STRUCTURE OF HIBISCUS LATENT SINGAPORE VIRUS

DETERMINED BY X-RAY FIBER DIFFRACTION

SUNIL KUMAR TEWARY

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPROE

2010

STRUCTURE OF HIBISCUS LATENT SINGAPORE VIRUS

DETERMINED BY X-RAY FIBER DIFFRACTION

SUNIL KUMAR TEWARY

MSc. (Biotech.), M.Tech. (Biotech. Biochem. Engg.)

A THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPROE

2010

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF PUBLICATIONS

Tewary, S. K., Oda, T., Kendall, A., Bian, W., Stubbs, G., Wong, S. M and Swaminathana, K. (2011). Structure determination of Hibiscus latent Singapore virus by X-ray fiber diffraction: a non-conserved His122 contributes to coat protein stability. Journal of Molecular Biology 406: 516-526.

LIST OF ABBREVIATIONS

BSMV	Barley stripe mosaic virus
CGMMV	Cucumber green mottle mosaic virus
CP	Coat protein
CsCl	Cesium chloride
EDTA	Ethylene-diamine-tetra-acetic acid
FD	Fiber diffraction
FT	Fourier transform
HLSV	Hibiscus latent Singapore virus
IR	Isomorphous replacement
MD	Molecular dynamics
MDIR	Multidimensional isomorphous replacement
MP	Movement protein
MR	Molecular replacemet
OAS	Origin of assembly sequence
PVX	Potato virus X
RdRp	RNA dependent RNA polymerase
RLS	Restrained least square
RMV	Ribgrass mosaic virus
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecylsulfate-polyacrylamide electrophoresis
SHMV	Sunn hemp mosaic virus
SIR	Single isomorphous replacement
TEM	Transmission electron microscope

TMV	Tobacco mosaic virus
TRV	Tobacco rattle virus
UTR	Untranslated region
UV	Ultra violet
VLPs	Virus like particles

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Hibiscus latent Singapore virus (HLSV) is a new member of the *Tobamovirus* family. The HLSV genome contains a unique poly(A) tract in its 3′-UTR which is absent in other *Tobamoviruses*. The virion is composed of a monomeric coat protein (CP) of 18 kDa. We have determined the HLSV structure at 3.5 Å by X-ray fiber diffraction with R factor of 0.096. The structure of HLSV CP resembles that of other *Tobamoviruses*, with a few unique differences. In other *Tobamoviruse* structure, CP sequence at position 122 contains a conserved Arg residue, while the HLSV and SHMV contain His residue. Also, His122 is followed by another positively charged amino acid residue Lys which is uncharged residue in other *Tobamoviruses*. There is a kink observed for the first time in the LR helix of HLSV due to the presence of the unique His122, which produces a bend in the helix in the non-Pro non-Gly bends. Also, the adjacent Lys123 may destabilize the helix by positive charge repulsion, making the kink more pronounced. In the HLSV structure, we are able to see Lys123 stabilizing the phosphate 1, hence balancing the protein-nucleic acid interactions. Another residue Arg92 from the Subunit -17 is believed to be involved in stabilizing the remaining phosphate 2 and phosphate 3. Arg122 is believed to regulate the guanine 1 recognition during assembly for all other existing structures of the *Tobamovirus*. Uniquely, His122 at this position showed a very strong salt bridge with the neighboring Asp88 from subunit -1, hence significantly stabilizing the loop adjacent to RR helix. The carboxyl-carboxylate interactions that drive viral disassembly are also seem to be different in HLSV. The nucleotide recognition mechanism for virus assembly is similar between HLSV and RMV but different from that of TMV and CGMMV.

By solving the structure of HLSV by X-ray fiber diffraction, we will be able to have a better understanding of the structural differences between HLSV and other *Tobamoviruses*. This research may also enhance our knowledge of virus structure at atomic details. By knowing the atomic details of this novel virus, we may be able to use it in future as a vector to express pathogenic epitopes (to develop vaccine) and to express economically important proteins.

CHAPTER 1. X-RAY FIBER DIFFRACTION TECHNIQUES

1.1 INTRODUCTION

Many biological molecules that are polymeric in nature are long, structurally helical and have a natural tendency to form fibers. This prevents the growth of single crystals from these polymeric fibers and even if crystals are grown, the molecular interactions in the crystals hardly represent any biologically significant interactions in the fibers. Conventional X-ray crystallography is therefore of very less use to fibrous biomolecules. These macromolecular helical aggregates are too large to be studied by nuclear magnetic resonance. Fiber diffraction (FD), however, is a powerful method for determining three dimensional (3-D) structural details of fibrous polymers. This technique has been used to study a wide variety of biopolymers, ranging from simple polypeptides, polynucleotides, and polysaccharides to very complex filamentous viruses and cytoskeletal filaments.

Fiber samples lack true 3-D crystalline structure. The key difference between fibers and crystals is that in fibers, structural aggregates, although parallel to each other, are randomly oriented (disordered) about the fiber axis. Consequently the diffraction pattern is cylindrically averaged. The cylindrical averaging is the defining property of the FD samples. The combination of cylindrical averaging and inherent disorder makes structural analysis of the fibrous filaments difficult. Hence, FD is not an appropriate method for studying molecules that do not naturally form filaments. Proteins that are fibrous in nature are important structural component of biological systems such as skin, bone, hair and tendons cannot be crystallized by the protein X-ray crystallographic technique. These samples are suitable for study by the method of

FD. One of the most active areas of research is the study of the relationship between structure and mechanical properties in spider dragline silk, using FD technique in conjunction with electron microscopy, spectroscopy and molecular biology methods (Grubb & Ji, 1999; Winkler et al., 1999). Studies of collagen fibers have been focused on its mechanical properties, in correlation to alignment pattern of fibrils with its different stages of development and stress (James et al., 1998; Purslow et al., 1998). Amyloid fibrils and other related fibrous aggregates have also been studied using the FD method (Inouye et al., 1998; Kirschner et al, 1998; Malinchik et al., 1998; Sunde et al., 1997). They are associated with various pathological conditions, including prion infections and Alzheimer's disease. They are formed when soluble proteins mis-fold to form insoluble cross-β structures. As a result of disorder in the fibrils, these insoluble proteins are often analyzed by using the FD technique. The relevance of these fibrils to medicine and fundamental questions of protein folding make it an interesting for research.

The experimental set-up for FD includes orientated fiber packed in a very thin quartz capillary tube and placed perpendicular to a collimated X-ray beam. A FD pattern is recorded on film. Fibers show helical symmetry rather than the 3-D symmetry assumed by crystals. The diffraction data recorded on film appears as layer lines as opposed to spots produced by a protein crystal. This difference in the FD pattern is due to the repeating nature of the polymer helix at a distance inversely proportional to the filament repeat distance. By analyzing the diffraction pattern from orientated fibers, one can find out the helical symmetry of the sample and may solve the structure. The following section discusses the theory of FD and its applications in research, focused mainly on filamentous viruses.

1.2 THEORY OF FIBER DIFFRACTION

1.2.1 Diffraction by a helical structure

Filamentous viruses, when exposed to X-rays, give rise to non-crystalline FD patterns. Therefore we discuss the theory of non-crystalline FD. However, much of the theory given here also applies to crystalline fibers, (Chandrasekaran & Stubbs, 2001). In general, intensity at any point in a diffraction pattern is given by

$$I = FF^* = |F|^2 \qquad (1)$$

where F is a structure factor (a complex number with amplitude and a phase). F is a function in 3-D reciprocal space, and is the Fourier transform (FT) of electron density (ρ, a function in real space) of the diffracting object (Rhodes, 2000). Equation (1) applies equally to FD and crystallographic diffraction patterns.

Helical molecules are described in cylindrical coordinates because of cylindrical symmetry. The coordinates are denoted as (r, φ, z) in real space and (R, ψ, Z) in reciprocal space. The z axis is the helix axis where the structure repeats exactly in a distance c (repeat distance) along this axis. The repeat contains u subunits in t turns of the helix. Diffraction from a helically symmetric molecule occurs only on layer lines, when $Z = l/c$, where l is layer line. On the equator, $Z = l = 0$. The line perpendicular to the equator and passing through the centers of all layer lines in the diffraction pattern is called the meridian. The distance from the meridian is denoted as R whereas Z is denoted as the distance from the equator. It was shown (Cochran and Crick, 1952) that the structure factor is given as,

$$F(R, \psi, Z) = \sum_j \cdot \sum_n f_j J_n(2\pi R r_j) \exp\left[i\left\{n\left(\psi + \frac{\pi}{2}\right) - n\varphi_j + 2\pi l z_j/c\right\}\right] \qquad (2)$$

where r_j, φ_j and z_j denote real space coordinates of atom j in the subunit and f_j is the

atomic scattering factor of that atom. J_n is the Bessel function of the first kind of order n. The value of n is restricted to satisfy the selection rule

$$l = tn + um$$

where *t* is number of turns in one repeat distance; *u* represents the number of subunits in one repeat distance; *n* and *m* are integers. The number of significant terms contributing to the structure factor is limited, since J_n(x) is generally negligible for x less than about n-2. Cylindrical averaging can be taken into account by re-writing equation (2) in terms of Fourier-Bessel structure factors G (Klug et al., 1958), where

$$Gn, l(R) = \sum_j f_j J_n(2\pi R r_j) \exp i(-n\varphi_j + 2\pi l z_j/c) \tag{3}$$

Gn,l(R) is independent of ψ. Then

$$F(R, \psi, Z) = \sum_n G_{n,l}(R) \exp[i\{n(\psi + \frac{\pi}{2})\}] \tag{4}$$

and it can be shown (Franklin and Klug 1955; Waser, 1955) that

$$I(R, l) = FF^* = \sum Gn, l(R)G^*_{nl}(R) = \sum |G_{nl}(R)|^2 \tag{5}$$

The number of significant terms in the summation is limited, e.g. in the diffraction pattern of Tobacco mosaic virus (TMV), the largest number of terms contributing to any intensity at a resolution of 3 Å is eight. Near the meridian, a single G term makes up the whole intensity. Equation (5) may be compared to the corresponding crystallographic equation (1). In crystallography, the diffracted intensity is the square of the amplitude of a single structure factor, whereas in FD the diffracted intensity is the sum of the squares of the Fourier-Bessel structure factors. The summation occurs because of the cylindrical averaging of the diffraction pattern, and may be thought of as the superimposition of the diffracted intensities. The electron density in a non-crystalline fiber may be calculated by means of a Fourier-Bessel synthesis, an analogue of Fourier synthesis used in crystallography and crystalline FD. The electron density at a point r, φ, z is

$$\rho(r, \varphi, z) = (1/c) \sum_j \cdot \sum_n g_{n,l}(r) \exp[i(-n\varphi_j + 2\pi l z_j/c)] \qquad (6)$$

where

$$gn, l(r) = \int_0^\infty \cdot G_{n,l}(R) J_n(2\pi R r_j) 2\pi R \, dR \qquad (7)$$

The inner part of the equator in FD is derived from a single G term, $G_{0,0,}$ whose phase is either 0 or 180°; that is, $G_{0,0}$ simply has a positive or negative sign. If the signs are known, or can be estimated, this part of the equator can be transformed to obtain a radial density distribution.

1.2.2 Nature of fiber diffraction

Fundamental structural aggregates of a fibrous sample are regarded as filaments, and complete diffraction from specimen as fiber. Fibers are collection of nearly parallel filaments, which are randomly oriented about the fiber axis. These structural aggregates might be individual virions in a fiber or an oriented solution, individual molecules of DNA or some other chain molecule. Diffraction from a fiber sample is confined to layer lines, because of the repeating nature of the filament helix at spacing inversely proportional to the repeat distance c. The layer lines are continuous and correspond to the cylindrical average of the FT of a single molecule. In a FD experiment, individual helical fibers are not perfectly aligned and not parallel to the fiber axis. This deviation from parallelism is called disorientation and it causes reflections from the fibers to spread into arcs. The layer lines in reciprocal space are perpendicular to the fiber axis in real space. The layer line that passes through the origin in reciprocal space is called the equator or zero layer line. The direction normal to the equator is called the meridian. The pitch of a helix is the width of one complete turn, measured parallel to the helical axis. The number of helical turns that marks a unique repeating section of the sample is known as helix repeat and the corresponding

distance is called repeat distance (c). In TMV, the pitch height is 23 Å with 49 subunits in three turns (u=49; t=3). So, the repeat distance c, is 69 Å. For a helical sample, the layer line separation is proportional to the reciprocal of the pitch of the helix. However, the pitch of a helix may not be simply the reciprocal of the layer line separation since there might be two or more helical turns in one repeat distance. Furthermore, two other parameters, fiber specimen tilt and twist must be accurately determined for proper data reduction. The specimen tilt is defined as the deviation of the fiber axis from an absolute normal to the X-ray beam and the twist is the deviation of the fiber axis from the plane of the detector, Fig. 1.1 (Kendall et al., 2007).

Fiber diffraction data contain less information than the equivalent from protein crystals. Firstly, cylindrical averaging reduces the information content of diffraction. The amount of information lost depends on the size and symmetry of filaments and also on the resolution of the diffraction data. Cylindrical averaging affects high resolution data that are away from the axis of rotation much more than that close to the axis. In fact, for highly symmetrical helical filaments, the data near the axis of rotation are cylindrically symmetric, and no data loss occurs due to cylindrical averaging. For filaments of high symmetry viruses, e.g. TMV, cylindrical averaging reduces the effective number of observable diffraction data at 3 Å resolution by a factor of about 2.5, while for Pf1, the corresponding factor is only 1.7 (Makowski, 1982). There is no significant data loss for TMV at 15 Å and for Pf1 at 7 Å resolution. Secondly, the resolution of the diffraction pattern is limited by disorientation of the fiber sample. In practice, the filaments are not perfectly parallel and the mean deviation of the filament axis from the fiber sample axis is typically 1 or 2°, even for a well oriented fiber. As result, the layer lines fan out (overlap), running into each other at high resolution. The severity of this effect obviously depends on the distance

between the layer lines, but in general, useful FD patterns hardly extend beyond 3 Å

resolution. Nevertheless, there are still sufficient data to solve structures.
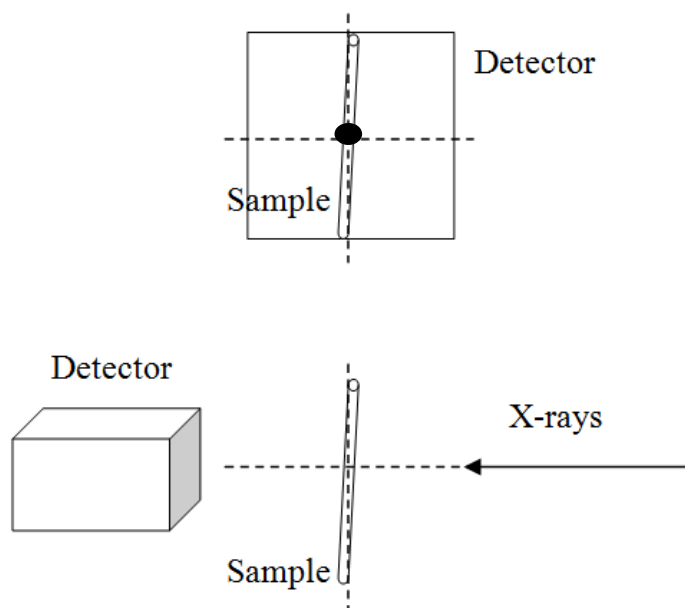
Figure 1.1. **Tilt and twist of a sample**. In the upper panel, the sample is viewed from the X-ray (solid circle) source. The misalignment of the fiber axis, represented as a helical tube, is called tilt. In the lower panel, in a view perpendicular to the upper panel, the twist is the misalignment of the fiber axis.

1.2.3   Crystalline and non crystalline fiber

Fibers in general are classified as crystalline and non-crystalline, and in some

cases intermediate between the two. In non-crystalline fibers, filaments are

approximately parallel to each other, but their positions and orientations are

uncorrelated. Diffraction patters from these fibers are limited to layer lines at spacings

inversely proportional to the filament repeat distance, but are otherwise continuous

and represent the cylindrical average of the diffraction pattern of a single filament.

Most filamentous plant viruses are non-crystalline in nature. The crystalline fiber

filaments form fully ordered microcrystals, usually of a very elongated form, and each

fiber consists of many such crystals randomly oriented about the fiber axis. In

diffraction patterns from the crystalline fibers, the layer lines are sampled to from separate reflection and the diffraction pattern represents the cylindrical average of the diffraction pattern from a single crystal. The A form of DNA is crystalline fiber in nature whereas the B form is non-crystalline fiber. Many filamentous bacteriophage fibers are intermediate because these fibers are usually made by drying (McDonald et al., 2008). The diffraction pattern from this type of fiber shows individual reflections at the center (low resolution) of the equator, but continuous data on all other layer lines.

1.3   STRUCTURE DETERMINATION USING FIBER DIFFRACTION

It is evident from Eqs. 5-7 that in order to determine electron density, the observed intensities in a diffraction pattern must be separated into their component G terms, and the phase of each G term must also be determined. Equivalently, the real imaginary parts of each G term, contributing to each point in the diffraction pattern must be determined. This is the FD equivalent of the phase problem in crystallography, determining the phase for each F. In studies of filamentous viruses, several different methods have been used for this determination.

1.3.1   Multidimensional isomorphous replacement (MDIR)

This method is an extension of the protein crystallographic method of isomorphous replacement (IR) (Namba & Stubbs 1985; Namba & Stubbs, 1987a; Stubbs & Diamond, 1975). If a heavy atom (i.e. an atom for which fj is significantly greater than the other atoms in a structure) can be introduced into the molecule without significantly perturbing the rest of the structure, the diffraction pattern from the resulting derivative can be used together with the original (native) diffraction

pattern to determine the phase of each intensity. Typically, heavy metal atoms such as mercury, platinum and uranium are used. If G = A + iB, then, Eq. 5 can be re-written as

$$I = \sum_{k=1}^{N} \cdot [A^2_k + B^2_k]$$
(8)

omitting the arguments (*R, l*) for clarity and denoting the number of significant G terms contributing to the intensity as N. Each A_k and B_k can be considered to be a component of a 2N-dimensional vector. If a derivative of the diffracting filament contains a heavy atom that contributes $a_k + ib_k$ to the kth G term at (*R, l*), the diffracted intensity from the derivative fiber will be

$$I = \sum_{k=1}^{N} \cdot [(A_k + a_k)^2 + (B_k + b_k)^2]$$
(9)

If the position of the heavy atom is known, a_k and ib_k can be calculated using equation (3). The heavy atom positions can be determined by search methods and refined by least-squares methods (Namba & Stubbs 1985; Stubbs & Diamond, 1975). The cylindrically averaged Patterson function is complicated and the Patterson methods that are used to locate heavy atoms in crystallography are not very useful in FD (McGillavry & Bruins, 1948; Stubbs, 1987). Clearly, given sufficient derivatives, each one giving an equation like equation (9), it is possible to solve the set of equations (8) and (9) for all of the A_n and B_n. At least twice as many heavy atom derivatives as there are G terms are required (e.g. for resolving 2 G terms, 4 derivatives are necessary). The errors in equations (9) can be minimized by using equation (8) as a constraint (Stubbs & Diamond, 1975). Alternatively, a least-squares procedure has been used to minimize the errors in equations (8) and (9) (Namba & Stubbs, 1985). Heavy atom derivatives can be difficult to make, particularly in the numbers required for MDIR. A method analogous to protein crystallographic single isomorphous replacement (SIR) has been developed for use in FD (Wang, 1985;

Namba & Stubbs, 1987a). It was used to determine the structure of Cucumber green mottle mosaic virus (CGMMV) at 5 Å resolution (Lobert & Stubbs, 1990). In this method, the magnitudes of the G terms contributing to a given intensity are either estimated from a model of a related structure (sometimes called the 'proportional amplitude' method) or they are assumed to be equal (Wang & Stubbs, 1994). The equal-amplitude assumption is preferable, as it avoids model bias. The phases are then determined by conventional crystallographic IR or by SIR.

Electron density maps of CGMMV, determined using the equal-amplitude assumption, were as good as maps determined using amplitudes based on the related TMV structure, even at 4.5 Å resolution (Lobert & Stubbs, 1990; Wang & Stubbs, 1994). But at higher resolutions the equal-amplitude assumption may not always be sufficient to obtain a fully interpretable map (Namba & Stubbs 1987a and 1987b). An additional source of information is in the positions of the layer lines in the diffraction pattern. The layer line spacing is determined by the size of the helical repeat, and when the repeat is approximate (fractional value), the G terms in each layer line do not fall at exactly the same values of Z (where Z is the distance from the equator in the reciprocal space), and the layer lines are said to be split (Tollin et al., 1968). Layer lines have a finite thickness and the disorientation in the diffracting specimen increases this thickness, so the splitting is seen only as a small shift in the apparent position of the line. The magnitude and direction of the shift depend on the relative magnitudes of the contributing G terms. The shifts can be measured for each heavy atom derivative and used in the equation (Makowski, 1978),

$$\Theta I = \sum_{k=1}^{N}. \; \Sigma q \theta_k [(A_k + a_k)^2 + (B_k + b_k)^2 \qquad (10)$$

where Θ is the shift observed in the layer line position, measured as an angular displacement about the centre of the diffraction pattern, θ_k is the calculated shift for

the kth G term and q is the ratio of the magnitude of splitting in the derivative to the magnitude of splitting in the native filament (Stubbs & Makowski, 1982). An equation like (10) can be obtained from the native (with $a_k = b_k = 0$) and from each derivative that is oriented sufficiently well to determine accurately. Equation (10) is independent of equation (9) and can be combined with them to determine the values of A_k and B_k. This method was particularly important in determining the structure of TMV for which several extremely well oriented derivatives were available (Namba & Stubbs, 1986).

1.3.2   Molecular replacement

The molecular replacement (MR) method is also used in fiber structure determination when a structure closely related to the structure under investigation is already known. A known structure can be used to estimate the phases and relative magnitudes of the G terms. These phases and relative magnitudes can be applied to the observed data and electron density maps can be calculated. A new model is built, based on the generated electron density by several cycles of refinement until the structure converges. This approach must be used with caution because the data is cylindrically averaged due to which fewer data are available, and the phase solution is consequently more dependent on the initial model. The map may therefore tend to resemble the initial model regardless of the true differences in the structure. This was a problem in earlier structure determinations. Improved methods of refinement, particularly molecular dynamics (MD) refinement have greatly reduced the problem of model bias (Wang & Stubbs, 1993). For example, the structure of the U2 strain of TMV could not at first be determined by MR from the TMV structure, even though these two virus structures are very similar. With the development of MD refinement,

the U2 model was finally solved satisfactorily (Pattanayek & Stubbs, 1992). With MD refinement, the structure of the much difficult Ribgrass mosaic virus (RMV) determined by MR from TMV (Wang et al., 1997) as initial model. Also the structure of the bacteriophage Pf3, determined from the Pf1as a initial model were refined (Welsh et al., 1998). The use of a limited number of heavy atom derivatives in this way may be a useful corrective measurement against excessive bias toward a model structure in molecular replacement.

1.4   REFINEMENT OF FIBER STRUCTURES

Two methods of restrained refinement have been used in FD studies of filamentous viruses, namely, restrained least squares (RLS) and MD. The RLS method was adapted from the protein crystallographic refinement method (Hendrickson, 1985; Stubbs et al., 1986). RLS has been effectively used for refinement of plant virus (Namba et al., 1989) structures. A closely related Jack-Levitt refinement has been used to refine bacteriophage structures (Jack & Levitt, 1978). The radius of convergence of these methods is limited and successful refinement depends on the accuracy of the starting model. The MD is the major choice of refining fiber structures. The most useful application of this method has been simulated annealing (SA), in which the structure is heated to a temperature of 3000-4000 K and then potential energy is minimized as the structure is cooled in small decrements. At higher temperatures, energy barriers between the starting model and the structures of the lower potential can be overcome. In this way the radius of convergence of the refinement is increased. Simulated annealing has been used to refine both bacteriphages and filamentous viruses (Gonzalez, 1995; Wang et al., 1997; Wang & Stubbs 1994; Welsh et al., 1998).

