Infection, Genetics and Evolution 98 (2022) 105219

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Cassava Brown Streak Viruses express second 6-kilodalton (6K2) protein with varied polarity and three dimensional (3D) structures: Basis for trait discrepancy between the virus species

Herieth Rhodes Mero [a,c,*], Sylvester Leonard Lyantagaye [b,1], Erik Bongcam-Rudloff [c,1]

[a] University of Dar es Salaam, Mkwawa University College of Education (MUCE), P. O. Box 2513, Iringa, Tanzania
[b] University of Dar es Salaam, Mbeya College of Health and Allied Science, P. O. BOX 608, Mbeya, Tanzania
[c] Swedish University of Agricultural Sciences (SLU), SLU-Global Bioinformatics Centre, Department of Animal Breeding and Genetics, P. 0. BOX 7054, 750 07 Uppsala, Sweden

ARTICLE INFO

Keywords:
Modelling
Homology
Ab initio (de novo)
Second 6-KiloDalton protein (6K2)
Homo-tetramer
Homo-trimer
Oligomerization
Cassava Brown Streak Viruses (CBSVs)

ABSTRACT

Cassava Brown Streak Virus (CBSV) and Ugandan Cassava Brown Streak Virus (UCBSV) are the two among six virus species speculated to cause the most catastrophic Brown Streak Disease of Cassava (CBSD) in Africa and Asia. Cassava Brown Streak Virus (CBSV) is hard to breed resistance for compared to Ugandan Cassava Brown Streak Virus (UCBSV) species. This is exemplified by incidences of CBSV species rather than UCBSV species in elite breeding line, KBH 2006/0026 at Bagamoyo, Tanzania. It is not yet understood as to why CBSV species could breakdown CBSD-resistance in the KBH 2006/0026 unlike the UCBSV species. This marks the first in silico study conducted to understand molecular basis for the trait discrepancy between CBSV and UCBSV species from structural biology view point. Following ab initio modelling and analysis of physical-chemical properties of second 6-kilodalton (6K2) protein encoded by CBSV and UCBSV species, using ROBETTA server and Protein Parameters tool, respectively we report that; three dimensional (3D) structures and polarity of the protein differs significantly between the two virus species. (95% and 5%) and (85% and 15%) strains of 20 CBSV and 20 UCBSV species respectively, expressed the protein in homo-trimeric and homo-tetrameric forms, correspondingly. 95% and 85% of studied strain population of the two virus species expressed hydrophilic and hydrophobic 6K2, respectively. Based on findings of the curent study, we hypothesize that; (i) The hydrophilic 6K2 expressed by the CBSV species, favour its faster systemic movement via vascular tissues of cassava host and hence result into higher tissue titres than the UCBSV species encoding hydrophobic form of the protein. t and (ii) The hydrophilic 6K2 expressed byCBSV species have additional interaction advantage with Nuclear Inclusion b protease domain (NIb) and Viral genome-linked protein (VPg), components of Virus Replication Complex (VRC) and hence contributing to faster replication of viral genome than the hydrophobic 6K2 expressed by the UCBSV species. Experimental studies are needed to resolve the 3D structures of the 6K2, VPg and NIb and comprehend complex molecular interactions between them. We suggest that, the 6K2 gene should be targeted for improvement of RNA interference (RNAi)-directed transgenesis of virus-resistant cassava as a more effective way to control the CBSD besides breeding.

1. Introduction

At least six virus species are suspected to cause the most debilitating brown streak disease of cassava (CBSD) with up to 100% yield loss, thus contributing to food insecurity and poverty in cassava producer countries (Mukiibi et al., 2019). Among the six species, two ipomoviruses *(Cassava Brown Streak Virus (CBSV)* and *Ugandan Cassava Brown Streak Virus (UCBSV))* of *Potyviridae* family and two ampeloviruses *(Manihot esculenta associated virus, type 1 (MEaV-1)* and *M. esculenta associated virus, type 2 (MEaV-2))* of *Closteroviridae* family were isolated

* Corresponding author at: University of Dar es Salaam, Mkwawa University College of Education (MUCE), Department of Biological Sciences, P. O. Box 2513, Iringa, Tanzania.
E-mail addresses: herieth.mero@slu.se, heriethmero@gmail.com (H.R. Mero).
[1] These authors contributed equally in producing this article.

https://doi.org/10.1016/j.meegid.2022.105219
Received 17 April 2021; Received in revised form 15 December 2021; Accepted 17 January 2022
Available online 20 January 2022

from cassava displaying foliar chlorotic and brown necrotic lesions (Kwibuka et al., 2021; Monger et al., 2001; Winter et al., 2010). The remaining two species predicted to have emerged from the UCBSV clade (Ndunguru et al., 2015), have remained unknown till to date and are yet to be isolated from either symptomatic or asymptomatic CBSD-cases of cassava.

The use of transgenic, virus-resistant cassava engineered by RNA interference (RNAi) technology, is a promising method applicable for controlling the CBSD in sub-Saharan Africa, besides breeding (Ogwok et al., 2012; Patil et al., 2011; Yadav et al., 2011).

Nonetheless, CBSV species is difficulty to breed resistance for compared to the UCBSV species (Alicai et al., 2016). For instance, based on field observation and routine diagnosis by reverse transcription polymerase chain reaction (RT-PCR) assay, the CBSV species was confirmed to cause symptomatic infection in a virus-resistant KBH 2006/0026 genotype unlike the UCBSV species. This indicates that, the former has succeeded to break virus-resistance of the KBH 2006/0026 whilst the latter has not. RNA viruses with a larger genetic landscape (i.e broader genetic diversity evidenced by greater amino acid usage, higher number of nonsynonymous mutations sites and faster evolution rate across entire genome) like the CBSV species, tend to adapt to host's genetic changes and overcome its resistance more easily than those with smaller genetic landscape like the UCBSV species (Alicai et al. 2016). However, no sufficient information is available on how disparity in amino acid usage between the CBSV and UCBSV species dictates their differed ability in breaking virus resistance in new breeding lines of cassava

A plethora of traits also differs between the two viruses whereby the CBSV species is more pathogenic, hyper-virulent and accumulate in tissue at higher titres, compared to the UCBSV species (Alicai et al., 2016; Mohammed et al., 2016; Ndunguru et al., 2015). Reason(s) for the trait difference between CBSV and UCBSV species are unknown. Only a single study by Alicai et al. (2016) has hypothesized on differed virulence levels between the two species. Alicai et al. (2016) stated that "possession of more genetic arsenals by CBSV species, is the reason why it is more virulent than UCBSV species". Nonetheless, no experimental reports exists to prove validity of the stated hypothesis (Mero et al., 2021). Other traits (tissue titers and pathogenicity) differing between the two viruses have remained uninvestigated.

For the past three decades, a lot of genomic data were generated in the process of scrutinizing genetic diversity and evolution of the CBSV and UCBSV species (Mbanzibwa et al., 2009b; Monger et al., 2010; Winter et al., 2010; Patil et al., 2011; Ndunguru et al., 2015; Alicai et al., 2016 and Amisse et al., 2019b). Equally, transcriptomic data were generated in scrutinizing mechanism used by distinct cassava genotypes to resist virus infections (Maruthi et al., 2014; Anjanappa et al., 2017; Amuge et al., 2017; Anjanappa et al., 2018).

Not all biological questions on trait inconsistency between CBSV and UCBSV species as well as host-and-pathogen interaction, can be answered through genomics and/or transcriptomics. In other words, sequences alone cannot describe how viral protein interact with the host inside the cell because it is the 3D protein structure with a given conformation that performs functions, rather than a primary sequence (Faure et al., 2019). The manner by which primary protein structures encoded by CBSV and UCBSV are folded post-translation, dictates how each of the two viruses interacts with the host's immunity factors. This in turn determine field performance of the host, resistance-breakdown being an example.

Considering the fact that, CBSVs express same type of genes but with differed percentage homology, we postulate that, mechanisms underlying trait difference between CBSV and UCBSV species are centered on discrepancy in either of the following: post-translational modification processes, subsequent folding of three-dimensional (3D) structures of protein and mechanisms by which the two viruses regulate their gene expression levels. Neither mechanisms underlying gene expression regulation nor post-translational modification (PTM) processes are clearly known for both CBSV and UCBSV species.

As for post-translational modification processing, we hypothesize that protein folding patterns differs between CBSV and UCBSV species due to variation in protein sequence and differential autocatalytic cleavage of a single polyproteins by virus-encoded serine protease (P1). Since P1 is the most genetically diverse protein of all (10) expressed by both CBSV and UCSBV species, we further question how does its enormous variation impacts trait discrepancy between and within strains of the two virus species.

With exception of the atypical putative nucleoside triphosphate pyrophosphatase (Ham 1) protein whose function was recently determined by complete gene deletion using recombinant DNA (rDNA) technology (Tomlinson et al., 2019), function(s) of 9 remaining proteins [Coat Protein (CP), Cylindrical Inclusion (CI), first Potyviral Protein (P1), third Potyviral Protein (P3), first 6-kDa protein (6K1), second 6-kDa protein (6K2), nuclear inclusion a protease domain (NIa), nuclear inclusion b protease domain (NIb) and Viral genome-linked protein (VPg)] encoded by CBSV and UCBSV species, were extrapolated basing on nucleotide sequence homology with few *Ipomovirus* species of the *Potyviridae* family (Monger et al., 2001; Mbanzibwa et al., 2009a, 2009b; Monger et al., 2010; Winter et al., 2010).

The present study argue that even though CBSV and UCBSV species show certain percentage of identity or similarity in nucleotide/amino acid sequences with other *Ipomovirus* species, that doesn't necessarily infer functional homology between them.

Knowledge on structural biology of proteins encoded by the CBSV and UCBSV species is missing as no sufficient investigations have been conducted to determine the 3D structures of viral proteins experimentally. This is evidenced by lack of publication of 3D structures of the 10 viral encoded proteins in highly recognizable database resources for experimentally determined protein structures such as Protein Databank (PDB) and Structural Classification of Proteins (SCOP) (Andreeva et al., 2014 and Andreeva et al., 2020).

That being the case, we suspect that some of the 10 ipomoviral proteins (P1, P3, 6K1, 6K2, CI, NIa, NIb, VPg, Ham 1 and CP) encoded by CBSV and UCBSV species, performs versatile functions. Nonetheless, versatile functions of those proteins are yet to be unraveled.

With reference to the work done by Monger et al. (2001) and Mbanzibwa et al. (2009b), only western blotting was used to isolate Coat Protein (CP) of CBSV species and determine its molecular weight. Another proteomic study by Qin et al. (2017) was conducted to assess Post harvest Physiological Deterioration (PPD) losses in cassava using two-dimensional electrophoresis (2-DE) in combination with Matrix-Assisted Laser Desorption Ionization tandem time of flight mass Spectrometer (MALDI-TOF-TOF-MS/MS).

Several reasons contribute to lack of 3D protein structures of CBSV and UCBSV in Protein Data Bank (PDB), one being high analysis cost by any of the following technologies (Cryo-Transmission Electron Microscopy (Cryo-TEM), Nuclear Magnetic Resonance (NMR), High-Resolution Atomic-Force Spectroscopy and X-ray Crystallography) in chemistry (Gotte and Libonati, 2014; Studer et al., 2019). Besides, experiments for determination of 3D protein structure being time

consuming, getting experimental samples with insufficient protein concentration owing to low tissue titers and un-equal virus distribution in the CBSD-cases of cassava, are other associated challenges to analysis of 3D- structures of CBSV and UCBSV species.

As an alternative to expensive methods for determining 3D structures of proteins said above, the present in silico study was conducted to address the following research questions related to unknown reasons for trait discrepancy between CBSV and UCBSV species by using bioinformatics tools and protein sequences published in Genbank: (i) Which protein (s) differ in 3D structures between CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains? (ii) What are the predicted versatile functions of 10 proteins (P1, P3, 6K1, 6K2, CI, NIa, NIb, VPg, Ham 1 and CP) among Cassava Brown Streak Viruses (CBSVs)? (iii) What is the difference in 3D structures of the Second 6-kilodalton (6K2) protein between and within additional 19 and 19 strains of CBSV and UCBSV species, respectively? (iv) What is the difference in physical-chemical properties of 10 protein(s) encoded by CBSV: TZ_MAF 49 versus UCBSV: TZ_MAF 58 strains? (v) What is the difference in physical-chemical properties of the Second 6-kilodalton (6K2) protein between additional 19 and 19 strains of CBSV and UCBSV species, respectively? (vi) Which mutations governs variation of protein (s) with remarkable discrepancy in 3D structures between additional 19 and 19 strains of CBSV and UCBSV species, respectively (vii) How are the Second 6-kilodalton (6K2) proteins of 20 strains of CBSV evolutionary distant from 20 strains of UCBSV species?

Knowledge gathered from this study on reasons for trait discrepancy between CBSV and UCBSV species, will serve as a road map towards improvement of ongoing breeding programs and RNA interference (RNAi)-mediated genetic engineering of cassava in sub-Saharan Africa. This will consequently reduce sporadic outbreaks of the CBSD, yield loss and food insecurity, subsequently.

2. Protein biology and methods for modelling of 3D structures of viral proteins

2.1. Functional proteomics of cassava Brown Streak Viruses (CBSVs)

Both CBSV and UCBSV species expresses sum of 10 mature proteins namely: the CP, CI, P1, P3, 6K1, 6K2, NIa, NIb, Ham 1 and VPg (Mbanzibwa et al., 2009b; Ndunguru et al., 2015; Alicai et al., 2016). The two viruses express these proteins by direct translation of its monopartite, positive sense single stranded genomic RNA (+ssRNA) to form a single polyprotein (2900–2912 amino acid long) (Mbanzibwa et al., 2009a; Winter et al., 2010; Mbanzibwa et al., 2011; Mbewe et al., 2017; Tomlinson et al., 2018). At specific sites, the polyprotein is autocatalytically cleaved using viral Serine Protease (P1) (Mbewe et al., 2017; Ndunguru et al., 2015). The P1 with highest nucleotide variation serve as RNA silencing suppressor (RSS) and it is also suspected to play crucial role in vector-borne transmission of the two viruses, by binding to stylet of Whiteflies (Alicai et al., 2016; Ivanov et al., 2014; Mäkinen and Hafrén, 2014; Mbewe et al., 2017). Besides encasing the viral genomic RNA, the CP is a pathogenicity determinant, as it interacts with cellular structures during infection as well as vectors species (Whitefly and Aphids) during disease transmission (Ateka et al., 2017). To conserve molecular functions of CBSVs, the atypical Ham 1 protein serve as “mutation brake” by hydrolyzing non-canonical, mutagenic nucleotide triphosphates (NTPs) and prevent their incorporation into viral genome during replication (Mbanzibwa et al., 2009b; Tomlinson et al., 2019). VPg replicates viral genomic RNA, once it form a large Virus Replication Complex (VPg-NIb-6K2) with 6K2 and NIb proteins. Replication of genomic (+)ssRNA of CBSVs is catalyzed by the NIb, which is synonymously known as RNA dependent RNA Polymerase (Mbewe et al., 2017). The function of 6K1 is not clear, but long distant cell-to-cell movement of Cassava Brown Streak Viruses (CBSVs) is facilitated by 6K2 which interacts with cellular membranes (Alicai et al., 2016). Function(s) of additional P3N-PIPO protein which is produced exceptionally from another open reading frame (ORF) after a + 2 Ribosomal frame-shift in the P3 sequence of CBSVs, is not known. However, it has been hypothesized to play crucial role in regulation of gene expression (Ivanov et al., 2014).

2.2. Comparative homology modelling of three dimensional (3D) structures of proteins

Comparative homology modelling of proteins is the simulation of three dimensional (3D) structure of target protein sequences by using pre-solved template structures (references) stored in databases through four major sequential steps: template identification, template selection, model building and model quality estimation (Studer et al., 2019; Walker, 2009; Waterhouse et al., 2018; Xiang et al., 2020).

Being the most popular web-server for homology modelling of 3D structures of proteins (Waterhouse et al., 2018), the present study hereby discuss principles underling SWISS-MODEL algorithm. Besides offering an option to model multiple proteins in parallel, SWISS-MODEL accepts query sequences with not more than 5000 amino acid residues.

In SWISS-MODEL, 3D-structures of query proteins are created by copying information (coordinates) of homologous and evolutionary related structures stored in SWISS-MODEL Template Library (SMTL) after similarity search between them using Basic Local Alignment (BLAST) tool. The SMTL is routinely synchronized with the highly reputable PDB resource, containing manually curated experimentally determined protein structures generated by X-ray crystallography, Electron Microscopy (EM) and NMR. HHBlits package tool uses Hidden Markov Model (HMM)-HMM to include secondary structure information predicted by Position-Specific Iterative Basic Local Alignment Search (PSI-BLAST)-based secondary structure PREDiction (PSIPRED) tool to the model (Studer et al., 2019). Promod3 and MODELLER algorithms uses annotated homo-oligomeric structures of selected template stored in SMTL to construct and refine the final 3D structure of protein (Biasini et al., 2014). Quality of the selected template and simulated protein structure is measured in terms of GMEAN and QMEAN *Z*-score value ranging between 0 and 1 (Waterhouse et al., 2018). QMEAN Z-scores around zero indicate good agreement between the model structure and experimental structures of similar size. Nonetheless, the higher the score values, the higher the quality of the simulated model. Models with ≤ -4.0 QMEAN Z-scores are considered to be of low quality (https://swissmodel.expasy.org/docs/help) (Studer et al., 2019).

The major drawback of comparative homology modelling is that, it doesn’t give an insight on why a protein adopts a certain 3D structure conformation because the model is built by reproducing the pre-existed reference 3D structure (Rigden, 2017).

2.3. De novo modelling of 3D structures of proteins

De novo modelling softwares (ROSETTA, I-TASSER, QUARK, ASTRO-FOLD, UNRES and AMBER/CHARMM/OPLS) predicts 3D structures of protein from the scratch on basis of energy functions of amino acids constituting primary structure of protein, rather than relying on experimentally solved reference structures (Rigden, 2017). The said tools uses either of the two (physics- or knowledge-based)

Table 1
40 strains of Cassava Brown Streak viruses (CBSVs) collected in CBSD-endemic countries (Tanzania, Kenya, Uganda, Mozambique and Malawi) between 1997 and 2014.

Accession number (GenBank)	Virus Species: Strain	Genome size (nt)	Geographical origin	Sample collection date	Publication
GQ329864	CBSV: Tanzanian strain	9008	Musoma district, Kiobakari, Tanzania	2008	Monger et al., 2010
FJ185044	UCBSV: Ugandan strain	9070	Mukono, Uganda	2006	Monger et al., 2010
FJ039520	UCBSV:MLB3	9069	Muleba, Tanzania	2007	Mbanzibwa et al., 2009a
HM181930	UCBSV: UG[Uganda: Namulonge:2004]	9070	Namulonge, Uganda	2004	Mbanzibwa et al., 2011
GU563327	CBSV KOR6 strain	8995	Korogwe, Tanga, Tanzania	2008	Mbanzibwa et al., 2011
FN433930	UCBSV:Ke_125	9070	Kilifi, Kenya	1999	Winter et al., 2010
FN433931	UCBSV: Ke_54 isolate	9070	Kilifi, Kenya	1997	Winter et al., 2010
FN434109	UCBSV: Ug_23	9070	Namulonge, Uganda,	2007	Winter et al., 2010
FN433932	UCBSV Ma_42	9070	Chitendze, Malawi	2007	Winter et al., 2010
FN433933	UCBSV Ma_43	9070	Salima, Malawi	2007	Winter et al., 2010
FN434436	CBSV Mo_83	9008	Namcurra, Mozambique	2007	Winter et al., 2010
HG965222	UCBSV-[UG;Kab;07]	9020	Kabanyoro, Uganda	2007	Beyene et al. (Unpublished)
KR108838	UCBSV TZ_Ser_5	8703	Serengeti, Tanzania	01-May-2013	Ndunguru et al., 2015
KR108830	CBSV TZ_Ser_6	8748	Serengeti, Tanzania	01-April-2013	Ndunguru et al., 2015
KR108837	UCBSV:TZ_Ser_6	8703	Serengeti, Tanzania	01-May-2013	Ndunguru et al., 2015
KR108832	CBSV:TZ_Tan_19_2	8748	Tanga, Tanzania	01-April-2013	Ndunguru et al., 2015
KR108839	UCBSV: TZ_Tan_23	8706	Tanga,Tanzania	01-April-2013	Ndunguru et al., 2015
KR108833	CBSV:TZ_Tan_26	8742	Tanga, Tanzania	01-April-2013	Ndunguru et al., 2015
KR108831	CBSV:TZ_Nya_36	8748	Nyasa, Tanzania	01-April-2013	Ndunguru et al., 2015
KR108829	CBSV: TZ_Nya_38	8748	Nyasa, Tanzania	01-April-2013	Ndunguru et al., 2015
KR108828	CBSV:TZ_MAF_49	8748	Mafia, Tanzania	01-May-2013	Ndunguru et al., 2015
KR108836	UCBSV:TZ_MAF_51	8706	Mafia, Tanzania	01-May-2013	Ndunguru et al., 2015
KR108835	UCBSV:TZ_MAF_58	8706	Mafia, Tanzania	01-May-2013	Ndunguru et al., 2015
LT577537	CBSV:UG_Muk_1	8751	Mukonongo district, Uganda	01-July-2013	Alicai et al., 2016
LT577538	CBSV:UG_Muk_4	8751	Mukonongo district, Uganda	01-July-2013	Alicai et al., 2016
LT577539	UCBSV:UG-May_8	8703	Mayuge district, Uganda	01-July-2013	Alicai et al., 2016
MG387659	CBSV; F6S1	8748	Kenya	08-May-2013	Ateka et al., 2017
MG387657	CBSV:F7S3	8748	Kenya	06-May-2013	Ateka et al., 2017
MG387653	UCBSV:F14S3	8703	Kenya	02-May-2013	Ateka et al., 2017
MG387654	UCBSV:F14S4	8703	Kenya	03-May-2013	Ateka et al., 2017
MG387655	UCBSV:F12S1	8307	Kenya	04-May-2013	Ateka et al., 2017
MG387652	UCBSV:F13S1	8313	Kenya	01-May-2013	Ateka et al., 2017
MG387656	UCBSV:F14S3	8667	Kenya	05-May-2013	Ateka et al., 2017
KY563367	CBSV: CBSV_Nama_Moz_4_14	8655	Namapa, Mozambique	05-June-2014	Amisse et al., 2019a
KY563362	CBSV: CBSV_Nama_Moz_5_14	8749	Namapa, Mozambique	05-June-2014	Amisse et al., 2019a
KY563361	CBSV: CBSV_Namp_Moz_8_14	8748	Nampula, Mozambique	05-June-2014	Amisse et al., 2019a
KY563366	CBSV: CBSV_Alt_Mz_16_14	8748	Alto Molocue, Mozambique	05-June-2014	Amisse et al., 2019a
KY563363	CBSV: CBSV_Mocu_Moz_20_14	8748	Mocuba, Mozambique	05-June-2014	Amisse et al., 2019a
KY563364	CBSV: CBSV_Mocu_Moz_22_14	8748	Mocuba, Mozambique	05-June-2014	Amisse et al., 2019a
KY563365	CBSV: CBSV_Que_Moz_23_14	8748	Quelimane, Mozambique	05-June-2014	Amisse et al., 2019a

energy function types, to simulate models owing to whether statistics from pre-determined 3D structures in PDB will be applicable during modelling iterations (Chivian et al., 2005; Xiang, 2006; Higgs and Stantic, 2008; Zhang, 2008; Yang et al., 2014; Rigden, 2017). Among the mentioned software packages, ROSETTA and I-TASSER are popular and used more frequently for de novo modelling (Higgs and Stantic, 2008). In this context the present study made use of ROBETTA, an open source server for a linux-based Rosetta (Chivian et al., 2005), because it uses both physics- and energy-based force field types (energy functions) and permits parallelization of dual jobs unlike I-TASSER server which uses only knowledge-based energy function and permits only a single job run at a time. Nonetheless, both ROSETTA and I-TESSER use Monte Carlo, Clustering/filtering methods to search and select appropriate ab initio models (Chivian et al., 2005; Rigden, 2017; Zhang, 2008). Usually the final model is selected from thousands of structural conformations (decoys), which were constructed after query sequence matched to aligned energy functions.

Accurate energy function with which the native structure of a protein corresponds to the most thermodynamically stable state compared to all possible decoy structures, efficient search method which can quickly identify the low-energy states through conformational search and strategy that can select near-native models from a pool of decoy structures; are the critical factors which pre-determine efficacy of the ab initio modelling method (Rigden, 2017).

In both ROBETTA and I-TASSER, quality of the predicted 3D-structures of proteins is assessed in terms of computed Confidence (C) score (0–1). The higher the Confidence (C) score value, the confident is the model (Chivian et al., 2005; Rigden, 2017; Zhang, 2008). Unlike homology modelling, de novo (ab initio) modelling gives an insight on why and how primary protein adapts a certain 3D structure.

