Determining Proteins Responsible for the
Host Range of Potyviruses

by

Adam Scanlon

Canterbury Christ Church University

Thesis submitted

for the degree of MSc by Research

2022

i. Abstract

The host range of a virus is important as it determines the characteristics of the virus and the distribution of the virus. The aim of this study was to use molecular dynamics simulations to model the interaction of specific virus proteins with plant proteins. Then validate these results through a yeast two-hybrid assay. The nerine yellow stripe virus's (NeYSV) host range is unknown. However, the tobacco etch virus (TEV) is well studied and its hosts include *Nicotiana benthamiana* and *Arabidopsis*. The NIa-pro of the TEV virus is predicted to interact with a variety of host proteins. This includes the protein SAHH2 which plays an important role in the creating siRNA, which defend against plant viruses. Potyviral proteins that interact with SAHH disable RNA silencing. This allows a virus to resist the hosts defences and become virulent, expanding the host range. This study modeled the interaction of NIa-pro from NeYSV with the SAHH proteins from *N. benthamiana* and *Arabidopsis*. The molecular dynamics results suggested that NeYSV NIa-pro would not bind to *Arabidopsis* SAHH2 or NbSAHH, suggesting that neither plants are suitable hosts of NeYSV. However, the molecular dynamics results showed that TEV NIa-pro would bind to *Arabidopsis* SAHH2, supporting previous studies, but not NbSAHH. As both of these plants are hosts of TEV the interaction between NIa-pro and SAHH protein cannot be a determinant of host range. Due to auto-activation, caused by NIa-pro, the molecular dynamics result were not able to be verified through the use of a yeast two-hybrid assay.

ii. Contents

1. Introduction

This introduction provides context to the widespread damage caused by potyviruses (Overview of Potyviruses 1.1), the role viral proteins play in plant infection (Potyvirus proteins and host proteins 1.2), the specific role NIa-pro plays in plant infection (NIa-pro from PRSV interactions 1.3; NIa-pro from TEV interactions 1.4) and NIa-pro's relation to host range (PRSV NIa-pro and host range 1.5). This section also contains a short summary of the experimental method and what we hoped to discover through these procedures (Aims of the investigation 1.6).

1.1 Overview of Potyviruses

Potyviruses are aphid spread viruses that are widespread in the agricultural and horticultural environment (Gibson et al., 2004). The most economically devastating members of the potyvirus family occur in tropical and subtropical regions. They dramatically reduce crop productivity, endangering food security in developing countries (Gowda et al., 2015). For example, maize is an important cereal crop and the maize lethal necrosis disease (MLND) has emerged as a serious threat to maize production in Eastern Africa. This disease is a combination of the maize virus and the sugar cane virus. These potyviruses work synergistically to destroy the host. Almost 90 percent of all pre-commercial and commercial maize plants are susceptible to MLND (Gowda et al., 2015). Bananas are also a staple food crop in developing countries and a major cash crop(Ordonez et al., 2015). Banana bract mosaic virus results in yield losses of up to 70% and has become a threat to Australia's banana industry (Sankaranarayanan et al., 2020). All commercially available bananas are grown from cuttings (Ordonez et al., 2015). Therefore, all bananas seen on supermarket shelves are genetically identical. Consequently, when resistance to pathogens breaks down in one banana, the resistance has been overturned across the whole commercially grown variety. This has lead to whole strains of bananas being unavailable for commercial use (Ordonez et al., 2015).

Potyviruses are also a problem in the horticultural industry. In 2005, the United States horticultural industry was worth 45 billion dollars (Jordan, Guaragna and Putnam, 2011). A large portion of the income, from this industry, comes from the distribution and sales of ornamental plants such as daffodils, tulips and lilies. Lily plants are frequently infected with potyviruses causing chlorosis as well as streaking and leaf deformation. This affects the appearance of plants and can consequently prevent sales (Jordan, Guaragna and Putnam, 2011). In the United Kingdom, reports have shown that potyviruses are affecting ornamental *Allium* species (Scrace, Denton and Clover, 2015). These viruses affected the appearances of the plants and subsequently affected sales. However, until the findings of this study were reported, the *Allium* species had not been identified as hosts of the potyviruses (Scrace,

Denton and Clover, 2015). With the correct tests, the potyviruses may have been identified and the host plants isolated and destroyed before they were placed on the market. Furthermore, Nerine Lilys are grown commercially in Europe and are traded internationally (Beddoe et al., 2020) . Recently Nerine lilies were discovered to display a mild to severe mosaic, darker green oval spots and yellow stripes (Beddoe et al., 2020). The cause of these symptoms was the nerine yellow stripe virus (NeYSV) (Beddoe et al., 2020). NeYSV appears to be confined to the Amaryllidaceae family (Beddoe et al., 2020). However, the complete hostrange of this virus remains to be determined. Some studies have predicted that NeYSV may affect *Nicotiana benthamiana* (*N.benthamiana*) (Beddoe et al., 2020). With infections reported in multiple countries, understanding the host range of the viruses may prevent further contamination to ornamental and commercial crops (Beddoe et al., 2020).

1.2 <u>Potyvirus proteins and host proteins</u>

The entire potyvirus' genome is contained within a +ssRNA, which has only one open reading frame (Adams, Antoniw and Beaudoin, 2005) . When potyviruses infect a plant, they transcribe their entire genome which produces a polyprotein as depicted in Figure 1.1. All these proteins play an important role in viral infection as outlined in Table 1.1 . However, certain viral proteins need to interact with host proteins in order for the virus to replicate, move and transmit to other plants (Mäkinen and Hafrén, 2014). It was reported by Oliveira et al., (2019) that the potyvirus-derived VPg protein from potato virus Y (PVY) associates with the host transcription factor Eukaryotic translation initiation factor 4, on the subunit G (AtEIF4G). RNAs contain a m^{7}G cap on the 5′ end. This helps stabilise the RNA and also provides a region for host proteins to bind to. RNA's m^{7}G cap associates with AtEIF4Es cap-binding site which enables the transcription factor to transcribe genes. VPg can substitute for the m^{7}G cap and outcompete RNA's for the cap binding site. Therefore, the transcription of genes associated with EIF4E can be controlled throughdirect competition with its cap-binding site. This interaction is essential as it leads to successful viral infection.

P1	HC-Pro	P3N-PIPO / P3	6K1	CI	6K2	VPg	NIa-pro	NIb	CP
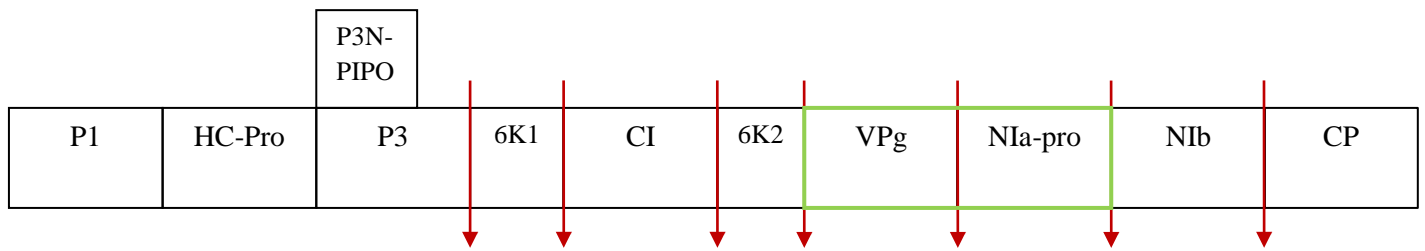
Figure 1.1 Image representing a potyviral polyprotein. Each box is labelled with the protein produced by potyviruses. The red arrows represent regions of the polyprotein cleaved by NIa-pro. The green box represents the NIA protein (comprised of the VPg and NIa-pro proteins) (Adams, Antoniw and Beaudoin, 2005).

Viral Protein	Abbreviation	Function
First Protein	P1	Genome amplification (Verchot and Carrington, 1995)
Helper component proteinase	HC-Pro	Viral movement and host gene silencing (Urcuqui-Inchima, Haenni and Bernardi, 2001)
Third protein	P3	Plant pathogenicity (Urcuqui-Inchima, Haenni and Bernardi, 2001)
First 6K protein	6K1	Viral Replication (Cui and Wang, 2016)
Cylindrical inclusion protein	CI	Viral movement through cells (Urcuqui-Inchima, Haenni and Bernardi, 2001)
Second 6K protein	6K2	Viral Replication (Urcuqui-Inchima, Haenni and Bernardi, 2001)
Viral genome-linked protein	VPg	Viral replication acts as a primer during RNA synthesis (Goodfellow, 2011)
Nuclear inclusion a protein	NIa	Cleavage of most sites in the polyprotein (Sun et al., 2010)
Nuclear inclusion protein b	NIb	Viral replication as an RNA replicase recruits several host proteins into the viral replication complexes (Shen et al., 2020)
Coat protein	CP	Cell-to-cell and systemic movement (Urcuqui-Inchima, Haenni and Bernardi, 2001)
Pretty interesting protein	P3N-PIPO	Acts as a suppressor of RNA-mediated gene silencing allows efficient cell to cell propagation (Vijayapalani et al., 2012)

Table 1.1 List of proteins produced by potyviruses, their abbreviated names and the function they play in the virus.

The HC-pro from PVY was also found to associate with EIF4G (Valli et al., 2017). This interaction was predicted to alter the structure of EIF4G, preventing the transcription of virus related defences genes (Valli et al., 2017). This, in turn, led to an increase in viral replication (Ala-Poikela et al., 2011). Furthermore, VPg from the papaya ring spot virus (PRSV) bound to the papaya calreticulin proteins calcium-binding domain (Valli et al., 2017). Ca^{2+} is an

essential second messenger that participates in many plant signal pathways including defence signalling (Valli et al., 2017). Valli (2017), proposes that the interaction downregulates plant defence.

The RNA-induced silencing complex (RISC), is a multi-protein complex which plays a role in silencing the expression of unwanted genes (Pratt and MacRae, 2009). For example, viral RNA expression. The Dicer enzyme facilitates the reaction of the RISC complex by breaking down viral dsRNA. The dsRNA fragments are then brought to the RISC complex where they are bound by Argonaute (AGO4). Whilst in this complex, the viral dsRNA is unwound and one strandof the RNA is removed .The remaining strand is the 'guider strand' and is complementary to the transcribed viral sequence. The guider stand can then 'guide' the RISC complex to the original mRNA sequence (of which it is complementary to). This results in the cleavage and destruction of the original sequence. These guiding strands of mRNA are known as siRNAs (small interfering RNAs) (Pratt and MacRae, 2009). A 2020 study discovered that NIa-pro, from the pepper mottle virus, was able to cause transcriptional changes to genes by altering the methylation of the plants genome (Gong et al., 2020). NIa-pro was found to downregulate proteins associated with plant defence, for example AGO4 (Gong et al., 2020). The method behind this alteration of gene expression is not fully understood (Gong et al., 2020). This paper argued that the interaction of NIa-pro with host proteins caused a change in DNA methylation which altered gene expression (Gong et al., 2020).

S-adenosylhomocysteinase hydrolase (SAHH) has been found to interact with HC-pro protein from PVY (De et al., 2018). SAHH plays a role in altering the methylation patterns in the plants genome (De et al., 2018). This is done by binding methyl groups to DNA which reduces gene expression (De et al., 2018). The interaction, between SAHH and HC-pro, was found to alter the chromatin structure reducing gene transcription (De et al., 2018). This decreased the transcription of defence related genes (De et al., 2018). Furthermore, siRNAs are methylated and, thereby, stabilized by the activity of RNA methyltransferases which require methylation by SAHH to function (Cañizares et al., 2013). The tomato chlorosis virus CP and PVY HC-pro was found to bind to and downregulate the function of NbSAHH (Cañizares et al., 2013). The paper theorised that downregulation of SAHH activity inhibited the S-adenosyl-methionine-dependent RNA methyltransferase (Cañizares et al., 2013). As a consequence, the modification of siRNAs may be inhibited (Cañizares et al., 2013). Destabilising the siRNAs can affect viral gene silencing due to the role siRNAs play in the RISC complex (Cañizares et al., 2013).

9

1.3 NIA-pro from PRSV interactions

The potyvirus NIA protease consists of two domains, the VPg domain and NIa-pro domain, which are separated by a weak catalytic site (Martínez et al., 2016). This is due to the suboptimal cleavage sequence separating the VPg and NIa-pro domains. NIA exists equally in both processed (NIa-pro and VPg) and unprocessed forms (NIA) (Martínez et al., 2016). The NIA protease utilises the trypsin-like domain from NIa-pro to cleave itself and the polyprotein at specific cleavage sites (Nunn et al., 2005). This is performed through the cytosine-histidiene-aspartame catalytic triad which hydrolyses the sequence specific cleavage site (Nunn et al., 2005). This was widely considered to be the sole function of NIa-pro (Carrington et al., 1993). However, in 2012 it was discovered that NIa-pro, from the papaya ring spot virus (PRSV), was responsible for interacting with five host proteins (Gao et al., 2012a). These proteins were reported to play important roles in the complete and proper functioning of potyviruses (outlined in Table1.2) (Gao et al., 2012a). Further analysis determined that the C-terminus of PRSV NIa-pro (the non catalytic site) interacted with Methionine Sulfoxide Reductase B1 (PaMSRB1) on a the C-terminal fragment (between 100-200 amino acids) (Gao et al., 2012b). This C-terminus region was responsible for the nuclear transportation of the protein. Nuclear transportation is important as it allows for the shuttling of key proteins to the nucleus (Lee et al., 2012). MSRB1 is activated by the expression of oxidants such as hydrogen peroxide (Rey and Tarrago, 2018). Oxidants are produced by the plant during stress, for example, when the plant is under attack by a pathogen (Hernández et al., 2016). These can build up in the plant cell and react with other molecules causing damage (Hernández et al., 2016). MSRB1 functions by reducing the oxidant methionine sulfoxide into methionine (Vieira Dos Santos et al., 2005). Due to this anti-oxidant function, MSRB1 can repair and restore functions to proteins which have been damaged because of oxidation of their methioine residue (Rey and Tarrago, 2018). Oh et al (2010) discovered that when the expression for CaMSRB2 in pepper plants was silenced, there was an increase in the production of oxidants which led to accelerated cell- death. This was found to enhance the virulence of an otherwise weak pathogen strain (Oh et al., 2010). Preventing the nuclear transport of MSRB1 prevents MSRB1from oxidising and restoring the functionality of critical proteins in the nucleus (Rey and Tarrago, 2018). This may cause plant defences to downregulate, aiding in viral replication (Reyand Tarrago, 2018). The C-terminal region of PaMSRB1 also contains a zinc finger. The functions of these regions in MSRB1 is to bind to metals such as zinc ions to stabilise the fold in a protein (Gao et al., 2012b). Furthermore, Gao et al (2012b) determined that the interaction between PaEIF3G and NIa-pro also occurs between the C-terminus of NIa-pro and the centre of the host protein (Gao et al., 2014). This middle region of PaEIF3G also contains a zinc binding motif (Gao et al., 2014). Due to this

10

pattern, potentially, the NIa-pro may bind to the conserved zinc finger motifs in host proteins. This could misfold the protein. Post translational modification of proteins is a common defence strategy of viruses against its host(Knipe, Raja and Lee, 2017).

Papaya Host protein	Abbreviated Name	Function in reference to potyviruses
Eukaryotic translation initiation factor 3 subunit G	EIF3G	Inhibits host protein synthesis (Grzela et al., 2006)
Fructose-1, 6-bisphosphate aldolase	FBPA1	Plant Stress (Lu et al., 2012)
FK506 Binding protein	FK506BP	Contributes to symptom development (Yang et al., 2021)
GTP Binding Family Protein	GTPBP	Roles in virus RNA Replication (Nishikiori et al., 2011)
Methionine sulfoxide reductase B1	MSRB1	Antioxidant (Gao et al., 2012b)
Metallothionein-like protein	MTL	Controls potyviral replication (Yang et al., 2015)
Table 1.2 The host proteins discovered to interact with papaya ringspot virus NIa-pro in papaya plants and the function of these proteins in relation to potyviruses		

1.4 NIa-pro and TEV interactions

Research by Martínez et al., (2016) showed that tobacco etch virus (TEV) NIa-pro, potentially, interacted withover 70 host proteins in the model organism *Arabidopsis thaliana*. These host proteins may aid in viral replication and defence related responses. The Martínez et al (2016), showed several proteins of interest, from the cytoplasmand nucleus, bound to the TEV NIA-pro. For example, it was discovered that there was interaction between NIA-pro and adenosylhomocysteinase2 (AtSAHH2) (Martínez et al., 2016). This enzyme is a homologue of SAHH1: the protein is structurally and functionally identical. As seen previously, the PVY targets AtSAHH to downregulate the hosts defences (De et al., 2018). Potentially, the NIa-pro

11

from TEV downregulates important viral defence genes by targeting AtSAHH2 (Lee et al., 2012). Furthermore, the protein RISC-interacting clearing exoribonuclease (AtRICE1) was found to interact with TEV NIa-pro (Martínez et al., 2016). AtRICE1 is responsible for the normal functioning of the RISC complex which fights off viral infection (Zhang et al., 2017). Additionally, TEV NIa-pro was discovered to associatewith AtGSTF2 (Martínez et al., 2016). This protein is responsible for transporting biomolecules for plant defence (Gullner et al., 2018). The protein also functions as an anitoxidant (Gullner et al., 2018). Altering the functions of these proteins is detrimental to the plants viral defence systems (Knipe, Raja and Lee, 2017). Overall, this suggests that NIa-pro may play a role is downregualting the hosts defences.

Although, the Gao et al (2012a) paper states that NIA-pro from PRSPV interacts with PaMSRB1, the Martínez et al (2016) paper states that the VPg domain from TEV interacts with AtMSRB1. Both studies agreed that NIa-pro interacts with a fructose-bisphosphate aldolase (FBPA). Furthermore, the Martínez et al (2016) paper collected their results 2 weeks after infection and the authors questioned if different host proteins interact with the NIa-pro at different stages in the viral lifecycle (Martínez et al., 2016). Therefore, it may be argued that TEV NIa-pro binds to MSRB1 at an earlier stage in the infection lifecycle.

1.5 PRSV NIa-pro and host range

Studies have also predicted that NIa-pro may be responsible for the host range of viruses (Chen et al., 2008). In a 2008 study, two similar strains of PRSV were examined (Chen et al., 2008). The study showed that one strain was able to infect papaya, while the other was able to infect cucurbits (Chen et al., 2008). Using recombinant viruses, it was determined that the NIA region was important for the infection and host range of the viruses (Chen et al., 2008). Through analysis of the genetic structure, it was determined that much of the genetic variation, between the two viruses, occurred in the NIa-pro region (Chen et al., 2008). It was discovered that by mutating the residue 'lysine 27' from NIa-pro, the virus displayed no symptoms of infection on papaya. The study concluded, that this residue was responsible for the host range of the virus. This residue is not involved with the catalytic activity of the protein (Chen et al., 2008). The study reasoned that the inability of PRSV to infect papaya was due to the mutated NIa-Pro domain being unable to associate with papaya host proteins (Chen et al., 2008). Specifically, the paper argued that the NIa-pro may interact with the host factor Eukaryotic translation initiation factor 4 on the subunit E (EIF4E) (Chen et al., 2008). Gao et al, (2014) later determined that PRSV interacts with EIF3G. The interaction between the virus and transcription factors is considered key in viral replication (Chen et al., 2008). However, the

Martínez et al (2016) paper did not find the TEV NIa-pro to interact with either EIF3G or EIF4E. Instead, Martínez et al (2016). reported that TEV NIA-pro binds to Poly-A-Binding (PABP) proteins four and eight. This is important, as in the turnip mosaic virus,the affinity between viral proteins and EIF4B increased threefold when in the presence of the hosts poly-a-binding proteins (Khan and Goss, 2012). Therefore, NIa-pro may play a crucial role as an intermediary, supporting the interaction between and EIF4E and viral proteins by binding to PABPs.

1.6 Aims of the investigation

This study aimed to use molecular dynamics and 'yeast two-hybrid assays' to investigate the potential role of SAHH and determine the host range of NeYSV. The TEV infects both *Arabidopsis-Ler0* and *N. benthamiana*. In contrast, NeYSV has not been found to infect *Arabidopsis* and there are reports that it infects *N. benthamiana* (Beddoe et al., 2020). SAHH plays an important role in the regulation of plant viral defences (Lee et al., 2012). Therefore, the interaction and alteration to the protein may allow a virus to proliferate.Using the results of the 'yeast two-hybrid assay' and molecular dynamics results, we aimed to determine whether TEV and NeYSV NIa-pros interacts with proteins that play a significant role in modulating the hosts defence. Therefore, proteins found to interact with both TEV and NeYSV only in *N. benthamiana*, may be responsible for the host range of the virus. This study alsoaimed to identify the cleavage sequence of the NeYSV protease through computational methods. The proteases may interact with host proteins in the same way the protease interactswith the polyprotein, by recognising and binding to sequence motifs.

2. Method

Computational analysis

The method section outlines the computational and laboratory experiments performed in this study. The computational methods performed were sequence alignments (Sequence alignments 2.1) and structural alignments to identify important residues and regions of proteins (Structural alignments 2.2). This was followed by the molecular dynamics simulation for the interaction between viral proteins and plant proteins (Molecular dynamics 2.3). Finally, this section contains methods used for the analysis of the MD simulations to determine if the two proteins interact (MD analysis 2.4) . Afterwards, the method section contains the laboratory experimental procedures. This included the designing of plant and viral primers (Primer design 2.5) , the extraction of plant RNA the conversion to cDNA (Amplifying sequences of interest 2.7) 2.6), the PCR of the cDNA using the designed primers (Amplifying sequence of interest 2.7), finally, the method section outlines the technique used for the yeast transformation and yeast two hybrid assay (Yeast transformation and control experiments 2.8).

2.1 Sequence Alignments

In the Martínez et al (2016) paper, 70 *Arabidopsis* proteins were identified to interact with TEV NIa-pro. In this study, these *Arabidopsis* proteins were standard protein BLAST (Basic Local Alignment Search Tool) searched against the NCBI database using *N. benthamiana* as a model organism. The data showed which of the 70 proteins, identified in *Arabidopsis,* had complete sequence data in *N. benthamiana*. These proteins were used in this study. The NeYSV's NIa-pro was protein BLAST search to determine similar proteases. Using this information, the NIa-pro protein sequence from NeYSV was aligned to the NIa-pro from TuMV using MEGA7 (Kumar, Stecher and Tamura, 2016). This enabled the identification of conserved functional components in the NeYSV NIa-pro, for example, the self cleavage site and catalytic triad. Additionally, using MEGA7, the NeYSV protease was aligned to the papaya ring spot virus' NIa-pro (Kumar, Stecher and Tamura, 2016). This was to identify, in NeYSV NIa-pro, the location of the residue discovered by Chen et al (2008) responsible for host range. The aminoacid sequence from AtSAHH2 was aligned against human SAHH. Due to the conserved sequence of the protein, the identification of AtSAHH2 C-terminal tail was enabled (Wang etal., 2014)

The locations of the individual proteins inside the NeYSV polyprotein were previously identified by Beddoe (2020). The regions flanking the NeYSV proteins cleaved by NIa-pro were compared to the TuMV cleavage sites. This identified an optimal cleavage sequence for the NeYSV protease. The cleavage sites on the viral polyprotein are bound to by the protease (Adams, Antoniw and Beaudoin, 2005). Hypothetically, the protease may bind to the cleavage sequences on the host proteins.

2.2 Structural alignments

The NeYSV NIa-pro self cleavage region and catalytic residues were also determined by aligning the NeYSV proteases' homology model to the TEV protease crystal structure model (obtained from PDB: 1LVM), using PyMOL version 2.5.2 (Schrödinger and DeLano, 2020). Structural alignments of the docked viral and host protein structures were achieved using pYMOL (Krissinel and Henrick, 2004)

2.3 Molecular dynamics

The model of the TEV protease was taken from PDB entry 1Q31. However, due to 'the low electron density around the C-terminal' (between residues 223-234) a region was not modelled (Nunn et al., 2005). The 1Q31 crystal structure model also contains a mutation at the catalytic site to prevent autolysis (Nunn et al., 2005). First, using mutabind (Li et al., 2016), the alanine from position 151 was mutated back into its original catalytic cytosine. Secondly, the missing loop region was modelled using DeReUS-Loop (Karami et al., 2018). The conformation of this loop region was determined using Modloop (Fiser, Do and Šali, 2000).This enabled the generation of a complete TEV NIa-pro structure. The myc tagged TEV NIa-pro protein was created using the mutate function in PyMOL version 2.5.2 (Schrödinger & DeLano, 2020).The myc tag and cloning site residues from pGBKT7 were added to the N and C terminus of the protease.

The homology model of the NeYSV NIa-pro (QOD42425.1), AtSAHH (OAP04800.1) and NbSAHH (LC008356.1) was created using the online I-tasser server (Roy, Kucukural and Zhang, 2010). This server was used as it has consistently been ranked as one of the best homology model generators (Croll et al., 2019). This generated co-ordinate files of the proteins. These co-ordinate files were equilibrated before docking. The docking between NIa- pro and the host proteins were performed using ClusPro and verified by subsequent docking through H-dock, which utilize different methodologies of docking (Desta et al., 2020; Yan etal., 2020). Both the template and template free binding displayed similar results. However, when equilibrating H-dock models of AtSAHH2 and TEV, the TEV's NIa-pro C-terminal loopwas found to clip inside AtSAHH2 forming a structure that could not naturally occur.
Therefore, equilibration and subsequent steps were carried out using ClusPro generated docking models. The identified cleavage sequence was bound to the equilibrated structure ofthe NeYSV protease using HPepDock (Zhou et al., 2018).

To equilibrate the .pdb files from I-tasser and ClusPro, charmm 36 topology files were used to create topology files of the individual and interacting proteins (Huang et al., 2016). The topology files provide mapping between bonded and non-bonded interactions involving the various combinations of atom types found in the topology. This information is used to

determine how the atoms are connected to each other in the co-ordinate file. The generation of the psf file was performed using the psfgen plugin in VMD v1.9 (Humphrey et al, 1996). Using VMD's solvation plugin, the protein was solvated inside a waterbox. The waterbox contained 8Å of padding from the centre of the protein (van Lun, van der Spoel and Andersson, 2011). Using VMD, the waterbox was neutralised through the addition of NaCl. All proteins, examined in this study, were located in the plant cytoplasm which consist of water, ions and has a pH of around 7.5 (Martinière et al.,2018). Therefore, the pH and the ions of the waterbox, mimic the environment these proteins are found in. However, sodium chloride is not often found naturally in plants, in a high concentration, due to its toxicity (Álvarez-Aragón et al., 2016). Previous studies examining the molecular dynamics (MD) of plant proteins have determined that NaCl is suitable for neutralising a system without affecting plant protein structure (van Lun, van der Spoel and Andersson, 2011). Additionally, previous studies examining the molecular dynamics of NIa- pro have determined that a neutral pH is also optimal for establishing viral protein dynamics (Mathur et al., 2012). The structure was geometry optimised using CHARMMs conjugate gradient minimisation. This adjusts the structure to the force field and relaxes any possible steric clashes. Long-range electrostatic, and Van der Waals interactions, were computed with a cutoff of 12 Å and 15 Å respectively. The neighbour searching algorithm was used to quickly and efficiently find the pairs of atoms within the predefined cutoff radius. The particle-mesh Ewald method was used to calculate electrostatic forces (van Lun, van der Spoel and Andersson, 2011). The structure was minimised for 2500cycles until the energy of the system had levelled off sufficiently (Figure 6.1). A constraints file was created using the last frame of the minimisation cycle. Constraints maintain bonds at constant lengths and also diminish the high-frequency motions, decreasing processing times. This was computed using the constraint algorithm SHAKE which processes all constraints oneafter another. The temperature of the system was linearly increased from 0K to 295k. As plantsdo not thermoregulate, they remain at room temperature. Therefore, this temperature was seenas the optimal temperature for plant growth and ideal for modelling the interaction between proteins in plants (Michaletz et al., 2015). The temperature of the NeYSV NIa-pro was incremented by 0.0001Kp/s per 2950000 steps. This rate increase avoided any constraint failures in the rattle and shake algorithm. For the docking structures, the rate of heating was increased to 0.001kp/s to reduce computing times.

