFYNNTPLHILTPWK
KVLTSGSVFSANQVCNTVNLIPLDIAQ
RFRVVYMSITRLSDDGSYRIPRGVFEF
RSRNALAFNILVTIRVEGDV
DSSRGNLGMFKDYQATFMVHIGNFS
RKKNQAYSADYCKLKIEKMGLVFAL

GGIGGTSLHIRCTGKMSKALNAQLGF
KKILCYPLMEINEDLNRFLWRSEC
KIVRIQAVLQPSVPQDFRVYNDVIISD
DQGLFKIL

Table 5: Primary amino acid sequence of CDV F protein for which templates were searched and models were built.

MHRGIPKSSKTQTHTQQDRPPQPSTE
LEETRTSRARHSTTSAQRSTHYDPRTS
DRPVSYTMNRTRSRKQTSHRLKNIPV
HGNHEATIQHIPESVSKGARS
QIERRQPNAINSGSHCTWLVLWCLG
MASLFLCSKAQIHWDNLSTIGIIGTDN
VHYKIMTRPSHQYLVIKLIPNASLIEN
CTKAELGEYEKLLNSVLEPIN
QALTLMTKNVKPLQSLGSGRRQRRF
AGVVLAGVALGVATAAQITAGIALH
QSNLNAQAIQSLRTSLEQSNKAIEEIR
EATQETVIAVQGVQDYVNNELVP
AMQHMSCELVGQRLGLRLLRYYTEL
LSIFGPSLRDPISAEISIQALIYALGGEI
HKILEKLGYSGSDMIAILESRGIKTKIT
HVDLPGKFIILSISYPTL
SEVKGVIVHRLEAVSYNIGSQEWYTT
VPRYIATNGYLISNFDESSCVFVSESAI
CSQNSLYPMSPLLQQCIRGDTSSCART
LVSGTMGNKFILSKGNIVA
NCASILCKCYSTSTIINQSPDKLLTFIA
SDTCPLVEIDGATIQVGGRQYPDMVY
EGKVALGPAISLDRLDVGTNLGNALK
KLDDAKVLIDSSNQILETVR
RSSFNFGSLLSVPILSCTALALLLLIYC
CKRRYQQTLKQHTKVDPAFKPDLTG
TSKSYVRSL

Table 6: Primary amino acid sequence of CDV N protein for which templates were searched and models were built.

MFKRTRDQPPLASGSGGAIRGIKHVII
VLIPGDSSIITRSRLLDRLVRLVGDPEI
NGPKLTGILISILSLFVESPGQLIQRIID
DPDVSIKLVEVIPSIN
SVCGLTFASRGASLDSEADEFFKIVDE
GSKAQGQLGWLENKDIVDIEVDDAE
QFNILLASILAQIWILLAKAVTAPDTA
ADSEMRRWIKYTQQRRVVGEF
RMNKIWLDIVRNRIAEDLSLRRFMVA
LILDIKRSPGNKPRIAEMICDIDNYIVE
AGLASFILTIKFGIETMYPALGLHEFS
GELTTIESLMMLYQQMGET
APYMVILENSVQNKFSAGSYPLLWSY
AMGVGVELENSMGGLNFGRSYFDPA
YFRLGQEMVRRSAGKVSSALAAELGI
TKEEAQLVSEIASKTTEDRTIRA
AGPKQSQITFLHSERSDVTNQQPPTIN
KRSENQGGDRYSIHFSDERFPGYTPD
VNSSEWSESRYDTQTTQDGGNEDDR
KSMEAIAKMRMLTKMLSQTGTS
EESSPVYNDRELLN

Table 7: Primary amino acid sequence of CDV large polymerase protein for which templates were searched and models were built.

