Article:
Huynh, NT, Hesketh, EL, Saxena, P et al. (7 more authors) (2016) Crystal structure and proteomics analysis of empty virus-like particles of Cowpea mosaic virus. Structure, 24 (4). pp. 567-575. ISSN 0969-2126

https://doi.org/10.1016/j.str.2016.02.011

Crystal structure and proteomics analysis of empty virus-like particles of Cowpea mosaic virus

Nhung Huynh[1], Emma L. Hesketh[3], Pooja Saxena[2§], Yulia Meshcheriakova[2], You-Chan Ku[1], Linh Hoang[4], John E. Johnson[1], Neil A. Ranson[3], George P. Lomonossoff[2] and Vijay S. Reddy[1*]

[1]Integrative Structural and Computational Biology, [4]Scripps Center for Metabolics and Mass Spectrometry, The Scripps Research Institute, La Jolla, CA, 92037, USA.

[2]Dept. of Biological Chemistry, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK.

[3]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

* Corresponding author: Vijay S. Reddy; E-mail address: reddyv@scripps.edu

[§]Present address: Department of Chemistry, Indiana University, 800 E. Kirkwood Ave., Bloomington, IN 47405-7102, USA.

SUMMARY

Empty virus-like particles (eVLPs) of Cowpea mosaic virus (CPMV) are currently being utilized as reagents in various biomedical and nanotechnology applications. Here, we report the crystal structure of CPMV eVLPs determined using X-ray crystallography at 2.3 Å resolution and compare it with previously reported cryo-electron microscopy (cryo-EM) of eVLPs and virion crystal structures. Although the X-ray and cryo-EM structures of eVLPs are mostly similar, there exist significant differences at the C-terminus of the small (S) subunit. The intact C-terminus of the S subunit plays a critical role in enabling the efficient assembly of CPMV virions and eVLPs, but undergoes proteolysis after particle formation. In addition, we report the results of mass spectrometry-based proteomics analysis of coat protein subunits from CPMV eVLPs and virions that identify the C-termini of S subunits undergo proteolytic cleavages at multiple sites instead of a single cleavage site as previously observed.

Keywords: Cowpea mosaic virus, CPMV, eVLPs, Structure, Mass spectrometry, Proteomics

Highlights:

The crystal structure of eVLPs is extremely similar to that of CPMV virions.

After assembly, the surface exposed C-termini of S subunits undergo proteolysis at multiple sites.

Presence of different ratios of processed S subunits results in slow and fast electrophoretic forms of CPMV.

INTRODUCTION

Cowpea mosaic virus (CPMV), the type member of the Comoviridae subfamily of the plant-infecting Secoviridae, consists of single-stranded, positive-sense, bipartite RNA genome of RNA-1 (6kb) and RNA-2 (3.5kb). The two RNAs are encapsidated into separate particles and both RNAs required for virus infection. The virus particles are ~310 Å in diameter and display pseudo T=3 icosahedral symmetry (**Fig. 1A, B**). Each particle consists of 60 copies of a protomer comprising a large (L) and a small (S) coat protein (CP) subunits derived from a single CP precursor (VP60) encoded by the RNA-2 and processed by the RNA-1 encoded 24K proteinase. The L subunit (41kDa) is composed of two jellyroll β-barrel domains, and the S subunit (23kDa) consists of a single jellyroll β-barrel (**Fig. 1C**).

Due to simplicity and robustness of its capsid and the availability of an atomic resolution structure (Lin et al., 1999; Stauffacher et al., 1987), CPMV has been widely used in bio-nanotechnology applications. These include the display of antigenic peptides on the particle surface (Dalsgaard et al., 1997; Porta et al., 1994; Usha et al., 1993), the chemical coupling of a variety of moieties (Aljabali et al., 2010; Aljabali et al., 2012; Steinmetz et al., 2006a, b) and the deposition of minerals on the particle surface (Aljabali et al., 2011; Shah et al., 2009; Steinmetz et al., 2009) as well as development of particles to deliver drugs coupled to their outer surface to cells (Aljabali et al., 2013). A particular advantage of CPMV for in vivo applications is that particles are well tolerated in mammals (Destito et al., 2009; Rae et al., 2005). However, a disadvantage of using CPMV particles produced via infection is that approximately 90% of such particles contain one or other genomic RNAs. As a result preparations of such particles retain their infectivity to plants, with attendant bio-containment concerns, and the particles cannot be efficiently loaded with cargo. To address these issue a system for the production of RNA-free

empty virus-like particles (eVLPs) has been developed. These can be generated by co-expression of the coat protein precursor VP60 along with the 24K viral proteinase in Nicotiana benthamiana (Montague et al., 2011; Saunders et al., 2009). CPMV eVLPs generated this way have already proven useful as reagents for bio- and nanotechnology applications (Lebedev et al., 2016; Sainsbury et al., 2011; Sainsbury et al., 2014; Wen et al., 2012).

To gain insight into the structural similarities between eVLPs and native CPMV virions, we have conducted structural studies on CPMV eVLPs using X-ray crystallography. Furthermore, this allows comparison with the recently determined cryo-EM structure of eVLPs at 3.0 Å, performed independently (Hesketh et al., 2015), and facilitates the correlation of eVLP structures determined using two different methods. Hence, here we report the crystal structure of eVLPs at 2.3 Å resolution and compare it with the cryo-EM structure of eVLPs (Hesketh et al., 2015) and the crystal structure of CPMV virion (PDB-ID: 1NY7) (Lin et al., 1999; Stauffacher et al., 1987). In addition, we have carried out mass spectrometry-based proteomics analysis on the S subunits from CPMV eVLPs and virions to investigate in detail the location of the proteolytic cleavage sites of the S subunit that result in the occurrence of the slow and fast electrophoretic forms of the virus (Niblett and Semancik, 1969). Each protomer of CPMV that contains 587 amino acid (a.a.) residues undergoes proteolytic processing at residues Gln374-Gly375 by the viral protease prior or during assembly generating the large (L) and small (S) subunits (Franssen et al., 1982). The full-length L and S subunits contain 374 and 213 a.a respectively. It has been shown that the intact C-terminus of the S subunit is required for the efficient assembly of both CPMV virions and eVLPs (Sainsbury et al., 2011; Taylor et al., 1999). After particle assembly, the surface exposed C-terminus is proteolysed up to residue Leu563 (189) (Taylor et al., 1999). In the virion crystal structure (PDB-ID: 1NY7), the last visible C-terminal residue of the S-

subunit, Leu 563 (189) is partially exposed (Lin et al., 1999) (we have chosen to use the continuous numbering system for the S subunit to be consistent with the sequence databases, instead of the old numbering that is shown in parenthesis and starts from the residue number 1, which was used in the description of the virion structure (Lin et al., 1999)). However the rate of loss of the C-terminal amino acids appears to be slower with eVLPs than with virus (Sainsbury et al., 2011), raising the possibility that some of residues may be visible in the X-ray structure of eVLPs.

The results presented in this manuscript demonstrate that the overall structure of eVLPs is extremely similar to that of virions produced via infection, whose crystal structure was determined previously (Lin et al., 1999). Furthermore, with the exception of the C-terminus of S subunit, the crystal structure of eVLPs reported in the current study shows close agreement with the independently obtained cryo-EM structure (Hesketh et al, 2015). The proteomics analysis of eVLPs and of virions has revealed that the C-terminus of the S subunit of CPMV undergoes proteolytic processing at multiple sites suggesting the possibility that CPMV eVLPs and virions are composed of S subunits of variable lengths. This is in agreement with previous observations that the presence of different proportions of processed and unprocessed C-termini of S subunits are likely to result in fast and slow electrophoretic forms of CPMV virions (Niblett and Semancik, 1969).

RESULTS

Crystal Structure Determination of eVLPs

Crystals of CPMV eVLPs were grown under the following conditions: 0.1 M Sodium acetate pH 4.68, 3-4%(w/v) PEG3350, 0.3 M Ammonium sulfate and 5% (v/v) glycerol, at 20°C. The crystals belong to space group $I4_1$, (a = 655.97 Å, b= 655.97 Å, c = 571.45 Å, $\alpha = 90°$, $\beta =$

90°, γ = 90°) with two particles in the crystallographic asymmetric unit. The eVLP crystals are distinct from the crystals of CPMV virions (of space group I23), grown under different conditions 0.05M potassium phosphate pH 7.0, 0.4M Ammonium sulfate, 2%(w/v) PEG8000 (Lin et al., 1999). The statistics associated with diffraction data and model refinement are shown in Table 1. Binwise statistics for various resolution bins are provided in the Table S1. The details of data collection, data processing, structure determination and model refinement are described in the experimental procedures section. Briefly, the structure was determined using molecular replacement method using the program Phaser (McCoy et al., 2007) and refined using the program CNS (Brunger et al., 1998). The R-factor of the eVLP structure is relatively high (Rwork: 0.357; Rfree: 0.359), but is consistent with the molecular replacement R-factor (0.351) calculated between the F_{obs} and the F_{calcs} derived from back-transformed electron density obtained by imposing 120-fold non-crystallographic symmetry (NCS). The high model R-factor is likely due to imposition of 120-fold strict NCS constraints, in a high-symmetry space group ($I4_1$), resulting from 2 particles present in the crystallographic asymmetric unit of a high symmetry space group ($I4_1$). Further discussion on this topic can be found in the methods section.

A majority of the crystal structure of CPMV eVLP was built in strong electron density at a contour level of 1.0σ **(Fig. 2A)** with the exception of the last two residues (566-567) at the C-terminus of S subunit. The contour level was reduced to 0.5σ in the omit maps to model the residues, 566-567 **(Fig. 2B, C)**. However, all the reported residues, 1-369 (L subunit) and 375-567 (S subunit) are clearly visible in the final refined maps **(Fig. 2B).** The C-terminal residues 370-374 of the L subunit and the residues 568-587 (194-213) of the S subunit are disordered in the crystal structure. It is possible that some of the latter unidentified residues of the S subunit

could have been proteolytically removed prior to crystallization and hence likely to be absent in the crystals of eVLPs.