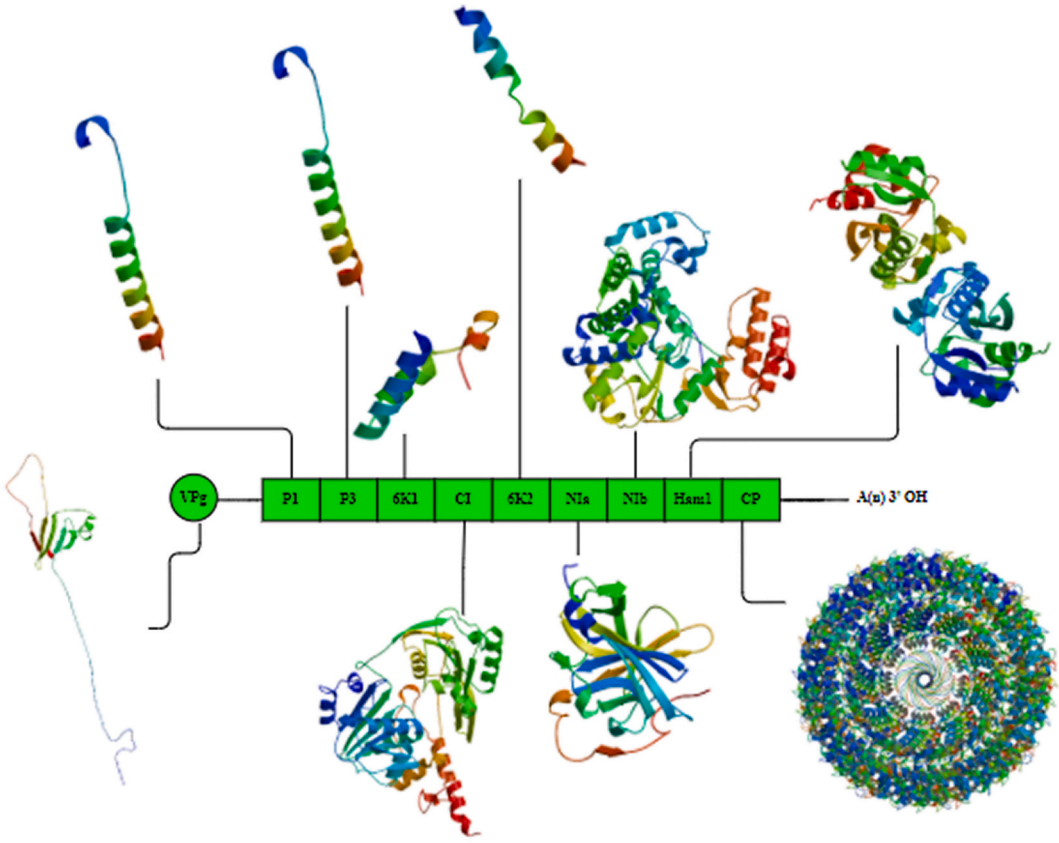

Fig. 1. Homology model of CBSV: TZ_MAF_49 strain (KR108828) predicted using SWISS-MODEL server.

3. Materials and methods

3.1. Source of sequence data

The present in silico-based study made use of protein sequence data of randomly selected 20 and 20 strains of CBSV and UCBSV species, published in National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) between 2009 and 2019 (Alicai et al., 2016; Amisse et al., 2019b; Ateka et al., 2017; Mbanzibwa et al., 2009b; Mbanzibwa et al., 2011; Monger et al., 2010; Ndunguru et al., 2015; Winter et al., 2010). Accession numbers, sample collection date and geographical origin of viruses used in this study are shown in Table 1.

3.2. Analysis of viral nucleotide and protein sequences

After determining nucleotide identity between CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains by Nucleotide BLAST (BLASTN), SWISS-MODEL (https://swissmodel.expasy.org/) and PROTEIN PARAMETERS tool (http://www.protparam.net/index.html) were used for comparative homology modelling and estimation of physical-chemical properties between the two virus species, correspondingly. By the time this study was conducted SWISS-MODEL server made use of 165,967 structures in PDB and predicted models basing on UniProtKB release 2021_01. Comparative homology modelling was further narrowed to only 6K2 protein of additional 19 and 19 strains of CBSV and UCBSV species, respectively. Sequence identity of 6K2 protein shared between and within 20 and 20 strains of CBSV and UCBSV species, respectively was determined using Protein BLAST (BLASTP) with default parameter set. To validate (confirm) results of homology modelling, ROBETTA web server (https://robetta.bakerlab.org/) was used for de novo (ab initio) modelling of 40 6K2 proteins before the generated three dimensional (3D) structures were superimposed against each other (without refinement) by *"(align Protein_1.pdb Protein_2.pdb, cycles = 0, transform = 0)"* command in PYMOL version 2. 4. 1 (https://pymol.org/2/). Pairwise alignment of 3D structures of 6K2 protein were refined by using *align* command *"(align protein1_to_ Protein 2)"* in PYMOL v. 2. 4. 1. MUSCLE in UGENE version 40.1, was used to determine genetic variation of the 6K2 protein between 20 and 20 strains of CBSV and UCBSV species separately. To determine evolutionary relatedness within and between 6K2 proteins of 20 and 20 variants of CBSV and UCBSV species respectively, phylogenetic tree was constructed by using PhyML-Maximum Likelihood method in UGENE v 40.1 software suite. During tree construction, all parameters were set default (substitution model (LG), equilibrium frequency (emperical), transition/transversion ratio (0), proportion of invariable sites (0.00), number of substitution rate category (4) and gama shape parameter (0)).

4. Results

4.1. Genome-wise homology modelling of three dimensional (3D) structures of proteins expressed by CBSV: TZ_MAF 49 versus UCBSV: TZ_MAF 58 strain

Genome-wise, comparative homology modelling of the CBSV:

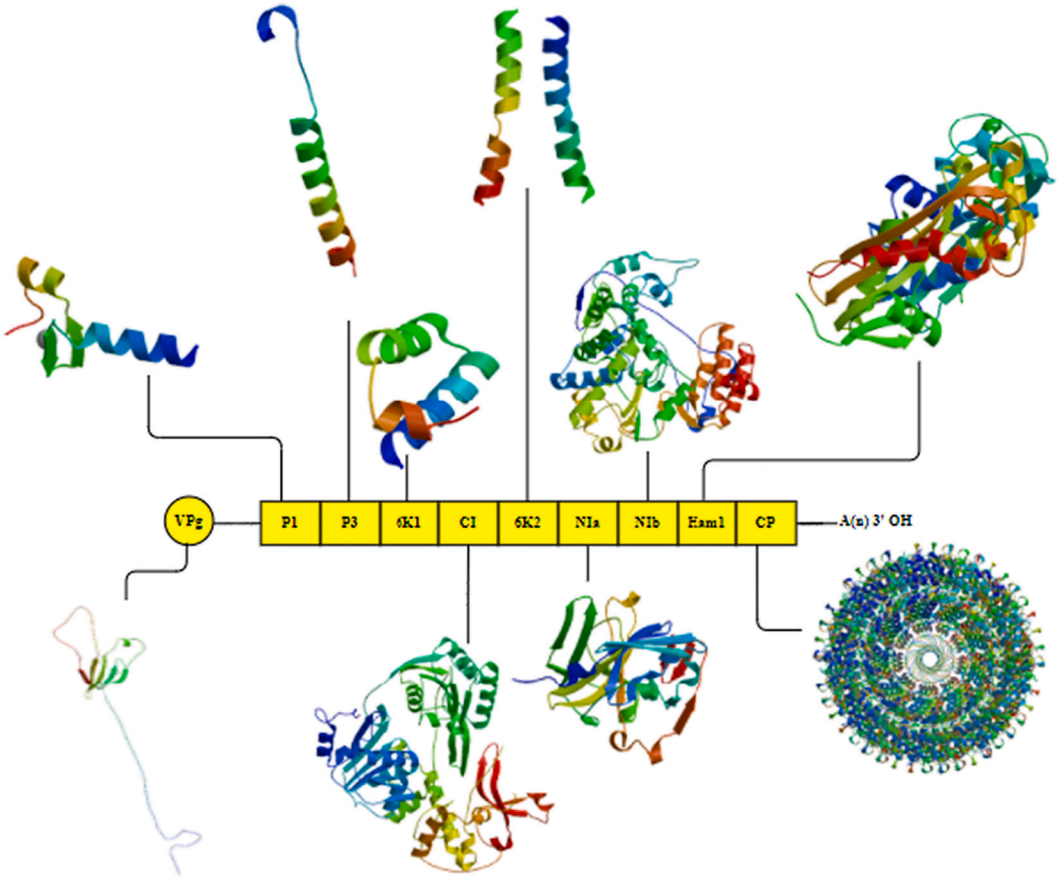

Fig. 2. Homology model of UCBSV: TZ_MAF_58 strain (KR108835) predicted using SWISS-MODEL server.

TZ_MAF 49 (8748 nucleotide long) and UCBSV: TZ_MAF 58 (8706 nucleotide long) strains with 72.19% nucleotide identity, done by using SWISS-MODEL server reveled that; at low confidence (QMEAN scores < (0–1)) the two virus species express nearly identical protein structures with exception of 6K2 protein (Fig. 1 and Fig. 2).

6K2 protein was predicted to be expressed as monomer and homo-dimer by the virulent CBSV: TZ_MAF 49 (QMEAN score = −1.51) and mild UCBSV: TZ_MAF 58 QMEAN score = −2.50) strains, correspondingly (Table 2 and Table 3). The 6K2 proteins of both CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains were identical to 3D structure of receptor tyrosine-protein kinase erbB-3 with PDB identification number (2I9u. 1. A) (Mineev et al., 2011). 0.48 coverage (18–42 range) was achieved when the template (2I9u. 1. A) was aligned to the 6K2 amino acid sequences of both CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains at (44 and 40) and (0.39 and 0.40) percentage (%) sequence identity and similarity, correspondingly.

Further homology modelling of 6K2 proteins of additional 19 and 19 strains of CBSV and UCBSV species respectively, revealed that; in total (1 (5.26%) and 18 (94.74%)) of the former and (5 (26.32%) and 14 (73.68%)) of the latter species expressed the same protein as a monomer and home-dimer correspondingly (Table 4 and Table 5).

Because 3D structural models of 6K2 proteins of the two virus species were predicted with low quality (i.e QMEAN score range for CBSV and UCBSV species were (−0.134 - -2.52) and (−1.99 - -2.51), respectively). in SWISS-MODEL server due to lack of reference structure in PDB with at least 60% amino acid sequence homology, de novo (ab initio) modelling by using ROBETTA server was done to confirm (validate) whether structural variation in 6K2 protein exists in a larger strain population of CBSV and UCBSV species or not.

The two viruses were found to express seven proteins (P1, P3, 6K1, NIa, NIb, CI and VPg) as monomers whilst the remaining two (Ham 1 and CP) proteins were expressed as homo-dimer and homo-35-mer, correspondingly. All predicted proteins of the two virus species had no ligands, with exception of P1 which contained Zinc (Zn) in its structure.

Even though none of the 10 proteins of the CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, matched with pre-determined 3D structures in PDB with at least 60% sequence identity, molecular functions of majority proteins were correctly anticipated in SWISS- MODEL as mentioned in previous publications (Alicai et al., 2016; Amisse et al., 2019a; Mbanzibwa et al., 2011; Mbewe et al., 2017; Monger et al., 2010; Ndunguru et al., 2015).

P3, P1, VPg, NIb, CP, Ham 1 and CI proteins expressed by both CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, were predicted to function as; regulatory focus of Yeast Kinetochore assembly, Transcription factor TFIIE of Yeast (*Saccharomyces cerevisiae*), Viral Protein genome linked (VPg) in recruiting eukaryotic translation initiation factors, RNA-directed RNA polymerase of Sapporo virus, Coat Protein of Potato Y virus, Inosine triphosphate pyrophosphatase (ITPA) P32T variant of

Table 2
Homology modelling results of CBSV: TZ_MAF 49 strain (KR108828) using SWISS-MODEL.

Protein identity	P1	P3	6K1	6K2	NIa
Protein size (aa residues)	362	294	52	52	324
Biounit Oligo-state	Homo-35-mer	Monomer	Monomer	Monomer	Monomer
GMQE	0.41	0	0.28	−0.15	0.56
Ligand	None	None	None	None	None
QMEAN	−5.87	−1.05	−2.96	−1.51	−2.33
Cβ	−3.96	−2.04	−2.02	−2.62	−1.92
All atoms	−2.27	0.9	−0.32	1.65	−2.49
Solvation	−2.24	−1.83	−2.9	−0.83	−1.28
Torsion	−4.34	0.09	−1.85	−0.96	−1.39
Number of Templates found in PDB	14	16	18	26	50
Templates matched with query sequence	3	4	1	1	2
Identity of the selected template (PDB accession)	6hxx.1.A	5t5I.1.A	1bkd.1.B	2I9u.1.A	3mmg.1.A
QSQE	0.09	0	0	0.08	0
Method of template generation	EM	X-ray, 2.20 A	X-ray, 2.80 A	NMR	X-ray, 1.70 A
Sequence Identity between selected template and query (%)	24.03	15.15	20	44	22.07
Sequence similarity between selected template and query	0.32	0.32	0.3	0.39	0.31
Coverage	0.62	0.11	0.77	0.48	0.95
Range	141–377	63–95	6.45	18–42	4–218

Protein identity	NIb	CI	Ham 1	CP	VPg
Protein size (aa residues)	502	630	226	378	185
Biounit Oligo-state	Monomer	Monomer	Homo-dimer	Homo-35-mer	Monomer
GMQE	0.43	0.2	0.58	0.41	0.46
Ligand	None	None	None	None	None
QMEAN	−4.93	−6.88	−2.1	−5.87	−6.87
Cβ	−4.03	−4.49	−0.52	−3.96	−5.82
All atoms	−2.41	−4.49	−0.44	−2.27	−4.4
Solvation	−0.97	−4.31	0.63	−2.24	−3.26
Torsion	−4.17	−4.89	−2.32	−4.34	−4.66
Number of Templates found in PDB	50	50	38	14	11
Templates matched with query sequence	1	2	1	3	3
Identity of the selected template (PDB accession)	3sfg.2.B	5gvu.1.A	4f95.1.A	6hxx.1.A	6nfw.1.A
QSQE	0	0	0.82	0.09	0
Method of template generation	X-ray, 2.21 A	X-ray, 2.82 A	X-ray, 2.07 A	EM	NMR
Sequence Identity between selected template and query (%)	16.92	31.64	50.54	24.03	26.75
Sequence similarity between selected template and query	0.28	0.36	0.44	0.32	0.33
Coverage	0.8	0.56	0.82	0.62	0.84
Range	63–476	62–457	24–207	141–377	29–179

human and Pre-mRNA-splicing factor and Adenosine Triphosphate (ATP)-dependent RNA helicase (PRP22) of Yeast (*S. cerevisiae*), correspondingly (de Oliveira et al., 2019; Dimitrova et al., 2016; Fica et al., 2017; Fullerton et al., 2007; Kežar et al., 2019; Schilbach et al., 2017; Simone et al., 2013). In the same order the mentioned proteins encoded by CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, matched to the identical 5t51.1.A, 5oqm.1.U, 6nfw.1.A, 2ckw.1.A, 6hxx.1.A, 4f95.1.A and 5mqo.1.A templates in PDB with (15.15% and 15.15%), (27.91% and 27.91%), (26.75% and 25.95%), (18.43% and 17.49%), (24.03% and 25.99%), (50.54% and 51.58%) and (25% and 23.33%) sequence identity, separately.

Surprisingly when CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains were compared, each of the two (NIa and 6K1) proteins matched to distinct reference structures in PDB.

6K1 protein sequences of CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains were identical to 2ije.1.A and 5wfo.1.C templates respectively, by 20% and 17.07% sequence identity, correspondingly (Freedman et al., 2006; Tsymbaliuk, 2019). In CBSV: TZ_MAF 49 strain, 6K1 was predicted to function as Guanine nucleotide releasing protein, crystal structure of RasGRFI whilst in UCBSV: TZ_MAF 58 strains the same protein is functionally similar to son of sevenless (SOS) Homolog 1, which perturbs RAS signaling in human upon binding to RAS:SOS:RAS (Freedman et al., 2006; Tsymbaliuk, 2019). From these findings we hypothesize that, 6K1 protein perform at least dual functions in CBSVs.

On the other hand, NIa protein of CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, showed 22.07% and 22.58% identity to 3mmg.2.A and 1Ivb.1.A templates in PDB, respectively. In CBSV: TZ_MAF 49 strain is predicted to function as a protease (nuclear inclusion protein a) of *Tobacco Vein Mottling Virus* (Sun et al., 2010). In UCBSV: TZ_MAF 58 strains, the same protein serve as a catalytic domain of the Nuclear Inclusion a (NIa) protease of Tobacco Etch Virus (Phan et al., 2002). The fact that, NIa performs more or less similar function in *Tobacco vein Mottling virus* and *Tobacco Etch Virus*, shows that the protein is functionally conserved among the said potyviruses.

4.2. De novo (ab initio) modelling of three dimensional (3D) structures of second 6-Kilodalton (6K2) proteins expressed by CBSV and UCBSV species

Findings of de novo (ab initio) modelling of 3D structure of 6K2 proteins using ROBETTA server were different from homology modelling by SWISS-MODEL server.

ROBETTA predicted 3D-structure of 6K2 proteins with better confidence compared to the SWISS-MODEL. Models of the same protein were predicted with (0.54–0.93) and (0.46–0.76) confidence score values for 20 and 20 strains of CBSV and UCBSV species, correspondingly (Table 6). Among top 5 ab initio models of the 6K2 protein predicted in ROBETTA server, the first model was selected as the best one despite the fact that; all of them had identical confidence score values. In this study, ab initio 3D structure models of 6K2 proteins with (0.5–1.0) confidence score values, were considered valid for both CBSV and UCBSV species.

Table 3
Homology modelling results of UCBSV: TZ_MAF 58 strain (KR108835) using SWISS-MODEL.

Protein identity	P1	P3	6K1	6K2	NIa
Protein size (aa residues)	362	294	52	52	324
Biounit Oligo-state	Zinc Ion	Monomer	Monomer	Homo-dimer	Monomer
GMQE	0	0	0.32	0.18	0.57
Ligand	None	None	None	None	None
QMEAN	−1.55	−1.14	−2.57	−2.5	−4.69
Cβ	−1.09	−2.05	−1.9	−2.95	−3.03
All atoms	−0.88	−0.02	−0.46	1.34	−2.42
Solvation	−0.82	−1.06	−2.47	−1.36	−2.02
Torsion	−1.34	−0.34	−1.57	−1.83	−3.34
Number of Templates found in PDB	45	16	18	29	50
Templates matched with query sequence	1	3	1	1	3
Identity of the selected template (PDB accession)	5oqm.1.U	3bcg.1.A	2ije.1.A	2I9u.1.A	1q3I.1.A
QSQE	0	0	0	0	0
Method of template generation	EM	X-ray, 2.48 A	X-ray, 2.2 A	NMR	X-ray, 2.7 A
Sequence Identity between selected template and query (%)	27.91	17.86	22.5	40	21.92
Sequence similarity between selected template and query	0.33	0.29	0.33	0.4	0.31
Coverage	0.12	0.1	0.77	0.48	0.94
Range	108–150	103–130	Jul-46	18–42	6–232

Protein identity	NIb	CI	Ham 1	CP	VPg
Protein size (aa residues)	502	628	226	366	185
Biounit Oligo-state	Monomer	Monomer	Homo-dimer	Monomer	Monomer
GMQE	0.52	0.42	0.6	0.41	0.44
Ligand	None	None	None	None	None
QMEAN	−6.25	−5.84	−1.1	−5.72	−7.62
Cβ	−5.23	−3.74	0.41	−3.73	−3.16
All atoms	−2.61	−3.2	−0.88	−2.2	−3.93
Solvation	−1.74	−2.1	0.09	−2.36	−4.08
Torsion	−4.91	−4.31	−1.24	−4.2	−5.61
Number of Templates found in PDB	50	50	28	12	16
Templates matched with query sequence	1	1	1	3	3
Identity of the selected template (PDB accession)	2ckw.1.A	5xjc.1.Y	4f95.1.A	6hxx.1.A	6nfw.1.A
QSQE	0	0	0.86	0.1	0
Method of template generation	X-ray, 2.30 A	EM, 3.60 A	X-ray, 2.07 A	EM	NMR
Sequence Identity between selected template and query (%)	17.49	24.9	5158.00%	25.99	25.95
Sequence similarity between selected template and query	0.27	0.32	0.44	0.33	0.32
Coverage	0.89	0.81	0.84	0.62	0.85
Range	10–476	44–579	22–207	134–365	29–177

Ab initio modelling revealed that, different strains of each of CBSV and UCBSV species express 6K2 proteins with alternating 3D structures. Ab initio models of 6K2 proteins of(19 (95%) and 1(4.35%)) and (17 (85%) and 3(15%)) strains of CBSV and UCBSV species, respectively had corresponding homo-trimer and homo-tetramer oligomeric 3D structures, each composed of alpha α subunits (helices) with conspicuous variation in length.

In general among 40 strains examined in this study, only 4 strains expressed its 6K2 in homo-tetrameric state whilst the rest expressed the protein in homo-trimeric state. Among the four strains with homo-tetrameric structures, only 1 (UG_Muk_1) strain with 0.93 confidence score belonged to CBSV species whilst the remaining 3 (TZ_MAF_58, UG; Kab; 07 and F12S1) strains with (0.55, 0.50 and 0.69) confidence scores, respectively belonged to UCBSV species. Ab initio models of the 6K2 proteins of 20 and 20 strains of CBSV and UCBSV species are shown in Fig. 3: (A1-A20) and (A21-A40) correspondingly.

4.3. Similarity in 3D structure of 6K2 ab initio models between and within CBSV and UCBSV species

Ab initio 3D structure models of 20 and 20 variants of CBSV and UCBSV species correspondingly, sharing 75–88.46% amino acid sequence identity have remarkable variation in 3D structure. 1.082–13.814 range of Root Mean Square Deviation (RMSD) was obtained when each of 20 strains of CBSV were superimposed against each of 20 strains of UCBSV species in situ (without refinement). Such variation exist not only between CBSV and UCBSV species, but also between strains of each of the two virus species (Table 7).

Fig. 4: (H1-H48) show refined 3D structural alignment between several categories of strains. 1.807–4.615, 1.483–12.886, 0.464–9.992, 0.343–6.72, 6.995–7.908, 0.322–11.099, 0.502–8.601, 5.799–8.897 and 1.222–11.019 range of RMSD values were obtained following a refined alignment of 6K2 proteins between: homo-tetrameric structures of CBSV species and homo-trimeric structures of UCBSV species (Fig. 4: H1-H6), homo-tetrameric structures of CBSV species and homo-trimeric structures of CBSV species (Fig. 4: H7-H12), different homo-trimeric structures of CBSV species (Fig. 4: H13-H18), homo-tetrameric structures of UCBSV species and homo-trimeric structures of CBSV species (Fig. 4: H19-H24), homo-tetrameric structures of UCBSV species and homo-trimeric structures of UCBSV species (Fig. 4: H25-H30), different homo-trimeric structures of UCBSV species (Fig. 4: H31-H36), homo-tetrameric structures of CBSV species and homo-tetrameric structures of UCBSV species (Fig. 4: H37-H39), different homo-tetrameric structures of UCBSV species (Fig. 4: H40-H42) and homo-trimeric structures of UCBSV species and homo-trimeric structures of CBSV species (Fig. 4: H43-H48), correspondingly (Table 8).

Homo-tetrameric 3D structure of CBSV species couldn't match perfectly to diverse homo-trimeric 3D of UCBSV species. Significant superimposition differences were observed when a unique CBSV: UG_Muk_1 strain (Accession: LT577538) with a homo-tetrameric 3D structure was aligned to diverse homo-trimeric 3D of UCBSV species

Table 4
Homology modelling results of 6K2 protein of 19 strains of CBSV species using SWISS-MODEL.

Accession number (GenBank)	GQ329864	**GU563327**	FN434436	KR108834	KR108832	KR108833
Protein size (aa residues)	52	52	52	52	52	52
Biounit Oligo-state	Monomer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer
GMQE	0.18	0.15	0.15	0.18	0.18	0.18
QMEAN	−2.07	−0.134	−1.51	−1.69	−2.22	−2.07
Ligand	None	None	None	None	None	None
QMEAN	−2.07	−0.134	−1.51	−1.69	−2.22	−2.07
Cβ	−2.97	−2.54	−2.62	−3.08	−2.61	−2.97
All atoms	1.42	1.7	1.65	1.8	1.43	1.42
Solvation	−0.89	−0.95	−0.83	−0.97	−0.58	−0.89
Torsion	−1.5	−0.7	−0.96	−0.99	−2	−1.5
Number of Templates found in PDB	32	24	26	27	29	32
Templates matched with query sequence	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0.1	0.08	0.08	0.08	0.11	0.1
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	44	44	44	44	42.31	44
Sequence similarity between selected template and query	0.4	0.39	0.39	0.4	0.4	0.4
Coverage	0.48	0.48	0.48	0.48	0.5	0.48
Range	18–42	18–42	18–42	18–42	17–42	18–42

Accession number (GenBank)	**KR108831**	KR108829	**LT577537**	LT577538	MG387659	MG387657
Protein size (aa residues)	52	52	52	52	52	52
Biounit Oligo-state	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer
GMQE	0.15	−0.15	0.18	0.18	0.18	0.18
QMEAN	−1.34	−2.52	−2.07	−2.07	−2.07	−2.07
Ligand	None	None	None	None	None	None
QMEAN	−1.34	−2.52	−2.07	−2.07	−2.07	−2.07
Cβ	−2.54	−2.7	−2.97	−2.97	−2.97	−2.97
All atoms	1.7	1.3	1.42	1.42	1.42	1.42
Solvation	−0.95	−1.12	−0.89	−0.89	−0.89	−0.89
Torsion	−0.7	−2.04	−1.5	−1.5	−1.5	−1.5
Number of Templates found in PDB	26	28	32	32	32	32
Templates matched with query sequence	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0.08	0.11	0.1	0.1	0.1	0.1
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	44	44	44	44	44	44
Sequence similarity between selected template and query	0.39	0.4	0.4	0.4	0.4	0.4
Coverage	0.48	0.48	0.48	0.48	0.48	0.48
Range	18–42	18–42	18–42	18–42	18–42	18–42

Accession number (GenBank)	KY563367	KY563362	KY563361	KY563366	KY563363	KY563364	KY563365
Protein size (aa residues)	52	52	52	52	52	52	52
Biounit Oligo-state	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer
GMQE	0.18	0.15	0.15	0.15	0.15	0.17	0.17
QMEAN	−2.3	−1.51	−1.51	−1.51	−1.51	−1.54	−1.54
Ligand	None	None	None	None	None	None	None
QMEAN	−2.3	−1.51	−1.51	−1.51	−1.51	−1.54	−1.54
Cβ	−2.83	−2.62	−2.62	−2.62	−2.62	−2.9	−2.9
All atoms	1.38	1.65	1.65	1.65	1.65	1.66	1.66
Solvation	−0.59	−0.83	−0.83	−0.83	−0.83	−0.8	−0.8
Torsion	−2.03	−0.96	−0.96	−0.96	−0.96	−0.92	−0.92
Number of Templates found in PDB	29	26	26	26	26	30	30
Templates matched with query sequence	1	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0.11	0.08	0.08	0.08	0.08	0.08	0.08
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	42.31	44	44	44	44	44	44
Sequence similarity between selected template and query	0.39	0.39	0.39	0.39	0.39	0.39	0.39
Coverage	0.5	0.48	0.48	0.48	0.48	0.48	0.48
Range	17–42	18–42	18–42	18–42	18–42	18–42	18–42

(RMSD = 1.807–4.615) and vice versa (RMSD = 0.343–6.72).

