Biologia
https://doi.org/10.2478/s11756-020-00665-x

ORIGINAL ARTICLE

In silico tertiary structure prediction and evolutionary analysis of two DNA-binding proteins (DBP-1 and DBP-2) from *Hyposidra talaca* nucleopolyhedrovirus (HytaNPV)

Bhabesh Deka[1] · **Chittaranjan Baruah**[2] · **Manash Barthakur**[3]

Received: 15 June 2020 / Accepted: 2 December 2020

Abstract

Hyposidra talaca is a vicious pest of tea plants in the Eastern Himalayan's Darjeeling foothills and NE India. The infestation of this pest leads to crop loss, as early instars prefer to feed the young harvestable leaves and late instars feed the matured leaves, which leads to loss of photosynthetic capacity of the entire tea bush. Among the few accessible methods to control *H. talaca* is the baculovirus *H. talaca* nucleopolyhedrovirus (HytaNPV). DNA-binding protein (DBP) plays a significant role in HytaNPV during viral replication and transcription. The present study attempted to predict the structure and the functional analysis of two crucial DNA-binding proteins (DBP-1 and DBP-2) in the absence of experimental structures. Analysis of sequence, prediction of structure, functional characterization, and evolutionary analysis based on UniProtKB studied the amino acid sequences of DBP-1 and DBP-2 proteins. Modeling of these two proteins was presented using ab-initio modeling. QMEANDisCo 4.0.0 global and local per-residue quality estimates verified the structure as high quality. Phylogenetic analysis of both HytaNPV DBP-1 and DBP-2 proteins revealed a close evolutionary relationship with *Buzura suppressaria* nucleopolyhedrovirus. Tunnel analysis revealed multiple tunnels in DBP-1 (six) and DBP-2 (eleven), indicating a large number of transport pathways for small ligands that influence their reactivity. The theoretical structures and statistical verifications were successfully deposited in the Model Archive. They will be useful for advanced computational analysis of each protein's interactions for detailed functional analysis and understanding of viral pathogenesis in the absence of a complete experimental structure.

Keywords *Hyposidra talaca* · Baculovirus · HytaNPV · Ab-initio modeling · DNA-binding proteins · Evolutionary relationship

Abbreviations

NPV	nucleopolyhedrovirus
HytaNPV	*Hyposidra talaca* nucleopolyhedrovirus
Bp	base pair
DBP	DNA-binding protein
BmNPV	*Bombyx mori* nucleopolyhedrovirus
DBP-1	DNA-binding protein-1
DBP-2	DNA-binding protein-2

✉ Bhabesh Deka
bhabesh.deka@gmail.com

[1] North Bengal Regional Research and Development Centre, Nagrakata, West Bengal 735225, India

[2] Bioinformatics Laboratory (DBT-Star College), Postgraduate Department of Zoology, Darrang College (Gauhati University), Tezpur, Assam 784001, India

[3] Department of Zoology, Pub Kamrup College, Baihata Chariali, Assam 781381, India

Introduction

Hyposidra talaca (Lepidoptera: Geometridae), commonly known as the black inch looper, is considered a major tea pest, which causes considerable damage to tea plantations, leading to a loss in yield and the quality of manufactured tea as well. In general, chemical insecticides are being used for the control of tea pests, including this black inch looper by the tea planters (Hazarika et al. 2009). The use of chemical pesticides has many negative impacts, such as the development of resistance in insects, abolition of natural enemies, imbalance in natural ecosystems, and increasing environmental contamination besides affecting human health. Therefore, non-chemical control measures are extremely important for managing destructive pests (Deka et al. 2017; Nguyen et al. 2018). Bio-control agents play a vital role in managing pests, especially lepidopteran pests with well-balanced ecological systems, by helping in the reduced use of pesticides in major agricultural crops, including tea. The nucleopolyhedrovirus (NPV) is a baculovirus that infects insects, primarily moths, and butterflies and has a high demand as a

pesticide. A baculovirus of *H. talaca* nucleopolyhedrovirus (HytaNPV), which is reported to be an extremely specific and successful bio-control agent (Mukhopadhyay et al. 2010, 2011; Antony et al. 2011, 2012; Sinu et al. 2011; Ghosh et al. 2015; Nguyen et al. 2018). HytaNPV is a double-stranded circular DNA genome that consists of 139,089 bp and 39.6% GC content. It encodes141 open reading frames, including 37 baculovirus core genes, 7 genes unique to the HytaNPV genome, 25 genes conserved among lepidopteran baculoviruses 72 genes known in baculovirus. It is a group II alphabaculovirus that codes for F protein and lacks the gp64 genes present in group I alphabaculovirus viruses (Nguyen et al. 2018).

Although there is an accessibility of sequence information for a few proteins of HytaNPV, the molecular mechanism and biochemistry of their functions are still not clearly understood due to the absence of their specific structural information. For viral replication and transcription, a DNA-binding protein (DBP) is required in the NPV (Zhao et al. 2016). DBP has a major role in DNA replication in HytaNPV, and the viral particles are multiplied inside the host insect's alimentary canal. Even a single molecule can cause the death of *H. talaca* larvae after the multiplication of the viral particle. DBP is a conserved gene that is widely present in the baculovirus genome and is considered to be a core gene of the virus (Zhau et al. 2016). Most of the NPVs, including *Bombyx mori* nucleopolyhedrovirus (BmNPV), DBP is not only essential for early viral replication but also a viral gene that has a significant impact on transcription and expression during all periods of the baculovirus life cycle (Zhau et al. 2016).

Earlier studies reported that the HytaNPV genome consists of 37 core genes involved in replication, transcription, oral infectivity, viral assembly, packaging, and host protein interactions (Nguyen et al. 2018). Among the 37 core genes, 22 genes were identified to be associated with DNA replication. These genes included DNA polymerase (*DNA-pol* (*orf65*; 1035aa)), alkaline exonuclease (alk-exo (*orf127*; 400aa)), DNA helicases (*helicase-2* (*orf80*; 1238aa) and *helicase-1* (*orf60*; 523aa), DNA-binding proteins (*DBP-1* (*orf13*; 265aa), *DBP-2* (*orf30*; 321aa), *lef-3* (*orf63*; 426aa) and *p6.9* (*orf85*; 76aa). Baculovirus genes can be broadly classified as early, late, and very late based on the timing of its expression during the lifecycle of the virus. In HytaNPV, DBPs are encoded by the early expressed genes, (*dbp-1* (*orf13*; 265aa), *dbp-2* (*orf30*; 321aa) as it uses the host RNA polymerase for is transcription. On the other hand, the proteins encoded by *lef-4* (*orf75*; 457aa), *lef-8* (*orf57*; 879aa), *lef-9* (*orf39*; 505aa) and *p47* (*orf29*; 393aa) form the virally encoded RNA polymerase subunits, which are referred to as late expressed genes (Nguyen et al. 2018) (Fig. 1). Although the database search of the experimental protein structures in RCSB-PDB PDB has retrieved 26 entries of nucleopolyhedrovirus, most of which are from *Autographa californica* nucleopolyhedrovirus