CPMV eVLP Structure and Particle Organization

The overall structure, size and organization of CPMV eVLPs is similar to that of the CPMV virion structure (Lin et al., 1999). The CPMV protomer contains 587 a.a. residues and 60 copies of the protomer form an icosahedral particle with T=1 symmetry and an averaged diameter of ~310 Å. However, the same (T=1) icosahedral architecture can be described as displaying pseudo T=3 (pT=3) icosahedral symmetry, as the three β-barrel domains of the protomer occupy comparable positions seen in a T=3 quasi-equivalent surface lattice (**Fig. 1A, B**). Each protomer undergoes proteolytic processing at residues Gln374-Gly375 by the 24K viral protease prior or during assembly generating the large (L) and small (S) subunits (**Fig. 1C**). In the eVLP structure, the L subunit (41kDa) is composed of ordered residues 1-369 that form two jellyroll β-barrel domains, while the S subunit (23kDa) consisting of ordered residues 375-567 (1-193) forms a single β-barrel domain. The β-barrel (β3) of the small subunit is located near the five-fold axis, with the β-barrels (β1 and β2) of the large subunit surrounding the three-fold axis (**Fig. 1A, B**).

Comparison of CPMV eVLP and Virion Structures

The crystal and cryo-EM structures of eVLP show a high degree of similarity with each another and also with the CPMV virion crystal structure. The structures have an overall root-mean-square deviation (rmsd) of 0.518 Å between eVLP-X-ray and eVLP-EM; 0.527 Å between eVLP-X-ray and CPMV-virion (X-ray); and 0.445 Å between eVLP-EM and CPMV-virion (X-ray) for 550 aligned Cα atoms **(Fig. 3A)**. The structure superpositions and calculation of rmsd

values were carried out using the UCSF Chimera program (Pettersen et al., 2004) and the graphics program O (Jones et al., 1991). **Figure 4A-F** show the rmsd values for the large and small subunits between the crystal and cryo-EM structures. Other than at the C-terminus of the S subunit, minor differences are observed at the residue Lys 176 of the large subunit and Asp 418 (44) of the small subunit between the eVLP and virion structures due to differences in rotamer and alternate backbone conformations, respectively **(Fig. 4G, H)**. However, the similar side chain conformation seen for the residue Lys 176 in both the eVLP structures compared to the virion structure (**Fig. 4G**) and its location on the capsid interior, near the icosahedral 2-fold axis, suggests that it might potentially interact with genomic RNA in CPMV virions. Other small structural differences with rmsds in the range of 1-2Å occur due to small changes in the residue locations (**Fig. S2**). Notwithstanding these differences, the eVLP and virion structures of CPMV appear to be very similar. We also compared the average B-factors per residue of eVLP and virion structures. Even though, the magnitudes of B-factors differ between the two capsids, the variation of B-factors along the polypeptide chain is very similar between the two structures (**Fig. S3**). This suggests that there are no significant differences between in the dynamics of eVLPs and virions, influenced by the absence or presence of genomic RNA, respectively.

The major difference between the X-ray and cryo-EM structures of eVLPs is seen at the C-terminus of the S subunit. A helical segment of residues 564-576 (190-202) is ordered only in the eVLP cryo-EM structure, with preceding residues 558-563 (184-189) and succeeding residues 577-587 (203-213) that could not be modeled due to weak or disordered density **(Fig. 3B)** (Hesketh et al., 2015). In contrast, the C-termini of the S subunit are continuously ordered until residues Arg567 (193) and Leu563 (189) in the crystal structures of eVLPs and CPMV virion, respectively **(Fig. 3B)** and in spite of different conditions used to grow crystals of eVLPs

(pH 4.7) and virions (pH 7.0). Of note, the cryo-EM structure of eVLPs was determined at pH 7.0 (Hesketh et al., 2015). Furthermore, the coiled-structure formed by the visible C-terminal residues 557-567 of the S subunit, in the eVLP crystal structure, is consistent with secondary structure predictions (Buchan et al., 2013) (Fig. S1). Notably, the ordered helical segment of residues 568-576 (194-202) in the cryo-EM structure spatially superimposes onto residues 562-567 (188-193) in the eVLP crystal structure (**Fig. 3B**). These structural differences seen at C-terminus of the S subunit between the crystal and cryo-EM structures of the eVLPs may be attributable to different pHs 4.7 and 7.0 and/or different extents of proteolysis undergone by the eVLPs used in the respective studies.

Mass Spectrometry and Proteomics

The S subunit of CPMV eVLPs and virions is present in slow and fast electrophoretic forms. The slow form contains the full-length S subunits, while the fast form consists of proteolysed form(s) of S subunits. In freshly prepared samples of eVLPs and virions, which were dialyzed into 50mM Potassium Phosphate buffer (pH 7.0), both the slow and fast forms are present **(Fig. 5A)**. However, within 2-3 weeks after purification of the eVLPs stored at 4°C, the S subunit contains mostly the fast (cleaved) form (Sainsbury et al., 2011). The extent of proteolysis is also found to be different based on the type of plant host used for the expression of CPMV or eVLPs, such as cowpea (Vigna unguiculata) compared with N. benthaminana (**Fig. 5B**). Mass spectrometry and proteomics analysis of the electrophoretic slow and fast forms (bands) of the S subunit reveal that the slow (uncleaved) form contains the native C-terminus ending with the sequence Thr-Ala-Ala (**Fig. 5C**), while the fast form contains peptides with different C-termini ending with residue Leu563 (189), Leu564 (190) or Phe566 (192). The proteomics analysis on the fast form of S subunit of CPMV virions has identified the proteolysis

at Ph566 in addition to at Leu563 and Leu564 (**Fig. 5C**). It is important to emphasize that the slow electrophoretic band of the S subunit contains only the full length S subunit (213 a.a.), unlike the fast form that displays different C-termini and composed of a varying number of amino acids between 189-192.

DISCUSSION

CPMV particles are being used as reagents for various biomedical and nanotechnology applications. More recently, these applications have turned to using eVLPs rather than the infectious virus (Lebedev et al., 2016; Sainsbury et al., 2011; Sainsbury et al., 2014; Wen et al., 2012). Therefore, it is of importance to structurally characterize eVLPs and analyze any significant differences that occur between the wild-type virus and eVLPs of CPMV. Previous studies have shown that the C-terminal 24-residues of the S subunit are important for particle assembly (Hesketh et al., 2015; Sainsbury et al., 2011), RNA encapsidation (Taylor et al., 1999) and the suppression of gene silencing (Canizares et al., 2004). Based on the location of the visible C-terminus (567) of the S subunit (**Fig. 1B**), the C-terminal 24-residues are likely to play crucial role in stabilizing the inter-protomer contacts around the 5-fold axis during assembly. These protomer-protomer associations lead to the formation of the pentamers of the protomers, and 12 such pentamers assemble together to form CPMV particles. Since, RNA encapsidation of CPMV likely to involve assembling of pentamers around the viral RNA, the C-terminal residues would also play an indirect role in RNA packaging. This assessment is in agreement with the assembly mechanism proposed by the cryo-EM studies (Hesketh et al., 2015). After assembly, the C-termini of S subunits of CPMV virions are prone to proteolysis and lose up to 24 residues without adversely affecting the particle integrity (Geelen et al., 1973; Niblett and Semancik, 1970; Taylor et al., 1999). In freshly prepared samples, a significant proportion of the C-termini

of the S subunits are intact (Sainsbury et al., 2011), resulting in the occurrence of slow and fast electrophoretic forms of CPMV particles (Niblett and Semancik, 1969). However, most of the S subunits lose upto 24 a.a from their C-termini by about 2-3 weeks after the purification, when stored at 4°C (Sainsbury et al., 2011).

The proteomics analysis on different electrophoretic forms (bands) of the S subunits of CPMV eVLPs as well as of virions suggests that the fast (cleaved) form of S subunits contain different or multiple C-terminal residues: Leu563 (189), Leu564 (190) and Phe566 (192), as opposed to a single cleavage site as previously reported (Taylor et al., 1999). The previous results were designed to ascertain the limit of cleavage and were obtained from a preparation of virions, produced in cowpea, which was deliberately aged for several weeks. Furthermore, the technique used, electrospray mass spectrometry of the entire S protein, did not permit the resolution of the results reported here. The observation that the fast form contains different C-terminal residues implies that the CPMV eVLPs or virions might consist of the S subunits with their C-termini processed to varying lengths. However, the final cleavage site cannot be too far from Leu563 (189), otherwise the residues would be surface exposed and susceptible to further proteolysis. This is consistent with the observation that none of the peptides from fast form of the S subunit contain C-terminal residues beyond Phe566 (192) (**Fig. 5C**).

The superposition of eVLP structures onto the virion structure of CPMV reveals a high degree of structural similarity with the exception of differences at the C-terminus of the S subunit (**Fig. 3A**). This shows that the CPMV eVLPs are able to assemble independently in the absence of nucleic acid and form capsids virtually identical to CPMV virion particles. The eVLP crystal structure shows continuous electron density until Arg567 (193), four residues beyond the

last cleavage seen at residue Leu563 (189) of the S subunit, identified by the proteomics analysis (**Fig. 5C**). A bifurcated salt-bridge is observed in the eVLP crystal structure between residue Glu559 (185) at the C-terminus of the reference S subunit and the residues Arg469 (95) and Asp478 (104) of the S subunit of the 5-fold related (clockwise) protomer that stabilizes the pentameric interface. The identification of a few additional residues beyond the limit of the cleavage site found in virions indicates that the X-ray structure represents an average structure of eVLPs that are composed of cleaved, partially cleaved and possibly uncleaved forms of the S subunit. The remaining residues beyond Arg567 (193) are either proteolysed and totally absent from the crystals used for structure determination or if present are disordered in the crystal structure. In addition, the identification of peptide fragments with residues downstream of Leu563 (189) in the proteomics analysis (**Fig. 5C**) suggests that the proteolysis may be either sequential in nature or may occur at several different sites and independently on different S subunits, as there are 60 such sites per particle. It is likely that after particle assembly, the C-termini of the small subunit are continually proteolysed by host proteases at various basic or aromatic residues until the last cleavable/accessible residue Leu564 (190) or a residue close to it is reached.

