UNIVERSITY OF LATVIA

ILVA LIEKNIŅA

Virus like particles of ssRNA bacteriophages - versatile tools in vaccine development

DOCTORAL THESIS

Promotion to the degree of Doctor of Biology

Molecular Biology

Riga, 2021

This work has been carried out at the Latvian Biomedical Research and Study Centre from 2018 to 2021.

This research has been carried out within the European Research and Development Foundation project "Obtaining and characterization of virus-like particles of novel RNA phages" (1.1.1.1/16/A/104).

The form of this thesis is a collection of research papers in biology, subfield – molecular biology.

Supervisor: Professor, Dr. biol. Kaspars Tārs

Reviewers:

1) Dr. biol. Andris Zeltiņš, Latvian Biomedical Research and Study Centre

2) Professor, Dr. biol. Uldis Kalnenieks, University of Latvia

3) Professor, PhD. Nicola J. Stonehouse FRSB FRSA, University of Leeds

The thesis will be defended at the public session of the Doctoral Committee of Biology, University of Latvia, on 21th June 2021 at Latvian Biomedical Research and Study Centre, Rātsupītes street 1-k.1, Riga.

The summary of doctoral thesis is available at the Library of the University of Latvia, Riga, Kalpaka blv. 4.

Chairman of the Doctoral Committee: _____/Prof., Dr. biol. Jānis Kloviņš/

Secretary of the Doctoral Committee: _____/Daina Eze/

1. ABSTRACT

The single-stranded RNA (ssRNA) bacteriophages of *Leviviraidae* family are among the simplest known viruses with smallest genomes and high mutation rates which infect various gram-negative bacteria. For a long time, the number of discovered ssRNA phages was very low, but recent metagenomic studies have uncovered a huge variety of diverse ssRNA phages with genomes remarkably different from those known before. Due to their simplicity and robustness, virus-like particles (VLPs) derived from the coat proteins (CPs) of (ssRNA) bacteriophages have found various applications as tools in molecular biology and related fields. However, only a few different ssRNA phage derived VLP types have been available so far, resulting in certain limitations in development and adaptations of VLP technology. VLPs have many applications, for example they can be used as efficient carriers of various antigens with immunological epitopes or as nanocontainers with tissue-targeted "addresses" and therefore serve as attractive tools in vaccine development or in targeted bioimaging and drug delivery. Although VLPs of many different viruses can be used, VLPs of ssRNA phages have certain advantages due to their simplicity, efficient production levels in bacterial and yeast expression systems, RNA – binding properties, low production cost and easy and fast purification. Until now, the range of available ssRNA phage VLPs was rather narrow, somewhat limiting their potentials applications. Therefore, there was a need to find new members of ssRNA phage VLPs.

More than 150 hypothetical ssRNA phage CP sequences were sourced from metagenomic sequence data. We selected 110 CP-encoding sequences and produced the respective proteins in bacteria which in 80 cases resulted in formation of VLPs. Most CPs were soluble when produced at 37 °C, but in some cases soluble CPs could be formed only at lower temperatures. Almost all the soluble CP assembled into VLPs. According to electron microscopy, morphology of obtained particles was like that of the previously studied phages, although a few deviations such as elongated or smaller particles were noted in certain cases. Additionally, stabilizing inter-subunit disulfide bonds and capsid-bound metal ions were detected in several VLPs. Compared to the few previously characterized types of ssRNA phage VLPs, several dozens of new particles representing ten distinct similarity groups were identified with a notable potential for biotechnological applications. During the structural investigations of novel VLPs we also found several previously unseen features such as novel secondary structure elements and bound cellular dsRNA fragments.

In order to evaluate suitability of obtained VLPs as tools for vaccine development, a model peptide was attached to the N- and C-termini of selected CPs. As a model peptide, we used a triple repeat of 23 N-terminal residues of the ectodomain of the influenza M2 protein, exploited previously in the development of the flu vaccine. During examination of 43 novel phage CPs for their ability to form chimeric VLPs, we found ten new promising candidates for further vaccine design, five of which were tolerant to insertions at both the N- and C-termini. Furthermore, it was demonstrated that most of the chimeric VLPs have proper antigenic features as judged from their reactivity with anti-M2 antibodies.

We believe that the novel VLPs described in this thesis will provide the foundation for future development of new vaccines and other applications based on ssRNA bacteriophage VLPs.