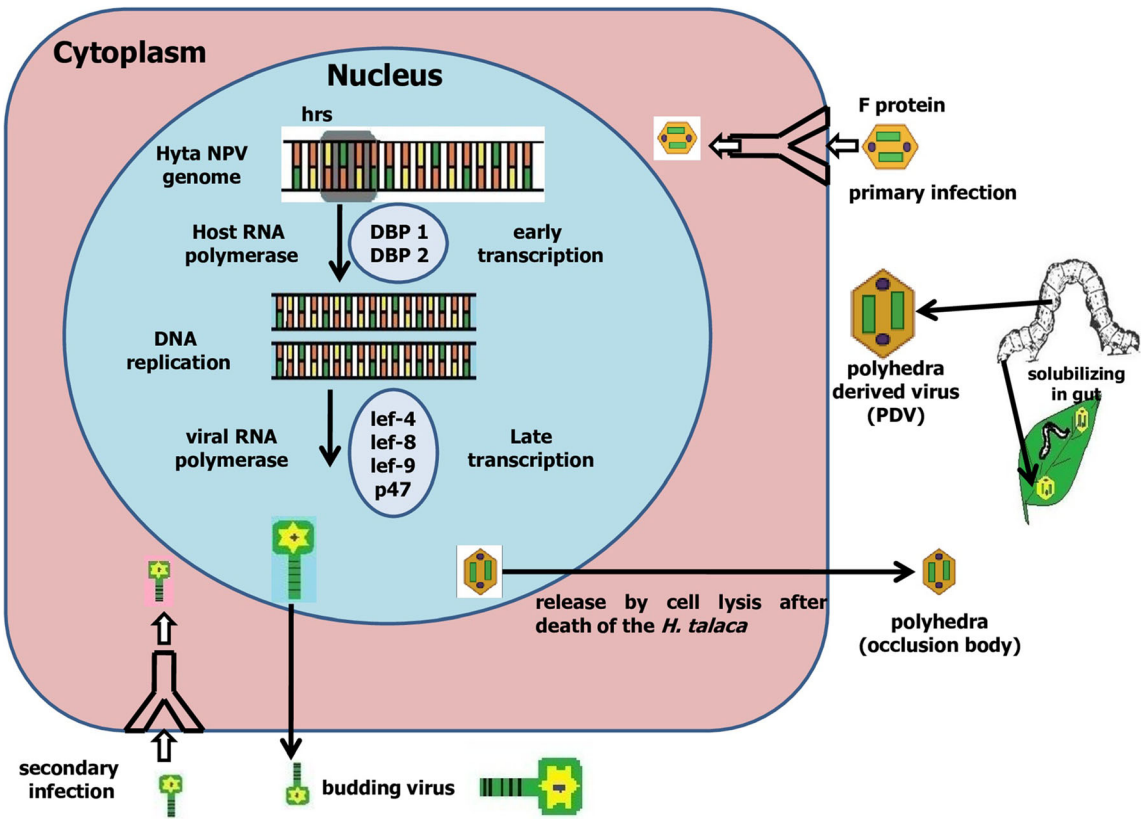


Fig. 1 Life-cycle of HytaNPV. Upon ingestion, the polyhedral are lysed inside the gut of the *H. talaca* larvae to release PDVs. Through primary infection, once the PDVs enter the nucleus, the viral DNA undergoes replication and transcription, generated budding virus (BV), which causes secondary infection. (Source: Li and Blissard 2011; Saxena et al. 2018; Sobhy 2017; Yin et al. 2013)

Table 1 The genome location of the genes encoding the b target proteins DBP-1 and DBP 2 of HytaNPV (GenBank: MH261376)

Protein name and UniProt Accession number	Length (aa)	Encoding gene, EMBL Accession no and Length	Genome Location & Locus Tag
DBP-1 (A0A2Z4HHX3)	265 aa	DBP-1 (AWW14373.1; length: 798 bp)	Complement (12,370..13167) (HytaNPV_gp013)
DBP-2 (A0A2Z4HHX6)	321 aa	DBP-2 (AWW14390.1 length: 966 bp)	Complement (31,566..32531) (HytaNPV_gp030)

(24 nos. of PDB entry), *Bombyx mori* nucleopolyhedrovirus (1 no. of PDB entry), *Wiseana signata* nucleopolyhedrovirus (1 no. of PDB entry), there is no structural information of any protein from *Hyposidra talaca* nucleopolyhedrovirus (HytaNPV). Therefore, the present study extended to amino acid sequence analyses, 3D structure prediction, and phylogenetic analysis of two impotent HytaNPV viral proteins, namely, DNA-binding protein-1 (DBP-1) and DNA-binding protein-2 (DBP-2). Proper characterization of these two proteins, encoded by DBP-*1* (*orf13*; 265aa) and DBP-*2* (*orf30*; 321aa) genes (Table 1; Fig. 2a and b), will help to understand their roles in DNA replication or encapsulation during the viral life cycle. The theoretical structures and statistical verifications will be useful for advanced computational analysis of each protein's interactions for detailed functional analysis and understanding of viral pathogenesis in the absence of a complete experimental structure.

Materials and methods

Acquisition and analysis of sequences

The translated amino acid sequence of HytaNPV, from the GenBank accession number MH261376, accessible as UniProtKB reviewed the amino acid sequences of DBP-1 (accession no. A0A2Z4HHX3), and DBP-2 (accession no. A0A2Z4HHX6) was used in the present study. The amino acid sequences of DBPfor different taxa were downloaded from UniProtKB for phylogenetic analysis based on BLASTp (Altschul et al. 1997) and FASTA hits (Pearson 1991).

Ab-initio tertiary structure prediction

BlastP and FASTA searches were performed separately with PDB to know the existing structure from the PDB and to decide modeling requirements (Table 1) (Fig. 2a and b). The output of the BLAST results analyzed the e-value generated by the BLAST family of search algorithm and query coverage. Ab-initio modeling was done in Baker Rosetta Server (https://robetta.bakerlab.org/). The loop regions were modeled using the de novo modeling procedure of Modeller and ModLoop (Fiser et al. 2000). The final optimized 3D structures with absolute coordinates were obtained by optimizing the molecular probability density functions of Modeller 9.24 (Webb and Sali 2016). The structures were verified by using global and local (per-residue) quality; an estimate of QMEAND is Co 4.0.0 (Haas et al. 2019) (supplementary Fig. S7 and S8). All the graphic representations of the three-dimensional structures were prepared using Chimera version 1.8.1 (Pettersen et al. 2004).

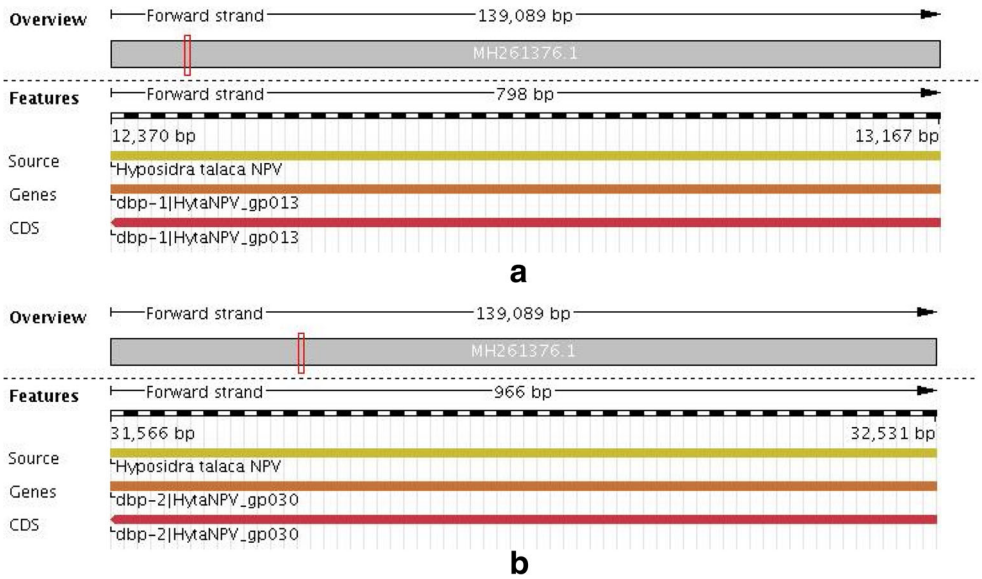


Fig. 2 Genome location of genes encoding two proteins of HytaNPV of present study. **a**: 798 bp DBP-1 gene (encoding DBP-1 protein) (Coding: AWW14373.1); **b**: 966 bp DBP-2 gene (encoding DBP-2 protein) (Coding: AWW14390.1) of HytaNPV

Proteomics analysis

Proteomics analyses were carried out using ExPASy proteomic tools (https://www.expasy.org/tools). The data mining and sequence analyses of the physicochemical parameters were computed using ProtParam (Gasteiger et al. 2005) and BioEdit (Hall 1999).