In the cryo-EM structure of eVLPs, a helical segment of thirteen residues 564-576 (190-202) of the S subunit is found to be ordered beyond the last proteolysed residue Leu563 (189) (Hesketh et al., 2015). However, some of the residues, Leu564 (190) - Arg569 (195), in this peptide segment that are known to undergo proteolysis, according to above proteomics results, are buried and inaccessible in the cryo-EM structure. Hence, if these terminal residues as seen in the cryo-EM structure were to become accessible and undergo proteolytic processing, it would require conformational reorganization of the C-terminus. Indirect evidence that such a

conformational change occurs on proteolysis is provided by the observation that deliberate digestion of eVLPs with chymotrypsin, which cuts after amino acids 565 (191), (567) 193 and 569 (195) of the S protein, sites similar those identified in this study, , significantly enhances their permeability to ions compared to the untreated eVLPs (Sainsbury et al., 2011).

The results presented in this manuscript demonstrate that the overall structure of eVLPs is extremely similar to that of virions produced via infection, whose crystal structure was determined previously (Lin et al., 1999). Furthermore, based on the residuewise B-factor comparisons along the polypeptide chain, the particle dynamics of CPMV virions and eVLPs appear to be similar irrespective of whether or not the genomic RNA is packaged (**Fig. S3**). With the exception of the C-terminus of S subunit, the crystal structure of eVLPs reported in this study shows close agreement with the independently obtained cryo-EM structure (Hesketh et al, 2015). Taken together, the information obtained from X-ray and cryo-EM structures along with the proteomics analysis of eVLPs has provided new insights into the structure and processing of the C-terminus of the S subunit of CPMV as well as particle assembly. Such knowledge is important for developing eVLPs for biotechnological applications as this region has been known to control assembly and accessibility (Sainsbury et al., 2011).

EXPERIMENTAL PROCEDURES

Expression, Purification and Crystallization of CPMV eVLPs

CPMV eVLPs were expressed and purified as previously described (Sainsbury et al., 2014). CPMV eVLPs, at a concentration of 6.0 mg/ml, were crystallized by hanging drop vapor

diffusion by mixing 2µl of eVLPs with 2µl of reservoir solution containing 3-4%(w/v) PEG3350, 0.3 M Ammonium sulfate, 5% (v/v) glycerol and 0.1 M Sodium acetate pH 4.68 and stored at room temperature (20°C). Diamond shaped crystals appeared within 1-2 days and grew to a maximum size of ~0.3mm in about 1-2 weeks. The crystals were sequentially transferred into a reservoir solution containing 10%, 15% and 20% (v/v) and flash cooled in liquid nitrogen prior to data collection.

Data Collection, Structure Determination, Model Building, and Refinement

The diffraction data from the CPMV-eVLP crystals were collected on a Pilatus detector at the beamline 23-ID-D, GMCA at APS, Chicago. The crystals diffracted better than 2.3 Å resolution. The data was processed using HKL2000 (Otwinowski and Minor, 1997). The final data scaling and merging of the unmerged data derived from SCALEPACK (HKL2000) was done using the SCALA program (Evans, 2006). All the reflections with partiality > 0.6 were included in the dataset. Details of the data statistics are shown in Table 1 and Table S1. The CPMV eVLP crystals belong to the space group $I4_1$ with unit cell dimensions **a**=**b**=655.97 Å, **c**=571.45 Å, **α**=**β**=**γ**=90° and contain two virus particles in the crystallographic asymmetric unit (16 particles in the unit cell), resulting in 120-fold non-crystallographic symmetry (NCS). The CMPV eVLP structure was determined by the molecular replacement method employing the program Phaser (McCoy et al., 2007) and using the poly-alanine model derived from the CMPV structure (PDB-ID: 1NY7). The details of the particle orientations and positions are given in Table 1. The latter polyalanine model was oriented and positioned according to the Phaser solution and rigid body refined using the program X-PLOR (Brunger, 1992). The Initial electron density was generated with the phases from the rigid body refined polyalanine model and improved by employing the phase refinement procedure with 120-fold NCS averaging using the

RAVE and CCP4 suite of programs (CCP4, 1994; Jones, 1992). This resulted in a readily interpretable map at 2.3 Å with a final R-factor of 0.351 and a correlation coefficient of 0.68, calculated between the F_{obs} and the F_{calcs} derived from back-transformed electron density. An initial model of CPMV protomer that occupies the icosahedral asymmetric unit, was built automatically using BUCCANEER program (Cowtan, 2008). The final model was adjusted manually into the electron density map contoured at 1.0σ using the programs COOT (Emsley et al., 2010) and O (Jones et al., 1991). The structure was refined using the program CNS imposing (strict) 120-fold NCS constraints (Brunger et al., 1998). Water molecules were identified using the program WATPEAK (CCP4, 1994) based on the difference (Fo-Fc) densities and later manually verified and adjusted using the graphics program COOT (Emsley et al., 2010). In the end 58 water molecules were included in the eVLP structure. The final refinement statistics are shown in Table 1. Model quality was analyzed using PROCHECK (Laskowski et al., 1993) and Stride (Heinig and Frishman, 2004) was used for the assignment of secondary structures.

The high model R-factor is likely due to imposition of 120-fold NCS constraints, resulting from 2 particles present in the crystallographic asymmetric unit of high symmetry space group ($I4_1$). To improve the R-factor, it may be necessary to release the strict NCS constraints and impose them as restraints at high resolutions (e.g, 2.3 Å), particularly when more than one particle is present in the crystallographic asymmetric unit. This phenomenon is even more pronounced in a high symmetry space group relative to low symmetry space group P1 as seen in few cases (Khayat et al., 2011; Tars et al., 2000a, b). Even though NCS restraints were used in the refinement of simple (T=1) viruses, where 60-fold NCS was present (Larson et al., 2014), it is difficult to implement the same with 120-fold NCS because of a greater number of chains (120x2 = 240) involved. In the end, this may not significantly improve the model except for

slightly better statistics (i.e., R-factor), as the model R-factor (0.359) is very similar to molecular replacement R-factor (0.351) calculated between the F_{obs} and the F_{calcs} derived from back-transformed electron density obtained by imposing 120-fold NCS. We tested if the imposition of 120-fold NCS constraints is the reason for high R-factor at 2.3Å, by calculating the R-factors at different high-resolution limits of the data. The R-factors calculated at various high-resolution limits of 3 Å, 4 Å and 5 Å yielded increasingly lower R-factors: 0.325, 0.294 and 0.291, respectively. This indicates that the imposition of 120-fold strict NCS is likely the cause of relatively higher R-factors at 2.3 Å resolution.

Mass Spectrometry and Proteomics

The capsid protein subunits (L and S) of CPMV virus particles from Cowpea and CPMV eVLPs were separated on a SDS PAGE gel as previously described (Sainsbury et al., 2011). The S subunit of CPMV virions and that of eVLPs contained two bands corresponding to the slow and fast forms (Fig. 5A,B). The bands corresponding to the slow and fast forms of the S subunit of eVLPs and of the fast form (only) of the virions were excised and submitted for proteomics mass spectrometry analysis by digestion with GluC protease. The excised gel bands were destained with a mixture of a 50:50 (v/v) dilution of acetonitrile-25 mM ammonium bicarbonate and dried with a SpeedVac system. The protein samples in the gel were then reduced in 25 µl of 10 mM d,l-dithiothreitol (Sigma) for 1 h and alkylated with 25 µl of 55 mM iodoacetamide (Sigma) for 30 min in the dark prior to an 18-h GluC digestion at 37°C using a 1:30 (w/w) enzyme-to-substrate ratio. The resulting peptides were extracted twice with a 50:45:5 (v/v) volume of acetonitrile-water-formic acid and concentrated to 30 µl before being analyzed. Peptides were analyzed by reverse-phase chromatography prior to mass spectrometry analysis

and MS/MS samples were analyzed by using Mascot (version 2.1.04; Matrix Science, London, United Kingdom) as described in Powers et al (Powers et al., 2011).

ACCESSION NUMBERS

The coordinates and structure factor amplitudes of the eVLP crystal structure have been deposited in the PDB with the accession code **5fmo**.

Acknowledgements

We acknowledge support from National Institutes of Health (NIH) grant AI070771 (to V.S.R), UK Biotechnological and Biological Sciences Research Council (BBSRC) Grant BB/L020955/1, the Institute Strategic Programme Grant 'Understanding and Exploiting Plant and Microbial Secondary Metabolism' (BB/J004596/1) and the John Innes Foundation (to G.P.L). GM/CA has been funded with US federal funds from the NIH (Y1-CO-1020, Y1-GN-11040) and the APS is supported by the US Department of Energy under contract DE-AC02-06CH11357.

Conflict of interest

GPL declares that he is a named inventor on granted patent WO 29087391 A1 which describes the transient expression system used in this manuscript to express eVLPs and also a named inventor on a filed patent "Production of viral capsids" PBL. US 61/186,970 which describes the production and uses of CPMV eVLPs.