Next, the structure was equilibrated. Velocities were assigned to the target temperature (295k) and were reset every cycle. This was to transfer the kinetic energy to potential energy in the system. As soon as the 'Total potential energy' levelled off, the equilibration stage was completed (Figure 6.3). Periodic boundary conditions were applied, and thermostats and barostats were introduced in the MD simulation, to keep the system temperature at 295k. This was in order to mimic experimental conditions: temperature and pressure were present and constant. This was performed using Langevin piston dynamics and Langevin thermostat. The pressure was set to one standard atmosphere to simulate room pressure. For the peptide

docked protease, the equilibrium ran for 10ns to preserve time. For all other structures, this ran for 25ns. For the docked structures, molecular dynamics simulations were performed by removingthe harmonic constraints and the models were simulated for 25ns. This was repeated three times to create a contentious model and determine any experimental outliers. For all structures, the final temperature was checked and tested to see if it matched the input temperature. The system velocities were analysed to determine if they matched the Maxwell Boltzmann normaldistribution. On all systems, normal distribution was found.

2.4 MD analysis

The VMD 'clustering tool' was used to cluster the MD frames. A cut-off value of 2.5Å was used. This created clusters, from the MD runs, with similar structures. The largest cluster, and therefore most stable version of the protein complex, was selected for further analysis. Using VMDs hydrogen bond analysis tool, interactions occurring between the viral and host proteins were determined. Likewise, the RMSD of the protein simulations were calculated using VMD's RMSD calculate tool. The RMSF of the simulation was calculated using VMDs' RMSF calculate tool. Frames from the cluster were selected and analysed using the server PDBePISA (Krissinel and Henrick, 2005). This was to determine the affinity of the viral protein to the host protein. Afterwards, the frames were loaded into the server MutaBind and alanine scanning was performed (Li et al., 2016). This determined the change in entropy any mutations may have upon the protein complex. The hydrophilic properties of the protease werecalculated using pYMOL's APBS plugin tool.

Yeast Two-Hybrid Analysis

2.5 Primer design

Primers were designed by examining the PGBKT7 and PGADT7 vectors that the TEV NIa-proand AtSAHH2 will be cloned into respectively. Restriction enzymes that cleaved the vectors, but did not cleave the protein sequence, were determined using the online tool; 'GenScript Restriction Enzyme Map Analysis' (www.genscript.com). This prevented the gene becoming fragmented during a restriction digest. Specific primers were designed which flanked the protein coding sequence. The restriction sites were added to the 5' and 3' ends of the primers. Added to the restriction sites were further base pairs to increase the rate of hybridization amplifying the rate of cleavage to the sequence (www.neb.com). The product, created by the PCR amplification and restriction, was compared to the cleaved vector to verify if the sequence could be cloned in-frame. The final primer sequences were inserted into 'Thermo Fisher multiple primer analyser' (www.thermofisher.com). This determined any primer dimers, the GC content (optimally between 40-60%) and the temperature of the primers (optimally between 40-70 °C).A nucleotide BLAST search was performed, on the primers, using their respective cDNA as a model organism. This determined any unwanted sequences from the genome that the primers may potentially amplify. When generating primers for the genes encoding *Arabidopsis* proteins of interest, a similar methodology was

performed with an additional final step. The *Arabidopsis* SAHH2 sequence was nucleotide BLAST searched, against the *Arabidopsis landsberg erecta* NCBI database, to determine any single nucleotide polymorphisms that may alter the primer function.

NIa-pro (MT396083.1) Primer (5'-3')		
Forward (5'-3')	NeYSV NIa-pro Forward	GATAAAC**CATATG**CACCATGAGAGTAAATC
Reverse (5'-3')	NeYSV NIa-pro Forward	CTGTGTATGAGCAAA**GGATCC**CAGATGGC
Table 2.1. Primers developed to amplify NIa-pro from NeYSV. The underlined regions represent the restriction sites from NdeI and BamHI respectively.		

| SAHH			
Host	Primer Name	Restriction Enzyme	Primer (5'-3')
Arabidopsis (NM_113286.3)	AtSAHH2 Forward	NdeI	GGAATTC**CATATG**CTTTGCTTGTAGAGAAAACCTC
	AtSAHH2 Reverse	BamHI	CG**GGATCC**GTACCTGTAGTGAACAGGCTTG
Nicotiana benthamiana (LC008356.1)	NbSAHH2 Forward	EcoRI	CG**GAATTC**ATGGCTCTATTAGTCGAGAA
	NbSAHH2 Reverse	BamHI	CG**GGATCC**GTACCTGTAGTGAGCAGG
Table 2.2 Primers developed to amplify sequences encoding for proteins predicted to interact with NIa-pro. The underlined regions represent the location of the restriction enzyme site.			

2.6 Genetic material extraction

N. benthamiana seeds were planted in a GoodHome peat-free multipurpose compost and the plants were grown in a growth cabinet for two months. The cabinet conditions were 60% humidity, 8h light/16h dark and a temperature of 23°C. The *Arabidopsis thaliana* Ler-0 seeds were planted in a multipurpose compost and cold stratified at -4°C for four days. Afterwards, they were placed in a growth cabinet set to 60% humidity, 12h light/12h dark and a temperature of 23 °C.They were grown for a month. Samples from the plants weighing less than 100mg were removed. These were dipped in liquid nitrogen and ground down using a pre-chilled pestle andmortar. The RNA was extracted using the QIAGEN RNeasy Plant kit. RTL buffer from the Plant kit and beta-mercaptoethanol were used to create a stock solution. 450µl was added tothe disrupted tissue. Through a series of column centrifuges and washes, the RNA was extracted. This was stored at -80 °C.

2.7 Amplifying sequences of interest

Using TakaraBio's RNA to cDNA EcoDry™ Premix (Double Primed), the extracted RNA wasconverted into cDNA. This cDNA was amplified, using the primers detailed on Table 2.1 and 2.2, using dreamTaq Mastermix (10X). The PCR reaction is outlined in Figure 2.1. The PCR fragment was excised from the gel and purified using QIAGEN's column purification. The restriction sites generated by the primers, along with the identical vector sites, were cleaved using the appropriate New England Biolab restriction enzymes. The volumes used for these reactions are listed in Table 2.3. Afterwards, the plasmids and DNA fragments would have been ligated together using Takara Bio's T4 DNA ligase (Table 2.4). Unfortunately, due to inconsistencies with the PCR results and sub-cloning, the *Arabidopsis thaliana-Col* plasmids U09076 (containing GSTF2) and C104785 (containing SAHH2) were purchased from the Arabidopsis Biological Resource Centre (ABRC). The genes of interests were subcloned from these plasmids into the PGADT7 plasmid. This was achieved through the use of GeneWiz's sub cloning services. Using a BLASTp search, the *Arabidopsis-Col* and *Arabidopsis-Ler0* SAHH2 were determined to have a 96% sequence similarity. Therefore, we can expect similar results to the Martínez et al (2016) paper. The full length NIa-pro from the tobacco etch virus was synthesised using GeneWiz's gene synthesis tool with the NCBI entry NC_001555.1 as a reference. This was undertaken to match the NIa-pro sequence outlined in the Martínez et al (2016) study. Through GeneWiz's sub cloning service, TEV NIa-pro was cloned into the PGBKT7 plasmid using the EcoRI and NotI restriction sites.
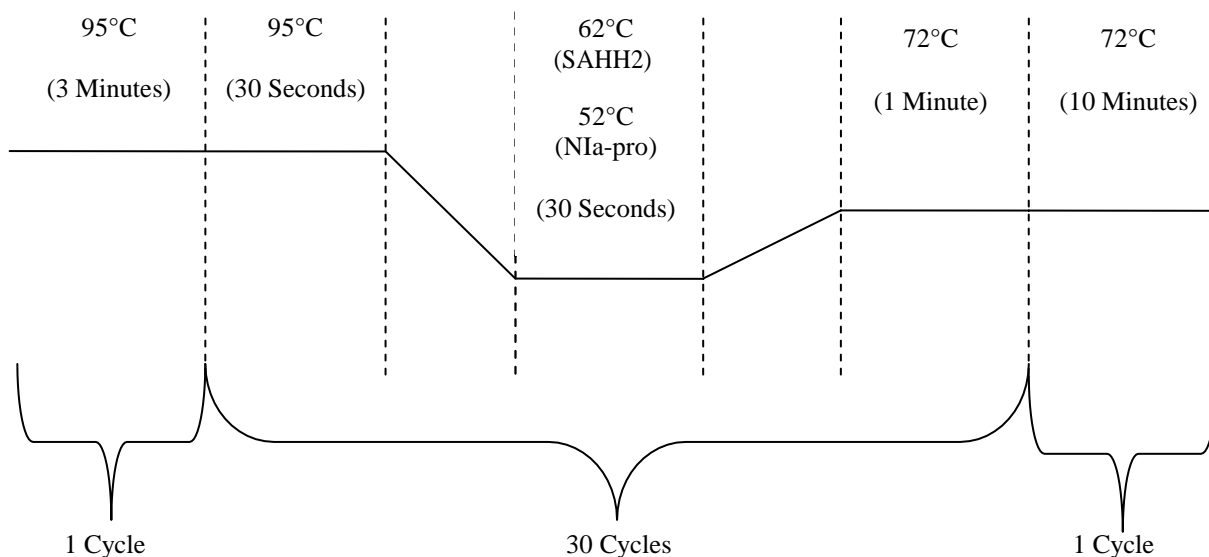
Figure 2.1 Timeline of PCR reactions outlining the temperature and duration of each step of the PCR reaction. The figure also displays the number of cycles the step was run for and the different annealing temperatures used.

Restriction Enzyme PCR product Double Digestion	
Product	Volume (µl)
Restriction Enzyme 1 (NdeI/EcoRI)	1
Restriction Enzyme 2 (BamHI)	1
Buffer	5
Water	25
PCR Product	23
Restriction Enzyme Plasmid Double Digestion	
Product	Volume (µl)
Restriction Enzyme 1 (NdeI/EcoRI)	1
Restriction Enzyme 2 (BamHI)	1
Buffer	5
Water	35
Plasmid	8
Table 2.3 quantities of reactants used in a digestion of PCR and plasmid products. Each reaction was incubated at 37°C.	

Ligation Reaction	
Product	Volume (µl)
10x T4 DNA Ligase Buffer	2
DNA Ligase	1
Vector DNA	11
Insert DNA	6

Table 2.4 Reaction volumes used for ligation of insert into vector using Thermo Fisher T4 DNA ligase. This was incubated at 22°C for 20 min.

2.8 Yeast Transformation and control experiments

The yeasts were transformed and co-transformed, with the PGBKT7 and PGADT7 plasmids, which contained the sequences of interest. The protocol outlined in the Takara Bio 'Yeast *Transformation* System 2 User *Manual'* was used (Yeastmaker™ Yeast Transformation System 2 User Manual, 2022). The PGBKT7 plasmid contains the Gal4 DNA-binding domain (DNA-BD) and enables the yeast strain to grow in tryptophan deficient media. The PGADT7 contains the AD GAl4 activation domain and enables the yeast to survive in leucine deficient media. When proteins cloned into the two plasmids are in close proximity to each other, the AD domain activates the Gal4 DNA binding domain switching on reporter genes. The reporter genes include AUR1-c which allows the yeast to resist the toxic drug Aureobasidin A (from Takara Bio); HIS3 which enables the yeast to grow on histidine deficient media; ADE2 allows which the yeast to grow on media deficient in adenine and finally, MEL1 whichcauses the yeast to turn blue when in the presence of the chromogenic substrate X-alpha-Gal. The PGBKT7 plasmids were transformed into the Y2HGold yeast strains while the PGADT7 plasmids were transformed Y187 yeast strain. The plasmids were co-transformed into the Y2HGold yeast strain. The transformed yeast were plated on their respective synthetically defined (SD) minimal media. The transformed yeast were also tested for auto activation and toxicity. For the yeast mating experiment, colonies were isolated from the single dropout media plates. Colonies containing the PGBKT7 and PGADT7 vectors were incubated in YPDA broth at 30°C overnight. The broth was pelleted and the YPDA broth was removed. The pellet was resuspended in molecular grade biological water and spread on to plates.
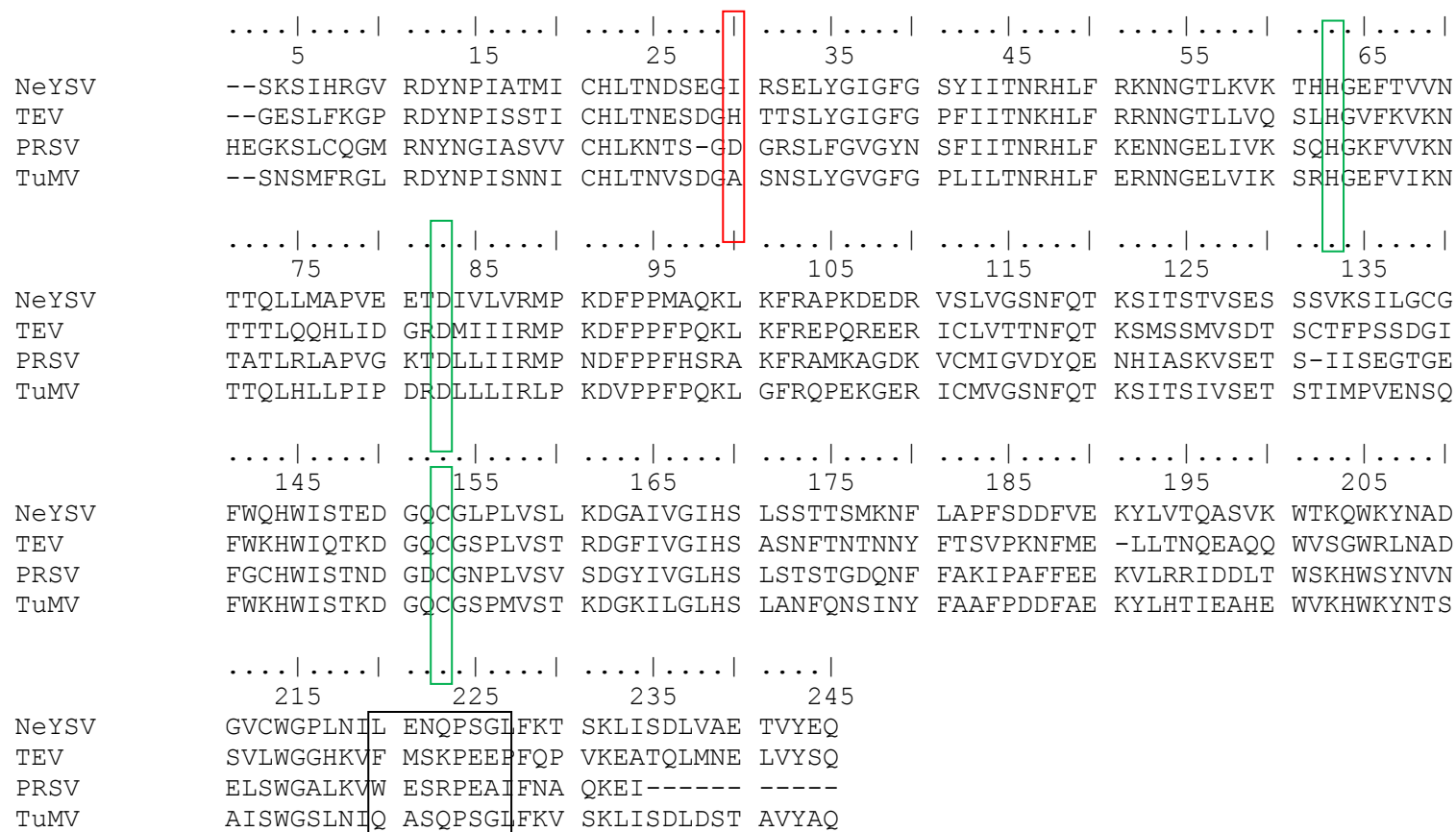
3.Results

This section outlines results gained from the experiments performed in section 2. Starting with the computational results. This includes identifying the catalytic triad and optimal cleavage sequence for the NeYSV protease (Identifying motifs 3.1). This is followed by the results from the structural alignments which supports the alignment results (Comparison between NeYSV and TEV NIa-pro structure 3.2).These sections are followed by the molecular dynamics results. The results of the possible optimal cleavage sequence docked to the NeYSV protease (Docking of peptide sequences of NeYSV 3.3). The computational results section ends on the molecular dynamics results between the viral proteases and plant proteins (Predicting interactions between proteases and SAHH 3.4). The next section of the results section contains the results of the yeast two hybrid assay (Yeast two-hybrid assay 3.5) used to support the results from the molecular dynamics results. 'myc tagged NIa-pro 3.6' attempts to determine the irregularities from the yeast two hybrid results by performing a molecular dynamics simulation on the protease with a myc tag to see if it will affect binding. Finally 'Identifying potential other interacting proteins 3.7' lists other proteins found to interact with TEV NIa-pro which have sequence data in *N. benthamiana*. These proteins may be interacted with by the proteases in both plant and so could be responsible for the host range of the virus.

3.1 Identifying motifs

Using a NCIB's protein BLAST search, it was discovered that the NeYSV's NIa-pro had a 60% similarity to the Hippeastrum mosaic virus' NIa-pro (Boratyn et al., 2019). However, there is very little information on this protease. In contrast, NeYSV's protease has a 57% similarity to Turnip Mosaic Virus (TuMV) NIa-pro. The Turnip Mosaic Virus (TuMV) is a well studied protein (Han et al, 2010; Khan and Goss, 2012; Léonard et al, 2004). By aligning TuMV NIa-pro to NeYSV NIa-pro, it was discovered that the TuMV proteases catalytic triad residues had aligned to identical residues in NeYSV protease (Figure3.1). Therefore, the catalytic triad of NeYSV NIa-pro was histidine 51, aspartic acid 81 and cytosine 151. It was also discovered, by aligning the sequences of the two proteases, that the self cleavage region of TuMV NIa-pro aligned to a region in the NeYSV protease (Figure3.1). This allowed for the identification of the NeYSV self cleavage site, which is LENQPS (Figure 3.1). The Chen et al (2008) paper determined that the lycine residue 27 of the PRSV protease was responsible for the host range of the virus. ThePRSV protease was aligned to the NeYSV protease and this residue matched to the amino acid isoleucine 28 (Figure3.1). Potentially, this residue is responsible for the host range in NeYSV.

Due to the sequence similarity between the NeYSV NIa-pro and the TuMV NIa-pro, it was predicted that the cleavage sequences of NeYSV NIa-pro would bear a strong resemblance to

22

the TuMV NIa-pro cleavage sequences. By comparing the identified cleavage sites of TuMV, with the regions separating the proteins in the NeYSV polyprotein, it was observed that the regions separating the proteins had a strong similarity to the cleavage sequences of TuMV (Table 3.1). The NeYSV predicted cleavage sites often have a valine at position four, a histadine at position two and either a glutamine or glutamic acid at position one (Table 3.1). Arguably, the NeYSV protease had a similar optimal cleavage recognition sequence as TuMV, the optimum cleavage sequence being V-X-H-G/S (Han et al., 2010). The protease may recognise this sequence and use it to bind to the polyprotein. Therefore, this sequence may be recognised on the host proteins and bound to. However, by re-examining Table 3.1, we can see that not all of the predicted cleavage sequence match the V-X-H-G/S pattern. In fact, the N-terminus of 6K1 cleavage sequence has no resemblance to the TuMV or TEV cleavage sites (Table 3.1).The NeYSV sequence contains no valine at position four, no histadine at position two and noglutamine or glutamic acid at position one. The NIb C-terminal cleavage site also contains no histadine at position two.

```
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ...|.|....|
               5         15        25        35        45        55        65
NeYSV     --SKSIHRGV RDYNPIATMI CHLTNDSEGI RSELYGIGFG SYIITNRHLF RKNNGTLKVK THHGEFTVVN
TEV       --GESLFKGP RDYNPISSTI CHLTNESDGH TTSLYGIGFG PFIITNKHLF RRNNGTLLVQ SLHGVFKVKN
PRSV      HEGKSLCQGM RNYNGIASVV CHLKNTS-GD GRSLFGVGYN SFIITNRHLF KENNGELIVK SQHGKFVVKN
TuMV      --SNSMFRGL RDYNPISNNI CHLTNVSDGA SNSLYGVGFG PLILTNRHLF ERNNGELVIK SRHGEFVIKN


          ....|....| ...|.|....| ....|....| ....|....| ....|....| ....|....| ...|.|....|
               75        85        95        105       115       125       135
NeYSV     TTQLLMAPVE ETDIVLVRMP KDFPPMAQKL KFRAPKDEDR VSLVGSNFQT KSITSTVSES SSVKSILGCG
TEV       TTTLQQHLID GRDMIIIRMP KDFPPFPQKL KFREPQREER ICLVTTNFQT KSMSSMVSDT SCTFPSSDGI
PRSV      TATLRLAPVG KTDLLIIRMP NDFPPFHSRA KFRAMKAGDK VCMIGVDYQE NHIASKVSET S-IISEGTGE
TuMV      TTQLHLLPIP DRDLLLIRLP KDVPPFPQKL GFRQPEKGER ICMVGSNFQT KSITSIVSET STIMPVENSQ


          ....|....| ...|.|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               145       155       165       175       185       195       205
NeYSV     FWQHWISTED GQCGLPLVSL KDGAIVGIHS LSSTTSMKNF LAPFSDDFVE KYLVTQASVK WTKQWKYNAD
TEV       FWKHWIQTKD GQCGSPLVST RDGFIVGIHS ASNFTNTNNY FTSVPKNFME -LLTNQEAQQ WVSGWRLNAD
PRSV      FGCHWISTND GDCGNPLVSV SDGYIVGLHS LSTSTGDQNF FAKIPAFFEE KVLRRIDDLT WSKHWSYNVN
TuMV      FWKHWISTKD GQCGSPMVST KDGKILGLHS LANFQNSINY FAAFPDDFAE KYLHTIEAHE WVKHWKYNTS


          ....|....| ...|.|....| ....|....| ....|
               215       225       235       245
NeYSV     GVCWGPLNIL ENQPSGLFKT SKLISDLVAE TVYEQ
TEV       SVLWGGHKVF MSKPEEPFQP VKEATQLMNE LVYSQ
PRSV      ELSWGALKVW ESRPEAIFNA QKEI------ -----
TuMV      AISWGSLNIQ ASQPSGLFKV SKLISDLDST AVYAQ
```

Figure 3.1 Alignment between the NeYSV (QOD42425.1), TEV(NC_001555.1), PRSV(APM86423.1) and TuMV(NP_734220.1) proteases. The PRSV lysine that is responsible for the host range has been mutated into aspartic acid. The location of this mutation and the residue from NeYSV that it is aligned to are outlined in red. The location of the catalytic triads of the proteases and their aligned residues are outlined in green. The location of the TuMV and NeYSV autocatalytic site is outlined in black.

24
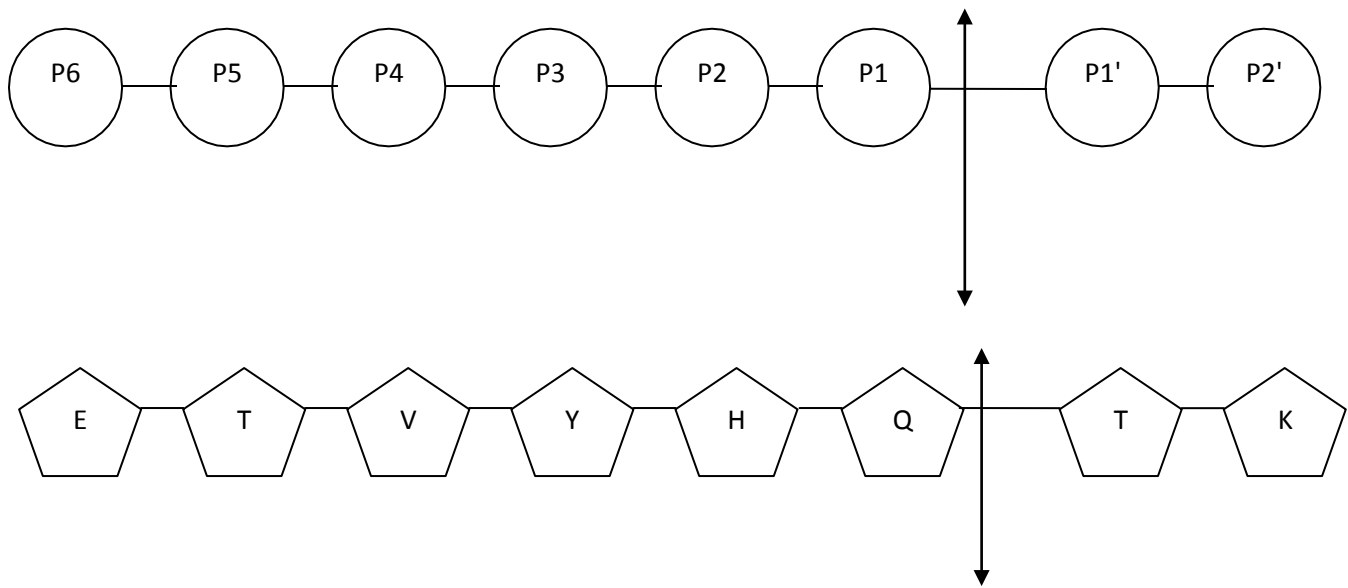
Figure 3.2 Diagrammatic representation of a protease and its substrate showing the nomenclature. The substrate residues are labelled progressively from P1 N-terminal of the cleavage site and P1′, etc., C-terminal of the cleavage site. This is followed by a diagrammatic representation of the NeYSV protease substrate for the Cl C-terminal (Adams, Antoniw and Beaudoin, 2005).