Sequence-based functional annotation was carried out in the sequence database using Pfam (*pfam.sanger.ac.uk/−*), PROSITE, PRINTS, InterProScan, and Hmmer version 3.3 (Finn et al. 2011). The MOLE 2.0 (Sehnal et al. 2013) and the Caver Web 1.0 (Stourac et al. 2019) were used for advanced analysis of biomacromolecular channels. The theoretical structures of the present study were used for tunnel analysis using Caver Web 1.0. Out of the top possible pockets with the highest relevance score (100%), largest volume ($Å^3$), and the highest estimated druggability value was considered for tunnel calculation. The tunnel bottleneck radius and length were calculated in Ångström (Å), and throughput (estimated tunnel importance) was calculated as e^{-cost}, where e is Euler's number. Throughput values range from 0 to 1; the higher the value, the greater the pathway's importance (Stourac et al. 2019).

Molecular phylogenetic analysis

The amino sequences of both the DBP-1 and DBP-2 genes used for phylogenetic analysis were aligned using ClustalW 1.6 (Thompson et al. 1994) integrated with MEGA X software (Kumar et al. 2018). The evolutionary history was deduced using maximum likelihood methods (Jones et al.1992). The total percentage of replicate trees in which the associated taxa were clustered together in the bootstrap tests (1000 replicates) (Felsenstein 2003). The initial tree(s) for the heuristic search were received automatically using the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances calculated using the JTT model, and then selecting the topology a higher log-likelihood value.

Results and discussion

Tertiary structures of DBP-1 and DBP-2

Although the database search of the experimental protein structures in RCSB-PDB retrieved 26 entries of NPV, most of which were from *Autographa californica* NPV (24 nos. of

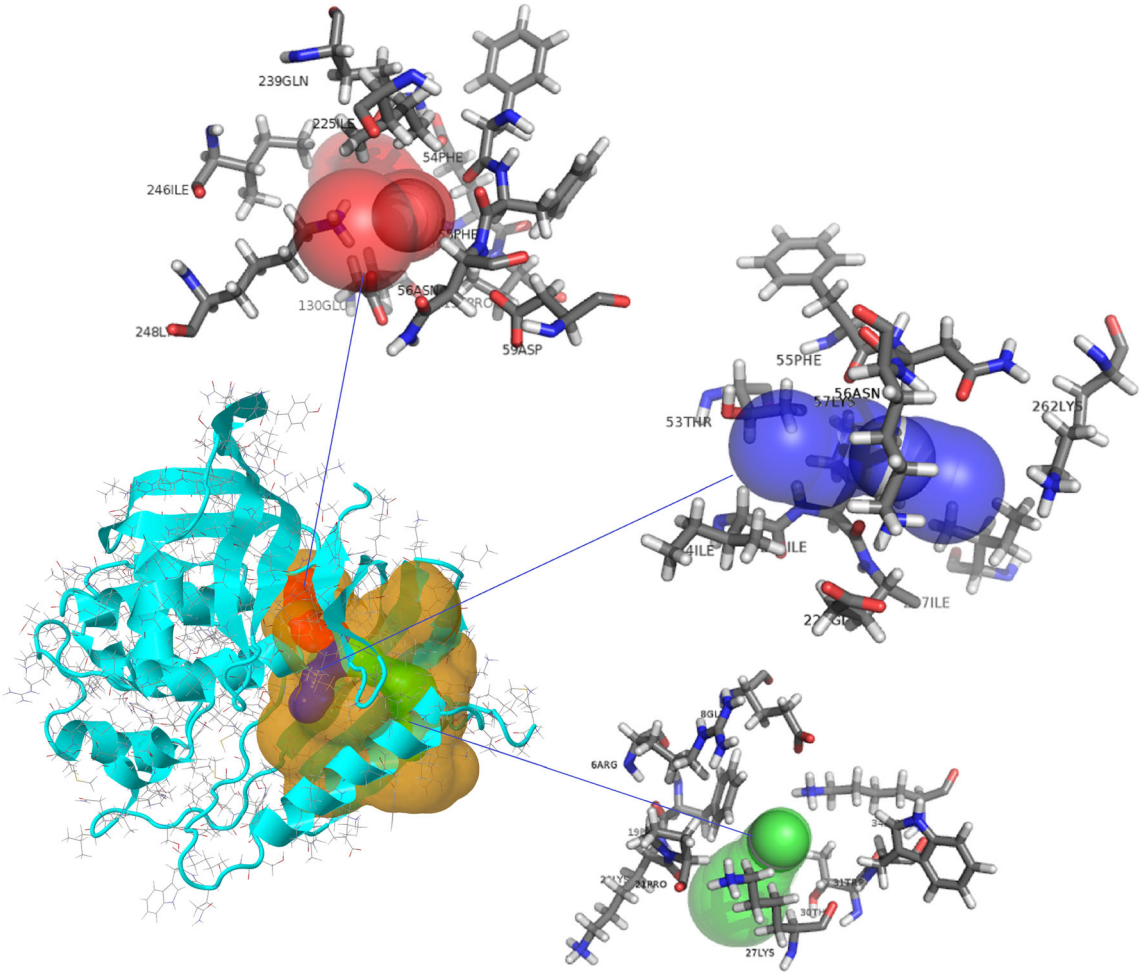


Fig. 3 Structure of DBP-1 protein of HytaNPV and its three high-throughput tunnels. Tunnels are colored on the basis of preferences in throughput values i.e. tunnel-1 (blue), and tunnel-2 (green). The high relevance pocket is shown (yellow)