Unlike the alignment between homo-tetrameric and homo-trimeric 3D structures of 6K2 protein, variation was higher when homo-trimeric 3D structures expressed by each of CBSV (RMSD = 0.464–9.992) and UCBSV (RMSD = 0.322–11.099) species were compared among themselves, separately.

Table 5
Homology modelling results of 6K2 protein of 19 strains of UCBSV species using SWISS-MODEL.

Accession number (GenBank)	FJ185044	FJ039520	HM181930	FN433930	FN433931	FN434109
Protein size (aa residues)	52	52	52	52	52	52
Biounit Oligo-state	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer	Homo-dimer
GMQE	0.19	0.18	0.19	0.17	0.17	0.17
QMEAN	−2.04	−2.51	−2.04	−2.51	−2.51	−2.51
Ligand	None	None	None	None	None	None
QMEAN	−2.04	−2.51	−2.04	−2.51	−2.51	−2.51
Cβ	−2.95	−2.97	−2.95	−2.97	−2.97	−2.97
All atoms	1.3	1.35	1.3	1.35	1.35	1.35
Solvation	−0.88	−1.33	−0.88	−1.33	−1.33	−1.33
Torsion	−1.51	−1.84	−1.51	−1.84	−1.84	−1.84
Number of Templates found in PDB	32	29	32	32	32	32
Templates matched with query sequence	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0	0	0	0	0	0
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	38.46	40	38.46	40	40	40
Sequence similarity between selected template and query	0.38	0.39	0.38	0.39	0.39	0.39
Coverage	0.5	0.48	0.5	0.48	0.48	0.48
Range	17–42	18–42	17–42	18–42	18–42	18–42

Accession number (GenBank)	FN433932	FN433933	HG965222	KR108838	KR108837	KR108839	KR108836
Protein size (aa residues)	52	52	52	52	52	52	52
Biounit Oligo-state	Homo-dimer	Homo-dimer	Homo-dimer	Monomer	Monomer	Homo-dimer	Homo-dimer
GMQE	0.2	0.2	0.17	0.19	0.19	0.17	0.18
QMEAN	−1.99	−1.99	−2.51	−2.29	−2.29	−2.51	−2.5
Ligand	None	None	None	None	None	None	None
QMEAN	−1.99	−1.99	−2.51	−2.29	−2.29	−2.51	−2.5
Cβ	−2.87	−2.87	-2.97	−3.37	−3.37	−2.97	−2.95
All atoms	1.34	1.34	1.35	1.55	1.55	1.35	1.34
Solvation	−0.75	−0.75	−1.33	−1.01	−1.01	−1.33	−1.36
Torsion	−1.53	−1.53	−1.84	−1.68	−1.68	−1.84	−1.83
Number of Templates found in PDB	33	33	32	28	28	31	29
Templates matched with query sequence	1	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0	0	0	0	0	0	0
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	38.46		40	38.46	38.46	40	40
Sequence similarity between selected template and query	0.38		0.39	0.38	0.38	0.39	0.4
Coverage	0.5		0.48	0.5	0.5	0.48	0.48
Range	17–42		18–42	17–42	17–42	18–42	18–42

Accession number (GenBank)	LT577539	MG387653	MG387654	MG387655	MG387652	MG387656
Virus strain (variant)	UG-May_8	F14S3	F14S4	F12S1	F13S1	F14S3
Protein size (aa residues)	52	52	52	52	52	52
Biounit Oligo-state	Homo-dimer	Monomer	Monomer	Homo-dimer	Monomer	Homo-dimer
GMQE	0.17	0.19	0.19	0.17	0.15	0.17
QMEAN	−2.51	−2.29	−2.29	−2.51	−1.51	−2.51
Ligand	None	None	None	None	None	None
QMEAN	−2.51	−2.29	−2.29	−2.51	−1.51	−2.51
Cβ	−2.97	−3.37	−3.37	−2.97	−2.62	−2.97
All atoms	1.35	1.55	1.55	1.35	1.65	1.35
Solvation	−1.33	−1.01	−1.01	−1.33	−0.83	−1.33
Torsion	−1.84	−1.68	−1.68	−1.84	−0.96	−1.84
Number of Templates found in PDB	32	28	28	32	26	32
Templates matched with query sequence	1	1	1	1	1	1
Identity of the selected template (PDB accession)	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A	2I9u.1.A
QSQE	0	0	0	0	0.08	0
Method of template generation	NMR	NMR	NMR	NMR	NMR	NMR
Sequence Identity between selected template and query (%)	40	38.46	38.46	40	44	40
Sequence similarity between selected template and query	0.39	0.38	0.38	0.39	0.39	0.39
Coverage	0.48	0.5	0.5	0.48	0.48	0.48
Range	18–42	17–42	17–42	18–42	18–42	18–42

Homo-tetrameric 3D structures of three UCBSV species were more dissimilar between themselves (RMSD = 5.799–8.897), compared to when they were compared with a sole CBSV: UG_Muk_1 strain (Accession: LT577538) expressing homo-tetrameric 3D structure of 6K2 protein (RMSD = 0.502–8.601).

Additionally, homo-trimeric 3D structures of UCBSV were different

Table 6
Results of ab initio modelling of 3D structure of the 6K2 protein encoded by 20 CBSV and 20 UCBSV species.

Accession number (GenBank)	Virus Species	Confidence score	Model Span (aa)
KR108830	CBSV	0.84	1–52
FN434436	CBSV	0.75	1-52
GQ329864	CBSV	0.93	1-52
GU563327	CBSV	0.58	1-52
KR108829	CBSV	0.57	1–52
KR108833	CBSV	0.93	1–52
KY563361	CBSV	0.93	1–52
KY563362	CBSV	0.87	1–52
KR108828	CBSV	0.82	1–52
KY563363	CBSV	0.72	1–52
KY563364	CBSV	0.54	1–52
KY563365	CBSV	0.75	1–52
KY563366	CBSV	0.72	1–52
KY563367	CBSV	0.56	1–52
LT577537	CBSV	0.94	1–52
LT577538	CBSV	0.93	1–52
MG387657	CBSV	0.86	1–52
MG387659	CBSV	0.76	1–52
KR108831	CBSV	0.85	1–52
KR108832	CBSV	0.68	1–52
KR108837	UCBSV	0.6	1–52
MG387654	UCBSV	0.51	1–52
FJ039520	UCBSV	0.56	1-52
FJ185044	UCBSV	0.59	1-52
KR108839	UCBSV	0.5	1–52
FN433930	UCBSV	0.52	1-52
FN433931	UCBSV	0.59	1-52
FN333932	UCBSV	0.56	1–52
FN333933	UCBSV	0.59	1–52
FN434109	UCBSV	0.57	1-52
HG965222	UCBSV	0.5	1-52
HM181930	UCBSV	0.76	1-52
KR108835	UCBSV	0.55	1–52
KR108836	UCBSV	0.56	1–52
KR108838	UCBSV	0. 46	1–52
LT577539	UCBSV	0.49	1–52
MG307652	UCBSV	0.72	1–52
MG387653	UCBSV	0.47	1–52
MG387655	UCBSV	0.69	1–52
MG387656	UCBSV	0.58	1–52

from those of UCBSV species (RMSD = 1.222–11.019).

Since the higher the RMSD, the higher the disparity between aligned 3D structures and vice versa is true (Maiorov and Crippen, 1994), above findings indicates that; additional subunit in homo-tetrameric 6K2 protein performs different function(s) compared to the homo-trimeric structure, regardless of species demarcation.

4.4. Variation in 6K2 protein sequence between CBSV and UCBSV species

35 and 17 amino acids of 6K2 protein were highly conserved and variable correspondingly, for both 20 CBSV and 20 UCBSV species. Cysteine (C), Alanine (A), Serine (S), Aspartic acid (D), Tyrosine (Y), Isoleucine (I), Glutamic acid (E), Lysine (K), Valine (V), Methionine (M), Guanine (G), Leucine (L), Threonine (T), Phenylalanine (F), Tryptophan (W), Arginine (R) and Glutamine (Q) were highly conserved at 1, (2, 36, 44 and 45), 3, 4, (5, 17, 18, 37 and 39), (6 and 22), (7, 8, and 47), (9,14 and 19), (10, 13 and 26), 11, (24, 27 and 33), 28, (32 and 34), 35, 38, (41 and 42) and 52 amino acid positions, respectively. Nonetheless, the two virus species compared vary significantly at 12, 15, 16, 20, 21, 23, 25, 29, 30, 31, 40, 43, 46, 48, 49, 50 and 51 amino acid positions of the 6K2 protein (Fig. 5 and Fig. 6).

3 (15%) strains (TZ_Tan_23, MLB3 and F13S1) of 20 UCBSV species had Asparagine (N) whilst the rest 17 (85%) of strains had Serine (S) at 12th amino acid position of the 6K2 protein. Conversely, at the same 12th amino acid position, 20 (100%) strains had Asparagine (N) residue.

1 (5%) of 20 strains of UCBSV species (F13S1) had K residue at 15th amino acid position of the 6K2 protein whilst the rest 19 (95%) had R residue at same position. Only 2 (10%) of 20 strains of CBSV species namely CBSV_Nama_Moz_4_14 and CBSV_Mocu_Moz_22_14, correspondingly have R residue whilst the rest 18 (90%) have K residue at the same 15th amino acid position of the protein.

18 (90%) and 2 (10%) (Ma_42 and Ma_43) of 20 strains of UCBSV species had N and S residues respectively at 16th amino acid position of the protein whilst 100% (20) strains of CBSV species had N residues at same amino acid position of the 6K2 protein.

15 (75%) and 5 (25%) (F14S4, F1453, F13S1, TZ_Ser_6 and TZ_Ser_5) strains of UCBSV species had S and P residues, respectively at 20th amino acid position of the protein whilst 19 (95%) and 1 (5%) (TZ_Nya_38) strains of CBSV species had P and S residues respectively at 20th amino acid position of the protein.

17 (85%) and 3 (15%) (TZ_MAF_51, TZ_MAF_58 and F13S1) of 20 strains of UCBSV species had L and I residue at 21st amino acid position of 6K2 protein. On the other hand, all 20 (100%) strains of CBSV species had I residues at 21st amino acid position of 6K2 protein.

1 (5%) (F13S1) and the rest 19 (95%) of 20 strains of UCBSV species had I and L residues respectively at 23rd amino acid position of the protein whilst 19 (95%) and the 1 (5%) (CBSV_Mocu_Moz_22_14) of 20 strains of CBSV species had I and V residues respectively at 23rd amino acid position of the protein.

15 (75%), 2 (10%) (Ugandan and UG [Uganda: Namulonge: 2004]), 2 (10%) (Ma_42 and Ma_43) and 1 (5%) (F13S1) of 20 strains of UCBSV species had M, V, I and L residues respectively at 25th amino acid position of the protein. 20(100%) strains of CBSV species had L residues at 25th amino acid position of the protein.

1 (5%) (F13S1) and the remaining 19 (95%) of 20 strains of UCBSV species had S and L residues respectively at 29th amino acid position of the protein whilst 9 (45%) (CBSV_Nama_Moz_5_14, CBSV_Namp_-Moz_8_14, CBSV_Mocu_Moz_22_14, CBSV_Que_Moz_23_14, TZ_Nya_38, TZ_MAF_49, Mo_83, CBSV_Mocu_Moz_20_14 and CBSV_Alt_Mz_16_14) and the rest 11 (55%) of 20 strains of CBSV species had A residues respectively at 29th amino acid position of the protein.

All (100%) of 20 strains of UCBSV species had V residues at 30th amino acid position of the protein whilst 18 (90%) and 2 (10%) (CBSV_Nama_Moz_4_14 and Kor 6) of 20 strains of CBSV species had V and I residues respectively at 30th amino acid position of the protein.

All (100%) of 20 strains of UCBSV species had A residues at 31st amino acid position of the protein. On the other hand, 7 (35%) (Tanzanian, TZ_Ser_6, TZ_Tan_26, CBSV_Nama_Moz_4_14, UG_Muk_4, F7S3 and F6S1) while the remaining 13 (65%) of 20 strains of CBSV species had T and A residues respectively at 31st amino acid position of the protein.

1 (5%) (F13S1) and 19 (95%) of 20 strains of UCBSV species had L and M residues respectively at 40th amino acid position of the protein whilst 10 (50%) and 10 (50%) strains of 20 strains of CBSV species had L and M residues respectively at 40th amino acid position of the protein

At 43rd amino acid position of 6K2 protein 18 (90%), 1 (5%) (F13S1) and 1 (5%) (MLB3) of 20 strains of UCBSV species had S, N and D residues, respectively whilst 18 (90%) and 2 (10%) (Kor 6 and CBSV_Nama_Moz_4_14) strains of 20 CBSV species had N and E residues respectively.

At 46th amino acid position of the protein; 18 (90%), 1 (5%) (F13S1) and 1 (5%) (MLB3) of 20 strains of UCBSV species had V, S, and I

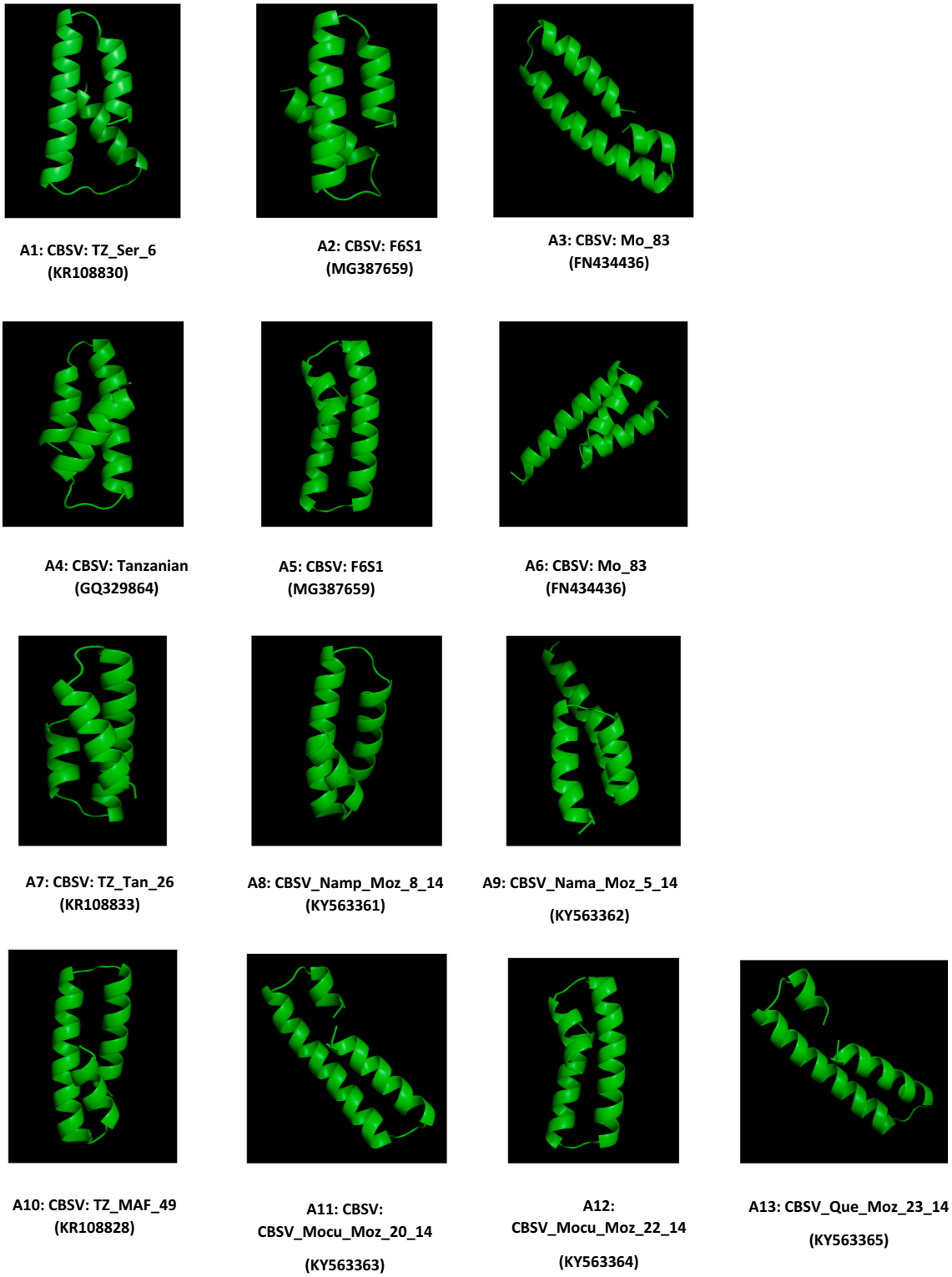

Fig. 3. Ab initio 3D structure models of 6K2 proteins of 20 CBSV (A1-A20) and 20 UCBSV (A21-A40) strains.

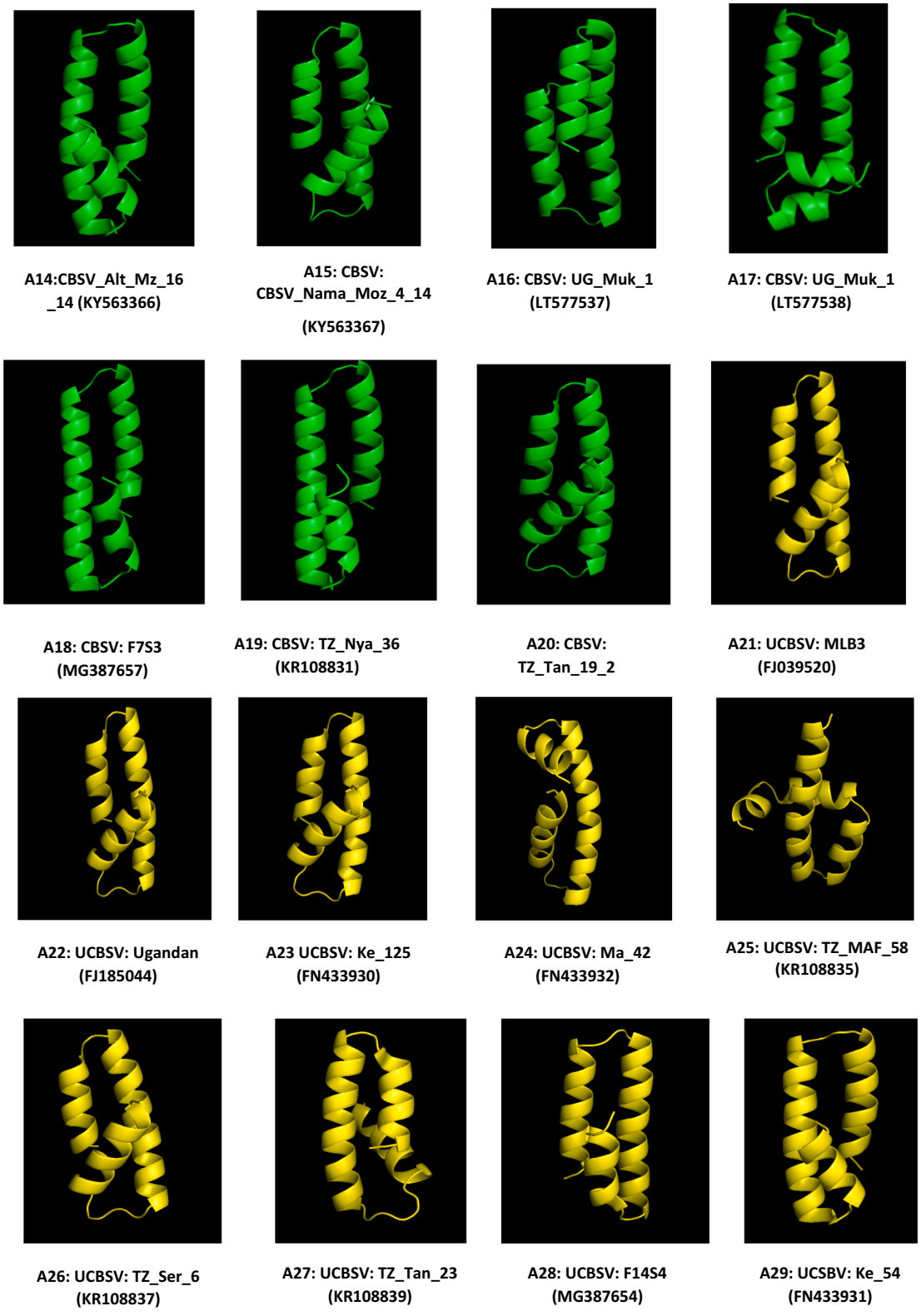

Fig. 3. (*continued*).

Fig. 3. (*continued*).

residues, respectively whereas 20 (100%) strains of CBSV species had S residues.

At 48th amino acid position of the protein; 19 (95%) and 1 (5%) (F13S1) of 20 strains of UCBSV species had L and V residues, respectively whereas 20(100%) strains of CBSV species had V residues.

At 49th amino acid position of the protein; 19 (95%) and 1 (5%) (TZ_Tan_19_2) of 20 strains of CBSV species had V and I residues, respectively whereas 20(100%) strains of UCBSV species had V residues.

At 50th amino acid position of the protein; 19 (95%) and 1 (5%) (Mo_83) of 20 strains of CBSV species had E and G residues, respectively whereas 20(100%) strains of UCBSV species had E residues.

At 51th amino acid position of the protein; 19 (95%) and 1 (5%) (F13S1) of 20 strains of UCBSV species had K and R residues, respectively whereas 16 (80%) and 4 (20%) (Kor 6, CBSV_Nama_-Moz_4_14, CBSV_Mocu_Moz_22_14 and CBSV_Que_Moz_23_14) of 20 strains of CBSV species had R and K residues.

4.5. Types of mutations influencing conformation of 3D-structure and function of 6K2 proteins in CBSV and UCBSV species

Among 17 genetically variable amino acid positions of the 6K2 protein, only substitution amino acid mutations were evidently observed between 20 CBSV and 20 UCBSV species.

Between CBSV and UCBSV species, diverse types of substitution mutations [N↔S, R↔K, S↔P, L↔I, L↔V, I↔V, M↔L, (L↔A, S↔A, and L↔S,), A↔T, (E↔S, E↔N, E↔D and N↔D), (S↔V and S↔I,) and E↔G]

Table 7
Results of unrefined pairwise alignment between 3D structures of 6K2 between and within 20 CBSV and 20 UCBSV species.