MDSVSVNQILYPEVHLDSPIVTNKLV
AILEYARIRHNYRLLDTTLVRNIKERI
SEGLSNQMIINCIETGSIVNQTLLSYPK
HNHVIYPNCNKLLFHAQDR
VISLRLRNIFKRGNSIYSKITDGVKKC
LNDINLSIGLGGVLDKTIGAKVDEAGI
IMQSSQWFEPFLLWFTIKTEMRSVIKS
STHNCRKRRQNPVFVRGES
FNVLVSRDLVCIIDITSHNVYYLTFEM
VLMYCDVIEGRLMTDTAMAIDHRYS
TLHVRIRYLWDLIDGFFLDLGNSTYQ
LVALLEPLSLAYLQLKDITFSL
RGAFLSHCFAEIQEILQDNGFYTEETF
QTLTQALDFVFITEDIHITGEIFSFFRSF
GHPRLEAITAAENVRKHMNQPKVVS
YETMMKGHAIFCGIIINGY
RDRHGGTWPPMDLPVHASPIIRNAHA
SGEGITYSQCIENWKSFAGIRFKCFMP
LSLDSDLTMYLKDKALAALKKEWDS
VYPKEFLRYNPPRSTESRRLVN
VFLEDSQFDPYNMIMYVISGQYLDDP
DFNLSYSLKEKEIKEVGRLFAKMTYK
MRACQVIAENLISNGIGKYFKDNGMA
KDEHDLTKALHTLAVSGVPKDK

KDSHRGLTNQCKSKKPTPYRGALHS
VSSPSSRYMDPNPNFCTSRREDNDIEI
YETVSAFITTDLKKYCLNWRYETISIF
AQRLNEIYGLPSFFQWLHRRL
EQSILYVSDPHCPPDLDRHVDLNTAP
NSQIFIKYPMGGVEGYCQKLWTISTIP
YLYLAAHESGVRIASLVQGDNQTIAV
TKRVPSTWSYALKKAEASRVT
TEYFIALRQRLHDVGHHLKANETIISS
HFFVYSKGIYYDGMLISQSLKSIARCV
FWSETIVDETRAACSNISTTLAKAIEK
GFDRYLAYALNILKIIQQV
LISLGFTINSAMTRDVIEPLLQDHCLL
TKMAILPAPIGGLNYLNMSRLFVRNI
GDPVTSSIADLKRMIRSGLLGVEILHQ
VMTQYPGDSSYLDWASDPYS
ANLPCVQSITRLLKNITARHVLINSPN
PMLKGLFHDESQDEDEALAAFLMDR
KIIIPRAAHEILDNTITGAREAIAGMLD
TTKGLIRASMKRGGLTPRII
NRLSTYDYEQFRAGIRLLSGKGHDPLI
DQDSCSVQLARALRNHMWAKLAKG
RPIYGLEVPDILESMKGYMIRRHESCL
LCASGSHNYGWFFVPANCQLDS
ITEGTSALRVPYIGSTTEERTDMKLAF
VKSPSRSLKSAVRIATVYSWAYGDDD
ESWQEAWTLAKQRANISLEELRMITP
ISTSTNLAHRLRDKSTQVKYS
GTSLIRVARYATISNDNLSFVIADKKV
DTNFIYQQGMLLGLGILEHLFRLSSTT
GDTNTVLHLHVETDCCVIPMSDHPRV
PGLRKVVIPRNICTNPLIYD
SNPIIEKDAVRLYNQSHRKHIVEFVT
WTTGQLYHVLAKSTAMSMVEMITKF
EKDHLNEVSALIGDDDINSFITEFLLV
EPRLFTVYLGQCAAINWGFEIH
YHRPSGKYQMGELLFSFLSRMSKGVF
KILTNALSHPKVYRRFWDSGMIEPVH
GPSLDSQNLHITVCNLIYNCYMIYLDL
LLNDELDDFSFILCESDEDVI
PERFDNIQARHLCILSDLYCNPRDCPQ
IRGLTPTQKCAVLSRYLKSKALESHV
GLTWNDKPILIDQYSCSLTYLRRGSIK
QIRLRVDPGFITDAVGCLEK
RPLRKSPISNASELKSEFDPPKDDLVK
LLSQLSTRTHNLPITGLGVRNYEVHSF
RRIGINSTACYKAVEIVSVIKNEFTSEE
HGLFLGEGSGAMLTVYKE
LLRLSRCYYNSGVSAEARTGQREISP
YPSEVSLVEHQLGLDKLVTVLFNGRP
EVTWVGSVDCYKYILSQISASSLGLIH
SDIESLPDKDIIEKLEELSAI
LSMTLILGKVGSVLVIKIMPASGDWV
QGFILYALPHFLRSYIIYPRYSNFVSTE
AYLVFTGLRAGRLVNPEGIKQQILRV
GIRTSPGLVGHILSSKQTAC
VQSLHGPPFQAKSFNPYLQGLTSIEKI
LINCGLTINGLKVCKNLLHHDISSGEE
GLKGSITILYRELARFKDNHQFSHGM
FHAYPVLIASQERELVSIIA
RKYCGYILLYSGDLYEITRIVRDLKAN
HIIFDLHRNLFMNNLSRSDRSLILTTIP
KRNWLFQLETKEIKEWFKLLGYSALI
RNH