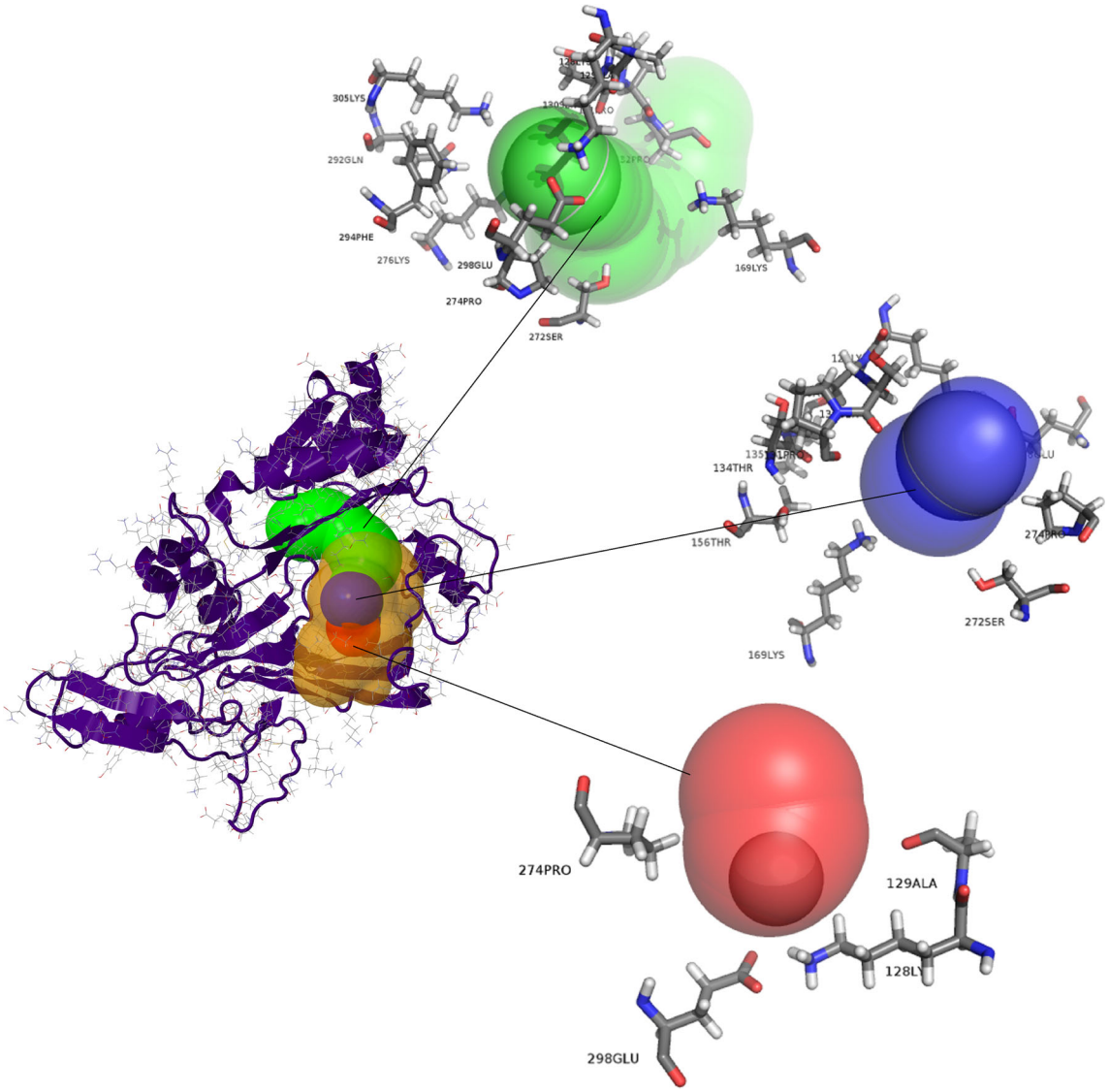


Fig. 4 Structure of DBP-2 protein of HytaNPV and its three high-throughput tunnels. Tunnels are colored on the basis of preferences in throughput values i.e. tunnel-1 (blue), and tunnel-2 (green). The high relevance pocket is shown (yellow)

PDB entry), *Bombyx mori* NPV (1 no. of PDB entry), and *Wiseana signata* NPV (1 no. of PDB entry). There is no structural information for any protein from HytaNPV. Therefore, the present study was designed to reveal the structural information of two important DBPs (DBP-1 and DBP-2) from HytaNPV.

The verified tertiary structure of DBP-1 had a qmean6 z score of −1.75 and 96.20% of amino acid residues in the favored region of the Ramachandran plot (Supplementary Table S1). The ProMotif analysis revealed that the structure of DBP-1 has 2 sheets, 1 beta alpha beta unit, 7 beta hairpins,4 beta bulges,11 strands, 10 helices,7 helix-helix interactions,19

Table 2 Physicochemical parameters of DBP-1 and DBP 2 of HytaNPV

Protein Name	MW (Da)	pI	Chemical Formula	Ext. coefficient Abs 0.1% (=1 g/l)	Instability index	Aliphatic index	GRAVY (Grand average of hydropathicity)
DBP-1 (A0A2Z4HHX3)	30,890.66	9.17	$C_{1379}H_{2185}N_{379}O_{398}S_{14}$	34,295	34.54	86.04	−0.383
DBP-2 (A0A2Z4HHX6)	37,010.48	6.11	$C_{1644}H_{2544}N_{428}O_{490}S_{27}$	33,975	54.28	73.15	−0.346

beta turns, and 7 gamma turns (Supplementary Fig. S1). Tunnel analysis has estimated 5 tunnels in the tertiary structure of DBP-1 (Fig. 3 and supplementary Fig. S3).

The verified tertiary structure of DBP-2 had a qmean6 z score of −1.86 and 96.55% of amino acid residues in the favored region of the Ramachandran plot (Supplementary Table S1). The ProMotif analysis revealed that the structure of DBP-2 has 5 sheets,8 beta hairpins, 3 beta bulges,13 strands,11 helices,2 helix-helix interactions,27 beta turns, and 6 gamma turns (Supplementary Fig. S2). Tunnel analysis estimated only one tunnel in the tertiary structure of DBP-2 (Fig. 4 and supplementary Fig. S4). After verification, the coordinate files of DBP-1 and DBP-2 were successfully deposited to Model Archive (https://www.modelarchive.org; Supplementary Table S1).

Protein tunnels are used to connect the functional hidden cavities with the immensity solvent and protein channels to enable transportation through biological membranes, and it characterizes the structural texture that directs the exchange rates of ions, water solvent, and ligands. These are present in an immense quantity and command on its function (Brezovsky et al. 2018). Of the top ten possible pockets of DBP-1, the pocket with relevance score (100%), volume 1911 $Å^3$, and estimated druggability of 0.49 were considered for tunnel analysis. Of the six observed tunnels, three high-throughput tunnels were analyzed. Tunnel 1(blue) with a

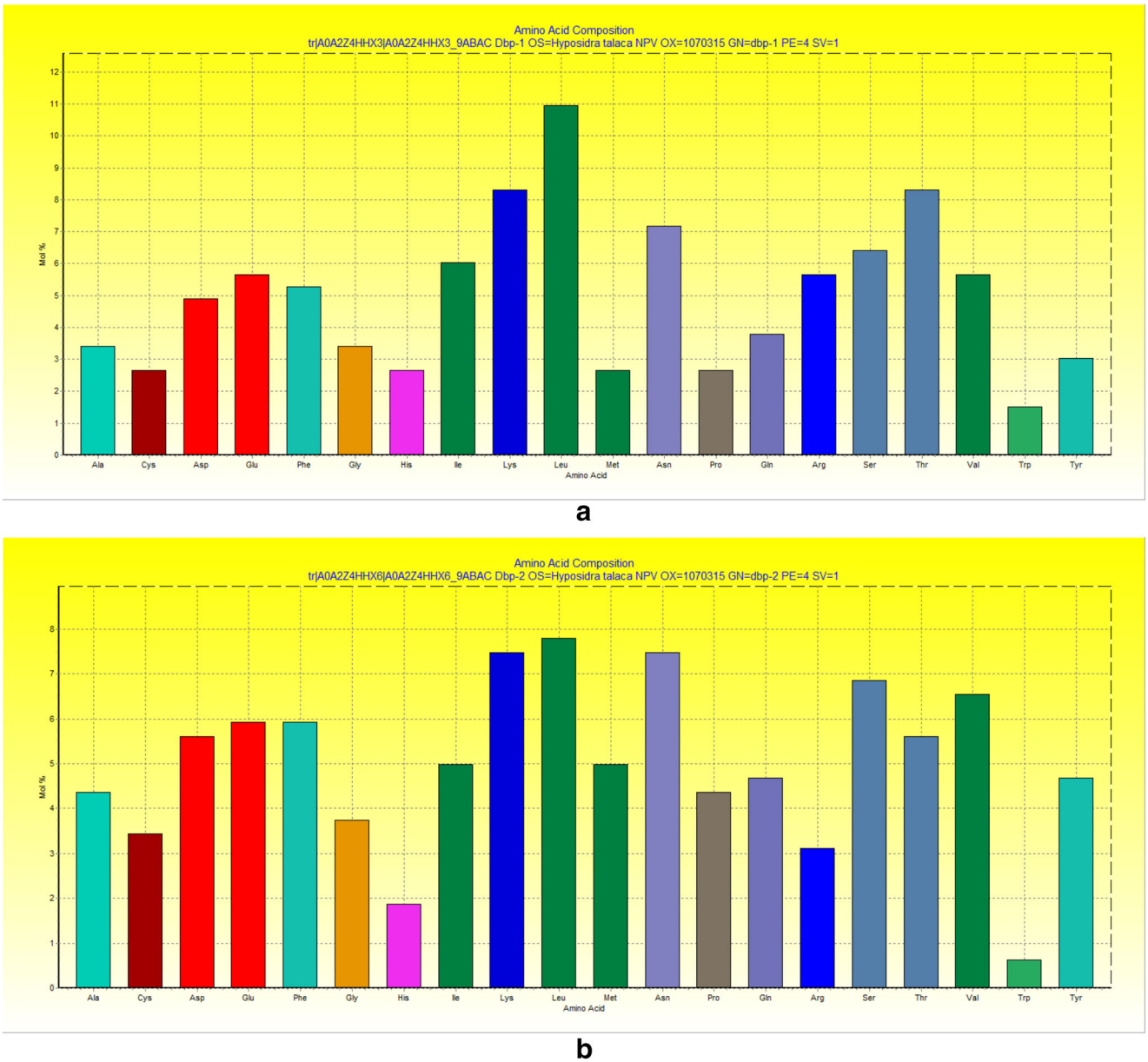


Fig. 5 Amino acid composition of two proteins of the present study from HytaNPV using BioEdit software. The colors represent the physicochemical property of each Amino Acid. **a**: DBP-1 protein and **b**: DBP-2 protein