	NIa-pro Recognition Sequence		
Region of cleavage site on polyprotein.	Turnip Mosaic Virus (UniProtKB - Q02597)	Nerene Yellow Stripe Virus (QOD42425.1)	Tobacco Etch Virus (NC_001555.1)
6K1 N-terminal	1168-EKVVH**Q** ↓ **V**-1176	1326-EVQYQH ↓ K-1333	1104-EDLVE**Q** ↓ **A**-1111
6K1 C-terminal	1220-**EPTVYHQ↓T**-1227	1376-EDNVVH**Q** ↓**S**-1382	1156-EREIIYT**Q** ↓**S**-1164
Cl C-terminal	1863-EAVNH**Q** ↓ **S**-1870	2010-ETVYH**Q** ↓ **T**-2017	1791-ETIYL**Q** ↓ **S**-1797
6K2 C-terminal	1918-EPVTH**E** ↓ **A**-1923	2063-EDVSF**E** ↓**A**-2070	1844-EP**V**YF**Q** ↓ **G**-1850
VPg C-terminal	2108-ELVP**V**DH**E** ↓ **S**-2116	2258-ESPVLINQVHH**E** ↓**S**-2271	2032-EDLTF**E** ↓**G**-2038
NIa-pro C-terminus	2351-DSTA**V**YA**Q** ↓ **T**-2359	2508-ET**V**YE**Q** ↓ **N**-2514	2274-EL**V**YS**Q** ↓ **G**-2280
NIa-self cleavage region	2333-QAS**Q**P**S** ↓**G**-**2340**	2485-LEN**Q**P**S** ↓ **G**-2494	2250-GHKVFM ↓S-2257
NIb C-terminus	2859-EAC**V**YH**Q** ↓ **A**-2876	3025-MVD**V**IF**Q** ↓**S**-3033	2786-ENLYF**Q** ↓**S**-2792

Table 3.1 The regions cleaved by TuMV NIa-pro, TEV NIa-pro and the predicted regions cleaved by NeYSV NIa-pro, determined thorough alignment of the polyproteins. The residues coloured in red represent identical residues that are in identical positions of the cleavage site. The blue downward facing arrow represents the region of the recognition sequence that is cleaved by the protease. The green highlight represents significantly important residues that do not fit with the identified cleavage recognition pattern.

3 .2 Comparison between NeYSV and TEV NIa-pro structure

By aligning the crystal structure of TEV protease to the NeYSV protease homology model, it showed that the predicted catalytic resides for NeYSV were in an identical physical location to the TEV proteases catalytic residues (Figure 3.3). This provides further data that the catalytic triads, identified in the sequences alignment, are correct. This data enables us to determine the location of the catalytic site on the protease. Further comparisons between regions of the

26

NeYSV protease to the TEV protease determined that the major variations between the proteases lie on the loops on C-terminus (191-245).
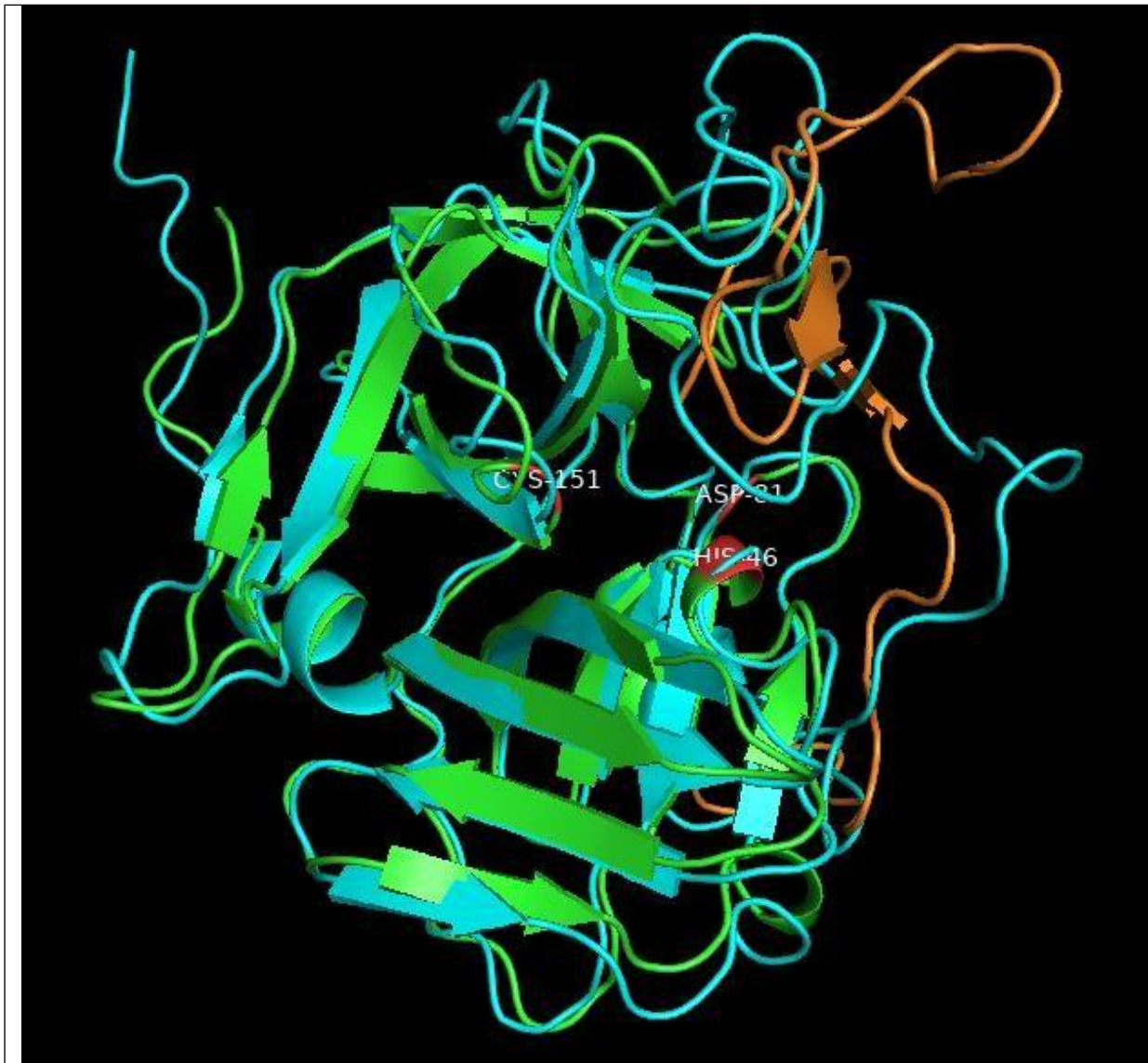
Figure 3.3 The equilibrated NeYSV protease in green aligned to the TEV crystal NIa-pro structure (blue). Residues from NeYSV located in the same region of TEV catalytic triad are highlighted in red. The C-terminal from TEV is depicted in orange. The RMSD of this alignment is 1.452 Å

3.3 Docking of Peptide Sequence to NeYSV

To confirm that the sequences identified in Table 3.1 were sequences that would be recognised and cleaved by the protease, a sequence matching the predicted optimal cleavage sequence (bearing a strong similarity to the TuMV cleavage sequence) was docked to, and equilibrated with the NeYSV protease. After equilibration, it was determined that the predicted NeYSV cleavage sequence was able to fit inside the NeYSV NIa-pro catalytic site (Figure 3.4b). As this peptide sequence can occupy the catalytic site, arguably, it is cleaved by the protease.

27

However, the peptide cleavage site lies 3.7 Å away from the nucliophilic cytosine (Figure 3.4c). This is too far for a hydrogen bond. Therefore, the electron cannot be donated to split the peptide (Figure 3.4c). This catalytic site is composed of three core hydrophobic residues Isoleucine 28, Leucine 32 and Leucine 47. These were identified using the electrostatic properties of the protein as negative energy change implies hydrophilicity (figure 3.4a). These residues also matched core hydrophobic residues as identified by Adams, Antoniw and Beaudoin, (2005).
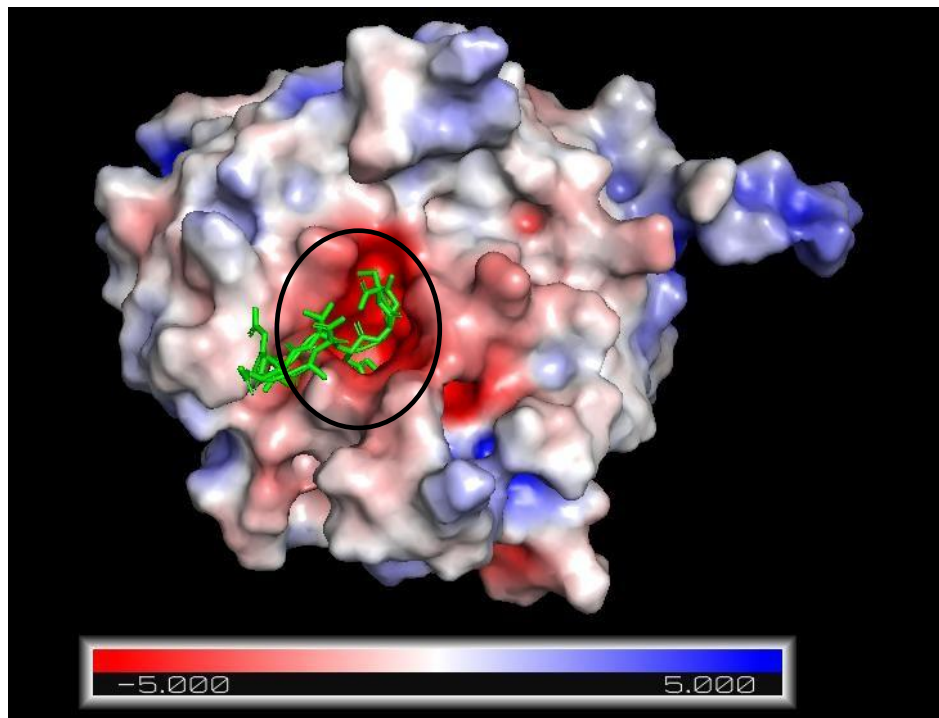
Figure 3.4a An equilibrated model of the NeYSV NIa-pro showing the surface of the protein. The electrostatic properties of the protein have been calculated. Areas in red are negatively charged areas whereas blue areas are positively charged. The peptide sequence of EDNVVH**Q** is displayed in green. The black circle represents the location of the hydrophillic residues.
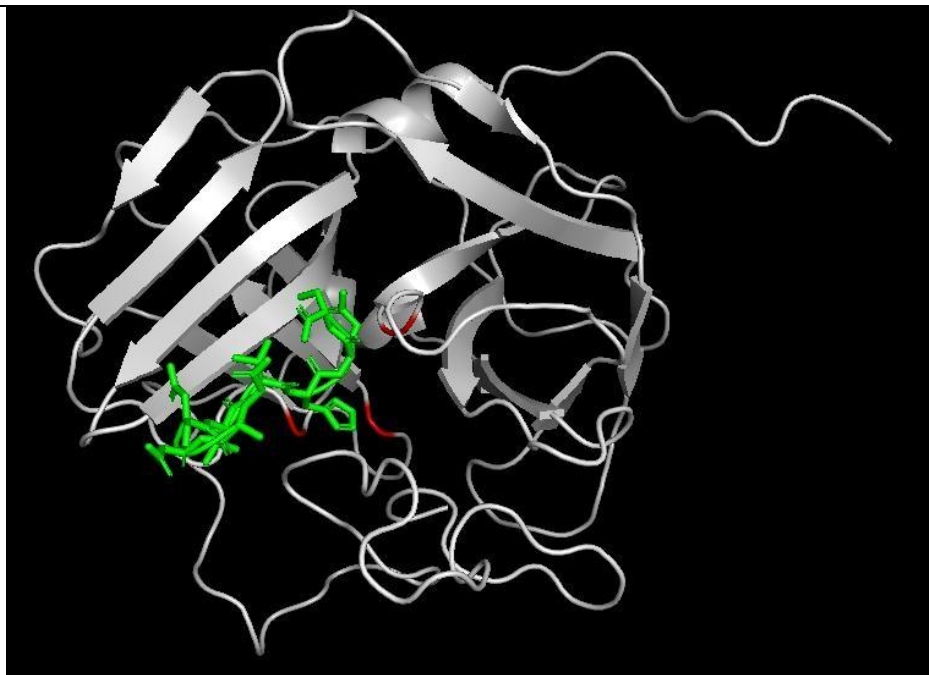
Figure 3.4b An equilibrated model of the NeYSV NIa-pro. The catalytic triad has been highlighted in red. The peptide sequence of EDNVVH**QS** is displayed in green. The catalytic triad is part of the proteins active site. Therefore the peptide, in this region, is within the active site.
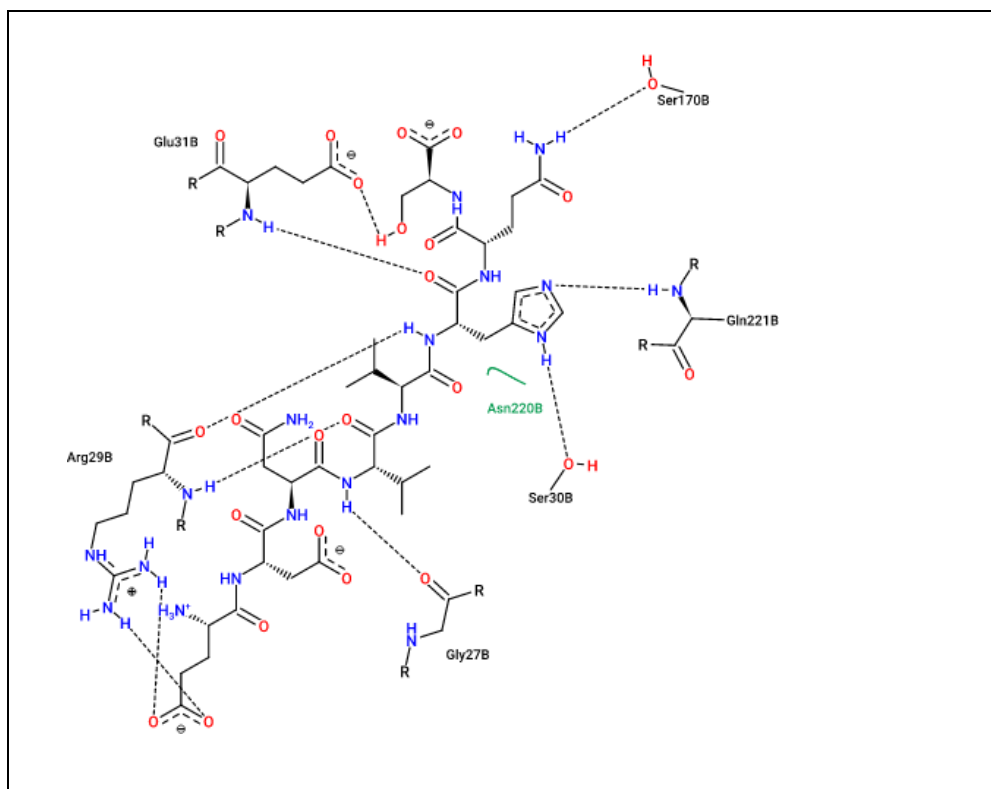
Figure 3.4c A 2D interaction diagram, created from the final frame of the equilibrated structure, between the peptide sequence and NeYSV NIa-pro. The image displays interactions that take place between the two. The C-terminus of the peptide, which is splitby the protease, is oriented towards the catalytic triad. This region is stabilised by glutamic acid 31, serine 170, glutamine 221, serine 30 and hydrophobic interactions with asparagine. The N-terminus of the peptide forms hydrogen bonds with arginine 29 and glycine 27.

3.4 Predicting interactions between proteases and SAHH

In order to predict the interaction between the viral proteases and host, AtSAHH2 was docked to TEV's NIa-pro and NeYSV NIa-pro. The structures were minimised, heated, equilibrated and finally the constraints were removed and an MD simulation was run for 25ns. Using frames from the clustering data of the MD simulation, the affinity between the protease and the host protein was determined. The cut-off values for the affinity interaction was 0 as the free energy must be negative for a reaction to be spontaneous (Noskov and Lim, 2001). With a positive ΔG there is a need for an input of energy in order for the interaction to take place (Noskov and Lim, 2001). The average affinity of the interaction between SAHH bound to NIa-pro was computed, by determining the protein-protein interaction affinity, for all the frames of the largest cluster in the simulation. The interaction between TEV and AtSAHH2 had an average affinity of -17.2 ΔG, kal/mol. In contrast, the average affinity between NeYSV NIa-pro and AtSAHH2 was determined to be -3.13ΔG, kal/mol. This interaction is five times

30

weaker than the TEV NIa-pro AtSAHH2 interaction. Using crystal structure data, previous papers have determined that trypsin complexed with bovine pancreatic trypsin inhibitor (an extremely stable and cleavage-resistant complex) has an affinity of -11.4ΔG, kal/mol at 298k (Noskov and Lim, 2001). Using this data as a benchmark, we could be confident that AtSAHH2 and TEV SAHH2 interaction takes place outside of a simulation. However, we cannot be confident that the NeYSV NIa-pro AtSAHH2 interaction could take place as the affinity is much lower.

To further support this hypothesis the Root Mean Squared Derivative (RMSD) model results show that between TEV and AtSAHH2, at the interacting surfaces, the RMSD was lower than the rest of the proteins (Figure 3.5). RMSD measures the difference between the structure of the protein at the start of the simulation and the end. The stability of the protein can be determined by how much the structure of the protein is altered from its original position. Therefore, the interacting surface is less subject to change and so the hydrogen bonds generated will be more stable. In the NeYSV and AtSAHH2 interaction model, the RMSD over the course of the simulation, is 1 angstrom larger than TEV AtSAHH2 (Figure 3.6b). This indicates that overall, the interaction was less stable as there was more movement of the structures. The Root Mean Square Fluctuation (RMSF) measures how much specific residues move in a simulation. The RMSF of NeYSV was much lower than the TEV protease (Figure 3.6a). Although the NeYSV structure was less stable over the simulation, the individual residues of TEV were not. This indicates that the TEV NIa-pro is more flexible than NeYSV NIa-pro.

a

b
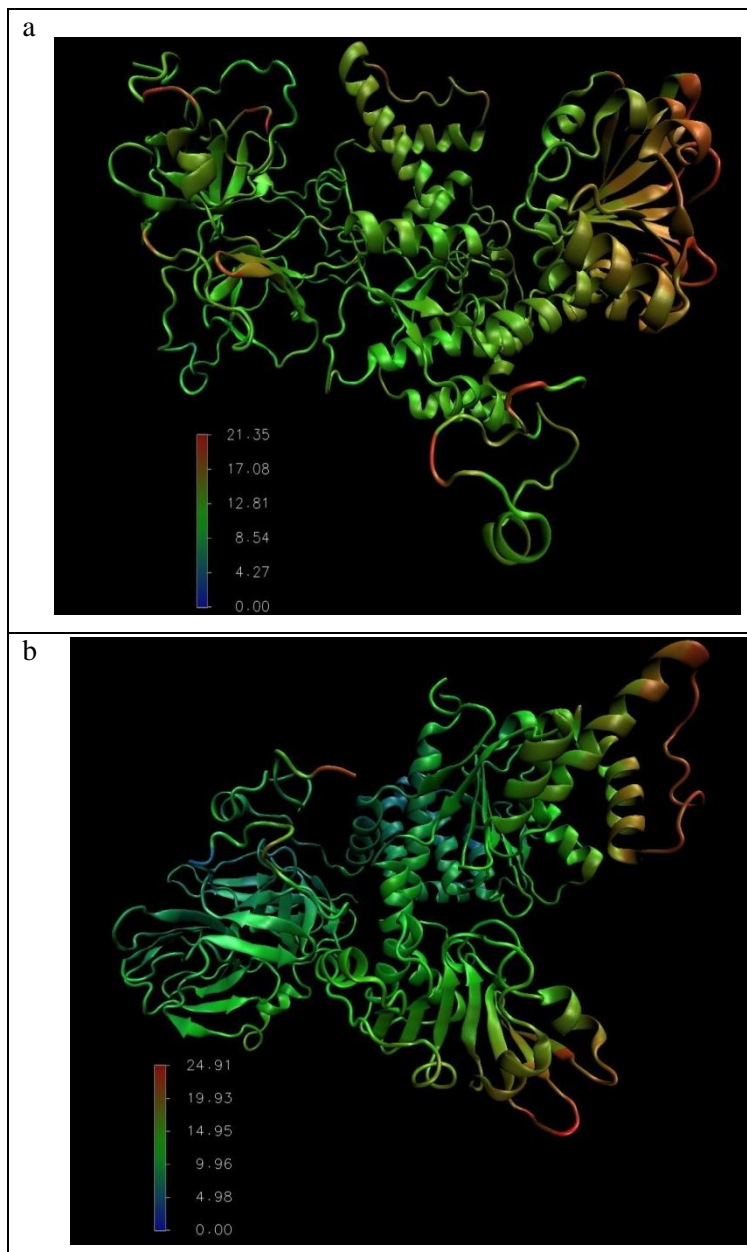
Figure 3.5 The RMSD values for the proteases conjoined with SAHH2 were calculated using VMD over the course of the simulation. These values were mapped onto the model. The scale below the model represents the relative RMSD (Å) value and the colour it is assigned. Image (a) shows AtSAHH (left) docked to NEYSV's protease (right). The colour of the NeYSV protease is predominantly green, indicating an average RMSD (Å) of 8. Image (b) shows TEV NIa-pro bound to AtSAHH2. The regions of the TEV protease interfacing with the AtSAHH model have a lower RMSD (Å) value as indicated in blue. The AtSAHH2 beta chain containing residue 304 has a blue colour indicating a low RMSD (Å). The TEV residue 135 also has a blue colour

RMSF of NySV's NIA-pro bound to AtSAHH2
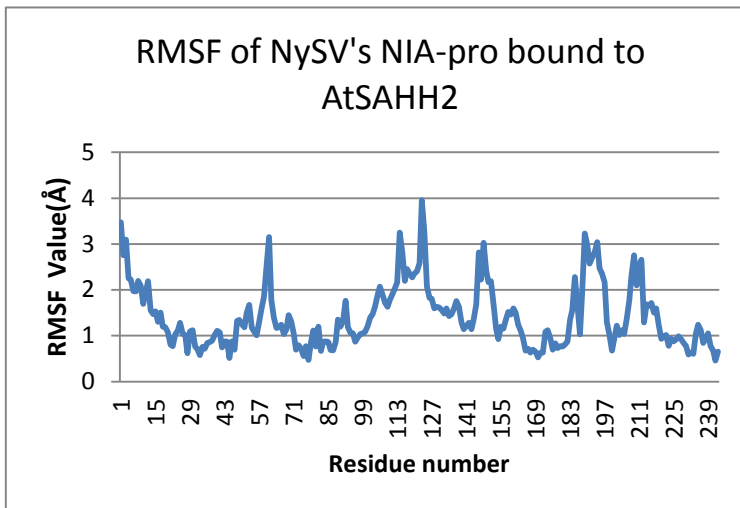
RMSD of NySV's NIa-pro docked to SAHH2 over the course of the simulation

RMSF of TEV's NIA-pro bound to AtSAHH2
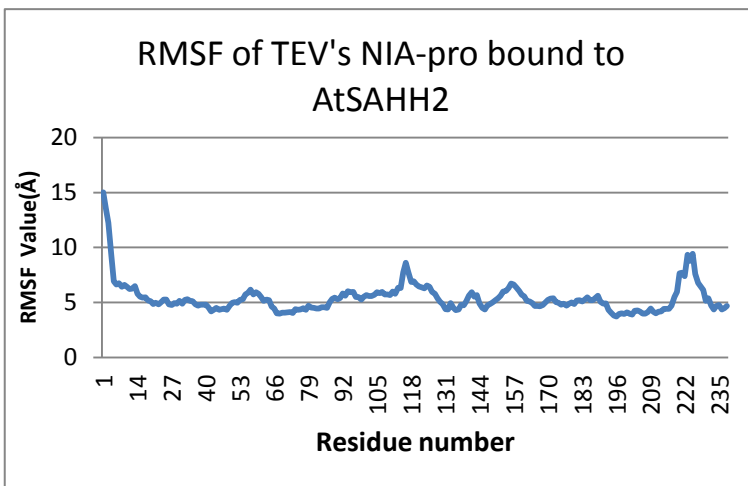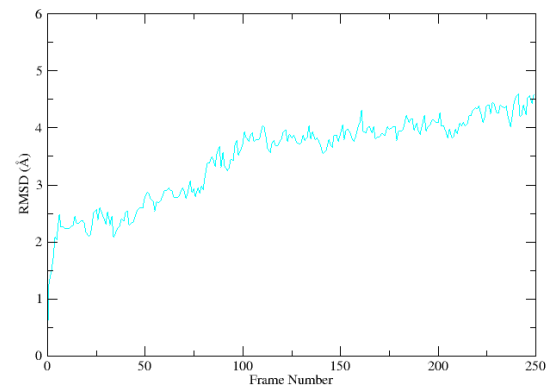
RMSD of TEV's NIa-pro docked to SAHH2 over the course of the simulation
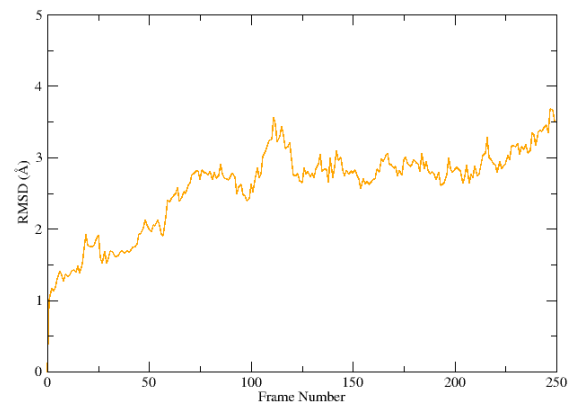
Figure 3.6a (left) Graph displaying the total RMSF of NIa-pro from the molecular dynamics simulation in which NIa-pro was bound to AtSAHH2. The graph depicts that the region in the NeYSV protein with thehighest degree of movement is residue 127 whereas, the lowest is residue 85. In the TEV protease, the RMSF of all the residues appears to be around 5 Å except for the first residue which is significantly higher. **Figure 3.6b (right)** Graphs displaying the total RMSD, over the course of the molecular dynamics simulation, in which NIa-pro was bound to SAHH2. The RMSD of the simulations appears to level off

By examining the binding data generated by the MD between the NeYSV protease and the host protein AtSAHH2, visual evidence depicted that the viral protein resided partly on the insertion region (amino acids 150-190) and mid section of the protein (amino acids 180-230) (Figure 3.7a). Unique hydrogen bond analysis between NeYSV NIa-pro and AtSAHH2

