Title

New Strategies for Functionalizing In Vitro Reconstituted Virus-like Particles with Protein Ligands

Author

Tisnado, Jerrell Ray

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

New Strategies for Functionalizing *In Vitro* Reconstituted Virus-like Particles with Protein Ligands

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Molecular Biology

by

Jerrell Ray Tisnado

2022

ABSTRACT OF THE DISSERTATION

New Strategies for Functionalizing *In Vitro* Reconstituted Virus-like Particles with Protein Ligands

by

Jerrell Ray Tisnado

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles, 2022

Professor William M. Gelbart, Chair

We have genetically engineered the capsid proteins (CPs) of two RNA-packaging plant viruses, BMV and TYMV, to facilitate the presentation of arbitrary protein ligands on the exterior surface of the virus-like particles (VLPs) that they form. This is important for the targeting of *in vitro* self-assembled mRNA-containing VLPs for *in vivo* gene delivery purposes. In particular, this platform will provide an alternative to lipid nanoparticle systems, with the advantages: of ensuring definite stoichiometry – one mRNA molecule per 180-protein capsid (as opposed to an indeterminate number of mRNAs in polymorphic/mixed-component liposomes); and of precise control over the site of conjugation of protein ligands, introducing genetic fusions (as opposed to promiscuous N-Hydroxysuccinimide (NHS) esters or sulfhydryl-maleimide chemistry).

More explicitly, we have developed two novel approaches to the functionalization of *in vitro* reconstituted VLPs with protein ligands. The first involves the genetic insertion of a poly-Glycine (polyG) moiety into an exterior loop of a CP, just downstream of a protease cut site. Cutting of the loop, followed by the addition of the Sortase A enzyme, will result in covalently linking – to the newly created polyG N-terminus – an arbitrary protein to whose C-terminus contains the Sortase A binding motif (LPXTG). The second strategy involves exploiting an analogy between the beta barrels of viral CPs and of green fluorescent protein (GFP) "split proteins." Here the idea is to make the split portions of CPs, genetically fuse the N-terminus of the C-terminal part (i.e., the tail end of the CP) to a protein ligand of interest, and then mix the two CP portions together so that they associate into a nicked mutant CP that, in turn, self-assembles into VLPs with protein ligands already displayed on their exterior surfaces. These two strategies are demonstrated for BMV and TYMV, respectively, and offer "proof-of-principle" examples of a potentially powerful means for *in vitro* synthesis of targeted mRNA-containing particles for gene delivery.

The dissertation of Jerrell Tisnado is approved.

Charles M. Knobler

Michael F. Carey

Harold G. Monbouquette

Feng Guo

William M. Gelbart, Committee Chair

University of California, Los Angeles

2022

DEDICATION

This dissertation is dedicated to my loving parents. They have been very supportive of my science throughout the years. I would not have had a chance to pursue a higher education in the sciences if not for their continued sacrifice and extreme patience. I love and miss them very much. I will show them the time spent away was not in vain.

Table of contents

1. Introduction.....1

References.....8

2. Bacterial capsid protein expression.....14

2.A. Capsid protein expression for *in vitro* assembly of BMV and CCMV

2.B. Human xcl1 expression and conjugation strategy using bifunctional linkers

Need to cleave N-terminal expression tag with enzyme that leaves an N-terminal valine, to preserve Xcr1 binding affinity

And need to acylate this valine so that the N-terminus is not a primary amine, to avoid its conjugation to bifunctional linker (hence need to switch to yeast expression)

References.....22

3. Yeast capsid protein expression.....24

3.A. General motivation relating to *in vivo* VLP formation involving BMV mutant capsid protein

Need to provide CP with suitable RNA molecule with which to form wildtype (T=3) VLPs

3.B. Specific BMV mutant CP – with polyG inserted in an exterior loop, along with proTEV cut site for Sortase A conjugation – which maintains ability of CP to package RNA and form VLPs *in vivo*

Specific BMV mutant CP – with polyG insert in exterior loop, and with proTEV cut site – for Sortase A conjugation, which maintains ability of CP to package RNA and form VLPs *in vivo*

Purification of these mutant CPs and *in vitro* assembly of VLPs using them

References.....30

4. **Creating Sortase A-labeled protein ligands and conjugating them to VLPs assembled from yeast-derived capsid proteins.....31**

mTFP1

GST

Hu-xcl1 ligand

Protein Z

References.....39

5. Enzyme and chemistry free presentation of protein ligands on virus-like particles, using self-complementing capsid protein fragments.....40

References.....64

6. **Materials and Methods.....67**

References.....70

LIST OF FIGURES

Figure 2-1: Schematic diagrams (A) and electron micrographs (B) of wild-type CCMV and BMV CPs and of the VLPs formed by them.....17

Figure 2-2: Click chemistry bifunctional linkers and protein conjugation.....20

Figure 2-3: Schematic diagram for a yeast derived human Xcl1 protein and its respective plasmid.....21

Figure 3-1: Flow chart indicating how a ligand can be ligated to a VLP via Sortase A enzyme....26

Figure 3-2: BMV candidate exterior loops for proTEV/polyG insertion.....27

Figure 3-3: BMV CP loop1 proTEV/polyG VLP sucrose gradient purification and gel imaging..28

Figure 3-4: Electron Micrographs of yeast-derived BMV loop1 proTEV/ polyG VLPs.....29

Figure 4-1: Diagram of Sortase A ligand cassette, with choice of 4 specific protein ligands.....32

Figure 4-2: 1% TAE agarose gel showing the effects of differing levels of proTEV cleaved BMV loop1 proTEV/polyG VLPs.....33

Figure 4-3: SDS PAGE of BMV CP loop1 proTEV/polyG to GST ligation via sortase A.....34

Figure 4-4: Crystal structure of Protein Z (aka “Protein A”) bound to the Fc region of human IgG (grey structure).....36

Figure 4-5: Flow chart indicating how a ligand can be ligated to mutant CP via Sortase A enzyme to avoid insolubility.....38

Figure 5-1: Ribbon diagrams of TYMV.....43

Figure 5-2: Electron Micrographs of TYMV wt.....44

Figure 5-3: Split protein components of the TYMV capsid protein (CP), involving a genetic fusion of the mTFP1 protein with the 8th and last beta strand of the TYMV CP.....45

Figure 5-4: Flow chart indicating how TYMV VLP assembly can occur *in vitro*.....47

Figure 5-5: Synthesis of *in vitro* assembled TYMV split-protein (sp) VLPs.....48

Figure 5-6: Different field-of-view images of 1:0 sp "monomer": wild type TYMV VLPs.....49

Figure 5-7: Flow chart indicating how TYMV VLP assembly can occur *in vitro*, with both mutant-split proteins and wild-type proteins.....50

Figure 5-8: Different field-of-view images of 1:3 split-protein "monomer": wild-type CP TYMV VLPs.....51

Figure 5-9: 1% TAE gel of 1:0 and 1:3 split-protein (sp) "monomer": wild type VLPs, Protein A, and wildtype CP VLPs, as imaged with by Coomassie stain (left) and fluorescence (right).....52

Figure 5-10: A schematic diagram showing the two TYMV split protein genes.....53

Figure 5-11: Mutant-split protein VLPs run in PBS sucrose gradient.....53

Figure 5-12: Mutant VLP bound Nickel resin.....54

Figure 5-13: 1% TAE gel of various (1:0, 1:3, and 1:5) sp "monomer": wildtype CP VLPs.....55

Figure 5-14: Different field-of-view images of 1:5 sp "monomer": wild type TYMV VLPs post sucrose-gradient purification and nickel-affinity chromatography.....56

Figure 5-15: Mutant-split protein VLPs run in PBS sucrose gradient.....57

Figure 5-16: Different field-of-view EM images of combined sample outlined in Figure 27.....58

Figure 5-17: Different field-of-view EM images of 1:5 sp "monomer": wildtype CP assembled VLPs.....59

Figure 5-18: A drawing outlining ideal conditions for split protein ligand integration with VLPs.....60

Figure 5-19: Split protein tool kit.....61

Figure 5-20: Crystal structures highlighting the different lengths of the final beta strand of different viruses.....63

ACKNOWLEDGMENTS

First and foremost, I want to thank my boss, friend, and racing partner Dr. William Gelbart. This man has made himself available whenever I needed consulting. He constantly went above and beyond to make sure I got the support I needed for my work. He occasionally indulges my unorthodox science ideas, gives me space to make mistakes, and has shown exceptional patience and grace watching me at times go down yet another rabbit hole. He is a natural born mentor, and I will always appreciate his advocacy.

I also want to thank my other boss, Dr. Charles Knobler. His help was always available just for the asking. Throughout the years, he equipped me with a steady supply of relevant reading material—always willing to go through the papers with me whenever I needed. I also appreciate his assistance in editing my writing, including this thesis.

To my lab partners Vladimir Musatov and Andreas Langenbacher, I want to say, Thank you so much! These two young scientists have been with me through thick and thin. They always believed in my vision and gave their best consistently. I would not be anywhere without their hard work.

A special thanks goes out to Dr. Anilú Duran Meza for being my kind friend and confidant during the years. She kept me sane during the tough times.

Nina Harpell also played a significant part in my success, gladly lending her skills whenever asked.

VITA

EDUCATION

2014 – Present	UCLA MBIDP Biochemistry, Biophysics, and Structural Biology graduate (BBSB) PhD program
2014 – Class of	Bachelor of Science degree: Chemistry – Emphasis in Biochemistry San Diego State University, San Diego CA Magna cum laude

RESEARCH EXPERIENCE

2015 Spring – Present	Prof. William Gelbart, UCLA Chemistry and Biochemistry Distinguished Professor, CA 90095
2012 Spring – 2014 Fall	Prof. Carl Carrano's BioInorganic Laboratory San Diego State University, CA 92182 Department of Chemistry and Biochemistry
2013 Summer	Harvard Stem Cell Institute, MA 02138
2011 Summer	Prometheus Labs San Diego, CA 92121

PUBLICATIONS

Barker RA, Tisnado JR, Lambert LA, et al. (2014). Molecular Characterization of a Homolog of the Ferric-uptake regulator, Fur, from the Marine Bacterium Marinobacter algicola DG893. *BioMetals*, doi: 10.1007/s10534-014-9815-7

Weerasinghe AJ, Amin SA, Barker RA, et al. (2013). Borate as a Synergistic Anion for *Marinobacter algicola* Ferric Binding Protein, FbpA: A Role for Boron in Iron Transport in Marine Life. *Journal of the American Chemical Society*, 135 (39), pp 14504–14507 doi: 10.1021/ja406609s

HONORS AND AWARDS

Spring 2015	NSF Graduate Research Fellowship
	Ford Foundation Fellowship
	AAAS/Science Program for Excellence in Science
Fall 2014	UCLA Cota-Robles Fellowship
	UCLA LSAMP Bridge to Doctorate Fellowship (NSF)
Summer 2014	UCLA Competitive Edge summer fellowship
Spring 2014	Louis Stokes Alliances for Minority Participation (LSAMP)
Program (NSF)	
Fall 2013	Student of the Game award
	SACNAS Travel Scholarship
Summer 2013	Harvard Stem Cell Institute Internship Program (HIP)
Year 2012-2013	SDSU Dean's List
Year 2013-2014	Harry E. Hamber Memorial Scholarship

Chapter 1: INTRODUCTION

1. Introduction

Plant viruses, or their virus-like-particle (VLP) counterparts, have become an exceedingly valuable source of biotemplates to develop increasingly sophisticated nanomaterials and delivery systems for a wide range of different fields and applications. By taking advantage of their inherit lack of lipids, lack of pathogenic crossover in mammals, ease of purification, homogeneity, small size, symmetry, well characterized structure(s), and self-assembly capability, it has been possible to functionalize the various particles for use in agriculture,[1] sensors,[2-3] enzymatic nanobioreactors,[4-5] multidimensional scaffold construction,[6-7] contrast reagents,[8] inorganic synthesis,[9] light harvesting materials,[10-11] electronics,[12-13] and in biomedical applications involving diagnostics[38-39] and therapeutics.[14-15]

In the field of therapeutics and vaccinology, three plant viruses stand out the most: icosahedral (spherical) Cowpea Chlorotic Mottle Virus (CCMV), icosahedral Brome Mosaic Virus (CCMV), and helical (cylindrical) Tobacco Mosaic Virus (TMV).

CCMV is a positive sense-RNA-genome Bromovirus that belongs to the Bromovirdae family. It has an outside diameter of 28nm with an 18nm interior cavity. Its spherical shell is composed of 180 copies of the same capsid protein (CP), which consists of 12 pentamers and 20 hexamers (with T=3 symmetry).[14] Trevor Douglas and Mark Young were the first to report its use as a nanoreactor for inorganic synthesis in 1998.[9] One of the most profound and useful features of this virus is that it is able to be disassembled and reassembled *in vitro* through straightforward alteration of ionic strength, divalent metal cation presence, and pH. It can also undergo a reversable swelling phenomenon at neutral pH, involving increase in its ~1.5nm holes that allows it to act like a gate, allowing materials to enter or exit the core.[16]

BMV is very similar to CCMV. It also belongs to the Bromovidae family, and is a spherical positive-sense RNA-genome virus that features a 28nm outside diameter, 18nm interior core. The (T=3) shell is made up of 180 capsid proteins that are also able to assemble

and disassemble under changing ionic strength, divalent metal cation presence, and pH.[17] It also undergoes the same reversible swelling mechanism as CCMV, induced at neutral pH conditions.[18] It was originally harvested from bromegrass back in 1942.[19]

TMV was the first virus to be imaged by electron microscopy[40] and to have its structure determined by X-ray crystallography.[41] It is also the plant virus that has seen the most modifications and use cases in the various fields of all plant viruses. It forms a rod-like structure 300nm in length, consisting of 2130 identical helically-ordered proteins that encapsidate its 6.4knt genome. It contains a 4nm cylindrical pore that runs down the center of the rod.[20] TMV is also stable in a wide range of pH values (pH 3.0-9.0) and also very thermostable—up to 90C. Additionally, it is resilient to many polar organic solutions including methanol, ethanol, acetone, DMSO, and tretrhydofuran.[21]

Other than being plant viruses, the one characteristic that CCMV,[22] BMV,[23] and TMV[24-25] all have in common is their ability to self-assemble *in vitro* from purified components. Furthermore, their capsid proteins can self-assemble around heterologous RNA, which provides robust RNase protection.[26] It is due to these unique benefits that these three viruses, and their associated virus-like particles (VLPs), have become the frontrunners for plant virus-based biomedical applications, especially in vaccinology and drug delivery.

For a more effective therapeutic (whether it involves delivery of diagnostic/imaging reagents, drugs, or mRNA), the VLP typically needs to be targeted to the right subset of cells. This means that a peptide, protein, or small functional moiety must be attached (conjugated) to the exterior surface of the VLP to enhance its specific uptake by a particular cell. This conjugation can be done with several methods, each with its own downside. To chemically conjugate protein ligands to VLPs by use of a cysteine, or lysine, or carboxylic acid is seen as the most popular method due to the abundance of available amino acid side chains and overall

reliability. But this method usually requires excess linkers that can alter the protein activity of both the ligand and the VLP. More significantly, the ligand to be bound can be attached to the VLP in the wrong orientation, or the linker itself could be attached within the active site of the protein ligand, resulting in lower functional activity in both cases.[27] The VLP can become overly stable after linker attachment, essentially crosslinking the VLP, keeping it from disassembling within the target cell and from releasing its RNA content. Direct protein fusions at the N or C termini of the CP can be tolerated if the protein or molecule is very small, but, again, larger proteins may result in losing the self-assembly capabilities of the fused CP.[28] For BMV and CCMV, the act of fusing proteins or peptides to either termini will not result in a surface-exposed ligand protein in any event. This is because neither termini faces outwards towards the solvent-accessible surface. Instead, the N-terminus is on the inside of the capsid, interacting strongly with the RNA content, and the C-terminus is involved in CP dimer formation. Click-chemistry techniques do offer better predictability regarding where the ligand will be attached on the VLP, but again, this will also require the ligand (or the VLP) to be labeled with linkers which can result in lower activity. Use of unnatural amino acids with click-chemistry modalities is an effective strategy, however sometimes its expression can lower the overall protein production yield.[29] Accordingly, there is a compelling need to develop an easy-to-use and stoichiometrically-lean modality that can covalently attached ligands to *in vitro* reconstituted VLPs.

To meet this demand, we have pursued two different (yet related) methods. The first is the use of the Sortase A enzyme derived from *Staphylococcus aureus*, which catalyzes the covalent attachment of any LPXTG peptide motif with any N-terminal poly glycine (polyG) moiety.[30] This strategy has proven effective in covalently labeling various VLPs with different ligands, with high specificity. The E2 (VLP) core of the pyruvate dehydrogenase complex was conjugated to elastin-like peptide (ELP), monomeric endogluconase (CelA), and tetrameric beta-glactosidase using the sortase A strategy, where the surface exposed N-termini of the VLP

was fused to a polyG moiety, and the ligands were fused to a LPETG motif.[30] Similarly, CCMV's N-terminus (interior exposed), containing the polyG moiety, was covalently bound to LPETG-containing fluorescent (FITC) molecules.[31] A reverse setup with the P22 VLP derived from the *Salmonella typhimurium* bacteriophage (where GFPs and hemagglutinin heads (ligands) were fused at the N-terminus with a polyG moiety, and the VLP's surface-exposed CP C-terminus was fused to the LPETG sortase binding motif) was also successful.[32] Similarly hepatitis B core protein VLPs were labeled via sortase A to small polyG containing ligand proteins.[33] But none of these cases involved the *in vitro* self-assembly of virus-like particles whose capsid proteins are conjugated to surface-exposed protein ligands, which is the goal of our work.

We know that the sortase A enzyme does an efficient job of covalently attaching ligands at specific points on VLPs, but as previously mentioned, unlike with other capsids the N and C termini of CCMV and BMV are not surface-exposed. Therefore, we have no accessible termini with which to use the sortase ligation strategy for VLP targeting. Genetic insertions into the surface-exposed protein loops is an attractive strategy, but when the insertion is too large it is known to cause a loss of self-assembly capability due to altered tertiary structure of the CP. Large genetic insertions, especially in the middle of the gene, also tend to produce insoluble proteins, making purification and subsequent self-assembly very challenging. Although a heterologous protein expression system like *E coli* is attractive, due to low cost and fast turnaround, this system does not result in self-assembly *in vivo*.[34-35] Therefore, the mutant proteins it can produce may or may not be self-assembly competent even if protein solubility were not a factor. A great testing ground for mutant CCMV or BMV CP self-assembly can be found in yeast (*P. Pastoris*). In particular, the groups of John Johnson[36] and Paul Ahiquist[37] found that they were able to produce *in cellulo* (in yeast) wild type VLPs that were virtually identical to their respective virions. In addition, Johnson's group was able to show that CCMV mutants could be produced that were also self-assembly competent; although some of these mutant VLPs lost RNA packaging capability, yet others did not.[36] One of the mutants they tested

included a short 11-amino-acid residue genetic insert within a surface exposed exterior protein loop (although the specifics of which loop was not clarified), and this mutant VLP was still able to package RNA cargo.[36]

Combining the utility of the yeast expression system along with the sortase A labeling strategy, we have developed a method to efficiently conjugate protein ligands to the exterior of BMV VLPs. This method involves creating an initial (new) polyG N-terminus at a specific surface-exposed protein loop, and using that moiety to covalently attach an LPETG-containing ligand of choice with the assistance of sortase A. This will allow the attached ligand to be docked at its C-terminus, thus allowing its N-terminus to be free and fully functional for targeting. We report here that we have successfully ligated a LEPTG-containing Glutathione S-transferase (GST) ligand (GST-LPETG-6xHis) to a purified yeast-derived BMV mutant VLP. This ligation strategy will be detailed in chapters 3 and 4.

The second approach we have developed is the "spit protein" method, offering an easy-to-use and stoichiometrically-lean modality that can present in a stable manner ligands on VLPs. This method excludes any chemical or enzyme involvement, but does involve–for preliminary "proof of principle" of the approach–another icosahedral plant virus, the Turnip Yellow Mosaic Virus (TYMV). Instead of utilizing enzymes, or NHS ester, or maleimide, or click-chemistry components to attach ligands, we introduce a fusion of the ligand of interest to the N-terminal end of the final C-terminal beta strand of the TYMV CP's beta barrel. This piece is then combined with another separate protein that contains the rest of the CP's beta barrel (more explicitly, the whole CP protein minus the last beta strand). The two pieces bind strongly but non-covalently together to form a new CP "monomer." And this monomer, in the presence of wildtype CPs, can self-assemble into mixed capsids with mutant "monomers" diluted among wildtype CPs. The two pieces of the split protein, both being soluble and stable due to cleavable hydrophilic tags (or hydrophilic ligand proteins), can be expressed in bacterial systems. We

report here that we have successfully created a split protein VLP that features an 6xHis-tagged fluorescent protein (6xHis-mTFP1) as the incorporated protein ligand. We purified this His-tagged split protein VLP using standard Nickel (Ni^{2+}) resin, eluted, and imaged with electron microscopy. This method is described and expanded in chapter 5.

The works reported here are preliminary. Although the data gathered so far indicates that both the sortase A and split protein conjugation strategies (ideas born in the latter part of my eight-year PhD training) are working, further validation and expansion is needed to fully develop them. My plan is to stay in the lab and complete work on them in the first year of my postdoc, so that I can publish a paper on each of them before applying for a research job with a pharmaceutical company.