Template Search

Template search with BLAST and HHBlits was performed against the SWISS-MODEL template library (SMTL, last update: 2020-04-15, last included PDB release: 2020-04-10). The target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. An initial HHblits profile was built using the procedure outlined in (Camacho et al., 2009; Remmert et al., 2012), followed by 1 iteration of HHblits against NR20. The obtained profile was then searched against all profiles of the SMTL. For each identified template, the template's quality was predicted from features of the target-template alignment. The templates with the highest quality were then selected for model building.

Model Building

Models were built based on the target-template alignment using ProMod3.

Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodelled using a fragment library. Sidechains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod3 fails, an alternative model is built with PROMOD-II (Guex et al., 2009; Waterhouse et al, 2018; Bienert et al, 2017). The global and per-residue model quality has been assessed using the QMEAN scoring function (Studer et al., 2020).

Ligand Modelling

Ligands present in the template structure are transferred by homology to the model when the following criteria are met: (a) The ligands are annotated as biologically relevant in the template library, (b) the ligand is in contact with the model, (c) the ligand is not clashing with the protein, (d) the residues in contact with the ligand are conserved between the target and the template. If any of these four criteria is not satisfied, a certain ligand will not be included in the model.

Oligomeric State Conservation

The quaternary structure annotation of the template is used to model the target sequence in its oligomeric form. The method (Bertoni et al., 2017) is based on a supervised machine learning algorithm, Support Vector Machines (SVM), which combines interface conservation, structural clustering, and other template features to provide a quaternary structure quality estimate (QSQE). The QSQE score is a number between 0 and 1, reflecting the expected accuracy of the interchain contacts for a model built based a given alignment and template. Higher numbers indicate higher reliability. This complements the GMQE score which estimates the accuracy of the tertiary structure of the resulting model.