Table 1. Data collection and Refinement Statistics

X-ray Source	23-ID-D, GMCA, APS
Wavelength (Å)	0.858
Space Group	$I4_1$
Cell parameters	a = 655.97 Å, b= 655.97 Å, c = 571.45 Å, α = 90°, β = 90°, γ = 90°
Number of particles	16/unit cell; 2/crystal asymmetric unit;
V_M (Å^3/Da)	2.75
Percent Solvent (%)	55
Resolution (Å)	50-2.3 (2.42 – 2.3)[§]
No. of unique reflections	3,192,399 (383,718) [§]
Completeness (%)	60.3 (49.6) [§]
R_{merge}[a] (%)	0.29 (0.452) [§]
I/σ	3.7 (1.7) [§]
Particle orientations (Phaser solution)	Particle-1: EULER 39.90, 71.85, 333.96 FRAC 0.252, 0.00, 0.0000 Particle-2: EULER 74.83, 89.08, 189.73 FRAC 0.500, 0.249, 0.253
Ordered residues	1-369 (L); 375-567 (S)
Waters	58
Refinement Statistics	
Resolution (Å)	20-2.3
No. of reflections	1,861,715
Degrees of NCS	120
R_{work}[b]	0.357
R_{free}[c]	0.359
RMS bond length (Å)	0.009
RMS bond angle (°)	1.527
Ramachandran Plot Statistics	
Most Favorable (%)	78.2
Allowed (%)	2.6
Disallowed (%)	0

[a]$R_{merge} = [\sum_h\sum_i|\mathbf{I}_h - \mathbf{I}_{hi}|/\sum_h\sum_i\mathbf{I}_{hi}]$ where $\mathbf{I}_h$ is the mean of $\mathbf{I}_{hi}$ observations of reflection h. Numbers in

[§]Shown in parenthesis are the statistics for the highest resolution shell.

[b]R_{work} and [c]$R_{free} = \sum||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|$ x 100 for 95% of recorded data ([b]R_{work}) or 5% data ([c]R_{free})

FIGURE LEGENDS

Figure 1. Structure and organization of CPMV eVLPs. A) An icosahedral (pseudo) T=3 cage showing the geometric organization of the large (L) and small (S) subunits in a CPMV capsid, shown in green and blue, respectively. The two β-barrels (β1 □□□ β2) of the L subunit encircle the 3-fold axis denoted by a yellow triangle. The β-barrel (β3) of the S subunit surround the icosahedral 5-fold vertices, identified by a yellow pentagon. The yellow diamond represents the location of icosahedral 2-fold axis. B) A surface representation of the subunit organization of the L and S subunits. The C-terminal residues (561-567) of the S subunit are shown in red colored surface. C) A ribbon diagram of the CPMV protomer depicting the secondary and tertiary structures of the L and S subunits. The N-termini of the L and S subunits are labeled as L-N and S-N, respectively and the visible C-termini are correspondingly labeled as L-C and S-C, respectively. Figures were generated using Chimera (Pettersen et al., 2004) and Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.)

Figure 2. Representative electron density of the eVLP crystal structure. A) The (3Fo-2Fc) electron density illustrated by blue mesh at residue Tyr100 contoured at 1.0σ. B) The continuous (3Fo-2Fc) density at the C-terminus of the S subunit comprising the terminal 16 residues (551-567) in the crystal structure. C) An omit map showing the density at the C-terminus of the S subunit, calculated by omitting residues 554-567. The β□”-strand of residues 421-425, adjacent to the C-terminal residues is shown as reference.

Figure 3. Superposition of CPMV protomers. A) A superposition of the eVLP-Xray (yellow), eVLP-cryoEM (blue) and native CPMV virion-X-ray (1NY7; green) structures. B) The close-up view of the superposition at the C-termini of the S subunits from three structures. Residues are

labeled in corresponding colors of the structures. A dashed line represents disordered residues, 558-563 (184-189) in the eVLP-cryoEM structure.

Figure 4. Graphs showing the root-mean-square deviation (rmsd) values between the crystal and cryoEM structures highlighting structural similarities and differences. Graphs: **A, B and C** show the rmsd values at each residue between the L subunits of the eVLP-Xray and CPMV (virion)-Xray (1NY7) ; eVLP-Xray and eVLP-cryoEM; eVLP-cryoEM and CPMV-Xray (1NY7) structures, respectively. The Cα atom deviations are shown by blue line and the average deviation of all atoms in each residue is indicated by red line. Graphs **D, E and F** show the rmsd values at each residue between the S subunits for the corresponding pairs as in graphs A, B and C, respectively. The residue numbers of S subunits, in graphs D, E and F, are shown from 1-193, which corresponds to residues 375-567. **G**) The side chain (rotamer) differences at residue Lys 176 of the L subunit between the three structures: eVLP-Xray (yellow), eVLP-cryoEM (blue) and CPMV-Xray (1NY7; green). **H**) The conformational differences in the backbone of the structures at residue Asp418 (44) of the S subunit.

Figure 5. SDS PAGE and proteomic analysis of CPMV eVLPs and virions. A) CPMV eVLPs, purified from N. benthamiana, Lane 1: molecular weight marker, Lane 2-CPMV eVLPs fresh prep, 3- CPMV eVLPs after storage at 4°C for more than 1 month (less slow forms of S subunit). The L subunit has a molecular weight of 41kDa and the S is subunit 23kDa. The lower case s and f letters (in gray) refer to slow (uncleaved) and fast (cleaved) bands of the S subunit. B) CPMV virion, Lane 1: molecular weight marker, Lane 2-4: samples purified from N. benthamiana and Lane 5-7: samples purified from cowpea plants. Different levels of proteolytic cleavage of S subunit are observed depending on the host plant. Samples from N. benthamiana look more

intact compared to samples from cowpea plants. The lower case s and f letters (in gray) refer to slow (uncleaved) and fast (cleaved) bands of the S subunit. C) List of identified peptides of the S subunit of CPMV virus and eVLPs after GluC protease digestion using proteomics mass spectrometry analysis. The cleavage point is indicated by a period (dot). The slow and fast forms of the S subunit were individually digested and the peptide sequences were identified. Selected amino acids are identified by their residue numbers. The sequence ending with TAA (587) corresponds to the native C-terminus of the full length of the S subunit (slow form).

REFERENCES

Aljabali, A.A., Barclay, J.E., Butt, J.N., Lomonossoff, G.P., and Evans, D.J. (2010). Redox-active ferrocene-modified Cowpea mosaic virus nanoparticles. Dalton Trans *39*, 7569-7574.

Aljabali, A.A., Barclay, J.E., Steinmetz, N.F., Lomonossoff, G.P., and Evans, D.J. (2012). Controlled immobilisation of active enzymes on the cowpea mosaic virus capsid. Nanoscale *4*, 5640-5645.

Aljabali, A.A., Shah, S.N., Evans-Gowing, R., Lomonossoff, G.P., and Evans, D.J. (2011). Chemically-coupled-peptide-promoted virus nanoparticle templated mineralization. Integr Biol (Camb) *3*, 119-125.

Aljabali, A.A., Shukla, S., Lomonossoff, G.P., Steinmetz, N.F., and Evans, D.J. (2013). CPMV-DOX delivers. Mol Pharm *10*, 3-10.

Brunger, A.T. (1992). X-PLOR, version 3.1: A system for X-ray crystallography and NMR (New Haven, Connecticut: Yale Univ. Press).

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr *54*, 905-921.

Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res *41*, W349-357.

Canizares, M.C., Taylor, K.M., and Lomonossoff, G.P. (2004). Surface-exposed C-terminal amino acids of the small coat protein of Cowpea mosaic virus are required for suppression of silencing. J Gen Virol *85*, 3431-3435.

CCP4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *50*, 760-763.

Cowtan, K. (2008). Fitting molecular fragments into electron density. Acta Crystallogr D Biol Crystallogr *64*, 83-89.

Dalsgaard, K., Uttenthal, A., Jones, T.D., Xu, F., Merryweather, A., Hamilton, W.D., Langeveld, J.P., Boshuizen, R.S., Kamstrup, S., Lomonossoff, G.P.*, et al.* (1997). Plant-derived vaccine protects target animals against a viral disease. Nat Biotechnol *15*, 248-252.

Destito, G., Schneemann, A., and Manchester, M. (2009). Biomedical nanotechnology using virus-based nanoparticles. Curr Top Microbiol Immunol *327*, 95-122.
Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.
Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.
Franssen, H., Goldbach, R., Broekhuijsen, M., Moerman, M., and van Kammen, A. (1982). Expression of Middle-Component RNA of Cowpea Mosaic Virus: In Vitro Generation of a Precursor to Both Capsid Proteins by a Bottom-Component RNA-Encoded Protease from Infected Cells. J Virol *41*, 8-17.
Geelen, J.L., Rezelman, G., and van Kammen, A. (1973). The infectivity of the two electrophoretic forms of cowpea mosaic virus. Virology *51*, 279-286.
Heinig, M., and Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res *32*, W500-502.
Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., Saxena, P., Thompson, R.F., Cockburn, J.J., Lomonossoff, G.P., and Ranson, N.A. (2015). Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. Nat Commun *6*, 10113.
Jones, T.A. (1992). a. yaap. asap, @#*? A set of Averaging Programs. In Molecular Replacement, E.J. Dodson, Gover, S. & Wolf, W., ed. (Warrington: SERC Daresbury Laboratory).
Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta crystallographica *47 (Pt 2)*, 110-119.
Khayat, R., Brunn, N., Speir, J.A., Hardham, J.M., Ankenbauer, R.G., Schneemann, A., and Johnson, J.E. (2011). The 2.3-angstrom structure of porcine circovirus 2. J Virol *85*, 7856-7862.
Larson, S.B., Day, J.S., and McPherson, A. (2014). Satellite tobacco mosaic virus refined to 1.4 A resolution. Acta Crystallogr D Biol Crystallogr *70*, 2316-2330.
Laskowski, R.A., MacArthur, M.W., and Thornton, J.M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst. *26*, 283-291.
Lebedev, N., Griva, I., Dressick, W.J., Phelps, J., Johnson, J.E., Meshcheriakova, Y., Lomonossoff, G.P., and Soto, C.M. (2016). A virus-based nanoplasmonic structure as a surface-enhanced Raman biosensor. Biosens Bioelectron *77*, 306-314.
Lin, T., Chen, Z., Usha, R., Stauffacher, C.V., Dai, J.B., Schmidt, T., and Johnson, J.E. (1999). The refined crystal structure of cowpea mosaic virus at 2.8 A resolution. Virology *265*, 20-34.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.
Montague, N.P., Thuenemann, E.C., Saxena, P., Saunders, K., Lenzi, P., and Lomonossoff, G.P. (2011). Recent advances of cowpea mosaic virus-based particle technology. Hum Vaccin *7*, 383-390.
Niblett, C.L., and Semancik, J.S. (1969). Conversion of the electrophoretic forms of cowpea mosaic virus in vivo and in vitro. Virology *38*, 685-693.
Niblett, C.L., and Semancik, J.S. (1970). The significance of the coat protein in infection by the electrophoretic forms of cowpea mosaic virus. Virology *41*, 201-207.
Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307-326.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry *25*, 1605-1612.