References

1. Avital, A., Muzika, N. S., Persky, Z., Bar, G., Michaeli, Y., Fridman, Y., Karny, A., Shklover, J., Shainsky, J., Savaldi-Goldstein, S., Weissman, H., Shoseyov, O., & Schroeder, A. (2021). Foliar Delivery of siRNA Particles for Treating Viral Infections in Agricultural Grapevines. *Advanced Functional Materials*, *31*(44), 2101003. https://doi.org/10.1002/adfm.202101003
2. Mao, C., Liu, A., & Cao, B. (2009). Virus-Based Chemical and Biological Sensing. *Angewandte Chemie International Edition*, *48*(37), 6790–6810. https://doi.org/10.1002/anie.200900231
3. Koch, C., Wabbel, K., Eber, F. J., Krolla-Sidenstein, P., Azucena, C., Gliemann, H., Eiben, S., Geiger, F., & Wege, C. (2015). Modified TMV Particles as Beneficial Scaffolds to Present Sensor Enzymes. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.01137
4. Minten, I. J., Claessen, V. I., Blank, K., Rowan, A. E., Nolte, R. J. M., & Cornelissen, J. J. L. M. (2011). Catalytic capsids: the art of confinement. Chem. Sci., 2(2), 358–362. https://doi.org/10.1039/c0sc00407c
5. Comellas-Aragonès, M., Engelkamp, H., Claessen, V. I., Sommerdijk, N. A. J. M., Rowan, A. E., Christianen, P. C. M., Maan, J. C., Verduin, B. J. M., Cornelissen, J. J. L. M., & Nolte, R. J. M. (2007). A virus-based single-enzyme nanoreactor. *Nature Nanotechnology*, *2*(10), 635–639. https://doi.org/10.1038/nnano.2007.299
6. Gerasopoulos, K., McCarthy, M., Banerjee, P., Fan, X., Culver, J. N., & Ghodssi, R. (2010). Biofabrication methods for the patterned assembly and synthesis of viral nanotemplates. *Nanotechnology*, *21*(5), 055304. https://doi.org/10.1088/0957-4484/21/5/055304
7. Klem, M. T., Willits, D., Young, M., & Douglas, T. (2003). 2-D Array Formation of Genetically Engineered Viral Cages on Au Surfaces and Imaging by Atomic Force

Microscopy. *Journal of the American Chemical Society*, *125*(36), 10806–10807. https://doi.org/10.1021/ja0363718

8. Liepold, L., Anderson, S., Willits, D., Oltrogge, L., Frank, J. A., Douglas, T., & Young, M. (2007). Viral capsids as MRI contrast agents. *Magnetic Resonance in Medicine*, *58*(5), 871–879. https://doi.org/10.1002/mrm.21307
9. Douglas, T., & Young, M. (1998). Host–guest encapsulation of materials by assembled virus protein cages. *Nature*, *393*(6681), 152–155. https://doi.org/10.1038/30211
10. Miller, R. A., Presley, A. D., & Francis, M. B. (2007). Self-Assembling Light-Harvesting Systems from Synthetically Modified Tobacco Mosaic Virus Coat Proteins. *Journal of the American Chemical Society*, *129*(11), 3104–3109. https://doi.org/10.1021/ja063887t
11. Ma, Y. Z., Miller, R. A., Fleming, G. R., & Francis, M. B. (2008). Energy Transfer Dynamics in Light-Harvesting Assemblies Templated by the Tobacco Mosaic Virus Coat Protein. *The Journal of Physical Chemistry B*, *112*(22), 6887–6892. https://doi.org/10.1021/jp8006393
12. Royston, E., Ghosh, A., Kofinas, P., Harris, M. T., & Culver, J. N. (2007). Self-Assembly of Virus-Structured High Surface Area Nanomaterials and Their Application as Battery Electrodes. *Langmuir*, *24*(3), 906–912. https://doi.org/10.1021/la7016424
13. Ghosh, A., Guo, J., Brown, A. D., Royston, E., Wang, C., Kofinas, P., & Culver, J. N. (2012). Virus-Assembled Flexible Electrode-Electrolyte Interfaces for Enhanced Polymer-Based Battery Applications. *Journal of Nanomaterials*, *2012*, 1–6. https://doi.org/10.1155/2012/795892
14. Zhang, Y., Dong, Y., Zhou, J., Li, X., & Wang, F. (2018). Application of Plant Viruses as a Biotemplate for Nanomaterial Fabrication. *Molecules*, *23*(9), 2311. https://doi.org/10.3390/molecules23092311
15. Abrahamian, P., Hammond, R. W., & Hammond, J. (2020). Plant Virus–Derived Vectors: Applications in Agricultural and Medical Biotechnology. *Annual Review of Virology*, *7*(1), 513–535. https://doi.org/10.1146/annurev-virology-010720-054958

16. Speir, J. A., Munshi, S., Wang, G., Baker, T. S., & Johnson, J. E. (1995). Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, *3*(1), 63–78. https://doi.org/10.1016/s0969-2126(01)00135-6
17. Rao, A. (2006). Genome Packaging by Spherical Plant RNA Viruses. *Annual Review of Phytopathology*, *44*(1), 61–87. https://doi.org/10.1146/annurev.phyto.44.070505.143334
18. Cuillel, M., Zulauf, M., & Jacrot, B. (1983). Self-assembly of brome mosaic virus protein into capsids. *Journal of Molecular Biology*, *164*(4), 589–603. https://doi.org/10.1016/0022-2836(83)90052-9
19. Lane, L. C. (1974). The Bromoviruses. *Advances in Virus Research Volume 19*, 151–220. https://doi.org/10.1016/s0065-3527(08)60660-0
20. Klug, A. (1999b). The tobacco mosaic virus particle: structure and assembly. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1383), 531–535. https://doi.org/10.1098/rstb.1999.0404
21. Alonso, J., Górzny, M., & Bittner, A. (2013). The physics of tobacco mosaic virus and virus-based devices in biotechnology. *Trends in Biotechnology*, *31*(9), 530–538. https://doi.org/10.1016/j.tibtech.2013.05.013
22. Vega-Acosta, J. R., Cadena-Nava, R. D., Gelbart, W. M., Knobler, C. M., & Ruiz-García, J. (2014). Electrophoretic Mobilities of a Viral Capsid, Its Capsid Protein, and Their Relation to Viral Assembly. *The Journal of Physical Chemistry B*, *118*(8), 1984–1989. https://doi.org/10.1021/jp407379t
23. Garmann, R. F., Goldfain, A. M., Tanimoto, C. R., Beren, C. E., Vasquez, F. F., Villarreal, D. A., Knobler, C. M., Gelbart, W. M., & Manoharan, V. N. (2022). Single-particle studies of the effects of RNA–protein interactions on the self-assembly of RNA virus particles. *Proceedings of the National Academy of Sciences*, *119*(39). https://doi.org/10.1073/pnas.2206292119

24. BUTLER, P. J. G., & KLUG, A. (1971). Assembly of the Particle of Tobacco Mosaic Virus from RNA and Disks of Protein. *Nature New Biology*, *229*(2), 47–50. https://doi.org/10.1038/newbio229047a0
25. Caspar, D. (1964). Assembly and Stability of the Tobacco Mosaic Virus Particle. *Advances in Protein Chemistry*, 37–121. https://doi.org/10.1016/s0065-3233(08)60268-5
26. Annamalai, P., & Rao, A. (2005). Dispensability of 3′ tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. *Virology*, *332*(2), 650–658. https://doi.org/10.1016/j.virol.2004.12.009
27. Parolo, C., de la Escosura-Muñiz, A., Polo, E., Grazú, V., de la Fuente, J. M., & Merkoçi, A. (2013). Design, Preparation, and Evaluation of a Fixed-Orientation Antibody/Gold-Nanoparticle Conjugate as an Immunosensing Label. *ACS Applied Materials &Amp; Interfaces*, *5*(21), 10753–10759. https://doi.org/10.1021/am4029153
28. Domingo, G. J., Orru', S., & Perham, R. N. (2001). Multiple Display of Peptides and Proteins on a Macromolecular Scaffold Derived from a Multienzyme Complex. *Journal of Molecular Biology*, *305*(2), 259–267. https://doi.org/10.1006/jmbi.2000.4311
29. Patel, K. G., & Swartz, J. R. (2011b). Surface Functionalization of Virus-Like Particles by Direct Conjugation Using Azide–Alkyne Click Chemistry. *Bioconjugate Chemistry*, *22*(3), 376–387. https://doi.org/10.1021/bc100367u
30. Chen, Q., Sun, Q., Molino, N. M., Wang, S. W., Boder, E. T., & Chen, W. (2015). Sortase A-mediated multi-functionalization of protein nanoparticles. *Chemical Communications*, *51*(60), 12107–12110. https://doi.org/10.1039/c5cc03769g
31. Schoonen, L., Pille, J., Borrmann, A., Nolte, R. J. M., & van Hest, J. C. M. (2015). Sortase A-Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly Efficient Cargo Loading. *Bioconjugate Chemistry*, *26*(12), 2429–2434. https://doi.org/10.1021/acs.bioconjchem.5b00485

32. Patterson, D., Schwarz, B., Avera, J., Western, B., Hicks, M., Krugler, P., Terra, M., Uchida, M., McCoy, K., & Douglas, T. (2017). Sortase-Mediated Ligation as a Modular Approach for the Covalent Attachment of Proteins to the Exterior of the Bacteriophage P22 Virus-like Particle. *Bioconjugate Chemistry*, *28*(8), 2114–2124. https://doi.org/10.1021/acs.bioconjchem.7b00296

33. Tang, S., Xuan, B., Ye, X., Huang, Z., & Qian, Z. (2016). A Modular Vaccine Development Platform Based on Sortase-Mediated Site-Specific Tagging of Antigens onto Virus-Like Particles. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep25741

34. Zhao, X., Fox, J. M., Olson, N. H., Baker, T. S., & Young, M. J. (1995). In Vitro Assembly of Cowpea Chlorotic Mottle Virus from Coat Protein Expressed in Escherichia coli and in Vitro-Transcribed Viral cDNA. *Virology*, *207*(2), 486–494. https://doi.org/10.1006/viro.1995.1108

35. Díaz-Valle, A., García-Salcedo, Y. M., Chávez-Calvillo, G., Silva-Rosales, L., & Carrillo-Tripp, M. (2015). Highly efficient strategy for the heterologous expression and purification of soluble Cowpea chlorotic mottle virus capsid protein and in vitro pH-dependent assembly of virus-like particles. *Journal of Virological Methods*, *225*, 23–29. https://doi.org/10.1016/j.jviromet.2015.08.023

36. Brumfield, S., Willits, D., Tang, L., Johnson, J. E., Douglas, T., & Young, M. (2004). Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *Journal of General Virology*, *85*(4), 1049–1053. https://doi.org/10.1099/vir.0.19688-0

37. Krol, M. A., Olson, N. H., Tate, J., Johnson, J. E., Baker, T. S., & Ahlquist, P. (1999). RNA-controlled polymorphism in the *in vivo* assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proceedings of the National Academy of Sciences*, *96*(24), 13650–13655. https://doi.org/10.1073/pnas.96.24.13650

38. Hwang, M. P., Lee, J. W., Lee, K. E., & Lee, K. H. (2013). Think Modular: A Simple Apoferritin-Based Platform for the Multifaceted Detection of Pancreatic Cancer. *ACS Nano*, *7*(9), 8167–8174. https://doi.org/10.1021/nn403465a
39. Destito, G., Yeh, R., Rae, C. S., Finn, M., & Manchester, M. (2007). Folic Acid-Mediated Targeting of Cowpea Mosaic Virus Particles to Tumor Cells. *Chemistry &Amp; Biology*, *14*(10), 1152–1162. https://doi.org/10.1016/j.chembiol.2007.08.015
40. Kausche, G. A., Pfankuch, E., & Ruska, H. (1939). Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Die Naturwissenschaften*, *27* (18), 292-299. https://doi.org/10.1007/bf01493353. Even without being able to read German, one can appreciate on p. 292 the first EM image of TMV (and one of the first EM images in general).
41. Franklin, R. E. (1955). Structure of Tobacco Mosaic Virus. *Nature,* 175, 379–381.

2. BACTERIAL CAPSID PROTEIN EXPRESSION

2.A. Capsid protein expression for *in vitro* assembly of BMV and CCMV

Interest in RNA therapeutics has peaked in the last few years due to the dramatic success of the mRNA vaccines developed to fight the COVID-19 pandemic. This type of mRNA vaccine was made popular by large Biotech firms like Pfizer and Moderna, both of which produced an effective lipid nanoparticle (LNP)/mRNA vaccine against the COVID-19 virus. In our lab we specialize in mRNA synthesis for the purpose of vaccine delivery to targeted cells, and instead of combining the mRNA with lipid components, we package and protect/encapsidate it in a shell made up of the capsid protein derived from a pair of plant viruses (CCMV or BMV). The proteins of both these plant viruses have the unique ability to package heterologous RNA between ~2700-4200nt in length into a single RNase-resistant virus-like particle (VLP).[1] These VLPs have far reaching therapeutic potential since they can package virtually any RNA (whose length is not too small [<2700nt] or too large [>4200nt]), each featuring different downstream effects that could include viral or cancer vaccine prophylaxis or therapy.

For the purpose of harvesting capsid protein (CP) for *in vitro* assembly, our lab began growing CCMV or BMV viruses in plants, purifying the virus, dissociating the virus, and recovering the CP. This process is very effective but it is long and tedious when compared to other more modern methods of protein production. Bacterial protein growth can also be effective in producing assembly-competent CP, and it requires a much shorter time compared to infecting plants (i.e,. purifying virus from them, disassembling the virus particles, and purifying CP from these mixes).[2-3] Unlike in plants, bacteria are unable to assemble *in vivo* the CP they produce; more explicitly, the CPs do not form wildtype spherical/ icosahedral shells consisting of 180 copies of the CP protein subunit.[4-5] Typically, the virus genome, single-stranded, positive-sense, mRNA in the case of BMV or CCMV, is packaged within the assembled shell, offering protection against its degradation. Within plants, this assembled protein/RNA structure results in a virion. But if this is done outside the natural plant host, and with

heterologous RNA, the result is a non-infectious VLP. While bacteria lack the ability to package RNA into a protein shell using the CP subunits, the assembly can be rescued *in vitro* by mixing the purified CP with RNA using particular choices (see below) of both the ionic strength and pH.

In vitro VLP assembly can be done in two general steps. To start, the CP, which is stored in high salt conditions (Buffer B: 50mM Tris-HCl, pH 7.2, 1M NaCl), is introduced to the mRNA of choice. These high-salt and neutral pH conditions keep the CPs from interacting with each other, and largely screen the interaction between the RNA and the arginine-rich N-terminal arms of the CP. Here, the major forces driving assembly are due to the electrostatic interaction between the protein (positive charge, from its N-terminal cationic residues) and the RNA (negative charge, from its backbone phosphates). The next step is to enhance this interaction by lowering, by dialysis, the salt concentration (using RNA binding buffer: 50mM Tris-HCl pH 7.2, 50mM NaCl, 1mM DTT), thereby reducing charge screening and driving the saturation of the RNA by CPs. The final step is to lower the pH (VSB - 50mM Sodium Acetate, pH 4.75) by dialysis or buffer exchange. The lower pH brings the proteins close to their isoelectric point, thereby enhancing lateral interactions between the bound proteins and allowing for a spontaneous protein assembly into VLP shells that contain the RNA cargo.

We created a bacterial expression plasmid for CCMV and BMV CP (see Figure 2-1A). We successfully produced and purified CP by standard methods,[2-3] and were able to verify that the BMV and CCMV CP were each assembly competent, just like plant-derived CP. In this way we were able to package mRNA up to 4200nt in length (see Figure 2-1B EMs)[1].

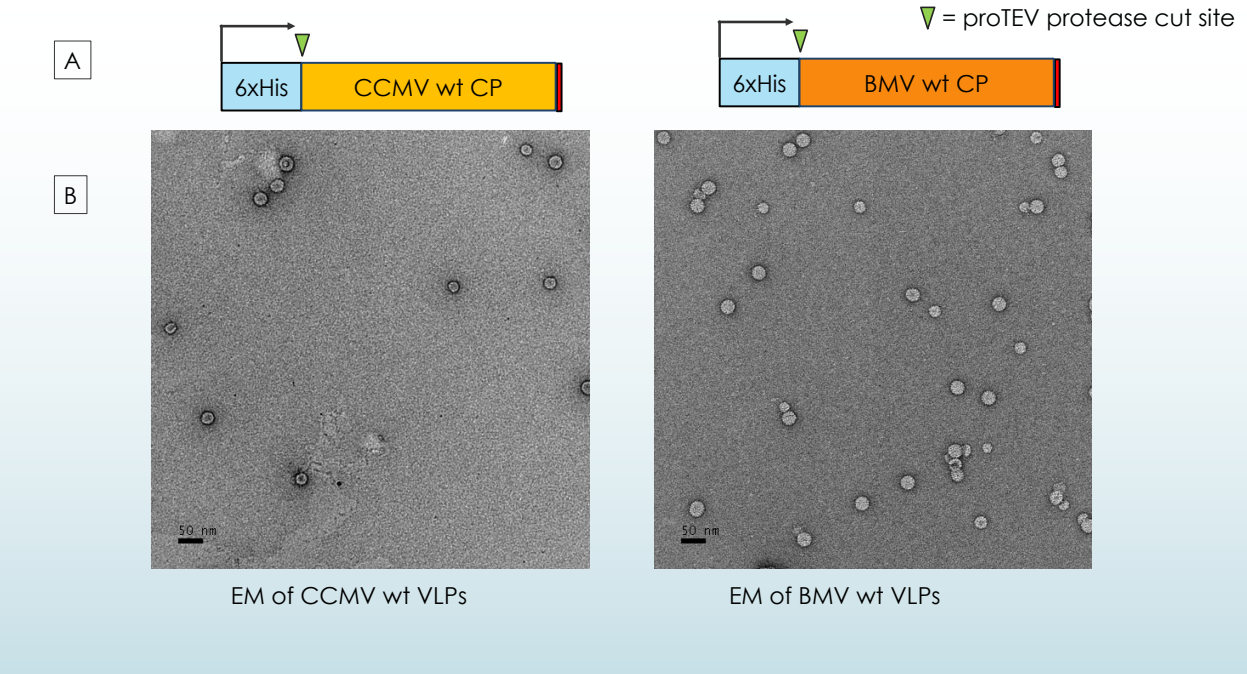

Figure 2-1: Schematic diagrams (A) and electron micrographs (B) of wild-type CCMV and BMV CPs and of the VLPs formed by them. CCMV and BMV CPs are both fused to an N-terminal Histag which is cleavable by proTEV (green arrow) (A). The respective EMs of CCMV wt and BMV wt VLPs, post CP purification and assembly around B1 (cognate RNA1 of BMV, ~3200nts) mRNA (B).

2.B. Human xcl1 expression and conjugation strategy using bifunctional linkers

Others in the lab are working toward an RNA-based CoVID-19 vaccine, and cancer vaccines, and they have been synthesizing mRNA that will express several specific antigenic regions of the CoVID-19 structural proteins (e.g., spike protein and nucleoprotein), and of model cancer antigens like ovalbumin [1]. These RNA molecules need to be *in vitro*-packaged into VLPs and directed to a specific subset of cross-presenting dendritic cells (DCs) for epitope priming and activation of downstream cytotoxic and helper T-cell response.[6] My task is to package the RNA into VLPs and to decorate their exterior with a targeting ligand, Xcl1, that is specific to a receptor uniquely expressed by a class of cross-presenting DCs. This cytokine ligand will bind with high affinity to its cognate receptor, Xcr1, found only on this subset of DCs. Xcl1 is highly structured *with the exception of its C-terminal tail*. However, unlike the disordered C-terminal tail, which is actually disposable and not required for Xcr1 activation, the first several amino acids on the N-terminus of Xcl1 are critical for cognate receptor binding.[7] This is

especially so for the first amino acid, valine. Switching valine to a glycine, or deleting valine, causes the affinity of the ligand for the receptor to drop significantly. Therefore, when producing this protein recombinantly we need to pay special attention to how we purify the target protein via tags, and in particular to what the resulting protein sequence is after cleavage of the tag. When using a TEV protease (proTEV) to remove a n N-terminal histag, for example, a glycine is produced at the N-terminus, which will have an unwanted effect because of it resulting in an N-terminal glycine. We can instead use a SUMO/SENPI (Small Ubiquitin-Related Modifier/Sentrin Specific Protease 1) cleavage mechanism[8-9] to produce a leading valine in the resulting protein post cleavage. But even if we have the correct sequence, we now have a new problem.

More explicitly, cleavage with SENPI, like other proteases (e.g., proTEV) will leave a primary amine group on the leading amino acid. This is something we want to avoid due to the "click-chemistry" mechanism of protein-to-VLP ligation we use. Here we ligate the ligand to the VLPs using two different linkers that will be joined via "click-chemistry" (see Figure 2-2A and 2-2B). The NHS ester "end" of these bifunctional linkers can react with any primary amine,[10] including the first amino acid of Xcl1, valine. And conjugated linkers attached to the N-terminal Xcl1 valine residue can adversely affect Xcl1 affinity for Xcr1, and thereby reduce uptake by DCs of the VLP and the subsequent expression and presentation of vaccine antigenic epitopes. Accordingly, an acylated valine (one without the primary amino group) is necessary to avoid conjugating the linkers to that particular amino acid. To get an acylated valine, we need to express this protein in a eukaryotic cell, like yeast, instead of bacteria. We are currently in the process of piecing together a construct for Xcl1 in yeast (see Figure 2-3).

For our first set of ligand conjugations using the bifunctional linkers, we chose to express mouse xcl1 in two different forms. Studies have shown that xcl1 is first expressed as a precursor protein that has its N-terminal "leader sequence" cleaved before it becomes functional as an

Xcr1-activating protein. This cleaved version is considered the full xcl1 ligand protein.[7] Other studies have shown that the business end of the protein is mainly in the first several amino acids, with the highly disordered C-terminal tail being very disposable.[7] In light of these facts, we made two versions and expressed them both in bacterial (Figure 2-2C). We labeled the smaller xcl1-dt protein (2-74aa) and the CCMV VLPs with their respective linkers (at 2:1 linker: protein ratio). We then conjugated the labeled proteins to one another by introducing them (after the excess linkers were removed) at different xcl1-dt:CP ratios (Figure 2-2D). It seems the higher the ligand presence (xcl1-dt), the more VLP is labeled with the ligand.