1.5   DIFFERENCE FOURIER AND OMIT MAP IN FIBER DIFFRACTION

Difference Fourier syntheses have been widely used in both protein and small-molecule crystallography to determine structures. It is also used in FD (Mandelkow et al., 1981). This use has been limited by the difficulties peculiar to FD which arise from the cylindrical averaging of FD patterns. Difference Fourier maps calculated from FD data, by direct analogy with crystallographic difference maps, tend to have high noise levels and found to be biased toward the known or model structure.

The crystallographic reflection has two components, one real and one imaginary. But the FD layer lines may have more Bessel terms as we move to higher resolution shell. The problem of FD is multidimensional unlike the two dimensional problem of crystallography. The number of significant Bessel terms (N) contributing to the diffraction intensity depends on the symmetry and dimensions of the diffracting particle and on the values of (R, l) (see Eq. 5). In FD, e.g. TMV, 10 Å resolution data can contribute to 1 Bessel term, but 2.9 Å data can contribute to 8 Bessel terms.

In crystallography the difference map is generally $2F_o - F_c$. The fiber equivalent of $2F_o$-F_c is $6\mathcal{G}_o - 5\mathcal{G}_c$ (Namba & Stubbs, 1987b). Although the difference maps described is satisfactory for most applications, it is sometimes desirable to minimize any possible bias towards a model structure. In recent years, omit maps have become popular as a means to eliminate model bias. Omit maps are calculated from observed structure factor amplitudes and calculated phases, but for phase calculation, the part of the structure under investigation is omitted. A series of map sections of the unit-cell can be systematically omitted (Artymiuk & Blake, 1981). Alternatively, parts of the model structure can be omitted e.g. at a time three amino acid residues can be omitted (Furey et al., 1986). A similar approach was taken in a FD study where electron density corresponding to amino acid side chains suspected of

changing conformation between two different forms of TMV protein were omitted (Mandelkow et al., 1981). In some cases omitted density may return or included density may disappear. In other cases, however, the noise level may be too high to allow unambiguous interpretations to be made. The size of the omitted structure has considerable bearing on the interpretability of omit maps. This is particularly true with FD data because the ratio of model observations to diffracted data observations is much higher than in crystallography. For very small omissions, such as a single side chain we may obtain satisfactory results but with larger omissions there is a significant loss of interpretability. It therefore appears that while omit maps can be of value to FD in answering questions about small regions of a molecule, they are not suitable for systematic examination of complete structures.

1.6 FIBER DIFFRACTION IN MACROMOLECULAR STRUCTURE DETERMINATION

Since the development of the theory of FD (Cochran et al., 1952a), it has been widely used for structure determination of helical aggregate forming biomolecules, notably synthetic polypeptides (Cochran & Crick, 1952b), deoxyribonucleic acid (Wilkins et al., 1953), TMV (Namba et al., 1989) and collagen (Cohen & Bear, 1953; Cowan et al., 1953).

1.6.1 Filamentous plant viruses

Filamentous plant viruses make up almost half of the plant virus genera. The *Potyvirus* genus alone has been described as including almost a third of known plant viruses (Riechmann et al., 1992) and is responsible for more than half viral crop damage in the world. A single *Potexvirus*, Potato virus X (PVX) destroys world

potato crop by 20% (White et al., 1994). Filamentous plant viruses can be grouped broadly into rigid (rod-shaped) and flexible viruses. The International Committee on Taxonomy of Viruses currently recognizes 8 genera of rigid filamentous plant viruses (type member, TMV) and 17 flexible species (type member, PVX) (Regenmortel et al., 2000). All existing filamentous plant viruses are RNA viruses that contain a single type of Coat protein (CP) encapsidating a single-stranded RNA molecule in a helical array. In some genera, the genome is divided among two or more RNA molecules in which the virus consists of multiple particles, typically of different lengths e.g. Tobacco rattle virus (TRV), a member of the *Tobravirus*, is a bipartite virus having particles of two lengths encapsidating the two RNA molecules that make up the TRV genome. Some filamentous plant virus genera are morphologically similar to each other at the electron microscopic level but most exhibit large differences in both morphology and chemical structure. A better argument has been presented that most filamentous plant viruses fall into one of two groups the rigid rods or the flexible filaments (Dolja et al., 1991).

Filamentous virus studies and the development of FD methods have always been synergistic. TMV and other *Tobamoviruses* have served as models for FD data processing. It also helps in method development including the method of angular deconvolution and phase determination (Makowski, 1978; Namba & Stubbs, 1985). Isomorphous replacement (IR) was used earlier to obtain radial density distributions of TMV (Caspar, 1956; Franklin, 1956). Later IR was developed to solve the multi-dimensional phase problem in structure determination by FD (Stubbs & Diamond, 1975). Other techniques developed on TMV included layer line splitting (Franklin & Klug, 1955). Methods of structure refinement and evaluation were also developed using TMV. RLS (Stubbs et al., 1986), MD refinement (Wang & Stubbs, 1993), using

the likelihood function as a target (Mu & Makowski, 2000) difference Fourier analysis (Namba & Stubbs, 1987b) and FD R factors (Millane, 1989; Stubbs, 1989) are part of improvement in FD structural study. The filamentous bacteriophage Pf1 has been important in FD method development. Angular deconvolution was first time applied to Pf1 structure (Makowski, 1978). Pf1 has also been important in the development of MD refinement (Gonzalez et al., 1995).

Other filamentous bacteriophages served as model systems for the development of background subtraction methods (Ivanova & Makowski, 1998; Marvin et al., 1987). Techniques for making oriented sols by shearing were originally developed for TMV (Bernal & Fankuchen, 1941; Gregory & Holmes, 1965). The use of magnetic fields for orienting FD specimens was first time applied using Pf1 (Torbet, 1987; Torbet & Maret, 1979). Magnetic orientation is now widely used in FD (Stubbs, 1999; Torbet, 1987).The combination of these two techniques with centrifugation (Ivanova & Makowski, 1998) showed exceptional promise for a number of FD systems including filamentous viruses (Oda et al., 1998; Stubbs et al., 2000; Yamashita et al., 1998b).

1.6.2 *Tobamovirus* structure determination by fiber diffraction

TMV, a rod shaped virus of the genus *Tobamovirus*, was the first virus to be discovered and subjected to structural studies using X-ray FD. Powder diffraction patterns from unoriented virus solutions had been obtained earlier (Wyckoff & Corey, 1936). Later it was shown that TMV can form highly oriented sols where rod shaped particles are oriented to within about 1° of each other (Bawden et al., 1936; Bernal & Fankuchen, 1941). These oriented sols yielded high quality FD patterns with virions aligned about their long axes. Also FD patterns from flexible

virus PVX was obtained although those patterns did not exhibit such a high degree of orientation (Tollin et al., 1980; Wilson & Tollin, 1969). The characteristic spacing of the layer lines in the TMV diffraction pattern and the higher intensity of every third layer line showed that the virion structure was periodically repeating every 69 Å and with an approximate repeat at every 23 Å. It was recognized that the pattern was typical of diffraction from a helical structure and 69 Å repeating unit of TMV must contain $3n + 1$ subunits in three turns of the helix, with n being an integer. It was shown for TMV that n = 16. i.e. 49 subunits in three turns (Franklin & Holmes, 1958). Using heavy atoms and IR, the equatorial diffraction patterns with and without heavy atom were compared for the radial density distribution. The center of the virion was hollow, along the virus axis, with a central hole of about radius of 20 Å and RNA was located about 40 Å from the viral axis. The method was later used for FD (Stubbs & Diamond, 1975) and used to determine the TMV structure (Holmes et al., 1975; Namba et al., 1989).

1.6.3   Other filamentous virus structure by fiber diffraction

The filamentous bacteriophages have been classified on the basis of structure into two classes, I and II (Marvin, 1998; Marvin et al., 1974a; Marvin et al., 1974b). They are morphologically similar at the electron microscopic level and are members of the *Inovirus* genus (family I*noviridae*) (Regenmortel et al., 2000).

A major development in their structure determination was the use of strong magnetic fields to induce the bacteriophage particles to orient parallel to each other (Torbet and Maret, 1979). This technique allows the production of exceptionally high-quality diffraction patterns. IR method was of little use in filamentous bacteriophage studies because most heavy atom compounds induced structural changes in the

virions. However, the simple α-helical structure of the CP (Marvin et al., 1974b) allows models to be built and refined against diffraction data. In this way, the structure of Pf1 was determined at 7 Å resolution (Makowski et al., 1980). The structures of several filamentous bacteriophages from both structural classes were determined later at resolutions as high as 3.0 Å (Welsh et al., 2000). The *Potexviruses* (type member, PVX) are flexible filamentous viruses about 5000 Å in length and 130 Å in diameter (Richardson et al., 1981; Tollin et al., 1980; White et al., 1994). The earliest FD studies of PVX showed that the diffraction patterns of virion have a periodicity of about 33Å (Bernal & Fankuchen, 1941). The FD patterns from *Potexviruses* have been interpreted and the symmetry of these viruses are found to be considerably more variable than rigid *Tobamoviruses* (Tollin et al., 1980). Recently oriented sols of PVX and Papaya mosaic virus have been obtained using techniques combining magnetic fields with centrifugal forces (Yamashita et al., 1988a). *Potexviruses* respond well to magnetic orientation in combination with centrifugation developed by Namba's group. *Tobraviruses* are rigid rod shaped bipartite viruses having particles of various lengths encapsidating two RNA molecules (MacFarlane, 1999; Mathis & Linthorst, 1994). Both types of particle have same CP with 230 Å diameters. On the basis of sequence comparisons with TMV, a model for the CP structure of *Tobraviruses* has been proposed (Goulden et al., 1992).

Diffraction studies have been described for oriented sols of TRV (Finch, 1965) and Pepper ringspot virus formerly known as the Campinas strain of TRV (Tollin & Wilson, 1971). Fiber diffraction study of a *Hordeivirus,* called Barley stripe mosaic virus (BSMV) has been reported (Finch, 1965). BSMV is a rigid rod 1250 Å in length and about 200 Å in diameter. In summary, structural study of a number of filamentous viruses (plant viruses and bacteriophages) is possible. However, crystallization of

filamentous viruses by common approaches for viral structure determination using protein crystallography is very difficult. In very rare cases, isolated CP can be crystallized. The TMV CP has been crystallized and its structure was determined at 2.8 Å resolution (Bloomer et al., 1978) and later at 2.4 Å (Bhyravbhatla et al., 1998). But the natural tendency of CP subunits is to form helical aggregates rather than crystals and attempts to crystallize the CP of other filamentous viruses have failed. In rare cases where crystals have been grown, protein–protein interactions in the crystal do not correspond to biologically significant interactions. In addition, protein-nucleic acid interactions are absent altogether (Bhyravbhatla et al., 1998). Thus, structural studies of filamentous viruses very much rely on the FD method.

CHAPTER 2. HIBISCUS LATENT SINGAPORE VIRUS

2.1   INTRODUCTION

Viruses can infect animals, plants and bacteria. Since first discovery of TMV (Beijerinck, 1898), many viruses have been discovered. Viruses are nucleoprotein complexes with their genetic material as DNA or RNA. The genetic material of a virus is protected by CP structure. Major category viruses are of two shapes, filamentous and icosahedral types. Plant viruses enter the host cells by damage of the host tissue. The cell to cell movement is regulated by movement protein (MP) whereas the long distance movement is regulated by CP.

Using various molecular biology methods, it is now possible to explore the genome organization and expression strategies of different viruses elaborately. This in turn help us to develop and design methods to combat crop losses resulting from viral epidemics in agricultural fields and their exploitation as vectors for expressing therapeutic proteins (Hamamoto et al., 1993; Wu et al., 2003). *Tobamoviruses* are rod-shaped with an approximate length of 3000 Å and a diameter of 180 Å. Its genome is positive-sense single-stranded RNA packed in a capsid of about 2100 CP subunits. This typical packaging forms a right handed helical virion with 49 CP subunits in three helical turns. Among *Tobamoviruses*, TMV is the most widely studied virus and it remains to be useful tools for understanding the fundamental processes of viral infection, replication and movement. *Tobamovirus* genus consists of several species, which can be classified into 2 major sub-groups based on their origin of assembly sequence (OAS). The OAS for the subgroup I and subgroup II located in the MP and CP, respectively. The complete genome sequences of the

various *Tobamoviruses* have been reported (Alonso et al., 1991; Chng et al., 1996; Goelet et al., 1982; Hamamoto et al., 1993; Heinze et al., 2006; Ikeda et al., 1993; Lartey et al., 1995; Meshi et al., 1981; Min et al., 2009; Rhie et al., 2007; Silver et al., 1996; Solis & Garcia-Arenal, 1990; Song et al., 2006; Srinivasan et al., 2002; Tan et al., 2000; Ugaki et al., 1991; Yoon et al., 2001; Yoon et al., 2002; Zhang et al., 2008).

2.2   HIBISCUS LATENT SINGAPORE VIRUS (HLSV)

Hibiscus latent Singapore virus (HLSV) is a newly discovered member of subgroup II *Tobamovirus*. It is a positive sense RNA virus comprising 6,474 nt (Genbank Accession No. NC 008310; Srinivasan et al., 2005). The helix has a repeat of 70.5 Å (helix pitch 23.5 Å), which is close to the 70.8 Å of CGMMV (Wang & Stubbs, 1994). Two other viruses of subgroup II are Cucumber green mottle mosaic virus (CGMMV) and Sun-hemp mosaic virus (SHMV). HLSV differs from CGMMV and SHMV in containing a 77-96 poly(A) tract at the 3′ untranslated region (UTR). The following sections give a preview of HLSV characterization.

2.2.1   General characterization of HLSV

The single molecule genome of HLSV is packed with 2,100 CP molecules (3 nucleotides per CP). The virus is 180 Å in diameter and the CP molecules are arranged helically to give a rigid rod shape. HLSV also makes two sub-genomic RNA for MP and CP (Srinivasan et al., 2002). HLSV induces chlorotic local lesions in *C. quinoa* and systemic infection in *Nicotiana benthamiana* (Srinivasan et al., 2002). HLSV host range is relatively limited as compared to TMV. HLSV forms particles of two length sizes, i.e. 34 nm and 307 nm during its life cycle of infection in plants

(Srinivasan et al., 2002). The longer particle packs the full length RNA genome HLSV. The shorter particles may contain its CP sub-genomic RNA, as reported for SHMV (Higgins et al., 1976).

2.2.2 HLSV gene structure, regulation and proteins function

The HLSV genome contains 5′ UTR, methyl transferase-helicase (128 kDa), RNA dependent RNA polymerase (RdRp, 186 kDa), movement protein (MP, 30 kDa), coat protein (CP, 18 kDa) and a 3′ UTR (Fig. 2.1). The 5′ UTR contains 58 nucleotides and is predicted to form a stem-loop structure. The (CAA)n repeat sequence of 5′ UTR is quite similar to the TMV 5′ UTR (CAA)n repeat sequence and it act as a translational enhancer (Gallie et al., 1987a and 1987b). The 3′ UTR poly(A) tract of 77-96 nt is followed by a t-RNA like structure. During *Tobamovirus* replication in plants, RdRp, MP and CP play an important role of virus replication, cell to cell movement and long distance movement, respectively (Asurmendi et al., 2004; Fujiki et al., 2006; Yamaji et al., 2006).
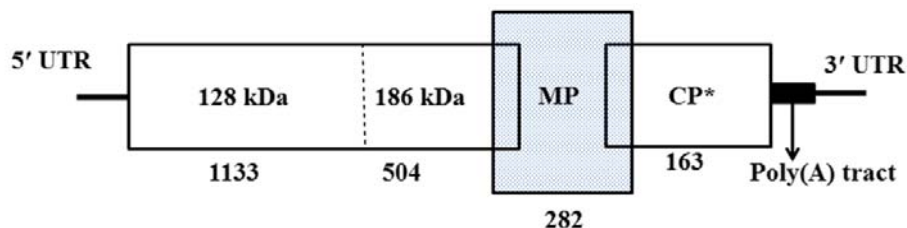
Figure 2.1. Genome organization of HLSV. Boxes represent coding regions for methyl transferase-helicase (128 kDa), RdRp (186 kDa), MP (30 kDa) and CP (18 kDa) protein with their amino acid lengths indicated. Dashed vertical line indicates amber stop codon for the RdRp. Asterisk in CP region indicates location of predicted OAS. Untranslated regions at the 5′and 3′ ends are represented by horizontal lines. Poly(A) tract is represented by solid box. Length of 3′ UTR is variable due to the variable length of poly(A) tract.

2.3   PREVIOUS STUDIES ON *TOBAMOVIRUS* STRUCTURES

Several *Tobamovirus* structures have been reported. The structure of TMV (Holmes et al., 1975; Namba et al., 1989), CGMMV (Wang & Stubbs, 1994), RMV (Wang et al., 1997) and U2 strain of TMV (Pattanayek & Stubbs, 1992) have already been solved by X-ray FD. Surprisingly, the TMV CP structure was solved at 2.4 Å resolution by X-ray crystallography method (Bhyravbhatla et al., 1998). But for other *Tobamovirus* structures X-ray crsyatllography approach was not successful. All other *Tobamovirus* structures are determined by X-ray FD.

2.3.1   Infection and stability of native virus capsid

For *Tobamoviruses*, it has been shown that negatively charged amino acid residues in the CP from different molecules are juxtaposed at subunit interfaces, at the low (proximal to virus axis) and high radius regions (distal to virus axis). This creates an electrostatic potential that is believed to drive disassembly and thus initiate the early stages of viral infection (Caspar, 1963; Lu et al., 1996; Namba et al., 1989). The carboxyl-carboxylate pair in the low radius region and phosphate-carboxylate pair also appear to bind calcium (Ca^{2+}). Electrostatic interactions have been recognized to be an important factor in the disassembly of helical and spherical plant viruses (Bancroft, 1970; Caspar, 1963). On the basis of titrations of isolated TMV CP under various conditions, anomalously titrating amino acids was suggested to exist in subunit interfaces (Shalaby & Lauffer, 1977). It was also shown that TMV has two sites which compete for calcium and protons ions (Gallagher & Lauffer, 1983a; Gallagher & Lauffer, 1983b). Ca^{2+} ion binding gives structural stability to many icosahedral viruses (Butler et al., 1977; Olson et al., 1983; Wada et al., 2008).

Structural details of TMV provide some interesting facts regarding molecular basis of assembly in *Tobamoviruses*. A general description of the assembly process of TMV has been described (Butler et al., 1977; Lebeurier et al., 1977). The OAS of the RNA binds to the 20 S aggregate of the CP and elongation of the viral rod proceeds by addition of the 20 S aggregates (Zimmern, 1977). For TMV, during infection, it is believed that low concentration of Ca^{2+} and high pH (relative to extracellular conditions) of the cell could considerably destabilize the close approach of the negative charges in the viral subunit interfaces as described earlier (Namba et al., 1989). However, these conditions are not sufficient for complete disassembly of virus under *in vitro* conditions. It was shown that with pretreatment of TMV at pH 8.0 particles could be dissociated under *in vitro* conditions by a preparation containing ribosomes (Wilson, 1984). This phenomenon, called co-translational disassembly, was later observed also for *in vivo* conditions (Wilson, 1984). This mechanism protects viral genome under unusual alkaline conditions.

In summary, the whole infection mechanism is described as; when virion enters into a plant cell, owing to low concentration of intracellular calcium (Ca^{2+}) and high pH (relative to the extracellular environment), protons and Ca^{2+} ions are removed from carboxyl-carboxylate pairs and phosphate-carboxylate positions. This allows electrostatic repulsive forces (from negative charges of the protein residues) to destabilize the intact virus and hence, disassembly. It has been proposed that protein-nucleic acid interactions involving first 69 nucleotides of 5´ UTR are weaker than the rest of the genome because of the presence of relatively less guanine bases (Douglas & Young, 1998). So, CP subunits forming about 1.5 turns of the virus helix at the 5´ UTR are lost easily. The first start codon is thus exposed and ribosomes bind and move toward the 3´ end during translation, competing with the CP and stripping the

rest of the genome, thereby beginning a new cycle of viral replication (Namba et al., 1989).

2.3.2 Evolutionary insights from virus structures

Three major forces that drive evolution of viruses are mutation, recombination and reassortment. The discovery of the GDD-sequence motif in a wide range of viral polymerases shows that viruses have genes that are related (Argos, 1988; Kamer & Argos, 1984). Koonin & Dolja (1993) and Zanotto and his team (1996) proposed that *Tobamoviruses* share their RNA polymerase genes with other species of a large group of viruses called the `α-like' virus group (Goldbach & De, 1994).

Tobamovirus CPs are also related in sequence and structure to those of other viruses with rod-shaped and filamentous virions. The 3-D architecture of a typical CPs of *Tobamovirus* consists of a bundle of four α-helices which is also a structural fold observed in several other proteins, and shows evidence of having arisen by duplication of a two-helix protein (McLachlan et al., 1980). In *Tobamoviruses*, protein-nucleic acid interactions are more conserved than protein-protein interactions. Firstly, among the 25 residues that are conserved in the *Tobamovirus* sequences, 11 are directly involved in RNA binding (Altschuh et al., 1987; Namba et al., 1989). Secondly, because evolutionary pressures would resist viable mutations in the nucleic acid binding site, mutations in other parts of the structure can be compensated for by complementary mutations in spatially nearby residues.

Another phenomenon, called Caspar carboxylates interaction, is an important feature of many viruses in disassembly process and is evolutionarily migrating among virus species (Wang et al., 1998; Caspar, 1963; Bancroft, 1970). It is seen as interactions between the side chains of specific amino acid residues of the CPs.

Carboxyl-carboxylate interactions are not conserved during evolution (Namba et al., 1989; Wang & Stubbs 1994; Wang et al., 1997). The complex and variable nature of protein-protein and protein-nucleic interactions in *Tobamoviruses* provides information about the relationship between structure and function. Carboxylate interactions allow structural features of inter-subunit interactions to be conserved, even while amino acid sequences are changing. This feature allows virus to change structure but retain function, is evolutionarily important and advantageous. Such evolutionary flexibility helps virus in evading host defensive responses. For example, host factors in tobacco are known to recognize the surface of the viral CP in many *Tobamoviruses* that triggers a hypersensitive response. This response is characterized by the death of cells close to the original site of viral infection thus preventing the virus to spread systemically and the rest of the plant is protected (Culver et al., 1994).

2.3.3 Coat protein interaction with genomic RNA

Of several interactions that stabilize a *Tobamovirus* virion, protein-nucleic acid interactions are very important. The entire genomic RNA of is encapsidated by CP through non base specific protein-nucleic acid interactions. Such interactions are non-directional, ionic and van der Waals forces. There are 4 α-helices in each CP, two radial and two slewed (Left Slewed, Right Slewed, Left Radial, Right Radial) (Champness et al., 1976).