Porta, C., Spall, V.E., Loveland, J., Johnson, J.E., Barker, P.J., and Lomonossoff, G.P. (1994). Development of cowpea mosaic virus as a high-yielding system for the presentation of foreign peptides. Virology *202*, 949-955.

Powers, M.E., Smith, P.A., Roberts, T.C., Fowler, B.J., King, C.C., Trauger, S.A., Siuzdak, G., and Romesberg, F.E. (2011). Type I signal peptidase and protein secretion in Staphylococcus epidermidis. J Bacteriol *193*, 340-348.

Rae, C.S., Khor, I.W., Wang, Q., Destito, G., Gonzalez, M.J., Singh, P., Thomas, D.M., Estrada, M.N., Powell, E., Finn, M.G.*, et al.* (2005). Systemic trafficking of plant virus nanoparticles in mice via the oral route. Virology *343*, 224-235.

Sainsbury, F., Saunders, K., Aljabali, A.A., Evans, D.J., and Lomonossoff, G.P. (2011). Peptide-controlled access to the interior surface of empty virus nanoparticles. Chembiochem *12*, 2435-2440.

Sainsbury, F., Saxena, P., Aljabali, A.A., Saunders, K., Evans, D.J., and Lomonossoff, G.P. (2014). Genetic engineering and characterization of Cowpea mosaic virus empty virus-like particles. Methods Mol Biol *1108*, 139-153.

Saunders, K., Sainsbury, F., and Lomonossoff, G.P. (2009). Efficient generation of cowpea mosaic virus empty virus-like particles by the proteolytic processing of precursors in insect cells and plants. Virology *393*, 329-337.

Shah, S.N., Steinmetz, N.F., Aljabali, A.A., Lomonossoff, G.P., and Evans, D.J. (2009). Environmentally benign synthesis of virus-templated, monodisperse, iron-platinum nanoparticles. Dalton Trans, 8479-8480.

Stauffacher, C.V., Usha, R., Harrington, M., Schmidt, T., Hosur, M., and Johnson, J.E. (1987). The structure of cowpea mosaic virus at 3.5Å resolution. In Crystallography in Molecular Biology, D. Moras, J. Drenth, B. Strandberg, D. Suck, and L. Wilson, eds. (New York: Plenum Publishing Corporation), pp. 293-308.

Steinmetz, N.F., Lomonossoff, G.P., and Evans, D.J. (2006a). Cowpea mosaic virus for material fabrication: addressable carboxylate groups on a programmable nanoscaffold. Langmuir *22*, 3488-3490.

Steinmetz, N.F., Lomonossoff, G.P., and Evans, D.J. (2006b). Decoration of cowpea mosaic virus with multiple, redox-active, organometallic complexes. Small *2*, 530-533.

Steinmetz, N.F., Shah, S.N., Barclay, J.E., Rallapalli, G., Lomonossoff, G.P., and Evans, D.J. (2009). Virus-templated silica nanoparticles. Small *5*, 813-816.

Tars, K., Fridborg, K., Bundule, M., and Liljas, L. (2000a). Structure determination of bacteriophage PP7 from Pseudomonas aeruginosa: from poor data to a good map. Acta Crystallogr D Biol Crystallogr *56*, 398-405.

Tars, K., Fridborg, K., Bundule, M., and Liljas, L. (2000b). The three-dimensional structure of bacteriophage PP7 from Pseudomonas aeruginosa at 3.7-A resolution. Virology *272*, 331-337.

Taylor, K.M., Spall, V.E., Butler, P.J., and Lomonossoff, G.P. (1999). The cleavable carboxyl-terminus of the small coat protein of cowpea mosaic virus is involved in RNA encapsidation. Virology *255*, 129-137.

Usha, R., Rohll, J.B., Spall, V.E., Shanks, M., Maule, A.J., Johnson, J.E., and Lomonossoff, G.P. (1993). Expression of an animal virus antigenic site on the surface of a plant virus particle. Virology *197*, 366-374.

Wen, A.M., Shukla, S., Saxena, P., Aljabali, A.A., Yildiz, I., Dey, S., Mealy, J.E., Yang, A.C., Evans, D.J., Lomonossoff, G.P.*, et al.* (2012). Interior engineering of a viral nanoparticle and its tumor homing properties. Biomacromolecules *13*, 3990-4001.