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
FJ039520	KY563367	86.54	241	782	2.015	52 vs 52
FJ039520	GU563327	84.62	239	779	8.176	52 vs 52
FJ039520	KR108829	82.69	237	773	12.207	52 vs 52
FJ039520	KR108832	80.77	234	770	1.798	52 vs 52
FJ039520	KR108831	80.77	232	764	8.333	52 vs 52
FJ039520	MG387657	80.77	231	768	7.583	52 vs 52
FJ039520	LT577538	80.77	231	768	7.532	52 vs 52
FJ039520	LT577537	80.77	231	768	7.945	52 vs 52
FJ039520	KR108833	80.77	231	768	8.943	52 vs 52
FJ039520	KR108830	80.77	231	768	4.039	52 vs 52
FJ039520	GQ329864	80.77	231	768	2.39	52 vs 52
FJ039520	MG387659	80.77	231	768	1.56	52 vs 52
FJ039520	KY563365	78.85	231	769	8.954	52 vs 52
FJ039520	KY563364	78.85	231	769	8.953	52 vs 52
FJ039520	KY563366	78.85	229	763	7.881	52 vs 52
FJ039520	KY563363	78.85	229	763	8.614	52 vs 52
FJ039520	KY563362	78.85	229	763	11.821	52 vs 52
FJ039520	KY563361	78.85	229	763	8.975	52 vs 52
FJ039520	KR108828	78.85	229	763	7.981	52 vs 52
FJ039520	FN434436	76.92	222	753	8.478	52 vs 52
FJ185044	KY563367	84.62	233	773	2.078	52 vs 52
FJ185044	GU563327	82.69	231	770	8.43	52 vs 52
FJ185044	KR108829	80.77	231	763	12.06	52 vs 52
FJ185044	KR108832	78.85	228	760	1.899	52 vs 52
FJ185044	KR108831	78.85	226	754	8.584	52 vs 52
FJ185044	MG387657	78.85	225	758	7.887	52 vs 52
FJ185044	LT577538	78.85	225	758	7.332	52 vs 52
FJ185044	LT577537	78.85	225	758	8.252	52 vs 52
FJ185044	KR108833	78.85	225	758	9.198	52 vs 52
FJ185044	KR108830	78.85	225	758	4.139	52 vs 52
FJ185044	GQ329864	78.85	225	758	1.824	52 vs 52
FJ185044	MG387659	78.85	225	758	1.674	52 vs 52
FJ185044	KY563365	76.92	225	759	9.288	52 vs 52
FJ185044	KY563364	76.92	225	759	9.324	52 vs 52
FJ185044	KY563366	76.92	223	753	8.092	52 vs 52
FJ185044	KY563363	76.92	223	753	8.925	52 vs 52
FJ185044	KY563362	76.92	223	753	11.654	52 vs 52
FJ185044	KY563361	76.92	223	753	9.207	52 vs 52
FJ185044	KR108828	76.92	223	753	8.268	52 vs 52
FJ185044	FN434436	75	216	743	8.748	52 vs 52
FN433930	KY563367	84.62	234	776	1.764	52 vs 52
FN433930	GU563327	82.69	232	773	8.303	52 vs 52
FN433930	KR108829	80.77	232	766	12.048	52 vs 52
FN433930	KR108832	78.85	229	763	1.674	52 vs 52
FN433930	KR108831	78.85	227	757	8.488	52 vs 52
FN433930	MG387657	78.85	226	761	7.74	52 vs 52
FN433930	LT577538	78.85	226	761	7.449	52 vs 52
FN433930	LT577537	78.85	226	761	8.183	52 vs 52
FN433930	KR108833	78.85	226	761	9.061	52 vs 52
FN433930	KR108830	78.85	226	761	4.208	52 vs 52
FN433930	GQ329864	78.85	226	761	2.263	52 vs 52
FN433930	MG387659	78.85	226	761	1.362	52 vs 52
FN433930	KY563365	76.92	226	762	9.209	52 vs 52
FN433930	KY563364	76.92	226	762	9.196	52 vs 52
FN433930	KY563366	76.92	224	756	8.019	52 vs 52
FN433930	KY563363	76.92	224	756	8.875	52 vs 52
FN433930	KY563362	76.92	224	756	11.672	52 vs 52
FN433930	KY563361	76.92	224	756	9.077	52 vs 52
FN433930	KR108828	76.92	224	756	8.136	52 vs 52
FN433930	FN434436	75	217	746	8.731	52 vs 52
FN433931	KY563367	84.62	234	776	8.532	52 vs 52
FN433931	GU563327	82.69	232	773	1.668	52 vs 52
FN433931	KR108829	80.77	232	766	10.974	52 vs 52
FN433931	KR108832	78.85	227	757	2.427	52 vs 52
FN433931	KR108831	78.85	229	763	8.65	52 vs 52
FN433931	MG387657	78.85	226	761	2.321	52 vs 52
FN433931	LT577538	78.85	226	761	6.693	52 vs 52
FN433931	LT577537	78.85	226	761	4.531	52 vs 52
FN433931	KR108833	78.85	226	761	3.192	52 vs 52
FN433931	KR108830	78.85	226	761	7.311	52 vs 52
FN433931	GQ329864	78.85	226	761	8.208	52 vs 52
FN433931	MG387659	78.85	226	761	8.183	52 vs 52
FN433931	KY563365	76.92	226	762	5.345	52 vs 52
FN433931	KY563364	76.92	226	762	5.457	52 vs 52

(*continued on next page*)

Table 7 (*continued*)

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
FN433931	KY563366	76.92	224	756	1.335	52 vs 52
FN433931	KY563363	76.92	224	756	5.135	52 vs 52
FN433931	KY563362	76.92	224	756	10.76	52 vs 52
FN433931	KY563361	76.92	224	756	3.403	52 vs 52
FN433931	KR108828	76.92	224	756	1.345	52 vs 52
FN433931	FN434436	75	217	746	4.999	52 vs 52
FN433932	KY563367	82.69	229	772	7.463	52 vs 52
FN433932	GU563327	78.85	232	766	10.974	52 vs 52
FN433932	KR108829	80.77	227	769	6.309	52 vs 52
FN433932	KR108832	76.92	224	759	7.424	52 vs 52
FN433932	KR108831	76.92	222	753	6.781	52 vs 52
FN433932	MG387657	76.92	221	757	5.897	52 vs 52
FN433932	LT577538	76.92	221	757	7.918	52 vs 52
FN433932	LT577537	76.92	221	757	6.643	52 vs 52
FN433932	KR108833	76.92	221	757	7.112	52 vs 52
FN433932	KR108830	76.92	221	757	7.611	52 vs 52
FN433932	GQ329864	76.92	221	757	7.671	52 vs 52
FN433932	MG387659	76.92	221	757	7.001	52 vs 52
FN433932	KY563365	75	221	758	6.688	52 vs 52
FN433932	KY563364	75	221	758	6.541	52 vs 52
FN433932	KY563366	75	219	752	6.103	52 vs 52
FN433932	KY563363	75	219	752	6.71	52 vs 52
FN433932	KY563362	75	219	752	12.074	52 vs 52
FN433932	KY563361	75	219	752	7.234	52 vs 52
FN433932	KR108828	75	219	752	6.1	52 vs 52
FN433932	FN434436	73.08	212	742	6.459	52 vs 52
FN333933	KY563367	82.69	229	772	2.088	52 vs 52
FN333933	GU563327	78.85	227	762	11.936	52 vs 52
FN333933	KR108829	80.77	227	769	8.734	52 vs 52
FN333933	KR108832	76.92	224	759	7.424	52 vs 52
FN333933	KR108831	76.92	222	753	6.781	52 vs 52
FN333933	MG387657	76.92	221	757	8.18	52 vs 52
FN333933	LT577538	76.92	221	757	7.692	52 vs 52
FN333933	LT577537	76.92	221	757	8.686	52 vs 52
FN333933	KR108833	76.92	221	757	9.403	52 vs 52
FN333933	KR108830	76.92	221	757	4.43	52 vs 52
FN333933	GQ329864	76.92	221	757	2.487	52 vs 52
FN333933	MG387659	76.92	221	757	1.925	52 vs 52
FN333933	KY563365	75	221	758	9.861	52 vs 52
FN333933	KY563364	75	221	758	8.881	52 vs 52
FN333933	KY563366	75	219	752	8.43	52 vs 52
FN333933	KY563363	75	219	752	9.5	52 vs 52
FN333933	KY563362	75	219	752	11.523	52 vs 52
FN333933	KY563361	75	219	752	9.347	52 vs 52
FN333933	KR108828	75	219	752	8.565	52 vs 52
FN333933	FN434436	73.08	212	742	9.361	52 vs 52
FN434109	KY563367	84.62	234	776	8.676	52 vs 52
FN434109	GU563327	82.69	232	773	2.049	52 vs 52
FN434109	KR108829	80.77	232	766	10.742	52 vs 52
FN434109	KR108832	78.85	229	763	8.794	52 vs 52
FN434109	KR108831	78.85	227	757	2.731	52 vs 52
FN434109	MG387657	78.85	226	761	2.935	52 vs 52
FN434109	LT577538	78.85	226	761	6.802	52 vs 52
FN434109	LT577537	78.85	226	761	4.682	52 vs 52
FN434109	KR108833	78.85	226	761	3.083	52 vs 52
FN434109	KR108830	78.85	226	761	7.568	52 vs 52
FN434109	GQ329864	78.85	226	761	8.321	52 vs 52
FN434109	MG387659	78.85	226	761	8.347	52 vs 52
FN434109	KY563365	76.92	226	762	5.801	52 vs 52
FN434109	KY563364	76.92	226	762	5.903	52 vs 52
FN434109	KY563366	76.92	224	756	1.907	52 vs 52
FN434109	KY563363	76.92	224	756	5.564	52 vs 52
FN434109	KY563362	76.92	224	756	10.501	52 vs 52
FN434109	KY563361	76.92	224	756	3.066	52 vs 52
FN434109	KR108828	76.92	224	756	1.973	52 vs 52
FN434109	FN434436	75	217	746	5.405	52 vs 52
HG965222	KY563367	84.62	234	776	10.695	52 vs 52
HG965222	GU563327	82.69	232	773	6.44	52 vs 52
HG965222	KR108829	80.77	232	766	8.708	52 vs 52
HG965222	KR108832	78.85	229	763	10.47	52 vs 52
HG965222	KR108831	78.85	227	757	7.147	52 vs 52
HG965222	MG387657	78.85	226	761	10.527	52 vs 52
HG965222	LT577538	78.85	226	761	7.492	52 vs 52
HG965222	LT577537	78.85	226	761	9.122	52 vs 52
HG965222	KR108833	78.85	226	761	7.708	52 vs 52
HG965222	KR108830	78.85	226	761	5.861	52 vs 52

(*continued on next page*)

Table 7 (*continued*)

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
HG965222	GQ329864	78.85	226	761	10.427	52 vs 52
HG965222	MG387659	78.85	226	761	10.164	52 vs 52
HG965222	KY563365	76.92	226	762	9.926	52 vs 52
HG965222	KY563364	76.92	226	762	9.892	52 vs 52
HG965222	KY563366	76.92	224	756	6.54	52 vs 52
HG965222	KY563363	76.92	224	756	9.725	52 vs 52
HG965222	KY563362	76.92	224	756	8.697	52 vs 52
HG965222	KY563361	76.92	224	756	5.677	52 vs 52
HG965222	KR108828	76.92	224	756	6.503	52 vs 52
HG965222	FN434436	75	217	746	9.504	52 vs 52
HM181930	KY563367	84.62	233	773	2.452	52 vs 52
HM181930	GU563327	82.69	231	770	8.311	52 vs 52
HM181930	KR108829	80.77	231	763	12.262	52 vs 52
HM181930	KR108832	78.85	226	754	8.482	52 vs 52
HM181930	KR108831	78.85	228	760	2.097	52 vs 52
HM181930	MG387657	78.85	225	758	7.689	52 vs 52
HM181930	LT577538	78.85	225	758	7.403	52 vs 52
HM181930	LT577537	78.85	225	758	8.139	52 vs 52
HM181930	KR108833	78.85	225	758	8.966	52 vs 52
HM181930	KR108830	78.85	225	758	4.137	52 vs 52
HM181930	GQ329864	78.85	225	758	2.166	52 vs 52
HM181930	MG387659	78.85	225	758	2.134	52 vs 52
HM181930	KY563365	76.92	225	759	9.416	52 vs 52
HM181930	KY563364	76.92	225	759	9.451	52 vs 52
HM181930	KY563366	76.92	223	753	7.958	52 vs 52
HM181930	KY563363	76.92	223	753	8.981	52 vs 52
HM181930	KY563362	76.92	223	753	11.843	52 vs 52
HM181930	KY563361	76.92	223	753	8.948	52 vs 52
HM181930	KR108828	76.92	223	753	8.082	52 vs 52
HM181930	FN434436	75	216	743	8.748	52 vs 52
KR108835	KY563367	86.54	20	324	8.351	52 vs 52
KR108835	GU563327	84.62	20.5	324	7.366	52 vs 52
KR108835	KR108829	82.69	20	91	1.817	52 vs 52
KR108835	KR108832	80.77	20	91	1.78	52 vs 52
KR108835	KR108831	80.77	20	91	2.359	52 vs 52
KR108835	MG387657	80.77	20	91	1.624	52 vs 52
KR108835	LT577538	80.77	20	91	1.768	52 vs 52
KR108835	LT577537	80.77	228	769	7.81	52 vs 52
KR108835	KR108833	80.77	20	91	1.885	52 vs 52
KR108835	KR108830	80.77	20	91	1.254	52 vs 52
KR108835	GQ329864	80.77	20	91	2.27	52 vs 52
KR108835	MG387659	80.77	20	91	2.34	52 vs 52
KR108835	KY563365	78.85	20	91	1.703	52 vs 52
KR108835	KY563364	78.85	20	91	1.722	52 vs 52
KR108835	KY563366	78.85	20	91	1.639	52 vs 52
KR108835	KY563363	78.85	20	91	1.712	52 vs 52
KR108835	KY563362	78.85	20	91	1.897	52 vs 52
KR108835	KY563361	78.85	20	91	1.712	52 vs 52
KR108835	KR108828	78.85	20	91	1.082	52 vs 52
KR108835	FN434436	76.92	20	91	1.777	52 vs 52
KR108836	KY563367	86.54	236	784	1.915	52 vs 52
KR108836	GU563327	84.62	234	781	8.177	52 vs 52
KR108836	KR108829	82.69	234	774	12.384	52 vs 52
KR108836	KR108832	80.77	231	771	1.874	52 vs 52
KR108836	KR108831	80.77	229	765	8.283	52 vs 52
KR108836	MG387657	80.77	228	769	7.557	52 vs 52
KR108836	LT577538	80.77	228	769	7.251	52 vs 52
KR108836	LT577537	80.77	228	769	781	52 vs 52
KR108836	KR108833	80.77	228	769	8.828	52 vs 52
KR108836	KR108830	80.77	228	769	4.138	52 vs 52
KR108836	GQ329864	80.77	228	769	2.204	52 vs 52
KR108836	MG387659	80.77	228	769	1.374	52 vs 52
KR108836	KY563365	78.85	228	770	9.087	52 vs 52
KR108836	KY563364	78.85	228	770	9.108	52 vs 52
KR108836	KY563366	78.85	226	764	7.8	52 vs 52
KR108836	KY563363	78.85	226	764	8.666	52 vs 52
KR108836	KY563362	78.85	226	764	12.038	52 vs 52
KR108836	KY563361	78.85	226	764	8.903	52 vs 52
KR108836	KR108828	78.85	226	764	7.91	52 vs 52
KR108836	FN434436	76.92	219	754	8.53	52 vs 52
KR108837	KY563367	86.54	242	782	1.966	52 vs 52
KR108837	GU563327	84.62	240	779	8.354	52 vs 52
KR108837	KR108829	80.77	237	769	1.864	52 vs 52
KR108837	KR108832	80.77	235	763	8.526	52 vs 52
KR108837	KR108831	80.77	234	767	7.822	52 vs 52
KR108837	MG387657	80.77	234	767	7.455	52 vs 52

(*continued on next page*)

Table 7 (*continued*)

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
KR108837	LT577538	80.77	234	767	8.042	52 vs 52
KR108837	LT577537	80.77	234	767	8.999	52 vs 52
KR108837	KR108833	80.77	234	767	4.165	52 vs 52
KR108837	KR108830	80.77	234	767	2.182	52 vs 52
KR108837	GQ329864	80.77	234	767	1.639	52 vs 52
KR108837	MG387659	78.85	234	768	9.289	52 vs 52
KR108837	KY563365	78.85	234	768	9.305	52 vs 52
KR108837	KY563364	78.85	232	762	8.052	52 vs 52
KR108837	KY563366	78.85	232	762	8.875	52 vs 52
KR108837	KY563363	78.85	232	762	11.862	52 vs 52
KR108837	KY563362	78.85	232	762	9.065	52 vs 52
KR108837	KY563361	78.85	232	762	8.173	52 vs 52
KR108837	KR108828	78.85	227	763	12.153	52 vs 52
KR108837	FN434436	76.92	225	752	8.72	52 vs 52
KR108838	KY563367	86.54	242	782	5.748	52 vs 52
KR108838	GU563327	84.62	240	779	9.004	52 vs 52
KR108838	KR108829	80.77	237	769	5.682	52 vs 52
KR108838	KR108832	80.77	235	763	9.243	52 vs 52
KR108838	KR108831	80.77	234	767	8.651	52 vs 52
KR108838	MG387657	80.77	234	767	6.74	52 vs 52
KR108838	LT577538	80.77	234	767	8.119	52 vs 52
KR108838	LT577537	80.77	234	767	8.805	52 vs 52
KR108838	KR108833	80.77	234	767	6.545	52 vs 52
KR108838	KR108830	80.77	234	767	5.483	52 vs 52
KR108838	GQ329864	80.77	234	767	5.473	52 vs 52
KR108838	MG387659	78.85	234	768	9.311	52 vs 52
KR108838	KY563365	78.85	234	768	9.378	52 vs 52
KR108838	KY563364	78.85	232	762	8.529	52 vs 52
KR108838	KY563366	78.85	232	762	8.892	52 vs 52
KR108838	KY563363	78.85	232	762	12.516	52 vs 52
KR108838	KY563362	78.85	232	762	8.859	52 vs 52
KR108838	KY563361	78.85	232	762	8.746	52 vs 52
KR108838	KR108828	78.85	227	763	12.698	52 vs 52
KR108838	FN434436	76.92	225	752	8.74	52 vs 52
KR108839	KY563367	86.54	239	781	1.732	52 vs 52
KR108839	GU563327	84.62	237	778	8.309	52 vs 52
KR108839	KR108829	82.69	237	771	12.102	52 vs 52
KR108839	KR108832	80.77	234	768	1.923	52 vs 52
KR108839	KR108831	80.77	232	762	8.456	52 vs 52
KR108839	MG387657	80.77	231	766	7.782	52 vs 52
KR108839	LT577538	80.77	231	766	7.314	52 vs 52
KR108839	LT577537	80.77	231	766	8.061	52 vs 52
KR108839	KR108833	80.77	231	766	9.014	52 vs 52
KR108839	KR108830	80.77	231	766	4.301	52 vs 52
KR108839	GQ329864	80.77	231	766	2.109	52 vs 52
KR108839	MG387659	80.77	231	766	1.387	52 vs 52
KR108839	KY563365	78.85	231	767	9.243	52 vs 52
KR108839	KY563364	78.85	231	767	9.299	52 vs 52
KR108839	KY563366	78.85	229	761	7.947	52 vs 52
KR108839	KY563363	78.85	229	761	8.862	52 vs 52
KR108839	KY563362	78.85	229	761	11.736	52 vs 52
KR108839	KY563361	78.85	229	761	9.057	52 vs 52
KR108839	KR108828	78.85	229	761	8.118	52 vs 52
KR108839	FN434436	76.92	217	746	8.541	52 vs 52
LT577539	KY563367	84.62	234	776	1.939	52 vs 52
LT577539	GU563327	82.69	232	773	8.213	52 vs 52
LT577539	KR108829	80.77	232	766	12.19	52 vs 52
LT577539	KR108832	78.85	229	763	1.796	52 vs 52
LT577539	KR108831	78.85	227	757	8.327	52 vs 52
LT577539	MG387657	78.85	226	761	7.68	52 vs 52
LT577539	LT577538	78.85	226	761	7.3	52 vs 52
LT577539	LT577537	78.85	226	761	7.906	52 vs 52
LT577539	KR108833	78.85	226	761	8.923	52 vs 52
LT577539	KR108830	78.85	226	761	4.049	52 vs 52
LT577539	GQ329864	78.85	226	761	1.993	52 vs 52
LT577539	MG387659	78.85	226	761	1.536	52 vs 52
LT577539	KY563365	76.92	226	762	9.091	52 vs 52
LT577539	KY563364	76.92	226	762	9.137	52 vs 52
LT577539	KY563366	76.92	224	756	7.856	52 vs 52
LT577539	KY563363	76.92	224	756	8.727	52 vs 52
LT577539	KY563362	76.92	224	756	11.809	52 vs 52
LT577539	KY563361	76.92	224	756	8.975	52 vs 52
LT577539	KR108828	76.92	224	756	8.028	52 vs 52
LT577539	FN434436	75	217	746	8.541	52 vs 52
MG307655	KY563367	84.62	234	776	7.424	52 vs 52
MG307655	GU563327	82.69	232	773	11.704	52 vs 52

(*continued on next page*)

Table 7 (*continued*)

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
MG307655	KR108829	80.77	232	766	10.034	52 vs 52
MG307655	KR108832	78.85	229	763	7.276	52 vs 52
MG307655	KR108831	78.85	227	757	11.608	52 vs 52
MG307655	MG387657	78.85	226	761	11.972	52 vs 52
MG307655	LT577538	78.85	226	761	9.736	52 vs 52
MG307655	LT577537	78.85	226	761	11.485	52 vs 52
MG307655	KR108833	78.85	226	761	10.878	52 vs 52
MG307655	KR108830	78.85	226	761	8.763	52 vs 52
MG307655	GQ329864	78.85	226	761	7.235	52 vs 52
MG307655	MG387659	78.85	226	761	7.538	52 vs 52
MG307655	KY563365	76.92	226	762	12.549	52 vs 52
MG307655	KY563364	76.92	226	762	12.565	52 vs 52
MG307655	KY563366	76.92	224	756	11.771	52 vs 52
MG307655	KY563363	76.92	224	756	12.448	52 vs 52
MG307655	KY563362	76.92	224	756	9.704	52 vs 52
MG307655	KY563361	76.92	224	756	10.96	52 vs 52
MG307655	KR108828	76.92	224	756	11.655	52 vs 52
MG307655	FN434436	75	217	746	12.239	52 vs 52
MG387652	KY563367	84.62	250	802	9.328	52 vs 52
MG387652	GU563327	82.69	253	815	5.308	52 vs 52
MG387652	KR108829	80.77	258	822	11.136	52 vs 52
MG387652	KR108832	78.85	262	825	9.378	52 vs 52
MG387652	KR108831	78.85	265	835	5.384	52 vs 52
MG387652	MG387657	78.85	259	823	5.122	52 vs 52
MG387652	LT577538	78.85	259	823	7.393	52 vs 52
MG387652	LT577537	78.85	259	823	5.446	52 vs 52
MG387652	KR108833	78.85	259	823	6.683	52 vs 52
MG387652	KR108830	78.85	259	823	8.06	52 vs 52
MG387652	GQ329864	78.85	259	823	9.351	52 vs 52
MG387652	MG387659	78.85	259	823	8.975	52 vs 52
MG387652	KY563365	76.92	264	827	2.766	52 vs 52
MG387652	KY563364	76.92	264	827	2.814	52 vs 52
MG387652	KY563366	76.92	268	837	5.24	52 vs 52
MG387652	KY563363	76.92	268	837	2.734	52 vs 52
MG387652	KY563362	76.92	268	837	11.111	52 vs 52
MG387652	KY563361	76.92	268	837	6.882	52 vs 52
MG387652	KR108828	76.92	268	837	5.26	52 vs 52
MG387652	FN434436	75	261	827	2.546	52 vs 52
MG387653	KY563367	86.54	242	782	1.728	52 vs 52
MG387653	GU563327	84.62	240	779	8.299	52 vs 52
MG387653	KR108829	80.77	237	769	1.573	52 vs 52
MG387653	KR108832	80.77	235	763	8.515	52 vs 52
MG387653	KR108831	80.77	234	767	7.748	52 vs 52
MG387653	MG387657	80.77	234	767	7.301	52 vs 52
MG387653	LT577538	80.77	234	767	8.099	52 vs 52
MG387653	LT577537	80.77	234	767	8.981	52 vs 52
MG387653	KR108833	80.77	234	767	4.16	52 vs 52
MG387653	KR108830	80.77	234	767	2.163	52 vs 52
MG387653	GQ329864	80.77	234	767	1.528	52 vs 52
MG387653	MG387659	78.85	234	768	9.218	52 vs 52
MG387653	KY563365	78.85	234	768	9.238	52 vs 52
MG387653	KY563364	78.85	232	762	7.983	52 vs 52
MG387653	KY563366	78.85	232	762	8.856	52 vs 52
MG387653	KY563363	78.85	232	762	11.927	52 vs 52
MG387653	KY563362	78.85	232	762	8.985	52 vs 52
MG387653	KY563361	78.85	232	762	8.093	52 vs 52
MG387653	KR108828	78.85	227	763	12.2	52 vs 52
MG387653	FN434436	76.92	225	752	8.667	52 vs 52
MG387654	KY563367	86.54	242	782	4.876	52 vs 52
MG387654	GU563327	84.62	240	779	7.382	52 vs 52
MG387654	KR108829	80.77	237	769	4.722	52 vs 52
MG387654	KR108832	80.77	235	763	7.328	52 vs 52
MG387654	KR108831	80.77	234	767	6.686	52 vs 52
MG387654	MG387657	80.77	234	767	6.025	52 vs 52
MG387654	LT577538	80.77	234	767	6.672	52 vs 52
MG387654	LT577537	80.77	234	767	7.641	52 vs 52
MG387654	KR108833	80.77	234	767	4.289	52 vs 52
MG387654	KR108830	80.77	234	767	4.138	52 vs 52
MG387654	GQ329864	80.77	234	767	4.537	52 vs 52
MG387654	MG387659	78.85	234	768	8.073	52 vs 52
MG387654	KY563365	78.85	234	768	8.266	52 vs 52
MG387654	KY563364	78.85	232	762	6.85	52 vs 52
MG387654	KY563366	78.85	232	762	7.535	52 vs 52
MG387654	KY563363	78.85	232	762	13.563	52 vs 52
MG387654	KY563362	78.85	232	762	7.812	52 vs 52
MG387654	KY563361	78.85	232	762	6.983	52 vs 52

(*continued on next page*)

Table 7 (*continued*)

Query sequence	Subject sequence	% identity of protein	Match Score	Executive align (aligned atoms)	RMSD	Aligned aa residues
MG387654	KR108828	78.85	227	763	13.814	52 vs 52
MG387654	FN434436	76.92	225	752	7.496	52 vs 52
MG387656	KY563367	84.62	234	776	11.928	52 vs 52
MG387656	GU563327	82.69	232	773	10.545	52 vs 52
MG387656	KR108829	80.77	232	766	1.986	52 vs 52
MG387656	KR108832	78.85	229	763	11.671	52 vs 52
MG387656	KR108831	78.85	227	757	10.798	52 vs 52
MG387656	MG387657	78.85	226	761	11.498	52 vs 52
MG387656	LT577538	78.85	226	761	13.662	52 vs 52
MG387656	LT577537	78.85	226	761	11.012	52 vs 52
MG387656	KR108833	78.85	226	761	10.18	52 vs 52
MG387656	KR108830	78.85	226	761	13.132	52 vs 52
MG387656	GQ329864	78.85	226	761	12.017	52 vs 52
MG387656	MG387659	78.85	226	761	12.027	52 vs 52
MG387656	KY563365	76.92	226	762	11.419	52 vs 52
MG387656	KY563364	76.92	226	762	11.469	52 vs 52
MG387656	KY563366	76.92	224	756	10.743	52 vs 52
MG387656	KY563363	76.92	224	756	11.579	52 vs 52
MG387656	KY563362	76.92	224	756	1.601	52 vs 52
MG387656	KY563361	76.92	224	756	10.079	52 vs 52
MG387656	KR108828	76.92	224	756	10.635	52 vs 52
MG387656	FN434436	75	217	746	11.349	52 vs 52

were evident at [(12, 16 and 43), (15 and 51), 20, (21, 23 and 25), (23, 25 and 48), (23, 30 and 49), (25, and 40), 29, 31, (43), (46) and 50], amino acid positions of the 6K2 protein, correspondingly.

In terms of physical chemical properties N↔S, R↔K, (L↔I, L↔V, I↔V, M↔L and L↔A), (S↔P, S↔A, L↔S, A↔T, S↔V and S↔I), E↔D, (E↔S, E↔N and N↔D) and E↔G substitutions occurred between; polar, positively charged, hydrophobic (non-polar), polar and hydrophobic, negatively charged, polar and negatively charged, negatively charged and hydrophobic residues, respectively.

4.6. Comparative molecular evolution of 6K2 protein between CBSV and UCBSV species

Phylogenetic analysis of 6K2 protein by PhyML-Maximum Likelihood in UGENE v 40.1 revealed that, the compared 20 CBSV and 20 UCBSV species share common ancestral origin, despite being clustered in separate clades (Fig. 7). A slight difference in branch length (0–0.035 and 0–0.036), was observed between CBSV and UCBSV species, respectively. UCBSV species had a slight higher number of substitutions compared to CBSV species, as represented by tree's branch lengths. TZ_Ser 6 and MLB3 strains of CBSV and UCBSV species, correspondingly had highest substitution mutations.