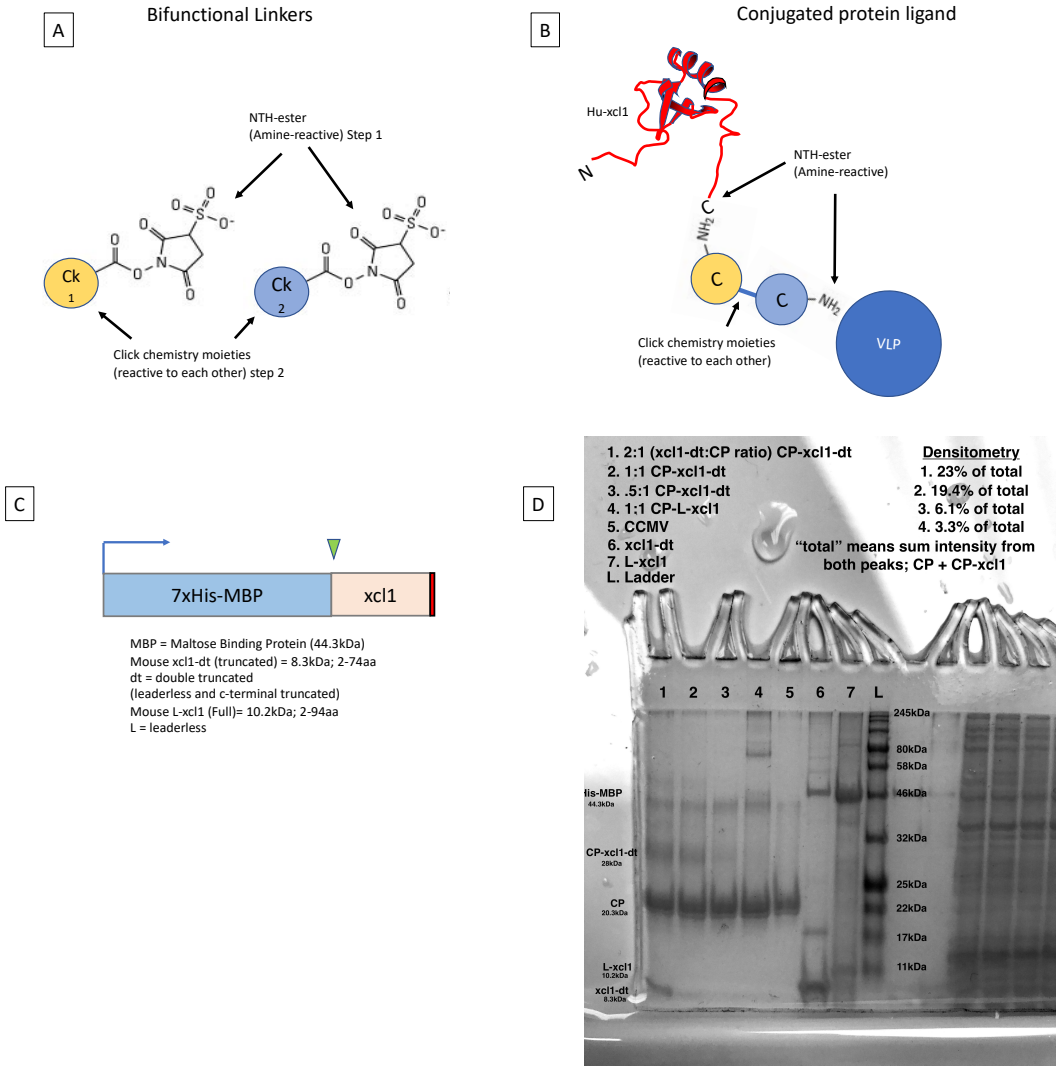

Figure 1-2: Click chemistry bifunctional linkers and protein conjugation. Two bifunctional linkers were used in the conjugation process, each with an NHS-ester at one end that is reactive to any free primary amine group, along with two different (but complementing) click-chemistry moieties (represented In yellow and blue) at the other end (A). Conjugation of the linkers to separate proteins (i.e. Xcl1 or a VLP CP) occurs first, then the two proteins are introduced to each other allowing binding via their click chemistry moieties (B) into a single covalently linked protein. Two xcl1 ligand proteins were bacterially expressed--mouse xcl1-dt (8.3kDa, 2-74aa) and mouse L-xcl1 (10.2kDa, 2-94aa). His-MBP tags were cleaved with proTEV (green arrows) and removed. The general schematic diagrams for the two xcl1 proteins are given. (C). Bacterial-derived CCMV wild type CP conjugated to xcl1-dt (mouse). SDS PAGE Gel (D) shows the different ligand (xcl1):CP conjugation ratio results. Conjugations using the bi-functional linker were introduced to xcl1-dt and CP separately at a linker:protein ratio of 2:1. Then the excess linkers were removed with a desalting column. Finally the two proteins were mixed at ligand:CP ratios that ranged from 0.5:1 to 2:1, with the 2:1 showing the heaviest xcl1 conjugations (lane 1) as determined by densitometry (D). L-xcl1 alone was included as a reference (lane 6), not conjugated to CP. L = leaderless; L-xcl1 = leaderless (mouse) xcl1 (Full untruncated).

In addition to producing wild-type human Xcl1, we plan to experiment with Xcl1 mutants that may prove to be even more affective Xcr1-targeting ligands. Several studies have established that by taking advantage of the critical role that the first several amino acids of Xcl1

have on Xcr1 binding, it is possible to enhance this affinity further by simply altering one or two amino acids (e.g., S6A, R9A, or S6A + R9A.[7])

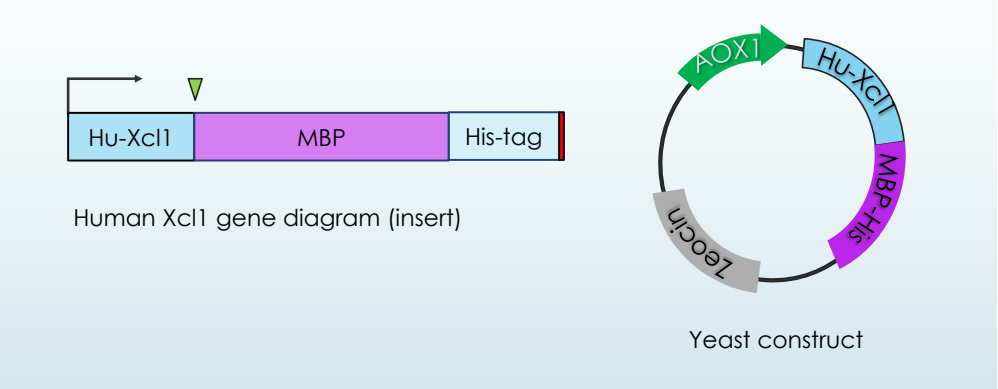

Figure 2-3: Schematic diagram for a yeast derived human Xcl1 protein and its respective plasmid. The human Xcl1 will be expressed with a C-terminal tail consisting of a very soluble MBP protein along with a histag. The histag is designed to allow us to remove the tag post cleavage (green arrow) (left). The plasmid is intended for a methanol inducible yeast strain (i.e., Pichia pastoris), and it features a Zeocin selectable marker along with the gene of interest (right).

References

1. Biddlecome, A., Habte, H. H., McGrath, K. M., Sambanthamoorthy, S., Wurm, M., Sykora, M. M., Knobler, C. M., Lorenz, I. C., Lasaro, M., Elbers, K., & Gelbart, W. M. (2019). Delivery of self-amplifying RNA vaccines in in vitro reconstituted virus-like particles. *PLOS ONE*, *14*(6), e0215031. https://doi.org/10.1371/journal.pone.0215031
2. Zhao, X., Fox, J. M., Olson, N. H., Baker, T. S., & Young, M. J. (1995). In Vitro Assembly of Cowpea Chlorotic Mottle Virus from Coat Protein Expressed in Escherichia coli and in Vitro-Transcribed Viral cDNA. *Virology*, *207*(2), 486–494. https://doi.org/10.1006/viro.1995.1108
3. Díaz-Valle, A., García-Salcedo, Y. M., Chávez-Calvillo, G., Silva-Rosales, L., & Carrillo-Tripp, M. (2015). Highly efficient strategy for the heterologous expression and purification of soluble Cowpea chlorotic mottle virus capsid protein and in vitro pH-dependent assembly of virus-like particles. *Journal of Virological Methods*, *225*, 23–29. https://doi.org/10.1016/j.jviromet.2015.08.023
4. Zhang, Y., Dong, Y., Zhou, J., Li, X., & Wang, F. (2018). Application of Plant Viruses as a Biotemplate for Nanomaterial Fabrication. *Molecules*, *23*(9), 2311. https://doi.org/10.3390/molecules23092311
5. Rao, A. (2006). Genome Packaging by Spherical Plant RNA Viruses. *Annual Review of Phytopathology*, *44*(1), 61–87. https://doi.org/10.1146/annurev.phyto.44.070505.143334
6. Le Gall, C., Cammarata, A., de Haas, L., Ramos-Tomillero, I., Cuenca-Escalona, J., Schouren, K., Wijfjes, Z., Becker, A. M. D., Bödder, J., Dölen, Y., de Vries, I. J. M., Figdor, C. G., Flórez-Grau, G., & Verdoes, M. (2022). Efficient targeting of NY-ESO-1 tumor antigen to human cDC1s by lymphotactin results in cross-presentation and antigen-specific T cell expansion. *Journal for ImmunoTherapy of Cancer*, *10*(4), e004309. https://doi.org/10.1136/jitc-2021-004309

7. Fox, J. C., Thomas, M. A., Dishman, A. F., Larsen, O., Nakayama, T., Yoshie, O., Rosenkilde, M. M., & Volkman, B. F. (2019). Structure-function guided modeling of chemokine-GPCR specificity for the chemokine XCL1 and its receptor XCR1. *Science Signaling*, *12*(597). https://doi.org/10.1126/scisignal.aat4128
8. Kuo, D., Nie, M., & Courey, A. J. (2014). SUMO as a Solubility Tag and In Vivo Cleavage of SUMO Fusion Proteins with Ulp1. *Protein Affinity Tags*, 71–80. https://doi.org/10.1007/978-1-4939-1034-2_6
9. Shen, L., Tatham, M. H., Dong, C., Zagórska, A., Naismith, J. H., & Hay, R. T. (2006). SUMO protease SENP1 induces isomerization of the scissile peptide bond. *Nature Structural &Amp; Molecular Biology*, *13*(12), 1069–1077. https://doi.org/10.1038/nsmb1172
10. Jiang, X., Hao, X., Jing, L., Wu, G., Kang, D., Liu, X., & Zhan, P. (2019). Recent applications of click chemistry in drug discovery. *Expert Opinion on Drug Discovery*, *14*(8), 779–789. https://doi.org/10.1080/17460441.2019.1614910

3. YEAST CAPSID PROTEIN EXPRESSION

3.A. General motivation relating to *in vivo* VLP formation involving BMV mutant capsid protein

As mentioned earlier, previous work (including our own) has shown that neither Cowpea Chlorotic Mottle Virus (CCMV) nor Brome Mosaic Virus (BMV) VLPs can be formed *in vivo* using *E. coli*, which is the most widely used system for recombinant protein production. Instead, success can be found using yeast, a single-cell eukaryotic protein-expression system. More specifically, BMV[1] and CCMV[2] coat proteins (CPs) can be expressed within yeast cells to generate intact VLPs (although the VLPs do require an adequately sized RNA to encapsidate). Moreover, this ability of yeast cells to assemble VLPs makes it a suitable testing ground for identifying potential mutant VLPs for in vitro assembly. The logic here is that if the yeast cells cannot form VLPs *in vivo,* given a specific CP mutant, then *in vitro* self-assembly of RNA using this CP expressed in bacteria will be unlikely. On the other hand, if we know that yeast can assemble a specific CP mutant into VLPs, then at least we know that assembly is *possible,* and we can then try to assemble them *in vitro* using CPs purified from a bulk bacterial growth. Alternatively, we can purify the CP components from the yeast cells, after disassembly of the VLPs, and then proceed to their *in vitro* assembly with RNA.

3.B. Specific BMV mutant CP – with polyG inserted in an exterior loop, along with proTEV cut site for Sortase A conjugation – which maintains ability of CP to package RNA and form VLPs *in vivo*

Here, we seek to have yeast cells make mutant VLPs whose CPs carry an insert within the CP that will allow for subsequent sortase-mediated ligation of a ligand of interest. More

specifically, we want to insert into a surface-accessible CP exterior loop a Tobacco Etch Virus (TEV) protease cut-site motif (ENLYFQ) followed by a polyglycine (polyG) (Figure 3-1A).

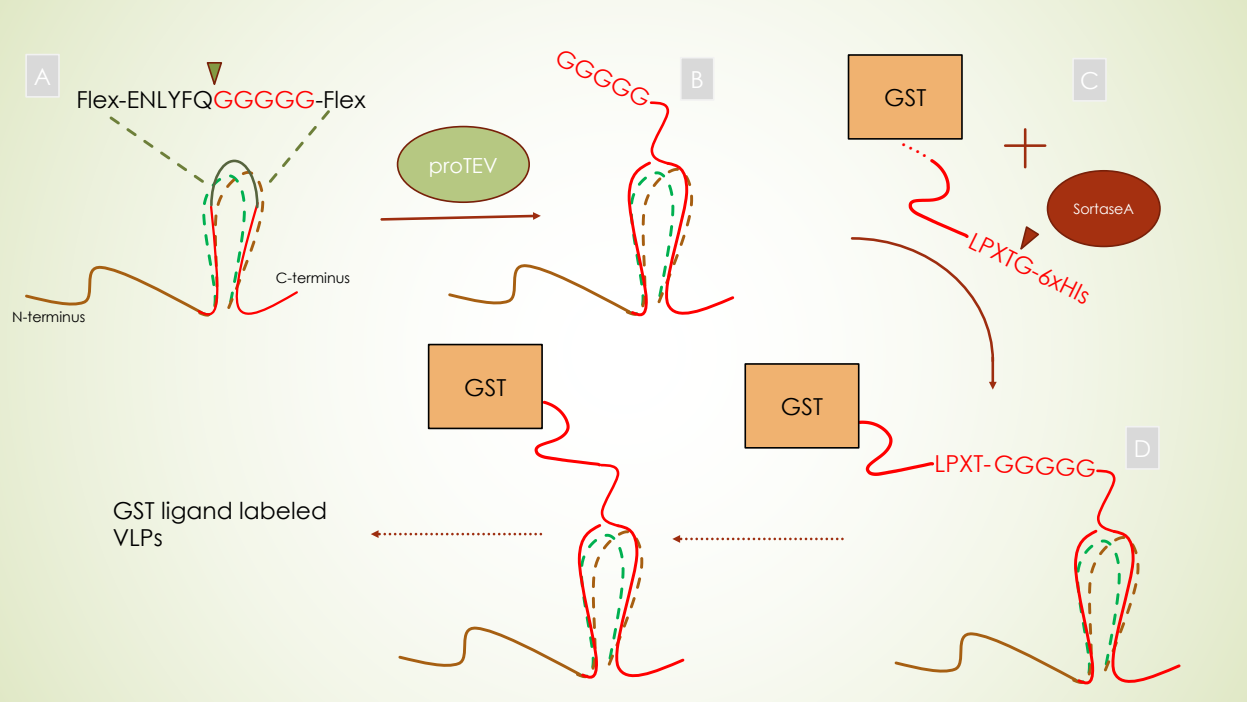

Figure 3-1: Flow chart indicating how a ligand can be ligated to a VLP via Sortase A enzyme. A genetic insert containing a proTEV cut site followed by a poly-Glycine is engineered into one of the surface exposed exterior loops of CP (A). Cleavage at the proTEV cut site (see the green arrow) produces a free new N-terminal polyG (B). A ligand (here GST) genetically fused to the sortase-specific sequence LPXTG is introduced along with the Sortase A enzyme. The enzyme binds the sortase-specific motif (LPXTG) at the red arrow (C). Then the polyG of the cleaved CP displaces the Sortase A enzyme, forming a covalent bond at the C-terminus of the ligand (D).

The VLPs will then be treated with TEV protease (proTEV), which will cut at their binding motif and produce a new polyG N-terminus at that site (Figure 3-1B). Another enzyme, Sortase A, will then be introduced along with a ligand containing a sortase A binding motif (LPXTG) at its C terminus. The Sortase A will covalently bind the LPETG motif at the Threonine amino acid, displacing any peptide after this position (Figure 3-1C). The surface-exposed polyG on the VLP(s) will then displace the sortase A enzyme, forming a new covalent bond between the C-terminus of the ligand and the N-terminus of the polyG (Figure 3-1D). The sortase A will then move on to repeat this cycle. Any ligand that has been bound by sortase A, but not displaced by a polyG, will eventually get hydrolyzed with an H_2O molecule, freeing the sortase A enzyme.[3] In

this way we can label VLPs with any ligand/protein of choice, as long as it contains an accessible C-terminal LPXTG sortase A motif. We refer to this as the Sortase A Labeling Strategy.

There are three exterior beta-strand turn loops within each CP. These three loops are surface-exposed and good candidates for a small genetic insertion. The loops are numbered 1-3 (Figure 3-2). The genetic insertion here is the proTEV cut site/ polyG, flanked by small flex

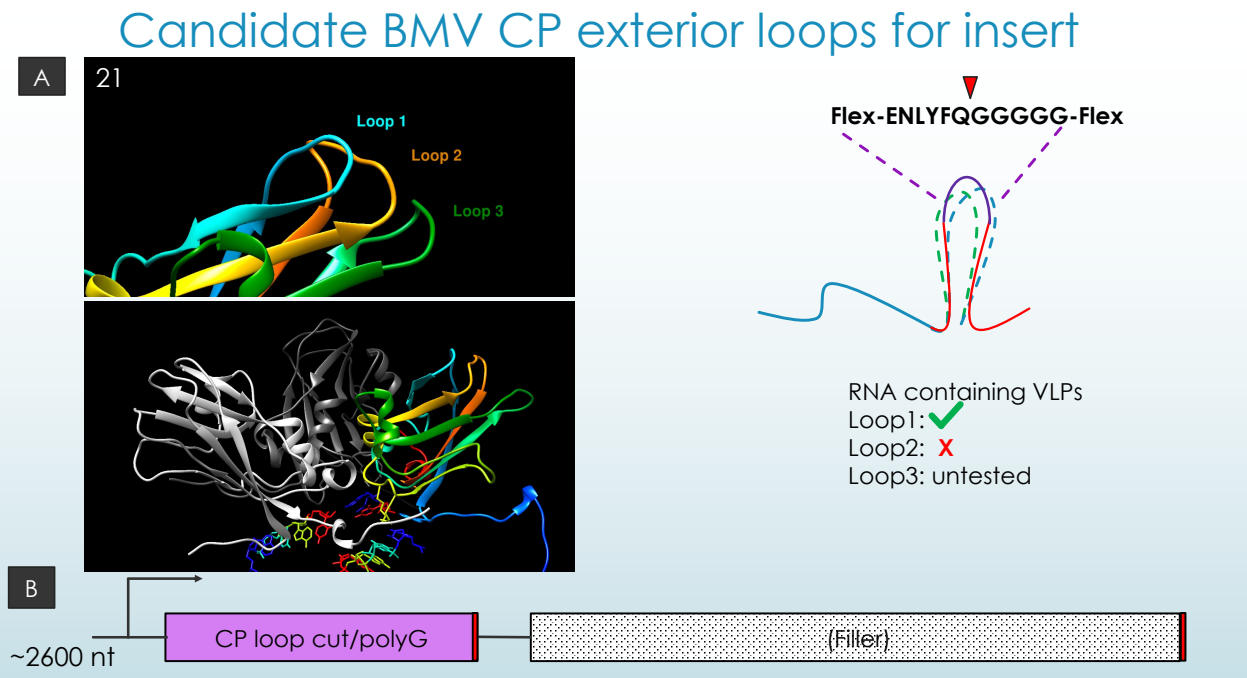

Figure 3-2: BMV candidate exterior loops for proTEV/polyG insertion. Two different ribbon diagram views of the CP containing the three primary candidate loops (which are the three protein turns connecting the beta strands of the beta barrel) for insertion: blue (A63), orange (P163), and green (V100) (A). The diagram of the CP gene containing the insertion at the different loop positions is upstream of the (non-expressed) filler gene, making a ~2600nt length total gene that will be transcribed and packaged into VLPs *in vivo* (B). Adequately sized RNA is necessary for T=3 size VLPs, and 2600nt should meet this requirement.

regions. We cloned the insertion into loop1 (E63 position) and the loop2 (A163) position. We found that only with loop1 can the yeast produce RNA-containing VLPs *in vivo*. This particular CP mutant still maintained its RNA packaging capability (i.e., the CP mutants can bind to and assemble around the RNA cargo). Typically, when the protein structure is disturbed with genetic insertions, the resulting CP tertiary structure is altered enough to also disturb the CP's ability to form VLPs, and even more likely the capability of packaging RNA. This appears to have

occurred with the loop2 insertion, but not for loop1. Yeast protein expression followed standard protocols.[4] We were able to validate the synthesis of VLPs using sucrose gradient (Figure 3-3A and 3-3B), agarose gel (Figure 3-3C and 3-3D), and EM (Figure 3-4) analyses. We call these VLPs "BMV CP loop1 proTEV/polyG" because they are derived from BMV CP and contain the proTEV motif within loop1 followed by a poly-glycine segment.

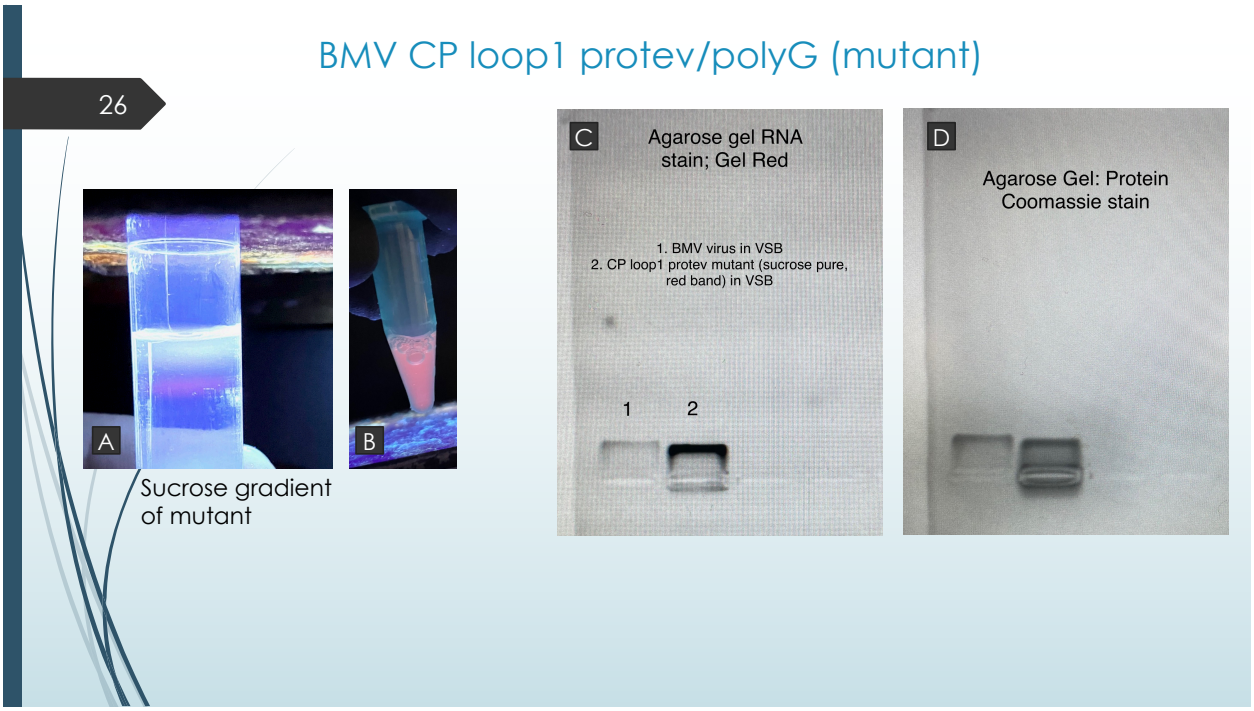

Figure 3-3: BMV CP loop1 proTEV/polyG VLP sucrose gradient purification and gel imaging. Yeast lysate was pre-stained with GelRed and run on an 10-40% VSB (50mM Sodium Acetate, pH 4.75) sucrose gradient (UV visible red band corresponds to the VLPs) (A). Red sucrose band (Red arrow) containing the VLPs was then isolated (B), and run on an 1% OVB (100mM Sodium Acetate, 1mM EDTA, pH 5.5) agarose gel (C &D).