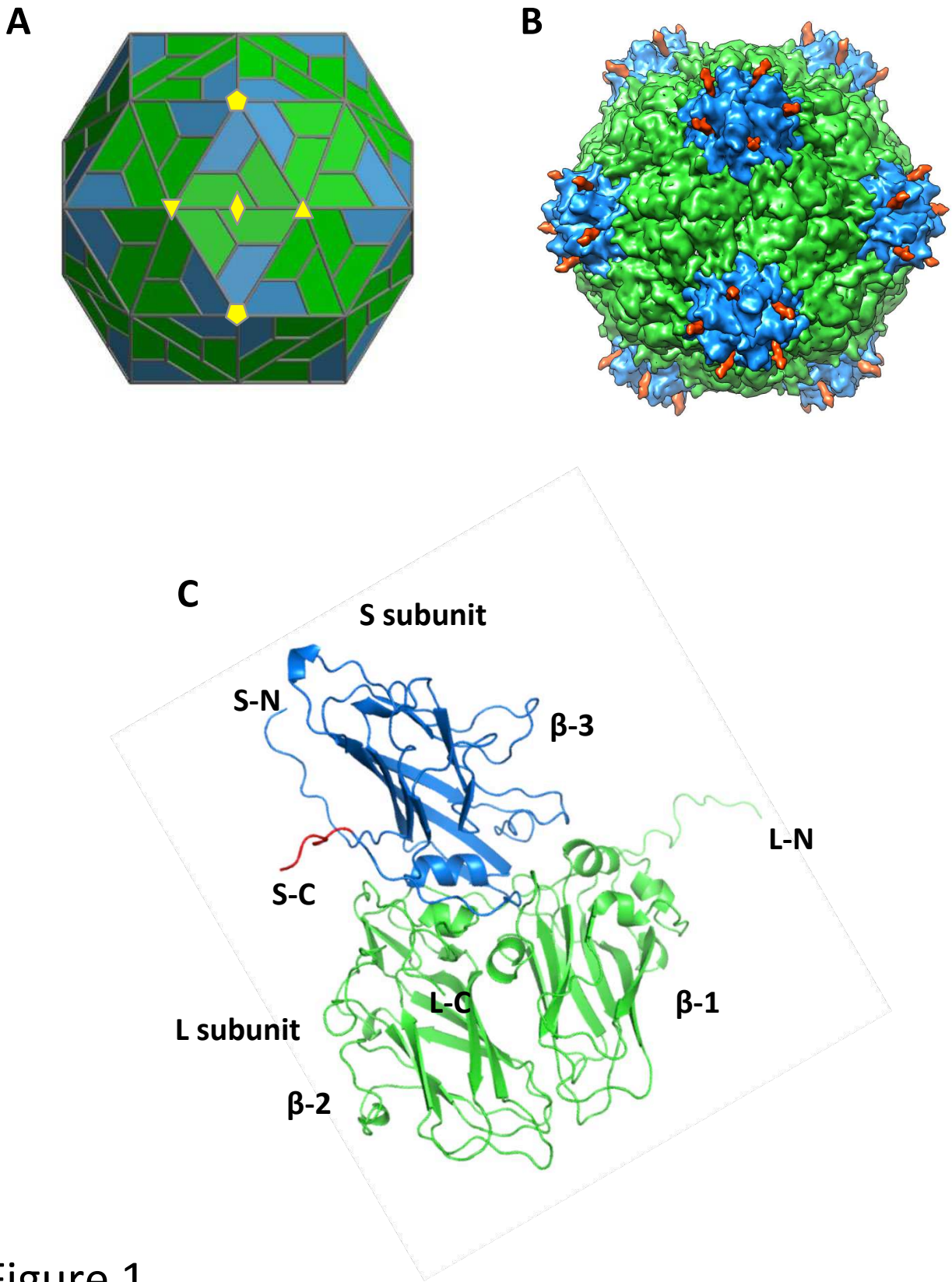


Figure 1

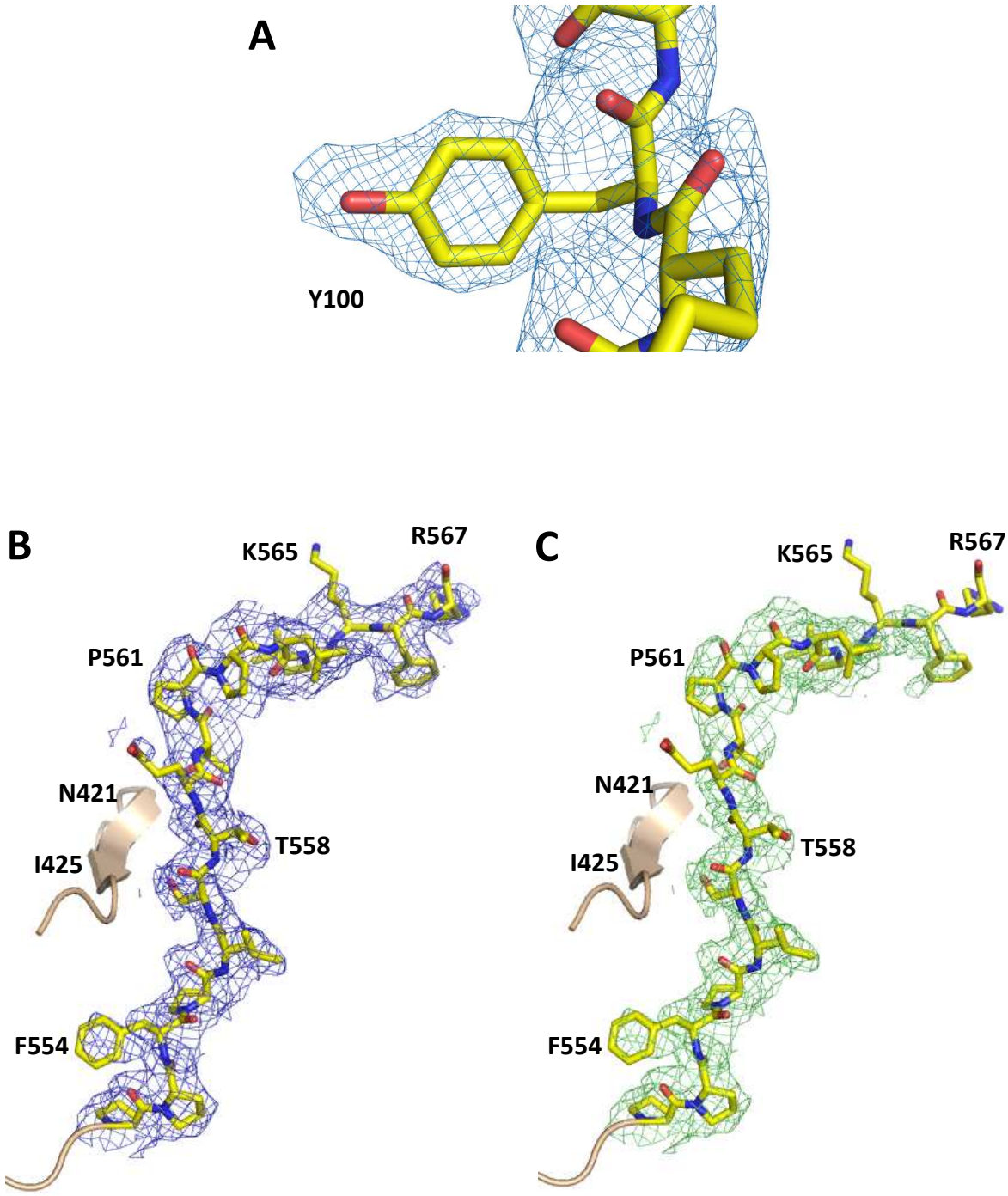

A
Y100
B
K565
R567
P561
N421
T558
I425
F554
C
K565
R567
P561
N421
T558
I425
F554

Figure 2

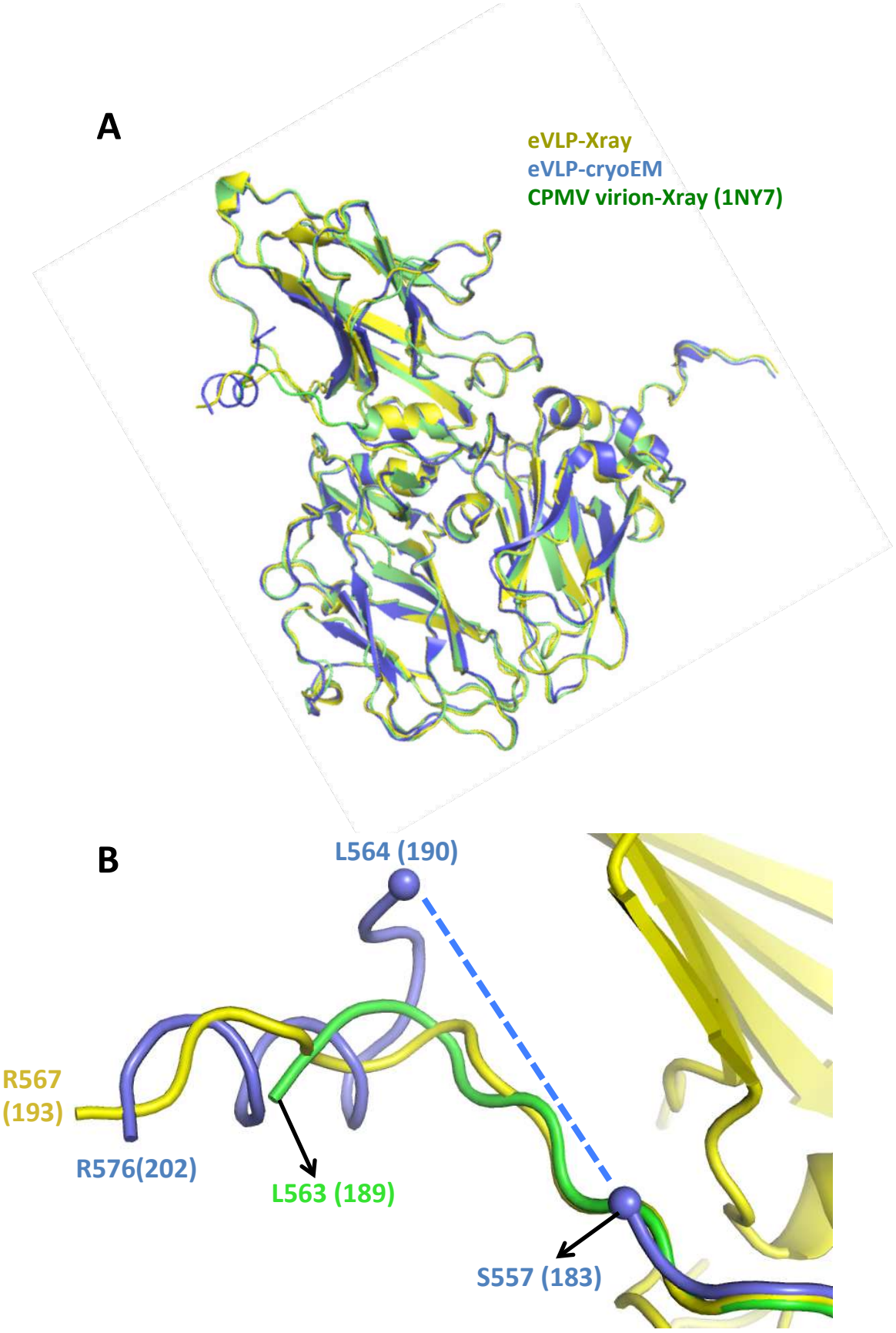

A
eVLP-Xray
eVLP-cryoEM
CPMV virion-Xray (1NY7)
B
L564 (190)
R567
(193)
R576(202)
L563 (189)
S557 (183)