Five and four clades of the CBSV and UCBSV species, respectively were obtained at 0–98.2% interval of bootstrap values. Regardless the species, clades obtained at 0% bootstrap value, indicates that, the clades are distantly related to other species in the tree owing to a lot of substitutional mutations. Clades constructed at 98.2% bootstrap value share similarity and are evolutionary related.

4.7. Comparative physical-chemical properties between CBSV: TZ_MAF 49 (KR108828) and UCBSV: TZ_MAF 58 (KR108835) strains

Genome-wise analysis of physical chemical properties of CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, revealed that 9 (P1, P3, 6K1, CI, NIa, NIb, VPg, Ham 1 and CP) proteins are hydrophilic in nature reflected by varying negative GRAVY values. However, UCBSV: TZ_MAF 58 and CBSV: TZ_MAF 49 strain expressed hydrophobic and hydrophilic forms of 6K2 proteins, respectively.

In terms of extinction coefficient, variation was observed for Ham 1, VPg, NIb, NIa, CI, P1 and P3 proteins with (22,460 and 16,960), (22,460 and 23,950), (80,790 and 85,260), (26,930 and 28,420), (58,330 and 56,840), (40,910 and 36,900), and (60,850 and 63,370) values for CBSV and UCBSV species, separately.

More details on how CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains differ in amino acid compositions are indicated in (Table 9) and (Table 10), correspondingly.

Other parameters such as Isoelectric Point (PI) and average molecular weight were more or less identical for all proteins between UCBSV: TZ_MAF 58 and CBSV: TZ_MAF 49 strains.

4.8. Comparative physical-chemical properties of 6K2 proteins expressed by CBSV versus UCBSV species

There's a significant difference in physical chemical properties of the 6K2 protein between CBSV and UCBSV species as shoown in (Table 11) and (Table 12), correspondingly.

20 CBSV and 20 UCBSV strains had a Grand Average of Hydropathicity (GRAVY) range of (−0.15–0.277) and (−0.115–0.087), correspondingly. 6K2 proteins of only 1 (5%) (Mo_83) strain of CBSV species was hydrophobic with 0.277 GRAVY index value whilst the remaining 19 (95%) strains expressed the same protein in hydrophilic nature. Contrary to CBSV strains, 17 (85%) of UCBSV strains were hydrophobic whilst 3 (15%) were hydrophilic in nature.

Aliphatic indices of 20 CBSV and 20 UCBSV strains were (86.35–97.5) and (90–97.50), respectively.

Nonetheless, 6K2 proteins of both 20 and 20 strains of CBSV and UCBSV species express the basic 6K2 protein but with differed degreed of alkalinity. Range of isoelectric points (Bjellqvist et al.) of the former and latter were (6.36–8.18) and (6.33–8.08), correspondingly. 2 CBSV strains (CBSV_Nama_Moz_4_14 and Kor 6) expressed the least alkaline 6K2 protein with 6.36 isoelectric points whilst 1 strain (Mo_83) expressed the most basic protein with 8.18 isoelectric point. For the UCBSV species, only a single strain (MLB3) expressed the least basic 6K2 protein with 6.33 isoelectric point whilst the rest of 19 strains expressed

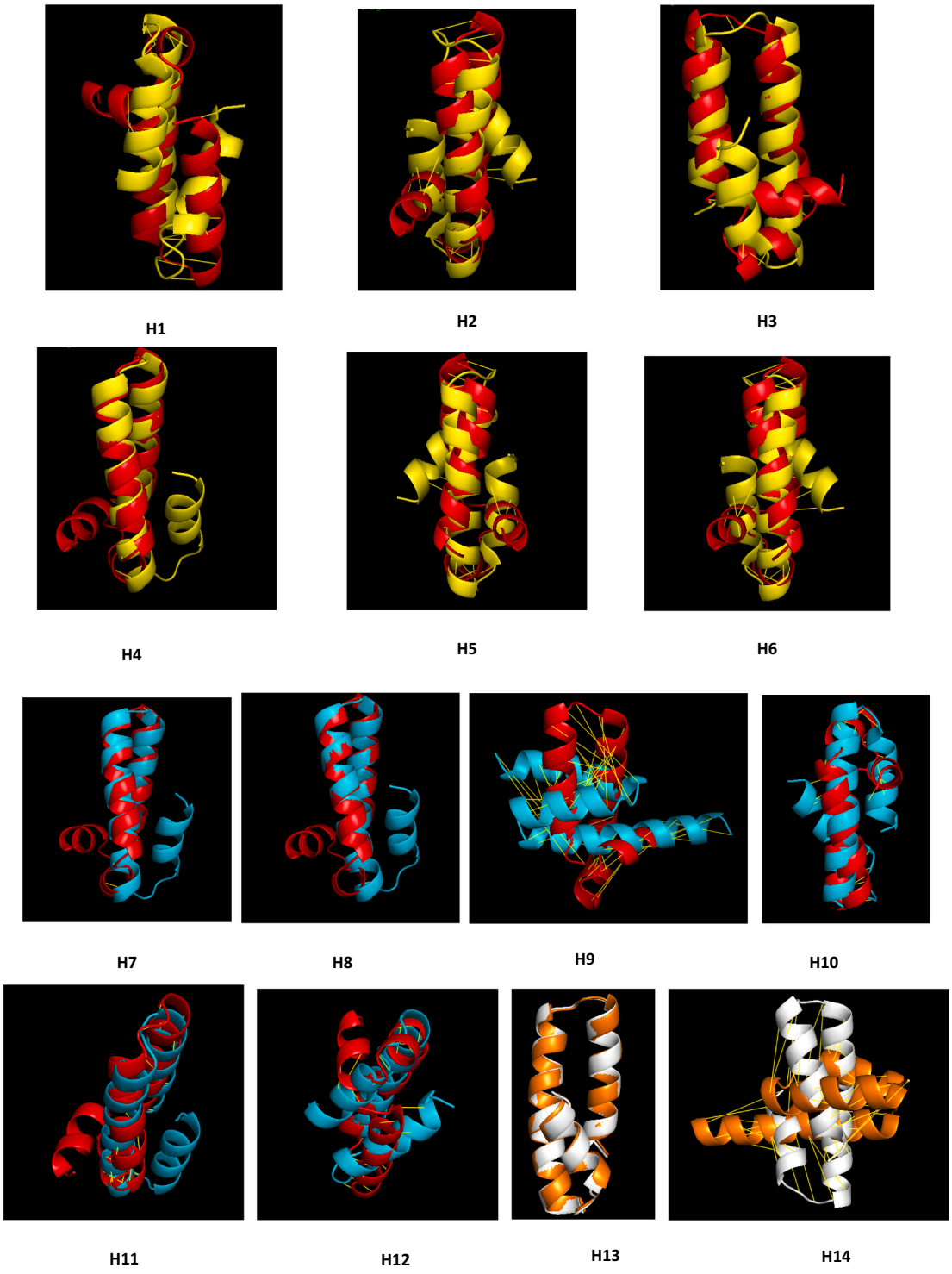

Fig. 4. Results of pairwise alignment of 3D structures of 6K2 proteins between and within strains of CBSV and UCBSV species (H1) CBSV: UG_Muk_4 (Red) and UCBSV: UG[Uganda:Namulonge:2004] (Yellow), (H2) CBSV: UG_Muk_4 (Red) and UCBSV: TZ_MAF_51 (Yellow), (H3) CBSV: UG_Muk_4 (Red) and UCBSV:F14S4 (Yellow), (H4) CBSV: UG_Muk_4 (Red) and UCBV: Ke_54 (Yellow), (H5) CBSV: UG_Muk_4 (Red) and UCBSV:MLB3 (Yellow), (H6) CBSV: UG_Muk_4 (Red) and UCBSV: TZ_Ser_6 strain (Yellow), (H7) CBSV: UG_Muk_4 (Red) and CBSV: TZ_MAF_49 (Cyan), (H8) CBSV: UG_Muk_4 (Red) and CBSV: CBSV_Nama_Moz_5_14 (Cyan), (H9) CBSV: UG_Muk_4 (Red) and CBSV: F6S1 (Cyan), (H10) CBSV: UG_Muk_4 (Red) and CBSV: Tan_70 (Cyan), (H11) CBSV: UG_Muk_4 (Red) and CBSV: Tanzanian (Cyan), (H12) CBSV: UG_Muk_4 (Red) and CBSV: CBSV_Nama_Moz_4_14 (Cyan), (H13) CBSV: TZ_MAF_49 (White) and CBSV: Kor 6 strain (Orange), (H14) CBSV: TZ_MAF_49 (White) and CBSV: CBSV_Nama_Moz_5_14 (Orange), (H15) CBSV: TZ_MAF_49 (White) and CBSV: F6S1 (Orange), (H16) CBSV: TZ_MAF_49 (White) and CBSV: Tan_70 (Orange), (H17) CBSV: TZ_MAF_49 (White) and CBSV: Tanzanian (Orange), (H18) CBSV: TZ_MAF_49 (White) and CBSV: CBSV_Nama_Moz_4_14 (Orange), (H19) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: TZ_MAF_49 (Cyan), (H20) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: CBSV_Nama_Moz_5_14 (Cyan), (H21) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: F6S1 (Cyan), (H22) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: Tan_70 (Cyan), (H23) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: Tanzanian (Cyan), (H24) UCBSV: TZ_MAF_58 (Hot-pink) and CBSV: CBSV_Nama_Moz_4_14 (Cyan), (H25) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV: UG[Uganda: Namulonge:2004] (Yellow), (H26) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV:TZ_MAF_51 (Yellow), (H27) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV:F14S4 (Yellow), (H28) UCBSV: TZ_MAF_58 (Hot-pink) and UCBV: Ke_54 (Yellow), (H29) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV: MLB3 (Yellow), (H30) UCBSV:

TZ_MAF_58 (Hot-pink) and UCBSV: TZ_Ser_6 (Yellow), (H31) UCBSV: UG[Uganda:Namulonge:2004] (Purple) and UCBSV:TZ_MAF_51 (Wheat), (H32) UCBSV: UG [Uganda:Namulonge:2004] (Purple) and UCBSV:F14S4 (Wheat), (H33) UCBSV: UG[Uganda:Namulonge:2004] (Purple) and UCBV: Ke_54 (Wheat), (H34) UCBSV: UG[Uganda:Namulonge:2004] (Purple) and UCBSV:MLB3 (Wheat), (H35) UCBSV: UG[Uganda:Namulonge:2004] (Purple) and UCBSV: TZ_Ser_6 (Wheat), (H36) UCBSV: UG[Uganda:Namulonge:2004] (Purple) and UCBSV: F14S3 (Wheat), (H37) CBSV: UG_Muk_4 (Red) and UCBSV: TZ_MAF_58 (Deep-olive), (H38) CBSV: UG_Muk_4 (Red) and UCBSV-[UG; Kab; 07] (Deep-olive), (H39) CBSV: UG_Muk_4 (Red) and UCBSV: F12S1 (Deep-olive), (H40) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV-[UG; Kab; 07] (Gray-90), (H41) UCBSV: TZ_MAF_58 (Hot-pink) and UCBSV: F12S1 (Deep-teal), (H42) UCBSV: F12S1 (Deep-teal) and UCBSV-[UG; Kab; 07] (Gray-90), (H43) UCBSV: F14S3 (Yellow) and CBSV: TZ_MAF_49 (Cyan), (H44) UCBSV:TZ_MAF_51 (Yellow) and CBSV: Kor 6 (Cyan), (H45) UCBSV:F14S4 (Yellow) and CBSV: CBSV_Nama_Moz_5_14 (Cyan), (H46) UCBV: Ke_54 (Yellow) and CBSV: F6S1 (Cyan), (H47) UCBSV: MLB3 (Yellow) and CBSV: Tan_70 (Cyan) and (H48) UCBSV: TZ_Ser_6 (Yellow) and CBSV: Tanzanian (Cyan). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

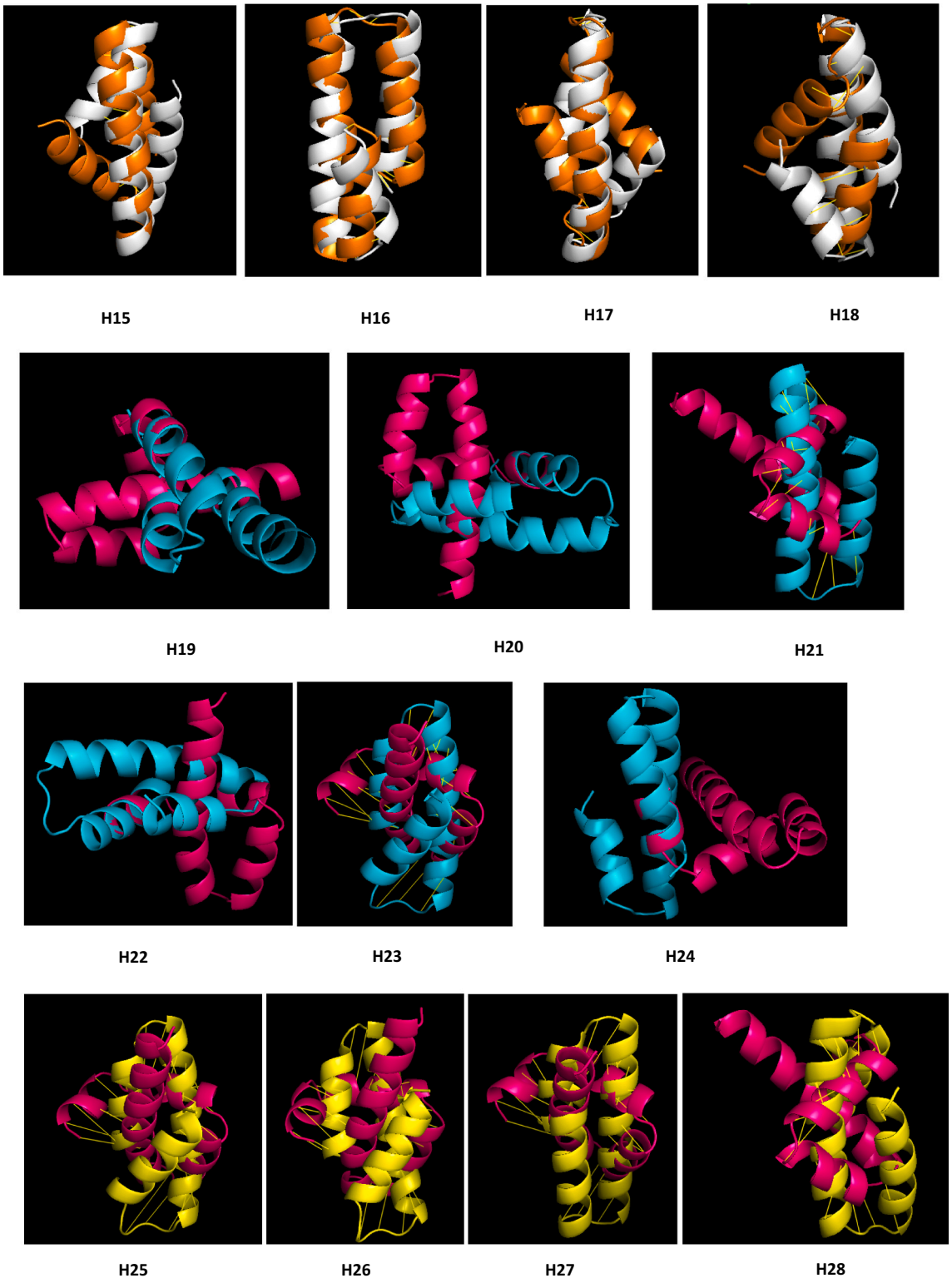

Fig. 4. (*continued*).

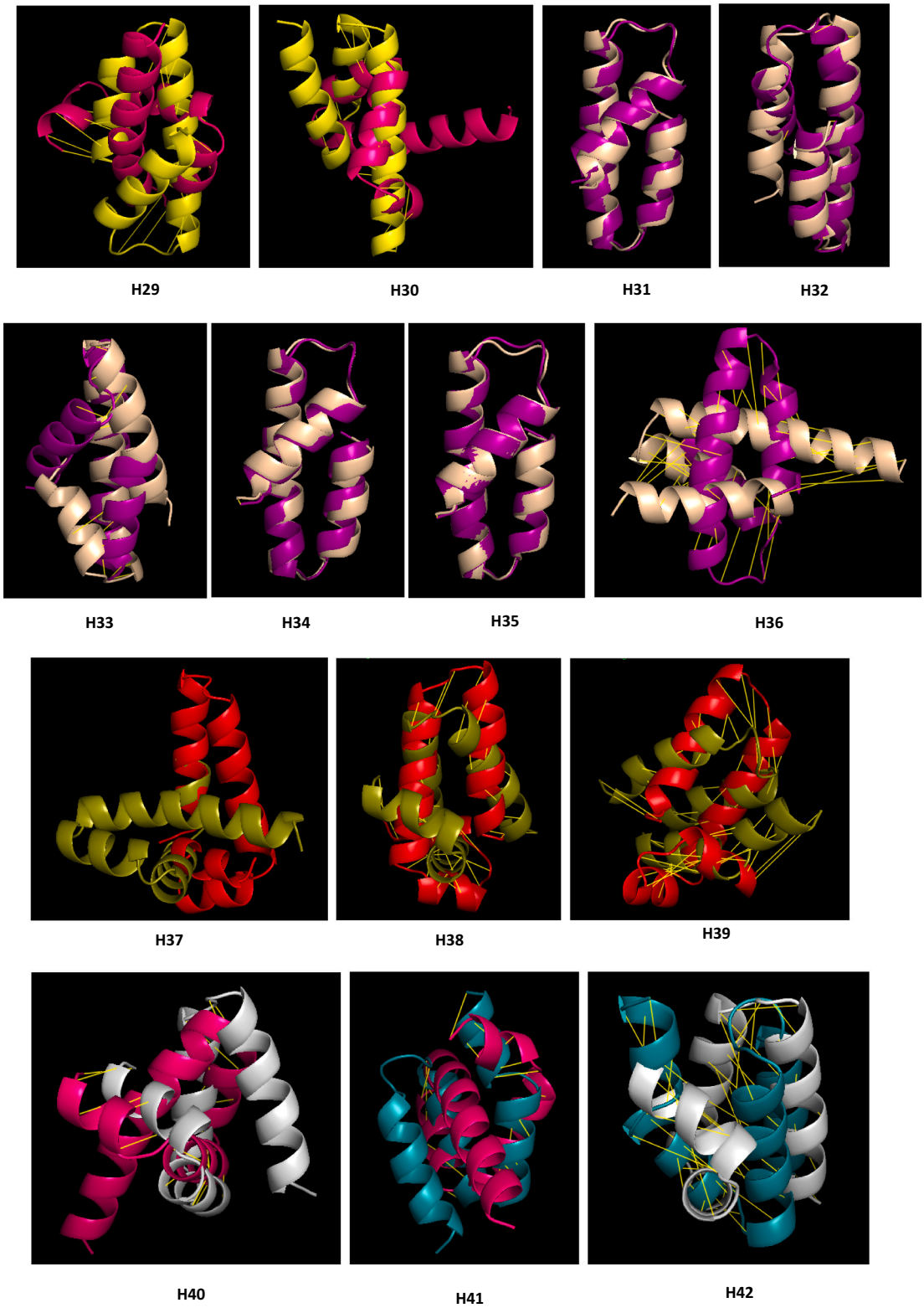

Fig. 4. (*continued*).

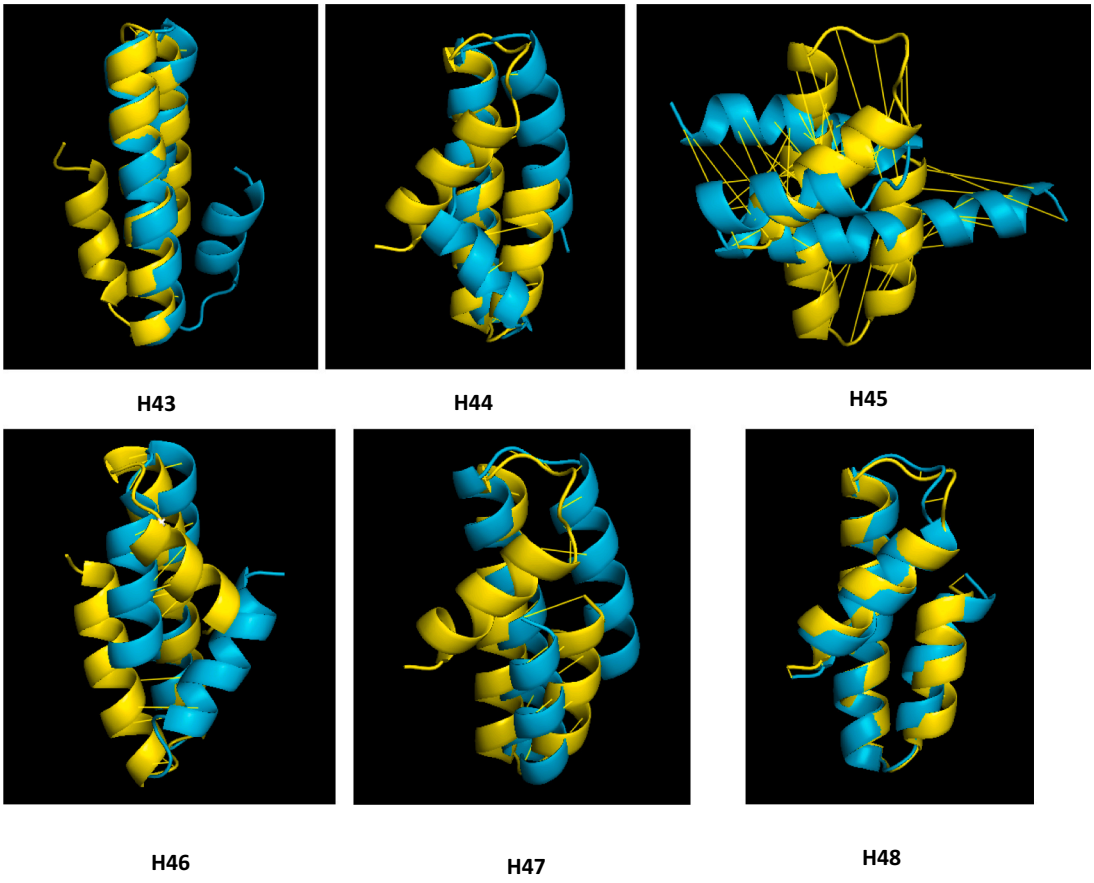

Fig. 4. (*continued*).

a more alkaline 6K2 protein with 8.08 isoelectric point.

There was a slight differences in molecular weights of 6K2 protein between CBSV and UCBSV species. Molecular weight range of 20 CBSV and 20 UCBSV strains were (5894.8036–5942.86) and (5847.8027–5961.9243), respectively.

5. Discussion

5.1. How disparity in polarity of the 6K2 protein influence tissue titers between CBSV and UCBSV species?

Basing on homology modelling results generated by SWISS-MODEL, 6K2 of both CBSV and UCBSV species was predicted to be a helical transmembrane protein, ErbB3 receptor tyrosine kinase of ErbB or HER family (Mineev et al., 2011). These findings are in agreement with study by Alicai et al. (2016) in which 6K2 protein was mentioned to associate with host's cell membrane and facilitate long distance systemic spread (cell-to-cell movement) of viruses.

Our analyses revealed that; majority strains of CBSV (95%) and UCBSV (85%) species, expressed hydrophilic and hydrophobic 6K2 proteins respectively. The fact that 6K2 proteins of CBSV and UCBSV species have notable opposite polarity properties, explains why the two viruses also exhibit different tissue titers. It was mentioned in Alicai et al. (2016) that, CBSV species accumulate in tissues at higher titers and have faster replication rate than the UCBSV species.

Evidences of CBSV's tropism to phloematic tissues of cassava and its model tobacco plant were reported previously in two histopathological studies by Munganyinka et al. (2018) and Saggaf et al. (2019). With RNA scope In Situ Hybridization (ISH) techniques CBSV-Mo83 (DSMZ PV-0949) strain was proved to exhibit distinctive tropism to stem tissues in tobacco model plant (*Nacotiana rustica*) and cassava (*Manihot esculenta* 7 (TME7), where by large quantity of the viruses was observed in phloematic and non-phloematic tissues of the latter and former species, separately (Munganyinka et al., 2018). After conducting an Immunohistochemical (IHC) study, Saggaf et al. (2019) confirmed that, precipitated CBSV-infected cells were seen in the phloem and along the adaxial and abaxial epidermal cells all around the leaf blades and central veins, contrary to healthy cells. Furthermore, the fact that both CBSV and UCBSV species can be transmitted artificially by sap inoculation and nodal/top-cleft grafting, confirms that these viruses are systemically spread through phloematic vascular tissues (Maruthi et al., 2014; Amuge et al., 2017; Anjanappa et al., 2017).

From that standpoint, we hypothesize that; hydrophilic 6K2 proteins gives CBSV species survival advantage over the UCBV species because it facilitate faster systemic spread of the virus through vascular tissues (phloem) which the host (cassava) uses to distribute water and nutrients throughput the plant. UCBSV species uses the same vascular tissues as its distribution portal throughout the plant but, virus dissemination rate is slower compared to CBSV species owing to hydrophobic nature of its 6K2 protein. We postulate that; because CBSV species are distributed

Table 8
Results of refined pairwise alignment between 3D structures of 6K2 between and within CBSV and UCBSV species.