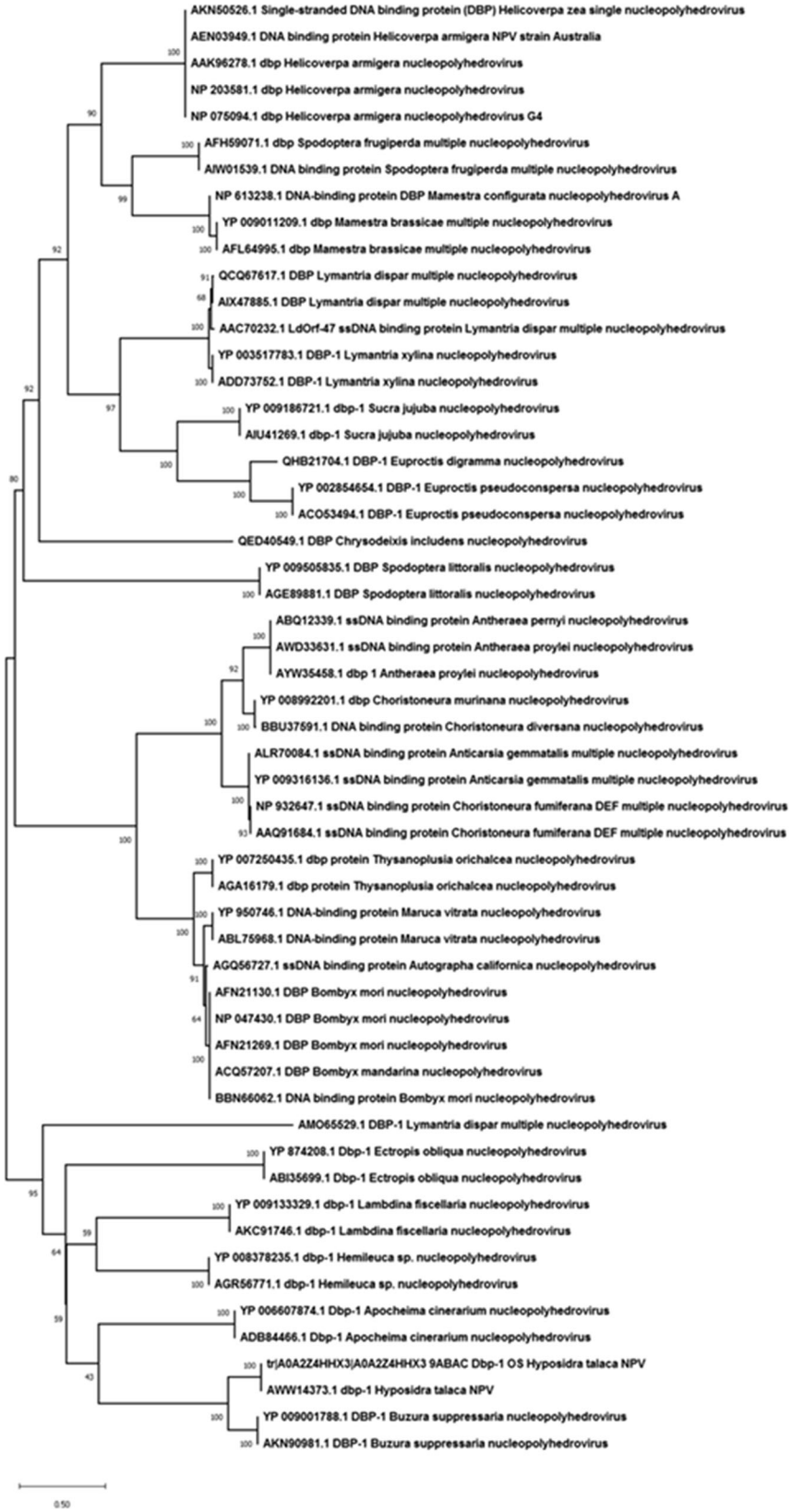
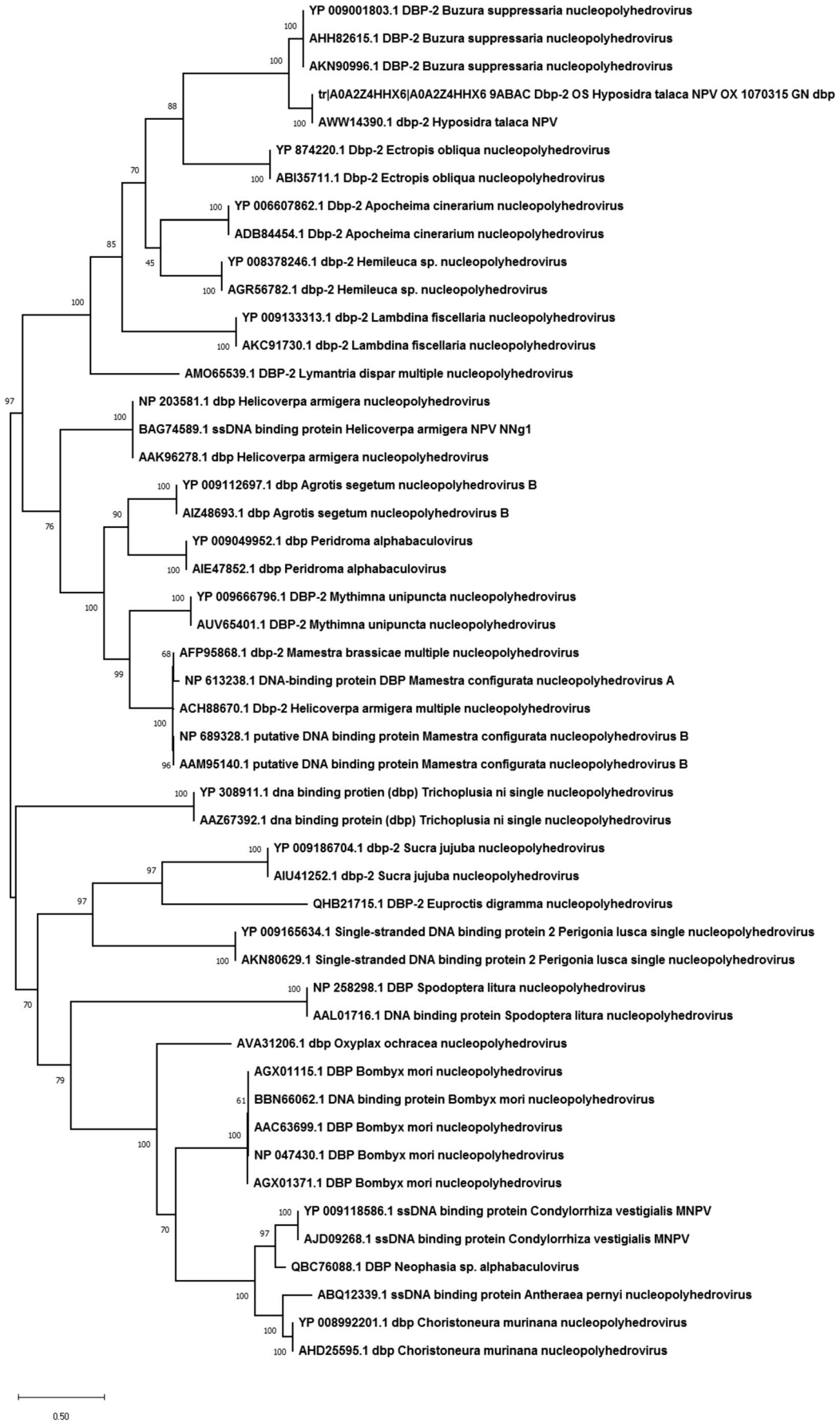


Fig. 6 Evolutionary analysis of DBP-1 by maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method and JTT matrix-based model (Jones et al. 1992). The tree with the maximum log likelihood (−13,075.55) is shown. This analysis implicated 55 amino acid sequences with 380 positions in the final dataset. Evolutionary analyses were carried out using MEGA X (Kumar et al. 2018)