Figure 3

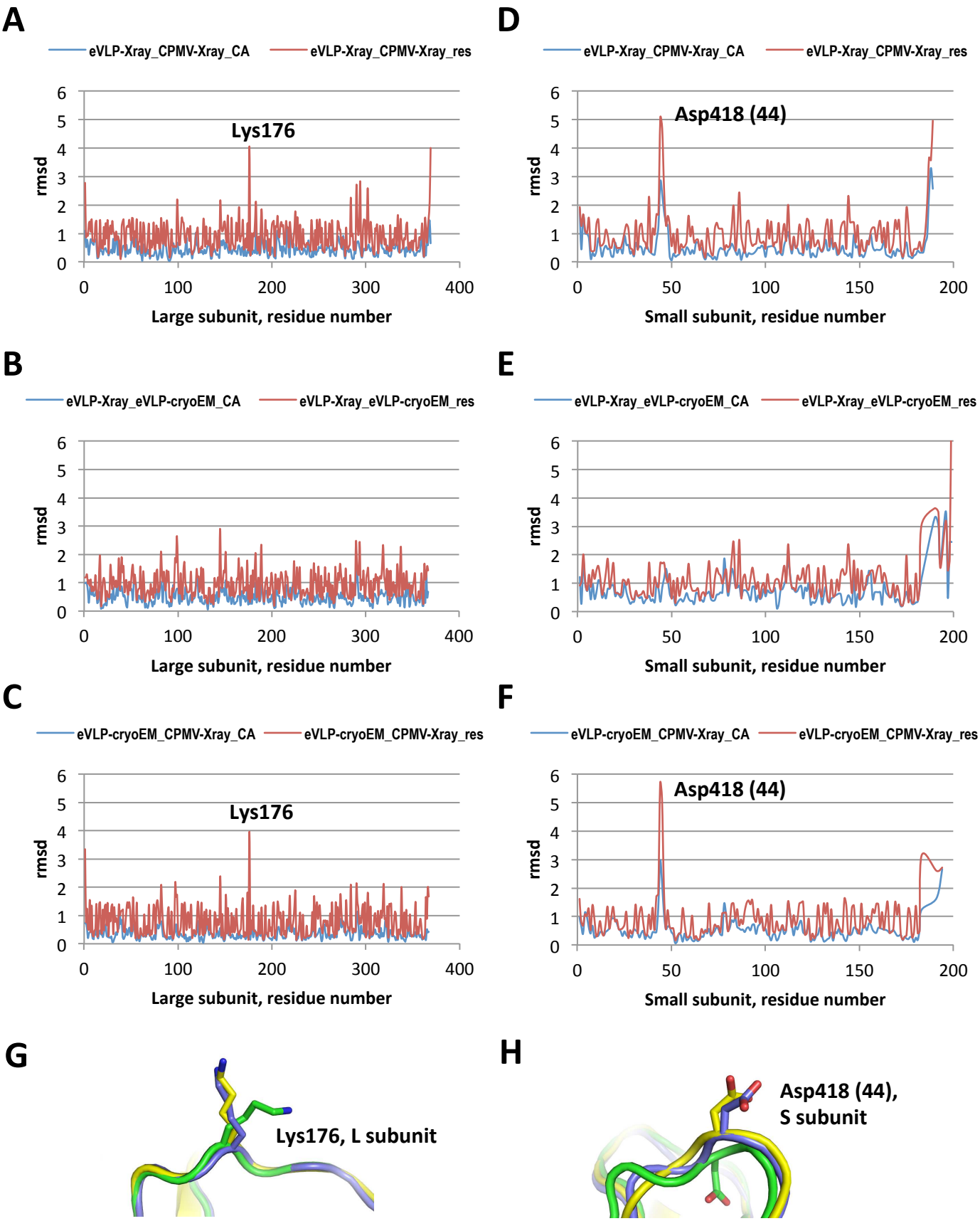
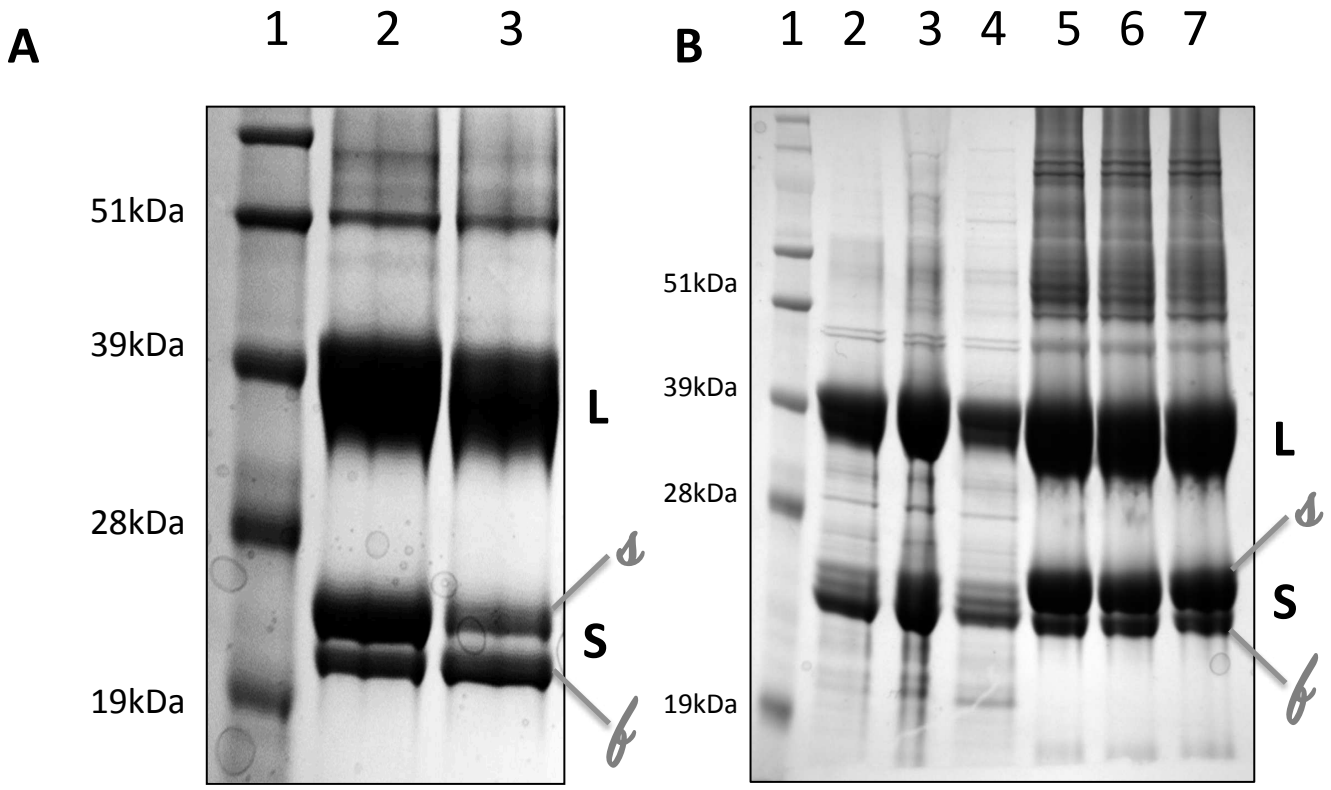

A
eVLP-Xray_CPMV-Xray_CA
eVLP-Xray_CPMV-Xray_res
Lys176
rmsd
Large subunit, residue number
D
Asp418 (44)
Small subunit, residue number
B
eVLP-Xray_eVLP-cryoEM_CA
eVLP-Xray_eVLP-cryoEM_res
E
C
eVLP-cryoEM_CPMV-Xray_CA
eVLP-cryoEM_CPMV-Xray_res
F
G
Lys176, L subunit
H
Asp418 (44),
S subunit

Figure 4



C

Small subunit	Type	Peptide sequence
CPMV Virus	Fast form	----------E.TPPLLKF.R | | 566 567 E.VNMRFDPNFRVAGNILMPPFPLSTETPPL.L E.VNMRFDPNFRVAGNILMPPFPLSTETPPLL.K | | 564 565
CPMV eVLP	Slow form	----------R.SVMVGHTATAA |	| 573 574 587
CPMV eVLP	Fast form	E.VNMRFDPNFRVAGNILMPPFPLSTETPPL.L E.VNMRFDPNFRVAGNILMPPFPLSTETPPLL.K | | 564 565

Figure 5

Table 1. Data collection and Refinement Statistics

X-ray Source	23-ID-D, GMCA, APS
Wavelength (Å)	0.858
Space Group	$I4_1$
Cell parameters	a = 655.97 Å, b= 655.97 Å, c = 571.45 Å, α = 90°, β = 90°, γ = 90°
Number of particles	16/unit cell; 2/crystal asymmetric unit;
V_M (Å^3/Da)	2.75
Percent Solvent (%)	55
Resolution (Å)	50-2.3 (2.42 – 2.3)[§]
No. of unique reflections	3,192,399 (383,718) [§]
Completeness (%)	60.3 (49.6) [§]
R_{merge}[a] (%)	0.29 (0.452) [§]
I/σ	3.7 (1.7) [§]
Particle orientations (Phaser solution)	Particle-1: EULER 39.90, 71.85, 333.96 FRAC 0.252, 0.00, 0.0000 Particle-2: EULER 74.83, 89.08, 189.73 FRAC 0.500, 0.249, 0.253
Ordered residues	1-369 (L); 375-567 (S)
Waters	58
Refinement Statistics	
Resolution (Å)	20-2.3
No. of reflections	1,861,715
Degrees of NCS	120
R_{work}[b]	0.357
R_{free}[c]	0.359
RMS bond length (Å)	0.009
RMS bond angle (°)	1.527
Ramachandran Plot Statistics	
Most Favorable (%)	78.2
Allowed (%)	2.6
Disallowed (%)	0

[a]$R_{merge} = [\sum_h\sum_i|\mathbf{I}_h - \mathbf{I}_{hi}|/\sum_h\sum_i\mathbf{I}_{hi}]$ where $\mathbf{I}_h$ is the mean of $\mathbf{I}_{hi}$ observations of reflection h. Numbers in

[§]Shown in parenthesis are the statistics for the highest resolution shell.

[b]R_{work} and [c]$R_{free} = \sum||F_{obs}| - |F_{calc}|| / \sum|F_{obs}|$ x 100 for 95% of recorded data ([b]R_{work}) or 5% data ([c]R_{free})

Supplemental Information

Crystal structure and proteomics analysis of empty virus-like particles of Cowpea mosaic virus

Nhung Huynh[1], Emma L. Hesketh[3], Pooja Saxena[2§], Yulia Meshcheriakova[2], You-Chan Ku[1], Linh Hoang[4], John E. Johnson[1], Neil A. Ranson[3], George P. Lomonossoff[2] and Vijay S. Reddy[1*]

[1]Integrative Structural and Computational Biology, [4]Scripps Center for Metabolics and Mass Spectrometry, The Scripps Research Institute, La Jolla, CA, 92037, USA.

[2]Dept. of Biological Chemistry, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK.

[3]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

* Corresponding author: Vijay S. Reddy; E-mail address: reddyv@scripps.edu

[§]Present address: Department of Chemistry, Indiana University, 800 E. Kirkwood Ave., Bloomington, IN 47405-7102, USA.

Table S1. Binwise data statistics of CPMV eVLP data set[§].

Dmax	Dmin	Nmeas	Nref	% completeness	Multiplicity	<I/>/sd	Rsym
50.0	7.27	519449	123415	73.8	4.2	23.5	0.106
7.27	5.14	800622	226437	74.0	3.5	8.4	0.258
5.14	4.20	924694	292341	73.8	3.2	6.2	0.310
4.20	3.64	767675	322157	68.7	2.4	3.2	0.358
3.64	3.25	744592	338852	63.7	2.2	2.2	0.398
3.25	2.97	708791	348780	59.3	2.0	1.9	0.415
2.97	2.75	719693	365886	57.2	2.0	1.8	0.423
2.75	2.57	796517	391997	57.0	2.0	1.8	0.431
2.57	2.42	764142	398816	54.5	1.9	1.7	0.432
2.42	2.30	681688	383718	49.6	1.8	1.7	0.452
Totals		**7427863**	**3192399**	**60.3**	**2.3**	**3.7**	**0.290**

[§]Scaling and merging was done using the SCALA program (Evans, 2006).

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr *62*, 72-82.