				RMSD per alignment cycle					Number of rejected atoms per alignment cycle						
Protein 1	Protein 2	Score	Aligned	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Matched	Executive
			atoms											atoms	RMSD
LT577538	HM181930	225	52 vs 52	5.49	4.96	None	None	None	2	2	None	None	None	48 vs 48	4.461
LT577538	KR108836	228	52 vs 52	5.51	4.98	None	None	None	2	2	None	None	None	48 vs 48	4.421
LT577538	MG387654	234	52 vs 52	4.39	3.34	3.22	3.1	2.99	3	1	1	1	2	44 vs 44	2.763
LT577538	FN433931	226	52 vs 52	6	4.9	3.89	2.94	2.25	3	3	3	2	2	39 vs 39	1.807
LT577538	FJ39520	231	52 vs 52	5.78	4.88	4.56	None	None	3	1	1	None	None	47 vs 47	4.393
LT577538	KR108837	234	52 vs 52	5.65	5.11	4.615	None	None	2	2	None	None	None	48 vs 48	4.615
LT577538	KR108828	259	52 vs 52	6.08	4.45	3.47	2.86	2.26	4	3	2	2	2	39 vs 39	1.921
LT577538	GU563327	250	52 vs 52	5.96	4.42	3.47	2.52	1.96	4	3	3	2	3	37 vs 37	1.483
LT577538	K563362	259	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	12.886
LT577538	MG387659	270	52 vs 52	5.64	5.08	None	None	None	2	2	None	None	None	48 vs 48	4.538
LT577538	FN434437	248	52 vs 52	6.51	5.49	4.8	4.12	3.48	3	2	3	2	1	41 vs 41	3.204
LT577538	GQ329864	270	52 vs 52	5.6	4.56	4.27	4.12	None	3	1	1	1	None	46 vs 46	3.962
KR108828	GU563327	248	52 vs 52	253	0.61	0.48	None	None	2	1	None	None	None	49 vs 49	0.464
KR108828	K563362	268	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	9.992
KR108828	MG387659	259	52 vs 52	7.14	6.1	5.48	4.5	3.68	3	2	4	3	2	38 vs 38	3.22
KR108828	FN434437	246	52 vs 52	2.44	None	None	None	None	1	None	None	None	None	51 vs 51	2.331
KR108828	GQ329864	259	52 vs 52	7.37	5.91	4.84	4.09	3.39	4	4	3	3	2	36 vs 36	2.884
KR108828	KY563367	250	52 vs 52	7.21	6.55	5.01	4.29	3.84	2	5	3	2	3	37 vs 37	3.192
KR108835	KR108828	20	52 vs 52	None	None	None	None	None	None	None	None	None	None	6 vs 6	0.63
KR108835	GU563327	20.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	6.49
KR108835	K563362	20	52 vs 52	None	None	None	None	None	None	None	None	None	None	6 vs 6	0.5
KR108835	MG387659	20	52 vs 52	None	None	None	None	None	None	None	None	None	None	6 vs 6	0.343
KR108835	FN434437	20.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	6.72
KR108835	GQ329864	20	52 vs 52	None	None	None	None	None	None	None	None	None	None	6 vs 6	0.428
KR108835	HM181930	23.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	7.908
KR108835	KR108836	23.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	7.585
KR108835	MG387654	22.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	6.995
KR108835	FN433931	23.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	7.068
KR108835	FJ39520	22	52 vs 52	None	None	None	None	None	None	None	None	None	None	28 vs 28	7.389
KR108835	KR108837	22.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	7.272
HM181930	KR108836	257	52 vs 52	0.9	0.83	0.77	0.74		2	2	1	1		46 vs 46	0.709
HM181931	MG387654	254	52 vs 52	3.33	2.92	2.26	1.86	1.7	2	4	3	2	1	40 vs 40	1.621
HM181932	FN433931	259	52 vs 52	7.22	6.22	5.33	4.55	3.65	3	3	3	4	3	36 vs 36	2.991
HM181933	FJ39520	251	52 vs 52	1.01	0.73	0.7	None	None	3	1	1	None	None	47 vs 47	0.672
HM181934	KR108837	254	52 vs 52	0.58	0.39	None	None	None	4	4	None	None	None	44 vs 44	0.322

(*continued on next page*)

Table 8 (*continued*)

Protein 1	Protein 2	Score	Aligned	RMSD per alignment cycle					Number of rejected atoms per alignment cycle					Matched	Executive
				Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5		
			atoms											atoms	RMSD
HM181935	MG387656	259	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	11.099
LT577538	KR108835	20	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	0.502
LT577538	HG9652222	226	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	8.3
LT577538	MG387655	226	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	8.601
KR108835	HG9652222	23.5	52 vs 52	None	None	None	None	None	None	None	None	None	None	30 vs 30	5.799
KR108835	MG387655	23.5	52 vs 52	6.29	None	None	None	None	1	None	None	None	None	29 VS 29	5.832
MG387655	HG9652222	263	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	8.897
MG347654	KR108828	232	52 vs 52	6.5	5.26	4.36	3.1	2.16	3	3	3	3	1	39 vs 39	2.028
KR108836	GU563327	234	52 vs 52	6.95	5.89	5.05	4.06	3.27	3	3	4	3	4	35 vs 35	2.496
FJ185044	K563362	223	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	11.019
FN433931	MG387659	226	52 vs 52	7.28	6.53	5.62	4.81	4.03	2	3	3	3	2	39 vs 39	3.543
FJ39520	FN434437	237	52 vs 52	7.01	6.28	5.28	4.72	4.26	2	3	2	2	1	42 vs 42	3.999
KR108837	GQ329864	234	52 vs 52	None	None	None	None	None	None	None	None	None	None	52 vs 52	1.222

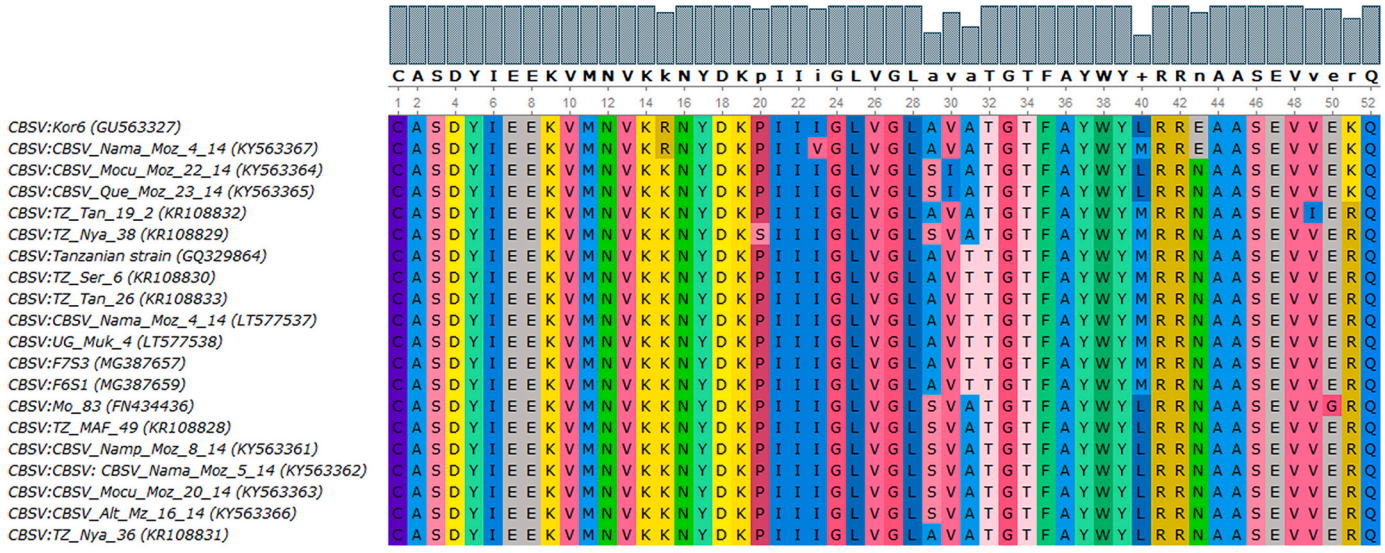

Fig. 5. Multiple sequence alignment of 6K2 protein expressed by 20 strains of CBSV species generated by MUSCLE in UGENE v 40.1 software suite.

faster through host's vascular systems, CBSV-infectedcassava are characterized by early onset of foliar and root CBSD-symptoms plus higher tissue titers compared to those infected by UCBSV species (Amisse et al., 2019a; Amuge et al., 2017; Anjanappa et al., 2017; Maruthi et al., 2014; Mulimbi et al., 2012; Saggaf et al., 2019). With that being said, we further hypothesize that; 6K2 protein is a pathogenicity determinant factor of CBSV and UCBSV species. Meaning that, without 6K2, the two viruses can not colonize its host's system.

Our assumptions aligns with what was previously reported by Anjanappa et al. (2017). The elite KBH 2006/0028 variety could resist single and mixed infection by CBSV and UCBSV species through blockage of its Plasmodesmata (PD) and subsequent inhibition of systemic movements of the viruses (Anjanappa et al., 2017). In our own perspective, we think occlusion of the PD cellular channels ended up restricting movement of 6K2, which is the carrier/distributor of viruses throughout the plant's vascular system. We postulate that; the halting of 6K2's expression might be another mechanism used by cassava varieties to block systemic spread of viruses during infection.

The reason why a minority strains population (5% and 15%) of CBSV and UCBSV species respectively, express 6K2 protein with hydrophobic

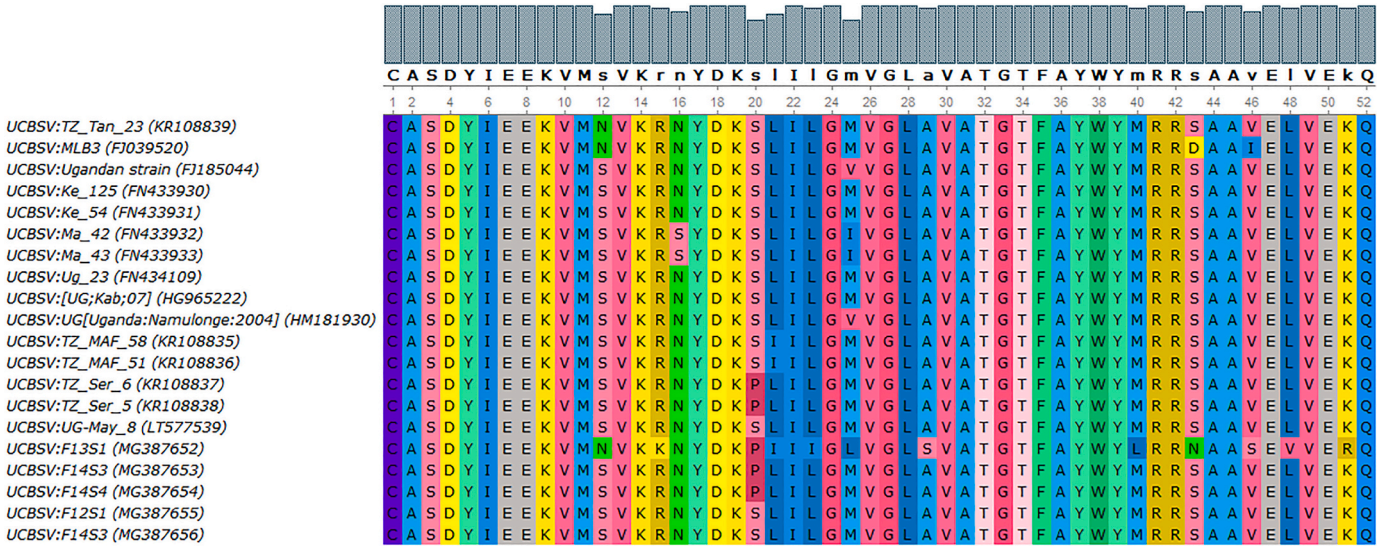

Fig. 6. Multiple sequence alignment of 6K2 protein expressed by 20 strains of UCBSV species generated by MUSCLE in UGENE v 40.1 software suite.

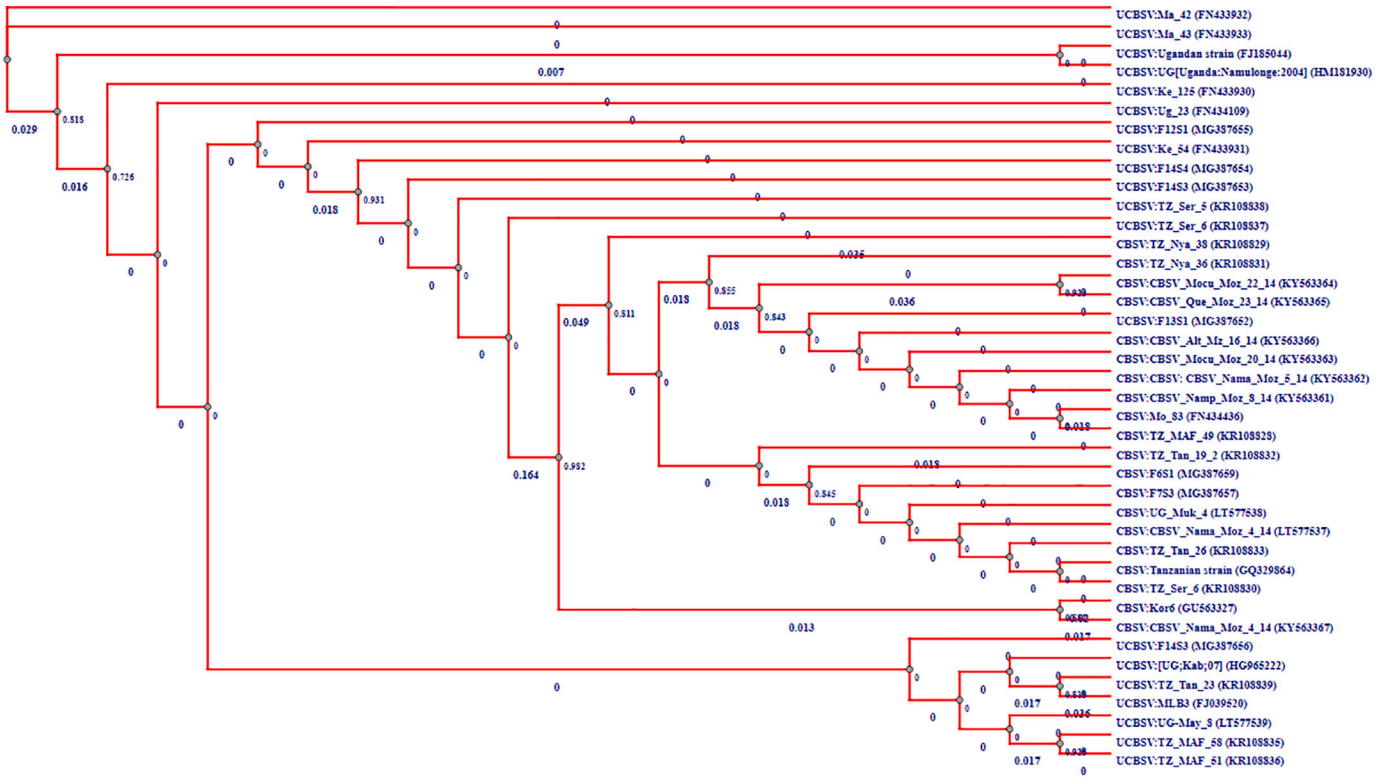

Fig. 7. Phylogenetic tree of 6K2 proteins of 20 CBSV and 20 UCBSV species constructed by using PhyML-Maximum Likelihood in UGENE v 40.1 software suite.

and hydrophilic characteristics is because they were negatively selected in the course of evolution but are possibly undergoing evolutionary changes.

5.2. How differential oligomerization of 6K2 protein influence functions and traits discrepancy between CBSV and UCBSV species?

The fact that, experimentally determined 3D structure (templates) in PDB have shown similarity and coverage to 6K2 by < 50% and < 50% respectively; explains why homology models of the protein were predicted in SWISS-MODEL as a monomer and homo-dimer for CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, correspondingly. On the other hand, ab initio models of the 6K2 protein were predicted in ROBETTA server as tetramer and homo-trimer because the whole protein (100% coverage) was used for construction of de novo 3D structures. That being the case, the present study considers the latter findings to be sounder (more precise) than the former. Generally, ab initio 6K2 protein models of (19 (95%) and 1(5%)) and (17(85%) and 3(15%)) population strains of CBSV and UCBSV species, respectively were homo-trimers and homo-tetramers, correspondingly. We are curious to understand whether oligomerization of 6K2 protein in CBSV and UCBSV species occurs by either covalent or non-covalent bonding (interaction between monomers).

Being a prerequisite to onset of physiological pathways in a living

Table 9
Physical-chemical properties of CBSV: TZ_MAF 49 strain (KR108828).

Protein identity		6K2	CP	Ham 1	VPg	NIb
Average molecular weight		5910.803	42,642.9202	25,320.4882	21,389.1826	57,901.2514
Monoisotopic molecular weight		5907.0494	42,616.6808	25,304.3063	21,375.7223	57,864.1987
Total number of atoms		836	5974	3610	2986	8110
Protein length		52	378	226	186	502
Negatively charged residues (D, E)		6 (11.54%)	59 (15.61%)	28 (12.39%)	32 (17.20%)	75 (14.94%)
Positively charged residues (K, R, H)		7 (13.46%)	62 (16.40%)	29 (12.83%)	35 (18.82%)	71 (14.14%)
Polar residues (C, S, Q, N, T, Y)		14 (26.92%)	93 (24.60%)	44 (19.47%)	42 (22.58%)	110 (21.91%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		25 (48.08%)	64 (43.39%)	125 (55.31%)	77 (41.40%)	246 (49.00%)
The aliphatic index		95.58	71.06	93.98	62.74	88.88
The Grand Average of Hydropathicity (GRAVY)		−0.115	−0.782	−0.035	−0.898	−0.289
amino acid pKa values from Bjellqvist et al		8.09	6.58	6.84	7.06	5.36
amino acid pKa values from Wikipedia		8.16	6.57	6.9	7.37	5.13
Extinction coefficient		11,460	46,410	22,460	22,460	80,790
Alanine (A)	Total count	5	39	13	6	21
	% count	9.62	10.32	5.75	3.23	4.18
Arginine (R)	Total count	3	26	9	14	29
	% count	5.77	6.88	3.98	7.53	5.78
Asparagine (N)	Total count	3	19	8	7	26
	% count	5.77	5.03	3.54	3.76	5.18
Aspartic acid (D)	Total count	2	24	11	13	33
	% count	3.85	6.35	4.87	6.99	6.67
Cysteine (C)	Total count	1	2	2	0	9
	% count	1.92	0.53	0.88	0	1.79
Glutamine (Q)	Total count	1	18	8	7	12
	% count	1.92	4.76	3.54	3.76	2.39
Glutamic acid (E)	Total count	4	35	17	19	42
	% count	7.69	9.26	7.52	10.22	8.37
Glycine (G)	Total count	3	16	17	13	37
	% count	5.77	4.23	7.52	6.99	7.37
Histidine (H)	Total count	0	4	1	3	11
	% count	0	1.06	0.44	1.61	2.19
Isoleucine (I)	Total count	4	23	18	3	36
	% count	7.69	6.08	7.96	1.61	7.17
Leucine (L)	Total count	3	21	19	12	47
	% count	5.77	5.56	8.41	6.45	9.36
Lysine (K)	Total count	4	32	19	18	31
	% count	7.69	8.47	8.41	9.68	6.18
Methionine (M)	Total count	1	7	6	9	13
	% count	1.92	1.85	2.65	4.84	2.59
Phenylalanine (F)	Total count	1	11	16	6	28
	% count	1.92	2.91	7.08	3.23	5.58
Proline (P)	Total count	1	21	14	7	20
	% count	1.92	5.56	6.19	3.76	3.98
Serine (S)	Total count	3	29	13	15	26
	% count	5.77	7.67	5.75	8.06	5.18
Threonine (T)	Total count	2	16	9	9	16
	% count	3.85	4.23	3.98	4.84	3.19
Tryptophan (W)	Total count	1	6	3	3	9
	% count	1.92	1.59	1.33	1.61	1.79
Tyrosine (Y)	Total count	4	9	4	4	21
	% count	7.69	2.38	1.77	2.15	4.18
Valine (V)	Total count	6	20	19	18	35
	% count	11.54	5.29	8.41	9.68	6.97

Protein identity		NIa	CI	6K1	P1	P3
Average molecular weight		26,043.8862	70,865.9973	5874.8176	41,284.2227	34,765.2433
Monoisotopic molecular weight		26,027.3234	70,821.259	5871.0112	41,257.8476	34,742.8379
Total number of atoms		3662	9970	828	5788	4902
Protein length		324	630	52	362	294
Negatively charged residues (D, E)		22 (9.40%)	82 (13.02%)	5 (9.62%)	54 (14.92%)	38 (12.93%)
Positively charged residues (K, R, H)		34 (14.53%)	99 (15.71%)	6 (11.54%)	58 (16.02%)	47 (15.99%)
Polar residues (C, S, Q, N, T, Y)		60 (25.64%)	152 (24.13%)	10 (19.23%)	89 (24.59%)	75 (25.51%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		118 (50.43%)	297 (47.14%)	31 (59.62%)	161 (44.48%)	134 (45.58%)
The aliphatic index		84.02	88.43	86.35	85.03	98.1
The Grand Average of Hydropathicity (GRAVY)		−0.144	−0.265	0.288	−0.354	−0.107
amino acid pKa values from Bjellqvist et al		8.76	6.89	8.15	5.99	8.34
amino acid pKa values from Wikipedia		8.44	6.87	8.18	5.97	7.88
Extinction coefficient		26,930	58,330	12,490	40,910	60,850
Alanine (A)	Total count	13	38	5	19	13
	% count	5.56	6.03	9.62	5.25	4.42
Arginine (R)	Total count	11	39	3	21	21
	% count	4.7	6.19	5.77	5.8	7.14
Asparagine (N)	Total count	13	28	1	17	9
	% count	5.56	4.44	1.92	4.7	3.06

(*continued on next page*)

Table 9 (*continued*)

Protein identity		NIa	CI	6K1	P1	P3
Aspartic acid (D)	Total count	10	44	2	22	18
	% count	4.27	6.98	3.85	6.08	6.12
Cysteine (C)	Total count	5	10	1	11	8
	% count	2.14	1.59	1.92	3.04	2.72
Glutamine (Q)	Total count	6	17	1	9	10
	% count	2.56	2.7	1.92	2.49	3.4
Glutamic acid (E)	Total count	12	38	3	32	20
	% count	5.13	6.03	5.77	8.84	6.8
Glycine (G)	Total count	18	40	5	18	11
	% count	7.69	6.35	9.62	4.97	3.74
Histidine (H)	Total count	8	19	0	10	6
	% count	3.42	3.02	0	2.76	2.04
Isoleucine (I)	Total count	13	28	4	27	26
	% count	5.56	4.44	7.69	7.46	8.84
Leucine (L)	Total count	14	59	4	24	29
	% count	5.98	9.37	7.69	6.63	9.86
Lysine (K)	Total count	15	41	3	27	20
	% count	6.41	6.51	5.77	7.46	6.8
Methionine (M)	Total count	4	15	1	8	6
	% count	1.72	2.38	1.92	2.21	2.04
Phenylalanine (F)	Total count	13	28	6	15	18
	% count	5.56	4.44	11.54	4.14	6.12
Proline (P)	Total count	13	21	1	14	3
	% count	5.56	3.33	1.92	3.87	1.02
Serine (S)	Total count	14	40	4	23	27
	% count	5.98	6.35	7.69	6.35	9.18
Threonine (T)	Total count	15	40	2	20	6
	% count	6.41	6.35	3.85	5.52	2.04
Tryptophan (W)	Total count	3	6	2	5	7
	% count	1.28	0.95	3.85	1.38	2.38
Tyrosine (Y)	Total count	7	17	1	9	15
	% count	2.99	2.7	1.92	2.49	5.1
Valine (V)	Total count	27	62	3	31	21
	% count	11.54	9.84	5.77	8.56	7.14

organism, homo- and hetero-oligomerization can be irreversible through covalent bond stabilization or reversible association through formation of hydrogen bonds, electrostatic and hydrophobic interactions (Higgs and Stantic, 2008). From that viewpoint, it is obvious that transition of 6K2 protein between homo-tetramer and homo-trimer, have significant impact on physiological functions of the protein between CBSV and UCBSV species and so does their trait discrepancy. Momentarily, we don't know which among the three and four alpha helices of homo-trimeric and homo-tetrameric 6K2 proteins respectively, serve as active site (s) for performing versatile functions in CBSV and UCBSV species, separately. Is it possible that, each monomer of the homo-trimeric and homo-tetrameric 6K2 proteins perform distinctive function? Could it be that the active sites of the homo-trimers and homo-tetramers protein also resides at the loops interspacing the alpha subunits? We recommend that; it is only through site directed mutagenesis, functions of all monomeric units and loops between them can be determined.

We are set to discuss the following research questions: How different is the mechanism underlying oligomerization of homo-trimeric versus homo-tetrameric structures of 6K2 protein? Regardless of species demarcation what survivor advantage (s) do strains expressing homo-tetrameric 6K2 proteins have, over those expressing homo-trimeric form of the same protein? Why do UCBSV species have more homo-tetrameric 6K2 protein than CBSV species?

Dimerization of two monomers to form homodimers which later self-associate to form homo-tetramers is the commonest homo-tetramerism mechanism in vivo, compared to direct association between distinct homo-dimers or sequential addition of monomer(s) to homo-dimer or homo-trimers (Powers and Powers, 2003). Contrariwise, homo-trimerism involves either association between a homo-dimer and a monomer or sequential association of three monomers. Basing on this explanation we suggest that; addition of monomer to homo-trimer, is the most likely mechanism by which homo-tetrameric structures of 6K2 proteins are formed in CBSV and UCBSV species.

Since for some proteins to be functional they need to be expressed in homo-tetrameric form (Powers and Powers, 2003), we suggest that; UCBSV strains express the homo-tetrameric form of 6K2 as a mechanism of advancing its function (s) in counteracting host's immune response compared to CBSV species. Findings of this study are in agreement with the fact that UCBSV species are currently evolving and undergoing speciation whilst CBSV species have already passed "survival for the flattest" theory (Ndunguru et al., 2015; Alicai et al., 2016; Amisse et al., 2019b). To be more certain about differential functions of the 6K2 proteins between CBSV and UCBSV species, we recommend that further experimental studies involving complete deletion or site directed mutagenesis of the 6K2 by recombinant DNA technology should be done. Basing on pairwise alignment of 3D structures, significant variation exists between and within homo-trimeric and homo-tetrameric 6K2 proteins of 20 CBSV and 20 UCBSV strains, respectively. That being the reality, this study hypothesize that; possession of 6K2 protein with homo-tetramer and homo-trimer structures influence virulence levels between strains of CBSV and UCBSV species. In other words we hypothesize that; besides being a pathogenic determinant, 6K2 is also a virulent factor in CBSVs. Further experimental studies involving complete deletion or site directed mutagenesis of the 6K2 by recombinant DNA technology, are equally needed to determine differential functions of homo-trimeric and homo-tetrameric 3D structures of the proteins in CBSV contrary to UCBSV species.