In TMV, nonspecific base binding is achieved by interactions between the base surface and left radial α-helix (Namba et al., 1989). Electrostatic interactions are considered complementary between electrostatic surfaces of the protein and the nucleic acid. The top surface of the helical array of the CP subunits creates a positively charged groove to accommodate the negatively charged phosphate groups

of RNA. TMV CP recognizes RNA with very high selectivity, efficiently encapsidating only its own or closely related RNA (Hirth & Richards, 1981; Meshi et al., 1983; Meshi et al., 1981). It was proposed that TMV assembly is initiated by a specific RNA sequence that includes AAGAAG as part of a sequence $(XXG)_6$ (Zimmern, 1977). It was proposed that this part of the genome has a highly base-paired secondary structure. In TMV and U2 strain, RNA recognition is achieved by a specific hydrogen bonding interaction between Arg122 and a guanine1 (Pattanayek & Stubbs, 1992; Namba et al., 1989). Arg122 is conserved in most *Tobamoviruses*. In CGMMV, guanine1 specificity is achieved by hydrogen bonding with Gln36 and this may be a characteristic of a subgroup II *Tobamovirus* (Wang & Stubbs, 1994). In TMV CP subunits, there is an electrostatic repulsion prevalent between Asp116 and phosphate2. There was electron density evidence for Ca^{2+} ions binding in this region shown (Namba et al., 1989). In general protein-nucleic acid interactions in a virus structure demonstrates an important principle that virus structure must achieve a metastable balance. All protein-nucleic acid interactions in viruses are required for assembly and dis-assembly in response to changes in their environment (Caspar, 1963; Bancroft, 1970).

2.3.4   Maturation processes of *Tobamoviruses*

Structural studies of CP are advantageous in understanding the self-assembly of a virus in terms of protein-protein and protein-nucleic acid interactions. These interactions are responsible for nucleation and subsequent growth of the virus. TMV CP is one of the first fibrous protein assemblies that was successfully crystallized, as a dimer of bilayer disks or also called a four layer aggregate having 68 subunits (each layer having 17 subunits) (Bloomer et al., 1978). This assembly shown a

sedimentation coefficient of 28 S (Bhyravbhatla et al., 1998) and the structure helped in understanding the self-assembly of TMV.

Since then, it was proposed that bilayer disk plays a key role through an initial RNA recognition reaction at a specific OAS, which then induces a structural switching from the cylindrical disk to helical subunit packing array (Bloomer & Butler, 1986; Butler & Durham, 1977; Hirth & Richards, 1981; Okada, 1986; Stubbs, 1984). A major role for the bilayer disk in the growth of TMV assembly has been proposed (Bloomer & Butler, 1986), but this point has been controversial for some time (Hirth & Richards, 1981; Okada, 1986; Schuster et al., 1980). Central to this controversy has been the assumption that the bilayer disk structure, which is seen in crystals as a dimer, also exists in solution and can undergo a direct structural change, resulting in helical packing of subunits. It has been shown that axial inter-subunit contacts in the bilayer disk are totally different from those in the helix (Champness et al., 1976). At pH 7.0, the CP exists as a mixture of 4 S and 20 S sedimentation aggregates. The CP species with a sedimentation coefficient of 20 S has been shown to be involved in formation of large helical protein aggregate (Durham et al., 1971; Schuster et al., 1979) as well as in the nucleation of viral assembly under *in vitro* experiments (Shire et al., 1979). The 20 S aggregate is present in low ionic strength solutions, either in equilibrium with 4 S aggregate at pH 7.0 or as metastable aggregate at pH 6.5 (Durham et al., 1971; Butler & Klug, 1971).

The 20 S aggregate from the crystallized four-layer aggregate, was further examined (Raghavendra et al., 1985) by circular dichorism technique. This study revealed that the aggregation assembly observed in the crystal differs from the structure present in the 20 S boundary in solution. As aggregates larger than 34 subunits cannot be formed in the closed cylindrical disk, the authors concluded that

two structures are not the same and that the four-layer aggregate is only one of the many self-assembling aggregates and not the nucleating aggregate of the virus.

2.4   RATIONALE AND OBJECTIVES

During virion assembly, *Tobamovirus* CP recognizes a repeating GXX sequence, which is located in the OAS of a viral RNA (Zimmern, 1977, Meshi et al., 1981, Meshi et al., 1983, Takamstusu et al., 1983). The guanine1 recognition during viral assembly is proposed for TMV, CGMMV and RMV (Namba et al., 1989, Wang and Stubbs, 1994, Wang et al., 1997). Arg122 is conserved in all *Tobamoviruses* except for HLSV and SHMV and it is believed to be involved in guanine1 hydrogen bond formation during virion assembly of TMV and U2 (Namba et al., 1989, Pattanayek and Stubbs, 1992). However, in HLSV and SHMV, the corresponding residue (Arg122) is replaced by histidine. Although histidine is a good hydrogen donor, it is unlikely to form a hydrogen bond with a base, since it does not have the flexible extended side chain to reach the base (Wang and Stubbs, 1994). Nevertheless, guanine specificity is believed to be required in SHMV (Meshi et al., 1981). What is the role of His122 and why is it not conserved as Arg122 during evolution in SHMV and HLSV? The recognition of the guanine1 in CGMMV and RMV is also believed to be dictated by Arg122 of the CP similar to TMV (Wang et al., 1997). If Arg122 is changed to His122, how do HLSV and SHMV CP recognize guanine1?

While carboxyl-carboxylate interactions in *Tobamoviruses* are the main driving force for virus disassembly (Wang et al., 1998), they are not conserved during the *Tobamovirus* evolution (Wang et al., 1997; Wang and Stubbs, 1994; Namba et al., 1989). Does this mean that these interactions in HLSV would be different from other *Tobamoviruses*?

HLSV was chosen for this study is because this virus is a novel *Tobamovirus*. We wanted to compare its structure with *Tobamoviruses* and also to attempt higher resolution for its structure.

By solving the structure of HLSV by X-ray FD, we will be able to have a better understanding of the differences between HLSV and other *Tobamoviruses*. This research may also enhance our knowledge of virus structure at atomic details. By knowing the atomic details of this novel virus we may able to use it in future as a vector to express pathogenic epitopes (to develop vaccine) and express economically important proteins.

CHAPTER 3. MATERIALS AND METHODS

3.1 MOLECULAR BIOLOGY

3.1.1 Cloning the HLSV c-DNA

The pBluescript II KS(+) vector (Stratagene) was used to clone the HLSV full length cDNA of 6474 bp (Cao Shishu, PhD. thesis, Natl. Univ. Singapore, 2007). The bacterial strain *Escherichia coli* DH5α was used for plasmid propagation.

3.1.2 *In vitro* transcript preparation

The pHLSV construct was linearized with *XhoI* and cleaned up using the Qiaquick gel extraction column following manufacturer's protocol (Qiagen). The mMessage mMachine kit (Ambion) was used for generating capped *in vitro* transcripts. The reaction mixture was as follows: 10 µl of 2x NTP/CAP, 2 µl of 10x reaction buffer, 1 µg linearized DNA template, 1µl GTP, 2 µl enzyme mix (T7 polymerase) and nuclease free water to make the reaction volume to 20 µl. The mixture was incubated at 37 °C for 2 h. The transcript was checked on 1.2% agarose gel. The transcript was cleaned using the manufacturer's protocol.

3.2 VIRUS PROPAGATION AND PURIFICATION

3.2.1 Plant inoculation

The full length c-DNA clone of HLSV (see section 3.1.2), linearized with *xhoI* at the 3' end was transcribed using mMessage mMachine® (Ambion). 10 µg of the transcribed RNA was mixed with 50 µl of nuclease free water and inoculated mechanically onto *Nicotiana benthamiana* leaves using carborundum. Later on,

31

HLSV-infected leaves were used as inoculums for re-inoculation to fresh plants. Mechanical inoculation using carborundum was carried out by grinding the infected leaves in 0.1 M borate buffer (pH 7.0), with a mortar and a pestle. *Nicotiana benthamiana* seeds were grown in plant growth room under 16 h light and 8 h dark at 25 °C. Newly germinated seedlings were transplanted into new pots and grown in a plant incubator for three weeks under the same environmental conditions. Plant inoculation was performed mechanically as described above. Subsequently, the inoculated leaves were rinsed with sterile water to remove buffer from the leaf surface. The infected plant was then transferred to a plant growth room and incubated for 4 weeks for virus multiplication. The propagation of virus from plant infection to final virus purification took two months.

3.2.2 Crude extraction of HLSV

HLSV infected *Nicotiana benthamiana* leaves were harvested and stored at -80 °C. Frozen infected leaves were homogenized using a blender in 3 volumes (w/v) of extraction ice cold buffer (0.1 M borate buffer, pH 7.0). Homogenization was carried out 3 times for 30 sec and between every cycle 1 min interval was given to avoid over heating. The homogenate was centrifuged at 7,000 x g for 30 min in a JA14 rotor (Beckman Coulter) at 4 °C. The supernatant was clarified with an equal volume of n-butanol/chloroform (1:1) and again centrifuged at 7,000 x g for 30 min. The supernatant was filtered using Whatman paper. The filtered supernatant was transferred into 94 ml polyallomer thin tubes (resistant to organic chemicals) and centrifuged at 22,000 x g for 2.5 h in a 45Ti rotor (in a Beckman Coulter L-100 xp ultracentrifuge) to pellet the virus. The pellet was resuspended overnight in 0.1 M borate buffer (pH 7.0) containing 5 mM EDTA.

3.2.3   Cesium chloride density gradient centrifugation

The crude virus preparation showed nucleic acid and other contamination from the plant tissue. To overcome this we used cesium chloride density gradient centrifugation. We Prepared 30% (w/v) CsCl solution in milli Q water with 5 mM EDTA, pH 7.0, to prevent virus aggregation. The crude virus sample was layered on the top of the solution in 30 ml ultra-clear centrifuge tube (with 3 mg virus per tube). The sample was centrifuged at 40,000 x g for 16 h at 20 °C in a 70.1Ti rotor (in a Beckman Coulter L-100 xp ultracentrifuge). The virus band was collected with a Pasteur pipette under fluorescent light. The collected virus, containing CsCl, was diluted 10 times using 0.1 M borate buffer containing 5 mM EDTA and transferred to 75 ml polycarbonate tubes and centrifuged at 22,000 x g for 2.5 h at 4 °C in a 45Ti rotor. The white clean virus pellet was resuspended in 20 mM borate buffer containing 5 mM EDTA overnight at 4 °C. The yield and quality of the virus was tested spectrophotometrically for $A_{260/280}$, using an extinction coefficient of 3.3.

3.2.4   Slow speed centrifugation to purify 300 nm long HLSV virion

To partially remove the 34 nm short particles from the long particles of 307 nm, low speed centrifugation was carried out. Initial attempts to separate the two particles at 8,000 x g for 18 h at 4 °C failed as it led to pelleting both virus particles at the bottom of the tube. In later attempts, the centrifugation speed was reduced to 6,000 x g for 18 h at 4 °C. The pellet was resuspended in 20 mM borate buffer (pH 7.0) containing 5 mM EDTA and the supernatant was re-centrifuged at 22,000 x g at 4 °C for 2.5 h to see the remaining visible pellet. A very small pellet was seen at the bottom of the tube. The pellet was resuspended in 20 mM borate buffer (pH 7.0). Both the pellet and supernatant were analyzed by electron microscopy.

3.2.5 Sephacryl 1000 gel filtration

In the next attempt, the Sephacryl S-1000 superfine (GE healthcare) size exclusion chromatography was carried out to separate the two particles. We loaded 30 mg CsCl purified HLSV at each FPLC run, with a flow rate of 0.2 ml/min. The virus was eluted from the column starting from 60th to 110th fractions. Three types of sample pooling were carried out: 60-74, 75-85 and 86-110. The three samples separately pooled and were again subjected to ultracentrifugation at 22,000 x g for 2.5 h at 4 °C to pellet the virus. The transparent virus pellet was resuspended in 20 mM borate buffer (pH 7.0) containing 5 mM EDTA overnight at 4 °C. The samples were stored at 4 °C for oriented sol preparation for X-ray FD study. Electron micrograms were taken on carbon coated copper grid with 2% uranyl acetate. For practical reason we use the peak fraction 75-85 ml pooled fraction for obtaining best sols as it consisted of longer particles only.

3.3 CHARACTERIZATION OF PURIFIED HLSV VIRION

3.3.1 Virus purity and concentration

The purity of the purified HLSV sample was detected first by measuring the $A_{260/280}$. The purified virus was diluted 100 times and the absorbance was recorded at Beckmann UV spectrophotometer.

3.3.2 Transmission electron microscope (TEM)

Copper grids (400-mesh) were prepared by covering them with formvar film a carbon layer was coated using a carbon evaporator. Purified virus was negatively stained with 1 µl of 10 mg/ml bacitracin (to prevent the positive staining of the highly

pure HLSV) and 1 μl 2.0% urine acetate and examined in a Jeol JEM-2010 TEM operated at 100 kV with magnification of 30,000.

3.3.3 Western blot

Virus purity was also checked on SDS-PAGE (Sodium dodecyl sulphate-polyacrylamide gel electrophoresis). 12 % SDS-polyacrylamide separating gel, with 5% stacking gel, was prepared. The samples were treated with 5 μl of 6x loading buffer and boiled at 100 °C for 5 min. Electrophoresis was run at 100 V for 2.5 h. The gel was stained by the Coomassie dye. In addition, Western blot analysis was performed with a HLSV polyclonal antibody (primary antibody) and an anti-rabbit antibody (secondary antibody), following the standard procedures.

3.4 ORIENTED SOL PREPARATION

For FD studies, the most effective specimen should be in a thick gel-like form, known as sol. The sol was prepared in a 0.5 mm quartz capillary tube using previous method (Gregory & Holmes, 1965). A virus pellet was exchanged to a suitable buffer (5 mM EDTA, pH 7.0) and the concentration was kept at 30 mg/ml. The resuspended virus was centrifuged at 10,000 $x\ g$ for 16 h at 4 °C in a 1.5 ml microfuge tube. The supernatant was pipetted out and the surface of the pellet was dried with the corner of a tissue paper. The virus pellet was mounted in thin 0.7 mm quartz capillary (Charles Supper) that was washed with 1 M HCl, reverse osmosis (RO) water and 5 mM EDTA, each for four times. During sol preparation, the wide mouth of the capillary was first connected with tubing (1/8", Nalgene). The thin sealed end was broken with a capillary cutter and the broken end was dipped into the virus pellet and the virus was drawn to a distance of 3 cm by withdrawing air by mouth through the PVC tube. The

virus column was pipetted back and forth several times gently for giving better orientation to the virus. Three µl 1 M KCl, which would increase the fluidity of the sample in the capillary, was used. The column of virus was drawn to the end of the capillary and the end was washed by alternate siphoning and blowing. This process was continued until the end of the capillary was completely clean. The capillary was cut at the broad end side also. The quality of the oriented sol was tested using a polarization microscope before sealing it. Well oriented sols show complete extinction (Yamashita et al., 1998a) in a polarizing microscope. Both ends were sealed using an oxygen flame torch. The sealed capillary, with the oriented sol, was kept upright in plasticine (Nalgene).

3.5   FIBER STRUCTURE DETERMINATION

3.5.1   Data collection

Unlike X-ray crystallography, in FD, two to five frames of data are enough for structure determination. Furthermore, importance of humidity in fiber sample preparation and significance of the chambers in controlling humidity have been demonstrated for filamentous plant viruses, yeast prions and *Potyviruses* (Hull, 2001; Wickner et al., 2007; Lo´pez-Moya & Garci´a, 1999). The oriented sols, in which the protein was obviously fully hydrated, gave rise to the strong meridional diffracted intensity at 4.7 Å resolution, characteristic of amyloid structures, but without diffracted intensity on the equator (McDonald et al., 2008).

3.5.2   Data processing

In X-ray crystallography, diffraction by a crystal produces reflections with designated lattice indices. The intensity of a reflection is related to all the atoms,

through FT. However, FD produces layer lines with intensity continuum. The intensity distribution function in FD is not related in a simple manner to the distribution of particle orientations. In addition, there are several factors that contribute to inaccuracies of intensity. Even for a well oriented sol specimen, there remains some degree of disorientation (Namba et al., 1989; Yamashita et al., 1998). There are other factors that give background to FD data which include X-ray scattering from the air, sample holder and nature of the solvent. With higher diffraction angle for a fiber specimen, the overlap of lines is very prominent and hence the intensity of a particular layer line contains data from neighbor layer lines. In FD, the data that are recorded on a detector represent real space data and must be converted to reciprocal space data. We used WCEN for data processing (Bian et al., 2006). In data reduction, any point on any layer line can be denoted by a vector with radius 'r' and angle 'θ' with respect to a reference horizontal axis, Fig. 3.1.
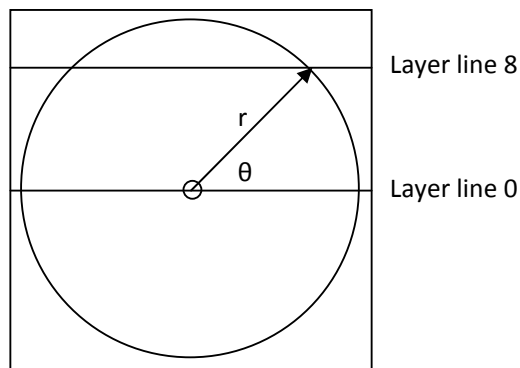
Figure. 3.1. Fiber diffraction data representation by WCEN. A point on any layer line can be represented by a radius of length 'r', positioned at angle 'θ' with respect to a horizontal axis.

The resolution d (in Å) at any point on the detector in a diffraction experimental setup is given by the equation,

$$d= \lambda/2 \sin[1/2 \tan^{-1}y/D] \qquad (11)$$

Where λ is the wavelength of X-rays (in Å), y is the distance of the point of interest from the center of the detector (in mm) and D is the distance between the specimen and detector (in mm). In WCEN, the maximum number of radius is 1200 and the raster value (or pixel size) for the ADSC Q315 detector is set as 0.1026 mm. Thus, in WCEN, the maximum value for y is given by,

$$y = 1200 \text{ x } 0.1026 = 123.12 \text{ mm} \qquad (12)$$

In the above equation (12), if y = 123.12 mm, λ = 0.9002 Å and D = 252.8 mm, then the maximum resolution (d) for an experiment will be 2.0 Å. It is worth noting here that even though the maximum y for an ADSC Q315 detector, according to the manufacturer, is 3072 pixel x 51 μm per pixel = 156.672 mm, WCEN takes only upto an effective maximum of 123.12 mm. Thus the raster value of WCEN is pre-set to match a maximum resolution of 2.0 Å and correlated to the specimen to detector distance of 252.8 mm.

When WCEN reads a diffraction frame, the entire diffraction image is first augmented into 4 equivalent quadrants. Depending on a user's choice, the 4 quadrants are first averaged or only one quadrant (by default the first quadrant, right and up) is taken. Also, the program is designed in a way that the effective detector half width (from the center to the edge) is divided into 1200 radii and each radius sweeps across the detector quadrant area from 0 to 90°. This means, the 1st quadrant area will be marked by 91 x 1200 data points. Thus the output of WCEN is a matrix of 91 x 1200 data points and all inter layer line points should have a theoretical intensity of 0. By the rule of inverse proportion, for our experimental setup, the relationship between a

radius number N and its corresponding resolution of diffraction is given by the equation

$$\text{Resolution} = (1200 \,/\, N) \times 2.0 \qquad (13)$$

The above equation (13) stems from the fact that the width between two adjacent radii is conveniently expressed as $0.001 \,\text{Å}^{-1}$ in the reciprocal space. This helps in the calculation of the resolution for any radius. For example, the resolution of radius 800 can be calculated as:

$$= (1200/800) \times 2 = 3 \,\text{Å}$$

The output of WCEN was corrected for angular convolution, a process known as deconvolution. The angular deconvolution method estimates the intensity from partially oriented specimens to the highest possible resolution. At higher diffraction angle noise will be large, compared to the intensity distribution at low resolution limits. The major problem in structural studies of oriented non-crystalline specimen is the insufficient number of data due to the cylindrical averaging of the intensities. The angular deconvolution of data allows collection of maximum amount of information from a given diffraction pattern. It defines the limiting resolution to which accurate intensity information can be obtained. We used the program RAD (Makoswski, 1978) for deconvolution. The output of RAD is further formatted by the program CORR (Makoswski, 1978), which presents FD data in the format layer line number, minimum radius, maximum radius (theoretically in the range 0 to the maximum radius allowed for the resolution of the data, Eq. 13) for structure determination and refinement.

3.5.3   Layer line splitting

Following earlier discussion in section 1.2.1, fiber particles are naturally disoriented. Due to this anomaly layer lines normally overlap with each other. So, each layer line may be associated with more than one Bessel term (section 1.5) and unlike the crystallographic counterpart, a phase is multidimensional in FD (section 1.3.1). The intensity at any point on a layer line is made of Fourier-Bessel structure factors. To solve a structure, contributions from each Bessel term must be determined precisely. Splitting of the layer line simply means a Bessel function does not exactly fall at the same distance from the equator (Franklin & Klug, 1955). This is generally the case as the repeat distance in one repeat (the number of helical turns) is not exactly repeated along a fiber specimen (Stubbs & Makowski, 1982). The degree of layer line splitting can be determined from exact axial repeat. In practice, due to disorientation of the particles in the specimen, splitting is not directly observed but the effect of layer line splitting causes shifts in the apparent position of the layer line which depend on the relative magnitude of the contributing Bessel terms. Nonetheless the splitting is very informative and can be a source of information about phase. This can be used to separate the contributions from the Bessel function which are superimposed on layer lines. It has been shown that shifts in layer line positions can be measured by angular deconvolution which can determine the contributions from two Bessel function terms that are superimposed on a layer line (Makowski, 1978).

3.6   HLSV STRUCTURE DETERMINATION

The structure of HLSV was solved using the molecular replacement (MR) method. Several frames (from different well oriented sols) of diffraction data at 0, 8 and 12° tilt of the specimen were collected. HLSV fiber sample diffracted X-rays to 2.8 Å. The diffraction data were processed using WCEN. Even though we could

extract data up to the 25th layer line (2.8 Å resolution), the effective data that could be used for refinement was limited to 3.5 Å (upto the 20th layer line). At shells higher than this resolution, layer lines smear/overlap heavily. Near meridian reflections were used for refining the tilt angle whereas other parameters were refined using well defined reflections, distributed throughout the diffraction image. The data were transformed to the polar reciprocal space with angular transformation and further were deconvoluted using RADFN which determines intensity along layer lines (Makowski, 1978). The input parameters for running RADFN are: specimen to detector distance (D) = 252.812 mm; helical repeat distance (C) = 70.51 Å; sigma G (SIGG), the disorientation limit = 1.3°; error limit (ELIMIT) = 1; maximum number of radii processed (NR) = 900; first radius to be processed (from the beam stop, IR1) = 50; background type (NBKGD) = 1; the switch value to rotate the pattern by 90 degree (IFV) = 0; number of allowed bad points for a radius (NULB) = 0; reciprocal space or film space (IFREC) = 0; detector resolution (the raster value, as defined by WCEN) (DR) = 0.1026. The output of RADFN was further formatted by CORR for structure determination and refinement.

A HLSV model was built based on the CGMMV structure (PDB code: 1cgm), but, the R-factor did not improve below 18.4%. The R factor in FD is inherently lower than that of protein X-ray crystallography because of the cylindrical averaging of diffracted intensities (Stubbs, 1989). Usually, the R-factor for a well refined FD structure will be in the range of 9.5%. The HLSV structure was finally determined at 3.5 Å using TMV as a model (PDB code: 2tmv). A trinucleotide (GAA) for RNA was used as discussed previously (Namba et al., 1989; Pattanayek & Stubbs, 1992) for other *Tobamovirus* structures. To model CP inter-subunit interactions, the coordinates of 6 neighboring subunits (+1, +16, +17, -1, -16, -17) around the helical axis from the

reference subunit (labeled '0' and named as the central subunit hereafter) calculated. These subunit interactions represent and correspond to all interactions in the virion. The blmax (Bessel function limit) value was set at 5. The initial model was refined by energy minimization (minimize) without any FD effective energy. The idealized model was subjected to repeated cycles of molecular dynamics simulated annealing (SA) refinement from an initial temperature of 3000 K to a final temperature of 300 K using the diffraction data between 10.0 and 3.5 Å resolution. The molecular dynamics refinement was followed by temperature factor refinement (B refinement). The model was adjusted using the recommended $6\mathcal{G}_{obs} - 5\mathcal{G}_{calc}$ maps (Wang & Stubbs, 1994; Namba et al., 1989). The $6\mathcal{G}_{obs} - 5\mathcal{G}_{calc}$ maps are the FD equivalent of the F_{obs}-F_{calc} maps used in protein crystallography. The larger coefficients are due to cylindrical averaging of the FD intensity and depend on the size and symmetry of the structure and resolution of the data (Namba & Stubbs, 1987b). The structure was refined by NIH-XPLOR (Schwieters et al, 2003). The model was fit to the map with the O (Jones et al., 1991) and COOT (Emsley & Cowtan, 2004) programs. The refinement process was repeated until the R-factor dropped to about 0.096.