2. KOPSAVILKUMS

Vienpavediena RNS (ssRNS) bakteriofāgi no *Leviviraidae* dzimtas ir vienkāršākie no zināmajiem vīrusiem ar īsu genomu, kurā novērots augsts mutāciju līmenis. Atklāto ssRNS fāgu skaits joprojām ir ļoti mazs, taču nesenie metagenomiskie pētījumi ir atklājuši milzīgu dažādu ssRNS fāgu sekvenču daudzumu, un to genomi ievērojami atšķiras no iepriekš zināmajiem. Vīrusveidīgās daļiņas (VLP), kas iegūtas no ssRNS bakteriofāga apvalka proteīna (CP), ir bijušas nozīmīgas un veiksmīgas pretendentes biotehnoloģijas jomā to vienkāršības un izturības dēļ. Tomēr ir pieejami tikai nedaudzi no ssRNS fāgiem atvasināti VLP veidi, kas uzliek noteiktus ierobežojumus šīs tehnoloģijas attīstītībai. VLP var izmantot kā efektīvus dažādu antigēnu vai "adrešu" nesējus, un tāpēc tie ir pievilcīgi instrumenti vakcīnu izstrādē. Šo daļiņu pielietojums ir plašs arī citās molekulārās bioloģijas jomās: mērķtiecīga bio-attēlveidošana dzīvos organismos, zāļu piegādes līdzekļi, peptīdu displeju sistēmas u.c. Pašlaik ir pieejams plašs dažādu vīrusu izcelsmes VLP klāsts, tomēr no ssRNS fāgu CP iegūtajām VLP ir būtiskas priekšrocības to unikālo īpašību dēļ. ssRNS fāgu VLP ir raksturīga efektīva proteīnu sintēze baktēriju un raugu ekspresijas sistēmās, zemas ražošanas izmaksas, kā arī viegla un ātra attīrīšanas metodoloģija. Pašlaik no ssRNS fāgu CP iegūto VLP klāsts ir ierobežots, tāpēc pētījumi šajā jomā ir aktuāli.

No metagenomu datu bāzēm tika iegūtas vairāk nekā 150 hipotētisko ssRNS fāgu CP sekvences. Turpmākajiem pētījumiem tika atlasītas 110 CP kodējošas sekvences un ekspresētas *E.coli* sistēmā. Gandrīz visu atlasīto gēnu ekspresija rezultējās šķīstošos proteīnos. Tika novērota sakarība, ka daļai CP, kuru sekvences iegūtas no vidēm, kur parasti ir zemāka temperatūra, funkcionālu proteīnu sintēze arī notika pazeminātā temperatūrā - +15 °C. Lielākā daļa šķīstošo CP veidoja VLP. Iegūto daļiņu morfoloģija elektronu mikroskopijā bija līdzīga iepriekš pētīto fāgu morfoloģijai, lai gan dažos gadījumos tika novērotas novirzes, piemēram, iegarenas vai mazākas daļiņas. Tika izveidota VLP kolekcija, kas sastāv no 80 jaunām VLP. Turklāt, vairākās VLP tika atklātas subvienības stabilizējošas disulfīdu saites un metāla jonu stabilizējošas struktūras. Salīdzinot ar dažiem iepriekš izmantotajiem ssRNS fāgu VLP veidiem, tagad ir pieejamas desmit jaunas CP grupas, ar ievērojamu potenciālu biotehnoloģiskos pielietojumos.

Sekojošajā šo CP un jauno VLP pētījumā tika pārbaudīta 43 dažādu vīrussveidīgo daļiņu toleranci ģenētiskām modifikācijām. Apvalka proteīni tika modificēti gan N-, gan C-terminālajās daļās, kur tika ievietots modeļpeptīds M2ex3, kas ir atvasināts no trīs apakštipu A tipa gripas jonu kanāla ārējā domēna. Pētījuma rezultātā tika atrasti desmit jauni daudzsološi kandidāti turpmākai vakcīnu izstrādei, no kuriem pieci bija toleranti pret ģenētisko modifikāciju gan N-, gan C-galos. Turklāt tika pierādīts, ka lielākajai daļai himēro VLP ir atbilstošas antigēnās īpašības, kuras izvērtētas pēc saistības ar anti-M2 antivielām.