RESULTS AND DISCUSSION

The CDV H protein was modelled using 22 templates and 4 models were generated. Model 01 showed 33.41% sequence identity with the template 2zb5.1.A (Fig. 2), while other models showed about 13.52 % seq identity with template. The P protein utilised 37 templates producing 5 models, model 01 Phosphoprotein, template 4c5q.1.A, showed seq identity 68.06% (Fig 3), model 03 RNA polymerase alpha subunit, template 1oks.1.A showed seq Identity 40 % (Fig. 4),model 05 Nucleoprotein, Phosphoprotein, template 5e4v.1.A showed seq Identity 46.81% (Fig. 5). CDV C protein using 15 templates produced 2 models, Model 01 cytoskeleton-associated protein 5, template 4qmi.1.A, seq identity 15.79% (Fig. 6), model 02 Bergeyella zoohelcum Cas13b (R1177A) mutant, template 6aay.1.A, seq identity 20.0% (Fig.7). Matrix protein, using 18 templates produced 2 models, model 01 hendra virus matrix protein, template 6bk6.1.A, seq identity 44.68% (Fig. 8), model 02 matrix protein, template 4g1g.1.A, seq identity 20.2% (Fig. 9). Fusion protein using 50 templates produced 3 models, model 01 fusion glycoprotein F0, template 5evm.1.A, seq identity 35.49% (Fig.11), model 02 fusion glycoprotein, template 1ztm.1.A, seq identity 27.75% (Fig 10), model 03 fusion glycoprotein F0, template 5ude.1.A, seq identity 16.4% (Fig. 12). N protein using 21 templates produced 1 model, model 01 nucleoprotein, phosphoprotein, template 5e4v.1.A seq

identity 81.75% (Fig.13). Large polymerase protein using 50 templates produced 6 models, model 01 RNA-directed RNA polymerase L, template 6v85.1.A, seq identity 29.77% (Fig. 14), model 02 large structural protein, template 6ueb.1.A, seq identity 17.37% (Fig.15), others are not shown here because of less seq identity.

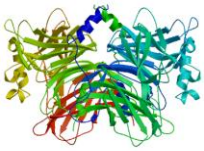

Fig. 2. H protein.

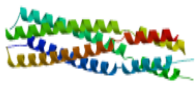

Fig. 3. P protein.

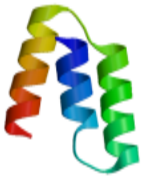

Fig. 4. P protein.

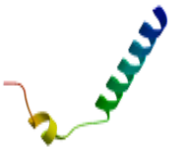

Fig. 5. Nucleoprotein Phosphoprotein

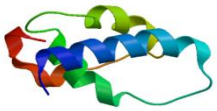

Fig. 6. C protein

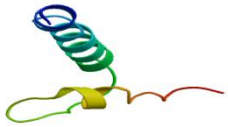

Fig. 7. C protein

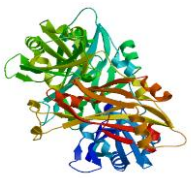

Fig. 8. M protein

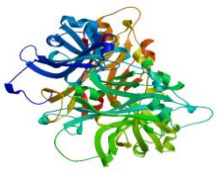

Fig. 9. M protein

Fig. 10. F protein

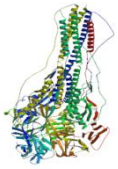

Fig 11. F protein

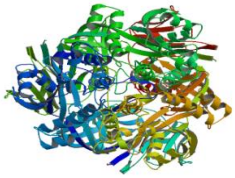

Fig. 12 F protein

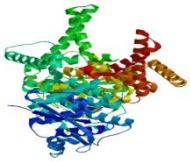

Fig. 13. N protein

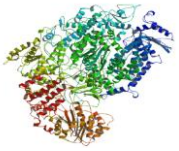

Fig.14.Large polymerase protein

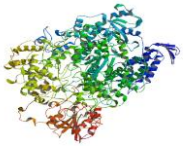

Fig. 15. Large structural protein

CONCLUSION

The viral proteins of CDV were modelled successfully using swiss-model software revealing their structural details. These structures show the properties of the proteins and their combining sites.

REFERENCES

Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T(2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T (2017). The SWISS-MODEL Repository - new features

and functionality. Nucleic Acids Res. 45:D313-D319.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009). BLAST+: architecture and applications. BMC Bioinformatics 10: 421-430.

Guex N, Peitsch MC, Schwede T (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30: S162-S173.

Remmert M, Biegert A, Hauser A, Söding J (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173-175.

Studer G, Rempfer C, Waterhouse AM, Gumienny G, Haas J, Schwede T (2020). QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics 36: 1765-1771.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1): W296-W303.