33

showed that the protease formed strong hydrogen bonds between the C-terminal of the protease and the insertion region of SAHH2. The lysine reside227 from NIa-pro bound to SAHH2's aspartate188 with an average occupancy of 50% (figure 3.7b). Also, in all three simulations, the NIa-pro residue arginine 29 was found to associate to glycine 130 (average 30% occupancy). Alanine scanning determined that mutations to the arginine residue may increase the entropy of the system by 1.42ΔG, kal/mol. However, the largest increase in entropy occurred via mutations to the lysine 227 which caused an increase in free energy by 1.79ΔG, kal/mol. Analysis by MutaBind suggests that mutations to both these residues may cause the complex to become unstable.
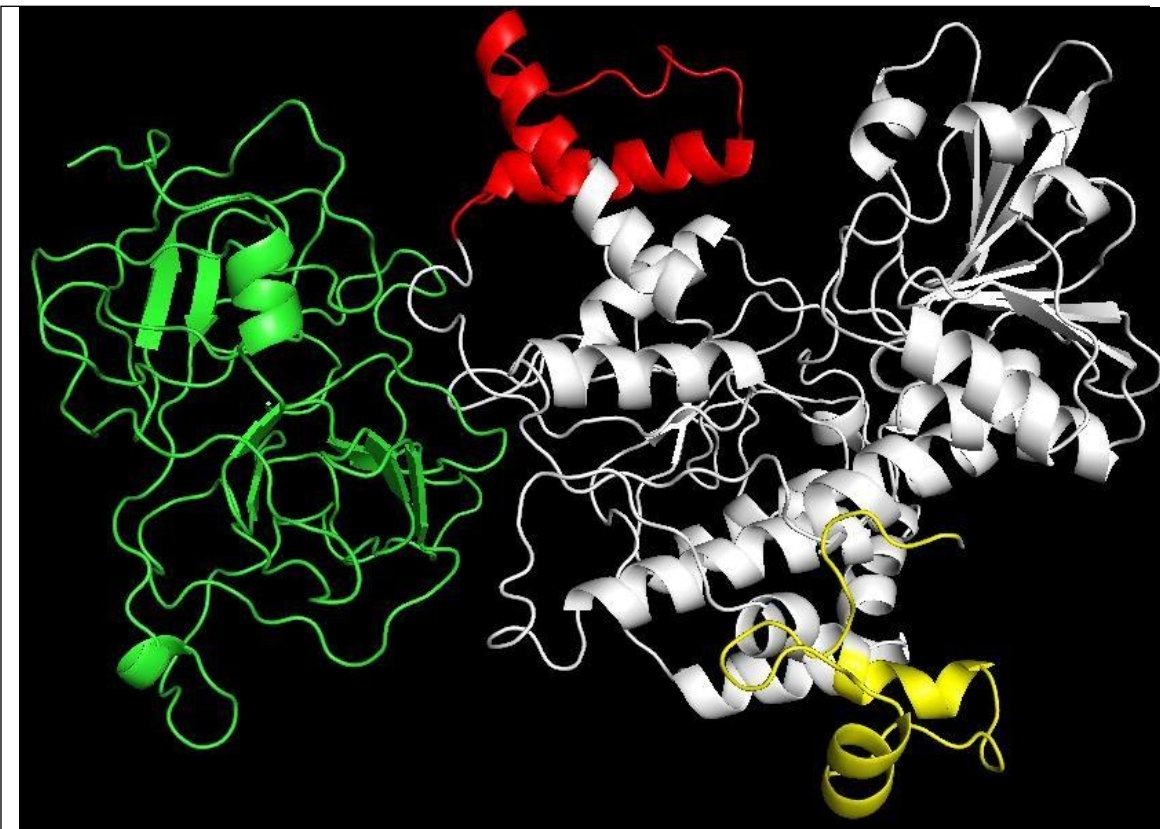
Figure 3.7a Frame of molecular dynamics performed on NeYSV NIa-pro docked to AtSAHH2. NIa-pro is depicted green, AtSAHH2 grey, the insertion region as red and the C-terminal as yellow. The affinity of this interaction is -2.6 ΔG, kal/mol
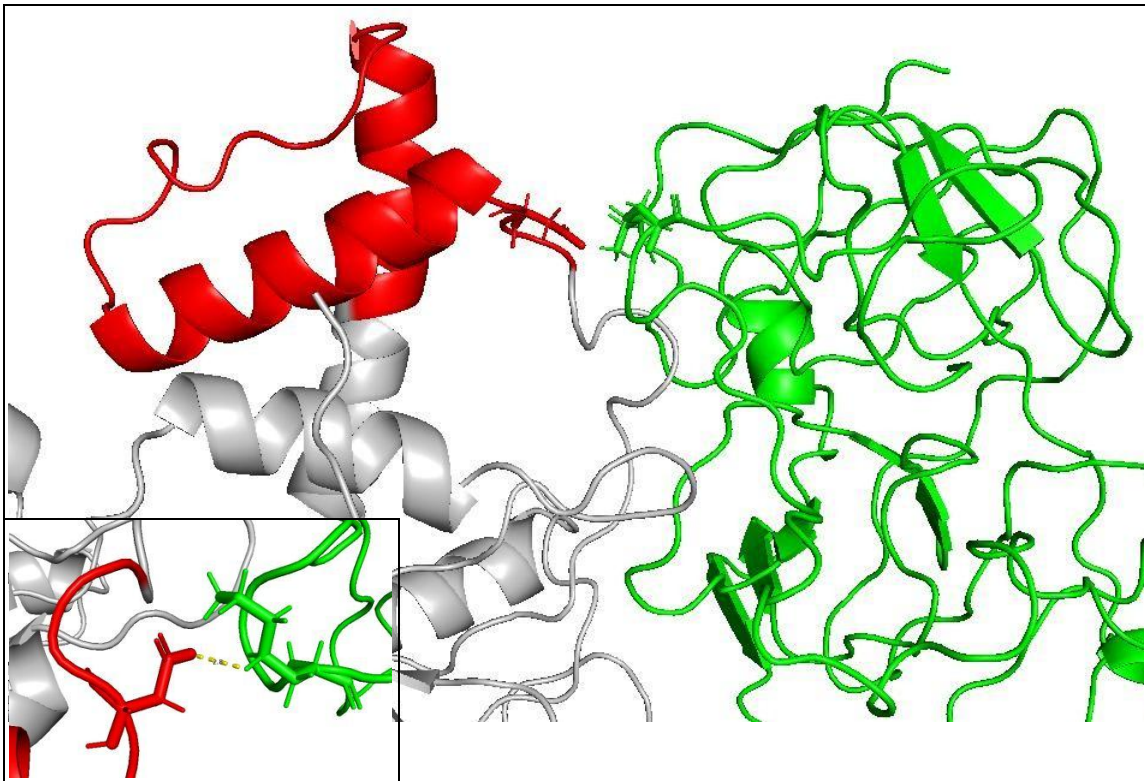
Figure 3.7b AtSAHH2 (grey) docked to NIa-pro from NeYSV. The image shows a hydrogen bond between the residues lysine227 from NIa-pro (green) and asparate 188 from SAHH2's insertion region (red). The length of this hydrogen bond is 1.6 Å.

In the interaction between TEV's NIa-pro and AtSAHH2, hydrogen bond analysis determined that the protease residue, serine 135, and residue 304 from AtSAHH2 had an occupancy of 53% (Figure 3.8b). The second strongest hydrogen bond occurred between arginine 105 (NIa-pro) and glutamine 35 (SAHH2) which had an 18% occupancy. Despite binding with residues associated with centre of AtSAHH2, physically, the TEV's NIa-pro occupies the C-terminal region (449-485) (Figure 3.8a). The C-terminal region was determined by aligning the *Arabidopsis* SAHH2 to human SAHH (Figure 6.11). Alanine scanning, performed in-silico using mutabind (Li et al., 2016), showed that mutations to the TEV NIa-pro's arginine 105 had a larger detrimental effect on the protein structure which increased the entropy of the system by 1.49 ΔG, kal/mol. In contrast, mutations to the TEV NIa-pro serine 135 increased the entropy by 0.87 ΔG, kal/mol. AtSAHH2 with NeYSV NIa-pro and ATSAHH2 bound to TEV NIa-pro models were aligned and by examining this data, it was determined that the two protease bind in completely different regions of AtSAHH2(Figure 3.8c).

35
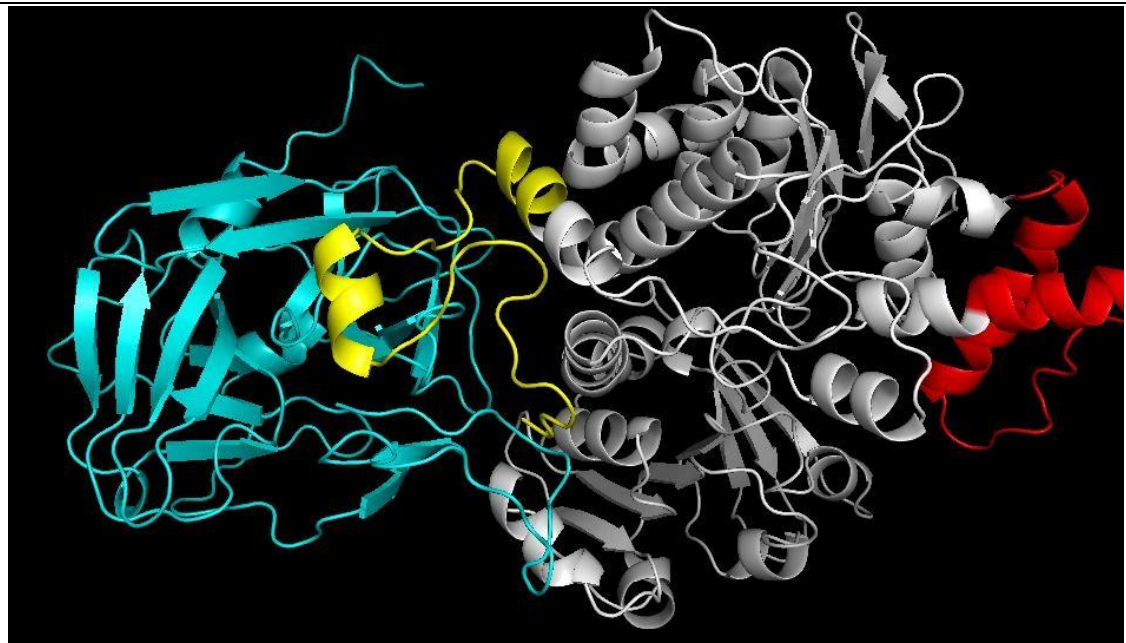
Figure 3.8a Model displaying the TEV NIa-pro bound to AtSAHH2 (grey). The C-terminal region is depicted yellow and the IR region is depicted as red. The affinity of this interaction is -17.4 ΔG, kal/mol
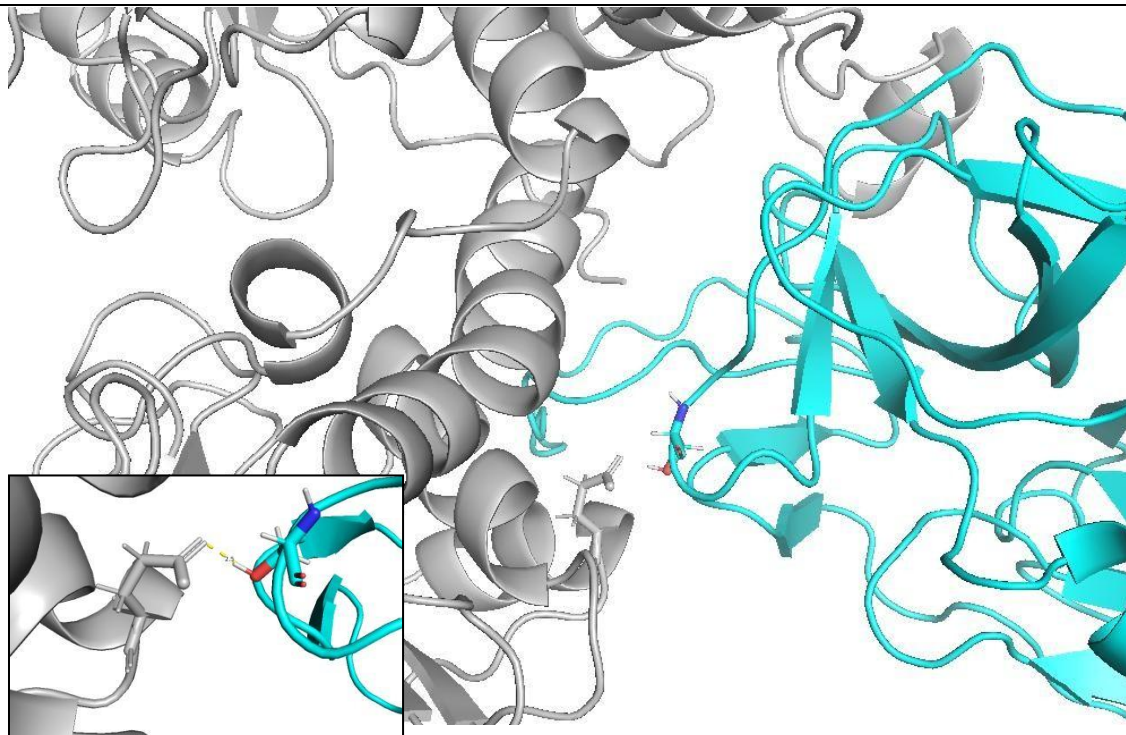
Figure 3.8b Model displaying TEV protease (blue) serine 135 residue hydrogen bonded to the AtSAHH2 (grey) glutamic acid 304 residue. The length of this hydrogen bond is 3.1 Å.
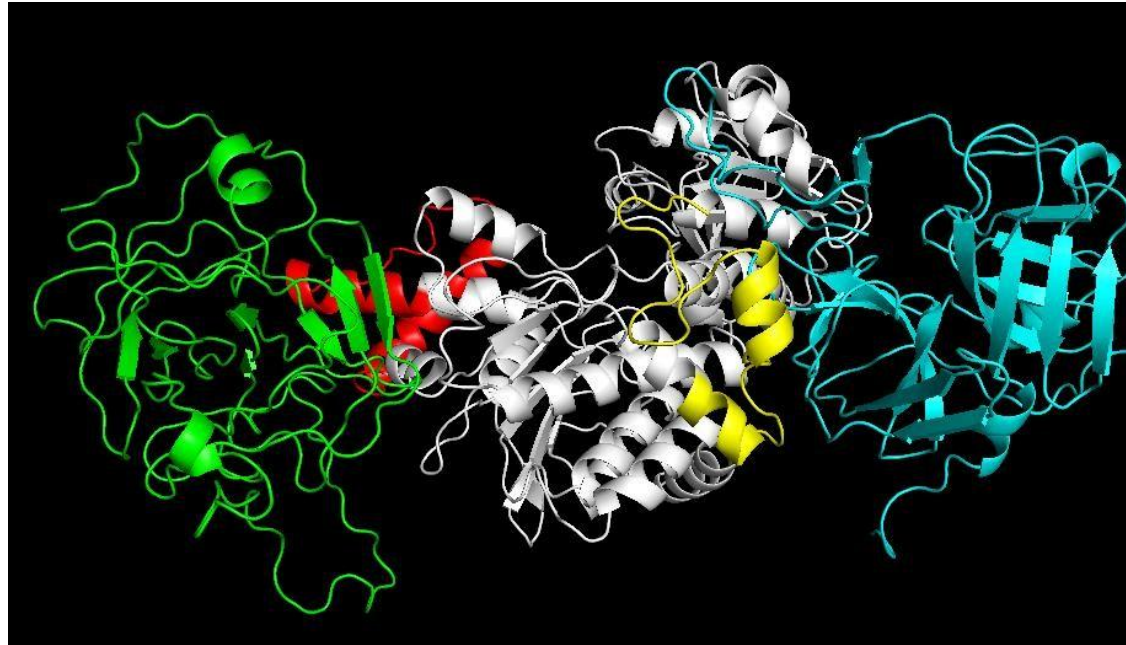
Figure 3.8c Image created by aligning the protein complexes. TEV protease is depicted as blue, the NeYSV protease green, AtSAHH white, the insertion region as red and the C-terminal tail as yellow.

The proteases were docked to *N. benthamiana* SAHH (NbSAHH) and underwent molecular dynamics simulations. The interaction between NeYSV and NbSAHH had the protease oriented in a similar position to the region it was located on AtSAHH2 (Figure 3.7a). The NeYSV NIa-pro was located on the C-terminal next to the insertion region and the C- terminal tail (Figure 3.9a). The average affinity between this interaction complex was determined to be -2.95 ΔG, kcal/mol. Although the negative delta G suggests a stable complex,the interaction energy is close to zero. This indicated that this interaction was weak and may not exist outside of the simulation. This weak interaction was supported by the RMSD and RMSF results. Visually, the RMSD of the NeYSV NIa-pro and NbSAHH is higher (and therefore more unstable) around the intersurface region (Figure 3.10). The RMSF and RMSD, over the course of the simulation, were also much higher in comparison to the docking with AtSAHH (Figure 3.11). This suggests that the interaction between NeYSV NIa-pro and NbSAHH was less stable. A single hydrogen bond was discovered to occur between threonine 65 (NIa-pro) and glutamic acid155 from the insertion region of NbSAHH. This hydrogen bond had an average occupancy 35.4%.

37
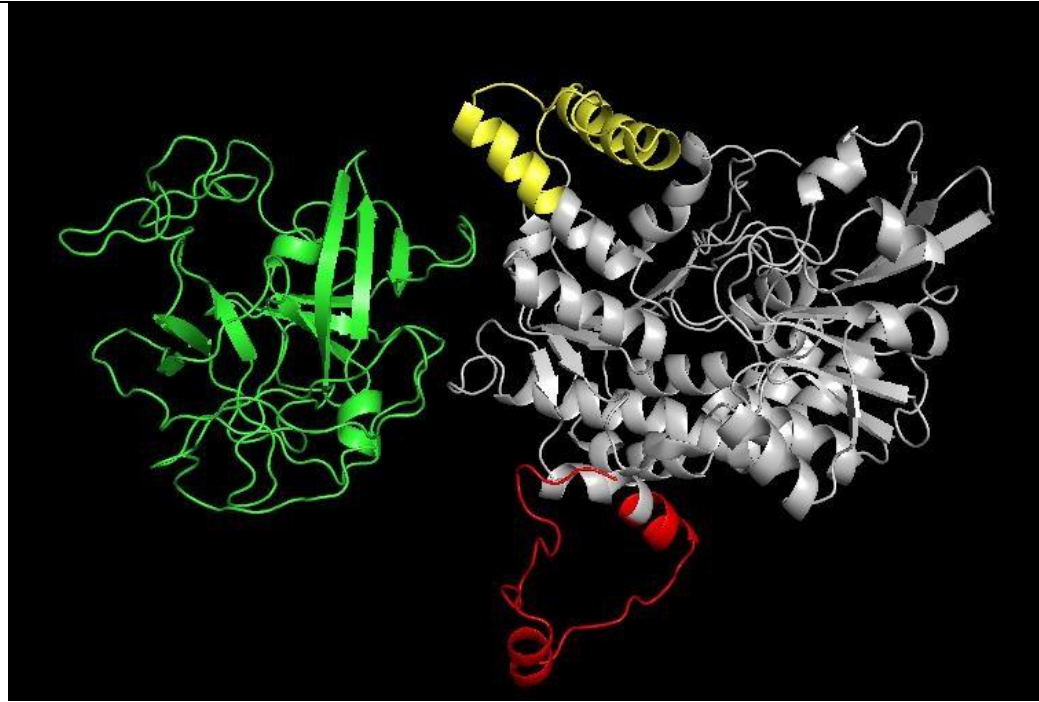
Figure 3.9a NeYSV protease (green) bound to NbSAHH (white) the insertion region is depicted yellow and the C-terminal tail red. The affinity of this interaction is -3.4 ΔG, kal/mol
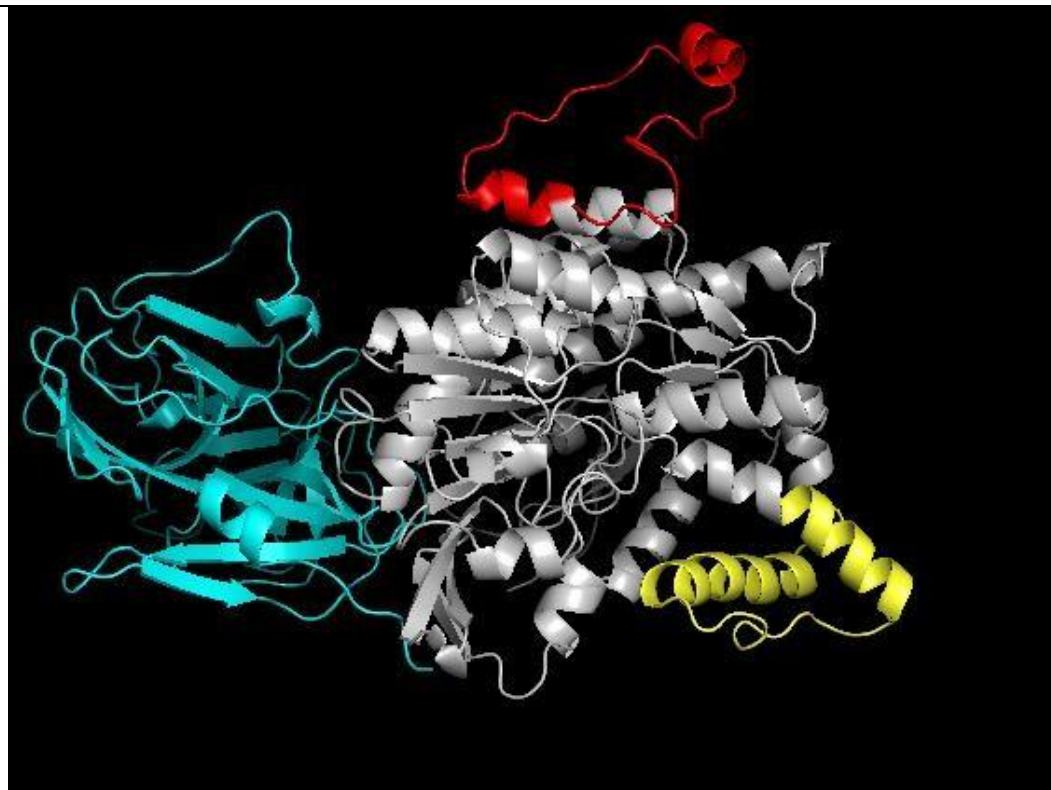

Figure 3.9b NeYSV protease (green) bound to NbSAHH (white), the insertion region is depicted in yellow and the C-terminal tail is depicted in red. The affinity of this interaction is -2.9 ΔG, kal/mol
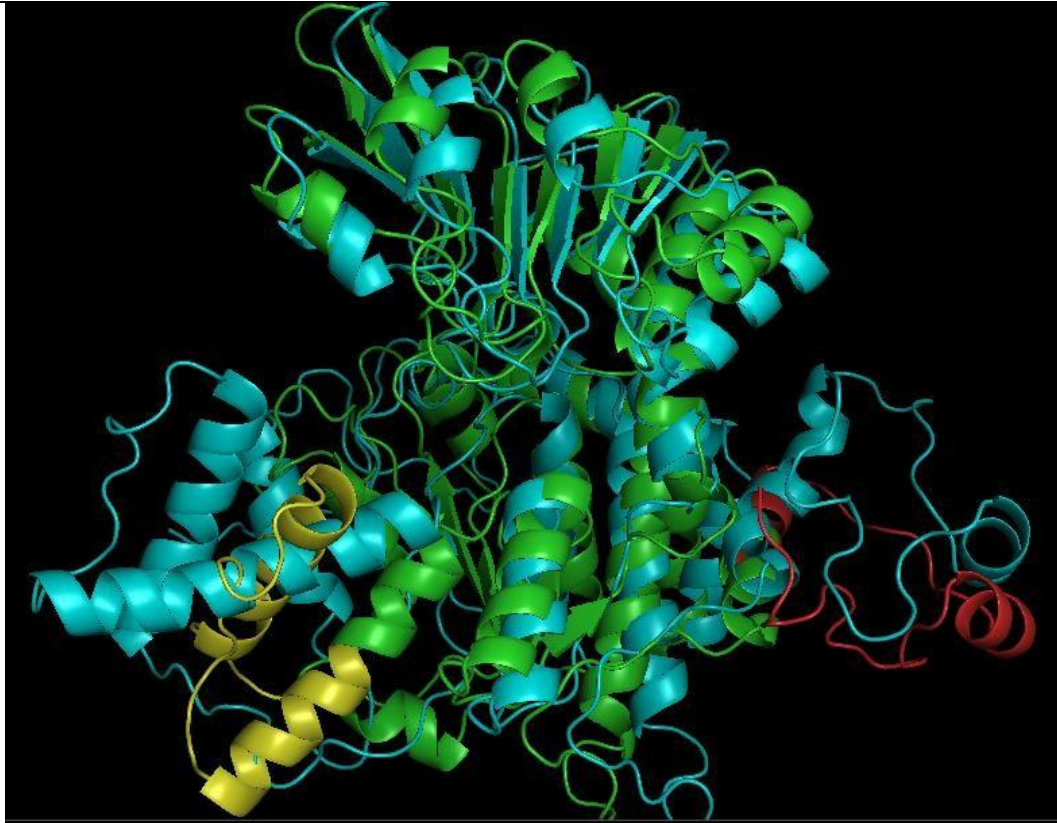
Figure 3.9c Alignment between NbSAHH (blue) and AtSAHH (green), the RMSDof the
alignment is 5.237 Å. Much of the variation between the structures appeared
to occur on the C-terminal tail region (red) and N-terminal insertion region (yellow).