Šajā darbā aprakstītās jaunās VLP paplašinās biotehnoloģisko līdzekļu klāstu vakcīnu platformu attīstībai, piegādes sistēmām un strukturāliem pētījumiem.

3. INDEX

1. Abstract 3
2. Kopsavilkums 4
3. Index 5
4. Introduction 7
5. Literature review 8
   5.1. ssRNA bacteriophages and their host diversity 8
      5.1.1. Taxonomy 8
      5.1.2. ssRNA phage hosts 11
      5.1.3. Genome organization of ssRNA bacteriophages 12
      5.1.4. Coat proteins and capsid 14
      5.1.5. Maturation protein 16
      5.1.6. Replicase 16
      5.1.7. Lysis protein 16
      5.1.8. RNA 16
      5.1.9. Infection process of ssRNA bacteriophages 18
      5.1.10. Studies of diversity of ssRNA bacteriophages 18
   5.2. Virus like particles 19
      5.2.1. Immunogenic properties of VLPs 19
      5.2.2. Applications of the ssRNA phage VLPs 21
6. Methods 27
      6.1.1. Construction of expression plasmids 27
      6.1.2. CP production 27
      6.1.3. Production and solubility assessment of fusion proteins 27
      6.1.4. VLP purification 28
      6.1.5. Purification of chimeric VLPs 28
      6.1.6. Electron microscopy 28
      6.1.7. Dynamic light scattering 28
      6.1.8. Detection of disulfides in VLPs 28
      6.1.9. Determination of VLP thermal stability 29
      6.1.10. ELISA 29
7. Results 30
   7.1. “Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data” 30
      7.1.1. Production of the CP genes of novel ssRNA phages 30
      7.1.2. Purification of proteins and characterization of VLP morphology 30

7.1.3. Stabilizing disulfide bonds in the novel VLPs 30
7.1.4. VLP thermal stability 31
7.2. “Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes” 46
7.2.1. Capsid shape and size 46
7.2.2. Intersubunit interactions 46
7.2.3. Potential carriers 46
7.3. “Novel ssRNA phage VLP platform for displaying foreign epitopes by genetic fusion” 58
7.3.1. CP selection for genetic modification 58
7.3.2. Ability of chimeric proteins to form VLPs 58
7.3.3. Stability of chimeric VLPs 58
7.3.4. Antigenicity of the chimeric particles 58
8. Discussion 68
9. Conclusions 73
10. Main thesis of defense 74
11. Publications 75
12. Approbation of the research 76
13. Acknowledgements 77
14. References 78

4. INTRODUCTION

The history of vaccines began three centuries ago with the first smallpox vaccine developed by Edward Jenner (Plotkin 2014). Vaccination is still recognized as the most effective way to fight and control infectious diseases caused by viruses. Most currently available vaccines are based on inactivated or attenuated pathogens. Despite the efficacy of these vaccines, there is a risk that inactivated or attenuated pathogens can regain virulence and cause true disease (Kushnir et al. 2012). Along with the development of molecular biology techniques, vaccines based on virus-like particles (VLPs) have emerged. VLPs are nanostructures that can self-assemble from individual coat proteins (CPs) of a virus. VLPs are structurally similar to the respective viruses but lack the genetic material needed for replication (Jennings and Bachmann 2008). VLPs can be easily produced in a variety of expression systems, including insect and mammalian cells, yeasts and bacteria. Additionally, VLPs are characterized by their ability to form high symmetry structures which can be rationally modified, providing the particles with desired functionality.

In vaccine development, the immune response against the carrier must be considered, and the narrow range of available CPs limits the number of potential vaccines that could be produced. Therefore, the discovery and characterization of novel ssRNA phage CPs and VLPs are of considerable interest for continued development in the area.

Recent metagenomic sequencing analyses have revealed the sequences of more than 150 new hypothetical ssRNA bacteriophages (Krishnamurthy et al. 2016) whose encoded envelope proteins can enrich the range of VLP-forming proteins.

The aim of study was to characterize novel VLPs derived from uncultured ssRNA bacteriophage CPs.

The tasks of the study:

1. Production of 120 new ssRNA phage CPs in *E.coli* and examination of the properties of obtained proteins - solubility, particle formation, and stability.
2. VLP purification for general characterization - including high resolution structural studies
3. Assessment of the tolerance of the most promising phage CPs to genetic fusions at the N- and C-termini using the model peptide M2e.