The quality and geometry of the final model was analyzed using PROCHECK (Laskowski et al., 1993). Also, omit maps (again, $6\mathcal{G}_{obs} - 5\mathcal{G}_{calc}$ maps) were calculated by omitting five residues at a time (in order to prevent model bias). The Ramachandran plot, shows some disallowed geometry, due to the low data resolution. All molecular diagrams were prepared with Pymol (DeLano, 2002). The coordinates for the six surrounding were built using a program written by Swaminathan for measuring protein-protein and protein-RNA interaction distances. We used the HELANAL program to compare the LR helix geometry of the *Tobamoviruses* (Bansal et al., 2000). Structure-based sequence alignment was made with the ESPript server

(http://espript.ibcp.fr/ESPript/ESPript/) and r.m.s.d. calculations were made with Superpose (http://wishart.biology.ualberta.ca/SuperPose/).

CHAPTER 4. RESULTS AND DISCUSSION

4.1   SAMPLE PREPARATION

4.1.1   Crude extraction of HLSV from plant tissue

Adequate amount of the HLSV virus was propagated by infecting healthy *N. benthamiana* plants with the full length c-DNA clone of HLSV in pBluescript KS II (+), linearized with xhoI at the 3′ end and transcribed using mMessage mMachine. HLSV was initially purified with the n-butanol/chloroform mixture. A260/280 of this crude HLSV was 1.38 to 1.6, as opposed to that of 1.19 for TMV. This is attributed to contamination by nucleic acids which could have come from the plant. We adopted additional purification steps to eliminate contamination and to obtain more homogeneous samples.

4.1.2   Purification using cesium chloride density gradient centrifugation

The virus was purified by the cesium chloride method. The purified virus formed a clear white band under yellow light (Fig. 4.1A). We could also see nucleic acids and other plant contamination as a higher band. Electron microscopy confirmed that the purified HLSV was of long intact particles (Fig. 4.1B). We also verified its purity by 12% SDS-PAGE and western blotting using an anti-HLSV antibody (Fig. 4.1C, D).

4.1.3   Slow speed centrifugation to purify the long virions

The crude HLSV preparation and the cesium chloride purified preparation contained two particles, the 307 nm long virus and 34 nm long shorter particles. For proper orientation of particles in sol, it is necessary to remove the short particles.

(A)

(B)

(C)

(D)
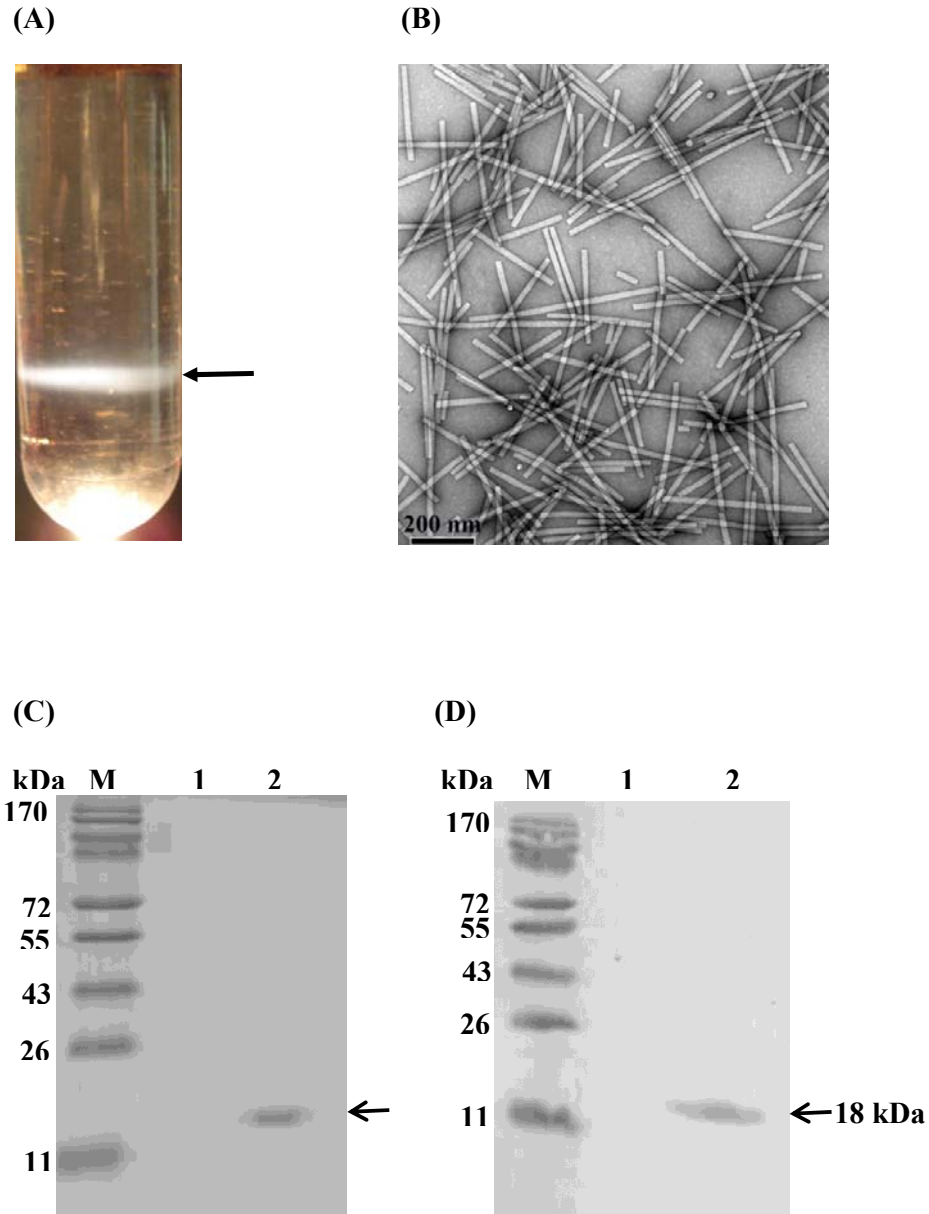
Figure 4.1. Cesium chloride density gradient centrifugation purification and Western blot analysis of purified HLSV. (**A**) Crude HLSV layered on top of 30% (w/v) CsCl solution and centrifuged at 40,000 *x g* for 16 h at 20 °C in a 70.1Ti rotor. HLSV virions (arrow) formed a white visible band. (**B**) CsCl purified HLSV particles were negatively stained with 2% uranyl acetate and viewed in a Jeol JEM-2010. (**C**) 12% SDS-PAGE and (**D**) Western blot analysis of the CsCl purified HLSV using an anti-HLSV antibody, showing a band of 18 kDa in lane 2. Lane M and 1 are a protein molecular weight marker and a negative control.

High homogeneity in fibrous samples is known to give better orientation. As a first step, we slow speed centrifugation was chosen to separate the two particles. Theoretically, at very low centrifugation speed, due to the difference in size and molecular weight of the two particles, the longer particle will pellet down faster and the smaller particle will remain in the supernatant. A clear small pellet was visible at the bottom of the centrifuge tube at the end of centrifugation. The samples from both the pellet and supernatant were analyzed by TEM (Fig. 4.2A, B, respectively). The pellet appeared to be of intact and long particles. However, we were also able to detect a significant amount of the longer and shorter virus particles in the supernatant which did not pellet out from the total 30 mg of virus. If we the centrifugation speed was increased, even the small particles would be pelleted out. Hence, separation of the two particles was not possible by this method. Owing to the low yield and intensive time requirement, gel filtration after the CsCl purification step was used.

(A) **(B)**
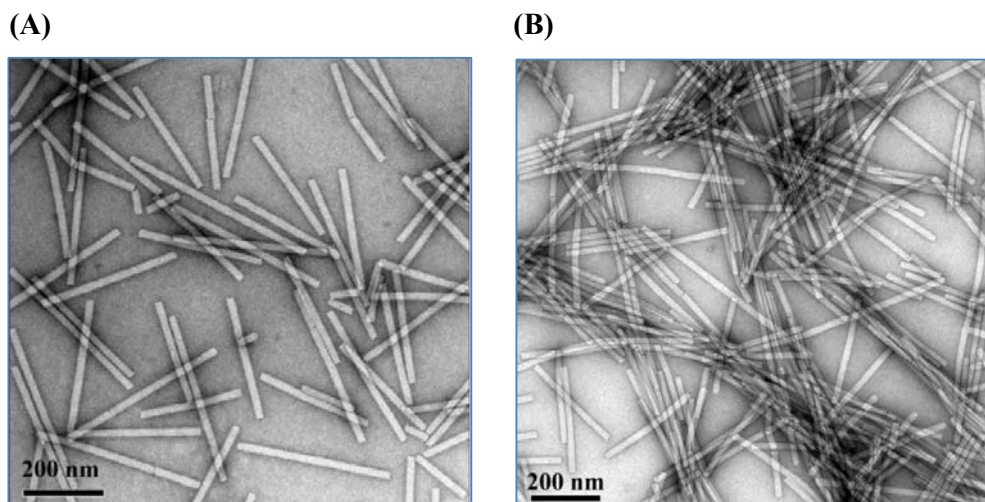
Figure 4.2. **TEM view of the HLSV after slow centrifugation**. (**A**) Pellet fraction and (**B**) supernatant fraction at 6,000 *x g* for 18 h at 4 °C. The virus was negative stained on a carbon coated grid with 2% uranyl acetate.

4.1.4 Sephacryl 1000 gel filtration chromatography

In the next attempt, gel filtration chromatography was run to remove the short HLSV particles. The virus was eluted from the S1000 column as an asymmetrical peak (Fig. 4.3A). The fractions (75-85 ml) that contained intact particles under TEM (Fig. 4.3B) were pooled and later used for the oriented sol of HLSV.

(A)

(B)
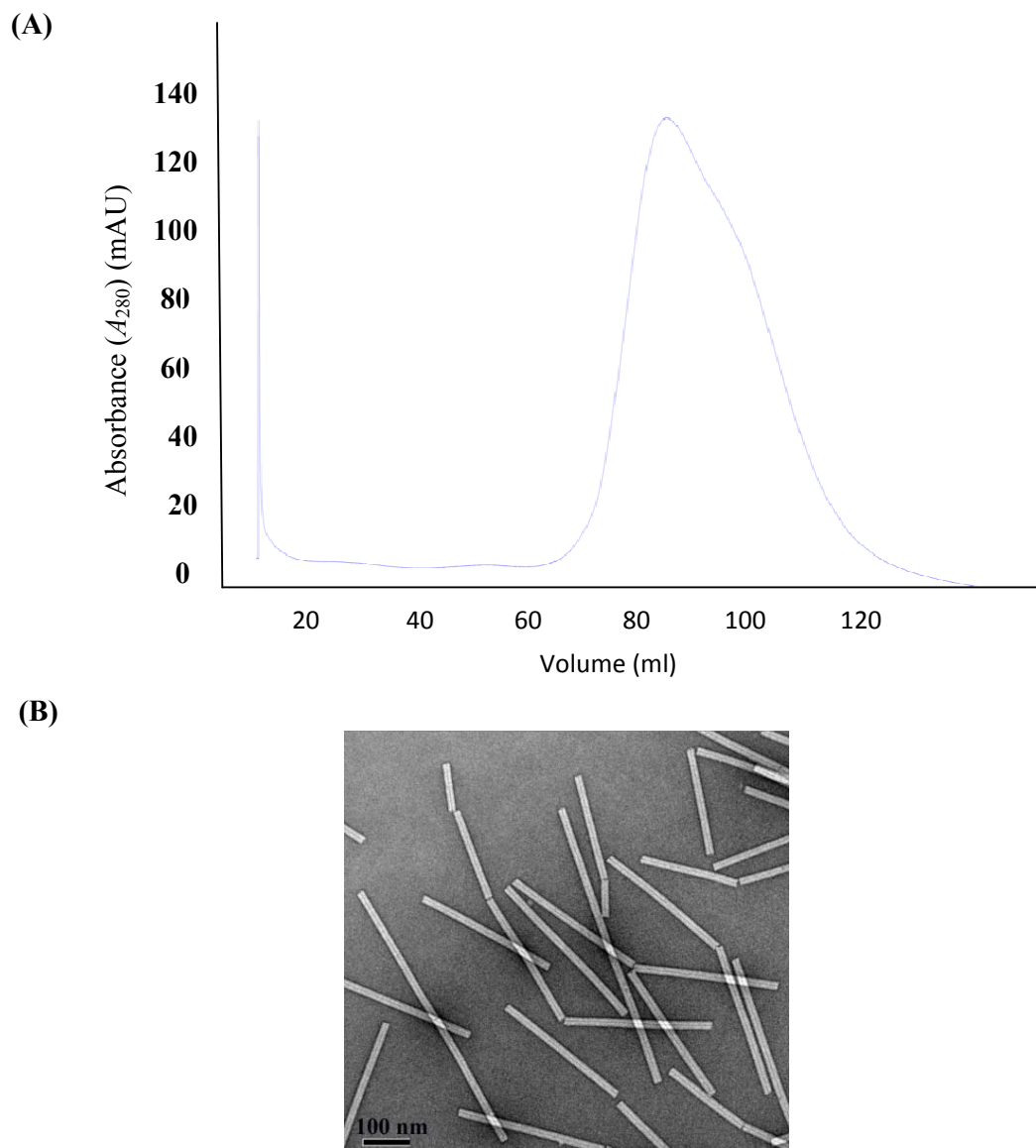
Figure 4.3. HLSV purification by Sephacryl 1000 gel filtration chromatography. (**A**) The chromatogram of HLSV purification showing an asymmetric peak (**B**) TEM of gel filtration purified virus, pooled from fractions 75-85 ml. 30 mg virus was loaded to the column and purified at a flow rate of 0.2 ml/min.

4.1.5 Sol preparation

The gel filtration purified HLSV was used to make oriented sol. The HLSV sol was prepared in a 0.7 quartz capillary (Fig. 4.4A). The quality of the prepared sol was examined by looking the sols under a polarization microscope. We observed that as the time pass the sols reorient themselves to give complete extinction under the polarization microscope (Fig. 4.4B).

(A)

(B)   **(i)**   **(ii)**   **(iii)**
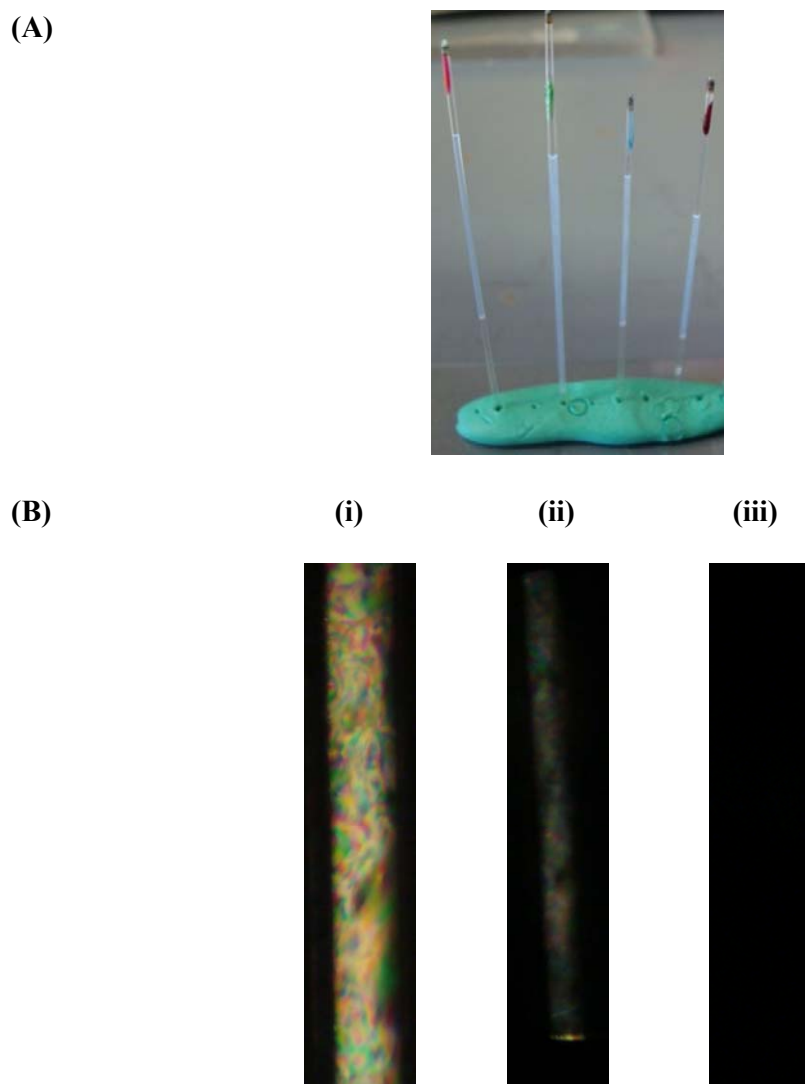
Figure 4.4. HLSV sol preparation and visualization under polarization microscope. (A) HLSV oriented sol in 0.7 mm quartz capillary tubes fixed on plasticine and (B) i, ii and iii visualization of the oriented sol on the day of preparation, 5 days and 10 days of preparation, respectively. The well oriented sol showed complete extinction. The images are shown under different magnification.

4.2   DATA COLLECTION AND STRUCTURE DETERMINATION

4.2.1   Fiber diffraction data collection

The HLSV X-rays fiber diffract to 2.8 Å at the 14 BM-C beam line, Advance

Photon Source, Argonne National Laboratory, USA (Fig. 4.5). The data collection

statistics are given in Table 1.
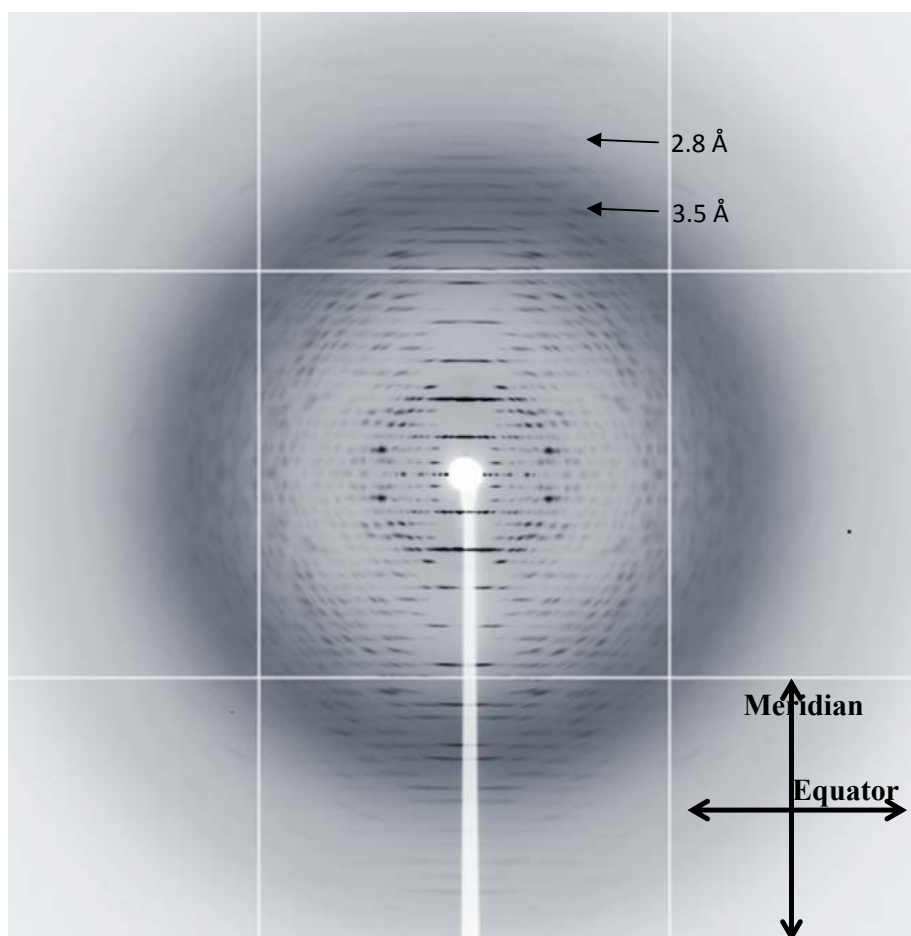
Figure 4.5. X-ray fiber diffraction frame of HLSV oriented sol.
Fiber diffraction pattern from oriented sol of native Hibiscus latent
Singapore virus. Even though the resolution limit along the meridian is
2.8 Å, the effective equatorial resolution is only 3.5 Å. The meridian
and equator are shown in the diagram.

Table 4.1 Data collection, processing, structure determination and refinement statistics.