The TEV NIA-pro interaction with NbSAHH showed the protease in a different location to the
AtSAHH simulations. The protease was oriented around the middle of the protease and did not
obscure or interact with any critical compartments of the host protein (Figure 3.9b). The
average affinity of this interaction was -3 ΔG kal/mol. This was lower, in comparison to TEV
NIa-pro bound to AtSAHH2, suggesting this interaction was weak and may not exist outside of
the simulation. Furthermore, the RMSD (over the course of the simulation) and visual RMSD
(in the intersurface areas) were higher than the TEV AtSAHH simulation (Figure 3.10; Figure
3.11b). In contrast, the RMSF of the TEV NIA-pro was lower when docked to NbSAHH
(Figure 3.11a). The study showed that the arginine 105 (NIa-pro) bound to glutamine 28 with
an occupancy of 28%. Glycine 7 (NIa-pro) was found to form hydrogen bonds with threonine
370 (NbSAHH). This bond had an average occupancy 14.82%.

a


```
58.62
46.89
35.17
23.45
11.72
 0.00
```

b


```
18.45
14.76
11.07
 7.38
 3.69
 0.00
```
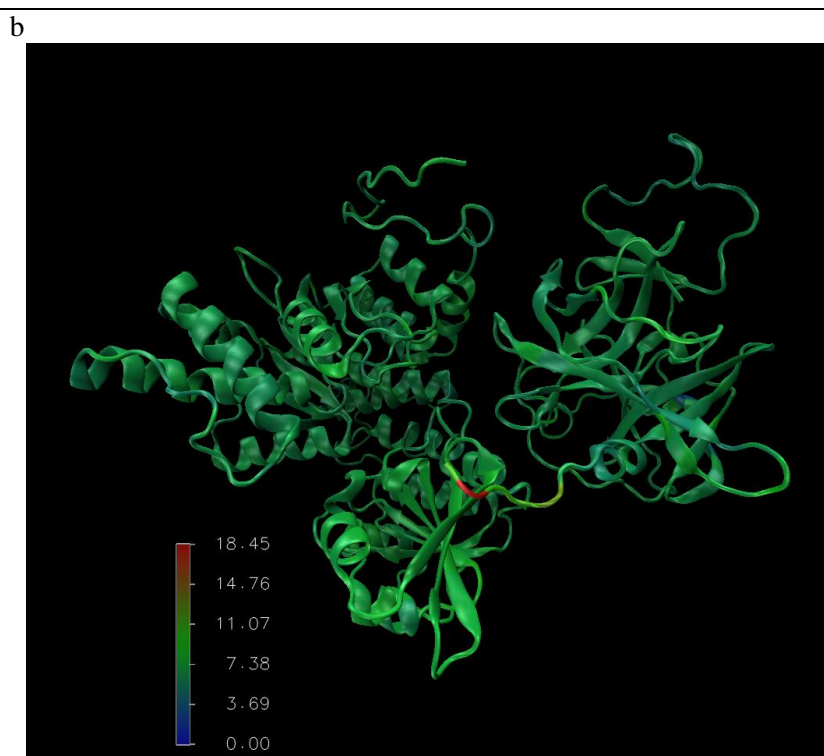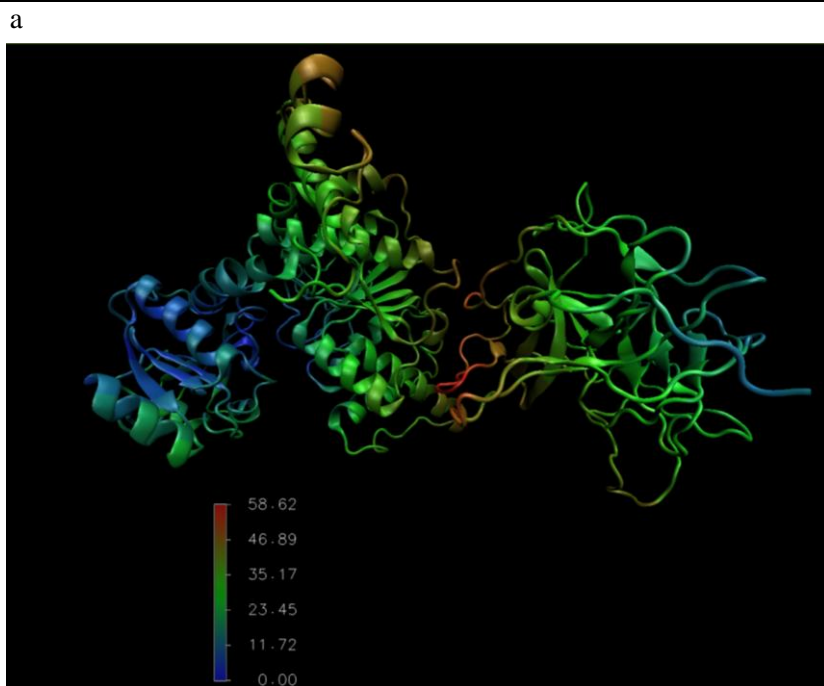
Figure 3.10 The RMSD values for the proteases conjoining with NbSAHH were calculated using VMD over the course of the simulation. These values were mapped ontothe model. The scale below the model represents the relative RMSD (Å) value and the colour it is assigned. Image (a) shows NbSAHH (right) docked to NeYSV's protease. The intersurface area of both NeYSV NIa-pro and NbSAHH appeared red indicating a high degree of movement during the simulation. Image (b) shows TEV NIa-pro (right) bound to NbSAHH. The colour of the TEV protease was predominantly green indicating an average RMSD(Å) of 8. However, the RMSD results determined that the C-terminal tail of the protease displayed a high degree of movement (18 Å), depicted as red in the image, this indicated a higher degree of movement in comparison to the majority of the protein.
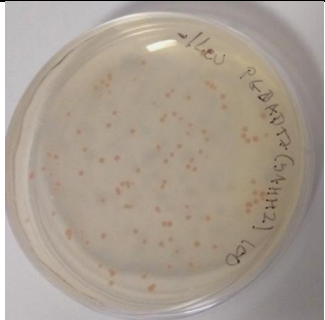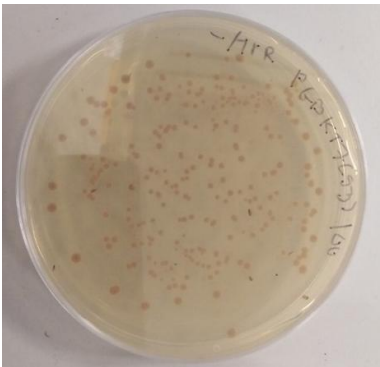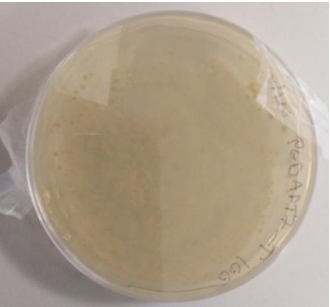
40
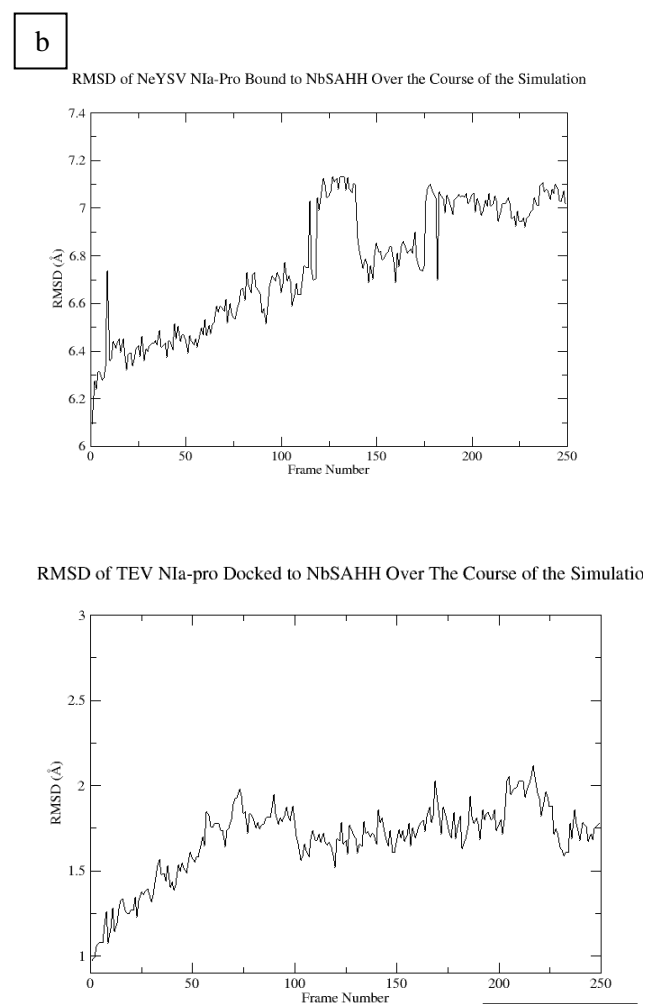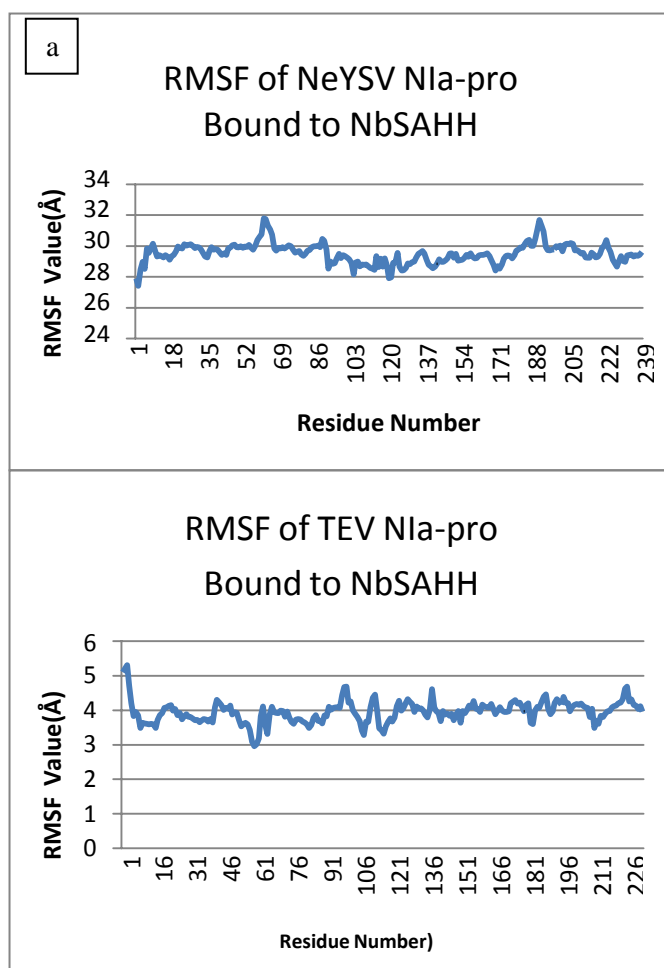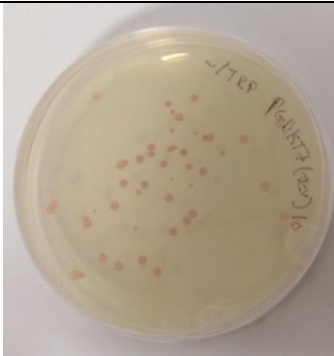
Figure 3.11a (left) Graph displaying the total RMSF of NIa-pro from the molecular dynamics simulation in which NIa-pro was bound to NbSAHH2. The RMSF graph shows that the NeYSV protein had the highest degree of movement. The RMSF, of all the residues, appeared to be around 30Å. In the TEV protease, the RMSF of all the residues appears to be around 4Å except for the first residue which is significantly higher.**Figure 3.11b (right)** Graphs displaying the total RMSD of the molecular dynamics simulation in which NIa-pro was bound to NbSAHH2. The RMSD of the TEV simulation appears to level off around frame100.However, the NeYSV RMSD does not level off until frame 200 and appears unstable.
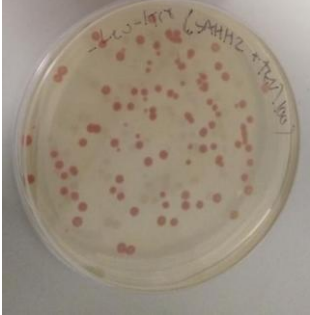
3.5 Yeast Two-Hybrid Assay

To validate the findings of the MD simulations, the TEV NIa-pro and AtSAHH2 were introduced to yeast plasmids then co-transformed into the Y2HGold Yeast Strain. Also, in order to perform the control experiments, the TEV NIa-pro and AtSAHH2 plasmids were then separately transformed into the Y2HGold Yeast Strain and Y187 Yeast Strain respectively. The TEV NIa-pro was cloned into the PGBKT7 (bait) vector. This vector enabled the Y2Hgold yeast strain to survive in agar plates which were deficient in the nutrient tryptophan (Figure 3.12). The AtSAHH2 was cloned into the PGADT7 plasmid. This enabled growth of Y187 yeast strain in media deficient of leucine (Figure 3.12). Co-transformed Y2HGold yeast containing both vectors and are resistant to both leucine and tryptophan deficiencies (Figure 3.12). The yeast strains Y2HGold and Y187 transformed with the PGBKT7 and PGADT7 vectors, which contained the viral NIa-pro and plant SAHH, were evaluated for toxicity by comparing the sizes of the cells to yeast cells transformed with the control plasmids PGBKT7-53 and PGADT7-T. As the size of the yeast colonies containing the viral and plant DNA were the same as the yeast colonies containing the control plasmids we determined that the viral and plant DNA had little to no toxicity on the yeast (Figure 3.12). Also, the transformation efficiencies of the yeast appeared to be adequate (Figure 3.12). This indicated that the transformed yeast cells were competent. When the co-transformed yeast was spread onto plates, with X-alpha-Gal, no cells appeared blue (Figure 3.13). When the two proteins expressed in the yeast vectors interact, the MEL1 gene is activated allowing for the digestion of X-alpha-Gal which results in a blue colour. Therefore, the white colonies on the plate suggested that the proteins did not interact. The yeast strains individually transformed with the plasmids were mated and applied to plates with X-alpha-Gal (Figure 3.13). These colonies grew but also appeared white. This supports the theory that the proteins do not interact. However, the yeast strain co-transformed with the protease, and host proteins were found to grow on quadruple dropout media, with the addition of Aureobasidin A (Figure 3.14). These plates only support yeast strains with protein-protein interactions. To explain, resistance to media deficient in histidine and adenine with Aureobasidin A is supported by reporter genes. These genes are only switched on by the interaction of the GAL4 and AD domains from the PGBKT7 and PGADT7 plasmids. Therefore, yeast strains that grow on quadruple dropout media with Aureobasidin A have proteins, cloned into the vectors, which interact. This suggests that, despite the X-alpha-Gal results, the proteins cloned in the bait and prey vectors have interacted.
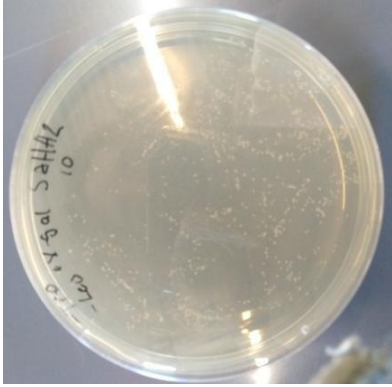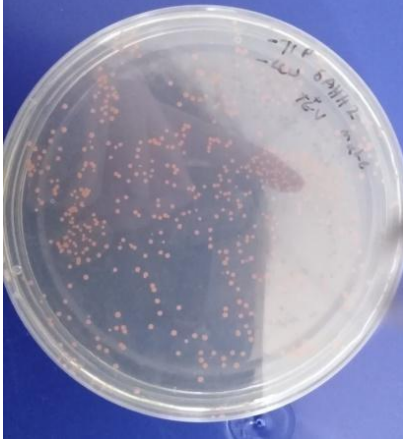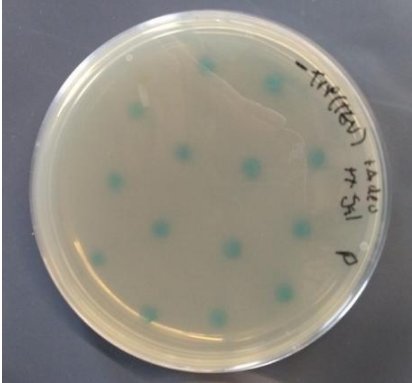
However, when the yeasts were transformed with only one of the plasmids and spread on agar containing X-alpha-Gal, all the colonies on the TEV NIa-pro plates were coloured blue (Figure 3.13). This suggests that the TEV protease was activating the reporter genes within the yeast.

42

This was further supported when the yeast containing TEV NIa-pro was spread on quadruple dropout media, containing X-alpha-Gal and Aureobasidin A, and blue colonies formed. The individual TEV NIa-pro protein appeared to activate all the reporter genes (Figure 3.14). Therefore, the co-transformation interaction cannot be confirmed due to the proteases ability to autoactivate genes. The plasmids containing the T-antigen were used as a negative control for protein-protein interactions.
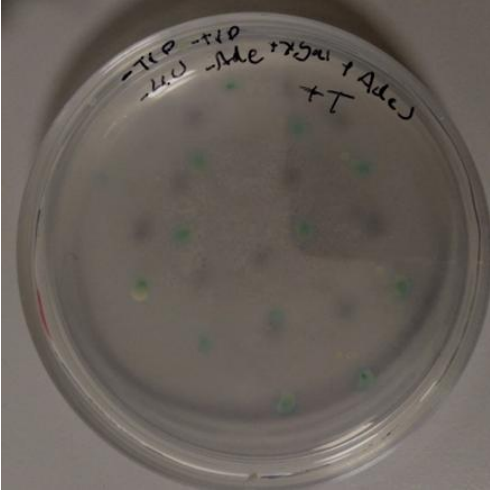
Determining the effective transformation of plasmids into the yeast			
Yeast Transformant	- Tryptophan plates	- Leucine plates	- Leucine - Tryptophanplates
TEV-NIa-pro $4x10^5$ cfu/µg		No Growth	No Growth
SAHH2 $1x10^5$ cfu/µg	No Growth		No Growth
pGBKT7-53(positive control) $4x10^6$ cfu/µg		No Growth	No Growth
pGADT7-T (positive control)	No Growth		No Growth

SAHH2 and TEV 1×10^5 cfu/µg			

Figure 3.12 Yeasts transformed with bait and prey plasmid spread out on SD minimal media. Y2Hgold yeast cells transformed with the PGBKT7 vector were spread on tryptophan deficient plants. Y187 yeasts cells transformed with the PGADT7 plasmid were spread out leucine deficient plates. Y2Hgold cells co-transformed with both plasmids were spread on tryptophan and leucine deficient media. The transformation efficiency of the yeasts were calculated.

Determining Protein-Protein interaction and auto activation via the addition of X-alpha-Gal	
Yeast Transformant	- Leucine - Tryptophan plates with X-alpha-Gal
AtSAHH2 and TEV NIa-pro (co-transfromed)	
AtSAHH2 and TEV NIa-pro (mated)	
TEV NIa-pro	

Figure 3.13 The co-transformed and mated yeast cells were spread on double drop out media containing no leucine and tryptophan. The media also contained X-alpha-Gal which turn yeast cells blue to indicate protein interactions. None of the yeast cells containing SAHH2 and TEV NIa-pro turned blue. Autoactivation test of the TEV NIa-pro in the PGBKT7 vector shows the yeast had turned blue. The TEV NIa-pro yeast had also grown on double dropout media.

Confirming Protein-Protein Interactions through quadruple knockout plates and fungicide Aureobasidin	
Yeast Transformant	Leucine plates - Tryptophan -Histidine -Adenine plates with Aureobasidin A and X-alpha-Gal
AtSAHH2 and TEV NIa-pro (co-transfromed)	
TEV NIa-pro	

Figure 3.14 Yeast cells from figure 3.13 were spread on quadruple dropout media containing no leucine, tryptophan, histidine and adenine. The media also contained Aureobasidin A. The figure shows that the co-transformed TEV NIA-pro grew colonies on this media. However, the individual TEV NIA-pro transformed yeast also grew colonies.

3.6 myc Tagged NIa-pro

Due to the inconsistencies with the yeast two-hybrid results, the TEV protease was modelled with the myc tag. This tag is attached to the protein when the sequence was introduced into the PGBKT7 plasmid. The myc tag is a short sequence of peptides which can significantly alter the structure or function of the protein (Morlacchi et al., 2012). After the tagged structure was equilibrated it was docked to AtSAHH2.

The results of Figure 3.15b shows that both the modified and un-modified NIa-pro bind to the same C-terminal region on the host protein (which matches the MD results). However, different regions of the NIa-pro associated with the host protein depending on whether the myc tag had been included. To explain, many of the residues that associate with the host protein on the unaltered NIa-pro were located on loops within the C-terminus (amino acids 150-250). In contrast, many of the residues on the myc tagged protein, which associate with SAHH2, were located on the N-terminus (amino acids 1-100). None of the residues in the modified protein, that associate with SAHH2, are from the myc tag. As seen from the alignment results, the myc tag altered the loop region between amino acids 150-245 (Figure 3.15a).

Figure 3.15a An equilibrated model of TEV NIa-pro (blue) containing the myc tagand additional residues that would be included during the proteins expression (red).This model was aligned to the equilibrated NeYSV NIa-pro (yellow) (RMSD scoreof 2.592 Å). Several regions of the protein are in alternative places due to the addition of the myc tag. The most apparent being the loop region on the C-terminus (200-240) which has been outlined by a red circle.
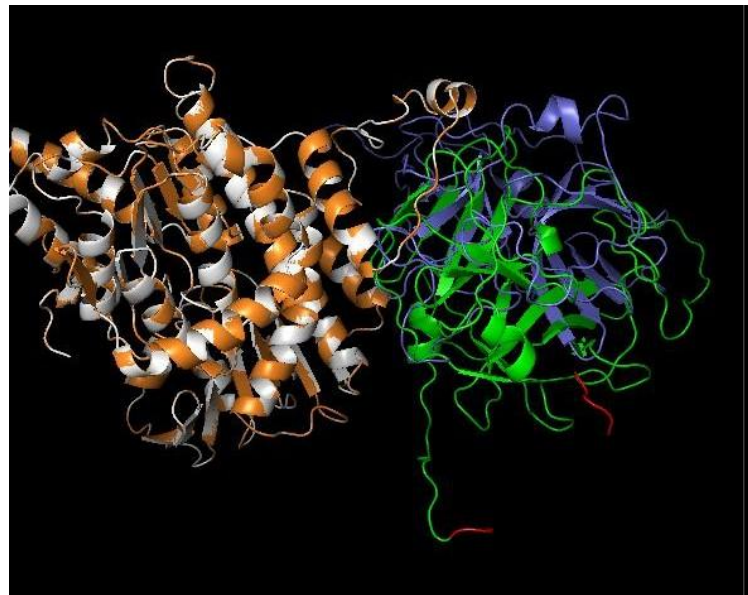
Figure 3.15b Image displays the equilibrated model of TEV NIa-pro (purple) docked to *Arabidopsis* SAHH2 (white).This structure has been aligned to the equilibrated TEV NIa-pro (green) which has been modified to include the myc tag (red). This tagged structure has been dockedto the AtSAHH2 (orange). The RMSD of the alignment is 0.001 Å.

3.7 Identifying potential other interacting proteins

Both TEV and NeYSV are reported to infect *N.benthamiana* (Collum & Culver, 2017 & Beddoe et al., 2020). The proteins from the Martínez et al (2016) paper identified as interacting with TEV NIa-pro were BLASTp searched against *N.benthamiana*. The results were then compiled in a table. These proteins may also interacts with TEV NIa-pro in *N.benthamiana* and therefore, could be responsible for the host range of the virus. A protein of interest is the DEAD-box ATP-dependentRNA helicase 38 which has a sequence similarity to the *N.benthamiana* ATP-dependent RNA helicase EIF4a. Chen et al (2008) reported that NIa-pro can interact with eukaryotic initiation factors and disturb nuclear transport affecting virulence. Gao et al (2014) also predicted that the interaction of NIa-pro with a eukaryotic initiation factor would be responsible for the host range of the virus. However, this match is only weak (35%) when compared to the predicted proteins from the *N.benthamiana* Genome v1.0.1, the protein matches an ATP-dependent RNA helicase (63%) (Fernandez-Pozo et al., 2014). But, eukaryotic initiation factors eIF-4A are part of the DEAD-box ATP-dependentRNA helicase family. Therefore, the DEAD-box ATP-dependentRNA helicase, which TEV NIa-pro interacts with, may also serve as a eukaryotic initiation factor in the plant. This protein could prove a strong addition to future related studies.

Arabidopsis Thaliana	NCBI ID	Nicotiana benthamiana	NCBI ID
Adenosylhomocysteinase 2	AT3G23810	S-adenosylhomocysteine hydrolase	BAR72292.1
ATP synthase CF1 alpha subunit	ATCG00120	ATP synthase F1 subunit 1	QIS68115.1
Carbonic Anhydrase	AT3G01500	plastid carbonic anhydrase	ASW15992.1
Catalase 3	AT1G20620	catalase	ACL27272.1
DEAD-box ATP-dependent RNA helicase 38	AT3G53110	ATP-dependent RNA helicase EIF4a	ANG84008.1
Delta(24)-sterol reductase	AT3G19820	sterol side chain reductase	BBE00760.1
Glyceraldehyde 3-phosphate dehydrogenase	AT1G42970	cytosolic glyceraldehyde-3-phosphate dehydrogenase	AKH15660.1
Heat shock protein 70	AT5G02490	HSP70	ARH56399.1
Oxygen-evolving enhancer protein 1-2	AT3G50820	chloroplast PsbO1 precursor	AFH57997.1
Peroxisomal (S)-2-hydroxy-acid oxidase GLO2	AT3G14415	glycolate oxidase	5ZBM_A
Phosphoglycerate kinase	AT1G79550	phosphoglycerate kinase	ADR71054.1
Photosystem I	AT1G31330	chloroplast photosystem I	UBK24965.1
ATP synthase	ATMG01190	ATP synthase F1	QIS68115.1
Tubulin	AT1G50010	alpha-tubulin	ALH22047.1

Table 3.2 BLASTp searches of proteins identified in the Martínez et al (2016) study to interact with TEV NIa-pro. Proteins excluded from this table either did not have a complete matching sequence identified in *Nicotiana benthamiana* .

4.Discussion

The results outlined in section 3 are discussed. 'Peptide cleavage results 4.1' examines the molecular dynamic results of the peptide cleavage sequence bound to the protease. The section also discusses if the mechanism in which the protease interacts with the peptide sequence is used to interact with host proteins. 'Molecular dynamic protein-protein interaction 4.2' discusses the molecular dynamics results between the viral proteases and plant proteins. Then the next section 'Yeast two-hybrid results 4.3' follows this up with a discussion of the yeast two-hybrid results. The results of the experiments as it pertains to the hypothesis are then discussed in section 'Host range 4.4'. Finally, the purpose of the experiment and results of the experiment are discussed in 'Further research/conclusion 4.5'. This section also discusses gaps in the experiment that could be filled in by future research.

4.1 Peptide Cleavage

The predicted optimal cleavage sequence, of the NeYSV protease, was determined to match the TuMV optimal cleavage sequence (Table 3.1). Although most potyviral proteases cleave using specific sequences, there were some variations within the NeYSV recognition sequences that were unusual. The NIb C-terminal cleavage site contained no histadine at position two. Instead, a larger phenalalanine was in place of the histidine. The 6K1 N-terminal cleavage site also had no valine at position 4. The NeYSV protease contained a valine, at position 169, in the binding pocket of the catalytic site. In previous studies, this residue increased the size of the catalytic site in comparison to the TEV protease. This may explain the variations found at position 4 of the predicted cleavage site (Adams, Antoniw and Beaudoin, 2005). This led to the conclusion that the NeYSV protease displays a wider range of peptide sequence recognition, in comparison to the TEV NIa-pro cleavage sequences. The interaction, between both NIa-pro's and SAHH, did not involve NIa-pro binding to residues in the host protein matching cleavage sequences in the polyprotein (Figure 3.11 and Figure 3.12).

The ability of NIa-pro to target host proteins based on the recognition of the cleavage sequence may be revealed in future research. The AtGSTF2 peptide sequence contains no motifs matching any of the NIa-pros cleavage sites. Docking between the crystal structure of GSTF2 and TEV NIa-pro had the protease bind to exposed residues of GSTF2. Theresidues from GSTF2 bound to NIa-pro partially matched the TEV 6K1 C-terminal cleavage site. This interaction blocked the active site of the protease.