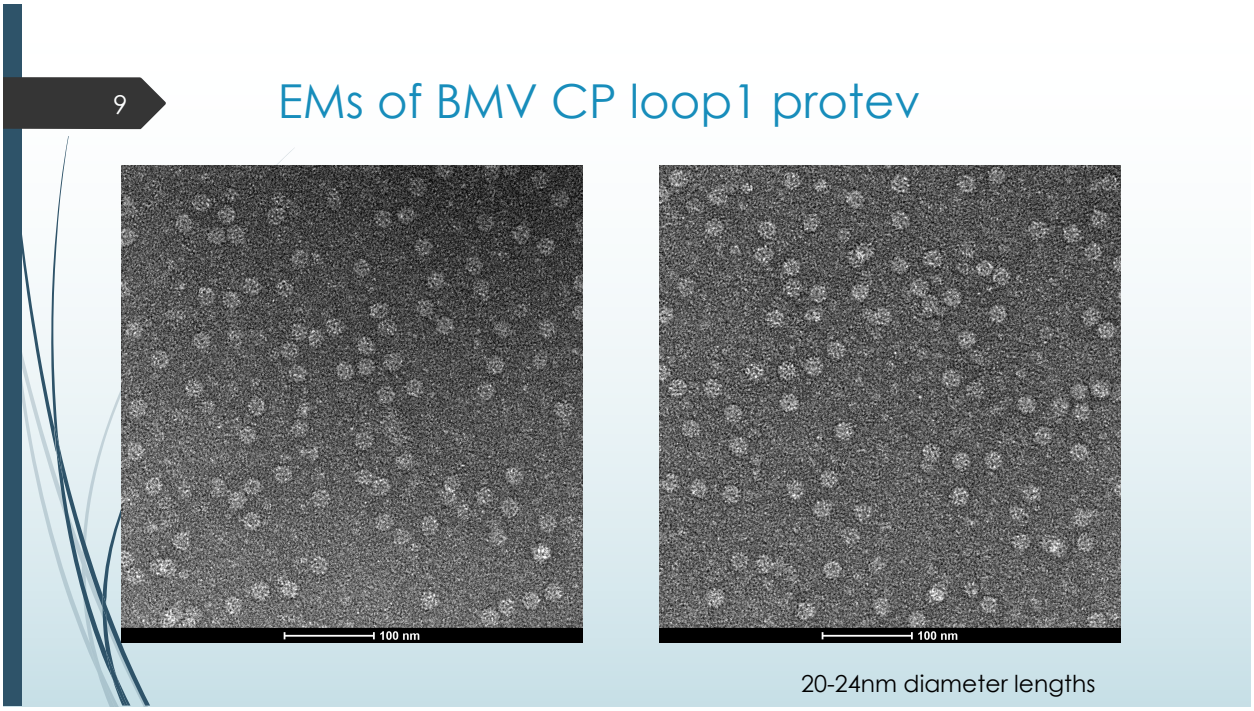

Figure 3-4: Electron Micrographs of yeast-derived BMV loop1 proTEV/ polyG VLPs. EMs of two different fields of view showing 20-24nm diameter VLPs. These yeast-derived VLPs were isolated from yeast lysate that was pre-stained with GelRed and run on a 10-40% VSB (50mM Sodium Acetate, pH 4.75) sucrose gradient as outlined in Figure 3-3. The VLP-containing sucrose was then AAmiconAmicon rinsed several times with just VSB, resulting in a sucrose-free sample that could be EM imaged.

References

1. Krol, M. A., Olson, N. H., Tate, J., Johnson, J. E., Baker, T. S., & Ahlquist, P. (1999). RNA-controlled polymorphism in the *in vivo* assembly of 180-subunit and 120-subunit virions from a single capsid protein. *Proceedings of the National Academy of Sciences*, *96*(24), 13650–13655. https://doi.org/10.1073/pnas.96.24.13650
2. Brumfield, S., Willits, D., Tang, L., Johnson, J. E., Douglas, T., & Young, M. (2004). Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *Journal of General Virology*, *85*(4), 1049–1053. https://doi.org/10.1099/vir.0.19688-0
3. Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F., & Schneewind, O. (1999). Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proceedings of the National Academy of Sciences*, *96*(22), 12424–12429. https://doi.org/10.1073/pnas.96.22.12424
4. Nielsen, K. H. (2014). Protein Expression-Yeast. *Methods in Enzymology*, 133–147. https://doi.org/10.1016/b978-0-12-420070-8.00012-x

4. CREATING SORTASE A-LABELED PROTEIN LIGANDS AND CONJUGATING THEM TO VLPs ASSEMBLED FROM YEAST-DERIVED CAPSID PROTEINS

We then set out to make a set of test ligands with which we could potentially label these VLPs. We created a sortase A ligand cassette that includes an LPETG pentapeptide with a standard 6xHistag C-terminal tail. The N-terminus of the cassette can be swapped with any protein of interest (Figure 4-1). Here we make our first two (test) ligand proteins, a fluorescent protein mTFP1 and the well-known Glutathione S-transferase (GST), both fused to LPETG-histag (Figure 4-1). These two protein fusions were expressed and purified from bacteria. Ligands (e.g., xcl1 and protein Z, see Figure 4-1) used for cell targeting will be introduced after these initial test ligands validate the Sortase A Labeling Strategy.

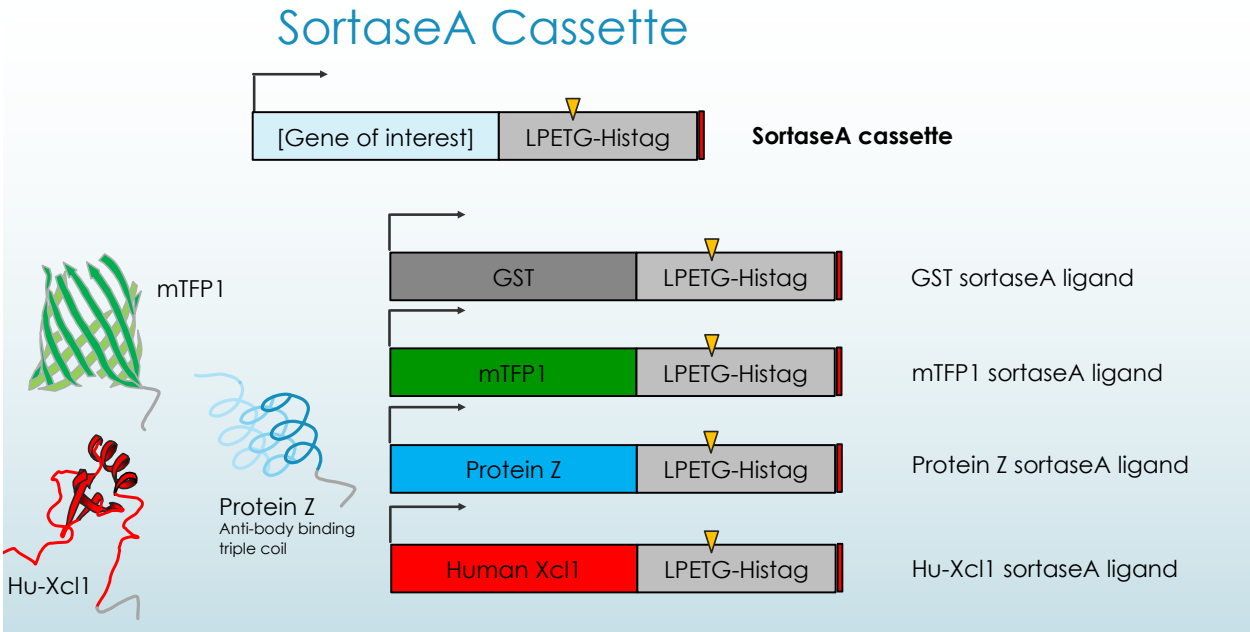

Figure 4-1: Diagram of Sortase A ligand cassette, with choice of 4 specific protein ligands . Each diagram shown depicts a different possible ligand that can be sortase A ligated to a polyG motif. mTP1 was chosen as the most suitable fluorescent protein, with the most intensity at the lowest pH compared to other fluorescent proteins. Protein Z is a highly useful ligand as it has high binding affinity for the Fc region of different antibodies—making this protein a type of universal antibody-binding-ligand (see later discussion). GST - Glutathione S-transferase is an ideal test ligand for this sortase A-mediated ligation strategy as it is not only very soluble but it can bind GSH resin. Orange arrow indicates Sortase A docking/ligation position within the binding motif.

In order to test the VLP sortase A labeling strategy, we begin by confirming that the proTEV cut site inserted within each CP can be cleaved by proTEV. We found that different cutting efficiencies were achieved with different incubation temperatures, when carried out in

virus suspension buffer (VSB) (50mM Sodium Acetate, pH 4.75). The most extensive cleavage is found when proTEV is incubated with the VLP at 37C. These results were confirmed via agarose gel analysis, with the most cleaved VLP running the slowest. The cleavage causing the most change to the overall surface charge compared to uncut VLPs (Figure 4-2).

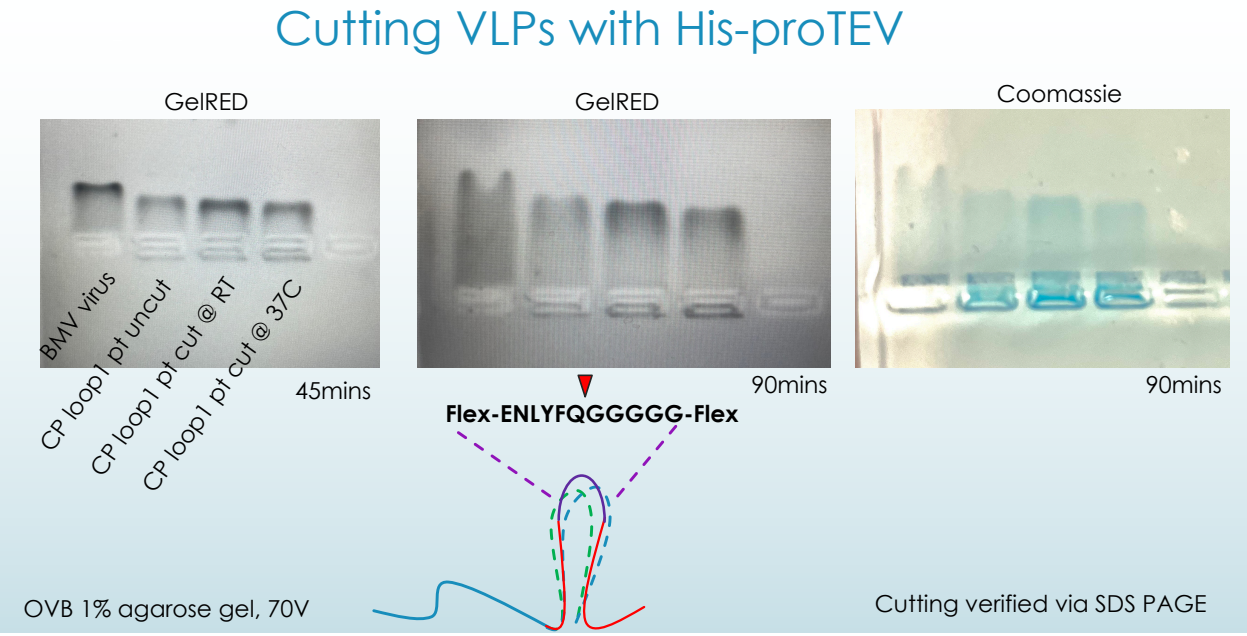

Figure 4-2: 1% TAE agarose gel showing the effects of differing levels of proTEV cleaved BMV loop1 proTEV/polyG VLPs. Runtimes are indicated below the gels.

ProTEV-treated BMV CP loop1 proTEV/polyG VLPs were then mixed with GST-LPETG-Histag (ligand) along with Sortase A enzyme. The mix was dialyzed against a neutral buffer (50mM Tris-HCl pH 7.5, 350mM NaCl) and at optimal temperature for sortase A activation (30C) overnight. The mix was then applied to a 100kDa Amicon filter and rinsed five times with the neutral buffer. The same experiment was repeated, minus the sortase A enzyme. Both solutions above the filter were then added to an SDS PAGE (Figure 4-3). Data indicates that we are getting successful conjugation to the mutant VLPs, albeit limited. Another ligand, mTFP1-LPETG-histag, was also tested for sortase A ligation to the cut BMV CP loop1 proTEV VLPs, but we were unable to detect conjugation via SDS PAGE. This indicates that not all proteins/ligands

that contain an LPETG-Histag C-terminal tail can be conjugated to the mutant loop1 proTEV/polyG BMV CP VLPs with equal efficiency. Each ligand must be tested individually.

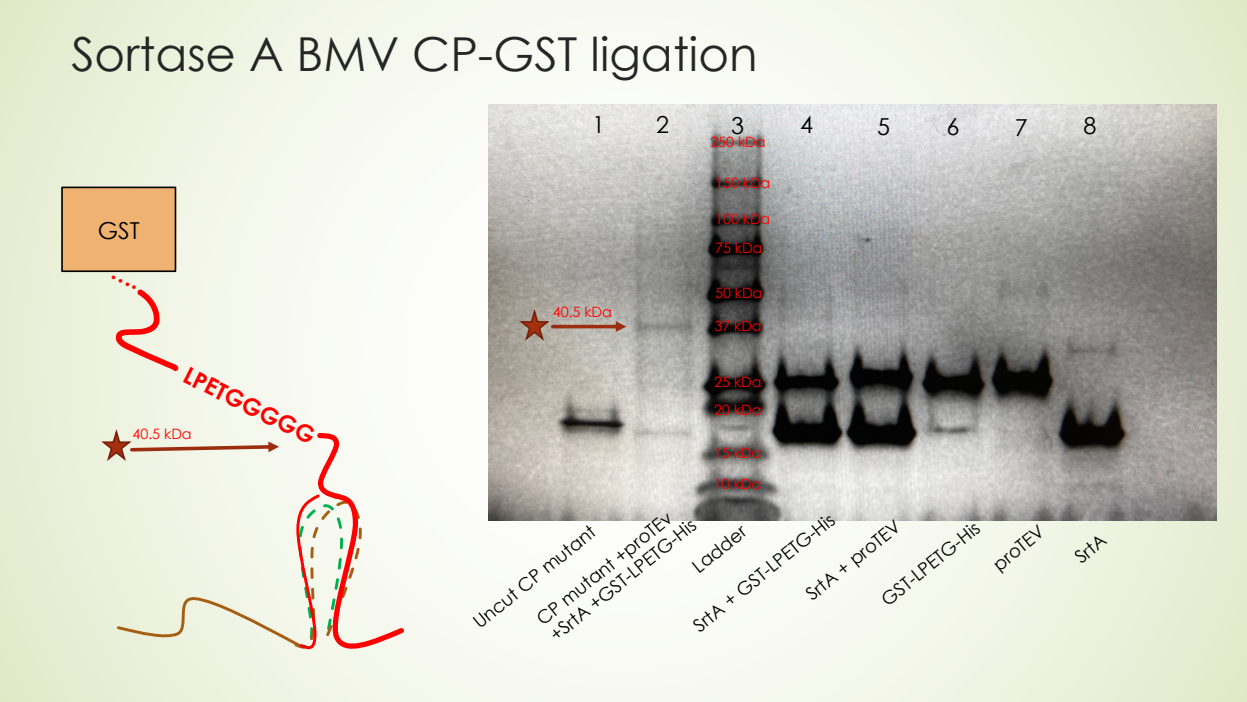

Figure 4-3: SDS PAGE of BMV CP loop1 proTEV/polyG to GST ligation via sortase A. Lane 1 is the CP loop1 proTEV/polyG VLPs uncut (negative control); Lane 2 (CP-GST protein diagram on the left is what reflected in lane 2) It contains the mutant CP + SrtA + proTEV + GST-LPETG-His; Lane 3 is the ladder; Lane 4 is SrtA + GST-LPETG-His; Lane 5 is SrtA + proTEV; Lane 6 is just GST-LPETG-His ; Lane 7 just the proTEV; Lane 8 is just the SrtA. CP mutant = BMV CP loop1 proTEV/polyG VLPs; SrtA = Sortase A enzyme; proTEV = TEV protease; GST = Glutathione S-transferase. The red arrow on the gel reflects the successful sortase A-mediated conjugation between the GST ligand and the BMV mutant VLP, the protein (half) depicted on the left in bold red (40.5 kDa final MW).

Since GST-LPETG-Histag has been successfully ligated to the mutant VLPs, we should be able to bind the VLPs to GSH resin because of the high affinity of GST tags for GSH resin. We should be able to first bind the VLPs to the resin, rinse the resin with buffer, and then elute the VLPs with a neutral buffer containing GSH (50mM Tris-HCl pH 7.5, 350mM NaCl, 10mM GSH). Finally, we can isolate the VLPs on an agarose gel and follow that up with EM. This would be another confirmation of successful ligation, as well as a purification step, isolating conjugated VLPs from non-GST conjugated VLPs.

Attaching GST ligands to VLPs is done merely to prove that the Sortase A conjugation can be effective. Once this is accomplished, we will switch to a more practical ligand protein like the human Xcl1 protein mentioned in Chapter 2 (Figure 2-3). As previously stated, the Xcl1 is highly structured with the exception of its C-terminal tail,[1] which is thus ideal for genetic fusion with the LPETG-Histag tail necessary for sortase A-mediated conjugation to VLPs. Additionally, in this case, we can ignore any leading primary amine that the protein may carry because here we are not using amine-reactive linkers; therefore, the protein can be bacterially expressed without issue. The xcl1 presence will be verified via spot/western blots.

Another immensely useful ligand that could prove invaluable for VLP targeting is "Protein Z", a 3-helix/59-residue polypeptide (see Figure 4-4) from *Staph. Aureus*. As mentioned earlier, this small protein has a high affinity for the Fc region of immunoglobins (antibodies),[2] thereby providing a "universal" ligand for conjugating antibodies to VLPs (indirectly). Significantly, only two of the three helices of Protein Z make direct contact with the Fc region of antibodies. The helix that is not involved in the interaction is the C-terminus of the Z protein,[2] making it ideal for fusing with LPETG-Histag tail as described above. With protein Z conjugated to the exterior, we could then functionalize the VLPs, non-covalently, with any antibody of our choosing, simply by incubating the protein-Z-conjugated VLPs with the antibody.

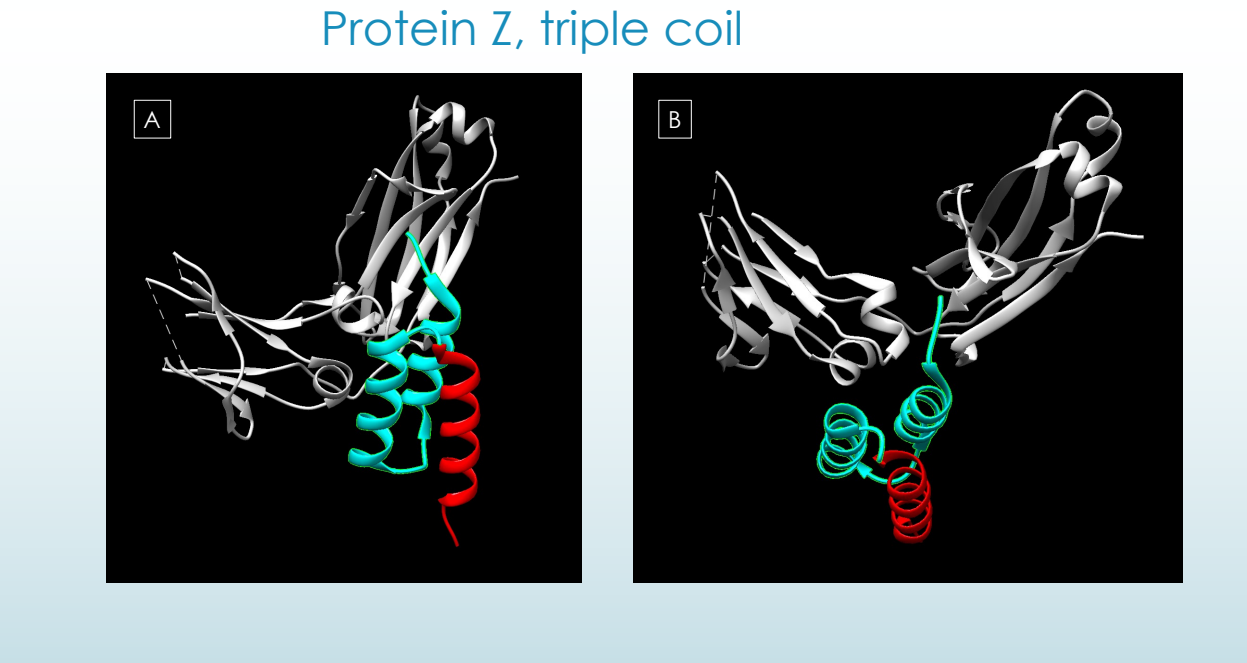

Figure 4-4: Crystal structure of Protein Z (aka "Protein A") bound to the Fc region of human IgG (grey structure). Protein Z is composed of three coils, with only the first two involved with antigen binding (cyan). Thus, a sortase A binding motif (LPXTG) could be fused the C-terminal tail of the final coil (Red), and used as a ligand for antibody conjugation. A and B represent two different angles of the same interaction.