Synchrotron	APS (USA)
Beamline (BioCARS)	14 BM-C
Detector	ADSC Q315
Raster	0.1026
X-ray Wavelength (Å)	0.9002
Exposure time (sec)	200
Specimen to detector distance (mm)	252.812
Resolution range (Å)	10-3.5
Completeness (%)	99
R	0.096
Protein atoms	1308
RNA atoms	65
Temperature factors ($Å^2$):	
Main chain atoms	27.63
Side chain atoms	25.19
Ramachandran plot (%):	
Residues in most favoured regions [A,B,L]	50.3
Residues in additional allowed regions [a,b,l,p]	35.4
Residues in generously allowed regions [~a,~b,~l,~p]	10.9
Residues in disallowed regions	3.4

4.2.2   Structure determination

The structure of HLSV was determined by X-ray fiber diffraction at 3.5 Å resolution. The refined HLSV model contains 163 amino acids (Fig. 4.6), with a theoretical *pI* of 4.84 and a molecular weight of 18 kDa) and 3 RNA nucleotides (Fig. 4.7A). The final R factor 0.096 is comparable to that of TMV, CGMMV and RMV structures (Namba et al., 1989; Wang & Stubbs, 1994; Wang et al., 1997). There are some geometrical disorders in the final model, owing to the low resolution of the data (Fig. 4.7B). The CP structure is a 4 helical bundle fold. The helices are named LS (left slewed), RS (right slewed), LR (left radial) and RR (right radial) (Champness et al., 1976). In addition, HLSV contains two short helices at the N- and C-termini, Leu10 to Gly15 and Thr142 to Ser148, respectively.
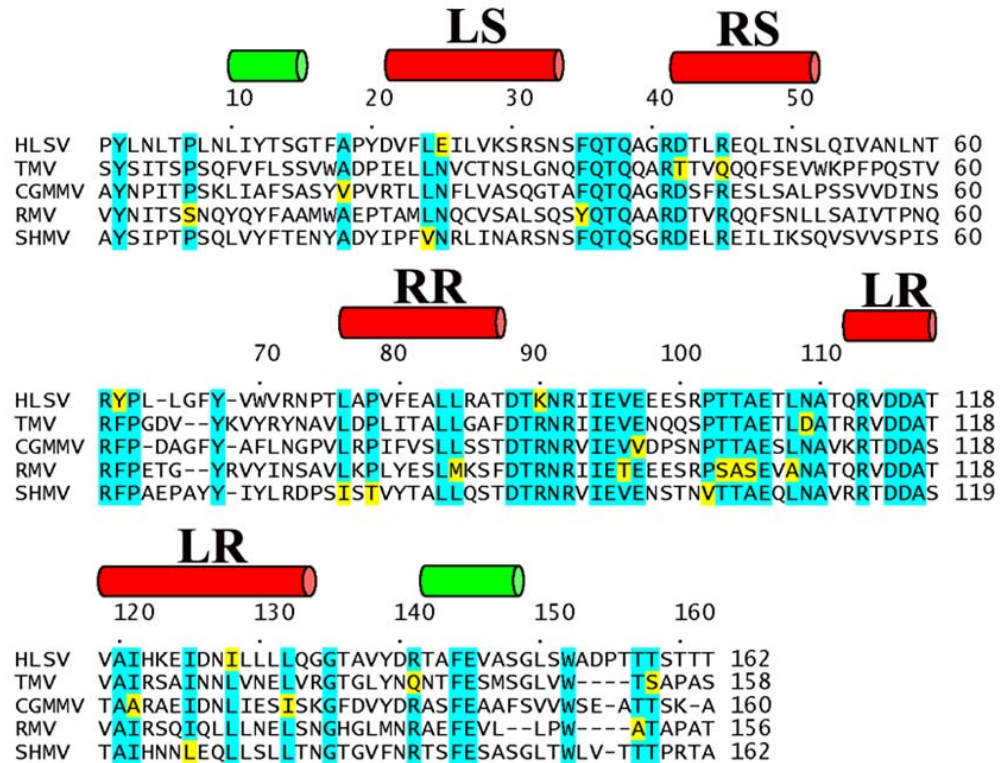
Figure 4.6. CP sequence alignment of HLSV, TMV, CGMMV, RMV and SHMV. The coat protein of HLSV aligns with TMV, RMV, CGMMV and SHMV with a sequence identity of 43%, 44%, 46% and 48%, respectively. Apart from the four core helices (red), there are two short helices (green) at the N- (Leu10-Gly15) and C- (Thr142-Ser148) termini. Fully conserved residues are shaded in blue and a single amino acid difference is shown in yellow. The secondary structure ranges are based on the HLSV structure.

(A)

(B)
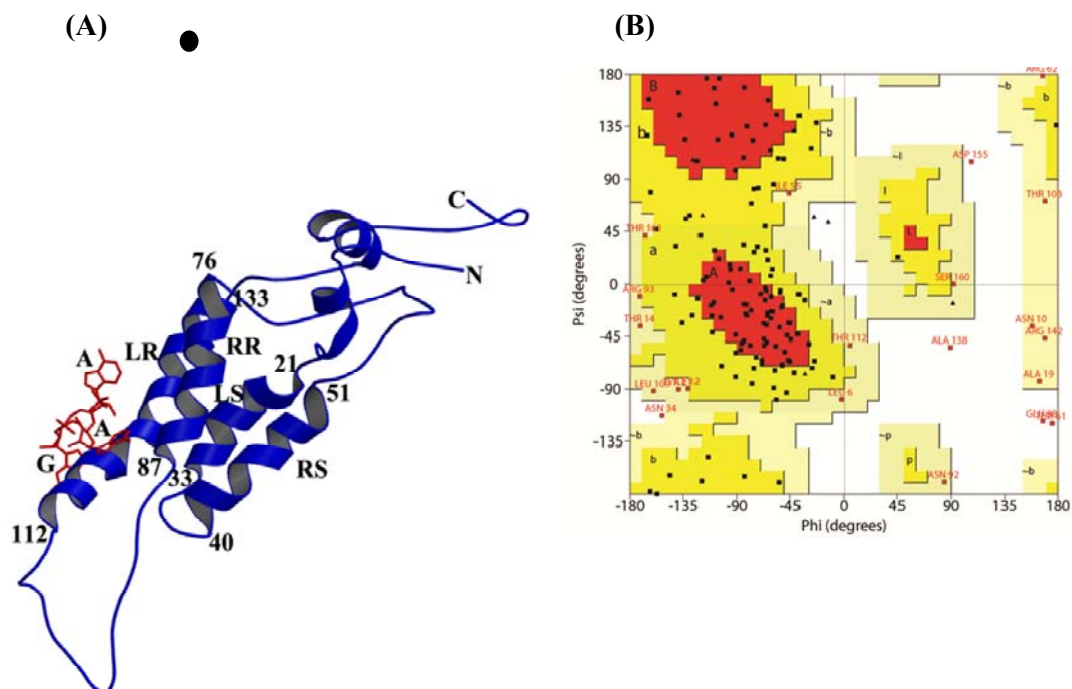
Figure 4.7. HLSV CP structure. (**A**) The ribbon diagram of the central (0th) subunit of the HLSV coat protein (CP), shown in blue. Based on the earlier convention of *Tobamovirus* structure publications, the first methionine is omitted and numbering starts from the second residue (Pro1-Thr162). The α-helices are labeled as left and right slewed (LS, RS) and left and right radial (LR, RR), a nomenclature introduced by Champness et al., 1976 and the corresponding amino acids number are marked. The 3 RNA nucleotides, GAA, are represented as red sticks. The black circle represents the location of the viral axis, which is perpendicular to the plane of the paper. (**B**) The Ramachandran plot of the HLSV CP structure. The residues that occupy the generously allowed and disallowed regions are labeled. The overall statistics is comparable to that of structures at 3.5 Å resolution.

4.2.3   HLSV CP contains a kink in the LR α-helix

The CP structure of HLSV is compared to that of TMV, CGMMV and RMV, (Fig. 4.8). The structure of HLSV CP is similar to that of TMV, CGMMV and RMV, with some noticeable differences.

(A)

(B)

(C)
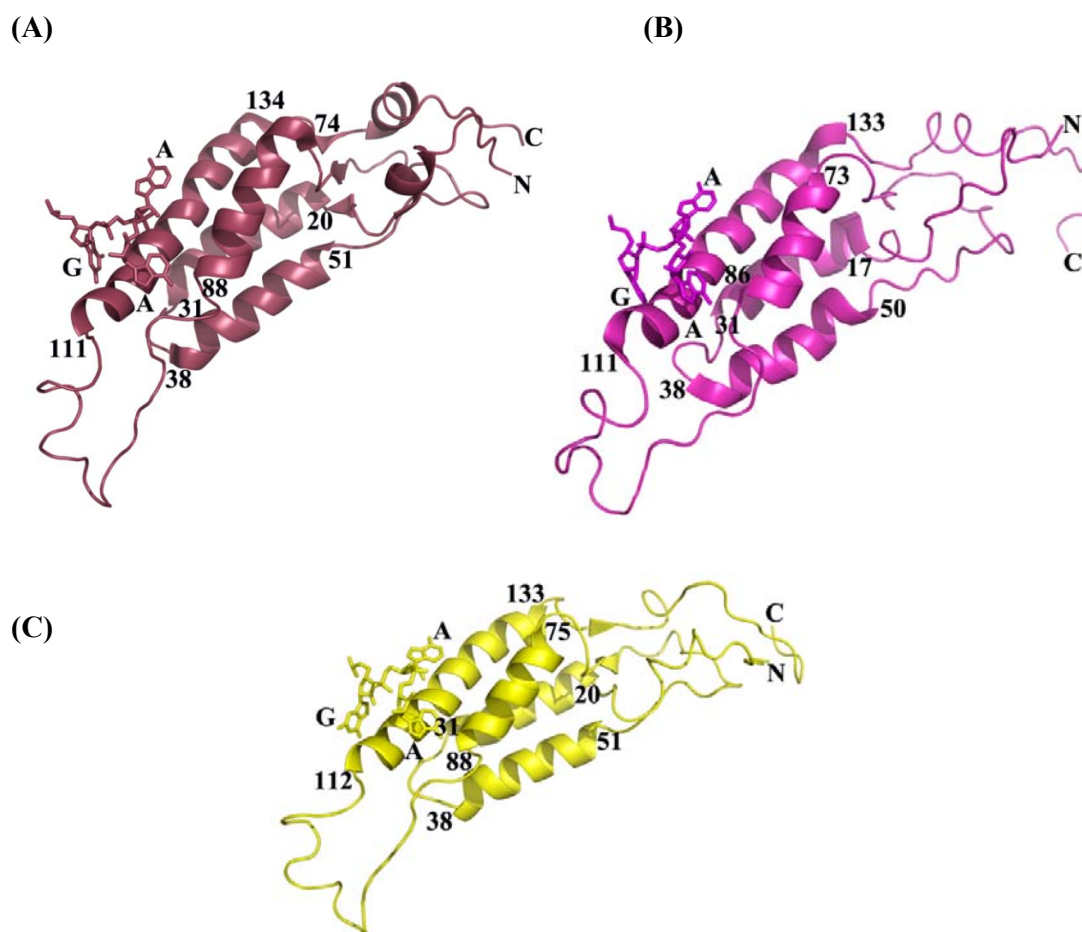
Figure 4.8. Comparison of *Tobamovirus* CP structures. The ribbon diagrams of (**A**) TMV (raspberry), (**B**) CGMMV (magenta) and (**C**) RMV (yellow) CPs. The RNA is shown as sticks in the corresponding color. Note the longest LR α-helix (Thr112 to Leu133) is nearly straight in TMV, CGMMV and RMV, whereas it has a kink in HLSV (Fig. 4.9).

The longest helix, LR, is slightly bent in TMV, CGMMV and RMV, but the bending is more pronounced in HLSV, presumably owing to the presence of His122 (Fig. 4.9). In most protein-nucleic acid structures, an interacting helix shows a bend at the interaction site (Sreekanth et al., 2008). An α-helix may be distorted for several reasons such as the occurrence of a proline (Barlow et al., 1988; Love et al., 1971; Richards and Kundrot, 1980), solvent induced distortions (Blundell et al., 1983) and peptide bond distortions (Chakrabarti et al., 1986; MacArthur and Thornton, 1996).
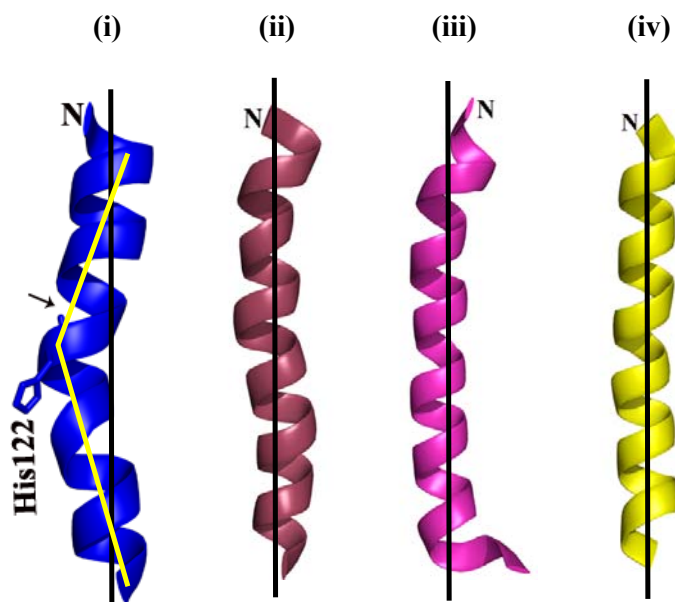
Figure 4.9 Comparison of the LR helix of *Tobamoviruses*. (i) HLSV (blue) **(ii)** TMV (raspberry) **(iii)** CGMMV (magenta) **(iv)** RMV (yellow). The bend in the HLSV LR helix is seen in the region Val119 to His122. This nonproline and nonglycine region of α-helix has a tendency for bending or kink. The kink (46.1°) in HLSV is indicated by an arrow. A straight helical axis is drawn to highlight the level of helical bend. The N terminus and His122 are labeled. His122 is shown as stick diagram.

Due to these distortions, they often become non-linear (Richard and Kundrot, 1980). Most of the α-helices in globular proteins are in fact curved (Barlow et al., 1988; Bernstein et al., 1977). The C-O…N-H distance is sensitive to the inclination of the peptide units with respect to the helix axis (Kumar and Bansal, 1996). In the case of kinked helices, the correlation is strong and at least two hydrogen bonds are broken at or near the site of the kink, which is clearly identified by large bending angles. This finding can be utilized to identify bends and kinks in helices, especially when non-proline or non-glycine residues, like His, Trp, Ile and Ala, are located in the region of a kink (Kumar and Bansal, 1996; Kumar and Bansal, 1998). In fact, the largest kink in a helix (local bending angle 54.4°) is observed in residues Thr60, His61, Trp62, and

Ala63 in the Thr44-Asp74 helix of the fertilization protein lysin (PDB code ILIS) (Shaw et al., 1993).

Using the helix analysis program HELANAL (Bansal et al., 2000), comparisons of the LR helix geometry of HLSV with that of TMV, CGMMV and RMV were obtained. The non-proline, non-glycine zone (residues [117]ATVAIH[122]) of the HLSV LR helix produces a kink. The maximum bending was found between residues Ala120 and Lys123, with a bending angle of 46.1° at His122 (Table 4.2). Lys123 may produce positive charge-charge repulsion with His122, destabilizing the helix further and making the kink more pronounced. The average LR helix bending angle between successive local helix axes for TMV, CGMMV and RMV is between 13.6 and 16.1° (Table 4.3). The location of maximum bending for these viruses is mostly localized at the ends of the helix but for HLSV, it is located in the middle of the helix.

Table 4.2. Properties of the LR helix of *Tobamoviruses*

PDB code	Helix start-end	n	h	Aver BA	Max. BA	Radius		rmsC	rmsL	r^2
3PDM	Q112- Q133	3.60	1.54	25.3	46.1	35	0.49	0.97	0.86	
Std. Dev.		0.32	0.40	12.1						
Geometry: K (at H122)										
2tmv	T111-G135	3.62	1.51	16.1	24.8	39	0.99	0.60	0.93	
Std. Dev.		0.25	0.2	4.9						
Geometry: K (at V130)										
1cgm	A110-G135	3.77	1.54	15.7	36.4	46	0.55	1.38	0.63	
Std. Dev.		0.37	0.4	7.0						
Geometry: K (at G135)										
1rmv	T111-N134	3.69	1.54	13.6	23.3	50	0.35	1.09	0.58	
Std. Dev.		0.27	0.2	6.0						
Geometry: K (at A117)										

n: The average number of residues per turn.

h: Average height per turn (Å).

Aver. BA: Average Bending Angle between successive local helix axes (°).

Max. BA: Maximum Bending Angle between succesive local helix axes (°).

Radius: Radius of least squares circle fitted to the local helix origins (Å).

rmsC: Root Mean Square Deviation for least squares circle fitted to the local helix origins (Å).

rmsL: Root Mean Square Deviation for least squares line fitted to the local helix origins (Å).

r^2: Square of linear correlation coefficient for least squares line fitted to the local helix origins.

Geometry: Overall geometry of the helix, Linear (L), Curved (C), Kinked (K)

Std. Dev.: Standard Deviations of the average parameters for the helix.

Note: The Kink of the helix is mainly decided by the bending angle. If the bending angle is >20° it is regarded as kink in the helix. In case the bending is <20° the helix geometry is decided by the ratio of rmsL/rmsC i.e. if r^2 is >1 it is considered curved; if r^2 is ≥0.8 it is considered linear; if r^2 is ≤0.5 helix is curved; for r^2 between 0.5-0.8 no geometry is assigned.

Superimposition of HLSV with TMV, CGMMV and RMV gives root mean square deviations (RMSD) of 1.51 (154 Cα atoms), 3.77 (159 Cα atoms), and 3.16 Å (155 Cα atoms), respectively (Fig. 4.10). There is also a marked difference at the C-terminus. The C-terminus (residues 155-162) of the HLSV CP forms a loop and points towards the helices, mainly because of the presence of Pro155. This differs from other *Tobamoviruses*. This region is flexible, as indicated by high temperature factors, and makes considerably less contact with the rest of the molecule.

(A)

(B)

(C)
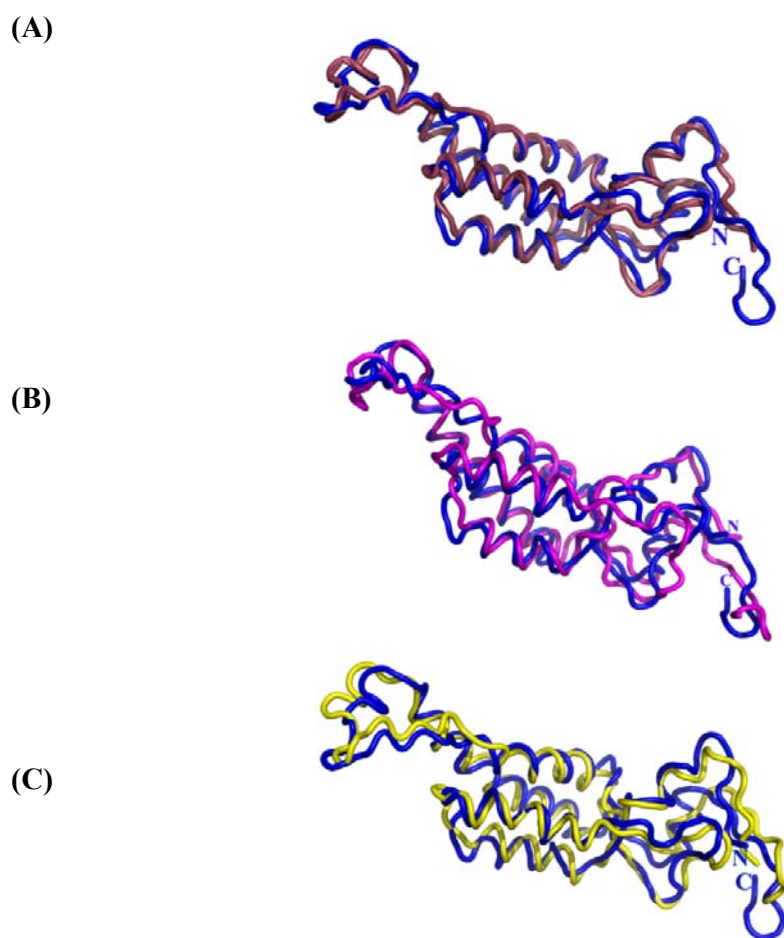
Figure 4.10. Superimposition of the HLSV CP (blue) with that of other *Tobamoviruses* **(A)** TMV (raspberry) **(B)** CGMMV (magenta) and **(C)** RMV (yellow).The RMSD values are 1.51, 3.77 and 3.16 Å for TMV (for 154 Cα atoms), CGMMV (159 Cα atoms) and RMV (155 Cα atoms), respectively.

4.2.4   Nucleic acid structure

The nucleotide triplet GAA, which is used as a representation of the entire genome for *Tobamovirus* structures (Namba et al., 1989), is also modeled in HLSV. The conformation of HLSV RNA is largely similar to that of TMV, CGMMV and RMV (Fig. 4.11), with a slight rearrangement of the bases. Specifically, guanine1 and adenine2 are in the *anti* conformation whereas adenine3 is in *syn* conformation. The ribose rings of guanine1 and adenine3 adopt the C3'-endo and C2'-exo conformations, respectively, while the adenine2 ribose is in the C4'-exo conformation.

(A)

(B)

(C)
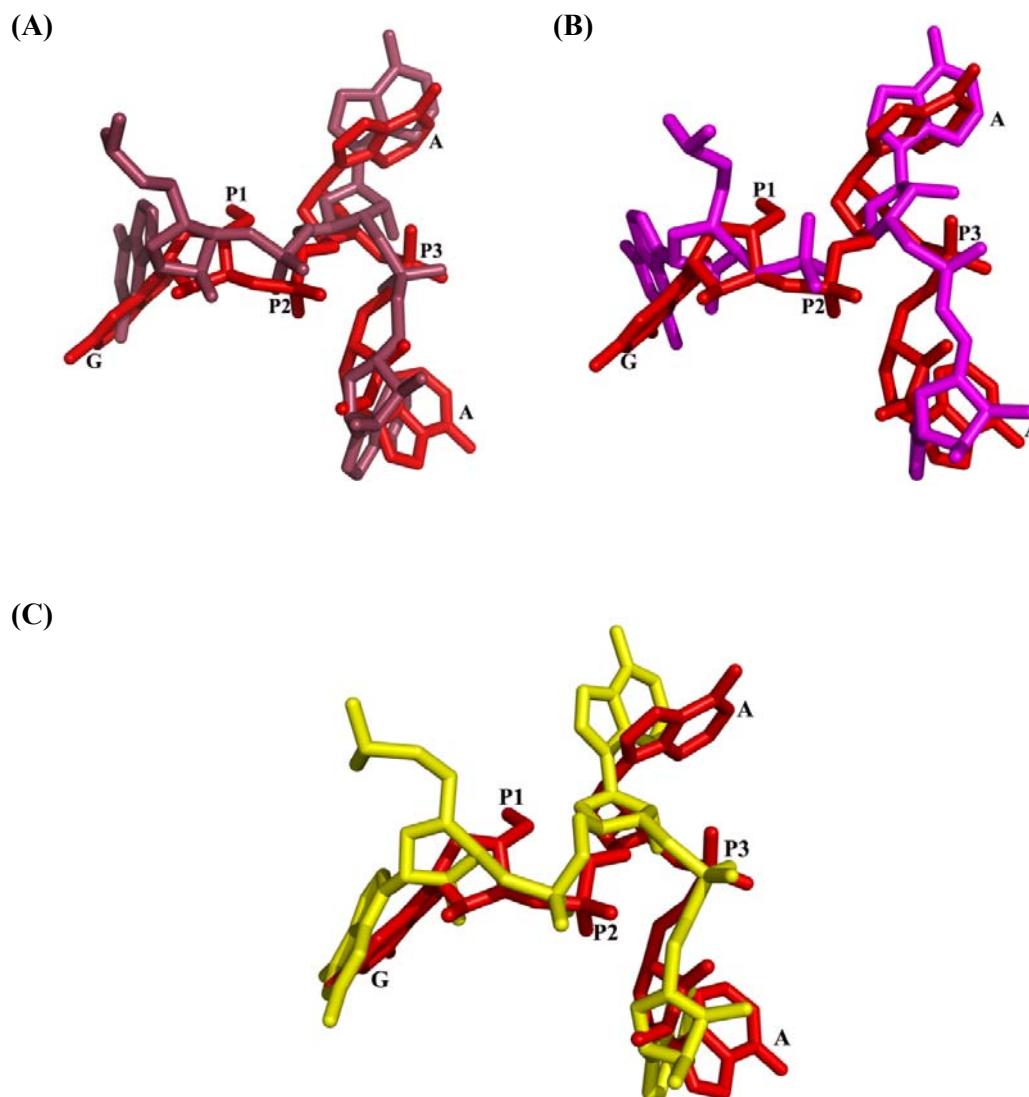
Figure 4.11. Superimposition of the HLSV RNA (red) with that in **(A)** TMV (raspberry) **(B)** CGMMV (magenta) and **(C)** RMV (yellow). Nucleotides 1 (guanine) and 2 (adenine) are in the *anti-* conformation whereas nucleotide 3 (adenine) is in the *syn-* conformation. Note that the ribose rings of nucleotides 1 and 3 adopt the C3'-endo and C2'-exo conformations, respectively while nucleotide 2 is in the C4'-exo conformation.