52
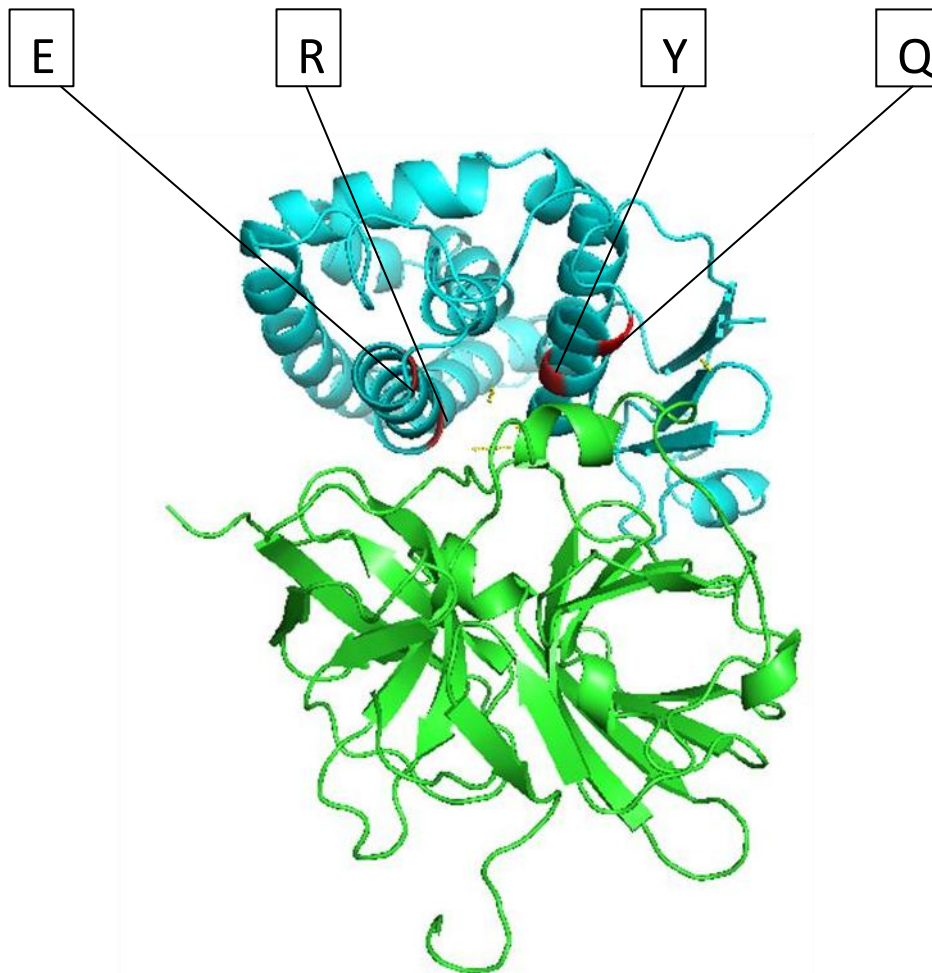
Figure 4.1 Crystal structure of TEV NIa-pro docked to AtGSTFT2. The residues TEV NIa-pro hydrogen bonds to are shown as red. The amino acid FASTA representation of these bound residues are displayed at the top of the image. This sequence partially matched the TEV 6K1 N-terminal cleavage region (EREIIYTQS).

4.2 Molecular Dynamics Protein-Protein Interactions

The interaction between NeYSV NIa-pro and ATSAHH2 determined that residues from the protease bound to part of the insertion region of AtSAHH2 (Figure 3.11) . The insertion region is an exposed surface of the outside of the protein and is responsible for interactions between SAHH2 and other proteins (Lee et al., 2012). Lee et al (2012) argued that this region may serve as an interaction domain for associations with proteins that can direct SAHH to different cellular compartments. Gao et al (2012b) determined that NIa-pro from PRSV is responsible for altering the nuclear transport of Methionine Sulfoxide Reductase B1 (PaMSRB1) via interactions with the C-terminal residues. Arguably, NeYSV NIa-pro may alter protein interactions and prevent nuclear transport by binding to the insertion region of SAHH2. Also this insertion region binds to the mRNA cap (Lee et al., 2012). The VPg protein from PVY was also found to bind to regions of proteins that associate with the mRNA cap (Oliveira et al., 2019). However, analysis of this interaction showed a low affinity (-3.13 ΔG, kal/mol). In fact,

53

the prediction software stated that most probably, 'the structures do not form a complex in solution'. This is also supported by the RMSD models in which the intersurface regions displayed a high degree of movement. This suggests an unstable interaction occurred.

In contrast to NeYSV's NIa-pro, TEV NIa-pro was found to strongly associate with the host protein (-17.2 ΔG, kal/mol). The residues serine 135 and argenine 105 from the protease were found to associate with AtSAHH2. Both the RMSD and affinity results suggest that the protein complex is stable and may exist outside a simulation. Arguably, none of the residues identified through hydrogen bond analysis (from either the protease or AtSAHH2) play any role in the protein other than structural (Lee et al., 2012). However, the association of the protease with the AtSAHH2 residues appeared to orient the protease to obscure the C-terminal of SAHH2. The tail region of SAHH2 interlocked with other SAHH2 proteins forming a protein complex responsible for the breakdown of S-adenosylhomocysteine (Wang et al., 2014). Therefore, by indirectly associating with the tail of AtSAHH2, TEV's NIa-pro may prevent diamerization, perturbing the protein's function. The potential inhibition of protein diamerization formation by binding to oligomerization sites is also seen in the docking of the TEV and NeYSV's NIa-pro to the AtRICE1 crystal structure (Figure 4.2). Docking determined that both structures bound to the oligomerization sites (Zhang et al., 2017).

In contrast, the MD simulations between the proteases and NbSAHH suggests that the proteases do not interact with NbSAHH. The similarity between NbSAHH and AtSAHH2 sequences is 96%. Due to these minor sequence alterations, the structure of SAHH has been altered to an extent in which the TEV NIa-pro is unlikely to bind to the protein. Alignments, between AtSAHH2 and NbSAHH, showed most of the variation in structure stems from the C-terminal tail (Figure 3.9c). The tomato chlorosis virus CP and PVY HC-pro has been found to bind to and downregulate the function of NbSAHH (Cañizares et al., 2013). Possibly, the binding to NbSAHH is fulfilled by the same viral protein in TEV. However, it's feasible that the proteases interact differently with NbSAHH and bind to the complex formed by NbSAHH proteins joining together. Arguably, it isnot possible to determine the host range of viruses through interacts between NIa-pro and SAHH. Proteins responsible for host range may bind to the NeYSV residue isoleucine 28. As this residue was identified as potentially being responsible for host range by alignments with PRSV NIa-pro.
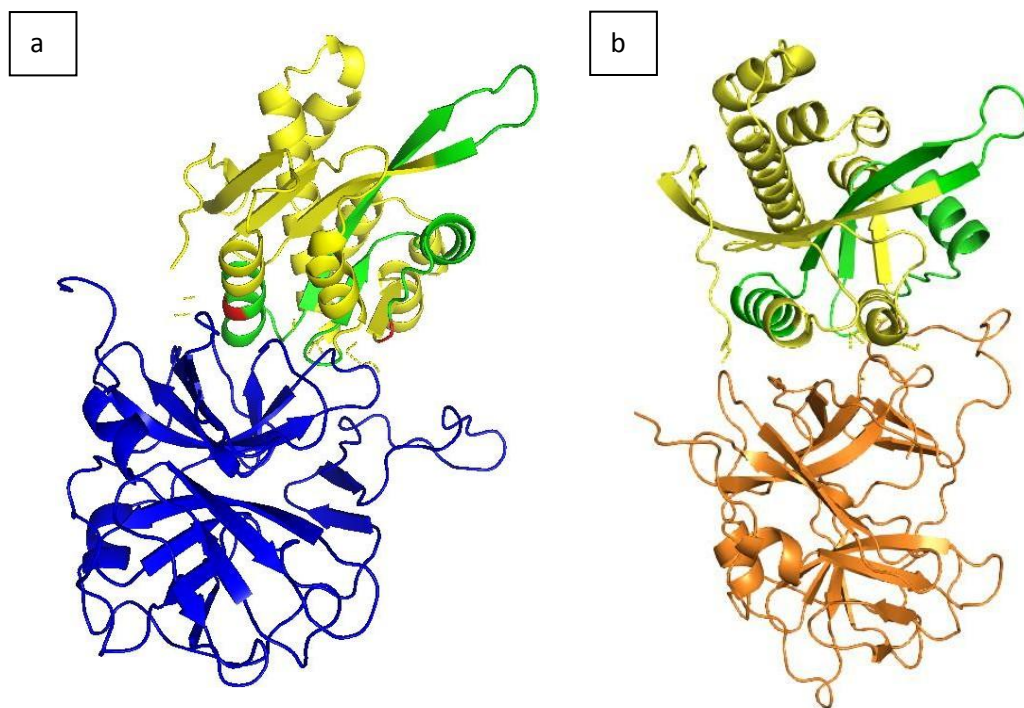
Figure 4.2 The crystal structure of RICE1 (yellow) docked to the TEV (blue) and NeYSV (orange) protease. The oligomerization sites are displayed green. Image a (left) shows the TEVNIa-pro binding to residues in the oligimirisation site 1 (red) of RICE1. Image b (right) shows NeYSV has formed hydrogen bonds with oligimirisation sites 1 and 2 in RICE1. The regions hydrogen bonded by the protease are highlighted in red.

4.3 Yeast two-hybrid assay

Despite the MD results, the initial yeast-two hybrid results suggest that the TEV NIa-pro and AtSAHH2 do not interact(Figure 3.8a; Figure 3.13). However, the autoactivation tests on the TEV NIa-pro showed definitive signs of autoactivation (Figure3.14) . In fact, the individually transformed yeast with NIA-pro grew on quadruple knockout medium with Aureobasidin A. This indicates that TEV NIa-pro can activate all the reporter genes on its own. Aureobasidin A is considered, by the Takara Bio Matchmaker® Gold Yeast Two-Hybrid System User Manual, to be a more accurate predictor of protein-protein interactions in contrast to the use of X-alpha-Gal used (Yeastmaker™ Yeast Transformation System 2 User Manual, 2022). The individually transformed yeast grows on quadruple knockout media with Aureobasidin A. The protein interactions of the co-transformed yeast were not confirmed because the individually transformed NIa-pro activates all the reporter genes. Potentially, the TEV protein binds to transcription factors in the yeast. The interaction with transcription factors can alter the results of a yeast two-hybrid assay according to the Takara Bio Matchmaker® Gold Yeast Two-Hybrid System User Manual used (Yeastmaker™ Yeast Transformation System 2 User Manual, 2022). The Schaad, Anderberg and Carrington (2000) study determined that, through yeast two-hybrid

analysis, the TEV NIA interacts with AtEIF4E. However, this was both the VPg and NIa-pro domains. Gao et al (2014) also determined that the papaya NIa-pro can bind to EIF3G through the C-terminal. Possibly, through the same C-terminal binding, the TEV protease binds to the yeast transcription factor. The Martínez et al (2016) study also determined that TEV NIa-pro interacts with EIF2 from *Arabidopsis*. To navigate the autoactivation, sequences encoding the interacting proteins from the bait and prey plasmids may be swapped. Due to time restrains this was not possible.

Due to the results gained through the yeast two-hybrid assay, the use of further MD analysis was deployed. Attaching the myc tag to NeYSV NIa-pro appeared to alter the binding of the protein to AtSAHH2. This study and Gao et al (2012b), determined that the C-terminus of NIa-pro was responsible for interactions with host proteins. However, attaching a myc tag appears to alter the C-terminal loop region and favour the N-terminal of NIa-pro for binding to host proteins. Therefore, attaching the myc tag in the process of a yeast two-hybrid experiment may indirectly affect protein-protein interactions. This may explain why the cells, co-transformed with TEV NIa-pro and SAHH2, appeared to display no signs of protein-protein interaction on X-alpha-Gal treated media. NIa-pro occurs in three forms within the plant: un-restricted, restricted and truncated (through auto-catalyzation) (Carrington 2000). When the protease is unrestricted, the VPg domain is bound to the NIa-pro domain on the C-terminus via a weak catalytic site (Martínez et al 2016). This protein complex is crucial for the infection of the virus as it binds to host proteins (Carrington 2000). Arguably, the addition of the myc tag mimics the binding of the NIa-pro to VPg altering the C-terminal structure.

4.4 Host Range

The affect of TEV in *Arabidopsis* has been well studied, there is a large amount of information on the plant proteins associated with virulence. NeYSV, reportedly, will infect *N. benthamiana* and not *Arabidopsis*. TEV will infect both *N. benthamiana* and *Arabidopsis* (Collum & Culver, 2017; Beddoe et al., 2020). The interaction between viral and host proteins determines the host range of the virus (Chen et al., 2008). Studies have argued that the interaction between NIa-pro and host proteins is a determinant of the host range of potyviruses (Chen et al., 2008). As TEV and NeYSV share a common host, any protein that TEV interacts with in both *N. benthamiana* and *Arabidopsis thaliana* is important for host range. Therefore, if this protein was to interact with NeYSV in both *N. benthamiana* and *Arabidopsi*s then *Arabidopsis* would potentially be a host for NeYSV.

This study determined through molecular dynamics that NIa-pro from NeYSV did not interact with SAHH from *N. benthamiana* or *Arabidopsis* suggesting that NeYSV does not infect either of these plants. However, while the TEV NIa-pro interacts with SAHH2 in *Arabidopsis* it does

not interact with SAHH in *N. benthamiana*. This suggests that the interaction between SAHH and NIa-pro is not a determinant of the host range of the virus. This study attempted to validate the MD interaction data through a yeast two hybrid assay. However, auto-activation prevented any accurate data from being obtained. This is potentially though NIa-pro interacting with a transcription factor in the yeast. However, further molecular dynamics results suggested that the interaction between the viral and host proteins could be altered by the addition of the his and myc tag further suggesting that a yeast two-hybrid analysis is not suitable to determine the interaction between NIa-pro and plant proteins. The Martínez et al (2016) study used a Twin-Strep-tag and determined that the location of this tag did not affect the viral infectivity. Therefore, the polypeptide tag would not have affected the binding of NIa-pro to host proteins. Gao et al., (2014) used a Sos recruitment yeast two-hybrid system to determine the interactions between PRSV and host proteins. The plasmid used in this system does not produce a his or myc tags attached to the protein. This may prevent the protein binding from being altered. The Sos recruitment yeast two-hybrid system also is not based on a transcriptional readout (Aronheim et al., 1997). Therefore, this method is suitable for identifying protein-protein interactions with proteins that can also bind to transcription factors. However, it is worth noting that the (Lin et al., 2009) study found no autoactivation of both the Pinellia and Shallot potyviral NIa-pro's using the Matchmaker Gold Yeast Two-Hybrid System. This system is the same one used in this study. Based on the molecular dynamics and yeast two hybrid results, in this study, the data from Lin et al (2009) data may be unreliable.

4.5 Further Research/Conclusion

This paper has outlined a novel way to determine the host range of potyviruses. The success of this technique could allow for the computational analysis, between plant proteins and a protease, to determine the host range of a virus. Therefore, saving time and equipment in the lab. This would also allow for a rapid testing of a variety of plants to determine a host range without having to import or grow the plants. However, this study was not able to draw any solid conclusions in terms of how the interaction between NIa-pro and plant-proteins affects host range. But, from the table outlined in Section 3.2 there are a wide variety of proteins that TEV NIa-pro interacts with that already has full and validated sequence data in *N. benthamiana*. In order to expand on the work performed in this study, these proteins could be swapped instead of SAHH2. Future studies could also use the Sos two hybrid system. To prevent auto-activation and polypeptide protein tags affecting binding. However, a pull-down assay could be deployed rather than a yeast two-hybrid assay. This would allow the entire proteome from the plant to interact with NIa-pro. Therefore, any protein which interacts with TEV NIa-pro in *N. benthamiana* and *Arabidopsis thaliana,* could be responsible for the host range of the virus. This data could be supported by a pull-down assay in which NeYSV NIa-pro interacts with the *N. benthamiana* and *Arabidopsis thaliana,* proteasome. If the protein which determined the host

range was found to interact with NeYSV NIa-pro in *N. benthamiana* but not in *Arabidopsis*. We could determine that *Arabidopsis* is not part of NeYSV host range. This information could be followed up using molecular dynamics to determine the effect the protease has on the host protein and how this may be detrimental to the plant's health. This could then be followed up further by generating mutant lines using virus-induced gene silencing to knock down the expression of host genes of interest. These knock down lines could then be infected and if no symptoms occur, then we could argue that the protein is responsible for the host range of the virus. Additionally, the cleavage sequence of the NeYSV protease has yet to be experimentally confirmed. This could be performed using the protocol laid out by Zheng et al (2008). Finally, the reason behind TEV NIa-pros targeting of SAHH2 was not experimentally proven, while the molecular dynamics results suggest that this interaction may prevent dimerization, no experiment has discovered TEV NIa-pro has this ability.

5. Bibliography

Adams, M.J., Antoniw, J.F. and Beaudoin, F. (2005). Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology*, [online] 6(4), pp.471–487. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1364-3703.2005.00296.x.

Ala-Poikela, M., Goytia, E., Haikonen, T., Rajamaki, M.-L. . and Valkonen, J.P.T. (2011). Helper Component Proteinase of the Genus Potyvirus Is an Interaction Partner of Translation Initiation Factors eIF(iso)4E and eIF4E and Contains a 4E Binding Motif. *Journal of Virology*, [online] 85(13), pp.6784–6794. Available at: https://jvi.asm.org/content/jvi/85/13/6784.full.pdf [Accessed 28 Feb. 2021].

Álvarez-Aragón, R., Haro, R., Benito, B. and Rodríguez-Navarro, A. (2016). Salt intolerance in Arabidopsis: shoot and root sodium toxicity, and inhibition by sodium-plus-potassium overaccumulation. *Planta*, [online] 243(1), pp.97–114. Available at: https://pubmed.ncbi.nlm.nih.gov/26345991/ [Accessed 27 Jan. 2022].

Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J. and Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Molecular and Cellular Biology*, [online] 17(6), pp.3094–3102. doi:10.1128/mcb.17.6.3094.

Beauchemin, C., Boutet, N. and Laliberté, J.-F. (2007). Visualization of the interaction between the precursors of VPg, the viral protein linked to the genome of turnip mosaic virus, and the translation eukaryotic initiation factor iso 4E in Planta. *Journal of Virology*, [online] 81(2), pp.775–782. Available at: https://pubmed.ncbi.nlm.nih.gov/17079311/ [Accessed 2 Mar. 2021].

Beddoe, N., Adams, I.P., McGreig, S. and Forsyth, A. (2020). Identification and full genomic sequence of nerine yellow stripe virus. *Archives of Virology*, [online] 165(12), pp.2967–2971. Available at: https://link.springer.com/article/10.1007/s00705-020-04776-3 [Accessed 9 Nov. 2020].

Bedhomme, S. and Elena, S.F. (2011). Virus Infection Suppresses Nicotiana benthamiana Adaptive Phenotypic Plasticity. *PLoS ONE*, [online] 6(2), p.e17275. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017275 [Accessed 2 May 2020].

Boratyn, G.M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. and Madden, T.L. (2019). Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*, [online] 20(1). Available at: https://pubmed.ncbi.nlm.nih.gov/31345161/ [Accessed 12 Sep. 2019].

Cañizares, M.C., Lozano-Durán, R., Canto, T., Bejarano, E.R., Bisaro, D.M., Navas-Castillo, J. and Moriones, E. (2013). Effects of the crinivirus coat protein-interacting plant protein SAHH on post-transcriptional RNA silencing and its suppression. *Molecular plant-microbe interactions: MPMI*, [online] 26(9), pp.1004–1015. Available at: https://pubmed.ncbi.nlm.nih.gov/23697374/ [Accessed 4 Mar. 2021].

Carrington, J.C., Haldeman, R., Dolja, V.V. and Restrepo-Hartwig, M.A. (1993). Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo. *Journal of Virology*, [online] 67(12), pp.6995–7000. Available at: https://pubmed.ncbi.nlm.nih.gov/8230423/ [Accessed 15 Dec. 2020].

Chen, I.-H., Tsai, A.Y., Huang, Y.-P., Wu, I.-F., Cheng, S.-F., Hsu, Y.-H. and Tsai, C.-H. (2017). Nuclear-Encoded Plastidal Carbonic Anhydrase Is Involved in Replication of Bamboo mosaic virus RNA in Nicotiana benthamiana. *Frontiers in Microbiology*, [online] 8. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5651272/ [Accessed 2 Mar. 2021].

Chen, K.-C., Chiang, C.-H., Raja, J.A.J., Liu, F.-L., Tai, C.-H. and Yeh, S.-D. (2008). A Single Amino Acid of NIaPro of Papaya ringspot virus Determines Host Specificity for Infection of Papaya. *Molecular Plant-Microbe Interactions®*, [online] 21(8), pp.1046–1057. Available at: https://apsjournals.apsnet.org/doi/pdf/10.1094/MPMI-21-8-1046 [Accessed 6 Aug. 2020].

Collum, T.D. and Culver, J.N. (2017) "Tobacco mosaic virus infection disproportionately impacts phloem associated translatomes in Arabidopsis thaliana and Nicotiana benthamiana," *Virology*, 510, pp. 76–89. Available at: https://doi.org/10.1016/j.virol.2017.07.002 [Accessed 6 Aug. 2020].

Croll, T.I., Sammito, M.D., Kryshtafovych, A. and Read, R.J. (2019). Evaluation of template-based modeling in CASP13. *Proteins: Structure, Function, and Bioinformatics*, 87(12), pp.1113–1127. doi:10.1002/prot.25800

Cui, H. and Wang, A. (2016). Plum Pox Virus 6K1 Protein Is Required for Viral Replication and Targets the Viral Replication Complex at the Early Stage of Infection. *Journal of Virology*, [online] 90(10), pp.5119–5131. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4859702/ [Accessed 28 Feb. 2022].

De, S., Chavez-Calvillo, G., Wahlsten, M. and Mäkinen, K. (2018). Disruption of the methionine cycle and reduced cellular gluthathione levels underlie potex–potyvirus synergism in Nicotiana benthamiana. *Molecular Plant Pathology*, [online] 19(8), pp.1820–1835. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6638099/ [Accessed 4 Mar. 2021].

Desta, I.T., Porter, K.A., Xia, B., Kozakov, D. and Vajda, S. (2020). Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure*, [online] 28(9), pp.1071-1081.e3. Available at: https://www.cell.com/structure/fulltext/S0969-2126(20)30209-4?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0969212620302094%3Fshowall%3Dtrue [Accessed 8 Feb. 2022].

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A. and Mueller, L.A. (2014). The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Research*, [online] 43(D1), pp.D1036–D1041. doi:10.1093/nar/gku1195.

Fiser, A., Do, R.K.G. and Šali, A. (2000). Modeling of loops in protein structures. *Protein Science*, [online] 9(9), pp.1753–1773. Available at: https://pubmed.ncbi.nlm.nih.gov/11045621/ [Accessed 8 May 2021].

Gao, L., Shen, W., Yan, P., Tuo, D., Li, X. and Zhou, P. (2012a). NIa-pro of Papaya ringspot virus interacts with papaya methionine sulfoxide reductase B1. *Virology*, [online] 434(1), pp.78–87. doi:10.1016/j.virol.2012.09.007.

Gao, L., Shen, W.T., Yan, P., Tuo, D.C., Li, X.Y. and Zhou, P. (2012b). A set of host proteins interacting with papaya ringspot virus NIa-Pro protein identified in a yeast two-hybrid system. *Acta Virologica*, [online] 56(1), pp.25–30. doi:10.4149/av_2012_01_25.

Gao, L., Tuo, D., Shen, W., Yan, P., Li, X. and Zhou, P. (2014). NIa-Pro of Papaya ringspot virus interacts with Carica papaya eukaryotic translation initiation factor 3 subunit G (CpeIF3G). *Virus Genes*, [online] 50(1), pp.97–103. doi:10.1007/s11262-014-1145-x

Gibson, R.W., Aritua, V., Byamukama, E., Mpembe, I. and Kayongo, J. (2004). Control strategies for sweet potato virus disease in Africa. *Virus Research*, [online] 100(1), pp.115–122. Available at: https://pubmed.ncbi.nlm.nih.gov/15036842/ [Accessed 27 Nov. 2019].

Gong, Y.-N., Tang, R.-Q., Zhang, Y., Peng, J., Xian, O., Zhang, Z.-H., Zhang, S.-B., Zhang, D.-Y., Liu, H., Luo, X.-W. and Liu, Y. (2020). The NIa-Protease Protein Encoded by the Pepper Mottle Virus Is a Pathogenicity Determinant and Releases DNA Methylation of Nicotiana benthamiana. *Frontiers in Microbiology*, 11. doi:10.3389/fmicb.2020.00102.

Goodfellow, I. (2011). The genome-linked protein VPg of vertebrate viruses — a multifaceted protein. *Current Opinion in Virology*, [online] 1(5), pp.355–362. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3541522/ [Accessed 19 Jan. 2020].

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M.S., Bright, J.M., Beyene, Y. and Prasanna, B.M. (2015). Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. *Theoretical and Applied Genetics*, [online] 128(10), pp.1957–1968. Available at: https://link.springer.com/article/10.1007/s00122-015-2559-0 [Accessed 27 Feb. 2021].

Grzela, R., Strokovska, L., Andrieu, J.-P. ., Dublet, B., Zagorski, W. and Chroboczek, J. (2006). Potyvirus terminal protein VPg, effector of host eukaryotic initiation factor eIF4E. *Biochimie*, [online] 88(7), pp.887–896. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0300908406000186 [Accessed 12 Oct. 2019].

Gullner, G., Komives, T., Király, L. and Schröder, P. (2018). Glutathione S-Transferase Enzymes in Plant-Pathogen Interactions. *Frontiers in Plant Science*, [online] 9. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2018.01836/full [Accessed 11 Aug. 2019]Guo, D., Rajamäki, M.-L., Saarma, M. and Valkonen, J.P.T. (2001). Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system. *Journal of General Virology*, [online] 82(4), pp.935–939. Availableat: https://sci-hub.yncjkj.com/10.1099/0022-1317-82-4-935 [Accessed 19 May 2021].

Han, H.-E., Sellamuthu, S., Shin, B.H., Lee, Y.J., Song, S., Seo, J.-S., Baek, I.-S., Bae, J., Kim, H., Yoo, Y.J., Jung, Y.-K., Song, W.K., Han, P.-L. and Park, W.J. (2010). The nuclear inclusion a (NIa) protease of turnip mosaic virus (TuMV) cleaves amyloid-β. *PloS One*, [online] 5(12), p.e15645. Available at: https://pubmed.ncbi.nlm.nih.gov/21187975/.

Hernández, J.A., Gullner, G., Clemente-Moreno, M.J., Künstler, A., Juhász, C., Díaz-Vivancos, P. and Király, L. (2016). Oxidative stress and antioxidative responses in plant–virus interactions. *Physiological and Molecular Plant Pathology*, [online] 94, pp.134–148. Available at: https://digital.csic.es/bitstream/10261/152207/1/oxidative_stress_Hernandez.pdf [Accessed

27 Nov. 2020].

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H. and MacKerell, A.D. (2016). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, [online] 14(1), pp.71–73. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5199616/ [Accessed 2 Feb. 2021].

Humphrey, W., Dalke, A. and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), pp.33–38. doi:10.1016/0263-7855(96)00018-5.

Jordan, R., Guaragna, M.A. and Putnam, M. (2011). DETECTION AND MOLECULAR CHARACTERIZATION OF NEW AND EMERGING POTYVIRUSES OF ORNAMENTAL PLANTS. *Acta Horticulturae*, [online] (901), pp.159–166. Available at: https://www.researchgate.net/publication/284809865_Detection_and_molecular_characterizatio n_of_new_and_emerging_potyviruses_of_ornamental_plants [Accessed 27 Feb. 2021].

Karami, Y., Guyon, F., De Vries, S. and Tufféry, P. (2018). DaReUS-Loop: accurate loop modeling using fragments from remote or unrelated proteins. *Scientific Reports*, [online] 8(1), p.13673. Available at: https://pubmed.ncbi.nlm.nih.gov/30209260/ [Accessed 9 Nov. 2021].

Khan, M.A. and Goss, D.J. (2012). Poly(A)-Binding Protein Increases the Binding Affinity and Kinetic Rates of Interaction of Viral Protein Linked to Genome with Translation Initiation Factors eIFiso4F and eIFiso4F·4B Complex. *Biochemistry*, [online] 51(7), pp.1388–1395. Available at: https://pubmed.ncbi.nlm.nih.gov/22299678/ [Accessed 15 Dec. 2020].