YP 009001803.1 DBP-2 Buzura suppressaria nucleopolyhedrovirus
AHH82615.1 DBP-2 Buzura suppressaria nucleopolyhedrovirus
AKN90996.1 DBP-2 Buzura suppressaria nucleopolyhedrovirus
tr|A0A2Z4HHX6|A0A2Z4HHX6 9ABAC Dbp-2 OS Hyposidra talaca NPV OX 1070315 GN dbp
AWW14390.1 dbp-2 Hyposidra talaca NPV
YP 874220.1 Dbp-2 Ectropis obliqua nucleopolyhedrovirus
ABI35711.1 Dbp-2 Ectropis obliqua nucleopolyhedrovirus
YP 006607862.1 Dbp-2 Apocheima cinerarium nucleopolyhedrovirus
ADB84454.1 Dbp-2 Apocheima cinerarium nucleopolyhedrovirus
YP 008378246.1 dbp-2 Hemileuca sp. nucleopolyhedrovirus
AGR56782.1 dbp-2 Hemileuca sp. nucleopolyhedrovirus
YP 009133313.1 dbp-2 Lambdina fiscellaria nucleopolyhedrovirus
AKC91730.1 dbp-2 Lambdina fiscellaria nucleopolyhedrovirus
AMO65539.1 DBP-2 Lymantria dispar multiple nucleopolyhedrovirus
NP 203581.1 dbp Helicoverpa armigera nucleopolyhedrovirus
BAG74589.1 ssDNA binding protein Helicoverpa armigera NPV NNg1
AAK96278.1 dbp Helicoverpa armigera nucleopolyhedrovirus
YP 009112697.1 dbp Agrotis segetum nucleopolyhedrovirus B
AIZ48693.1 dbp Agrotis segetum nucleopolyhedrovirus B
YP 009049952.1 dbp Peridroma alphabaculovirus
AIE47852.1 dbp Peridroma alphabaculovirus
YP 009666796.1 DBP-2 Mythimna unipuncta nucleopolyhedrovirus
AUV65401.1 DBP-2 Mythimna unipuncta nucleopolyhedrovirus
AFP95868.1 dbp-2 Mamestra brassicae multiple nucleopolyhedrovirus
NP 613238.1 DNA-binding protein DBP Mamestra configurata nucleopolyhedrovirus A
ACH88670.1 Dbp-2 Helicoverpa armigera multiple nucleopolyhedrovirus
NP 689328.1 putative DNA binding protein Mamestra configurata nucleopolyhedrovirus B
AAM95140.1 putative DNA binding protein Mamestra configurata nucleopolyhedrovirus B
YP 308911.1 dna binding protien (dbp) Trichoplusia ni single nucleopolyhedrovirus
AAZ67392.1 dna binding protein (dbp) Trichoplusia ni single nucleopolyhedrovirus
YP 009186704.1 dbp-2 Sucra jujuba nucleopolyhedrovirus
AIU41252.1 dbp-2 Sucra jujuba nucleopolyhedrovirus
QHB21715.1 DBP-2 Euproctis digramma nucleopolyhedrovirus
YP 009165634.1 Single-stranded DNA binding protein 2 Perigonia lusca single nucleopolyhedrovirus
AKN80629.1 Single-stranded DNA binding protein 2 Perigonia lusca single nucleopolyhedrovirus
NP 258298.1 DBP Spodoptera litura nucleopolyhedrovirus
AAL01716.1 DNA binding protein Spodoptera litura nucleopolyhedrovirus
AVA31206.1 dbp Oxyplax ochracea nucleopolyhedrovirus
AGX01115.1 DBP Bombyx mori nucleopolyhedrovirus
BBN66062.1 DNA binding protein Bombyx mori nucleopolyhedrovirus
AAC63699.1 DBP Bombyx mori nucleopolyhedrovirus
NP 047430.1 DBP Bombyx mori nucleopolyhedrovirus
AGX01371.1 DBP Bombyx mori nucleopolyhedrovirus
YP 009118586.1 ssDNA binding protein Condylorrhiza vestigialis MNPV
AJD09268.1 ssDNA binding protein Condylorrhiza vestigialis MNPV
QBC76088.1 DBP Neophasia sp. alphabaculovirus
ABQ12339.1 ssDNA binding protein Antheraea pernyi nucleopolyhedrovirus
YP 008992201.1 dbp Choristoneura murinana nucleopolyhedrovirus
AHD25595.1 dbp Choristoneura murinana nucleopolyhedrovirus
0.50

◀ **Fig. 7** Evolutionary analysis of DBP 2 by maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method and JTT matrix-based model (Jones et al. 1992). The tree with the maximum log likelihood (−14,125.17) is revealed. This analysis implicated 49 amino acid sequences with 410 positions in the final dataset. Evolutionary analyses were carried out using MEGA X (Kumar et al. 2018)

bottleneck radius of 1.7 Å, length of 7.2 Å, distance to the surface of 6.6 Å, curvature of 1.1, the throughput of 0.78, and a number of residues 17. Tunnel 2 (green) with a bottleneck radius of 1.7 Å, length of 15.2 Å, distance to the surface of 12.2 Å, curvature of 1.2, the throughput of 0.68, and a number of residues 27. Tunnel 3 (red) with a bottleneck radius of 1.6 Å, length of 14.9 Å, distance to the surface of 12.6 Å, curvature of 1.2, the throughput of 0.68, and a number of residues 26 (Fig. 3).

Of the top ten possible pockets of DBP-2, the pocket with relevance score (100%), volume 1436 $Å^{3,}$ and highest estimated druggability of 0.70 was considered for tunnel analysis. Of the eleven observed tunnels, three high-throughput tunnels were analyzed. Tunnel 1(blue) with a bottleneck radius of 3.4 Å, length of 6.3 Å, distance to the surface of 5.7 Å, the curvature of 1.1, the throughput of 0.92, and a number of residues 22. Tunnel 2 (green) with a bottleneck radius of 3.5 Å, length of 23.0 Å, distance to the surface of 16.6 Å, curvature of 1.4, the throughput of 0.86, and a number of residues 41. Tunnel 3 (red) with a bottleneck radius of 2.0 Å, length of 7.3 Å, distance to the surface of 5.5 Å, curvature of 1.3, the throughput of 0.86, and a number of residues 18 (Fig. 4).

Tunnel analysis of the 3D structures of the present study has estimated the presence of multiple tunnels in DBP-1 and DBP-2. The presence of multiple tunnels in this so far uncharacterized protein may take a key role in a large number of transport pathways for small ligands influencing their reactivity. It has been experimentally demonstrated that the tunnels and their properties can define many important protein characteristics like substrate specificity, enantioselectivity, stability, and activity (Brezovsky et al. 2018).

Proteomics profiles of DBP-1 and DBP-2

DBP-1 protein of HytaNPV has MW = 30,889.12 Da and rich in amino acids leucine (10.94%), lysine (8.30%), and threonine (8.30%) (Fig. 5a).DBP-2 of HytaNPV (mw = 37,008.72 Da) is rich in leucine (7.79%), lysine (7.48%) and asparagine (7.48%) (Fig. 5b). Inter ProScan search results showed that both DBP-1 and DBP-2 are members of the ssDNA-binding protein baculovirus (IPR006871) family baculovirus ssDNA-binding proteins (*ssDNA-bd_baculovirus)*. HMMER (Biosequence analysis using profile hidden Markov Models) analysis of both proteins revealed that DBP-1 has a maximum hit match with Pfam ID PF04786.12 ssDNA-binding protein (aa 31–264) and DBP-2 has a maximum hit match with Pfam ID PF04786.12 ssDNA-binding protein (aa 71–320). Grand average hydropathicity values indicated that both proteins were hydrophilic (Table 2) (supplementary Fig. S5 and S6). The instability index value indicates that DBP-1 is a stable protein, but DBP-2 is an unstable protein (Table 2). The aliphatic index value indicates that DBP-1 has more thermostability than DBP-2 (Table 2).