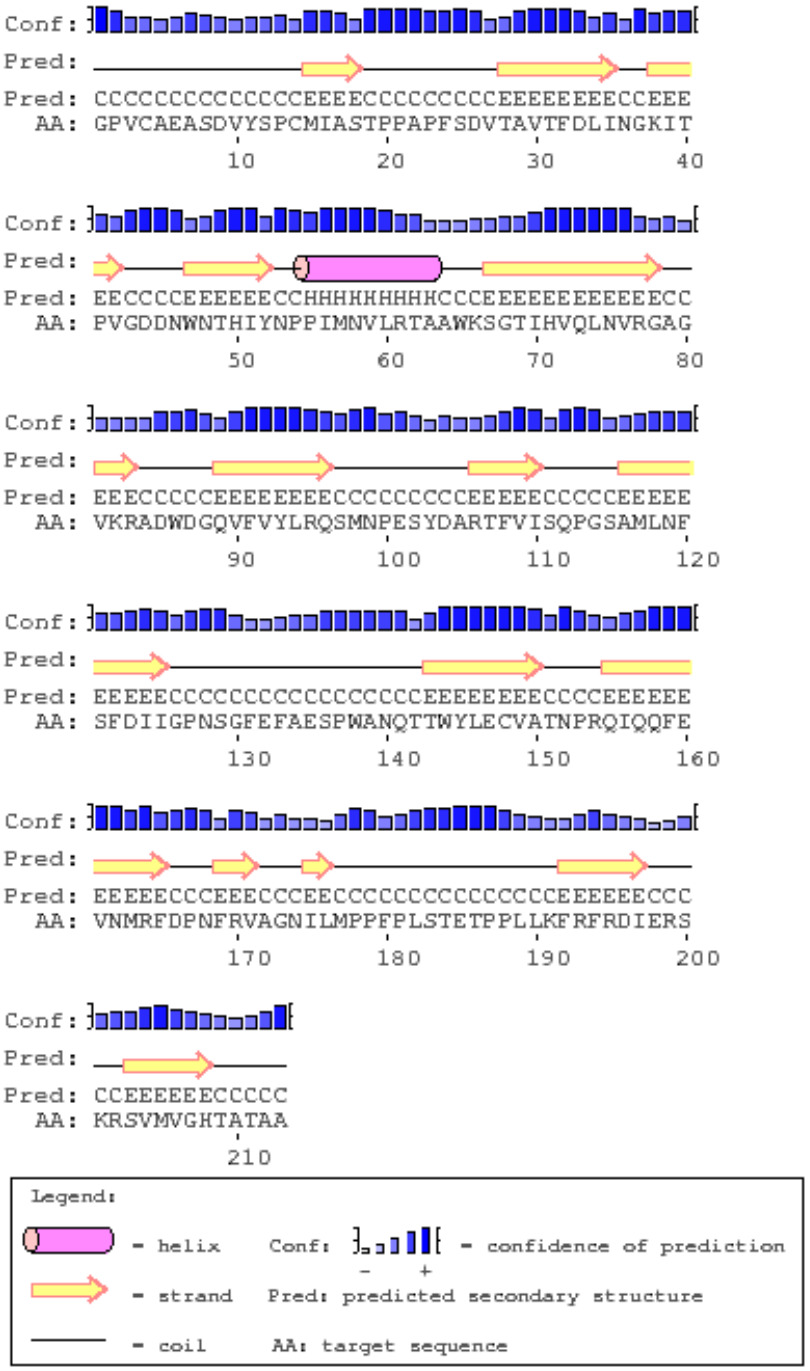


Figure S1. Sequence based secondary structure prediction (SSP) of S subunit of CPMV. The C-terminal region displays mainly coil and beta-strand regions consistent with the X-ray structure of eVLPs. The old residue numbering is shown here. The residues 180-213 corresponds to residues 554-587 according to the new numbering. The SSP was obtained using the PSIPRED server http://bioinf.cs.ucl.ac.uk/psipred/ (Buchan et al., 2013).

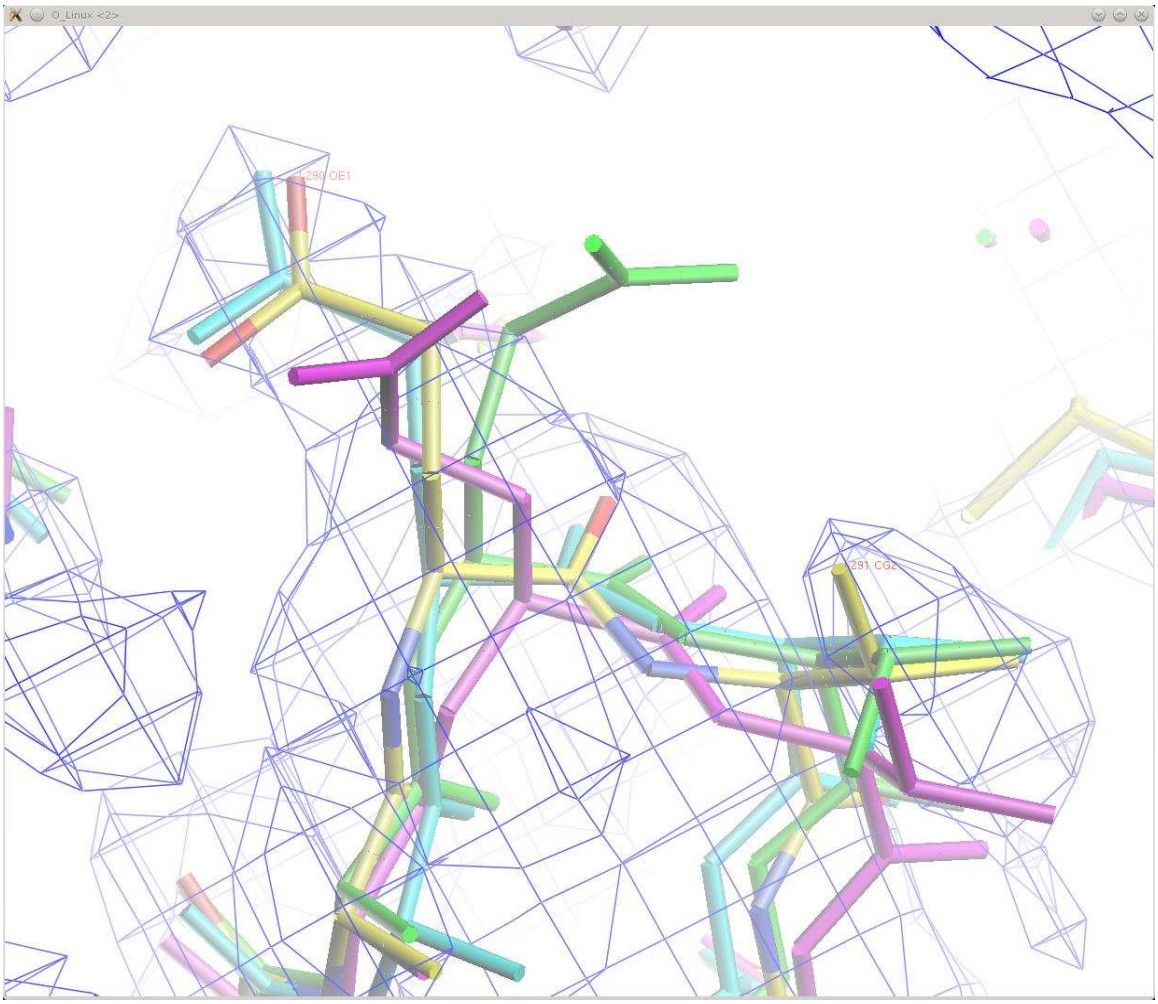


Figure S2. An example illustration of small structural variations with rmsds 1-2Å between different CPMV structures. Shown in atomic colors (yellow (C), red (O), blue (N)) is eVLP X-ray structure refined with CNS, shown in cyan Is eVLP X-ray structure refined with X-PLOR, in magenta is eVLP cryo-EM structure and the virion X-ray structure (PDB-ID: 1NY7) is show in green.

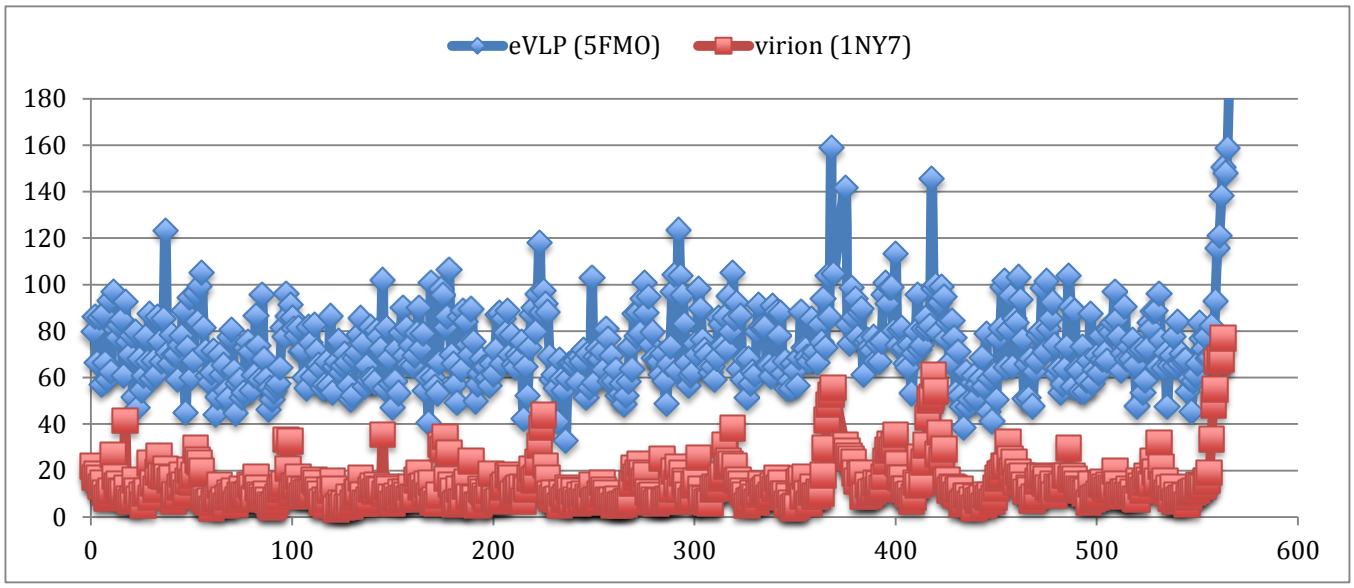


Figure S3. Plot comparing the average residue B-factors of the eVLP (blue line) and virion (red line) X-ray structures. Even though absolute values (magnitudes) differ between the two structures, the relative variations of the B-factors along the polypeptide chains appear to be similar.