4.2.5   Protein-protein interaction

Six protein monomers, subunits +1, +16, +17, -1, -16, -17 around the helical axis, interact with a reference protein monomer (labeled '0' or named as 'central subunit' hereafter) as well as among themselves (Fig. 4.12). This arrangement is repeated throughout the virion axis and represents the intermolecular interactions that occur throughout the entire virus structure. The lateral inter-subunit contacts in HLSV are different from those in other *Tobamovirus* structures. The lateral interface between the central subunit and subunit 1 of TMV contains two ion pairs, Arg113-Asp115 and Arg122-Asp88 (Namba et al., 1989). These two salt bridges are present in RMV also (Wang et al., 1997). However, in CGMMV, the Arg113-Asp115 ion pair is present but the Arg122-Asp88 salt bridge is not formed even though these residues are conserved. In addition, CGMMV has an Arg77-Glu130 lateral inter-subunit ion pair, which is not present in TMV (Wang and Stubbs, 1994). In HLSV CP, residues Arg113, Asp115 and Asp88 were conserved. The HLSV CP did not form any salt bridge between Arg113 and Asp115, whereas the His122-Asp88 salt bridge is confirmed with a distance of 2.8 Å from subunit -1 (Fig. 4.13). Another strong salt bridge Arg31-Glu81 (of subunit -1), was also formed in HLSV and the measured distance was 2.3 Å.

(A)

(B)
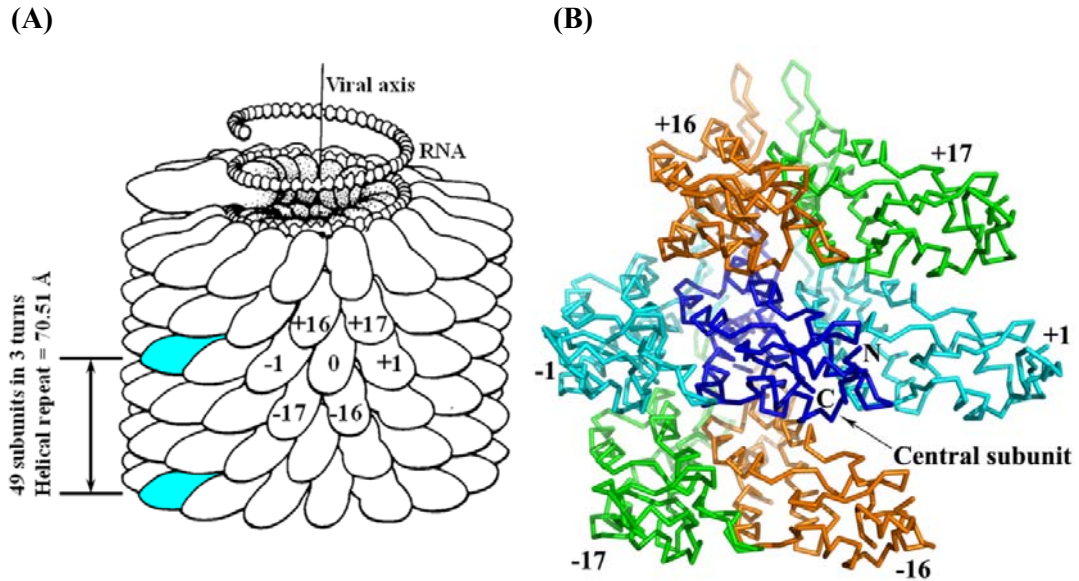
Figure 4.12. Pictorial representation of a partial HLSV virion. (A) The coat protein (CP) molecules form a right handed helical arrangement. The rod-shaped virus has a length of 3000 Å and a diameter of 180 Å. About 2,150 protein molecules pack a genome of about 6,474 RNA nucleotides. The helix has a repeat distance of 70.51 Å and 49 CP subunits in three turns. The virion is stabilized by intersubunit interactions between the central molecule (marked 0) and other subunits marked as +1, +16, +17, -1, -16, -17. The other subunits also interact among themselves. This figure is reproduced with the additional details from: Klug, A and Caspar, DLD. (1961). The structure of small viruses. Advances in Virus Research, **7**, 225-325, with permission from Elsevier B.V. (**B**) Arrangement of the interacting HLSV subunits, following the scheme of panel A. The relative subunit positions are marked. The RNA is omitted for clarity. The subunits +1 and -1 are shown in cyan, subunits +16 and -16 in orange and subunits +17 and -17 in green.
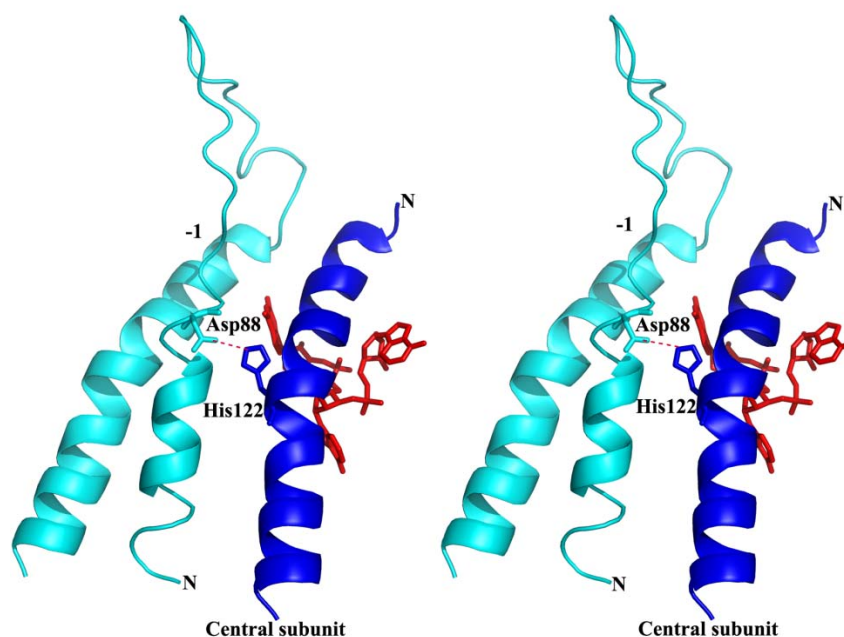
Figure 4.13. HLSV protein-protein (CP) interaction. Hydrogen bonding between His123 and Asp88 (of subunit -1) stabilizes the loop adjacent to RR helix and gives significant stability to the interaction CP at the lateral interface. The N-termini of the subunits are labeled. Asp88 and His122 are shown as sticks. Subunits 0 and -1 are colored blue and cyan, respectively.

Axial interactions of CP subunits are not the same among TMV, CGMMV and RMV (Namba et al., 1989; Wang and Stubbs, 1994; Wang et al., 1997). In TMV, there are two ion pairs in the axial interface between the central subunit and its neighboring subunit 17 along the virion helix: Glu50-Arg134 and Glu95-Arg112 (Namba et al., 1989). CGMMV also has two axial intersubunit ion pairs: Arg122-Asp42 and Lys134-Asp57 from subunit 17 (Wang and Stubbs, 1994). The Arg122-Asp42 contact is also present in RMV (Wang et al., 1997; Wang and Stubbs, 1994). In HLSV, Glu95 is conserved but amino acid residue at position 112 is a glutamine, as in RMV. The Glu50-Arg134 ion pair of TMV is not found in HLSV. Instead, the two axial salt bridges are Arg45-Asp126 and Asp116-Arg92 from subunit +17 and -17, respectively. Apart from salt bridge formation, Arg 92 is involved in protein-RNA interaction in HLSV to stabilize the virus particle.

Carboxyl-carboxylate interactions form the main driving force for virus disassembly in TMV, CGMMV and RMV (Wang et al., 1998; Caspar, 1963). In TMV, there are two carboxyl-carboxylate pairs: high radius, involving axial Glu50-Asp77 (subunit 16) and low radius, involving lateral inter-subunit interactions Glu95/Glu97/Asp109-Glu106 (subunit 1). The influence of carboxylates in viral disassembly has been well studied in TMV (Culver et al., 1995; Lu et al., 1996). In addition, in HLSV, intersubunit carboxyl-carboxylate interactions are also observed (Figs. 4.14, 4.15 and Table 4.3).
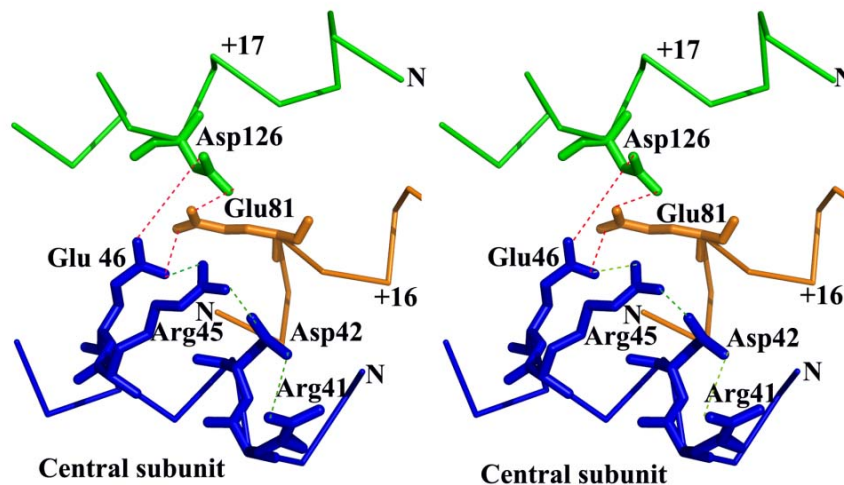
Figure 4.14. Stereo view of HLSV axial intersubunit carboxyl-carboxylate interactions at high radius region. Glu46 with Asp126 (of subunit +17 at 5.1 Å), Glu46 with Glu81 (of subunit +16 at 6.3 Å), and Glu81 with Glu46 and Asp126. Also, some important salt bridges are: Arg45 with Glu46 (4.0 Å) and Asp42 (3.1 Å); Arg41 with Asp42 (at 4.0 Å) which stabilizes the RS helix. The salt bridges are shown in green dashed lines whereas carboxyl-carboxylate interactions are shown in red dashed lines.
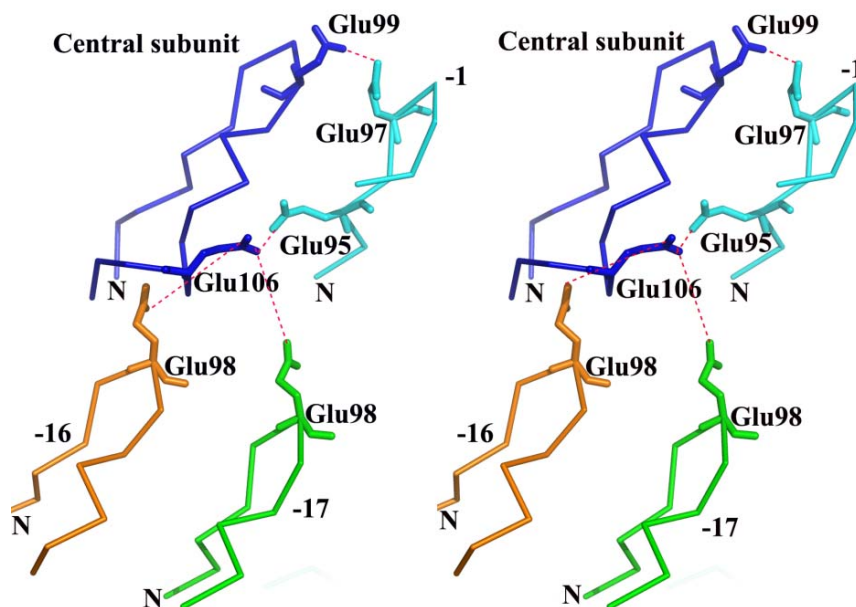
Figure 4.15. **Stereo diagram of HLSV carboxyl-carboxylate interactions at the low radius loop region.** Glu99 with Glu97 (of subunit -1 at 2.7 Å), Glu106 with Glu95 (of subunit -1 at 4.7 Å), Glu106 with Glu98 (of subunit -17 at 6.0 Å) and Glu98 with Glu106 (of subunit -16 at 7.5 Å). The carbonyl-carboxylate interactions are shown in red dashed lines.

Table 4.3. Carboxyl-carboxylate interactions (< 8.0 Å) in *Tobamoviruses*. The subunit number is given in superscripts. For interactions involving 3 residues, distances are not provided.

Virus	High radius Interactions (Å)	Low radius interactions (Å)
HLSV	Glu46-Asp126+17-Glu81[+16]	Glu99-Glu97[-1] (2.7)
	Glu46-Asp126[+17] (5.1)	Glu106-Glu95[-1]-Glu98[-1]Glu98[-16]
	Glu46-Glu81[+16] (6.3)	Glu106-Glu95[-1] (4.7)
		Glu106-Glu98[-17] (6.0)
		Glu106-Glu98[-16] (7.5)
TMV	Glu50-Asp77[16] (4.0)	Glu95-Glu106[1] (5.0)
CGMMV	Glu46-Glu130[17]-Asp126[17]	Glu95-Asp98[1]-Glu106*
	Glu46-Asp126[17] (3.0)	Glu46-Glu130[17] (7.2)
RMV	Glu19-Glu143[16] (6.0)	Glu97-Glu98-Glu99-Glu95-Glu106*

*the interacting subunit number is not provided

4.2.6 **Protein-nucleic acid interaction**

The OAS of HLSV is predicted to be located in the CP ORF (Srinivasan et al., 2002; Srinivasan et al., 2005). It contains GAA repeats, similar to TMV, CGMMV and RMV. The hydrogen bonding interactions between the nucleotide bases and CP at the OAS in HLSV assembly are different from those in TMV and CGMMV, but identical to those in RMV. In CGMMV, guanine1 forms a hydrogen bond with Arg113. Wang and Stubbs (1994) argue that any base at this position would make a hydrogen bond with Arg113 and particularly, a guanine at this position would make an additional hydrogen bond with the side chain oxygen of Gln36. However, in HLSV, Arg113 does not form any hydrogen bond. Instead, there is a hydrogen bond between guanine1 and Asp115, which is also observed in RMV (Wang et al., 1997). This interaction has been proposed to contribute to nucleotide base specificity. In TMV, Arg122 plays an important role in selection and stabilization of guanine1 during virus assembly (Namba et al., 1989). The same was also proposed for CGMMV and RMV (Wang et al., 1997). In HLSV, the residue corresponding to Arg122 is His122 and its adjacent residue Lys123 stabilizes the phosphate of guanine1 (Fig. 4.16). In CGMMV, the charges on the RNA phosphate groups are neutralized by the interactions with neighboring side chains: Arg41 forms a hydrogen bond with phosphate1, Lys112 and Asn91 form hydrogen bonds with phosphate2 and the charge on phosphate3 is partly stabilized by the dipole of the RS helix (with the NH groups of Gln38 and Ala39, both well placed to form hydrogen bonds with the phosphate oxygen atoms) (Wang and Stubbs, 1994). In HLSV, the charges on phosphate2 and phosphate3 are neutralized by Arg92 from the axial position at neighboring subunit -17, which is positioned at a distance of 3.8 Å and 3.7 Å, respectively (Fig. 4.17). Furthermore, Thr37 of subunit -16 makes a hydrogen bond

with N6 of base2 (with a distance of 2.8 Å). In HLSV, His122 does not have any influence over stabilization and recognition of guanine1 interactions. The phosphate-carboxylate interaction of TMV, CGMMV and RMV (Namba et al., 1989; Wang and Stubbs, 1994; Wang et al., 1997) is also observed in HLSV. In HLSV, phosphate2 is 2.9 Å away from the carboxylate group of Asp116 (this residue is conserved among all *Tobamoviruses*) (Fig. 4.18). In TMV, this region is shown as a potential calcium binding site (Namba et al., 1989).
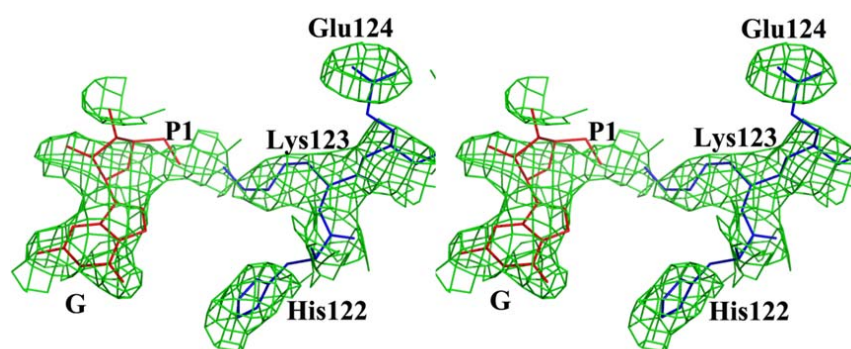
Figure 4.16. **Electron density map of HLSV CP-nucleic acid intrerraction.** Stereo views of the $6\mathcal{G}_{obs} - 5\mathcal{G}_{calc}$ electron density maps showing the Lys123-phosphate1 interaction.

The electron density map (of HLSV CP-nucleic acid) does not show any significant electron density in the vicinity of the Asp116 carboxylate and phosphate groups, since our preparation of virus sols contained EDTA. As a result, EDTA could have sequestered all Ca^{+2} ions. The nucleotide bases of HLSV are flat against the LR helix, similar to TMV, CGMMV and RMV, but with a little shift in their positions. Their faces form hydrophobic interactions with some of the side chain methyl groups. Thr118 is at a distance of 2.8 Å from guanine1 (within the same subunit) and stabilizes it by its methyl group hydrophobic interaction whereas Ala39 and Thr37 are at a distance of 2.8 Å each, respectively, from neighboring subunit -16 and stabilize base2. Guanine1 is stabilized by Val119 within same subunit (at a distance 2.6 Å),

while base3 is stabilized by Ala86, Thr89, and Ala117 at 3.9, 4 and 5 Å within the same subunit. The summary of all important CP-RNA interaction are given in Fig. 4.19.
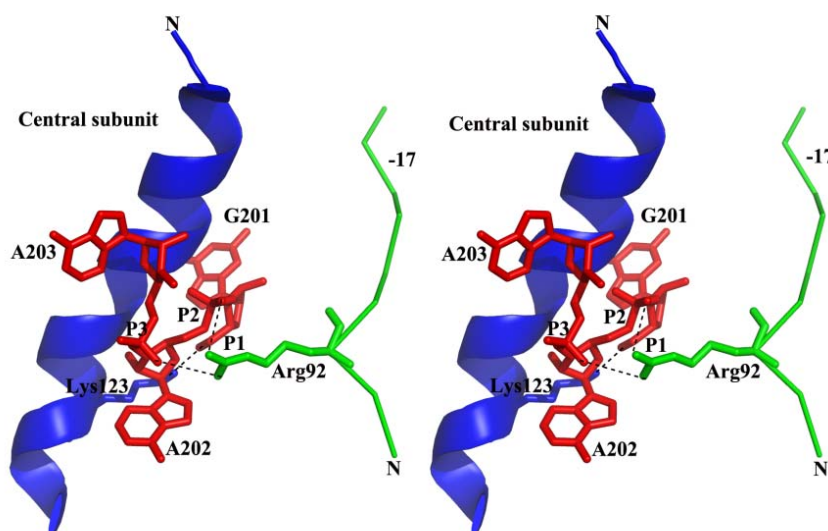
Figure 4.17. HLSV CP-RNA interaction. Arg92 (of subunit -17) makes contacts with phosphate2 (P2) and phosphate3 (P3). Lys123 makes an intramolecular hydrogen bond with phosphate1 (P1). The N-termini of subunits 0 and -17 are labeled. The RNA and Arg92 are shown as sticks and also labeled.
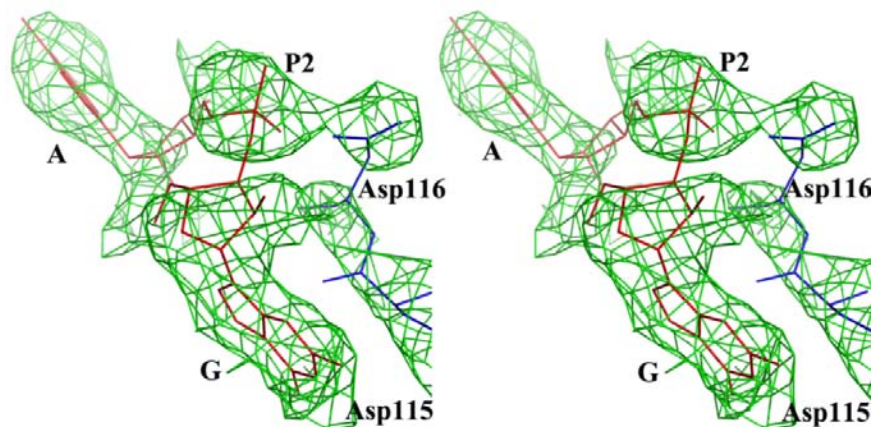
Figure 4.18. Electron density map showing phosphate-carboxylate interactions in the HLSV CP. Asp116-phosphate2 interactions are shown as sticks. The electron density is contoured at the 1σ level. The colors of protein and RNA are blue and red, respectively.
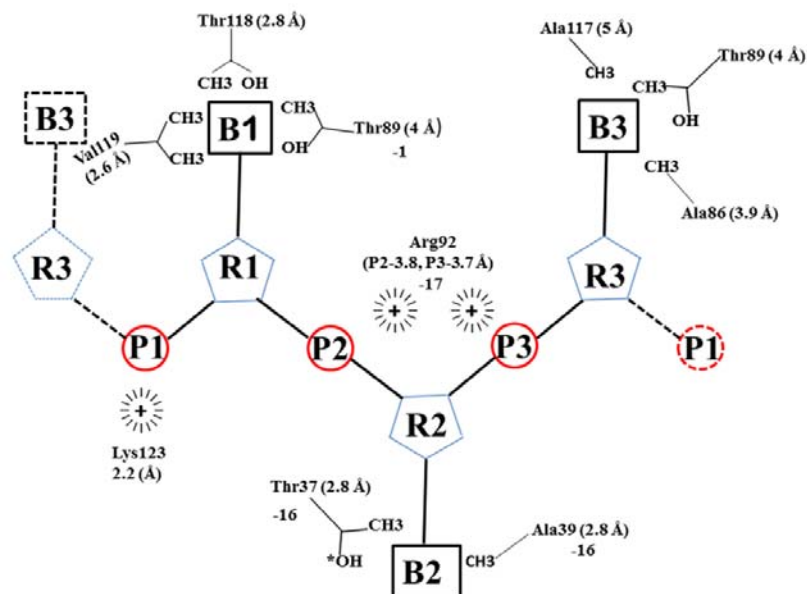
Figure 4.19. Summary of CP-RNA interactions of HLSV (Schematic diagram). The RNA bases are named B1, B2 and B3 whereas the ribose sugars and phosphates are labeled as R1, R2, R3 and P1, P2, P3, respectively. The interacting residues and subunits are also labeled. Subunits are not labeled for in-subunit interactions. The charge-charge interactions are shown as .

4.3   DISCUSSION

4.3.1   Protein-RNA interactions

There are several unique features in the protein-RNA interaction in HLSV. Firstly, Arg90 which is conserved among all *Tobamoviruses* is replaced by a lysine. In TMV, three arginines (Arg41, Arg90 and Arg92) in the CP interact with viral RNA and provide stability (Namba et al., 1989), while in CGMMV, such stability is maintained by Arg41, Lys112 and Arg122 (Wang and Stubbs, 1994). In a protein disk study, where RNA is absent, Arg92 is disordered (Champness et al., 1976; Jardetzky et al., 1978) but is ordered in an intact virus (Stubbs *et al.*, 1977). In HLSV, the protein-RNA interactions are stabilized by Ag92 and Lys123. Secondly, Thr37 forms a hydrogen bond with N6 of base2 (of subunit -16). Thirdly, the location of the

interaction of Lys123 with phosphate1 is unique, as the corresponding residue in other viruses is uncharged.