Knipe, D.M., Raja, P. and Lee, J. (2017). Viral Gene Products Actively Promote Latent Infection by Epigenetic Silencing Mechanisms. *Current opinion in virology*, [online] 23, pp.68–74. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5475406/ [Accessed 3 Mar. 2021].

Krissinel, E. and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, [online] 60(12), pp.2256–2268. Available at: https://scripts.iucr.org/cgi-bin/paper?ba5056 [Accessed 11 Jun. 2020].

Krissinel, E. and Henrick, K. (2005). *Detection of Protein Assemblies in Crystals*. [online] Springer Link. Available at: https://link.springer.com/chapter/10.1007%2F11560500_15 [Accessed 26 Jan. 2022].

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, [online] 33(7),

pp.1870–1874. Available at: https://pubmed.ncbi.nlm.nih.gov/27004904/ [Accessed 14 Apr. 2019].

Lee, S., Doxey, A.C., McConkey, B.J. and Moffatt, B.A. (2012). Nuclear Targeting of Methyl-Recycling Enzymes in Arabidopsis thaliana Is Mediated by Specific Protein Interactions. *Molecular Plant*, [online] 5(1), pp.231–248. Available at: https://www.sciencedirect.com/science/article/pii/S1674205214606602 [Accessed 1 Feb. 2022]

Léonard, S., Viel, C., Beauchemin, C., Daigneault, N., Fortin, M.G. and Laliberté, J.-F. (2004). Interaction of VPg-Pro of turnip mosaic virus with the translation initiation factor 4E and the poly(A)-binding protein in planta. *The Journal of General Virology*, [online] 85(Pt 4), pp.1055–1063. Available at: https://pubmed.ncbi.nlm.nih.gov/15039548/ [Accessed 6 Mar. 2022].

Li, M., Simonetti, F.L., Goncearenco, A. and Panchenko, A.R. (2016). MutaBind estimates and interprets the effects of sequence variants on protein–protein interactions. *Nucleic Acids Research*, [online] 44(Web Server issue), pp.W494–W501. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987923/ [Accessed 14 Dec. 2020].

Lin, L., Shi, Y., Luo, Z., Lu, Y., Zheng, H., Yan, F., Chen, J., Chen, J., Adams, M.J. and Wu, Y. (2009). Protein-protein interactions in two potyviruses using the yeast two-hybrid system. *Virus Research*, [online] 142(1-2), pp.36–40. doi:10.1016/j.virusres.2009.01.006.

Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., Zheng, C. and Wu, C. (2012). Identification and characterization of fructose 1,6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses. *Gene*, [online] 503(1), pp.65–74. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0378111912004684?via%3Dihub [Accessed 15 Dec. 2020].

Mäkinen, K. and Hafrén, A. (2014). Intracellular coordination of potyviral RNA functions in infection. *Frontiers in Plant Science*, [online] 5. Available at: https://www.frontiersin.org/articles/10.3389/fpls.2014.00110/full [Accessed 23 Jul. 2019].

Martínez, F., Rodrigo, G., Aragonés, V., Ruiz, M., Lodewijk, I., Fernández, U., Elena, S.F. and Daròs, J.-A. (2016). Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. *BMC Genomics*, [online] 17(1). doi:10.1186/s12864-016-2394-y

Martinière, A., Gibrat, R., Sentenac, H., Dumont, X., Gaillard, I. and Paris, N. (2018). Uncovering pH at both sides of the root plasma membrane interface using noninvasive imaging.

Proceedings of the National Academy of Sciences, [online] 115(25), pp.6488–6493. Available at: https://www.pnas.org/content/115/25/6488 [Accessed 26 Jan. 2022].

Mathur, C. *et al.* (2012) "Functional regulation of PVBV nuclear inclusion protein-a protease activity upon interaction with viral protein genome-linked and phosphorylation," *Virology*, 422(2), pp. 254–264. Available at: https://doi.org/10.1016/j.virol.2011.10.009.

Michaletz, S.T., Weiser, M.D., Zhou, J., Kaspari, M., Helliker, B.R. and Enquist, B.J. (2015). Plant Thermoregulation: Energetics, Trait–Environment Interactions, and Carbon Economics. *Trends in Ecology & Evolution*, [online] 30(12), pp.714–724. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0169534715002396 [Accessed 26 Jan. 2022].

Morlacchi, S., Sciandra, F., Bigotti, M.G., Bozzi, M., Hübner, W., Galtieri, A., Giardina, B. and Brancaccio, A. (2012). Insertion of a myc-tag within α-dystroglycan domains improves its biochemical and microscopic detection. *BMC Biochemistry*, [online] 13(1). Available at: https://bmcbiochem.biomedcentral.com/articles/10.1186/1471-2091-13-14 [Accessed 6 Mar. 2022].

Nishikiori, M., Mori, M., Dohi, K., Okamura, H., Katoh, E., Naito, S., Meshi, T. and Ishikawa, M. (2011). A Host Small GTP-binding Protein ARL8 Plays Crucial Roles in Tobamovirus RNA Replication. *PLoS Pathogens*, 7(12), p.e1002409.

Noskov, S.Yu. and Lim, C. (2001). Free Energy Decomposition of Protein-Protein Interactions. *Biophysical Journal*, [online] 81(2), pp.737–750. Available at: https://www.sciencedirect.com/science/article/pii/S0006349501757384 [Accessed 23 Feb. 2022].

Nunn, C.M., Jeeves, M., Cliff, M.J., Urquhart, G.T., George, R.R., Chao, L.H., Tscuchia, Y. and Djordjevic, S. (2005). Crystal Structure of Tobacco Etch Virus Protease Shows the Protein C Terminus Bound within the Active Site. *Journal of Molecular Biology*, [online] 350(1), pp.145–155. Available at: https://www.researchgate.net/publication/7822871_Crystal_Structure_of_Tobacco_Etch_Virus_Protease_Shows_the_Protein_C_Terminus_Bound_within_the_Active_Site [Accessed 15 Dec. 2020].

Oliveira, L.C. de, Volpon, L., Rahardjo, A.K., Osborne, M.J., Culjkovic-Kraljacic, B., Trahan, C., Oeffinger, M., Kwok, B.H. and Borden, K.L.B. (2019). Structural studies of the eIF4E–VPg complex reveal a direct competition for capped RNA: Implications for translation. *Proceedings*

of the National Academy of Sciences, [online] 116(48), pp.24056–24065. Available at: https://www.pnas.org/content/116/48/24056 [Accessed 11 Feb. 2022].

Ordonez, N., Seidl, M.F., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B.P.H.J., Ploetz, R.C. and Kema, G.H.J. (2015). Worse Comes to Worst: Bananas and Panama Disease—When Plant and Pathogen Clones Meet. *PLOS Pathogens*, [online] 11(11), p.e1005197. Available at: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005197 [Accessed 2 Mar. 2021].

Pratt, A.J. and MacRae, I.J. (2009). The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *Journal of Biological Chemistry*, [online] 284(27), pp.17897–17901. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2709356/.

Rey, P. and Tarrago, L. (2018). Physiological Roles of Plant Methionine Sulfoxide Reductases in Redox Homeostasis and Signaling. *Antioxidants*, [online] 7(9), p.114. doi:10.3390/antiox7090114

Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, [online] 5(4), pp.725–738. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2849174/ [Accessed 10 May 2019].

Sankaranarayanan, R., Palani, S.N., Kumar, A., Selvakumar A. S., P. and Tennyson, J. (2020). Prediction and experimental confirmation of banana bract mosaic virus encoding miRNAs and their targets. *ExRNA*, [online] 2(1). Available at: https://exrna.biomedcentral.com/articles/10.1186/s41544-019-0044-7 [Accessed 27 Feb. 2021].

Schaad, M.C., Anderberg, R.J. and Carrington, J.C. (2000). Strain-Specific Interaction of the Tobacco Etch Virus NIa Protein with the Translation Initiation Factor eIF4E in the Yeast Two-Hybrid System. *Virology*, [online] 273(2), pp.300–306. Available at: https://www.sciencedirect.com/science/article/pii/S0042682200904161?via%3Dihub [Accessed 12 Oct. 2019].

Schrödinger, L. & DeLano, W., 2020. *PyMOL*, Available at: http://www.pymol.org/pymol.

Scrace, J., Denton, J.O. and Clover, G.R.G. (2015). Potyviruses and tobraviruses infecting ornamental Allium species in the United Kingdom. *New Disease Reports*, [online] 32, pp.13–13. Available at: https://www.ndrs.org.uk/article.php?id=032013 [Accessed 23 Feb. 2022].

Shen, W., Shi, Y., Dai, Z. and Wang, A. (2020). The RNA-Dependent RNA Polymerase NIb of Potyviruses Plays Multifunctional, Contrasting Roles during Viral Infection. *Viruses*, [online] 12(1), p.77. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7019339/ [Accessed8

66

Mar. 2020].

Sun, P., Austin, B.P., Tözsér, J. and Waugh, D.S. (2010). Structural determinants of tobacco vein mottling virus protease substrate specificity. *Protein Science*, [online] 19(11), pp.2240–2251. doi:10.1002/pro.506.

Takarabio.com. 2022. *Yeastmaker™ Yeast Transformation System 2 User Manual*. [online] Available at: <https://www.takarabio.com/documents/User%20Manual/PT1172/PT1172-1.pdf> [Accessed 8 October 2022].

Urcuqui-Inchima, S., Haenni, A.-L. and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research*, 74(1-2), pp.157–175.

van Lun, M., van der Spoel, D. and Andersson, I. (2011). Subunit interface dynamics in hexadecameric rubisco. *Journal of Molecular Biology*, [online] 411(5), pp.1083–1098. doi:10.1016/j.jmb.2011.06.052.

Valli, A.A., Gallo, A., Rodamilans, B., López-Moya, J.J. and García, J.A. (2017). The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. *Molecular Plant Pathology*, [online] 19(3), pp.744–763. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6638112/ [Accessed 28 Feb. 2021].

Verchot, J. and Carrington, J.C. (1995). Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *Journal of Virology*, [online] 69(6), pp.3668–3674. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC189082/ [Accessed 2 Mar. 2022].

Vieira Dos Santos, C., Cuiné, S., Rouhier, N. and Rey, P. (2005). The Arabidopsis Plastidic Methionine Sulfoxide Reductase B Proteins. Sequence and Activity Characteristics, Comparison of the Expression with Plastidic Methionine Sulfoxide Reductase A, and Induction by Photooxidative Stress. *Plant Physiology*, [online] 138(2), pp.909–922. doi:10.1104/pp.105.062430.

Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N. and Miller, W.A. (2012). Interaction of the trans-frame potyvirus protein P3N-PIPO with host protein PCaP1 facilitates potyvirus movement. *PLoS pathogens*, [online] 8(4), p.e1002639. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22511869 [Accessed 22 Feb. 2020].

Wang, Y., Kavran, J.M., Chen, Z., Karukurichi, K.R., Leahy, D.J. and Cole, P.A. (2014).

Regulation of S-adenosylhomocysteine hydrolase by lysine acetylation. *The Journal of Biological Chemistry*, [online] 289(45), pp.31361–31372. Available at: https://pubmed.ncbi.nlm.nih.gov/25248746/ [Accessed 31 May 2021].

www.genscript.com. (n.d.). *Restriction Enzyme Map Analysis*. [online] Available at: https://www.genscript.com/tools/restriction-enzyme-map-analysis [Accessed 23 Feb. 2022]

www.neb.com. (n.d.). *Cleavage Close to the End of DNA Fragments*. [online] Available at: https://www.neb.com/~/media/nebus/files/chart%20image/cleavage_olignucleotides_old.pdf.

www.thermofisher.com. (n.d.). *Multiple Primer Analyzer - US*. [online] Available at: https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html [Accessed 19 Jul. 2020].

Yang, M., Zhang, F., Wang, F., Dong, Z., Cao, Q. and Chen, M. (2015). Characterization of a Type 1 Metallothionein Gene from the Stresses-Tolerant Plant Ziziphus jujuba. *International Journal of Molecular Sciences*, [online] 16(8), pp.16750–16762. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4581167/ [Accessed 28 Feb. 2022].

Yang, Z., Zhang, Y., Wang, G., Wen, S., Wang, Y., Li, L., Xiao, F. and Hong, N. (2021). The p23 of Citrus Tristeza Virus Interacts with Host FKBP-Type Peptidyl-Prolylcis-Trans Isomerase 17-2 and Is Involved in the Intracellular Movement of the Viral Coat Protein. *Cells*, [online] 10(4), p.934. doi:10.3390/cells10040934.

Yan, Y., Tao, H., He, J. and Huang, S.-Y. (2020). The HDOCK server for integrated protein–protein docking. *Nature Protocols*, [online] 15(5), pp.1829–1852. Available at: https://www.nature.com/articles/s41596-020-0312-x [Accessed 25 Oct. 2020].
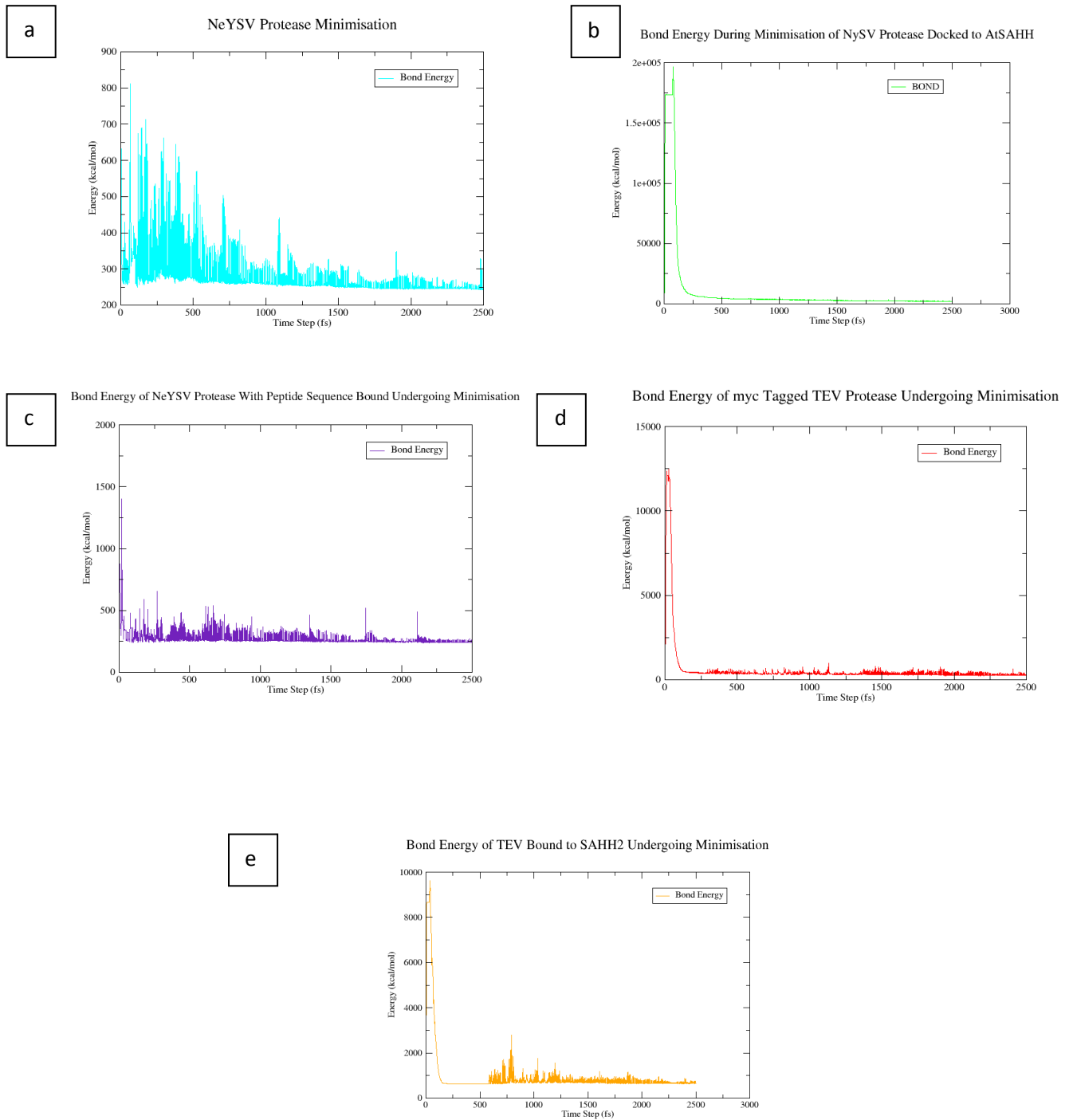
Zhang, Z., Hu, F., Sung, M.W., Shu, C., Castillo-González, C., Koiwa, H., Tang, G., Dickman, M., Li, P. and Zhang, X. (2017). RISC-interacting clearing 3'- 5' exoribonucleases (RICEs) degrade uridylated cleavage fragments to maintain functional RISC in Arabidopsis thaliana. *eLife*, [online] 6. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5451212/ [Accessed 1 Dec. 2020].

Zheng, N., Pérez, J. de J., Zhang, Z., Domínguez, E., Garcia, J.A. and Xie, Q. (2008). Specific and efficient cleavage of fusion proteins by recombinant plum pox virus NIa protease. *Protein Expression and Purification*, [online] 57(2), pp.153–162. Available at: https://www.sciencedirect.com/science/article/pii/S1046592807002501?via%3Dihub [Accessed

31 May 2021].

Zhou, P., Jin, B., Li, H. and Huang, S.-Y. (2018). HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic Acids Research*, [online] 46(W1), pp.W443–W450. Available at: https://academic.oup.com/nar/article/46/W1/W443/4994210 [Accessed 27 May 2021].
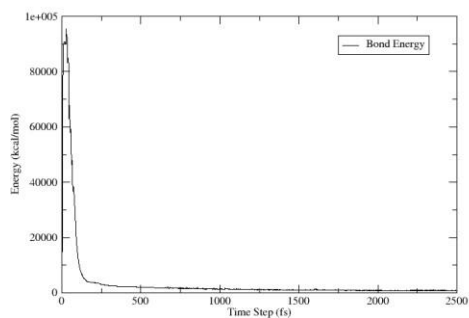
6. Supplementary Data

a NeYSV Protease Minimisation

b Bond Energy During Minimisation of NySV Protease Docked to AtSAHH

c Bond Energy of NeYSV Protease With Peptide Sequence Bound Undergoing Minimisation

d Bond Energy of myc Tagged TEV Protease Undergoing Minimisation

e Bond Energy of TEV Bound to SAHH2 Undergoing Minimisation

Figure 6.1. Results of energy minimisation performed on NeYSV NIa-pro (a),NeYSV-pro bound to SAHH2 (b), NeYSV NIa-pro docked to the peptide sequence (c), TEV NIa-pro with Myc tag added (d) and TEV NIa-pro bound to SAHH2 (e). The point where the energy plateaus and a final local energy minimum has been found is at the points 2000,100,2200,2000 and 100 respectively.

Figure 6.2. Results of energy minimisation performed on TEV (b) and NeYSV (a) NIa-pro bound to NbSAHH2. The point where the energy plateaus, and a final local energy minimum has been found, is at the time step 50 in both simulations.
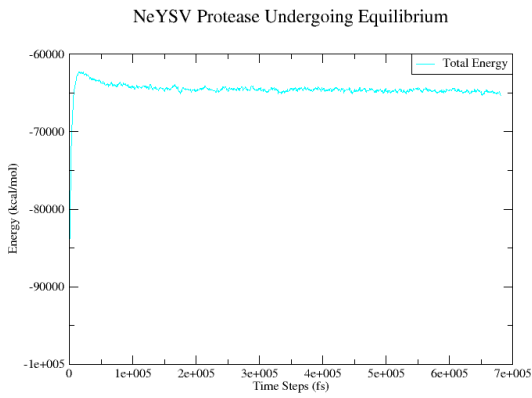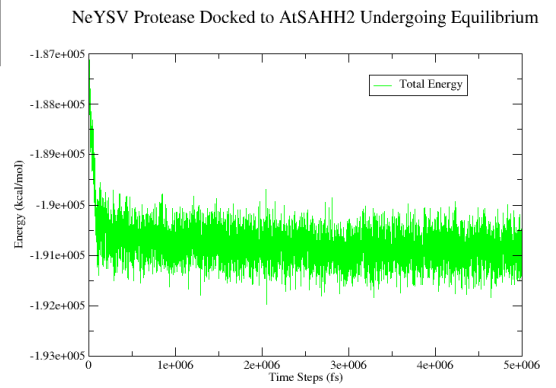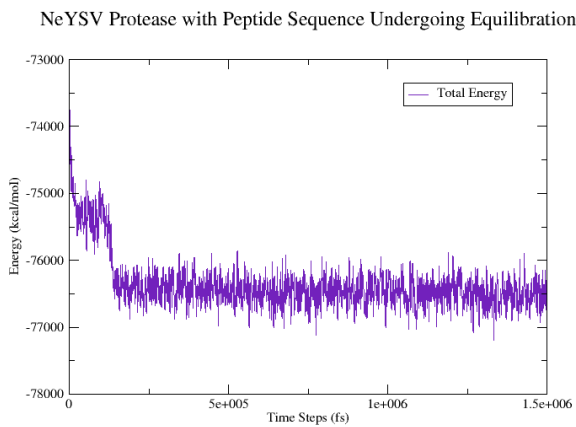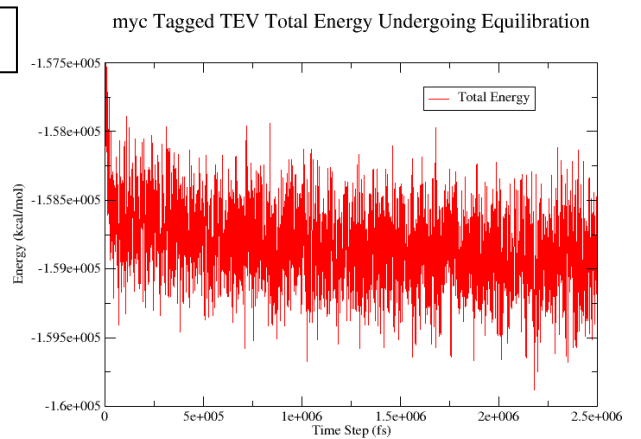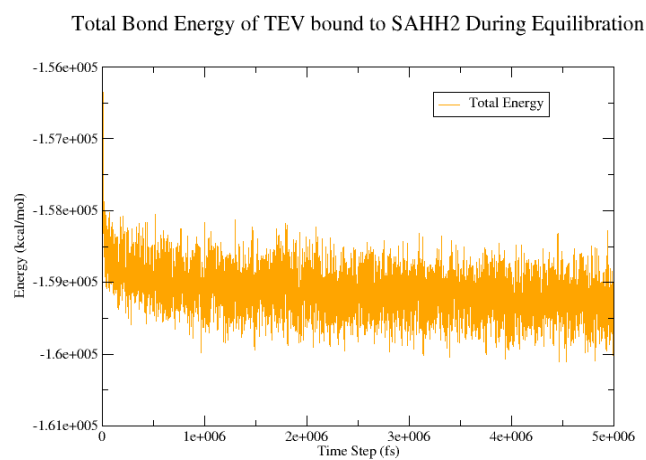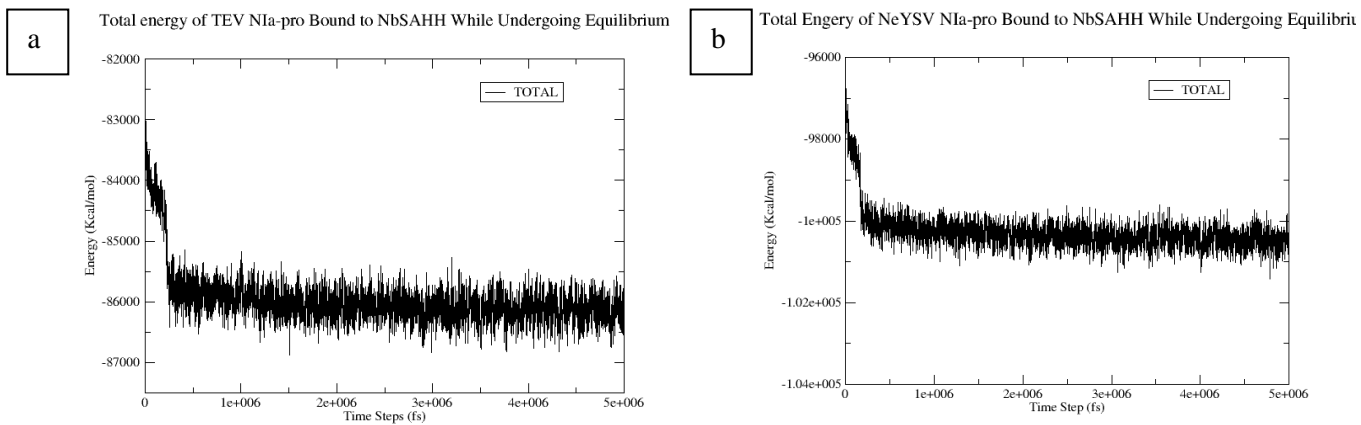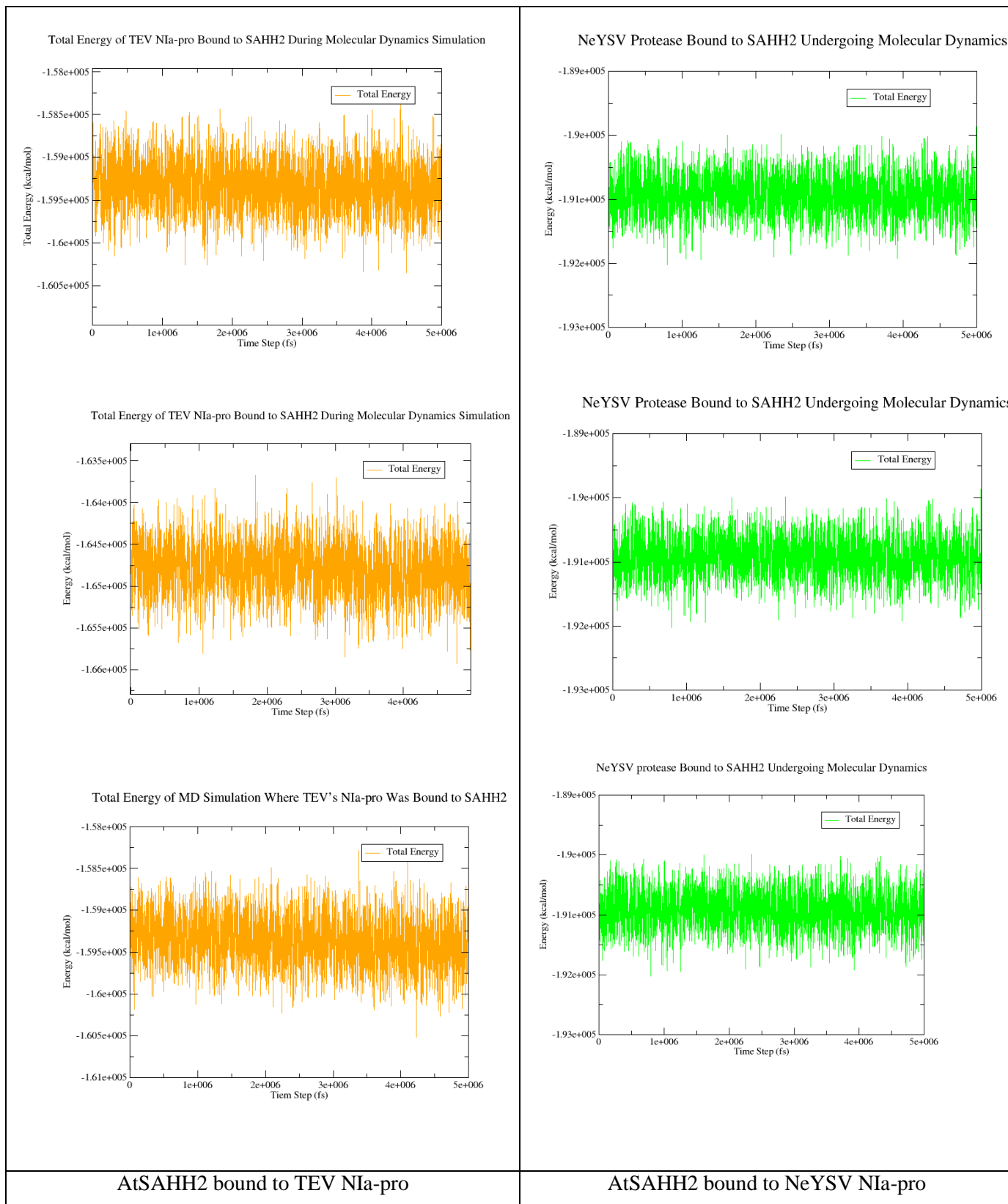
Figure 6.3. Results displaying the total energy from the equilibration performed on the NeYSV protease (a), the NeYSV NIa-pro docked to AtSAHH2 (b), the NeYSV NIa-pro docked to the peptide sequence (c), the myc tagged TEV NIa-pro (d) and AtSAHH2 bound to TEV NIa-pro (e).

a Total energy of TEV NIa-pro Bound to NbSAHH While Undergoing Equilibrium

**b** Total Engery of NeYSV NIa-pro Bound to NbSAHH While Undergoing Equilibriu

Figure 6.4. Results displaying the total energy from the equilibration simulation performed on NbSAHH2 bound to TEV NIa-pro (a) and NbSAHH2 bound to NeYSV NIa-pro (b).