Yeast-derived VLPs offer important guidance for which CP mutants will be *in vitro* assembly competent (able to self-assemble into RNase resistant protein shells around RNA), as previously mentioned. They also provide raw materials (CP) that we can use to *in vitro* assembly around our *in vitro* transcribed RNA of choice, provided we express this protein in sufficiently high levels. We would disassemble the yeast-derived VLPs into component parts using established protocols that we routinely employ with wildtype CCMV or BMV virions.[3] The challenging issue is that the CP mutants (or mutants in general) are typically far more unstable in solution (when unassembled) compared to the wildtype protein. Although this is not an issue for yeast cells where they assemble *in vivo*, this is an issue when attempting to assemble *in vitro*. More explicitly, the protein is difficult (but not necessarily impossible) to keep soluble at suitable protein concentrations that are needed for *in vitro* assembly with RNA, even at higher salt and with various detergents. So, part of the challenge of making the Sortase A labeled CP mutant VLPs to

be a viable end product is getting the CP mutants to behave properly, be soluble enough, so that they can successfully assemble around RNA as a protein shell *in vitro*. Once we can do that, then labeling post-assembly should be straightforward and simple. But how o we make the CP mutant soluble?

To achieve our goal of a functionalized VLP containing an *in vitro* transcribed RNA, there are a several approaches. Two are outlined here. Of the two, the simplest is to harvest the CP from the yeast-derived VLPs, then *in vitro* assemble around new RNA. But as previously mentioned, this will likely be difficult because the CP mutant will not be very soluble after disassembly. One potential remedy is to titrate wild type into solution with the mutant CP. This will effectively "dilute" the mutant protein to sustainable levels. This has the added benefit of exploiting the protein "crowding effect." a phenomenon associated with enhancing the solubility of mutant proteins by essentially sequestering them from one another in the presence of a much larger number of soluble (wildtype) proteins.[4] Also, although the yeast experiments showed that loop1 was more accommodating to genetic insertions while maintaining RNA packaging capabilities, we do not necessarily need (or want) a VLP that is entirely composed of labeled mutant CPs. Of the 180 CPs that typically make up a VLP, we actually prefer to have as few as five to ten to be labeled. A few is enough for VLP targeting, and also maintains the efficiency with which RNA content is made accessible to the ribosomal machinery. Therefore, diluting the mutant CP with wild type not only keeps the mutant CP more soluble but also is advantageous for establishing an ideal number of ligands per VLP.

Another approach is to N-terminally fuse a soluble tag like the Maltose Binding protein (MBP), to the CP mutant. This protein can be expressed in a bacterial system for convenience. The MBP can be cleaved by a protein different from the one that cleaves the proTEV/polyG site (Figure 4-5). Significantly, assembly with the MBP-fused CP is not possible because of its size; accordingly, it must be removed before *in vitro* assembly is attempted. In this scheme the CP

mutant is made more soluble before assembly. To do this we can first cut the loop1 proTEV/polyG BMV CP with proTEV (Figure 4-5a), then ligate our ligand with sortase A (Figure 4-5b). This is followed by cleavage with Xa-protease to free the labeled CP of its MBP tag (Figure 4-5c), which should be more soluble now that it is covalently attached to the ligand. From there we can mix it with wild type CP and attempt assembly around an RNA of choice. These two strategies are as yet untested. But the protein outlined above in Figure 4-5 has been expressed, purified, and is ready for testing.

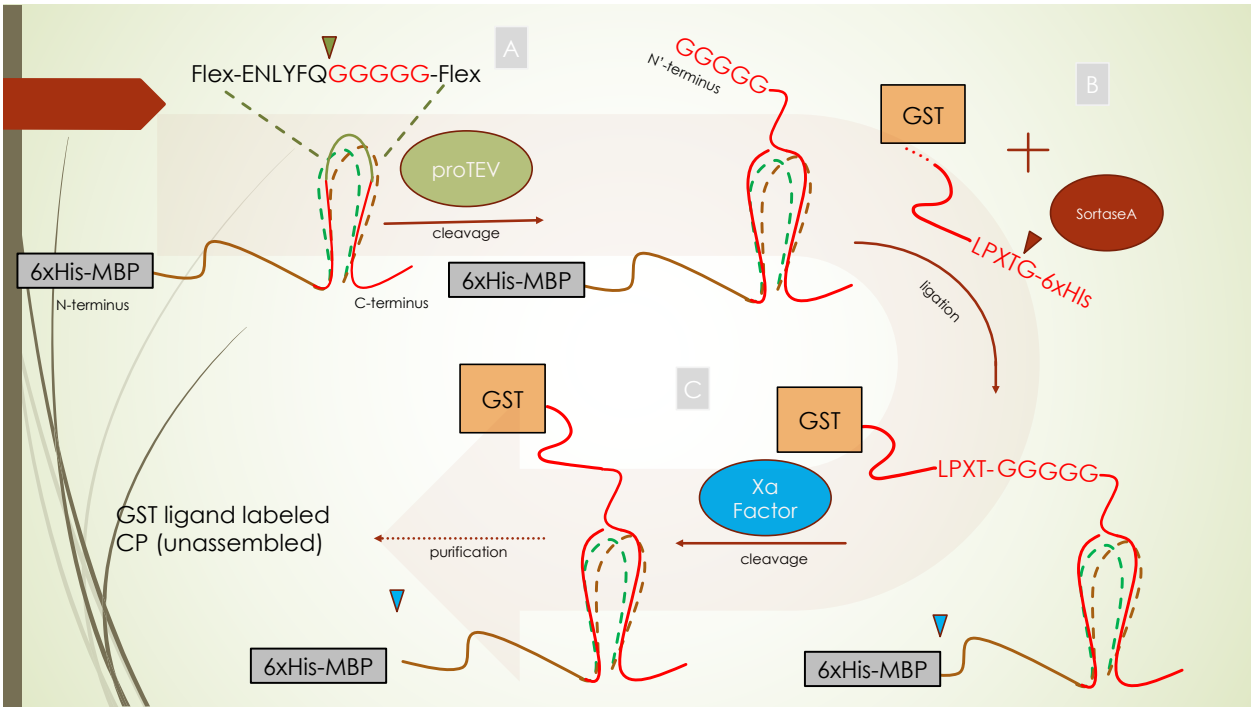

Figure 4-5: Flow chart indicating how a ligand can be ligated to mutant CP via Sortase A enzyme to avoid insolubility. A CP containing the genetic proTEV/poly-Glycine insert is cleaved at the proTEV cut site (A) (green arrow) producing a free new N-terminal polyG . A ligand (here GST) genetically fused to the sortase-specific binding motif LPXTG is introduced along with the Sortase A enzyme. The enzyme binds the sortase-specific motif (LPXTG) at the red arrow (B). Then the polyG of the cut CP displaces the Sortase A enzyme, forming a covalent bond at the C-terminus of the ligand. Next, another protease is introduced (Xa factor), which cleaves off the His-MBP tag (C). This results in a (more) soluble CP protein that is also labeled, ready for downstream assembly.

References

1. Fox, J. C., Thomas, M. A., Dishman, A. F., Larsen, O., Nakayama, T., Yoshie, O., Rosenkilde, M. M., & Volkman, B. F. (2019b). Structure-function guided modeling of chemokine-GPCR specificity for the chemokine XCL1 and its receptor XCR1. *Science Signaling*, *12*(597). https://doi.org/10.1126/scisignal.aat4128
2. Braisted, A. C., & Wells, J. A. (1996b). Minimizing a binding domain from protein A. *Proceedings of the National Academy of Sciences*, *93*(12), 5688–5692. https://doi.org/10.1073/pnas.93.12.5688
3. Garmann, R. F., Comas-Garcia, M., Knobler, C. M., & Gelbart, W. M. (2015). Physical Principles in the Self-Assembly of a Simple Spherical Virus. *Accounts of Chemical Research*, *49*(1), 48–55. https://doi.org/10.1021/acs.accounts.5b00350
4. Christiansen, A., Wang, Q., Cheung, M. S., & Wittung-Stafshede, P. (2013). Effects of macromolecular crowding agents on protein folding in vitro and in silico. *Biophysical Reviews*, *5*(2), 137–145. https://doi.org/10.1007/s12551-013-0108-0

5. ENZYME AND CHEMISTRY FREE PRESENTATION OF PROTEIN LIGANDS ON VIRUS-LIKE PARTICES USING SELF-COMPLEMENTING CAPSID PROTEIN FRAGMENTS

Self-complementing (or "split") proteins can best be illustrated by the unique properties of the Green Fluorescent Protein (GFP).[1-2] The tertiary structure of GFP is that of a beta barrel housing a chromophore inside, with the outer barrel composed of 11 beta strands. These strands serve to create a unique environment that acts on the small peptide in its core, refolding and maturing the peptide into an active chromophore whose fluorescence ensues from its interaction with the beta barrel.[1] Since the chromophore does not become active without the final (11th) beta strand, this makes GFP an ideal proximity sensor, as long as the final strand can be introduced separately. And this is possible due to the high stability of this protein's beta barrel structure. More explicitly, when GFP_{1-10}[3-4] (i.e., containing only beta strands 1-10) is expressed without its final strand, it is stable in this truncated form, awaiting its missing piece. Further, the final piece is a small beta strand that can be fused to any protein of interest (POI). This POI-11^{th}strand fusion then binds GFP_{1-10}, completing the GFP beta barrel and providing a scaffold for the POI, while at the same time "turning on" the fluorescence of the peptide chromophore.[1-2] This strategy has been exploited in a wide range of applications, including visualization of subcellular protein co-localization,[5] quantification of protein aggregation,[6] detection of cytosolic peptide delivery,[7] identification of cell-cell contacts,[4] and target protein scaffolding assemblies[3].

In the present work, we exploit an analogy between the 11-strand beta barrel of GFP and the 8-strand beta barrels of many viral capsid proteins (CPs). Instead of fluorescence "reporting" of the re-association of split GFP pieces, the complementation of split CP pieces results in the assembly of virus-like particles (VLPs). As in the GFP scenario, the "missing" beta strand is fused to a POI, which can be an antibody, antigen, cytokine (or a GFP!), et cetera. And because of the position of the "missing" beta strand (in particular the "break point" between it and the rest of the CP) in the quaternary structure of the viral capsid (see Figure 5-1a), the POI ends up being present on the outside of the capsid, positioned just where we want it for binding to target proteins.

Turnip Yellow Mosaic Virus (TYMV) is a well-studied 28nm-diameter icosahedrally-symmetric plant virus whose capsid is composed of 180 identical proteins.[8] The individual capsid proteins are, in essence, each a pseudo beta barrel not too dissimilar to that of GFP. Specifically, this barrel features the typical 8-stranded "jelly roll" fold that is found in many viral proteins, particularly in viral capsid proteins, and is composed of two distinctly curved beta sheets that mimic a GFP beta barrel.[9] The principal stability factor in TYMV virions resides in the strong protein-protein interactions. As a consequence, unlike other plant virus capsid proteins, the TYMV capsid protein will assemble into empty VLPs when expressed in bacteria.[8] Our aim is to temporarily halt this process by fusing a large protein tag (i.e., Maltose Binding Protein [MBP]) to the N-terminus of the CP. The assembly can then be rescued with the cleavage of the tag, under typical physiological conditions (Figure 5-2). It is because the TYMV capsid proteins have such strong affinity for each other that controlled *in vitro* assembly of their empty capsids is possible.[8,10] Further, because of the exceptionally strong non-covalent interaction between its truncated (7-strand) beta barrel and its missing (8th) beta strand, this particular capsid protein is a good model system for the new VLP labeling strategy that we report here. The idea is to then apply this approach to other virus capsid proteins–and, in particular, ones from the spherical/icosahedral CCMV and BMV viruses–that are uniquely capable of *in vitro* assembly when combined with purified genomic or heterologous RNA. The ultimate goal is to exploit the split-protein/ligand-fusion strategy for viral capsid proteins that (unlike TYMV CP) *in vitro* package heterologous RNA into nucleocapsids, thereby allowing the reconstitution of VLPs conjugated to an arbitrary protein of interest and that contain arbitrary therapeutic mRNA.

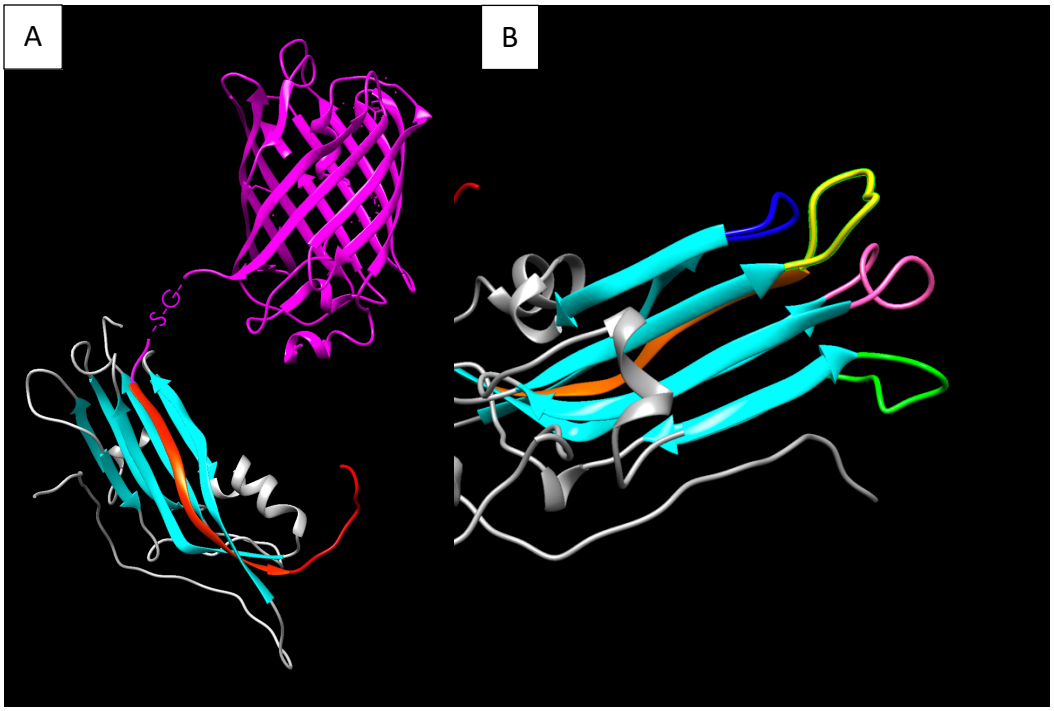

Figure 5-1: Ribbon diagrams of TYMV. Left (A): A ribbon diagram depicting the binding of TYMV Protein A with Protein B. Protein A, 1- 164aa, is labeled in grey with its beta strands in cyan color; the N-terminus is labeled with a grey arrow and the C'-terminus is labeled with a cyan arrow. Protein B, mTFP1-166-191aa, is shown in magenta with the beta strand tail in orange; the C-terminus (red) is indicated with a blue arrow. The S-G is a short flex linker that serves as a bridge between the two (8^{th}-beta-strand and mTFP1) domains, and it is shown with a white arrow. Protein B provides the missing 8^{th} (and last) beta strand, thus completing the beta barrel. The beta strand "locks" into place within barrel of Protein A, in the same manner as in the intact wildtype capsid protein. Right (B): TYMV CP showcasing the different surface exposed loops. Loop 2 (yellow) is the most peripheral (radially-outermost) of all four surface loops. The blue 45, yellow P165, pink P85, and G120 are the amino acid representatives of protein turn-loops 1, 2, 3, and 4, respectively.

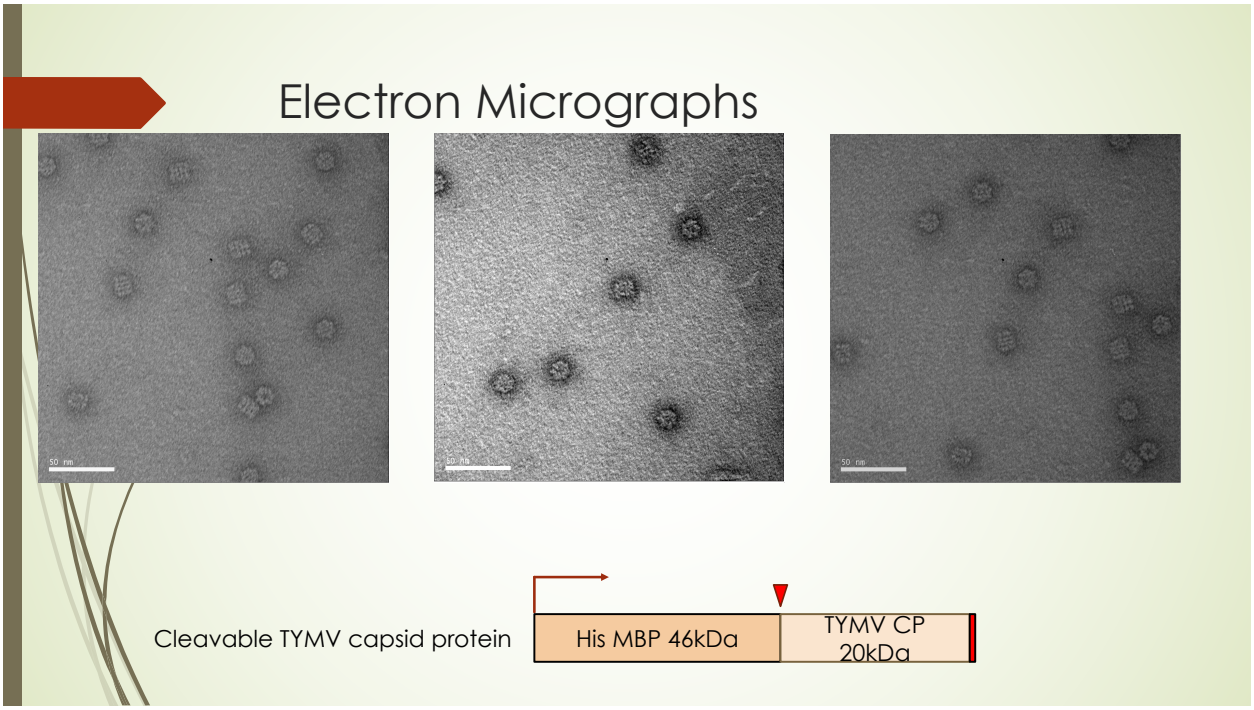

Figure 5-2: Electron Micrographs of TYMV wt. BOTTOM: A schematic diagram showing the wildtype TYMV CP with an N-terminal His-MBP tag. This tag is cleavable (red arrow) by a proTEV enzyme, and cleavage of the tag initiates *in vitro* VLP formation when the protein is incubated in 350 NaCl/50mM Tris HCl, pH 7.5 buffer solution. TOP: Images reflecting different fields-of-view of the resulting TYMV-wild-type-CP VLPs. The scale bar is 50nm, and the monodisperse, ordered, VLPs are 15-20nm in diameter.

Some protein turns of the beta strands in the TYMV beta barrel are exposed at the outer surface of the capsid shell. This is especially true for the last protein turn connected to the last beta strand of the beta barrel. Adopting the idea of GFP split protein described above, it is possible to: "remove" (i.e., delete from the CP expressed in bacteria) only the last beta strand of TYMV capsid protein (in particular, the C-terminal end consisting of residues 165-189, see Figure 5-3); to fuse this last strand to a protein of interest (e.g., a targeting or fluorescent ligand [see mTFP1 in Figure 5-3]); to express, *separately*, this fusion and the remaining (7-beta-strand N-terminal) portion of the TYMV capsid protein (residues 1-164); and to then mix them to allow for the self-complementation of both parts to form a new "monomer"/"intact" protein. We can do this in bulk solution, forming 180 such "monomers" that are able to assemble together into a VLP. This VLP will display 180 ligand proteins on its outside surface, each fused to the 8th beta strand of a capsid protein. Unlike GFP, which exposes all beta strands at the protein surface, TYMV's beta strands

are not exposed on the VLP surface but are instead hidden and tucked between adjacent coat proteins. This is especially the case for the capsid protein C-terminal tail that is an extension of the last beta strand. This tail is well hidden within the "body" of the TYMV protein shell. Accordingly, it is very unlikely that this strand, and its fused protein partner, will be able to escape its beta barrel, and by extension the quaternary structure of VLP post assembly. This will make this enzyme-less and chemistry-less split protein conjugation as reliable as other modalities that involve covalent bond attachments.

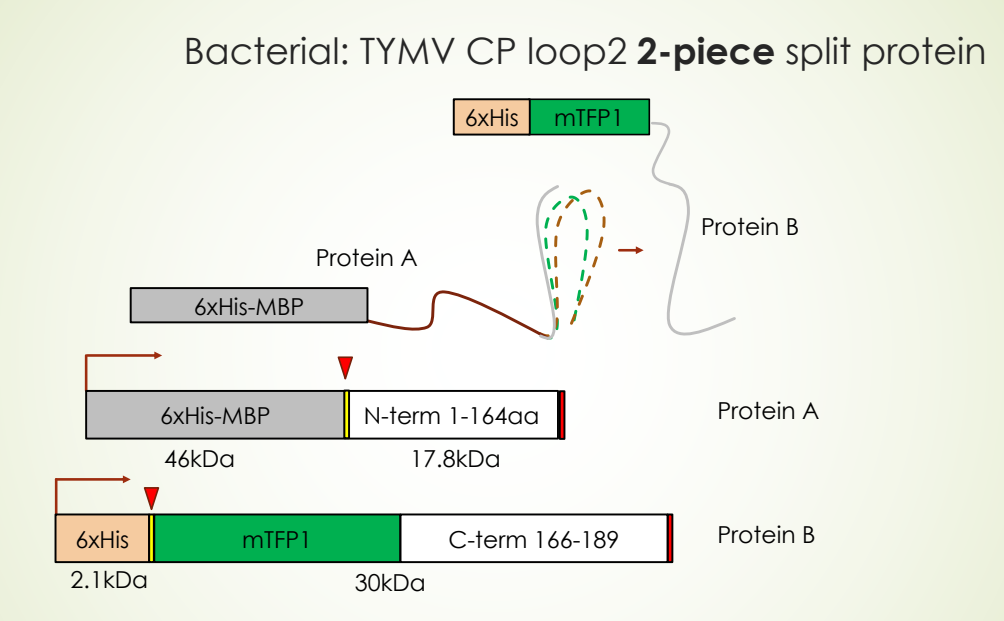

Figure 5-3: Split protein components of the TYMV capsid protein (CP), involving a genetic fusion of the mTFP1 protein with the 8th and last beta strand of the TYMV CP. BOTTOM: A schematic diagram showing the two TYMV split protein genes (not drawn to scale. Protein A features a his-MBP N-terminal tag fused to TYMV 1-164aa. Protein B features a histag fused to a mTFP1 fluorescent protein followed by the C-terminal 166-189aa of TYMV. TOP: Simplistic view of how the two proteins will come together to form a split protein (sp) "monomer". Red arrows indicate proTEV cut sites.