Table 10
Physical-chemical properties of UCBSV: TZ_MAF 58 strains (KR108835).

Protein identity		6K2	CP	Ham1	VPg	NIb
Average molecular weight		5892.8615	41,429.4495	25,082.2136	21,649.6552	58,230.6499
Monoisotopic molecular weight		5888.9768	41,404.0714	25,066.0728	21,635.7415	58,193.3403
Total number of atoms		830	5800	3565	3015	8154
Protein length		52	366	226	185	502
Negatively charged residues (D, E)		6 (11.54%)	56 (15.30%)	28 (12.39%)	36 (19.46%)	76 (15.14%)
Positively charged residues (K, R, H)		7 (13.46%)	57 (15.57%)	33 (14.60%)	37 (20.00%)	73 (14.54%)
Polar residues (C, S, Q, N, T, Y)		13 (25.00%)	93 (25.41%)	45 (19.91%)	32 (17.30%)	111 (22.11%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		26 (50.00%)	160 (43.72%)	120 (53.10%)	80 (43.24%)	242 (48.21%)
The aliphatic index		90.00	77.84	90.93	60.59	89.1
The Grand Average of Hydropathicity (GRAVY)		0.056	−0.72	−0.158	−0.942	−0.296
amino acid pKa values from Bjellqvist et al		8.09	5.91	8.24	5.91	5.41
amino acid pKa values from Wikipedia		8.16	5.80	7.88	5.83	5.20
Extinction coefficient		11,460	46,410	16,960	23,950	85,260
Alanine (A)	Total count	6	32	14	10	25
	% count	11.54	8.74	6.19	5.41	4.98
Arginine (R)	Total count	3	25	7	13	28
	% count	5.77	6.83	3.1	7.03	5.58
Asparagine (N)	Total count	1	24	13	7	25
	% count	1.92	6.56	5.75	3.78	4.98
Aspartic acid (D)	Total count	2	29	13	17	34
	% count	3.85	7.92	5.75	9.19	6.77
Cysteine (C)	Total count	1	2	5	0	10
	% count	1.92	0.55	2.21	0	1.99
Glutamine (Q)	Total count	1	19	6	5	13
	% count	1.92	5.19	2.65	2.7	2.59
Glutamic acid (E)	Total count	4	27	15	19	42
	% count	7.69	7.38	6.64	10.27	8.37
Glycine (G)	Total count	3	17	17	11	33
	% count	5.77	4.64	7.52	5.95	6.57
Histidine (H)	Total count	0	5	3	4	12
	% count	0	1.37	1.33	2.16	2.39
Isoleucine (I)	Total count	3	24	11	8	36
	% count	5.77	6.56	4.87	4.32	7.17
Leucine (L)	Total count	3	23	21	10	47
	% count	5.77	6.28	9.29	5.41	9.36
Lysine (K)	Total count	4	27	23	20	33
	% count	7.69	7.38	10.18	10.81	6.57
Methionine (M)	Total count	3	5	6	12	12
	% count	5.77	1.37	2.65	6.49	2.39
Phenylalanine (F)	Total count	1	11	12	7	28
	% count	1.92	3.01	5.31	3.78	5.58
Proline (P)	Total count	0	18	14	8	18
	% count	0	4.92	6.19	4.32	3.59
Serine (S)	Total count	4	24	8	7	23
	% count	7.69	6.56	3.54	3.78	4.58
Threonine (T)	Total count	2	15	9	8	16
	% count	3.85	4.1	3.98	4.32	3.19
Tryptophan (W)	Total count	1	6	2	3	9
	% count	1.92	1.64	0.88	1.62	1.79
Tyrosine (Y)	Total count	4	9	4	5	24
	% count	7.69	2.46	1.77	2.7	4.78
Valine (V)	Total count	6	24	23	11	34
	% count	11.54	6.56	10.18	5.95	6.77

Protein identity		NIa	CI	6K1	P1	P3
Average molecular weight		26,210.1424	70,890.0269	5788.7799	41,231.2517	34,665.3676
Monoisotopic molecular weight		26,193.3469	70,845.0814	5784.9553	41,204.8271	34,642.8789
Total number of atoms		3681	9956	818	5787	4903
Protein length		234	628	52	362	294
Negatively charged residues (D, E)		23 (9.83%)	81 (12.90%)	5 (9.62%)	60 (16.57%)	35 (11.90%)
Positively charged residues (K, R, H)		33 (14.10%)	97 (15.45%)	6 (11.54%)	60 (16.57%)	49 (16.67%)
Polar residues (C, S, Q, N, T, Y)		60 (25.64%)	156 (24.84%)	11 (21.15%)	79 (21.82%)	78 (26.53%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		118 (50.43%)	294 (46.82%	30 (57.69%)	163 (45.03%)	132 (44.90%)
The aliphatic index		82.35	87.15	88.27	85.11	98.06
The Grand Average of Hydropathicity (GRAVY)		−0.192	−0.27	0.335	−0.397	−0.123
amino acid pKa values from Bjellqvist et al		8.70	6.73	8.00	5.69	8.76
amino acid pKa values from Wikipedia		8.28	6.73	7.88	5.54	8.43
Extinction coefficient		28,420	56,840	12,490	36,900	63,370
Alanine (A)	Total count	12	33	6	23	10
	% count	5.13	5.25	11.54	6.35	3.4
Arginine (R)	Total count	13	40	1	19	16
	% count	5.56	6.37	1.92	5.25	5.44
Asparagine (N)	Total count	12	27	1.92	18	13

(*continued on next page*)

Table 10 (*continued*)

Protein identity		NIa	CI	6K1	P1	P3
	% count	5.13	4.3	2	4.97	4.42
Aspartic acid (D)	Total count	13	42	3.85	28	11
	% count	5.56	6.69	2	7.73	3.74
Cysteine (C)	Total count	6	10	3.85	11	8
	% count	2.56	1.59	1	3.04	2.72
Glutamine (Q)	Total count	8	20	1.92	7	7
	% count	3.42	3.18	3	1.93	2.38
Glutamic acid (E)	Total count	10	39	5.77	32	24
	% count	4.27	6.21	4	8.84	8.16
Glycine (G)	Total count	18	39	7.69	22	10
	% count	7.69	6.21	0	6.08	3.4
Histidine (H)	Total count	6	19	0	7	8
	% count	2.56	3.03	3	1.93	2.72
Isoleucine (I)	Total count	14	36	5.77	29	26
	% count	5.98	5.73	5	8.01	8.84
Leucine (L)	Total count	13	52	9.62	27	29
	% count	5.56	8.28	5	7.46	9.86
Lysine (K)	Total count	14	38	9.62	34	25
	% count	5.98	6.05	5	9.39	8.5
Methionine (M)	Total count	5	18	1	9	8
	% count	2.14	2.87	1.92	2.49	2.72
Phenylalanine (F)	Total count	12	29	5	16	15
	% count	5.13	4.62	9.62	4.42	5.1
Proline (P)	Total count	15	22	1	10	4
	% count	6.41	3.5	1.92	2.76	1.36
Serine (S)	Total count	10	47	4	17	29
	% count	4.27	7.48	7.69	4.7	9.86
Threonine (T)	Total count	16	36	2	16	8
	% count	6.84	5.73	3.85	4.42	2.72
Tryptophan (W)	Total count	3	6	2	4	8
	% count	1.28	0.96	3.85	1.1	2.72
Tyrosine (Y)	Total count	8	16	1	10	13
	% count	3.42	2.55	1.92	2.76	4.42
Valine (V)	Total count	26	59	3	23	22
	% count	11.11	9.39	5.77	6.35	7.48

It wasdemonstrated that, ErbB3 receptor tyrosine kinase transitioned its oligomeric state from a monomer to homo-dimer when subjected to high concentrations of dodecylphosphocholine (DPC-d38) detergent (Mineev et al., 2011). Considering the fact that, 6K2 is an integral component of the Virus Replication complex (VRC) besides playing crucial role in long distance systemic spread of CBSVs; it is obvious that, the protein moves between extracellular and intracellular (Cytoplast and nucleus) compartments of the host with different PH and intracellular fluid concentrations. From that standpoint we propose that; physical-chemical properties of the intracellular compartments in which the 6K2 protein is located, mayinfluence its oligomerization between homo-tetramer and homo-trimer state. Since regulation of enzymes' activity is among functions of homo-oligomeric proteins besides cell-to-cell adhesions (Nishi et al., 2013), we hypothesize that CBSV and UCBSV species regulate NIb-directed replication of its genomic (+)ssRNA, through transition of 6K2 protein between homo-trimeric and homo-tetrameric oligomeric states. Such transition of the 6K2 between homo-trimeric and homo-tetrameric states depends largely on types of mutations the viruses undergo, in response to CBSD-resistance mechanisms exerted by distinct cassava genotypes under different agro-ecological conditions. Even a single mutation (substitution/deletion of amino acids) can have a huge impact on oligomerization of 6K2 protein as well as its transitions between homo-trimeric and homo-tetrameric states.

Diverse organisms are able to use few proteins expressed from small genomes to perform multiple functions through protein homo-oligomerism, which is in turn governed by mutations (Nishi et al., 2013). Therefore, contribution of substitutional mutations to adoption of certain physical-chemical properties by 6K2 protein and its subsequent oligomerization, cannot be left undiscussed when differential tissue titers and replication rate exists between CBSV and UCBSV species are concerned. Among the compared 20 CBSV and 20 UCBSV species N↔S, R↔K, (L↔I, L↔V, I↔V, M↔L and L↔A), (S↔P, S↔A, L↔S, A↔T, S↔V and S↔I), E↔D, (E↔S, E↔N and N↔D) and E↔G substitutions occurred between; polar, positively charged, hydrophobic (non-polar), polar and hydrophobic, negatively charged, polar and negatively charged, negatively charged and hydrophobic residues, respectively. From these findings, it is obvious that; substitution mutations of hydrophobic have surpassed hydrophilic and charged amino acid residues in CBSV and UCBSV species. We therefore postulate that; substitution mutations of hydrophobic are dominant over other mutation types for the sake of maintaining functional diversification (pathogenicity and virulence) of 6K2 protein, in both CBSV and UCBSV species. On the other hand we hypothesize that; substitution mutations of hydrophilic amino acid residues, were minor in 6K2 protein because they are detrimental to the two virus species and they had to be restricted in low frequency in order to conserve crucial functions of the protein.

This study propose that differential Post Translational Modifications (phosphorylation, acetylation, glycosylaton etc.) of 52 amino acid residues constituting the 6K2 protein, is the reason why different strains of CBSV and UCBSV species express the same protein but with differed molecular weights ranging between (5894.8036–5942.86) and UCBSV (5847.8027–5961.9243), respectively.

Table 11
Physical-chemical properties of 6K2 proteins of 19 CBSV species.

Accession number (GenBank)		MG387659	GQ329864	KR108829	KR108830	KR108832	KR108833
Average molecular weight		5942.8631	5942.8631	5918.7977	5942.8631	5926.8636	5942.8631
Monoisotopic molecular weight		5939.0215	5939.0215	5914.9851	5939.0215	5923.0266	5939.0215
Total number of atoms		837	837	831	837	836	837
Protein length		52	52	52	52	52	52
Negatively charged residues (D, E)		6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)
Positively charged residues (K, R, H)		7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		14 (26.92%)	14 (26.92%)	15 (28.85%)	14 (26.92%)	13 (25.00%)	14 (26.92%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		25 (48.08%)	25 (48.08%)	24 (46.15%)	25 (48.08%)	26 (50.00%)	25 (48.08%)
The aliphatic index		88.08	88.08	88.08	88.08	91.92	88.08
The Grand Average of Hydropathicity (GRAVY)		−0.15	−0.15	−0.137	−0.150	−0.096	−0.150
amino acid pKa values from Bjellqvist et al		8.09	8.09	8.09	8.09	8.09	8.09
amino acid pKa values from Wikipedia		8.16	8.16	8.16	8.16	8.16	8.16
Extinction coefficient		11,460	11,460	11,460	11,460	11,460	11,460
Alanine (A)	Total count	5	5	5	5	6	5
	% count	9.62	9.62	9.62	9.62	11.54	9.62
Arginine (R)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Asparagine (N)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Aspartic acid (D)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Cysteine (C)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamine (Q)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamic acid (E)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Glycine (G)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Histidine (H)	Total count	0	0	0	0	0	0
	% count	0	0	0	0	0	0
Isoleucine (I)	Total count	4	4	4	4	5	4
	% count	7.69	7.69	7.69	7.69	9.62	7.69
Leucine (L)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Lysine (K)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Methionine (M)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Phenylalanine (F)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Proline (P)	Total count	1	1	0	1	1	1
	% count	1.92	1.92	0	1.92	1.92	1.92
Serine (S)	Total count	2	2	4	2	2	2
	% count	3.85	3.85	7.69	3.85	3.85	3.85
Threonine (T)	Total count	3	3	2	3	2	3
	% count	5.77	5.77	3.85	5.77	3.85	5.77
Tryptophan (W)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Tyrosine (Y)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Valine (V)	Total count	6	6	6	6	5	6
	% count	11.54	11.54	11.54	11.54	9.62	11.54

Accession number (GenBank)		KY563362	KY563366	KY563367	LT577537	LT577538	MG387657	GU563327
Average molecular weight		5910.803	5910.803	5913.8217	5942.8631	5942.8631	5942.8631	5909.8153
monisotopic molecular weight		5907.0494	5907.0494	5909.9949	5939.0215	5939.0215	5939.0215	5906.0542
Total number of atoms		836	836	832	837	837	837	837
Protein length		52	52	52	52	52	52	52
Negatively charged residues (D, E)		6 (11.54%)	6 (11.54%)	7 (13.46%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	7 (13.46%)
Positively charged residues (K, R, H)		7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		14 (26.92%)	14 (26.92%)	12 (23.08%)	14 (26.92%)	14 (26.92%)	14 (26.92%)	12 (23.08%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		25 (48.08%)	25 (48.08%)	26 (50.00%)	25 (48.08%)	25 (48.08%)	25 (48.08%)	26 (50.00%)
The aliphatic index		95.58	95.58	88.08	88.08	88.08	88.08	97.5
The Grand Average of Hydropathicity (GRAVY)		−0.115	−0.115	−0.108	−0.150	−0.150	−0.15	−0.065
amino acid pKa values from Bjellqvist et al		8.09	8.09	6.36	8.09	8.09	8.09	6.36
amino acid pKa values from Wikipedia		8.16	8.16	6.39	8.16	8.16	8.16	6.39
Extinction coefficient		11,460	11,460	11,460	11,460	11,460	11,460	
Alanine (A)	Total count	5	5	6	5	5	5	6
	% count	9.62	9.62	11.54	9.62	9.62	9.62	11.54
Arginine (R)	Total count	3	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77	5.77
Asparagine (N)	Total count	3	3	2	3	3	3	2

(*continued on next page*)

Table 11 (*continued*)

Accession number (GenBank)		KY563362	KY563366	KY563367	LT577537	LT577538	MG387657	GU563327
Aspartic acid (D)	% count	5.77	5.77	3.85	5.77	5.77	5.77	3.85
	Total count	2	2	2	2	2	2	2
Cysteine (C)	% count	3.85	3.85	3.85	3.85	3.85	3.85	3.85
	Total count	1	1	1	1	1	1	1
Glutamine (Q)	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	1	1	1	1	1	1	1
Glutamic acid (E)	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	4	4	5	4	4	4	5
Glycine (G)	% count	7.69	7.69	9.62	7.69	7.69	7.69	9.62
	Total count	3	3	3	3	3	3	3
Histidine (H)	% count	5.77	5.77	5.77	5.77	5.77	5.77	5.77
	Total count	0	0	0	0	0	0	0
Isoleucine (I)	% count	0	0	0	0	0	0	0
	Total count	4	4	3	4	4	4	4
Leucine (L)	% count	7.69	7.69	5.77	7.69	7.69	7.69	7.69
	Total count	3	3	2	2	2	2	3
Lysine (K)	% count	5.77	5.77	3.85	3.85	3.85	3.85	5.77
	Total count	4	4	4	4	4	4	4
Methionine (M)	% count	7.69	7.69	7.69	7.69	7.69	7.69	7.69
	Total count	1	1	2	2	2	2	1
Phenylalanine (F)	% count	1.92	1.92	3.85	3.85	3.85	3.85	1.92
	Total count	1	1	1	1	1	1	1
Proline (P)	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	1	1	1	1	1	1	1
Serine (S)	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	3	3	2	2	2	2	2
Threonine (T)	% count	5.77	5.77	3.85	3.85	3.85	3.85	3.85
	Total count	2	2	2	3	3	3	2
Tryptophan (W)	% count	3.85	3.85	3.85	5.77	5.77	5.77	2.85
	Total count	1	1	1	1	1	1	1
Tyrosine (Y)	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	4	4	4	4	4	4	4
Valine (V)	% count	7.69	7.69	7.69	7.69	7.69	7.69	7.69
	Total count	6	6	7	6	6	6	6
	% count	11.54	11.54	13.46	11.54	11.54	11.54	11.69

Accession number (GenBank)		FN434436	KR108831	KY563361	KY563363	KY563364	KY563365
Average molecular weight		5902.831	5894.8036	5910.803	5910.803	5896.8164	5896.8164
Monisotopic molecular weight		5899.0173	5891.0545	5907.0494	5907.0494	5893.0589	5893.0589
Total number of atoms		830	835	836	836	837	837
Protein length		52	52	52	52	52	52
Negatively charged residues (D, E)		5 (9.62%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)
Positively charged residues (K, R, H)		6 (11.54%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		10 (19.23%)	13 (25.00%)	14 (26.92%)	14 (26.92%)	14 (26.92%)	14 (26.92%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		31 (59.62%)	26 (50.00%)	25 (48.08%)	25 (48.08%)	25 (48.08%)	25 (48.08%)
The aliphatic index		86.35	97.5	95.58	95.58	97.5	97.5
The Grand Average of Hydropathicity (GRAVY)		0.277	−0.065	−0.115	−0.115	−0.098	−0.098
amino acid pKa values from Bjellqvist et al		8.18	8.09	8.09	8.09	8.07	8.07
amino acid pKa values from Wikipedia		8.18	8.16	8.16	8.16	8.16	8.16
Extinction coefficient							
Alanine (A)	Total count	5	6	5	5	5	5
	% count	9.62	11.54	9.62	9.62	9.62	9.62
Arginine (R)	Total count	4	3	3	3	2	2
	% count	7.69	5.77	5.77	5.77	3.85	3.85
Asparagine (N)	Total count	1	3	3	3	3	3
	% count	1.92	5.77	5.77	5.77	5.77	5.77
Aspartic acid (D)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Cysteine (C)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamine (Q)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamic acid (E)	Total count	3	4	4	4	4	4
	% count	5.77	7.69	7.69	7.69	7.69	7.69
Glycine (G)	Total count	5	3	3	3	3	3
	% count	9.62	5.77	5.77	5.77	5.77	5.77
Histidine (H)	Total count	0	0	0	0	0	0
	% count	0	0	0	0	0	0
Isoleucine (I)	Total count	4	4	4	4	5	5
	% count	7.69	7.69	7.69	7.69	9.62	9.62
Leucine (L)	Total count	4	3	3	3	3	3
	% count	7.69	5.77	5.77	5.77	5.77	5.77
Lysine (K)	Total count	2	4	4	4	5	5

(*continued on next page*)

Table 11 (*continued*)

Accession number (GenBank)		FN434436	KR108831	KY563361	KY563363	KY563364	KY563365
Methionine (M)	% count	3.85	7.69	7.69	7.69	9.62	9.62
	Total count	1	1	1	1	1	1
Phenylalanine (F)	% count	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	6	1	1	1	1	1
Proline (P)	% count	11.54	1.92	1.92	1.92	1.92	1.92
	Total count	1	1	1	1	1	1
Serine (S)	% count	1.92	1.92	1.92	1.92	1.92	1.92
	Total count	4	2	3	3	3	3
Threonine (T)	% count	7.69	3.85	5.77	5.77	5.77	5.77
	Total count	2	2	2	2	2	2
Tryptophan (W)	% count	3.85	3.85	3.85	3.85	3.85	3.85
	Total count	2	1	1	1	1	1
Tyrosine (Y)	% count	3.85	1.92	1.92	1.92	1.92	1.92
	Total count	1	4	4	4	4	4
Valine (V)	% count	1.92	7.69	7.69	7.69	7.69	7.69
	Total count	3	6	6	6	5	5
	% count	5.77	11.54	11.54	11.54	9.62	9.62

6. Conclusion

Basing on ab initio modelling, the present study conclude that, three dimensional (3D) structures and polarity of the second 6-kDa protein (6K2) protein varies significantly between CBSV and UCBSV species and hence the suspected basis for differed tissue titres and replication rate between the two virus species. The investigated 20 CBSV and 20 UCBSV species, express homo-trimeric and homo-tetrameric structures of 6K2 at higher (95% and 85%) and lower (5% and 15%) proportions, respectively. In addditon, 95% and 85% of 20 CBSV and 20 UCBV express its 6K2 protein in hydrophilic and hydrophobic form, respectively. Based on findings of the current study, we raised two hypotheses: (i) Hydrophilic 6K2 protein expressed by CBSV species favour its faster systemic movement via vascular tissues of cassava and hence results into higher tissue titres than the hydrophobic 6K2 expressed by UCBSV species (ii) Hydrophilic 6K2 protein expressed by CBSV species have additional interaction advantage with Nuclear Inclusion b (NIb) protease domain and viral genome-linked protein (VPg), components of Virus Replication Complex (VRC) and hence contributing to faster replication of viral genome than hydrophobic 6K2 expressed by UCBSV species.

With exception of the 6K2 protein, no disparity in 3D structures of the rest of proteins (6K1, P1, P3, VPg, NIa, NIb, CP, Ham 1 and CI) was observed between genome-wise, homology models of the CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains. In terms of physical-chemical characteristics, the two strains differed significantly in extinction coefficient for Ham 1, VPg, NIb, NIa, CI, P1 and P3 proteins.

P3, 6K2, P1, VPg, NIb, CP, Ham 1 and CI proteins expressed by both CBSV: TZ_MAF 49 and UCBSV: TZ_MAF 58 strains, were predicted to function as; regulatory focus of Yeast Kinetochore assembly, Helical Transmembrane protein, ErbB3 receptor tyrosine kinase of ErbB or HER family, Transcription factor TFIIE of Yeast (*Saccharomyces cerevisiae*), Viral Protein genome linked (VPg) in recruiting eukaryotic translation initiation factors, RNA-directed RNA polymerase of *Sapporo virus*, Coat Protein of *Potato Y virus* and Inosine triphosphate pyrophosphatase (ITPA) P32T variant of human and Pre-mRNA-splicing factor Adenosine Triphosphate (ATP)-dependent RNA helicase (PRP22) of Yeast (*S. cerevisiae*), correspondingly. Nonetheless, 6K1 was predicted to function as Guanine nucleotide releasing protein, crystal structure of RasGRFI whilst in UCBSV: TZ_MAF 58 strains the same protein was functionally similar to Son of Sevenless (SOS) homolog 1, which perturbs RAS signaling in human upon binding to RAS:SOS:RAS. In UCBSV: TZ_MAF 58 and CBSV: TZ_MAF 49 strains, NIa was predicted to function as a protease (nuclear inclusion protein a) of *Tobacco vein Mottling virus* and a catalytic domain of the Nuclear Inclusion a (NIa) protease of *Tobacco Etch Virus*, respectively.

Since precision of ab initio methods for modelling 3D structure of proteins is limited (restricted) to small sized proteins with $\leq$150 amino acids long, genome-wise modelling of CBSV- and UCBSV-encoded proteins (CP, CI, P1, P3, NIa, NIb, Ham 1 and VPg) with (366–378, 628–630, 362, 294, 324, 502, 226 and 185 amino acid long, respectively) couldn't be accomplished in this study. Therefore, we suggest that, more research investment should be put in designing more advanced bioinformatics tools for modelling large-sized proteins ($\geq$ 150 amino acid long). Further scrutiny of the remaining, small-sized 6K1 (52 aa) encoded by the CBSVs is still ongoing. Poor quality of nucleotide sequences of CBSVs is among challenges encountered in the present in silico study, therefore there's a need to re-sequence some viruses and ensure that, sequence data of upcoming novel virus strains are of high quality. In order to understand better how homo-trimer and homo-tetramer forms of 6K2 protein interacts differently with other protein members (VPg and NIb) of Viral Replication Complex (VRC) we suggest that, further experimental studies involving recombinant DNA technology should be conducted. Considering its small size (52 amino acid long) and functional versatility, we recommed that the 6K2 gene should be targeted for RNA interference (RNAi)-derived transgenes of cassava varieties which are uniquely resistant to infection by CBSV and UCBSV species, separately. Improvement of virus-resistance in transgenic cassava produced by 6K2-specific RNAi technology, will reduce CBSD-epidemics, economic losses among farmers (poverty) and food insecurity in Africa.

Conflict of interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' contributions

HRM conceptualized and synthesized the whole research ideas, literature search and write-up of the article. SLL and EBR provided supervisory support and guidance during conduction of this study.

Funding

The present study was funded by the Swedish International

Table 12
Physical-chemical properties of 6K2 proteins of 19 UCBSV species.