Phosphate-carboxylate interactions are common among all *Tobamoviruses*. In TMV, this location is believed to be involved in binding of Ca^{2+} ions. The interaction among phosphate, carboxylate and Ca^{2+} ions with TMV is presumed to play an important role in its assembly and disassembly (Namba et al., 1989). Electron density of calcium was not observed in the X-ray FD map of HLSV, most likely owing to the use of EDTA in virus purification.

Tobamoviruses recognize viral RNA by means of a repeating GXX sequence in a loop at the OAS (Zimmern, 1977; Meshi et al., 1981; Meshi et al., 1983; Takamstsu et al., 1983). This was demonstrated *in vitro* with the CGMMV CP and TMV RNA. In both TMV and U2, RNA guanine1 is assumed to be recognized by a hydrogen bonding interaction with Arg122 (Namba et al., 1989; Pattanayek and Stubbs, 1992). In TMV CP, Asp115 forms a hydrogen bond with N2 of the guanine1, contributing further to base recognition specificity (Namba et al., 1989). In CGMMV CP, Arg122 is involved in phosphate group stabilization and its base specificity is achieved by hydrogen bonding between Gln36 and N2 of guanine. In contrast, HLSV follows RMV CP in its hydrogen with guanine1 (Wang et al., 1997). Electron density for Asp115 and guanine1 shows that they are in close proximity, suggesting nucleotide selection specificity during virus assembly.

4.3.2 HLSV CP protein-protein interaction

Protein-protein interactions have been extensively studied in the *Tobamovirus* genus. A salient feature of *Tobamoviruses* is the carboxyl-carboxylate interaction. The HLSV CP architecture is well conserved through evolution. The assembly and

disassembly of *Tobamoviruses* are believed to be controlled by electrostatic interactions between the negatively charged carboxyl-carboxylate groups, mainly between subunits (Caspar, 1963; Bancroft, 1970). TMV has been shown to bind calcium ions (Loring et al., 1962; Gallagher and Lauffer, 1983a; Gallagher and Lauffer, 1983b) and its carboxylate groups are characteristic calcium-binding sites (Einspahr and Bugg, 1984). After comparing the structure of HLSV with those of TMV, CGMMV, RMV and ORSV, it is observed that the process of such interactions is dynamic (Table 4.2). The axial interaction of CGMMV is contributed from subunit 17, whereas for the other 3 *Tobamoviruses*, the interaction is from subunit 16. The potential high radius axial carboxyl-carboxylate groups in HLSV are Glu46 (central subunit), Asp126 (subunit +17) and Glu81 (subunit +16). The Glu46-Asp126 carboxyl-carboxylate pair is also conserved in CGMMV (Wang and Stubbs, 1994), but in HLSV, it is further strengthened by Glu81 from the subunit +16. The HLSV low radius carboxylates cluster in the loop region, between Glu99-Glu97 and Glu106-Glu95 (subunit -1). In addition, Glu98 from subunits -16 and -17 interacts with central subunit (Fig. 4.20 and Table 4.3). Hence, the low radius carboxyl-carboxylate interactions of HLSV are more complex, like those of RMV (Wang et al., 1997; Wang et al., 1998).
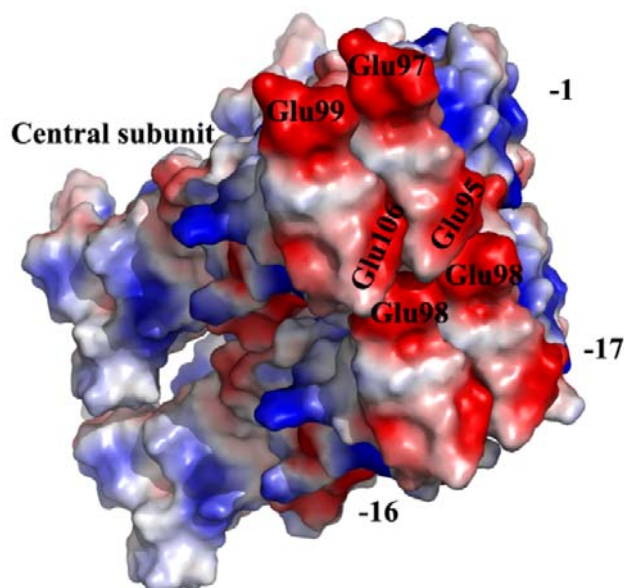
Figure 4.20. Surface density diagram of low radius carboxyl-carboxylate interactions of HLSV coat protein (CP). HLSV CP shows carboxyl-carboxylate interactions between the central and subunits -1, -16 and -17, as in Fig. 4.12. The red and blue colors represent a charge distribution of -55.5 and +55.5 KT/e, respectively.

4.3.3　Other structural features of HLSV

In the CGMMV CP, its residue 87 is a threonine and it contributes to the shortening of the RR α-helix significantly. In HLSV, Thr87 breaks the corresponding helix. Its Thr75 further shortens the RR helix, which is the shortest when compared to the TMV, CGMMV and RMV structures. In addition, the kink in the LR helix of HLSV is increased by His122. This is a major structural difference.

Understanding the recognition of GAA by the CP is very important to appreciate RNA specificity and virus assembly. Asp115 and Arg122 form guanine specific hydrogen bonds in TMV and U2, but not in CGMMV (Namba et al., 1989; Pattanayek & Stubbs, 1992). The structural basis of RNA guanine1 recognition is less obvious in CGMMV (Wang and Stubbs, 1994) and RMV. Arg122 is believed to be the key player in base recognition during virus assembly both in CGMMV and RMV (Wang et al., 1997). The considerable flexibility of the arginine side chain would

easily allow it to extend during assembly process (Wang and Stubbs 1994). This otherwise conserved Arg122 in *Tobamoviruses* is changed to His122 in HLSV and SHMV. The presence of two consecutive positively charged amino acids at this position is unique to HLSV. His122 is involved in a salt bridge and Lys123 is involved in a base independent phosphate interaction. In TMV and CGMMV, guanine1 selection during assembly is achieved by hydrogen bonding with viral CP Arg122 and Gln36, respectively (Namba et al., 1989; Wang and Stubbs, 1994). However, guanine1 selection in HLSV is achieved by Asp115, as in RMV (Wang et al., 1997). More studies should be undertaken to understand the importance of guanine1 recognition and viral assembly.

The four residues at positions 25, 62, 90 and 129 are conserved in all other *Tobamoviruses*, except for HLSV (Fig. 4.6). Glu25 in the HLSV CP makes an intramolecular salt bridge with Lys29 and stabilizes the RS α-helix. Also, Tyr62 functions only as a hydrophobic residue and fails to execute its second role as a charge stabilizer, as it is not a part of any helix and no charge residues are seen in its vicinity. Lys90 and Ile128 replace Arg and Leu in other *Tobamoviruses*, respectively. Lys90 is at a distance of 5.9 Å from guanine1 and probably it could be one of the hydrogen bonding candidates during viral assembly before attaining a fixed conformation in a virion, similar to the proposed method for Arg122 in CGMMV and RMV (Wang et al., 1997). This is only a speculation and needs validation with mutational experiments**.** Ile128 is part of the LR helix and it is likely to assume the same role of Leu128 in other viruses.

In conclusion, the kink at the LR helix of HLSV is a unique feature among *Tobamoviruses*. *Tobamovirus* protein-nucleic acid base specificity interaction during virus assembly is proposed to be achieved by viral CP Arg122 residue (Namba et al.,

1989; Wang et al., 1997). This position is occupied by a histidine in HLSV which is not close to any of the RNA bases. However, it interacts with Asp88 of the neighboring lateral subunit 1 and stabilizes the less ordered loop through a salt bridge. Furthermore, Lys123 in HLSV is close to phosphate1, suggesting its possible involvement with base recognition. However, the exact nature and mechanism of RNA recognition by *Tobamoviruses* are yet to be determined.

The intersubunit carboxyl-carboxylate interactions of HLSV partially resemble those of CGMMV, but remain distinct from other *Tobamoviruses*. The carboxyl-carboxylate interactions include involvement of both subunit +16 and +17 in HLSV. This is another difference observed, as the high radius interactions are usually confined to the axial subunit, either 16 or 17 in other *Tobamoviruses*. This unique axial contribution shows more intermolecular carboxyl-carboxylate interactions occurring in the HLSV CP. The low radius inter-subunit interactions in other *Tobamoviruses* are confined only to the lateral inter-subunit direction but in HLSV they extend in both lateral and axial directions.

4.4 FUTURE DIRECTIONS

Viruses are nucleoprotein complexes. Due to their small size and rapid replication rate, they are very useful in biotechnological applications. A virus CP has the capacity to pack other molecules, apart from nucleic acids, so that it could be exploited as a potential drug packaging protein nano cage. (Douglas & Young, 1998; Ren et al., 2006). These viruses like particles (VLPs) can be used as a novel delivery platform for drugs, for example Doxorubicin, an anti-cancer drug (Ren et al., 2007).

Gallie and his team are able to pack multiply dispersed copies of RNA containing the OAS to form pseudovirus like particles (Gallie et al., 1987a). In an *in*

vitro study, a short OAS, located in the MP gene, was shown to nucleate encapsidation of the 6395 nucleotide long genome by the TMV CP (Butler, 1984; Lomonossoff & Wilson, 1985). *Agrobacterium tumefaciens* was used to incorporate the TMV OAS and was found to assemble into a stable 'pseudovirus' like particle *in vivo* during systemic infection by TMV (as a helper). These results may further be extended for packaging foreign genes into a virus. As a route to protect, accumulate and recover a specific mRNA *in vivo*, in transgenic plant cells, this new method may be useful in developmental and plant molecular biology (Sleat et al., 1988; Sleat et al., 1986).

To control the mosquito *Aedes*, trypsin-modulating oostatic factor was expressed successfully on the virion of TMV as a potential larvicide (Borovsky et al., 2006). Similar approach can be adopted with our structural details of the HLSV and may be used as a biocontrol in future. TMV has also been used in the development of vaccines. The Canine oral papillomavirus L2 protein, displayed on the surface of the TMV CP along with streptavidin, was found to be more immunogenic than uncoupled antigen when tested in mice (Sleat et al., 1986). TMV was also used to display protein A, which was densely packed as nanoparticles (>2,100 copies per viral particle) on its virion. This can be a new approach as an immunoadsorbent to purify monoclonal antibodies (mAb) with less cost (Werner et al., 2006). With the structural details of HLSV we can also adopt similar approach and carry out more experiments to see its suitability as an immunoadsorbent. Furthermore, a new TMV based vector has been used to produce systemically an angiotensin-I converting enzyme inhibitor in transgenic tobacco and tomato (Hamamoto et al., 1993). In another study two epitopes of the foot and mouth disease virus are successfully expressed by a TMV based vector (Wu et al., 2003). This study can be useful in developing vaccines

against many pathogenic viruses and bacteria. The full length clone of CGMMV was used to express Hepatitis B surface antigen (Ooi et al., 2006). The study showed threefold increase in the level of anti HBsAg immunoglobulin, suggesting the possible application of new chimeric virus as an effective Hepatitis B vaccine. HLSV can also be exploited in similar studies.

We believe that with the current structure of HLSV, we can do more functional studies to better understand the virus structure function relation. We can ask some basic questions like for example what will happen to virus if the we mutate the His122 of the CP? Will the kink still exist if we mutant His122? What kind of amino acid substitution will increase the degree of the bent? Will the virus LR helix remain structurally the same? Will the virus be as infective as the wild type? Can the knowledge be used in designing nano-molecule for anti-cancer drug delivery?

REFERENCES

Alonso, E., Garcia-Luque, I., De-la Cruz, A., Wicke, B., Avila-Rincon, M. J., Serra, M. T., Castresana, C. & Diaz-Ruiz, J. R. (1991). Nucleotide sequence of the genomic RNA of Pepper mild mottle virus, a resistance-breaking *Tobamovirus* in pepper. *J. Gen. Virol.* **72,** 2875-2884.

Altschuh. D., Lesk, A. M., Bloomer, A. C. & Klug, A. (1987). Correlations of co-ordinated amino acid substitutions with function in viruses related to TMV. *J. Mol. Biol.* **193,** 693-707.

Argos, P. (1988). A sequence motif in many polymerases. *Nucl. Acids Res.* **16,** 9909-9916.

Artymiuk, P. J. & Blake, C. C. F. (1981). Refinement of human lysozyme at 1.5 Å resolution analysis of non-bonded and hydrogen-bond interactions. *J. Mol. Biol.* **152,** 737-762.

Asurmendi, S., Berg, R. H., Koo, J. C. & Beachy, R. N. (2004). Coat protein regulates formation of replication complexes during TMV infection. *Proc. Natl. Acad. Sci. USA.* **101,** 1415-1420.

Bancroft, J.R. (1970). The self-assembly of spherical plant viruses. *Adv. Virus Res.* **16,** 99-134.

Bansal, M., Kumar, S. & Velavan, R. (2000). HELANAL: a program to characterize helix geometry in proteins. *J. Biomol. Struct. Dyn.* **17,** 811–819.

Barlow, D. & Thornton, J. M. (1988). Helix geometry in protein. *J. Mol. Biol.* **201,** 601-619.

Bawden, F. C., Pirie, N. W., Bernal, J. D. & Fankuchen, I. (1936). Liquid crystalline substances from virus-infected plants. *Nature* **138,** 1051-1052.

Beijerinck, M. W. (1898). Concerning a contagium vivum fluidium as a cause of the spot-disease of tobacco leaves. Reprint from: Phytopathology Classics, Number 7. 1942.

Bernal, J. D. & Fankuchen, I. (1941). X-ray and crystallographic studies of plant virus preparations. III. *J. Gen. Physiol.* **25,** 147-165.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Jr-Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi T. & Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112,** 535-542.

Bhyravbhatla, B., Watowich, S. & Caspar, D. L. D. (1998). Refined atomic model of the four-layer aggregate of the TMV coat protein at 2.4 Å resolution. *Biophys. J.* **74,** 604-615.

Bian, W., Wang, H., McCullough, I. & Stubbs, G. (2006). WCEN: a computer program for initial processing of fiber diffraction patterns *J. Appl. Crystallogr.* **39,** 752-756.

Bloomer, A. C. & Butler, P. J. G. (1986). The Plant Viruses (Van Regenmortel, M.H. V., Fraenkel-Conrat, H., Eds.) Vol. 2, pp 19-57, Plenum, New York.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). Protein disk of TMV at 2.8 Å resolution showing the interactions within and between subunits. *Nature* **276,** 362-368.

Blundell, T. L., Barlow, D., Barkakoti N. & Thornton, J. M., 1983. Solvent-induced distortions and the curvature of α-helices. *Nature* **306,** 281-283.

Borovsky, D., Rabindran, S., Dawson, W. O., Powell, C. A., Iannotti, D. A., Morris, T. J., Shabanowitz, J., Hunt, F. D., Debondt, H. L. & DeLoof, A. (2006). Expression of aedes trypsin-modulating oostatic factor on the virion of TMV: a potential larvicide 2006. *Proc. Natl. Acad. Sci. USA.* **103,** 18963-18968.

Butler, P. J. G. (1984). The current picture of the structure and assembly of TMV. *J. Gen. Virol.* **65,** 253-279.

Butler, P. J. G. & Durham, A. C. H. (1977). TMV protein aggregation and the virus assembly. *Adv. Protein Chem.* **31,** 187-251.

Butler, P. J. G., Finch, J. T. & Zimmern, D. (1977). Configuration of TMV RNA during virus assembly. *Nature* **265,** 217-219.

Butler, P. J. G. & Klug, A. (1971). Assembly of the particle of TMV from RNA and disks of protein. *Nature New Biol.* **229,** 47-50.

Caspar, D. L. D. (1956). Structure of TMV: radial density distribution in the TMV Particle. *Nature* **177,** 928.

Caspar, D. L. D. (1963). Assembly and stability of the Tobacco mosiac virus Particle. *Adv. Protein Chem.* **18,** 37-121.

Chakarabarti, P., Bernard, M. & Rees, D. C. (1986). Peptide bond distortions and the curvature of α-helices. *Biopolymers* **25,** 1087-1093.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G. & Klug, A. (1976). The structure of the protein disk of TMV to 5 Å resolution. *Nature* **259,** 20-24.

Chandrasekaran, R. & Stubbs, G. (2001). Fibre diffraction in international tables for crystallography, Vol. F: Crystallography of biological macromolecules (Rossman, M.G. and Arnold, E., eds.), Kluwer Academic Publishers, The Netherlands, 444-450.

Chng, C. G., Wong, S. M., Mahtani, P. H., Loh, C. S., Goh, C. J., Kao, M. C. C., Chung, M. C. M. & Watanabe, Y. (1996). The complete sequence of a Singapore

isolate of Odontoglossum ringspot virus and comparison with other *Tobamoviruses*. *Gene* **171,** 155-161.

Cochran, W. & Crick, F. H. C. (1952a). Evidence for the Pauling–corey α-helix in synthetic polypeptides. *Nature* **169,** 234-235.

Cochran, W., Crick, F. H. C. & Vand, V. (1952b). The structure of synthetic polypeptides. 1. The transform of atoms on a helix. *Acta Crystallogr.* **5,** 581-586.

Cohen C. & Bear, R. S. (1953). Helical polypeptide chain configuration in collagen. *J. Amer. Chem. Soc.* **75,** 2783-2784.

Cowan, P., North, A. C. W. & Randall, J. T. (1953). Nature and structure of collagen. pp 241. London: Butterworths.

Culver, J. N., Dawson, W. O., Plonk, K. & Stubbs, G. (1995). Site-directed mutagenesis confirms the involvement of carboxylate groups in the disassembly of TMV. *Virology* **206,** 724-730.

Culver, J. N., Stubbs, G. & Dawson, W. O. (1994). Structure-function relationship between Tobacco mosaic virus coat protein and hypersensitivity in *Nicotiana sylvestris*. *J. Mol. Biol.* **242,** 130-138.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. http://www.pymol.org.

Dolja, V. V., Boyko, V. P., Agranovsky, A. A. & Koonin, E. V, (1991). Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. *Virology* **184,** 79-86.

Douglas, T. & Young, M. (1998). Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **393,** 152-155.

Durham, A. C. H., Finch, J. T. & Klug, A. (1971). States of aggregation of TMV protein. *Nature New Biol.* **229,** 37-42.

Einspahr, H. & Bugg, C. E. (1984). Metal ions in biological systems: calcium and its role in biology (Sigel, H., ed.), Vol 17, 51-97, Dekker, New York.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Biol. Crystallogr. D* **60,** 2126-2132.

Finch, J. T. (1965). Preliminary X-ray diffraction studies on Tobacco rattle and barley stripe mosaic viruses. *J. Mol. Biol.* **12,** 612-619.

Franklin, R. E. & Klug, A. (1955). The splitting of layer lines in X-ray fibre diagrams of helical structures: application to TMV. *Acta Crystallogr.* **8,** 777-780.

Franklin, R. E. (1956). Structure of TMV: location of the ribonucleic acid in the TMV Particle. *Nature* **177,** 928-930.

Franklin, R. E. & Holmes K. C. (1958). TMV: application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. *Acta Crystallogr.* **11,** 213-220.

Fujiki, M., Kawakami, S., Kim, R. W. & Beachy, R. N. (2006). Domains of TMV movement protein essential for its membrane association. *J. Gen. Virol.* **87,** 2699-2707.

Furey, W. F., Robbins, A. H., Clancy, L. L., Winge, D. R., Wang, B. C. & Stout C. D. (1986). Crystal structure of Cd, Zn metallothionein. *Science* **231,** 704-710.

Gallagher, W. H. & Lauffer, M. A. (1983a). Calcium ion binding by TMV. *J. Mol. Biol.* **170,** 905-919.

Gallagher, W. H. & Lauffer, M. A. (1983b). Calcium ion binding by TMV. *J. Mol. Biol.* **170,** 921-929.

Gallie, D. R., Plaskitt, K. A., Michael, T. & Wilson, A. (1987a). The effect of multiple dispersed copies of the origin of assembly sequence from TMV RNA on the morphology of pseudovirus particles assembled *in vitro*. *Virology* **158,** 473-476.

Gallie, D. R., Sleat, D. E., Watts, J. W., Turner, P. C. & Wilson, T. M. (1987b). The 5' leader sequence of TMV RNA enhances the expression of foreign gene transcripts in vitro and in vivo. *Nucleic Acids Res*. **15,** 3257-3273.

Goelet, P., Lomonossoff, G. P., Butler, P. J., Akam, M. E., Gait, M. J. & Karn, J. (1982). Nucleotide sequence of Tobacco mosaic virus RNA. *Proc Natl. Acad. Sci. USA*. **79,** 5818-5822.

Goldbach, R. & De, H. P. (1994). RNA viral supergroups and the evolution of RNA viruses. In The evolutionary biology of viruses (ed. S. Morse) pp. 161-184. New York: Raven Press.

Gonzalez, A., Nave, C. & Marvin, D. A. (1995). Pf1 filamentous bacteriophage: refinement of a molecular model by simulated annealing using 3.3 Å resolution X-ray fibre diffraction data. *Acta Crystallogr. D* **51,** 792-804.

Goulden, M. G., Davies, J. W., Wood, K. R. & Lomonossoff, G. P. (1992). Structure of *Tobraviral* particles: a model suggested from sequence conservation in *Tobraviral* and *Tobamoviral* coat proteins. *J. Mol. Biol.* **227,** 1-8.

Gregory, J. & Holmes, K. C. (1965). Methods of preparing orientated Tobacco mosaic virus sols for X-ray diffraction. *J. Mol. Biol.* **13,** 796-801.

Grubb, D. T. & Ji, G. (1999). Molecular chain orientation in supercontracted and re-extended spider silk. *Int. J. Biol. Macromol.* **24,** 203-210.

Hamamoto, H., Sugiyama, Y., Nakagawa, N., Hashida, E., Matsunaga, Y., Takemoto, S., Watanabe, Y. & Okada, Y. (1993). A new Tobacco mosaic virus vector and its use for the systemic production of angiotensin-I-converting enzyme inhibitor in transgenic tobacco and tomato. *Nature Biotech.* **11,** 930-932.

Heinze, C., Lesemann, D. E., Ilmberger, N., Willingmann, P. & Adam, G. (2006). The phylogenetic structure of the cluster of tobamovirus species serologically related to ribgrass mosaic virus (RMV) and the sequence of streptocarpus flower break virus (SFBV). *Arch. Virol.* **151,** 763-774.

Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115,** 252-270.

Higgins, T. J. V., Goodwin, P. B. & Whitfield, P. R. (1976). Occurence of short particles in beans infested with cowpea strain of TMV. II. Evidence that short particles contain the cistron for coat-protein. *Virology* **71,** 486-497.

Hirth, L. & Richards, K. E. (1981). Tobacco mosaic virus: model for structure and function of a simple virus. *Adv. Virus Res.* **26,** 145-199.

Holmes, K. C., Stubbs, G. J., Mandelkow, E. & Gallwitz, U. (1975). Structure of Tobacco mosaic virus at 6.7 Å resolution. *Nature* **254,** 192-196.

Hull, R. (2001). Matthews' Plant Virology, 4th ed. San Diego: Academic Press.

Ikeda, R., Watanabe, E., Watanabe, Y. & Okada, Y. (1993). Nucleotide sequence of *Tobamovirus* Ob which can spread systemically in N gene tobacco. *J. Gen. Virol.* **74,** 1939-1944.

Inouye, H. & Kirschner, D. A. (1998). Polypeptide chain folding in the hydrophobic core of hamster scrapie prion: analysis by X-ray diffraction. *J. Struct. Biol.* **122,** 247-255.

Ivanova, M. I. & Makowski, L. (1998). Iterative low-pass filtering for estimation of the background in fiber diffraction patterns. *Acta Crystallogr. A* **54,** 626-631.

Jack, A. & Levitt, M. (1978). Refinement of large structures by simultaneous minimization of energy and R factor. *Acta Crystallogr. A* **34,** 931-935.