Of the four surface-exposed protein turn-loops, loop2 (numbered from the center of the capsid hexamer/pentamer outward, radially, see Figure 5-1B), which contains amino acid 165, is by far the most peripheral (most radially outermost) of the four and thus offers the most ideal position for inserting a fused ligand. Accordingly, we expressed the TYMV protein in two separate proteins, each with its own N-terminal fused tag. MBP protein is ideal for the N-terminus of the TYMV capsid protein (1-164aa) because we have found that it is large enough to

inhibit assembly. This protein is termed Protein A. The C-terminus of TYMV (166-189aa) contains a fused ligand (here referenced fluorescent protein mTFP1, Figure 5-3). This ligand-fused protein is termed Protein B. Amino acid P165 was not included in either protein in order to make room for ligand protein backbone.

Proteins A and B were expressed in bacteria separately. This separation was done because we were not sure if Protein B (His-mTFP1-166-189) was soluble enough to handle low salt conditions upon bacterial cell lysis. Typically, mutant proteins have trouble staying soluble, and to help solubility, we usually increase the salt concentration. This would run contrary to our goal of Protein A/B association which requires low salt conditions. It was surmised that if we lysed the bacteria in low salt, Protein B would crash out immediately. Protein A does not have this issue since MBP is very soluble in most salt conditions, including low salt. Therefore, it was best to use higher salt to purify Protein B alone, then dial down the salt afterwards. We would get more yield this way. Once proteins A & B were in hand, I mixed the two together at a ~1:2 mole ratio; for every Protein A, there are two of the Protein B present (Figure 5-4A). We wanted complete saturation (binding) of Protein A, with Protein B in excess. This mixture was then dialyzed overnight into low salt (50mM NaCl, Tris-HCl 7.5), with the result that proteins A and B associate (Figure 5-1 and 5-4B). Next we added the proTEV cutting enzyme to cleave off the MBP tag of Protein A and the 6xHis tag of Protein B (Figure 5-4C). The cleavage reaction was allowed to run for 9 hours, and during this time the newly associated “monomers” undergo spontaneous assembly into VLPs (Figure 5-4D).

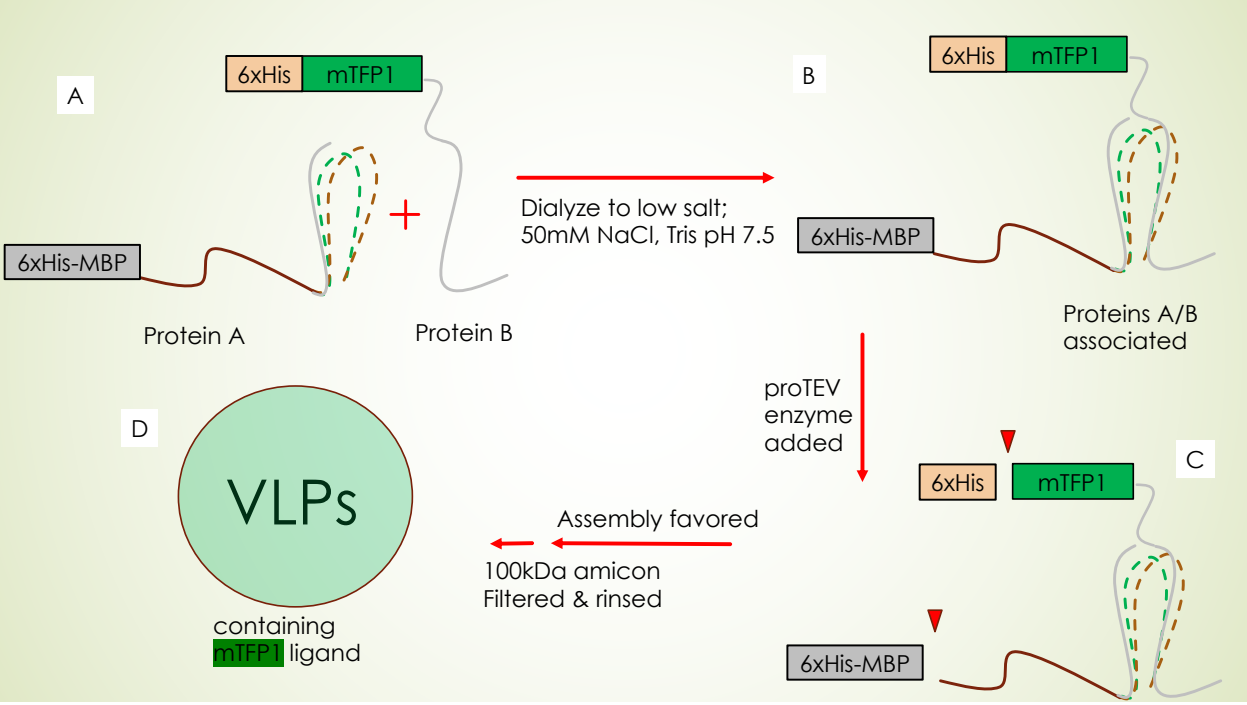

Figure 5-4: Flow chart indicating how TYMV VLP assembly can occur in vitro. Proteins A and B (16A) come together under low salt conditions (16B). Then we cleave the tags off both proteins by adding in a proTEV enzyme (16C). From here the assembly is spontaneous (16D). The putative VLPs can be isolated from an excess (unassembled) proteins via Amicon filtering.

This putative assembly mix, which included (along with the newly assembled VLPs) all "spectating" or non assembled proteins (uncleaved and cleaved Protein A, Protein B, 6xHis and 6xHis-MBP tags, and proTEV enzyme) was then applied over a 100kDa Amicon. The Amicon was rinsed ~9x times with a suitable neutral pH buffer (50mM Tris-HCl pH 7.5, 350mM NaCl). This allows for the removal of anything smaller than 100kDa, i.e., everything but the newly assembled VLPs. Since the solution above the filter was still green, it appears that this solution contains mTFP1-integrated TYMV VLPs (Figure 5-5A).

Figure 5-5: Synthesis of in vitro assembled TYMV split-protein (sp) VLPs. Left, A: mTFP1-containing VLPS after proTEV cleavage, Amicon filtering, and several washes. Middle, B: 1% TAE agarose gel with VLPs and Amicon flow through. Right, C: VLPs extracted from gel.

The solution recovered from the filter, and the mTFP1 proteins in the Amicon flow-through, were applied separately to a 1% TAE agarose gel, where the putative VLPs run faster than the Protein B recovered from the flow through (Figure 5-5B). Finally, we gel-extracted the band in lane 1, electro-eluted the proteins, and concentrated them using a new Amicon 100kDa (Figure 5-5C).

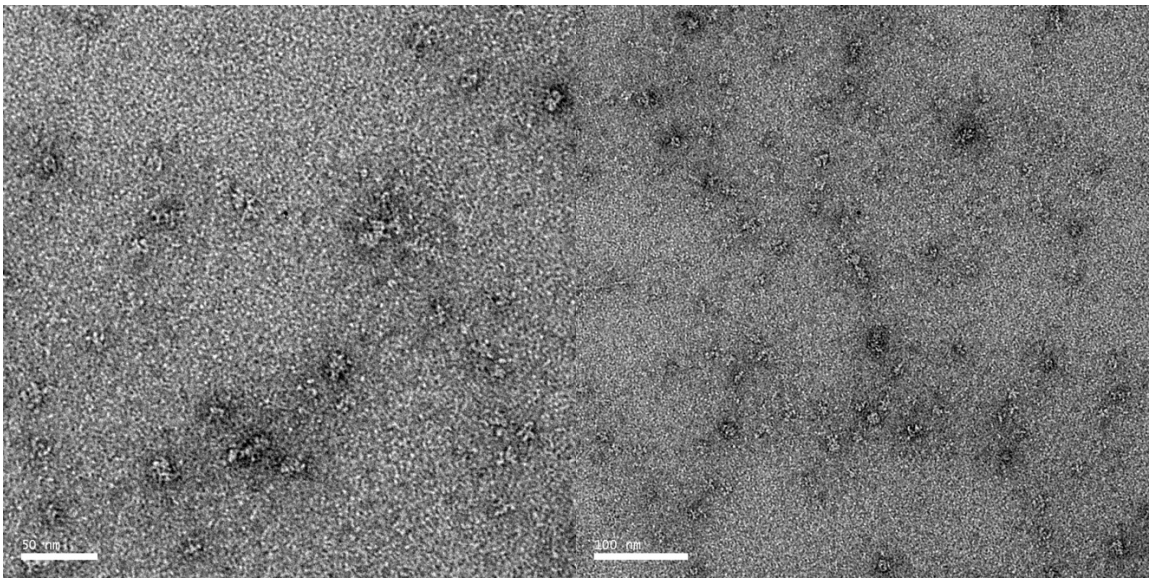

Figure 5-6: Different field-of-view images of 1:0 sp "monomer": wild type TYMV VLPs. The sample consists mostly of disordered protein aggregates, due to steric hindrance effects associated with the large TFP1 ligand – see discussion in text.

Although it is evident that the “monomers” (heterodimers!) of associated Protein A/B can form structures greater than 100kDa, it is also possible that they cannot form complete VLPs due to mTFP1 steric interference. Indeed, according to electron micrographs (Figure 5-6), there seems to be mostly protein aggregation with only a few malformed VLPs. These VLPs appear smaller than the wild type VLPs. *In vitro* assembled empty wild type VLPs are smaller than plant derived virions, ranging in diameter from ~20-24nm (Figure 5-2). Therefore, it is reasonable that the split protein “monomer,” each of which carries an mTFP1 ligand that ~27kDa, would have trouble forming intact VLPs due to sterics, i.e., there simply is no room for the bulky 27kDa ligands to be part of neighboring capsid proteins.

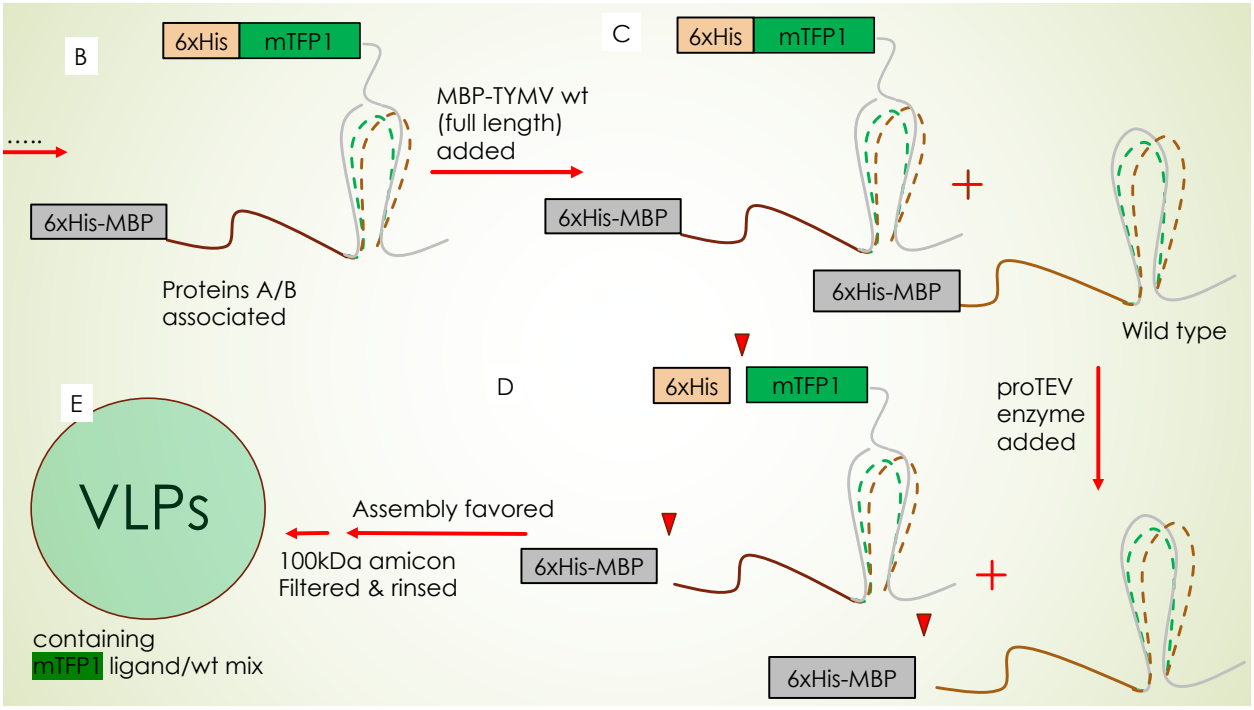

Figure 5-7: Flow chart indicating how TYMV VLP assembly can occur *in vitro,* with both mutant-split proteins and wild-type proteins. Proteins A and B come together under low salt conditions, as described previously, to form mutant-split proteins (19B). Then we can titrate in wild-type CP (19C) at any particular mole equivalent (e.g., 1:3, and 1:5. Next, we cleave the tags off all protein (including wild-type) by adding in a proTEV enzyme (19D). From here the assembly is spontaneous; ideally, the wild type is integrated with the mutant proteins to form VLPs (19E).

To alleviate this potential crowding, I have introduced MBP-TYMV wild type (full length) into the associated Protein A/B mixture (Figure 5-7C). I've done this at a ~1:3 mole ratio; for every mutant "monomer" there should be three wild type monomers present. Theoretically, this should form VLPs that are peppered with mutants upon proTEV cleavage (Figure 5-7D), i.e., upon release of N-terminal MBP tags (on both mutant and wildtype proteins) and subsequent assembly of the VLP. This is intended as a method of dispersing the mTFP1 ligand (~27kDa)-labeled split capsid proteins amongst wild-type capsid proteins (~20kDa), increasing the odds of a successful assembly that includes the mTFP1 ligand (Figure 5-7E). The negative-stain electron micrograph of this sample shows us a definite improvement in VLP formation (Figure 5-8). Although we still see malformed VLPs, these VLPs are mostly intact and more ideal looking VLPs (ones with good size and symmetry) can also be found although they are few in number.

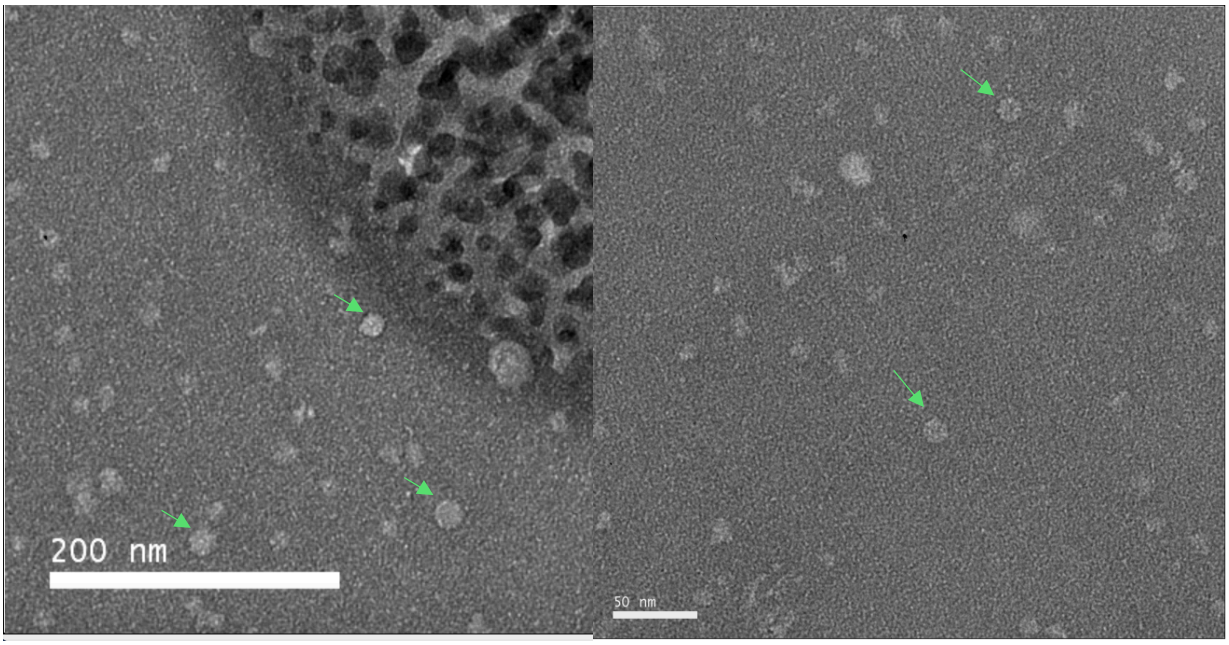

Figure 5-8: Different field-of-view images of 1:3 split-protein "monomer": wild-type CP TYMV VLPs. Green arrows indicate well-formed spherical VLPs. All other VLPs are malformed to different extents. Both images include a mixture of both kinds of VLPS, malformed and well-formed.

Clearly it is important to verify that the particles we see in our EMs (Figure 5-6 and Figure 5-8) are VLPs (or amorphous particles) which have both mTFP1-containing Protein B and Protein A, and not just Protein A alone, or wildtype protein alone, or a combination of Protein A and wildtype protein. Accordingly, we ran an 1% TAE agarose gel with both split-protein samples (without and with wildtype protein) while including two negative controls, Protein A and wildtype protein. Our goal is to see that the split protein (Protein B/mTFP1-containing) VLPs run differently down the gel compared to any VLP (or superstructure) that Protein A or wild type, alone, produces. Indeed, split protein VLPs run much slower in the gel compared to the negative control samples (Figure 5-9). Since this is the case, and since we specifically gel-extracted and imaged only the fluorescing bands, we can conclude with confidence that the VLPs we see in the EM images are true split protein VLPs.

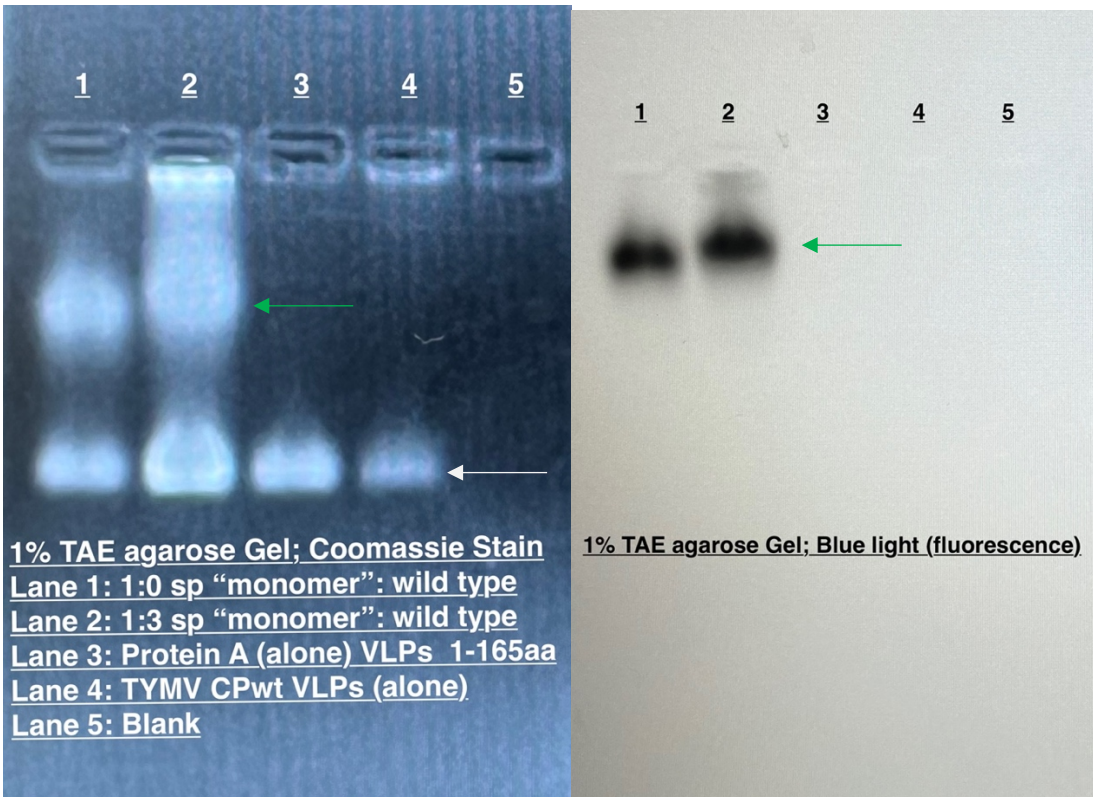

Figure 5-9: 1% TAE gel of 1:0 and 1:3 split-protein (sp) "monomer": wild type VLPs, Protein A, and wildtype CP VLPs, as imaged with by Coomassie stain (left) and fluorescence (right). Samples were run for 45min at 70V. Green and white arrows indicate VLP structures that *have* and have *not* integrated a fluorescent Protein B mTFP1 molecule.

It appears that as we alleviate steric crowding of the split protein "monomer" by diluting them with wildtype proteins, we are allowing for more ideal VLP formation. This idea has been tested by increasing the mutant:wt mole ratio to 1:5.

Since we were confident a 1:5 sample would give us VLPs, we also wanted to increase the rigor by which we test for ligand display. To this end we utilized an Xa-factor protease cut site that was adjacent to and on the N-terminal side of the proTEV cut site in Protein A in order to selectively cleave the MBP from Protein A *and at the same time leave the 6xHis fused to the N-terminus of mTFP1 on Protein B* (see Figure 5-10). This preserves the his tag on the mTFP1 ligand for nickel resin binding.