Acession number (GenBank)		MG387654	MG387656	KR108839	FJ039520	FJ185044	FN433930
Average molecular weight		5902.9000	5892.8615	5919.8871	5961.9243	5860.8015	5892.8615
Monisotopic molecular weight		5898.9976	5888.9768	5915.9877	5957.9983	5857.0048	5888.9768
Total number of atoms		833	830	833	838	829	830
Protein length		52	52	52	52	52	52
Negatively charged residues (D, E)		6 (11.54%)	6 (11.54%)	6 (11.54%)	7 (13.46%)	6 (11.54%)	6 (11.54%)
Positively charged residues (K, R, H)		7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		12 (23.08%)	13 (25.00%)	13 (25.00%)	12 (23.08%)	13 (25.00%)	13 (25.00%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		27 (51.92%)	26 (50.00%)	26 (50.00%)	26 (50.00%)	26 (50.00%)	26 (50.00%)
The aliphatic index		90.00	90	90.00	91.92	95.58	90.00
The Grand Average of Hydropathicity (GRAVY)		0.027	0.042	−0.01	−0.056	0.087	0.042
amino acid pKa values from Bjellqvist et al		8.09	8.09	8.09	6.33	8.09	8.09
amino acid pKa values from Wikipedia		8.16	8.16	8.16	6.38	8.16	8.16
Extinction coefficient		11,460	11,460	11,460	11,460	11,460	11,460
Alanine (A)	Total count	6	6	6	6	6	6
	% count	11.54	11.54	11.54	11.54	11.54	11.54
Arginine (R)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Asparagine (N)	Total count	1	1	2	2	1	1
	% count	1.92	1.92	3.85	3.85	1.92	1.92
Aspartic acid (D)	Total count	2	2	2	3	2	1
	% count	3.85	3.85	3.85	5.77	3.85	1.92
Cysteine (C)	Total count	1	1	1	1	1	4
	% count	1.92	1.92	1.92	1.92	1.92	7.69
Glutamine (Q)	Total count	1	1	1	1	1	3
	% count	1.92	1.92	1.92	1.92	1.92	5.77
Glutamic acid (E)	Total count	4	4	4	4	4	0
	% count	7.69	7.69	7.69	7.69	7.69	0
Glycine (G)	Total count	3	3	3	3	3	2
	% count	5.77	5.77	5.77	5.77	5.77	3.85
Histidine (H)	Total count	0	0	0	0	0	4
	% count	0	0	0	0	0	7.69
Isoleucine (I)	Total count	2	2	2	3	2	4
	% count	3.85	3.85	3.85	5.77	3.85	7.69
Leucine (L)	Total count	4	4	4	4	4	3
	% count	7.69	7.69	7.69	7.69	7.69	5.77
Lysine (K)	Total count	4	4	4	4	4	1
	% count	7.69	7.69	7.69	7.69	7.64	1.92
Methionine (M)	Total count	3	3	3	3	2	3
	% count	5.77	5.77	5.77	5.77	3.85	5.77
Phenylalanine (F)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Proline (P)	Total count	1	0	0	0	0	0
	% count	1.92	0	0	0	0	0
Serine (S)	Total count	3	4	3	2	4	4
	% count	5.77	7.69	5.77	3.85	7.69	7.69
Threonine (T)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Tryptophan (W)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Tyrosine (Y)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Valine (V)	Total count	6	6	6	5	7	6
	% count	11.54	11.54	11.54	9.62	13.46	11.54

Acession number (GenBank)		FN433931	FN433932	FN433933	FN434109	HG965222	HM181930
Average molecular weight		5892.8615	5847.8027	5847.8027	5892.8615	5892.8615	5860.8015
Monisotopic molecular weight		5888.9768	5844.0095	5844.0095	5888.9768	5888.9768	5857.0048
Total number of atoms		830	829	829	830	830	829
Protein length		52	52	52	52	52	52
Negatively charged residues (D, E)		6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)
Positively charged residues (K, R, H)		7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		13 (25.00%)	13 (25.00%)	13 (25.00%)	13 (25.00%)	13 (25.00%)	13 (25.00%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		26 (50.00%)	26 (50.00%)	26 (50.00%)	26 (50.00%)	26 (50.00%)	26 (50.00%)
The aliphatic index		90	97.5	97.5	90	90	95.58
The Grand Average of Hydropathicity (GRAVY)		0.042	0.144	0.144	0.042	0.042	0.087
amino acid pKa values from Bjellqvist et al		8.09	8.09	8.09	8.09	8.09	8.09
amino acid pKa values from Wikipedia		8.16	8.16	8.16	8.16	8.16	8.16
Extinction coefficient		11,460	11,460	11,460	11,460	11,460	11,460
Alanine (A)	Total count	6	6	6	6	6	6
	% count	11.54	11.54	11.54	11.54	11.54	11.54
Arginine (R)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Asparagine (N)	Total count	1	0	0	1	1	1

(*continued on next page*)

Table 12 (*continued*)

Acession number (GenBank)		FN433931	FN433932	FN433933	FN434109	HG965222	HM181930
	% count	1.92	0	0	1.92	1.92	1.92
Aspartic acid (D)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Cysteine (C)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamine (Q)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Glutamic acid (E)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Glycine (G)	Total count	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77
Histidine (H)	Total count	0	0	0	0	0	0
	% count	0	0	0	0	0	0
Isoleucine (I)	Total count	2	3	3	2	2	2
	% count	3.85	5.77	5.77	3.85	3.85	3.85
Leucine (L)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Lysine (K)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Methionine (M)	Total count	3	2	2	3	3	2
	% count	5.77	3.85	3.85	5.77	5.77	3.85
Phenylalanine (F)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Proline (P)	Total count	0	0	0	0	0	0
	% count	0	0	0	0	0	0
Serine (S)	Total count	4	5	5	4	4	4
	% count	7.69	9.62	9.62	7.69	7.94	7.69
Threonine (T)	Total count	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85
Tryptophan (W)	Total count	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92
Tyrosine (Y)	Total count	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69
Valine (V)	Total count	6	6	6	6	6	7
	% count	11.54	11.54	11.54	11.54	11.54	13.46

Acession number (GenBank)		KR108836	KR108837	KR108838	LT577539	MG387653	MG387655	MG387652
Average molecular weight		5892.8615	5902.9	5902.9	5892.8615	5902.9	5892.8615	5910.803
Monisotopic molecular weight		5888.9768	5898.9976	5898.9976	5888.9768	5898.9976	5888.9768	5907.0494
Total number of atoms		830	833	833	830	833	830	836
Protein length		52	52	52	52	52	52	52
Negatively charged residues (D, E)		6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)	6 (11.54%)
Positively charged residues (K, R, H)		7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)	7 (13.46%)
Polar residues (C, S, Q, N, T, Y)		13 (25.00%)	12 (23.08%)	12 (23.08%)	13 (25.00%)	12 (23.08%)	13 (25.00%)	14 (26.92%)
Hydrophobic residues (A, G, I, L, M, P, F, W, V)		26 (50.00%)	27 (51.92%)	27 (51.92%)	26 (50.00%)	27 (51.92%)	26 (50.00%)	25 (48.08%)
The aliphatic index		90	90	90	90.00	90	90	95.58
The Grand Average of Hydropathicity (GRAVY)		0.056	0.027	0.027	0.042	0.027	0.042	−0.115
amino acid pKa values from Bjellqvist et al		8.09	8.09	8.09	8.09	8.09	8.09	8.09
amino acid pKa values from Wikipedia		8.16	8.16	8.16	8.16	8.16	8.16	8.16
Extinction coefficient		11,460	11,460	11,460	11,460	11,460	11,460	11,460
Alanine (A)	Total count	6	6	6	6	6	6	5
	% count	11.54	11.54	11.54	11.54	11.54	11.54	9.62
Arginine (R)	Total count	3	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77	5.77
Asparagine (N)	Total count	1	1	1	1	1	1	3
	% count	1.92	1.92	1.92	1.92	1.92	1.92	5.77
Aspartic acid (D)	Total count	2	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85	3.85
Cysteine (C)	Total count	1	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
Glutamine (Q)	Total count	1	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
Glutamic acid (E)	Total count	4	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Glycine (G)	Total count	3	3	3	3	3	3	3
	% count	5.77	5.77	5.77	5.77	5.77	5.77	5.77
Histidine (H)	Total count	0	0	0	0	0	0	0
	% count	0	0	0	0	0	0	0
Isoleucine (I)	Total count	3	2	2	2	2	2	4
	% count	5.77	3.85	3.85	3.85	3.85	3.85	7.69
Leucine (L)	Total count	3	4	4	4	4	4	3
	% count	5.77	7.69	7.69	7.69	7.69	7.69	5.77
Lysine (K)	Total count	4	4	4	4	4	4	4

(*continued on next page*)

Table 12 (*continued*)

Acession number (GenBank)		KR108836	KR108837	KR108838	LT577539	MG387653	MG387655	MG387652
	% count	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Methionine (M)	Total count	3	3	3	3	3	3	1
	% count	5.77	5.77	5.77	5.77	5.77	5.77	1.92
Phenylalanine (F)	Total count	1	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
Proline (P)	Total count	0	1	1	0	1	0	1
	% count	0	1.92	1.92	0	1.92	0	1.92
Serine (S)	Total count	4	3	3	4	3	4	3
	% count	7.69	5.77	5.77	7.69	5.77	7.69	5.77
Threonine (T)	Total count	2	2	2	2	2	2	2
	% count	3.85	3.85	3.85	3.85	3.85	3.85	3.85
Tryptophan (W)	Total count	1	1	1	1	1	1	1
	% count	1.92	1.92	1.92	1.92	1.92	1.92	1.92
Tyrosine (Y)	Total count	4	4	4	4	4	4	4
	% count	7.69	7.69	7.69	7.69	7.69	7.69	7.69
Valine (V)	Total count	6	6	6	6	6	6	6
	% count	11.54	11.54	11.54	11.54	11.54	11.54	11.54

Development Cooperation Agency (SIDA), project number 2221.

Acknowledgement

Special thanks goes to SIDA, for funding the conduction of the present study through a sandwich PhD scholarship offered at the University of Dar es Salaam (UDSM),Tanzania in collaboration with Swedish University of Agricultural Sciences (SLU), Sweden. Dr. Juliette Hayer is highly appreciated for her immeasurable support and guidance during data analysis. All members of Swedish University of Agricultural Sciences (SLU), SLU-Global Bioinformatics Centre, Department of Animal Breeding and Genetics, Uppsala, Sweden are well-regarded for their assistance throughout conduction of this study.

References

Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M.A., Boykin, L.M., 2016. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. Sci. Rep. 6 https://doi.org/10.1038/srep36164.

Amisse, J.J.G., Ndunguru, J., Tairo, F., Ateka, E., Boykin, L.M., Kehoe, M.A., Cossa, N., Rey, C., Sseruwagi, P., 2019a. Analyses of seven new whole genome sequences of cassava brown streak viruses in Mozambique reveals two distinct clades: evidence for new species. Plant Pathol. 68, 1007–1018. https://doi.org/10.1111/ppa.13001.

Amisse, Jamisse J.G., Ndunguru, J., Tairo, F., Boykin, L.M., Kehoe, M.A., Cossa, N., Ateka, E., Rey, C., Sseruwagi, P., 2019b. First report of Cassava brown streak viruses on wild plant species in Mozambique. Physiol. Mol. Plant Pathol. 105, 88–95. https://doi.org/10.1016/j.pmpp.2018.10.005.

Amuge, T., Berger, D.K., Katari, M.S., Myburg, A.A., Goldman, S.L., Ferguson, M.E., 2017. A time series transcriptome analysis of cassava (Manihot esculenta Crantz) varieties challenged with Ugandan cassava brown streak virus. Sci. Rep. 7, 1–21. https://doi.org/10.1038/s41598-017-09617-z.

Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., Murzin, A.G., 2014. SCOP2 prototype: a new approach to protein structure mining. Nucleic Acids Research 42 (18), 11847. https://doi.org/10.1093/nar/gkt1242.

Andreeva, A., Kulesha, E., Gough, J., Murzin, A.G., 2020. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. Nucleic Acids Res. 48, D376–D382. https://doi.org/10.1093/nar/gkz1064.

Anjanappa, R., Mehta, D., Okoniewski, M., Szabelska, A., Gruissem, W., Vanderschuren, H., 2017. Early Transcriptome Analysis of the Brown Streak Virus–Cassava Pathosystem Provides Molecular Insights into Virus Susceptibility and Resistance, 36164, 100552. https://doi.org/10.1101/100552.

Anjanappa, R.B., Mehta, D., Okoniewski, M.J., Szabelska-Beręsewicz, A., Gruissem, W., Vanderschuren, H., 2018. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. Mol. Plant Pathol. 19, 476–489. https://doi.org/10.1111/mpp.12565.

Ateka, E., Alicai, T., Ndunguru, J., Tairo, F., Sseruwagi, P., Kiarie, S., Makori, T., Kehoe, M.A., Boykin, L.M., 2017. Unusual occurrence of a DAG motif in the Ipomovirus Cassava brown streak virus and implications for its vector transmission. PLoS One 12. https://doi.org/10.1371/journal.pone.0187883.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, 252–258. https://doi.org/10.1093/nar/gku340.

Chivian, D., Kim, D.E., Malmström, L., Schonbrun, J., Rohl, C.A., Baker, D., 2005. Prediction of CASP6 structures using automated Robetta protocols. Proteins Struct. Funct. Genet. 61, 157–166. https://doi.org/10.1002/prot.20733.

de Oliveira, L.C., Volpon, L., Rahardjo, A.K., Osborne, M.J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M., Kwok, B.H., Borden, K.L.B., 2019. Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: implications for translation. Proc. Natl. Acad. Sci. U. S. A. 116, 24056–24065. https://doi.org/10.1073/pnas.1904752116.

Dimitrova, Y.N., Jenni, S., Valverde, R., Khin, Y., Harrison, S.C., 2016. Structure of the MIND complex defines a regulatory focus for yeast kinetochore assembly. Cell 167, 1014–1027.e12. https://doi.org/10.1016/j.cell.2016.10.011.

Faure, G., Joseph, A.P., Craveur, P., Narwani, T.J., Srinivasan, N., Gelly, J.C., Rebehmed, J., De Brevern, A.G., 2019. IPBAvizu: a PyMOL plugin for an efficient 3D protein structure superimposition approach. Source Code Biol. Med. 14, 1–5. https://doi.org/10.1186/s13029-019-0075-3.

Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.C., Newman, A.J., Nagai, K., 2017. Structure of a spliceosome remodelled for exon ligation. Nature 542, 377–380. https://doi.org/10.1038/nature21078.

Freedman, T.S., Sondermann, H., Friedland, G.D., Kortemme, T., Bar-Sagi, D., Marqusee, S., Kuriyan, J., 2006. A Ras-induced conformational switch in the Ras activator Son of sevenless. Proc. Natl. Acad. Sci. U. S. A. 103, 16692–16697. https://doi.org/10.1073/pnas.0608127103.

Fullerton, S.W.B., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P.A., Rohayem, J., 2007. Structural and functional characterization of Sapovirus RNA-dependent RNA polymerase. J. Virol. 81, 1858–1871. https://doi.org/10.1128/jvi.01462-06.

Gotte, G., Libonati, M., 2014. Protein oligomerization. Oligomer. Chem. Biol. Compd. https://doi.org/10.5772/57489.

Higgs, T., Stantic, B., 2008. Hydrophobic-Hydrophilic Forces and Their Effects on Protein Structural Similarity. … Proc. [of the] Third …, pp. 1–12.

Ivanov, K.I., Eskelin, K., Lõhmus, A., Mäkinen, K., 2014. Molecular and cellular mechanisms underlying potyvirus infection. J. Gen. Virol. https://doi.org/10.1099/vir.0.064220-0.

Kežar, A., Kavčič, L., Polák, M., Nováček, J., Gutiérrez-Aguirre, I., Žnidarič, M.T., Coll, A., Stare, K., Gruden, K., Ravnikar, M., Pahovnik, D., Žagar, E., Merzel, F., Anderluh, G., Podobnik, M., 2019. Structural basis for the multitasking nature of the potato virus Y coat protein. Sci. Adv. 5, 1–14. https://doi.org/10.1126/sciadv.aaw3808.

Kwibuka, Y., Bisimwa, E., Blouin, A.G., Bragard, C., Candresse, T., Faure, C., Filloux, D., Lett, J., Maclot, F., Marais, A., Ravelomanantsoa, S., Shakir, S., Vanderschuren, H., Massart, S., 2021. Novel Ampeloviruses Infecting Cassava in Central Africa and the South-West Indian Ocean Islands, pp. 1–17.

Maiorov, V.N., Crippen, G.M., 1994. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. J. Mol. Biol. https://doi.org/10.1006/jmbi.1994.1017.

Mäkinen, K., Hafrén, A., 2014. Intracellular coordination of potyviral RNA functions in infection. Front. Plant Sci. https://doi.org/10.3389/fpls.2014.00110.

Maruthi, M.N., Bouvaine, S., Tufan, H.A., Mohammed, I.U., Hillocks, R.J., 2014. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One 9. https://doi.org/10.1371/journal.pone.0096642.

Mbanzibwa, Deusdedith R., Tian, Y., Mukasa, S.B., Valkonen, J.P.T., 2009a. Cassava Brown Streak Virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-pro. J. Virol. 83, 6934–6940. https://doi.org/10.1128/jvi.00537-09.

Mbanzibwa, D.R., Tian, Y.P., Tugume, A.K., Mukasa, S.B., Tairo, F., Kyamanywa, S., Kullaya, A., Valkonen, J.P.T., 2009b. Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch. Virol. 154, 353–359. https://doi.org/10.1007/s00705-008-0301-9.

Mbanzibwa, D.R., Tian, Y.P., Tugume, A.K., Patil, B.L., Yadav, J.S., Bagewadi, B., Abarshi, M.M., Alicai, T., Changadeya, W., Mkumbira, J., Muli, M.B., Mukasa, S.B., Tairo, F., Baguma, Y., Kyamanywa, S., Kullaya, A., Maruthi, M.N., Fauquet, C.M., Valkonen, J.P.T., 2011. Evolution of cassava brown streak disease-associated viruses. J. Gen. Virol. 92, 974–987. https://doi.org/10.1099/vir.0.026922-0.

Mbewe, W., Tairo, F., Sseruwagi, P., Ndunguru, J., Duffy, S., Mukasa, S., Benesi, I., Sheat, S., Koerbler, M., Winter, S., 2017. Variability in P1 gene redefines phylogenetic relationships among cassava brown streak viruses. Virol. J. 14, 1–7. https://doi.org/10.1186/s12985-017-0790-9.

Mero, H.R., Lyantagaye, S.L., Bongcam-Rudloff, E., 2021. Why has permanent control of cassava brown streak disease in sub-Saharan Africa remained a dream since the 1930s? Infect. Genet. Evol. 94, 105001 https://doi.org/10.1016/j.meegid.2021.105001.

Mineev, K.S., Khabibullina, N.F., Lyukmanova, E.N., Dolgikh, D.A., Kirpichnikov, M.P., Arseniev, A.S., 2011. Spatial structure and dimer-monomer equilibrium of the ErbB3 transmembrane domain in DPC micelles. Biochim. Biophys. Acta Biomembr. 1808, 2081–2088. https://doi.org/10.1016/j.bbamem.2011.04.017.

Mohammed, I.U., Ghosh, S., Maruthi, M.N., 2016. Host and virus effects on reversion in cassava affected by cassava brown streak disease. Plant Pathol. 65, 593–600. https://doi.org/10.1111/ppa.12458.

Monger, W.A., Seal, S., Isaac, A.M., Foster, G.D., 2001. Molecular characterization of the Cassava brown streak virus coat protein. Plant Pathol. 50, 527–534. https://doi.org/10.1046/j.1365-3059.2001.00589.x.

Monger, W.A., Alicai, T., Ndunguru, J., Kinyua, Z.M., Potts, M., Reeder, R.H., Miano, D. W., Adams, I.P., Boonham, N., Glover, R.H., Smith, J., 2010. The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. Arch. Virol. 155, 429–433. https://doi.org/10.1007/s00705-009-0581-8.

Mukiibi, D.R., Alicai, T., Kawuki, R., Okao-Okuja, G., Tairo, F., Sseruwagi, P., Ndunguru, J., Ateka, E.M., 2019. Resistance of advanced cassava breeding clones to infection by major viruses in Uganda. Crop Prot. 115 https://doi.org/10.1016/j.cropro.2018.09.015.

Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., Reeder, R., Legg, J.P., Laurenson, L., Weekes, R., Thom, F.E., 2012. First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. New Dis. Rep. 26, 11. https://doi.org/10.5197/j.2044-0588.2012.026.011.

Munganyinka, E., Margaria, P., Sheat, S., 2018. Localization of cassava brown streak virus in *Nicotiana rustica* and cassava *Manihot esculenta* (Crantz) using RNAscope® in situ hybridization. Virology Journal 15 (128), 1–11. https://doi.org/10.1186/s12985-018-1038-z.

Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., Kehoe, M., Boykin, L.M., Melcher, U., 2015. Analyses of twelve new whole genome sequences of cassava brown streak viruses and ugandan cassava brown streak viruses from East Africa: diversity, supercomputing and evidence for further speciation. PLoS One 10. https://doi.org/10.1371/journal.pone.0139321.

Nishi, H., Hashimoto, K., Madej, T., Panchenko, A.R., 2013. Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. In: Progress in Molecular Biology and Translational Science. Copyright © 2013, 1st ed. Elsevier Inc. https://doi.org/10.1016/B978-0-12-386931-9.00001-5. All Rights Reserved.

Ogwok, E., Odipio, J., Halsey, M., Gaitán-Solís, E., Bua, A., Taylor, N.J., Fauquet, C.M., Alicai, T., 2012. Transgenic RNA interference (RNAi)-derived field resistance to cassava brown streak disease. Mol. Plant Pathol. 13, 1019–1031. https://doi.org/10.1111/j.1364-3703.2012.00812.x.

Patil, B.L., Ogwok, E., Wagaba, H., Mohammed, I.U., Yadav, J.S., Bagewadi, B., Taylor, N.J., Kreuze, J.F., Maruthi, M.N., Alicai, T., Fauquet, C.M., 2011. RNAi-mediated resistance to diverse isolates belonging to two virus species involved in cassava brown streak disease. Mol. Plant Pathol. 12, 31–41. https://doi.org/10.1111/j.1364-3703.2010.00650.x.

Phan, J., Zdanov, A., Evdokimov, A.G., Tropea, J.E., Peters, H.K., Kapust, R.B., Li, M., Wlodawer, A., Waugh, D.S., 2002. Structural basis for the substrate specificity of tobacco etch virus protease. J. Biol. Chem. 277, 50564–50572. https://doi.org/10.1074/jbc.M207224200.

Powers, E.T., Powers, D.L., 2003. A perspective on mechanisms of protein tetramer formation. Biophys. J. 85, 3587–3599. https://doi.org/10.1016/S0006-3495(03)74777-8.

Qin, Y., Djabou, A.S.M., An, F., Li, K., Li, Z., Yang, L., Wang, X., Chen, S., 2017. Proteomic analysis of injured storage roots in cassava (Manihot esculenta Crantz) under postharvest physiological deterioration. PLoS One 12, 1–24. https://doi.org/10.1371/journal.pone.0174238.

Rigden, D.J., 2017. From protein structure to function with bioinformatics: Second Edition. In: From Protein Structure to Function with Bioinformatics, Second ed. https://doi.org/10.1007/978-94-024-1069-3

Saggaf, M.H., Ndunguru, J., Tairo, F., Sseruwagi, P., Ascencio-Ibáñez, J.T., Kilalo, D., Miano, D.W., 2019. Immunohistochemical localization of Cassava brown streak virus and its morphological effect on cassava leaves. Physiol. Mol. Plant Pathol. 105, 67–76. https://doi.org/10.1016/j.pmpp.2018.06.001.

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., Cramer, P., 2017. Structures of transcription pre-initiation complex with TFIIH and mediator. Nature 551, 204–209. https://doi.org/10.1038/nature24282.

Simone, P.D., Struble, L.R., Kellezi, A., Brown, C.A., Grabow, C.E., Khutsishvili, I., Marky, L.A., Pavlov, Y.I., Borgstahl, G.E.O., 2013. The human ITPA polymorphic variant P32T is destabilized by the unpacking of the hydrophobic core. J. Struct. Biol. 182, 197–208. https://doi.org/10.1016/j.jsb.2013.03.007.

Studer, G., Tauriello, G., Bienert, S., Waterhouse, A.M., Bertoni, M., Bordoli, L., 2019. Computational Methods in Protein Evolution, p. 1851. https://doi.org/10.1007/978-1-4939-8736-8.

Sun, P., Austin, B.P., Tözsér, J., Waugh, D.S., 2010. Structural determinants of tobacco vein mottling virus protease substrate specificity. Protein Sci. 19, 2240–2251. https://doi.org/10.1002/pro.506.

Tomlinson, K.R., Bailey, A.M., Alicai, T., Seal, S., Foster, G.D., 2018. Cassava brown streak disease: historical timeline, current knowledge and future prospects. Mol. Plant Pathol. 19, 1282–1294. https://doi.org/10.1111/mpp.12613.

Tomlinson, K.R., Pablo-Rodriguez, J.L., Bunawan, H., Nanyiti, S., Green, P., Miller, J., Alicai, T., Seal, S.E., Bailey, A.M., Foster, G.D., 2019. Cassava brown streak virus Ham1 protein hydrolyses mutagenic nucleotides and is a necrosis determinant. Molecular Plant Pathology 20 (8), 1080–1092. https://doi.org/10.1111/mpp.12813.

Tsymbaliuk, A., . © 2019. This Manuscript Version is Made Available under the Elsevier user License. https://www.elsevier.com/open-access/userlicense/1.0/1, pp. 1–21.

Walker, J.M., 2009. Homology modeling: methods and protocols. Methods in Molecular Biology 531 (1), 588.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427.

Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M., Butgereitt, A., 2010. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J. Gen. Virol. 91, 1365–1372. https://doi.org/10.1099/vir.0.014688-0.

Xiang, Z., 2006. Advances in homology protein structure modeling. Curr. Protein Pept. Sci. 7, 217–227. https://doi.org/10.2174/138920306777452312.

Xiang, Y., Zou, Q., Zhao, L., 2020. VPTMdb: a viral posttranslational modification database. Brief. Bioinform. 1–6 https://doi.org/10.1093/bib/bbaa251.

Yadav, J.S., Ogwok, E., Wagaba, H., Patil, B.L., Bagewadi, B., Alicai, T., Gaitan-Solis, E., Taylor, N.J., Fauquet, C.M., 2011. RNAi-mediated resistance to Cassava brown streak Uganda virus in transgenic cassava. Mol. Plant Pathol. 12, 677–687. https://doi.org/10.1111/j.1364-3703.2010.00700.x.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2014. The I-TASSER suite: protein structure and function prediction. Nat. Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213.

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinform. 9, 1–8. https://doi.org/10.1186/1471-2105-9-40.