Molecular phylogeny of DBP-1 and DBP-2

Evolutionary analysis of the DBP-1 and DBP-2 proteins of HytaNPV was based on the maximum likelihood (ML) method and JTT matrix-based model (Jones et al. 1992). The percentage of trees in which the associated taxa clustered simultaneously was revealed next to the branches. The ML phylogenetic tree based on the amino acid sequence of the DBP-1 protein revealed that it has the closest evolutionary relatedness with DBP-1 OS *H. talaca* NPV (UniProtKB accession numbers tr|AOA2Z4HHX3|AOA2Z4HHX3N9ABAC) along with the other species *Buzura suppressaria* NPV (UniProtKB accession numbers YP009001788.1 and AKN90981.1) that infect other tea pests, *B. suppressaria* and distantly related to *Apocheima cinerarium* NPV (UniProtKB accession number YP006607874.1 and ADB84466.1) (Fig. 6). However, the ML phylogenetic tree based on the amino acid sequence of the DBP-2 protein showed the closest evolutionary relationship with DBP-2OS *H. talaca* NPV OX 1070315 GN DBP-2 PE 4SV 1 (UniProtKB accession number tr|AOA2Z4HHX6|AOA2Z4HHX69ABAC) along with the other species *B. suppressaria* NPV (UniProtKB accession number YP00900103.1, AHH83615.1, and AKN90996) with 100% bootstrap support (Fig. 7).

The baculoviridae family is clustered into four different genera: alphabaculovirus (NPV isolated from Lepidoptera), betabaculovirus (granuloviruses isolated from Lepidoptera), gammabaculovirus (NPV isolated from Hymenoptera), and deltabaculovirus (NPV isolated from Diptera) (Jehle et al. 2006). Again, alphabaculoviruses are classified into two groups (I and II) based on the presence or absence of the gp64 gene, where the group I contains the gp64 gene and group II has a gene encoding fusion protein (F) (Hefferon et al. 1999; IJkel et al. 2000; Monsma et al. 1996; Pearson et al. 2000). Like *H. talaca, B. suppressaria* is also an insect pest of tea plants and some other host plants such as citrus, tung oil, and metasequoia plants. *B. suppressaria* NPV (BusuNPV) isolated from the NPV- infected dead *B. suppressaria* larva was used as an insecticide against this pest (Hu et al. 1993; Hu et al. 1998). Although the mechanism of baculovirus genome replication is not thoroughly understood, quite a few viral genes have been recognized as essential for DNA replication (Mikhailov 2003). Among them, BusuNPV encodes genes necessary for replication as well as DNA polymerase (Busu58), late expression

factor-1 (lef-1, Busu126), DNA helicase (Busu72), lef-2 (Busu104) and lef-3 (Busu56). Further genes associated with DNA replication comprise DBPs −1, 2 (DBP-1, Busu11, and DBP-2, Busu26) and very late factor-1 (vlf-1, Busu61) (Rohrmann 2011; Zhu et al. 2014). In lepidopteran baculoviruses, *DBP* is a conserved gene. Phylogenetic analysis showed that *DBP* duplicates of different baculoviruses might have evolved independently to the conserved *DBP* in alphabaculoviruses (Liu et al. 2014). In *Sucra jujube* NPV (SujuNPV), *DBP-2* appears to be closer to the *DBP* of betabaculovirus, where *DBP-1* (Suju30) *and DBP-2* (Suju13) encode 309 and 310 amino acid-long proteins, respectively, with 25% amino acid uniqueness (Liu et al. 2014). In *Bombyx mori* NPV (BmNPV), DBP is essential for virus replication (Mikhailov et al. 1998). Although the definite role of DBP in baculovirus replication remains unexplored, DNA binding proteins (DBP-1 and DBP-2) are required for normal levels of DNA synthesis or stability of nascent viral DNA as both are involved in the expression of many genes involved in growth, motility, metabolism, and virulence (Vanarsdall et al. 2007). Immunoelectron microscopic analysis of cells transfected with the DBP knockout revealed that DBP is required to produce normal-appearing nucleocapsids and the virogenic stroma generation (Vanarsdall et al. 2007). However, further studies of the properties and functions of DBP-1 and DBP-2 proteins are still needed to understand its involvement in virus pathogenesis. All contemporary studies on accessory proteins of NPV, including HytaNPV, recommend that they are crucial for virus replication (Zhu et al. 2014), and it is involved in the viral release, stability, and pathogenesis and finally contribute to virulence.

Baculoviruses are valuable natural control agents that consist of double-stranded DNA, and these viruses are highly specific and selective hosts (insect pests). Due to their narrow host range and slow killing speed, these viruses' utility is limited in agricultural applications. To accelerate the kill speed, baculoviruses have been genetically modified to express the foreign genes under powerful promoters (Inceoglu et al. 2001). Even though the recombinant viruses are synergized by synthetic pesticides, these viruses can reduce the synthetic pesticide load in the agriculture sector (Kamita et al. 2005). As the DBP is a core gene of the baculovirus, it is not only essential for early viral replication but also has a significant impact on transcription and expression during all periods of the baculovirus life cycle (Zhau et al. 2016). Prediction of the 3D structure of DBP proteins will help prepare genetically modified baculoviruses to accelerate the speed of kill to control of the *H. talaca,* the major pest of tea.

Conclusion

Hyposidra talaca NPV (HytaNPV), the biological control agent of a major tea pest consisting of 37 baculovirus core genes, 7 genes unique to the HytaNPV genome, 25 genes conserved among lepidopteran baculoviruses, and 72 genes known in baculovirus. The two DBPs (DBP-1 and DBP-2) of HytaNPV essential for replication and transcription are 798 bp and 966 bp in length, respectively. The present study reported theoretical modeling based on sequence and functional characterization based on the structure of two accessory proteins, DBP-1 and DBP-2. Phylogenetic analysis of both proteins revealed a close evolutionary relationship with a different species, *Buzura suppressaria* NPV, which infects *Buzura suppressaria,* another major tea pest. The presence of a large number of tunnels in the DBP-2 protein indicates its high reactivity. The theoretical structures and statistical verification reports were successfully deposited in the Model Archive. The theoretical structures would perhaps be useful for advanced computational analysis of each protein's interactions for detailed functional analysis and understanding of viral pathogenesis in the absence of a complete experimental structure.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.2478/s11756-020-00665-x.

Acknowledgments The authors are grateful to the Deputy Director (Research) North Bengal Regional R&D Centre, Nagrakata, West Bengal India and the Principal, Darrang College (Gauhati University), Tezpur (Assam), India, for supporting the research laboratory facility.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402. https://doi.org/10.1093/nar/25.17.3389

Antony B, Sinu PA, Das S (2011) New record of nucleopolyhedroviruses in tea looper caterpillars in India. J Invertebr Pathol 108:63–67. https://doi.org/10.1016/j.jip.2011.06.007

Antony B, Sinu PA, Rehman A (2012) Looper caterpillar invasion in north east Indian tea agro-ecosystem: change of weather and habitat loss may be possible causes? JTSR 2:1–5. https://doi.org/10.5376/jtsr.2012.02.0001

Brezovsky J, Kozlikova B, Damborsky J (2018) Computational analysis of protein tunnels and channels. In: Bornscheuer U, Höhne M (eds) Protein engineering. Methods in molecular biology, 1685. Humana Press, New York, pp 25–42. https://doi.org/10.1007/978-1-4939-7366-8_3