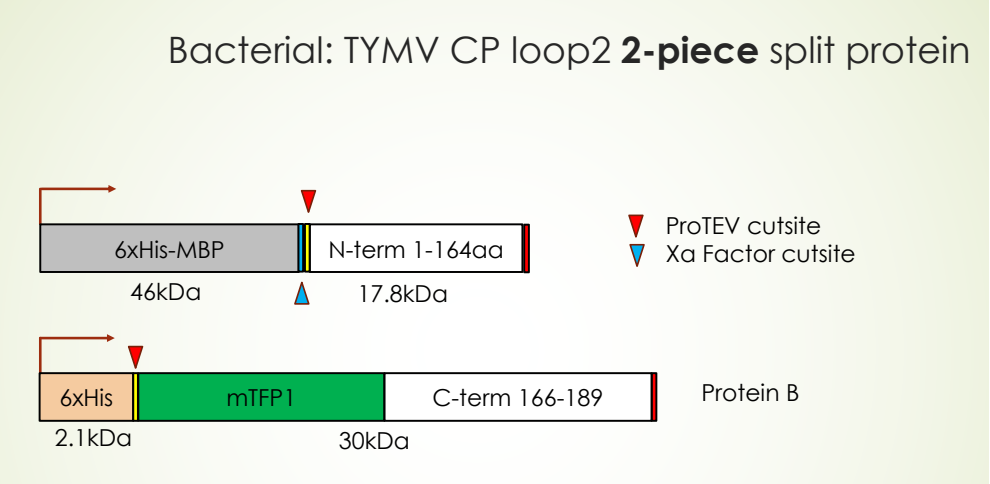

Figure 5-10: *A schematic diagram showing the two TYMV split protein genes*. The Xa factor cut site (blue arrow) is present just next to the proTEV cut site (red arrow) on Protein A. The Xa factor cut site is unique to Protein A.

After initiating VLP formation via Xa factor cleavage of Protein A (ridding of its MBP tag that inhibits self-assembly), we applied the mix to a PBS-sucrose gradient, and put the sample through ultracentrifugation for 6hr at 42000 RPM. The results were ideal, with the excess Protein B (top band) separating nicely from the His-mTFP1-containing VLPs (bottom band) (Figure 5-11).

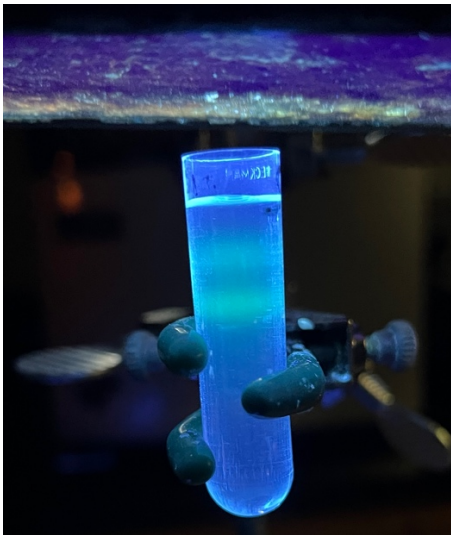

Figure 5-11: Mutant-split protein VLPs run in PBS sucrose gradient. Here the assembled 1:5 sp "monomer": wildtype CP VLPs (bottom band) are separated from excess Protein B (top band). Sample was spun at 42000 rpm for 6hrs. Both bands fluoresce under UV light.

The bottom band was collected and applied over nickel resin. The resin was rinsed and then washed with a neutral pH buffer containing 25mM Imidazole to ensure that the VLPs tightly bound. The results can be seen in Figure 5-12, which show an abundance of mTFP1 presence post wash.

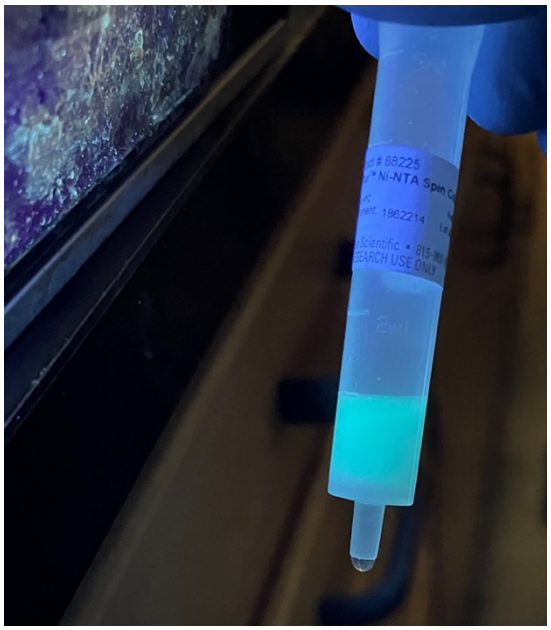

Figure 5-12: Mutant VLP bound Nickel resin. Here, 1:5 sp "monomer": wildtype CP VLPs are shown tightly bound to the Nickel resin post Imidazole wash. We can observe the mTFP1 labeled VLPs fluorescing (green color) under UV light.

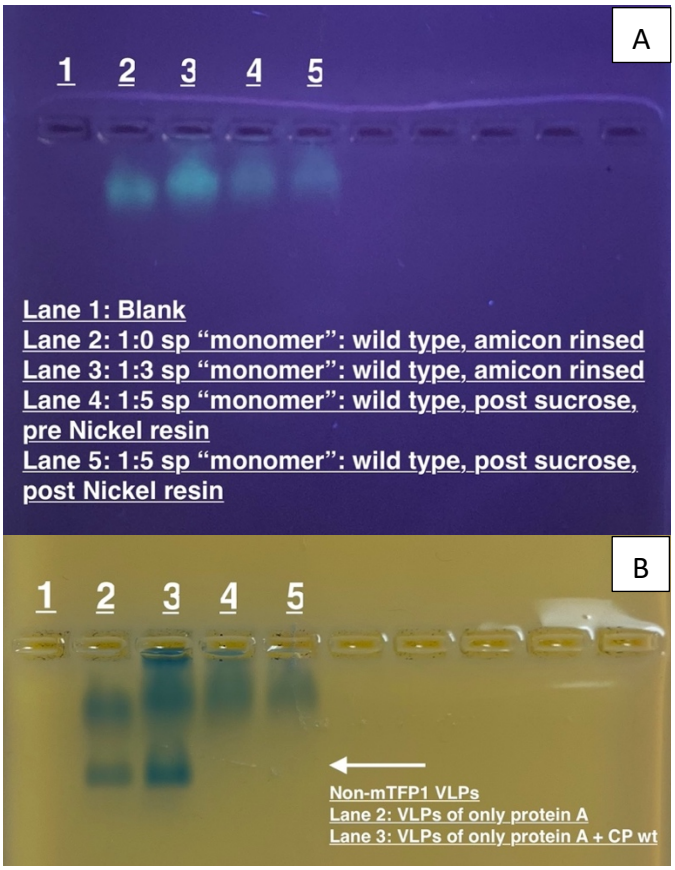

Figure 5-13: 1% TAE gel of various (1:0, 1:3, and 1:5) sp "monomer": wildtype CP VLPs. Samples were run for 45min at 70V. Lanes 2 and 3 involve VLPs whose formation is initiated by proTEV cleavage, whereas lanes 4 and 5 involve VLPs formed by Xa cleavage. TOP: VLPs showing fluorescence, illuminated under UV light. BOTTOM: Same gel as the top, but under Coomassie stain. White arrow indicates VLP structures that have not integrated a fluorescent mTFP1 molecule.

The sample was then eluted and run on a TAE 1% agarose gel (Figure 5-13) and EM imaged (Figure 5-14). The EM images show that most particles we see are malformed/aggregates of some similar mass compared to the few well-formed VLP highlighted with green arrows. To get clearer images, the samples needed a final dialysis with a saltless buffer of just 75mM Tris-HCl pH 7.5. It is uncertain at this time if this saltless buffer further deformed the malformed/aggregates, allowing only the most stable and symmetrical VLPS to survive.

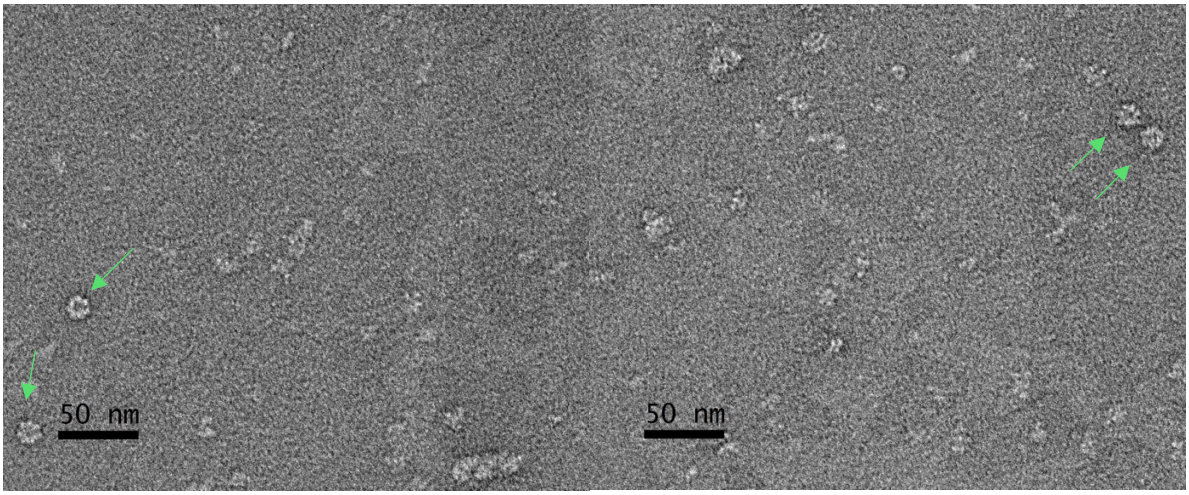

Figure 5-14: Different field-of-view images of 1:5 sp "monomer": wild type TYMV VLPs post sucrose-gradient purification and nickel-affinity chromatography. Green arrows indicate well-formed spherical VLPs. All other VLPs are disordered/malformed to different extents.

Throughout all the TYMV *in vitro* assembly work so far, we have produced VLPs that are significantly undersized relative to their native size of 28nm. This could potentially be remedied with assembling in higher salt, or including RNA during assembly, or both. The first of these options was tested. Up to this point we have induced assembly with a 50mM NaCl solution; now we are raising this to 350mM NaCl (50mM Tris-HCl). Also, we used three different protein ratios under these conditions: TYMV mutant:TYMV wt ratio of 1:0, 1:3, and 1:5. We then combined all three before applying them over a PBS sucrose gradient, as before. This time we used a blue light to illuminate the mTFP1-containing fluorescent bands. We found that under these conditions blue light is better for fluorescent band detection than UV light. Unlike with other conditions (or perhaps due to just the blue light usage), we saw three bands, with the bottom band being the least bright (Figure 5-15).

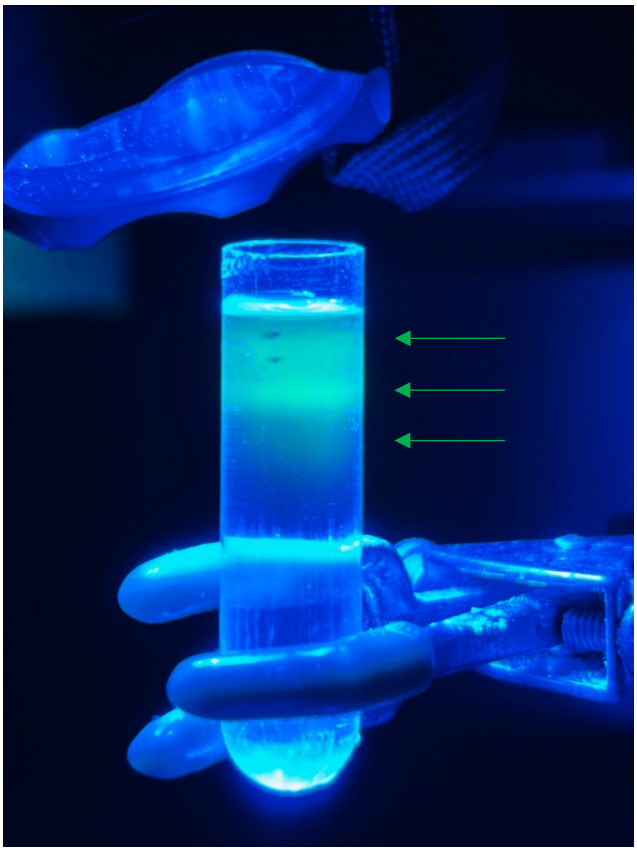

Figure 5-15: Mutant-split protein VLPs run in PBS sucrose gradient. Here the sample is a mixture of three different conditions (12ug total, each condition): 1:0, 1:3, and 1:5 sp "monomer": wildtype CP ratio assembled VLPs. All three conditions were combined and placed over a 10-40% PBS sucrose gradient. The tubes were spun at 40000 rpm for 3 hours. We observed the sample using a blue light and observed three bands (see green arrows). (Note that the light at the bottom is not fluorescence, but rather just a reflection of the blue flashlight.) We removed only the faint bottom band for further study. The top two bands here are the same two bands shown under UV light in *Figure 5-11*. Light on the very bottom of the tube is mere reflection off the centrifuge tube.

Of the three green bands, the bottom-most band was extracted, dialyzed against RAB (50mM Tris-HCl, 50mM NaCl), and imaged with EM (Figure 5-16).

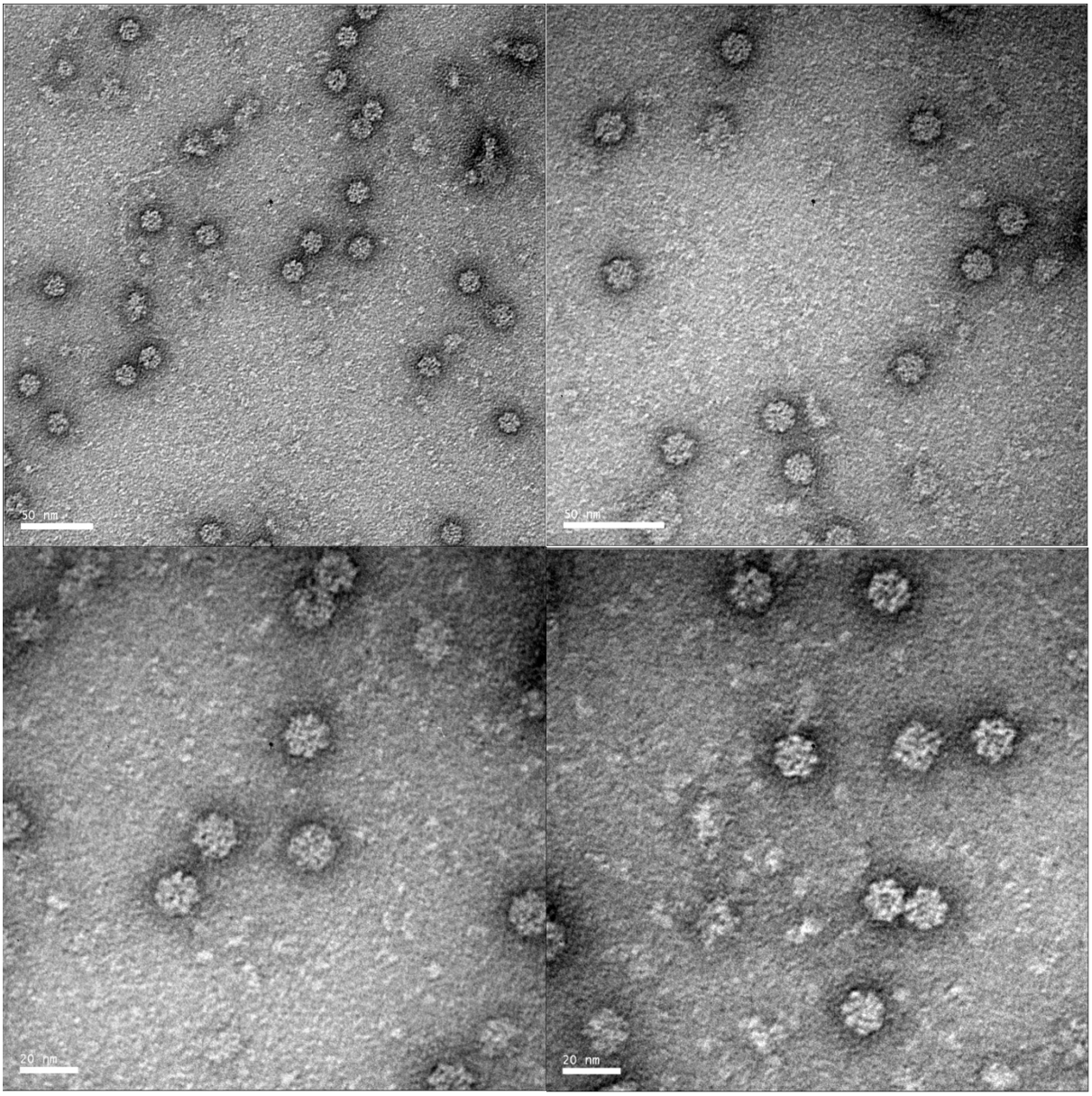

Figure 5-16: Different field-of-view EM images of combined sample outlined in Figure 5-15. Sample is a mixture of three different conditions (12ug total each condition) involving 1:0, 1:3, and 1:5 sp "monomer": wildtype-CP assembled VLPs. All three conditions were combined and placed over a 10-40% PBS sucrose gradient. The tubes were spun at 40000 rpm for 3 hours. We observed three bands. We removed and imaged the bottom band here.

Although we now have consistency in VLP formation comparable to that in Figure 5-2 (where the particles are formed from wildtype CP only), we still have undersized (compared to wildtype) VLPs with a diameter ~14nm instead of ~28nm. This whole process was repeated for only one of the three samples, the mutant:wt 1:5 (Figure 5-17).

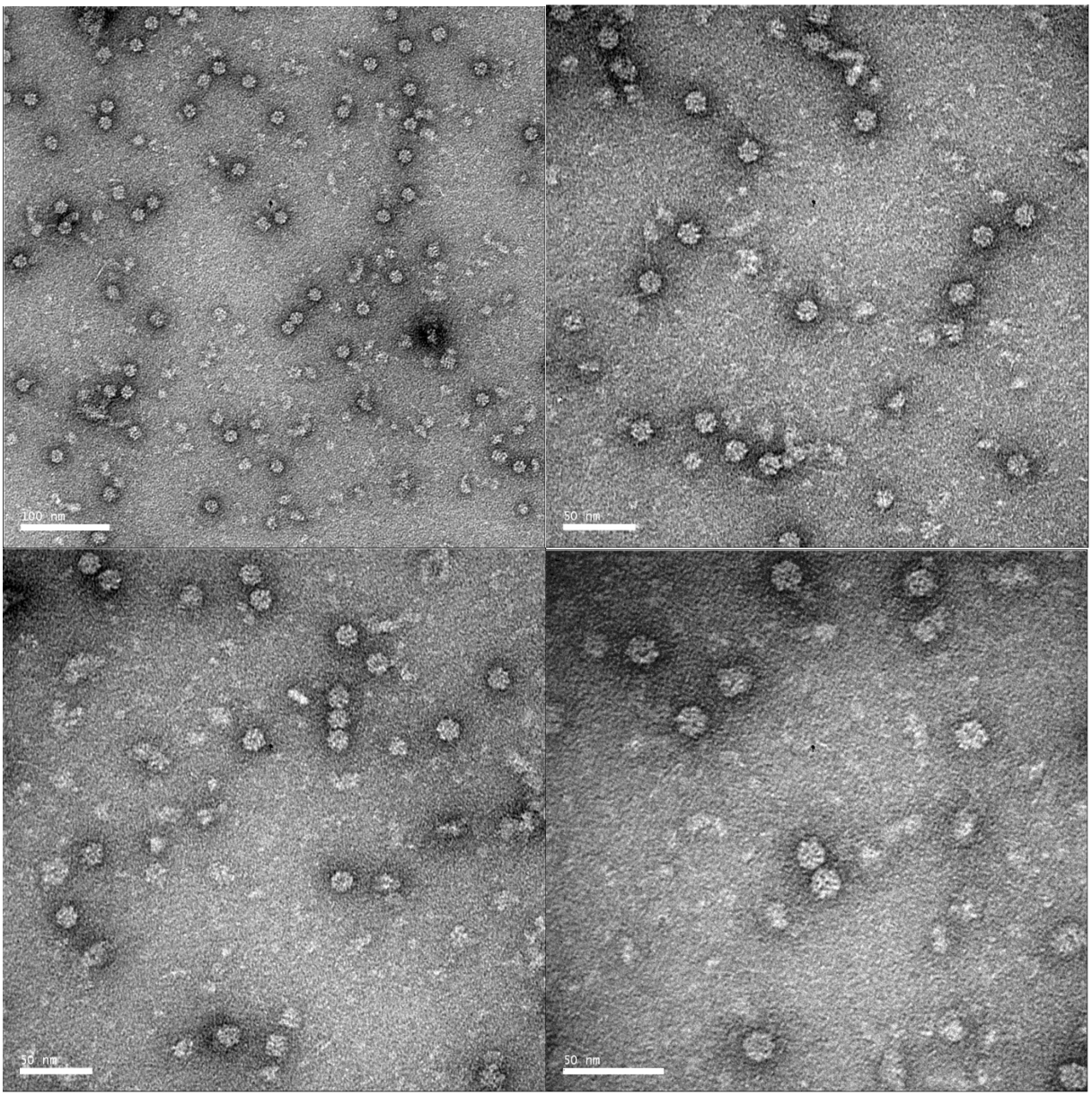

Figure 5-17: Different field-of-view EM images *of* 1:5 sp "monomer": wildtype CP assembled VLPs. All three conditions were combined and placed over a 10-40% PBS sucrose gradient. The tubes were spun at 40000 rpm for 3 hours. We observed three bands. We removed and imaged the bottom band here.

What we can gather from these data is that, at least to some extent the split protein concept is working—Protein A is associating with Protein B, and even forming (undersized) VLPs. Here, the ligand portion of Protein B is mTFP1, a ~27kDa protein, which is even larger than the wild type TYMV CP. Thus, integration of this Protein B is frustrated when it is forced into a VLP that is undersized, involving more steric hinderance. In other words, the larger the diameter of the VLP the more accommodating it will be toward the mTFP1 containing ligand

(Figure 5-18). Even without the consideration of split protein, VLPs smaller than wildtype virus are usually more unstable and less likely to effectively protect RNA cargo than one with native size (~28nm). Experiments set in the near future will involve assembly of TYMV CP around different length RNAs. We hypothesize that a suitable RNA will force the CP into VLPs native size (or near).