73

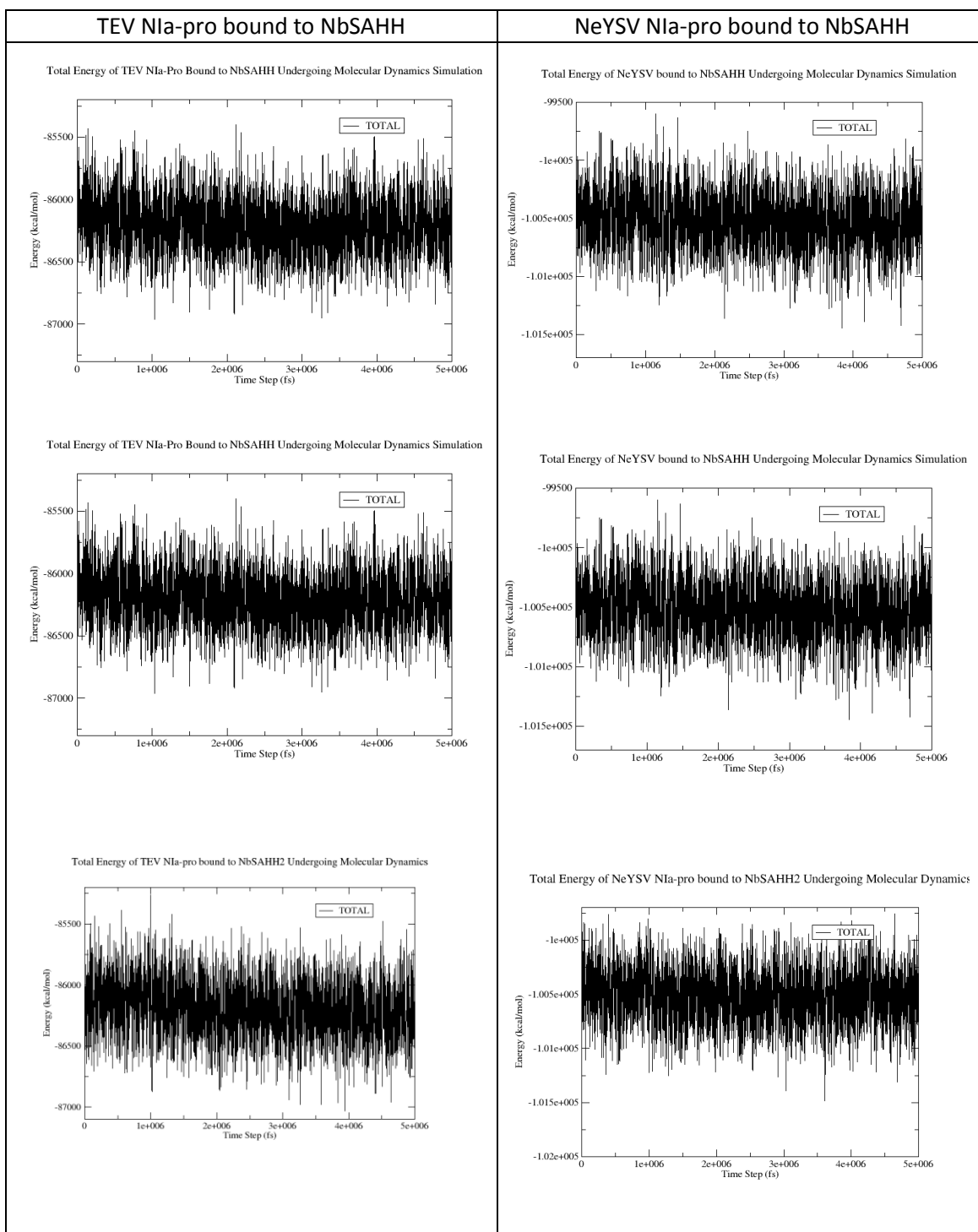| AtSAHH2 bound to TEV NIa-pro | AtSAHH2 bound to NeYSV NIa-pro |

Figure 6.5 Results displaying the total energy of the molecular dynamics simulation performed on AtSAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.

68
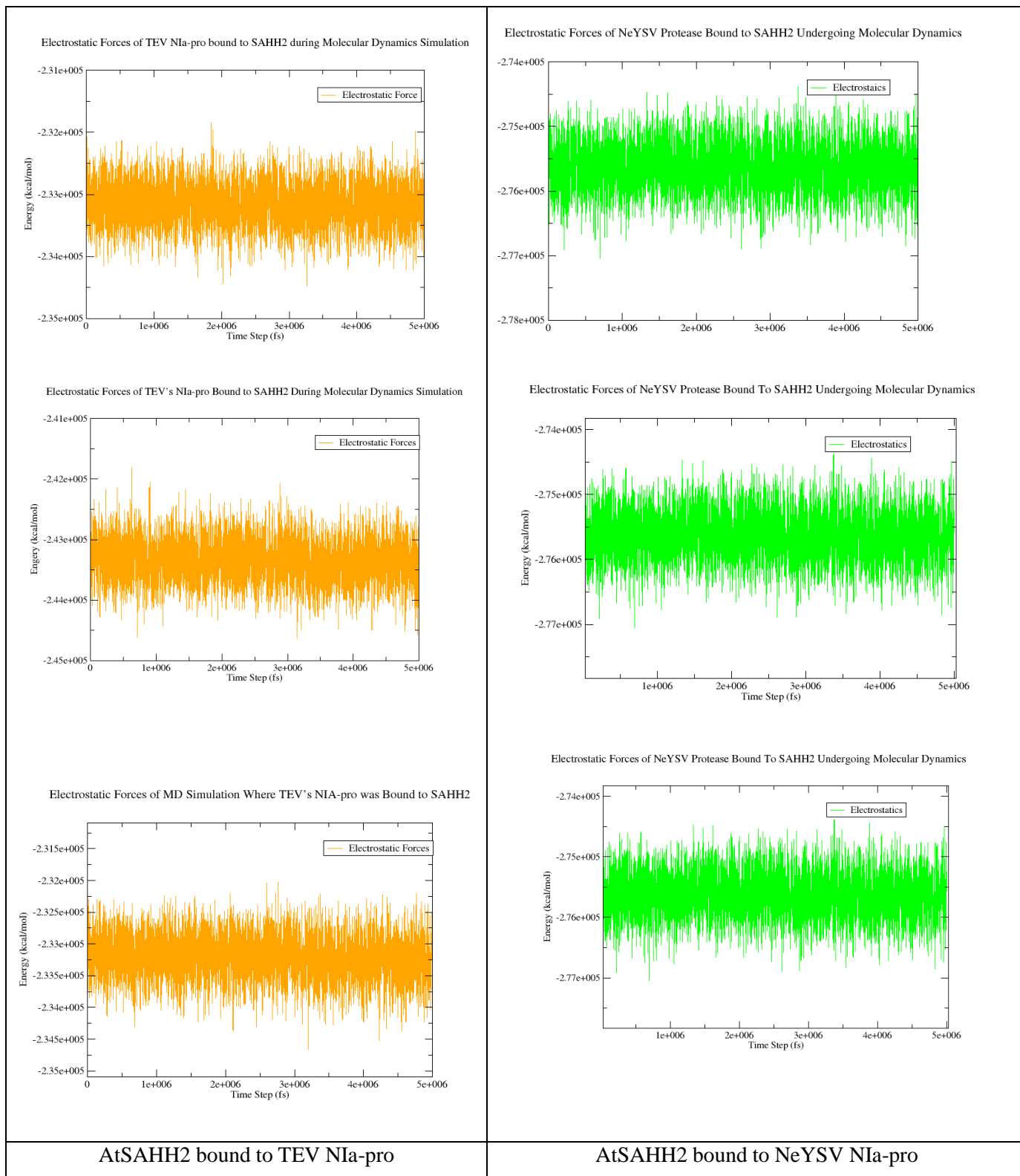
Figure 6.6. Results displaying the electrostatic forces of the molecular dynamics simulation performed on AtSAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.

69
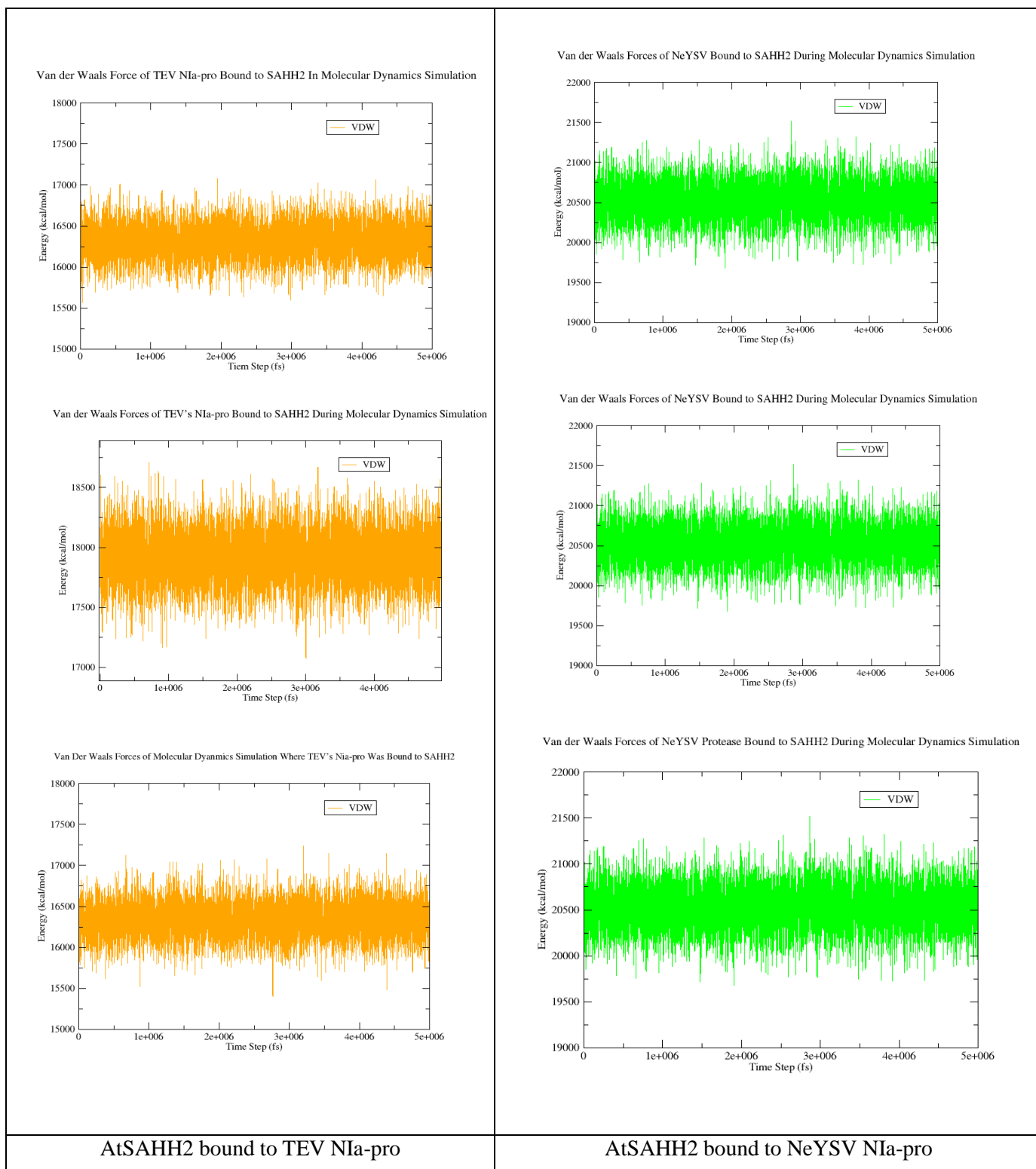
Figure 6.7. Results displaying the Van Der Waals forces of the molecular dynamics simulation performed on SAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.

Figure 6.8. Results displaying the total energy of the molecular dynamics simulation performed on AtSAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.
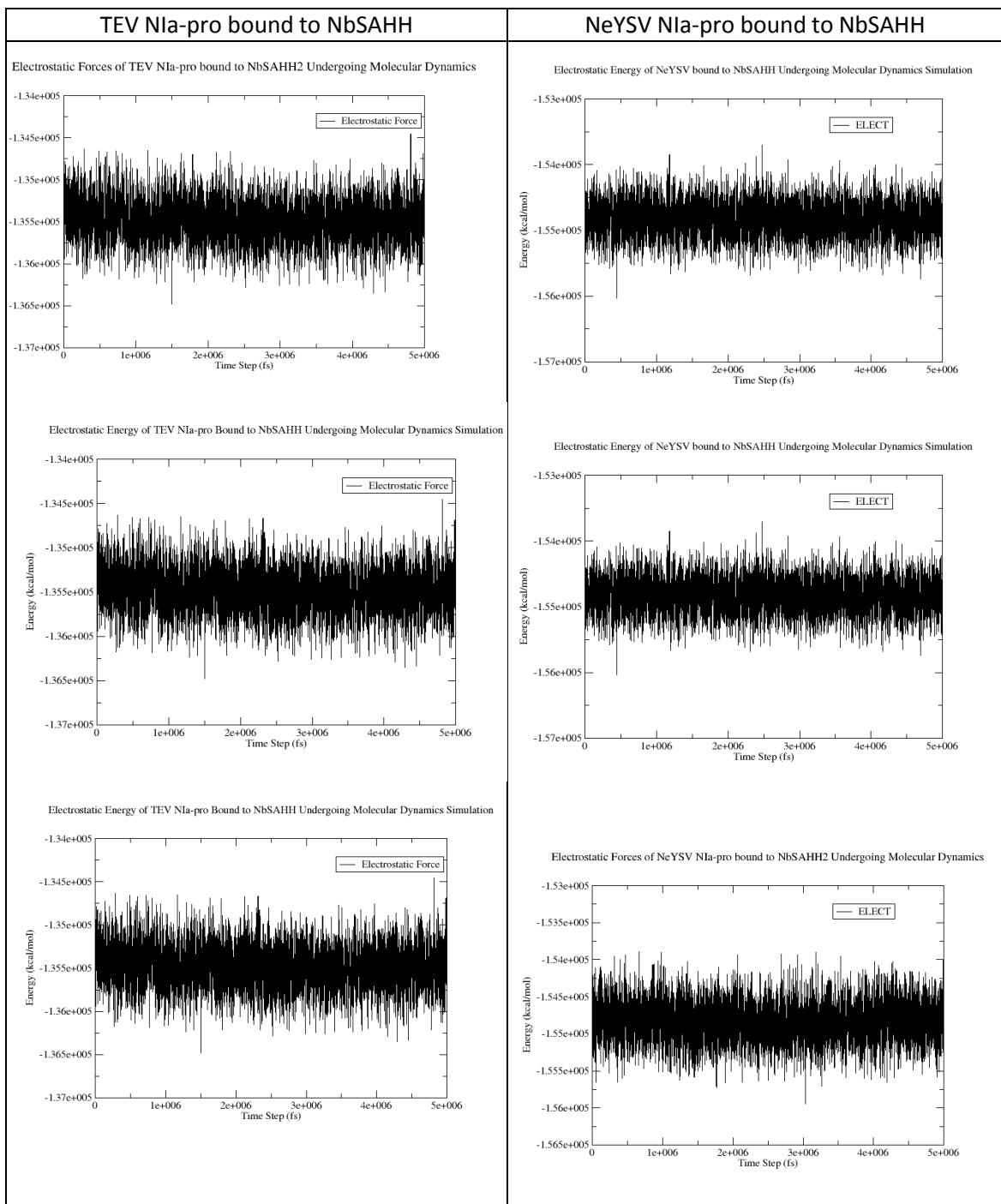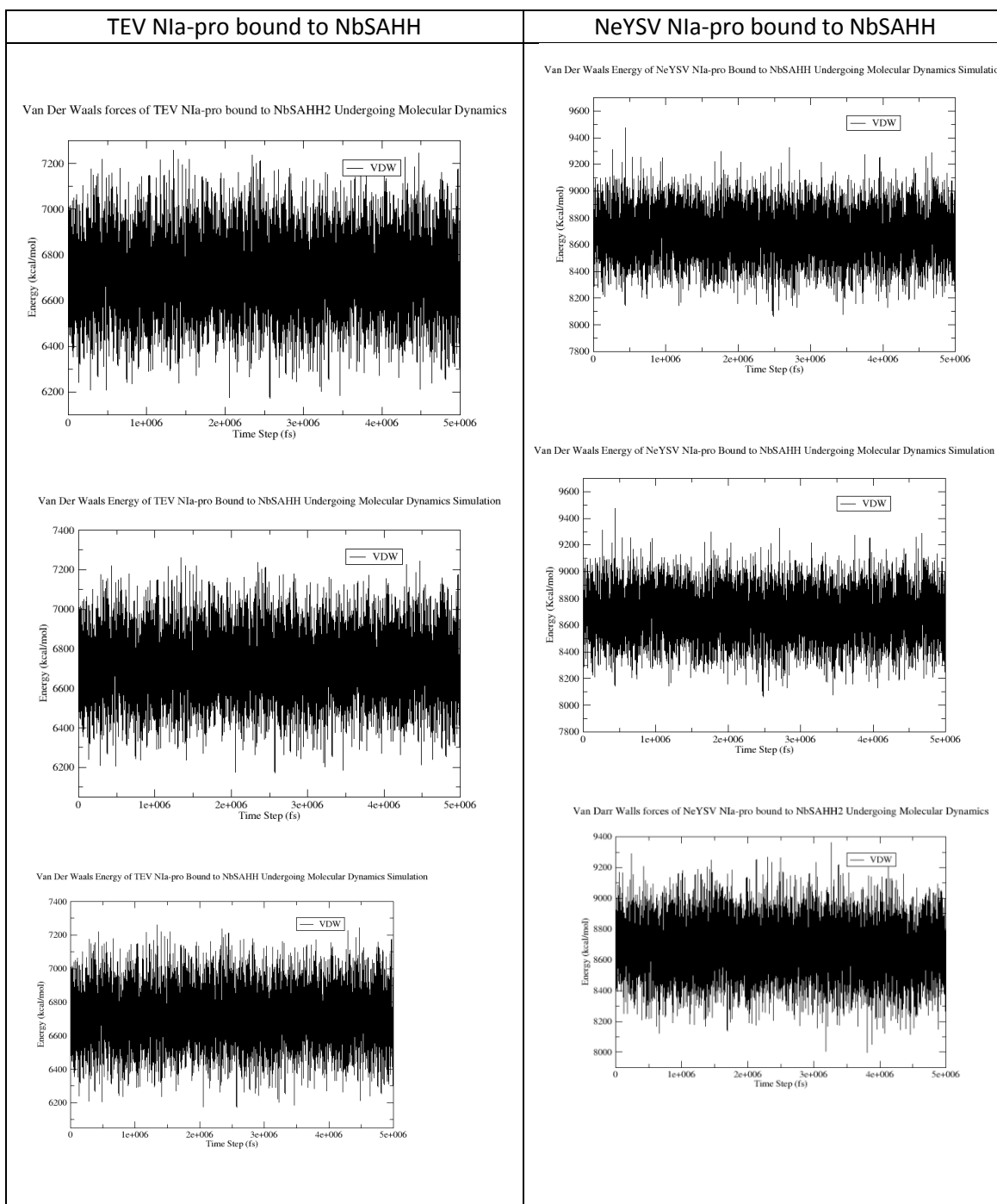
TEV NIa-pro bound to NbSAHH	NeYSV NIa-pro bound to NbSAHH

Figure 6.9. Results displaying the electrostatic forces of the molecular dynamics simulation performed on AtSAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.

| TEV NIa-pro bound to NbSAHH	NeYSV NIa-pro bound to NbSAHH

Figure 6.10. Results displaying the Van Darr Walls forces of the molecular dynamics simulation performed on AtSAHH2 bound to TEV NIa-pro and AtSAHH2 bound to NeYSV NIa-pro.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    5         15        25        35        45        55        65
SAHH_HUMAN      -------MSD KLPYKVADIG LAAWGRKALD IAENEMPGLM RMRERYSASK PLKGARIAGC LHMTVETAVL
SAHH2 Arab      MALLVEKTSS GREYKVKDMS QADFGRLEIE LAEVEMPGLV SCVTEFGPSQ PLKGARITGS LHMTIQTAVL


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    75        85        95        105       115       125       135
SAHH_HUMAN      IETLVTLGAE VQWSSCNIFS TQDHAAAAIA KAGIPVYAWK GETDEEYLWC IEQTLYFK-D GPLNMILDDG
SAHH2 Arab      IETLTALGAE VRWCSCNIFS TQDHAAAAIA RDSAAVFAWK GETLQEYWWC TERALDWGPG GGPDLIVDDG


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    145       155       165       175       185       195       205
SAHH_HUMAN      GDLTNLIHT- ---------- ---------- ---------- ---------- KYPQLLPGIR GISEETTTGV
SAHH2 Arab      GDATLLIHEG VKAEEIFAKN GTFPDPTSTD NPEFQIVLSI IKDGLQVDPK KYHKMKERLV GVSEETTTGV


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    215       225       235       245       255       265       275
SAHH_HUMAN      HNLYKMMANG ILKVPAINVN DSVTKSKFDN LYGCRESLID GIKRATDVMI AGKVAVVAGY GDVGKGCAQA
SAHH2 Arab      KRLYQMQETG ALLFPAINVN DSVTKSKFDN LYGCRHSLPD GLMRATDVMI AGKVAVICGY GDVGKGCAAA


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    285       295       305       315       325       335       345
SAHH_HUMAN      LRGFGARVII TEIDPINALQ AAMEGYEVTT MDEACQEGNI FVTTTGCIDI ILGRHFEQMK DDAIVCNIGH
SAHH2 Arab      MKTAGARVIV TEIDPICALQ ALMEGLQVLT LEDVVSEADI FCTTTGNKDI IMVDHMRKMK NNAIVCNIGH


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    355       365       375       385       395       405       415
SAHH_HUMAN      FDVEIDVKWL NEN-AVEKVN IKPQVDRYRL KNGRR-IILL AEGRLVNLGC AMGHPSFVMS NSFTNQVMAQ
SAHH2 Arab      FDNEIDMLGL ETYPGVKRIT IKPQTDRWVF PDTNSGIIVL AEGRLMNLGC ATGHPSFVMS CSFTNQVIAQ


                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
                    425       435       445       455       465       475       485
SAHH_HUMAN      IELWTHP--D KYPVGVHFLP KKLDEAVAEA HLGKLNVKLT KLTEKQAQYL GMSCDGPFKP DHYRY
SAHH2 Arab      LELWNEKSSG KYEKKVYVLP KHLDEKVAAL HLGKLGARLT KLTKDQSDYV SIPVEGPYKP VHYRY
```

Figure 6.11 Alignment between the *Arabidopsis* SAHH2 protein (OAP04800.1) and human SAHH protein (AAP36293.1). The insertion region of the AtSAHH2 has been outlined in green. The Human SAHH2 C-terminal tail and the AtSAHH2 residues that it aligned to are outlined in red.
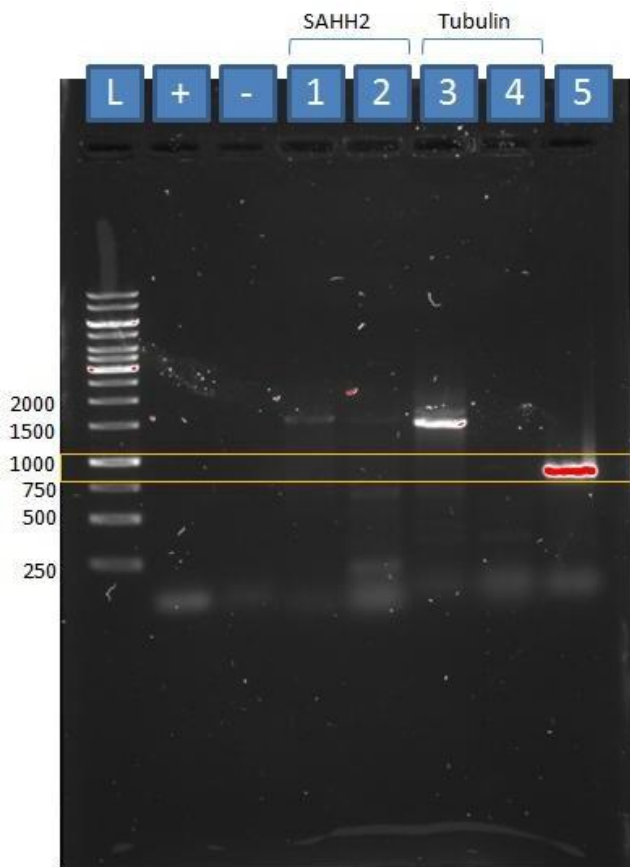
Figure 6.12 Gel Image of *Arabidopsis* (Ler-0), NeYSV and *Nicotiana* cDNA PCR amplified with the primers outlined in Table 2.1 and Table 2.2. Under the identity heading is the cDNA; primer type; temperature of the PCR reactions annealing stage and the expected band size of the PCR fragment. The gels showed that the *Arabidopsis* Tubulin and NeYSV NIa-pro PCRs formed strong bands around the expected size. The *Arabidopsis* and *Nicotiana* SAHH formed weak bands.