James, V. J., McConnell, J. F. & Amemiya, Y. (1998). Molecular structural changes in human fetal tissue during the early stages of embryogenesis. *Biochim. Biophys. Acta* **1379,** 282-288.

Jardetzky, O., Akasaka, K., Vogel, D., Morris, S. & Holmes, K. C. (1978). Unusual segmental flexibility in a region of Tobacco mosaic virus coat protein. *Nature* **273,** 564-566.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110-119.

Kamer, G. & Argos, P. (1984). Primary structural comparisons of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucl. Acids Res.* **12,** 7269-7282.

Kendall, A., McDonald, M. & Stubbs, G. (2007). Precise determination of the helical repeat of Tobacco mosaic virus. *Virology* **369,** 226-227.

Kirschner, D. A., Elliot-Bryant, R., Szumowski, K. E., Gonnerman, W. A., Kindy, M. S., Sipe, J. D. & Cathcart, E. S. (1998). *In vitro* amyloid fibril formation by synthetic peptides corresponding to the amino terminus of apo SAA isoforms from amyloid-susceptible and amyloid-resistant mice. *J. Struct. Biol.* **124,** 88-98.

Klug, A., Crick, F. H. C. & Wyckoff, H. W. (1958). Diffraction by helical structures. *Acta Crystallogr.* **11,** 199-213.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28,** 375-430.

Kumar, S. & Bansal, M. (1996). Structural and sequence characteristics of long a helices in globular proteins. *Biophy. J.* **71,** 1574-1586.

Kumar, S. & Bansal, M. (1998). Geometrical and sequence characteristics of α-helices in Globular Proteins. *Biophy J.* **75,** 1935-1944.

Lartey, R. T., Voss, T. C. & Melcher, U. (1995). Completion of a c-DNA sequence from a *Tobamovirus* pathogenic to crucifers. *Gene* **166,** 331-332.

Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. App. Crystallogr.* **26,** 283-291.

Lebeurier, G., Nicolaieff, A. & Richards, K. E. (1977). Inside-out model for self-assembly of Tobacco mosaic virus. *Proc. Natl. Acad. Sci. USA.* **74,** 149-153.

Lo´pez-Moya, J. J. & Garci´a, J. A. (1999). Encyclopedia of virology, edited by R. Webster & A. Granoff, pp. 1369-1375. London: Academic Press.

Lobert, S. & Stubbs, G. (1990). Fiber diffraction analysis of Cucumber green mottle mosaic virus using limited numbers of heavy atom derivatives. *Acta Crystallogr. A* **46,** 993-997.

Lomonossoff, G. P. & Wilson, T. M. A. (1985). In Davies, J. W. (ed), Mol. Plant Virol. CRC Press, Florida, Vol. 1, pp 43-83.

Loring, H. S., Fujimoto, Y. & Tu, A. T. (1962). Tobacco mosaic virus a calcium-magnesium coordination complex. *Virology* **16,** 30-40.

Love, W. F., Klock, P. A., Lattman, E. E., Padhan, E. A., Ward K. B. J. & Hendrickson, W. A. (1971). *Cold Spring Harbor Symp., Quan. Biol.* **36,** 349-357.

Lu, B., Stubbs, G. & Culver, J. N. (1996). Carboxylate interactions involved in the disassembly of Tobacco mosaic *Tobamovirus*. *Virology* **225,** 11-20.

MacArthur, W. M. & Thornton, J. M. (1996). Deviations from planarity of the peptide bond in peptides and proteins. *J. Mol. Biol.* **264,** 1180-1195.

MacFarlane, S. A. (1999). Molecular biology of the *Tobraviruses*. *J. Gen. Virol.* **80,** 2799-2807.

Makowski, L. (1978). Processing of X-ray diffraction data from partially oriented specimens. *J. Appl. Crystallogr.* **11,** 273-83.

Makowski, L. (1982). The use of continuous diffraction data as a phase constraint. II. application to fiber diffraction data. *Appl. Crystallogr.* **15,** 546-557.

Makowski, L., Caspar, D. L. D. & Marvin, D. A. (1980). Filamentous bacteriophage Pf1 structure determined at 7 Å resolution by refinement of models for the α-helical subunit. *J. Mol. Biol.* **140,** 149-181.

Malinchik, S. B., Inouye, H., Szumowski, K. E. & Kirschner, D. A. (1998). Structural analysis of Alzheimer's beta (1-40) amyloid: protofilament assembly of tubular fibrils. *Biophys. J.* **74,** 537-545.

Mandelkow, E., Stubbs, G. & Warren, S. (1981). Structures of the helical aggregates of Tobacco mosaic virus protein. *J. Mol. Biol.* **152,** 375-386.

Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8,** 150-158.

Marvin, D. A., Bryan, R. K. & Nave, C. (1987). Pf1 Inovirus: electron density distribution calculated by a maximum entropy algorithm from native fibre diffraction data to 3 Å resolution and single isomorphous replacement data to 5 Å resolution. *J. Mol. Biol.* **193,** 315-343.

Marvin, D. A., Pigram, W. J., Wiseman, R. L., Wachtel, E. J. & Marvin, F. J. (1974a). Filamentous bacterial viruses: XII. Molecular architecture of the class I (fd, If1, IKe) virion. *J. Mol. Biol.* **88,** 581-582.

Marvin, D. A., Wiseman, R. L. & Wachtel, E. J. (1974b). Filamentous bacterial viruses: XI. Molecular architecture of the class II (Pf1, Xf) virion. *J. Mol. Biol.* **82,** 121-130.

Mathis, A. & Linthorst, H. J. M. (1994). Encyclopedia of Virology vol 2, ed R G Webster and A Ganoff (London: Academic) pp 1442-1446.

McDonald, M., Kendall, A., Tanaka, M., Weissman, J. S. & Stubbs, G. (2008). Enclosed chambers for humidity control and sample containment in fiber diffraction. *J. Appl. Cryst.* **41,** 206-209.

McGillavry, C. H. & Bruins, E. M. (1948). On the Patterson transforms of fibre diagrams. *Acta Crystallogr.* **1,** 156-158.

McLachlan, A. D., Bloomer, A. C. & Butler, P. J. G. (1980). Structural repeats and evolution of Tobacco mosaic virus coat protein and RNA. *J. Mol. Biol.* **136,** 203-224.

Meshi, T., Kiyama, R., Ohno, T. & Okada, Y. (1983). Nucleotide sequence of the coat protein cistron and 3′ noncoding region of Cucumber green mottle mosaic virus (watermelon strain) RNA. *Virology* **127,** 54-64.

Meshi, T., Ohno, T., Iba, H. & Okada, Y. (1981). Nucleotide sequence of a cloned cDNA copy of TMV (cowpea strain) RNA, including the assembly origin, the coat protein cistron, and the 3′ non-coding region. *Mol. Gen. Genet.* **184,** 20-25.

Millane, R. P. (1989). R factors in X-ray fiber diffraction. II. Largest likely R factors. *Acta Crystallogr. A* **45,** 573-576.

Min, B. E., Song, Y. S. & Ryu, K.H. (2009). Complete sequence and genome structure of cactus mild mottle virus. *Arch. Virol.* **154,** 1371-1374.

Mu, X. Q. & Makowski, L. (2000). The likelihood function in fiber diffraction. *Acta Crystallogr. A* **56,** 168-177.

Namba, K., Pattanayek, R. & Stubbs, G. (1989). Visualization of protein-nucleic acid interactions in a virus: refined structure of intact Tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction. *J. Mol. Biol.* **208,** 307-325.

Namba, K. & Stubbs, G. (1985). Solving the phase problem in fiber diffraction. Application to Tobacco mosaic virus at 3.6 Å resolution. *Acta Crystallogr. A* **41,** 252-262.

Namba, K. & Stubbs, G. (1986). Structure of Tobacco mosaic virus at 3.6 Å resolution: implications for assembly. *Science* **231,** 1401-1406.

Namba, K. & Stubbs, G. (1987a). Isomorphous replacement in fiber diffraction using limited numbers of heavy-atom derivatives. *Acta Crystallogr. A* **43,** 64-69.

Namba, K. & Stubbs, G. (1987b). Difference Fourier syntheses in fiber diffraction. *Acta Crystallogr. A* **43,** 533-539.

Oda, T., Makino, K., Yamashita, I., Namba, K. & Ma´eda, Y. (1998). Effect of the length and effective diameter of F-actin on the filament orientation in liquid crystalline sols measured by x-Ray fiber diffraction. *Biophys. J.* **75,** 2672-2681.

Okada, Y. (1986). Molecular assembly of Tobacco mosaic virus *in vitro. Adv. Biophys.* **22,** 95-149.

Olson, A. J., Bricogne, G. & Harrison, S. C. (1983). Structure of Tomato busy stunt virus IV. The virus particle at 2.9 Å resolution. *J. Mol. Biol.* **171,** 61-93.

Ooi, A., Tan, S., Mohamedc, R., Rahmand, N. A. & Othmana, R. Y. (2006). The full-length clone of Cucumber green mottle mosaic virus and its application as an expression system for hepatitis B surface antigen. *J. Biotech.* **121,** 471-481.

Pattanayek, R. & Stubbs, G. (1992). Structure of the U2 strain of Tobacco mosaic virus refined at 3.5 Å resolution using X-ray fiber diffraction. *J. Mol. Biol.* **228,** 516-528.

Purslow, P. P., Wess, T. J. & Hukins, D. W. L. (1998). Collagen orientation and molecular spacing during creep and stress-relaxation in soft connective tissues. *J. Exp. Biol.* **201,** 135-142.

Raghavendra, K., Adams, M. L. & Schuster, T. M. (1985). Tobacco mosaic virus protein aggregates in solution: structural comparison of 20S aggregates with those near conditions for disk crystallization. *Biochemistry* **24,** 3298-3304.

Regenmortel, M. H. V. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., Mcgeoch, D. J., Pringle, C. R. & Wickner, R. B. (ed) (2000). Virus taxonomy: seventh report of the Int. committee on taxonomy of viruses (San Diego: Academic)

Ren, Y., Wong, S. M. & Lim, L. Y. (2006). In vitro-reassembled plant virus-like particles for loading of polyacids. *J. Gen. Virol.* **87,** 2749-2754.

Ren, Y., Wong, S. M. & Lim, L. Y. (2007). Folic acid-conjugated protein cages of a plant virus: novel delivery platform for doxorubicin. *Bioconjugate Chem.* **18,** 836-843.

Rhie, M. J., Min, B. E., Hong, J. S., Song, Y. S. & Ryu, K. H. (2007). Complete genome sequence supports bell pepper mottle virus as a species of the genus Tobamovirus. *Arch. Virol.* 152, 1401-1407.

Rhodes, G. (2000). Crystallography made crystal clear 2nd edn (San Diego: Academic).

Richards, F. M. & Kundrot, C. E. (1980). Identification of structural motifs from protein coordinate data: secondary structure and first level super-secondary structure. *Proteins* **3,** 71-84.

Richardson, J. F., Tollin, P. & Bancroft, J. B. (1981). The architecture of the *Potexviruses. Virology* **112,** 34-39.

Riechmann, J. L., La´ın, S. & Garcia, J. A. (1992). Highlights and prospects of *Potyvirus* molecular biology. *J. Gen. Virol.* **73,** 1-16.

Schuster, T., Scheele, R., Adams, M., Shire, S., Steckert, J. & Potschka, M. (1980). Studies on the mechanism of assembly of Tobacco mosaic virus. *Biophys. J.* **32,** 313-329.

Schuster, T. M., Scheele, R. B. & Khairallah, L. H. (1979). Mechanism of self-assembly of Tobacco mosaic virus protein. I. Nucleation-controlled kinetics of polymerization. *J. Mol. Biol.* **127,** 461-468.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *J. Mag. Res.* **160,** 65-73.

Shalaby. R. A. F. & Lauffer, M. A. (1977). Hydrogen ion uptake upon Tobacco mosaic virus protein polymerization. *J. Mol. Biol.* **116,** 709-725.

Shaw, A., McRee, D. E., Vacquir, V. D. & Stout, C. D. (1993). The crystal structure of lysin, a fertilization protein. *Science* **262,** 1864-1867.

Shire, S. J., Steckert, J. J. & Schuster, T. M. (1979). Mechanism of self-assembly of Tobacco mosaic virus protein. II. Characterization of the metastable polymerization nucleus and the initial stages of helix formation. *J. Mol. Biol.* **127,** 487-506.

Silver, S., Quan, S. & Deom, C. M. (1996). Completion of the nucleotide sequence of sunn-hemp mosaic virus: a *Tobamovirus* pathogenic to legumes. *Virus Genes* **13,** 83-85.

Sleat, D. E., Gallie, D. R., Watts, J. W., Deom, C. M., Turaer, P. C., Beachy, R. N. & Wilson, T. M. A. (1988). Selective recovery of foreign gene transcripts as virus-like particles in TMV-infected transgenic tobaccos. *Nucl. Acids Res.* **16,** 3127-3140.

Sleat, D. E., Turner, P. C., Finch, J. T., Butler, P. J. G. & Wilson, T. M. A. (1986). Packaging of recombinant RNA molecules into pseudovirus particles directed by the origin-of-assembly sequence from Tobacco mosaic virus RNA. *Virology* **155,** 299-308.

Solis, I. & Garcia-Arenal, F. (1990). The complete nucleotide sequence of the genomic RNA of the *Tobamovirus* Tobacco mild green mosaic virus. *Virology* **177,** 553-558.

Song, Y. S., Min, B. E., Hong, J. S., Rhie M. J, Kim, M. J & Ryu, K.H. (2006). Molecular evidence supporting the confirmation of maracuja mosaic virus as a species of the genus Tobamovirus and production of an infectious cDNA transcript (2006). *Arch. Virol.* **12,** 2337-48

Sreekanth, R., Pattabhi, V. & Rajan, S. S. (2008). Characterization of alpha helices interacting with nucleic acids. *Compt. Bio. Chem.* **32,** 378–381.

Srinivasan, K. G., Min, B. E., Ryu, K. H., Adkins, S. & Wong, S. M. (2005). Determination of complete nucleotide sequence of Hibiscus latent Singapore virus: evidence for the presence of an internal poly(A) trac. *Arch. Virol.* **150,** 153-166.

Srinivasan, K. G., Narendrakumar, R. & Wong, S. M. (2002). Hibiscus virus S is a new subgroup II *Tobamovirus*: evidence from its unique coat protein and movement protein sequences. *Arch. Virol.* **147,** 1585-1598.

Stubbs, G. (1984). In Biological Macromolecules and Assemblies, vol. 1. Virus Structures (Jurnak, F. A. & McPherson, A., eds), pp 149-202, Wiley, New York.

Stubbs, G. (1987). The Patterson function in fiber diffraction in Patterson and Pattersons (Glusker, J.P., Patterson, E.K., and Rossi, M., eds.), OUP, New York, 548-557.

Stubbs, G. (1989). The probability distributions of X-ray intensities in fiber diffraction: largest likely values for fiber diffraction R factors. *Acta Crystallogr. A* **45,** 254-258.

Stubbs, G. (1999). Developments in fiber diffraction. *Curr. Opin. Struct. Biol.* **9,** 615-619.

Stubbs, G. & Diamond, R. (1975). The phase problem for cylindrically averaged diffraction patterns. Solution by isomorphous replacement and application to Tobacco mosaic virus. *Acta Crystallogr. A* **31,** 709-718.

Stubbs, G., Ferrell, G., Reams, M. & Fletcher, N. (2000). Fibre diffraction and diversity in filamentous plant virus structure. *Fib. Diffract. Rev.* **9,** 24-28.

Stubbs, G. & Makowski, L. (1982). Coordinated use of isomorphous replacement and layer-line splitting in the phasing of fiber diffraction data. *Acta Crystallogr. A* **38,** 417-425.

Stubbs, G., Namba, K. & Makowski, L. (1986). Application of restrained least-squares refinement to fiber diffraction from macromolecular assemblies. *Biophys. J.* **49,** 58-60.

Stubbs, G. & Stauffachher, C. (1989). Structure of the RNA in Tobacco mosaic virus. *Acta Crystallogr. A* **31,** 709-718.

Stubbs, G., Warren, S. & Holmes, K. (1977). Structure of RNA and RNA binding site in Tobacco mosaic virus from 4 Å map calculated from X-ray fiber diagram. *Nature.* **267,** 216-221.

Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B. & Blake, C. C. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273,** 729-739.

Takamatsu, N., Ohno, T., Meshi, T. & Okada, Y. (1983). Molecular cloning and nucleotide sequence of the 30 K and the coat protein cistron of TMV (tomato strain) genomic RNA. *Nucl. Acids Res.* **11,** 3767-3778.

Tan, S. H., Nishiguchi, M., Murata, M. & Motoyoshi, F. (2000). The genome structure of Kyuri green mottle mosaic *Tobamovirus* and its comparison with that of Cucumber green mottle mosaic *Tobamovirus*. *Arch. Virol.* **145,** 1067-1079.

Tollin, P., Wilson, H. R. & Young, D. W. (1968). X-ray diffraction evidence of the helical structure of narcissus mosaic virus. *J. Mol. Biol.* **34,** 189-192.

Tollin, P. & Wilson, H. R. (1971). Observations on the structure of the campinas strain of Tobacco rattle virus. *J. Gen. Virol.* **13,** 433-440.

Tollin, P., Wilson, H. R. & Bancroft, J. B. (1980). Further observations on the structure of particles of Potato virus x. *J. Gen. Virol.* **49,** 407-410.

Torbet, J. (1987). Using magnetic orientation to study structure and assembly. *Trends Biochem. Sci.* **12,** 327-330.

Torbet, J. & Maret, G. (1979). Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field. *J. Mol. Biol.* **134,** 843-845.

Ugaki, M., Tomiyama, M., Kakutani, T., Hidaka, S., Kiguchi, T., Nagata, R., Sato, T., Motoyoshi, F. & Nishiguchi, M. (1991). The complete nucleotide sequence of Cucumber green mottle mosaic virus (SH strain) genomic RNA. *J. Gen. Virol.* **72,** 1487-1495.

Wada, Y., Tanaka, H., Yamashita, E., Kubo, C., Ichiki-Uehara, T., Nakazono-Nagaoka, E., Omurab, T. & Tsukihara, T. (2008). The structure of Melon necrotic spot virus determined at 2.8 Å resolution. *Acta Crystallogr. F* **64,** 8-13.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115,** 90-112.

Wang, H., Culver, J. N. & Stubbs, G. (1997). Structure of Ribgrass mosaic virus at 2.9 Å resolution evolution and taxonomy of *Tobamoviruses*. *J. Mol. Biol*. **269,** 769-779.

Wang, H., Planchart, A. & Stubbs, G. (1998). Caspar carboxylates: the structural basis of *Tobamovirus* disassembly. *Biophy. J.* **74,** 633-638.

Wang, H. & Stubbs, G. (1993). Molecular dynamics in refinement against fiber diffraction data. *Acta Crystallogr. A* **49,** 504-513.

Wang, H. & Stubbs, G. (1994). Structure determination of Cucumber green mottle *mosaic virus* by X-ray fiber diffraction. Significance for the evolution of *Tobamoviruses*. *J. Mol. Biol.* **239,** 371-384.

Waser, J. (1955). Fourier transforms and scattering intensities of tubular objects. *Acta Crystallogr.* **8,** 142-150.

Welsh, L. C., Symmons, M. F. & Marvin, D. A. (2000). The molecular structure and structural transition of the α-helical capsid in filamentous bacteriophage Pf1. *Acta Crystallogr. D* **56,** 137-150.

Welsh, L. C., Symmons, M. F., Sturtevant, J. M., Marvin, D. A. & Perham, R. N. (1998). Structure of the capsid of pf3 filamentous phage determined from X-ray fiber diffraction data at 3.1 Å resolution. *J. Mol. Biol.* **283,** 155-177.

Werner, S., Marillonnet, S., Hause, G., Klimyuk, V. & Gleba, Y. (2006). Immunoabsorbent nanoparticles based on a *Tobamovirus* displaying protein A. *Proc. Natl. Acad. Sci. U S A.* **103,** 17678-17683.

White, K. A., Rouleau, M., Bancroft, J. B. & Mackie, G. A. (1994). Potexviruses. In Encyclopedia of Virology (R. G. Webster and A. Ganoff, Eds.), Vol. 2, pp. 1142-1147. Academic Press, London.

Wickner, R. B., Edskes, H. K., Shewmaker, F. & Nakayashiki, T. (2007). Prions of fungi: inherited structures and biological roles. *Nat. Rev. Microbiol.* **5,** 611-618.

Wilkins, M. H. F., Seeds, W. E., Strokes, A. R. & Wilson, H. R. (1953). Helical Structure of crystalline deoxy-pentose nucleic Acid. *Nature* **171,** 759-762.

Wilson, H. R. & Tollin, P. (1969). Some observations on the structure of Potato virus X. *J. Gen. Virol.* **5,** 151-154.

Wilson, T. M. A. (1984). Cotranslational disassembly of Tobacco mosaic virus in vitro. *Virology* **137,** 255-265.

Winkler, S., Szela, S., Avtges, P., Valluzi, R., Kirschner, D. A. & Kaplan, D. (1999). Designing recombinant spider silk proteins to control assembly. *Int. J. Biol. Macromol.* **24,** 265-270.

Wu, L., Jiang, L., Zhou, Z., Fan, J., Zhang, Q., Zhu, H., Han, Q. & Xu, Z. (2003). Expression of Foot and mouth disease virus epitopes in tobacco by a Tobacco mosaic virus based vector. *Vaccine* **21,** 4390-4398.

Wyckoff, R. W. G. & Corey, R. B. (1936). X-ray diffraction pattern of crystalline Tobacco mosaic virus proteins. *J. Biol. Chem.* **116,** 51-55.

Yamaji, Y., Kobayashi, T., Hamada, K., Sakurai, K., Yoshii, A., Suzuki, M., Namba, S. & Hibi, T. (2006). In vivo interaction between Tobacco mosaic virus RNA-dependent RNA polymerase and host translation elongation factor 1A. *Virology* **347,** 100-108.

Yamashita, I., Hasegawa, K., Suzuki, H., Vonderviszt, F., Mimori-Kiyosue, Y. & Namba, K. (1998a). Structure and switching of bacterial flagellar filaments studied by X-ray fiber diffraction. *Nature Struct. Biol.* **5,** 125-132.

Yamashita, I., Suzuki, H. & Namba, K. (1998b). Multiple-step method for making exceptionally well-oriented liquid-crystalline sols of macromolecular assemblies. *J. Mol. Biol.* **278,** 609-615.

Yoon, J. Y., Min, B. E., Choi, S. H. & Ryu, K. H. (2001). Completion of nucleotide sequence and generation of highly infectious transcripts to cucurbits from full-length cDNA clone of Kyuri green mottle mosaic virus. *Arch. Virol.* **146,** 2085-96.

Yoon, J. Y., Min, B. E., Choi, J. K. & Ryu, K., H. (2002). Genome Structure and Production of Biologically Active In Vitro Transcripts of Cucurbit-Infecting Zucchini green mottle mosaic virus. *Phytopathology* **92,** 56-63.

Zanotto, P. M., de, A., Gibbs, M. J., Gould, E. A. & Holmes, E. C. (1996). A re-evaluation of the higher taxonomy of viruses based on RNA polymerases. *J. Virol.* **70,** 6083-6096.

Zhang, Z. C., Lei, C. Y., Zhang, L. F., Yang, X. X, Chen, R. & Zhang, D, S. (2008). The complete nucleotide sequence of a novel *Tobamovirus*, rehmannia mosaic virus. *Arch. Virol.* **153,** 595-599.

Zimmern. D. (1977). The nucleotide sequence at the origin for assembly on Tobacco mosaic virus RNA. *Cell* **11,** 463-482.