Deka B, Babu A, Sarkar S, Mandal S, Kundu N, Patanjali PK, Bhandari P (2017) Bioactivity of an organic synergist in enhancing the efficacy

of Neem kernel aqueous extracts (NKAE) and three recommended acaricides in controlling tea red spider mite (*Oligonychuscoffeae*). JTSR 7(4):40–45. https://doi.org/10.5376/jtsr.2017.07.0008
Felsenstein J (2003) Inferring phylogenies. Sinauer Associates, Sunderland, p 664
Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. R. D. Finn, J. Clements, S. R. Eddy. Nucleic Acids Res 39:29–37. https://doi.org/10.1093/nar/gkr367
Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9:1753–1773. https://doi.org/10.1110/ps.9.9.1753
Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM (ed) The Proteomics Protocols Handbook, Humana Press, pp 571–607
Ghosh B, Mukhopadhyay A, Das A, Bahadur M (2015) Restriction endonuclease fragment analysis of *Hyposidra talaca* nucleopolyhedrovirus genome. Int J Curr Res Acad Rev 3:81–87
Haas J, Gumienny R, Barbato A, Ackermann F, Tauriello G, Bertoni M, Studer G, Smolinski A, Schwede T (2019) Introducing "best single template" models as reference baseline for the continuous automated model evaluation (CAMEO). Proteins 87:1378–1387. https://doi.org/10.1002/prot.25815
Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
Hazarika LK, Bhyuan M, Hazarika BN (2009) Insect Pest of tea and their management. Annu Rev Entomol 54:267–284. https://doi.org/10.1146/annurev.ento.53.103106.093359
Hefferon KL, Oomens AG, Monsma SA, Finnerty CM, Blissard GW (1999) Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. Virology 258:455–468. https://doi.org/10.1006/viro.1999.9758
Hu ZH, Liu MF, Jin F, Wang ZX, Liu XY et al (1993) Nucleotide sequence of the *Buzura suppressaria* single nucleocapsid nuclear polyhedrosis virus polyhedrin gene. J Gen Virol 74:1617–1620. https://doi.org/10.1099/0022-1317-74-8-1617
Hu ZH, Arif BM, Jin F, Martens JW, Chen XW et al (1998) Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 79:2841–2851. https://doi.org/10.1099/0022-1317-79-11-2841
IJkel WF, Westenberg M, Goldbach RW, Blissard GW, Vlak JM et al (2000) A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. Virology 275:30–41. https://doi.org/10.1006/viro.2000.0483
Inceoglu AB, Kamita SG, Hinton AC, Huang Q, Severson TF, Kang K, Hammock BD (2001) Recombinant baculoviruses for insect control. Pest Manag Sci 57(10):981–987. https://doi.org/10.1002/ps.393
Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA et al (2006) On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol 151:1257–1266. https://doi.org/10.1007/s00705-006-0763-6
Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275
Kamita SG, Kang K-D, Hammock BD, Inceoglu AB (2005) Genetically Modified Baculoviruses for Pest Insect Control. Comprehensive Mol Insect Sci 6:271–322
Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. MBE 35:1547–1549. https://doi.org/10.1093/molbev/msy096
Kyte J, Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132. https://doi.org/10.1016/0022-2836(82)90515-0
Li Z, Blissard GW (2011) *Autographa californica* multiple nucleopolyhedrovirus GP64 protein: roles of histidine residues in triggering membrane fusion and fusion pore expansion. J Virol 85(23):12492–12504. https://doi.org/10.1128/JVI.05153-11
Liu X, Yin F, Zhu Z, Hou D, Wang J et al (2014) Genomic sequencing and analysis of *Sucra jujube* Nucleopolyhedrovirus. PLoSONE 9(10):e110023. https://doi.org/10.1371/journal.pone.0110023
Mikhailov V (2003) Replication of the baculovirus genome. Mol Biol 37:250–259
Mikhailov VS, Mikhailova AL, Iwanaga M, Gomi S, Maeda S (1998) *Bombyx mori*nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. J Virol 72(4):3107–3116. https://doi.org/10.1128/JVI.72.4.3107-3116.1998
Monsma SA, Oomens AG, Blissard GW (1996) The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. J Virol 70:4607–4616. https://doi.org/10.1128/JVI.70.7.4607-4616.1996
Mukhopadhyay A, De D, Khewa S (2010) Exploring the biocontrol potential of naturally occurring bacterial and viral entomopathogens of defoliating lepidopteran pests of tea plantations. J Biopestic 3:117–120
Mukhopadhyay A, Khewa S, De D (2011) Characteristics and virulence of nucleopolyhedrovirus isolated from *Hyposidra talaca* (Lepidoptera: Geometridae), a pest of tea in Darjeeling Terai, India. Int. J. Trop. Insect Sci 31:13–19. https://doi.org/10.1017/S1742758411000026
Nguyen TT, Suryamohan K, Kuriakose B, Janakiraman V, Reichelt M, Chaudhuri S, Guillory J, Divakaran N, Rabins PE, Goel R, Deka B, Sarkar S, Ekka P, Tsai Y, Vargas D, Santhosh S, Mohan S, Chin C, Korlach J, Thomas G, Babu A, Seshagiri S (2018) Comprehensive analysis of single molecule sequencing-derived complete genome and whole transcriptome of *Hyposidra talaca* nuclear polyhedrosis virus. Sci Rep (nature) 8:8924. https://doi.org/10.1038/s41598-018-27084-y
Pearson WR (1991) Searching protein sequence libraries: comparison of the sensitivity and selectivity of the smith-waterman and FASTA algorithms. Genomics 11(3):635–650. https://doi.org/10.1016/0888-7543(91)90071-l
Pearson MN, Groten C, Rohrmann GF (2000) Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. J Virol 74:6126–6131. https://doi.org/10.1128/jvi.74.13.6126-6131.2000
Pettersen E, Goddard T, Huang C, Couch G, Greenblatt D, Meng E, Ferrin T (2004) UCSF chimera–a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612. https://doi.org/10.1002/jcc.20084
Rohrmann GF (2011) Baculovirus molecular biology. 2nd edition, National Center for Biotechnology Information (US), Bethesda
Saxena A, Byram PK, Singh SK, Chakraborty J, Murhammer D, Giri L (2018) A structured review of baculovirus infection process: integration of mathematical models and biomolecular information on cell–virus interaction. J Gen Virol 99:1151–1171. https://doi.org/10.1099/jgv.0.001108
Sehnal D, Vařeková RS, Berka K, Pravda, Navrátilová V, Banáš P, Crina-Maria Ionescu CM, Otyepka M (2013) MOLE 2.0: advanced approach for analysis of biomacromolecular channels. J Cheminformatics 5 (39). https://doi.org/10.1186/1758-2946-5-39
Sinu PA, Mandal P, Antony B (2011) Range expansion of *Hyposidra talaca* (Geometridae: Lepidoptera), a major pest, to north eastern Indian tea plantations: change of weather and anti-predatory behaviour of the pest as possible causes. Int J Trop Insect Sci 31:242–248. https://doi.org/10.1017/S174275841100035X
Sobhy H (2017) A comparative review of viral entry and attachment during large and giant dsDNA virus infections. Arch Virol 162:3567–3585. https://doi.org/10.1007/s00705-017-3497-8
Stourac J, Vavra O, Kokkonen P, Filipovic J, Pinto G, Brezovsky J, Damborsky J, Bednar D (2019) Caver web 1.0: identification of

tunnels and channels in proteins and analysis of ligand transport. Nucleic Acids Res 47:W414–W422. https://doi.org/10.1093/nar/gkz378

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680. https://doi.org/10.1093/nar/22.22.4673

Vanarsdall AL, Mikhailov VS, Rohrmann GF (2007) Characterization of a baculovirus lacking the DBP (DNA-binding protein) gene. Virology 364(2):475–485. https://doi.org/10.1016/j.virol.2007.03.024

Webb B, Sali A (2016) Comparative protein structure modeling using Modeller. Current protocols in bioinformatics 54. John Wiley & Sons, Inc 5(6):1–5.6.37. https://doi.org/10.1002/cpbi.3

Yin F, Wang M, Tan Y, Deng F, Vlak JM, Hu Z, Wang H (2013) Betabaculovirus F proteins showed different efficiencies when rescuing the infectivity of gp64-null Autographa californica nucleopolyhedrovirus. Virology 436:59–66. https://doi.org/10.1016/j.virol.2012.10.017

Zhao C, Zhang C, Chen B, Shi Y, Quan Y, Nie Z et al (2016) A DNA binding protein is required for viral replication and transcription in *Bombyx mori* Nucleopolyhedrovirus. PLoSONE 11(7):e0159149. https://doi.org/10.1371/journal.pone.0159149

Zhu Z, Yin F, Liu X, Hou D, Wang J et al (2014) Genome sequence and analysis of *Buzura suppressaria* nucleopolyhedrovirus: a group II alphabaculovirus. PLoS One 9(1):e86450. https://doi.org/10.1371/journal.pone.0086450