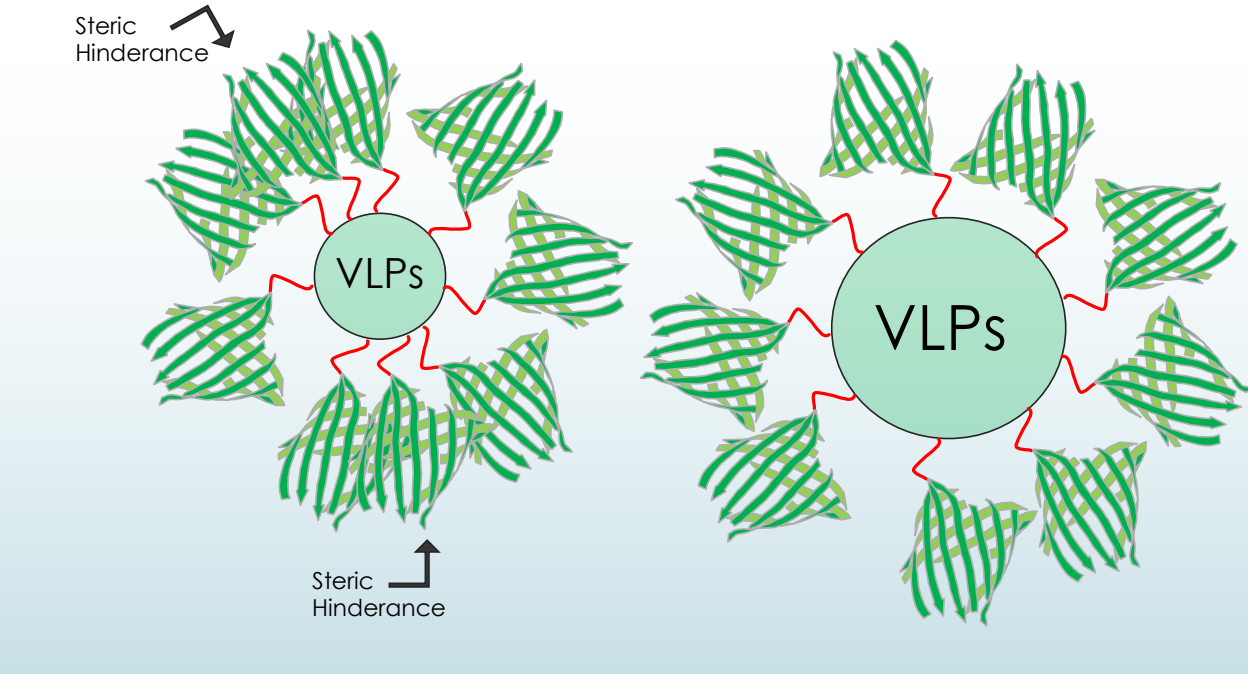

Figure 5-18: A drawing outlining ideal conditions for split protein ligand integration with VLPs. With smaller VLPs there is likelihood of VLP malformation and failed assembly due in part to steric interference during assembly. The ligands, here, mTFP1 proteins (left), are forced into an overly crowded local environment, which keeps them from forming ideal VLPs. And if smaller but relatively ideal VLPs are formed, there will be fewer ligands integrated versus a larger VLP (right). Getting TYMV to form VLPs at (or near) its native size (28nm) would be more ideal for mTFP1 ligand integration and overall stability by virtue of its larger size—less steric hinderance compared to the smaller particles we are currently observing.

Ultimately, we want an RNA-containing TYMV split-protein VLPs that will feature important ligands which have different functions. We want a "Protein B" tool kit, so to speak (Figure 5-19). Protein A will be compatible with a number of useful Protein B ligands. For example, we may need a specific antibody against an antigen, which could be met by a Single Chain Variable Fragment (ScFv) protein.[11] Or we may just employ protein Z to bind noncovalently with any antibody Fc region.[12] Xcl1 is another suitable protein to be used with this system, able to boost immune response to an antigen by targeting a subset of native DC cells.[13]

And if we want an even more universal docking feature, we could produce a N-terminal polyG in the same way we did in Chapter 4. A polyG would then be able to be connect to any LPETG containing protein ligand. The human SUMO that is fused to the tail of the polyG would be suitable as an adapter/extender of sorts. It brings the polyG farther away from the surface of the VLP, helping it become more accessible for ligation.

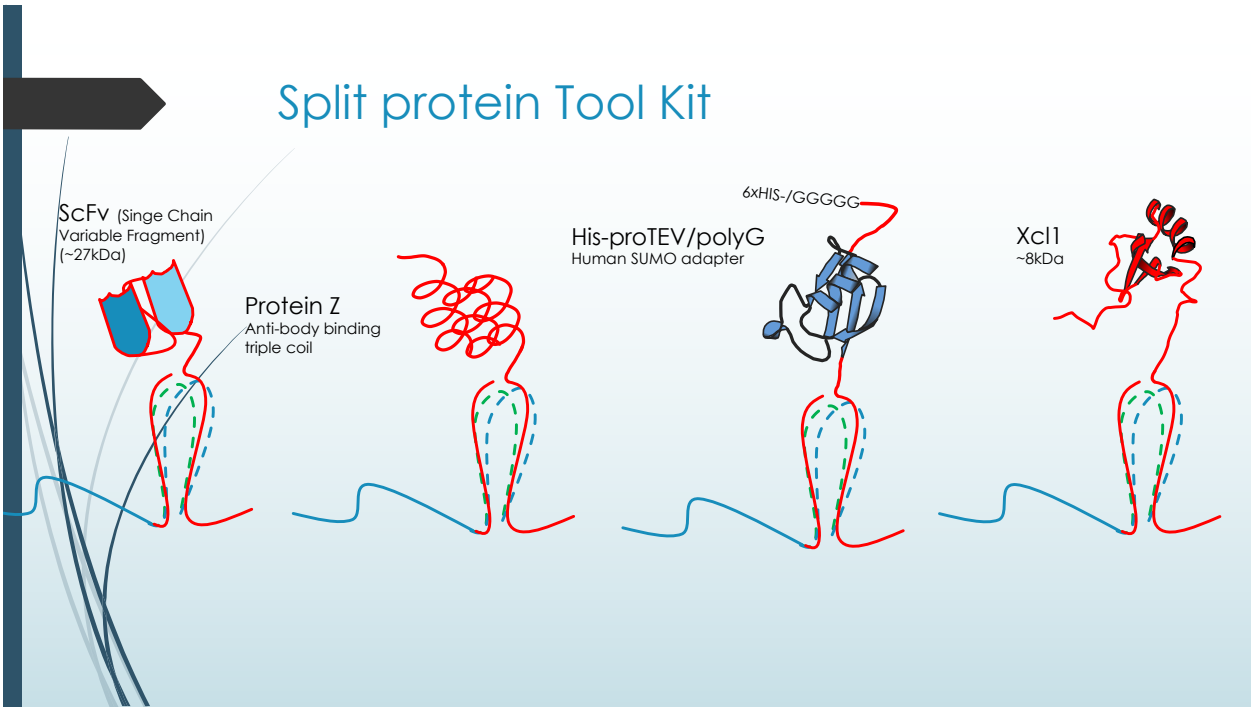

Figure 5-19: Split protein tool kit. These are some different Protein Bs that can be integrated into a split protein VLPs—each with useful functions. Single Chain Variable Fragments can be small enough to be ligands for split proteins, having high affinity for their respective antigen. Or, with use of Protein Z's triple coil, we can bind full size antibodies for the same purpose. His-proTEV/polyG ligands can act as an all-purpose adapter to which any LPXTG containing ligand can by ligated to (with Sortase A). Finally, Xcl1 is an excellent targeting ligand that can target a subset of dendritic cells (DCs) to prime the immune system against a specific antigen (one that could be encoded within the VLP's RNA cargo).

For the lab, introducing TYMV as a potential vaccine vector is particularly attractive for several additional reasons. TYMV virions are more pH stable than CCMV or BMV. They are stable at pH as low as 3.8[14] and as high as 8.5[10], and some particularly robust virions can persist even at higher pH of 10-11.[10,20] The stability at pH 8, which is the optimal pH for sortase A activity, means that TYMV would be an excellent candidate for the Sortase A Labeling Strategy. TYMV virions are even more resilient to higher ionic strength, as well as temperature, compared

to CCMV and BMV, given neutral pH conditions.[15-16] These attributes are further enhanced when dealing with empty TMYV shells, as it is understood that RNA packaging (for TYMV) actually makes the virions less stable.[10,17] Because TYMV protein-protein interactions are so strong, and because CP can assemble into empty VLPs spontaneously, we can likely loosen the RNA packaging requirements regarding protein-to-RNA ratios that were required for successful *in vitro* VLP assembly with CCMV and BMV CP.[18-19] The most important advantage TYMV would provide as a vaccine vector is that (with a native genome length of ~6.3knt) it has the potential to encapsidate RNA lengths much longer than CCMV or BMV (which maxed out at ~4200nt). RNA length capacity dictates how much genetic information we can carry per VLP. Lastly, TYMV is more suited to the split protein strategy than CCMV or BMV. As previously stated, TYMV involves more interactions (a higher number of hydrogen bonds) between the final beta strand of the beta barrel and its two flanking beta strands (Figure 5-20). The number of amino acids on the last beta strand involved in this binding for TYMV is twelve, whereas for CCMV the number of amino acids on the last beta strand involved in this binding is nine, and for BMV it is seven (Figure 5-20). The higher this number, the more hydrogen bonds are involved, and by extension the more likely the "missing" beta strand will "lock in" to position (i.e., see Figures 5-1, 5-3, 5-20). But it should be made clear that unlike BMV or CCMV, TYMV does not self-assemble around RNA, *in vitro*, from pure components. Therefore, it does not meet the criteria we have in place for pursuing the use of *in vitro* assembled VLPs for therapeutic purposes. However, we are currently researching different strategies for *in vitro* RNA packing using TYMV.

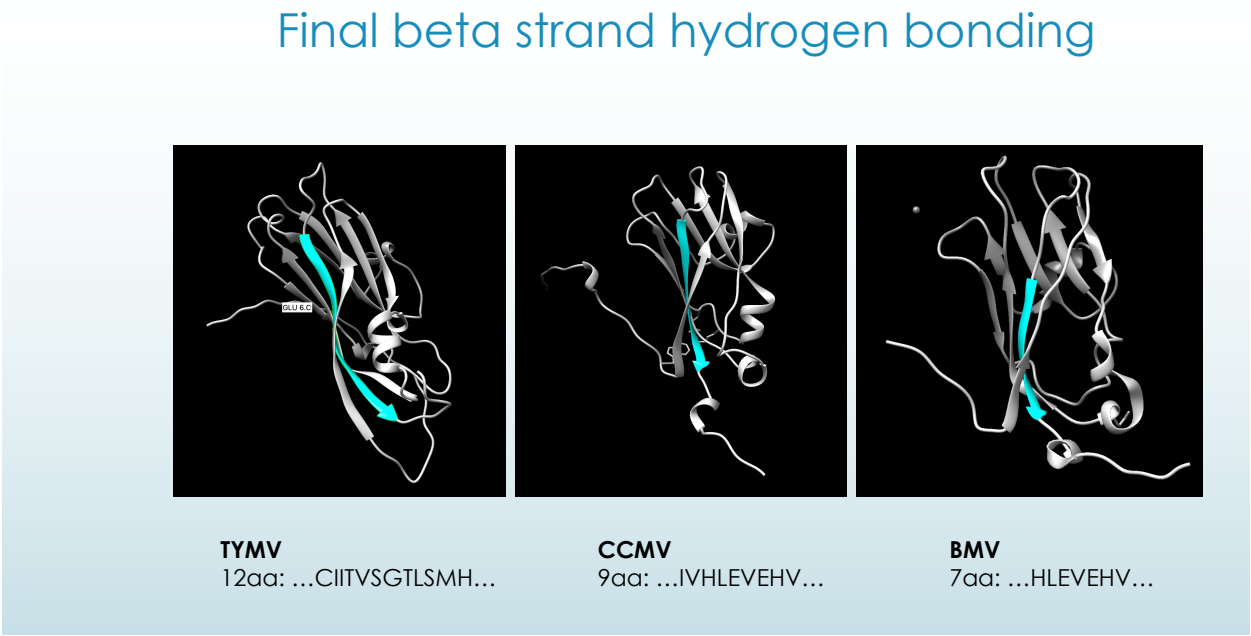

Figure 5-20: Crystal structures highlighting the different lengths of the final beta strand of different viruses. Each amino acid involved in a beta strand conformation within a beta sheet makes two hydrogen bonds with its neighbor (assuming it is flanked by a neighboring strand on each side). The more amino acids on a beta strand the greater the number of hydrogen bonds, and therefore, the more likely that beta strand is to "snap" into its cognate position within the beta sheet. For TYMV there are 12 amino acids involved in holding the final beta strand in place. So, that's 24 points of contact. For CCMV and BMV, there are only 9 and 7 amino acids in their final beta strands, respectively. So, it is TYMV that is mostly likely to have its final beta strand "snap" into place compared to the other two viruses. This indicates that TYMV may be the most suitable to the spit protein design.

References

1. Stepanenko, O. V., Stepanenko, O. V., Kuznetsova, I. M., Verkhusha, V. V., & Turoverov, K. K. (2013). Beta-Barrel Scaffold of Fluorescent Proteins. *International Review of Cell and Molecular Biology*, 221–278. https://doi.org/10.1016/b978-0-12-407699-0.00004-2
2. Kent, K. P., Oltrogge, L. M., & Boxer, S. G. (2009). Synthetic Control of Green Fluorescent Protein. *Journal of the American Chemical Society*, *131*(44), 15988–15989. https://doi.org/10.1021/ja906303f
3. Kim, Y. E., Kim, Y. N., Kim, J. A., Kim, H. M., & Jung, Y. (2015). Green fluorescent protein nanopolygons as monodisperse supramolecular assemblies of functional proteins with defined valency. *Nature Communications*, *6*(1). https://doi.org/10.1038/ncomms8134
4. Barnard, E., & Timson, D. J. (2010). Split-EGFP Screens for the Detection and Localisation of Protein–Protein Interactions in Living Yeast Cells. *Molecular and Cell Biology Methods for Fungi*, 303–317. https://doi.org/10.1007/978-1-60761-611-5_23
5. Ozawa, T., Takeuchi, M., Kaihara, A., Sato, M., & Umezawa, Y. (2001). Protein Splicing-Based Reconstitution of Split Green Fluorescent Protein for Monitoring Protein–Protein Interactions in Bacteria: Improved Sensitivity and Reduced Screening Time. *Analytical Chemistry*, *73*(24), 5866–5874. https://doi.org/10.1021/ac010717k
6. Chun, W., Waldo, G. S., & Johnson, G. V. W. (2010). Split GFP Complementation Assay for Quantitative Measurement of Tau Aggregation In Situ. *Methods in Molecular Biology*, 109–123. https://doi.org/10.1007/978-1-60761-744-0_9
7. Blakeley, B. D., Chapman, A. M., & McNaughton, B. R. (2012). Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein–protein interactions in vivo. *Molecular BioSystems*, *8*(8), 2036. https://doi.org/10.1039/c2mb25130b

8. Powell, J. D., Barbar, E., & Dreher, T. W. (2012). Turnip yellow mosaic virus forms infectious particles without the native beta-annulus structure and flexible coat protein N-terminus. *Virology*, *422*(2), 165–173. https://doi.org/10.1016/j.virol.2011.10.019
9. Canady, M. A., Larson, S. B., Day, J., & McPherson, A. (1996). Crystal structure of turnip yellow mosaic virus. *Nature Structural &Amp; Molecular Biology*, *3*(9), 771–781. https://doi.org/10.1038/nsb0996-771
10. Kuznetsov, Y. G., & McPherson, A. (2006). Atomic force microscopy investigation of Turnip Yellow Mosaic Virus capsid disruption and RNA extrusion. *Virology*, *352*(2), 329–337. https://doi.org/10.1016/j.virol.2006.04.008
11. Ahmad, Z. A., Yeap, S. K., Ali, A. M., Ho, W. Y., Alitheen, N. B. M., & Hamid, M. (2012). scFv Antibody: Principles and Clinical Application. *Clinical and Developmental Immunology*, *2012*, 1–15. https://doi.org/10.1155/2012/980250
12. Braisted, A. C., & Wells, J. A. (1996c). Minimizing a binding domain from protein A. *Proceedings of the National Academy of Sciences*, *93*(12), 5688–5692. https://doi.org/10.1073/pnas.93.12.5688
13. Fox, J. C., Thomas, M. A., Dishman, A. F., Larsen, O., Nakayama, T., Yoshie, O., Rosenkilde, M. M., & Volkman, B. F. (2019b). Structure-function guided modeling of chemokine-GPCR specificity for the chemokine XCL1 and its receptor XCR1. *Science Signaling*, *12*(597). https://doi.org/10.1126/scisignal.aat4128
14. Fukuto, M., Nguyen, Q. L., Vasilyev, O., Mank, N., Washington-Hughes, C. L., Kuzmenko, I., Checco, A., Mao, Y., Wang, Q., & Yang, L. (2013). Crystallization, structural diversity and anisotropy effects in 2D arrays of icosahedral viruses. *Soft Matter*, *9*(40), 9633. https://doi.org/10.1039/c3sm51853a
15. Speir, J. A., Munshi, S., Wang, G., Baker, T. S., & Johnson, J. E. (1995). Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, *3*(1), 63–78. https://doi.org/10.1016/s0969-2126(01)00135-6

16. Cuillel, M., Zulauf, M., & Jacrot, B. (1983). Self-assembly of brome mosaic virus protein into capsids. *Journal of Molecular Biology*, *164*(4), 589–603. https://doi.org/10.1016/0022-2836(83)90052-9
17. Virudachalam, R., Low, P. S., Argos, P., & Markley, J. L. (1985). Turnip yellow mosaic virus and its capsid have thermal stabilities with opposite ph dependence: studies by differential scanning calorimetry and 31P nuclear magnetic resonance spectroscopy. *Virology*, *146*(2), 213–220. https://doi.org/10.1016/0042-6822(85)90005-4
18. Vega-Acosta, J. R., Cadena-Nava, R. D., Gelbart, W. M., Knobler, C. M., & Ruiz-García, J. (2014). Electrophoretic Mobilities of a Viral Capsid, Its Capsid Protein, and Their Relation to Viral Assembly. *The Journal of Physical Chemistry B*, *118*(8), 1984–1989. https://doi.org/10.1021/jp407379t
19. Garmann, R. F., Goldfain, A. M., Tanimoto, C. R., Beren, C. E., Vasquez, F. F., Villarreal, D. A., Knobler, C. M., Gelbart, W. M., & Manoharan, V. N. (2022). Single-particle studies of the effects of RNA–protein interactions on the self-assembly of RNA virus particles. *Proceedings of the National Academy of Sciences*, *119*(39). https://doi.org/10.1073/pnas.2206292119
20. Virudachalam, R., Low, P. S., Argos, P., & Markley, J. L. (1985b). Turnip yellow mosaic virus and its capsid have thermal stabilities with opposite ph dependence: studies by differential scanning calorimetry and 31P nuclear magnetic resonance spectroscopy. *Virology*, *146*(2), 213–220. https://doi.org/10.1016/0042-6822(85)90005-4

6. MATERIALS AND METHODS

Genes and plasmids

We purchased MBP (item #8827), proTEV (item #8827), Sortase A (item #51141) constructs from Addgene.org. We expressed all recombinant proteins using BL21 (DE3) *E. coli* expression system (New England Biolabs). Cultures were grown for 16 hours at room temperature, and harvested with centrifugation at 9000 x g for 12mins. Cell lysis was performed with sonication, and the soluble fraction was isolated via centrifugation at 15000 x g for 45mins at 6°C. The clarified lysate was applied over a pre-equilibrated Nickel IMAC gravity column (HisPur™ Ni-NTA Resin, Thermo Scientific) and the His-tagged recombinant protein was eluted using 50mM Tris-HCl pH 7.5, 350mM NaCl, 250mM Imidazole. We performed a second round of purification using an size exclusion ÄKTA via gel-filtration chromatography (Superdex® 200 prep grade, GE Healthcare). Fractions were eluted in 50mM Tris-HCl pH 7.5, 350mMNaCl and pooled via Amicon 10kDa (Millipore-Sigma).

We purchased codon optimized gBlocks® from Integrated DNA Technologies (IDT) for BMVwt, CCMVwt, TYMVwt, and BMV proTEV/loop1/2/3, (mouse) xcl1, (human) Xcl1, and mTFP1. gBlocks® were cloned into either ampicillin resistant pET-15 MHL (Addgene #26092) or kanamycin resistant pET-30b (Addgene #51141) backbones for bacterial expression systems. Recombinant protein expression and purification is described above.

BMVwt and BMV proTEV/loop1/2/3 gBlocks® were cloned into an Ampicillin resistant and autotrophic HIS4 back bone (Addgene #25453) (Sears et al., 1998) meant for yeast derived protein expression.[1] *P. pastoris* GS115 cells were purchased though Thermofisher (item #C18100). Yeast derived protein expression was performed through previously described protocols (Huang et al., 2019) (Invitrogen *Pichia pastoris* USER GUIDE., 2019).[2]

Sucrose gradients

Sucrose gradients were made using a density range of 10-40% with a base PBS buffer pH 7.5 (for mTFP1 constructs) or VSB buffer (for BMV constructs). Samples were loaded onto a 50.1 SW rotor and spun at 42000 RPM for 3hrs. Yeast derived BMVwt and BMV mutant constructs were pre-stained using GelRED (Millipore-Sigma item #SCT123) before loading onto sucrose gradient(s). Samples were then illuminated with a UV handlamp and removed via glass pasteur pipette. All BMV samples were then dialyzed against VSB to remove sucrose. mTFP1 containing VLPs were dialyzed against 50mM Tris-HCl pH 7.5, 350mM NaCl.

VLP assemblies

BMVwt and CCMVwt *in vitro* assemblies were performed via canonical two-step assembly protocol (Cadena-Nava et al., 2012).[3] TYMV and mTFP1-containing split protein assemblies were described in-text.

References

1. Sears, I.B., O'Connor, J., Rossanese, O.W. and Glick, B.S. (1998), A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast, 14: 783-790. https://doi.org/10.1002/(SICI)1097-0061(19980615)14:8<783::AID-YEA272>3.0.CO;2-Y
2. Huang, Y., Zhang, Y., Li, S., Lin, T., Wu, J., & Lin, Y. (2019). Screening for functional IRESes using α-complementation system of β-galactosidase in Pichia pastoris. *Biotechnology for Biofuels*, *12*(1). https://doi.org/10.1186/s13068-019-1640-3
3. Cadena-Nava, R. D., Comas-Garcia, M., Garmann, R. F., Rao, A. L. N., Knobler, C. M., & Gelbart, W. M. (2012). Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *Journal of Virology*, *86*(6), 3318–3326. https://doi.org/10.1128/jvi.06566-11