5. LITERATURE REVIEW

5.1. ssRNA bacteriophages and their host diversity

Bacteriophages are tiny parasites that infect bacteria and archaea and form a major part of the vast viral world. Phages are found everywhere bacteria can thrive. They are found in huge numbers and are therefore considered to be the largest global reservoir of genetic diversity. Although most known bacteriophages have dsDNA genomes, there are other types of genomes found as well. Among them, ssRNA bacteriophages have several unique properties: 1) the genetic material is in the form of single-stranded positive sense RNA; 2) along with dsRNA phages belonging to family *Cystoviridae* they are the few representatives of *Riboviria* realm, infecting bacteria; 3) they are pili specific (Zinder 1980).

While the second half of the twentieth century is often called the "golden age" of virus discovery, the history of RNA bacteriophages dates back to the 1960s. Especially, Norton D. Zinder is considered the "father" of RNA phages, detection of RNA phages was accidental when studying *E. coli* bacteria containing F-pili (Zinder 1980). Due to their simplicity, RNA phages have played an important role in the development of molecular biology. For example, the MS2 phage is the first organism to have the entire genome sequenced (Fiers et al. 1976). Phages R17 and f2 were the first characterized virions with an icosahedral structure (Vasquez et al. 1966).

5.1.1. Taxonomy

In 2019, the latest ICTV (International Committee on Taxonomy of Viruses) virus taxonomy system, based on the composition of the genome packaged in a virion was published. This classification system continues the tradition introduced by D. Baltimore. RNA viruses are included in the *Riboviria* realm, which currently includes 2 kingdoms, 3 phyla, 2 subphyla, 21 classes, 28 orders, 8 suborders, 102 families, 43 subfamilies, 458 genera, 68 subgenera and 2686 species (Fig.1). ssRNA bacteriophages are assigned to the *Leviviridae* (from Latin: *levis* - light; *virus* - poison, toxic) family, which is further split into two coliphage genera - *Levivirus* and *Allolevivirus*. The *Levivirus* genus includes MS2, BZ13 and GA-like bacteriophages, while *Allolevivirus* includes Qβ, FI and SP-like bacteriophages (https://talk.ictvonline.org/taxonomy/).

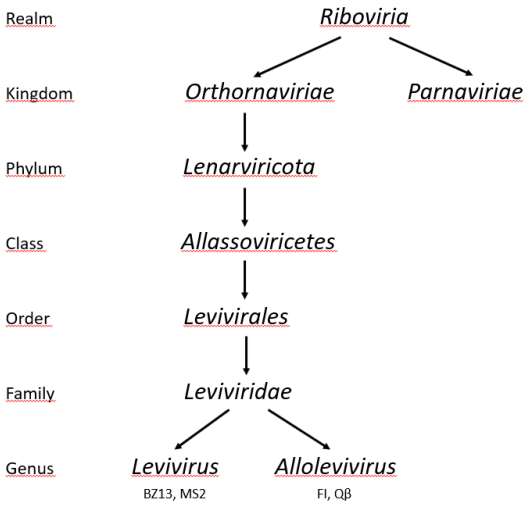

Fig.1. *Leviviridae* branch of *Riboviria* realm.

Genome organization, high similarity between RNA polymerases, identical use of host factors in replication, and similar mechanisms in translation regulation were the basis for combining the known ssRNA bacteriophages into the *Leviviridae* family (see (Bollback and Huelsenbeck 2001) for a review).

By the end of 2020, the GenBank contained the complete sequences of only 12 ssRNA bacteriophages, including MS2, GA, Qβ, AP205 and FI, and incomplete sequences of many other bacteriophages (Krishnamurthy et al. 2016).

Immunological characterization or serogrouping is considered to be important for virus classification. In the 1960s, when trying to find immunological differences for several RNA phages - f2, f4, fr, etc., it was found that they were all serologically related (Scott 1965). At the same time, another group of scientists found that the phage Qβ found in Japan was not serologically cross-reactive with MS2 (Overby et al. 1966). Based on these findings, three serological groups were introduced - I, II and III (Sakurai T. 1967). There is weak cross-reactivity between groups I and II, while group III phages do not display cross-reactivity with both groups mentioned above. Shortly afterwards, group IV and putative group V were introduced with the discovery of new phages. *Levivirus* members currently include serogroups I and II, and *Allolevivirus* members include serogroups III and IV, respectively. The following reference strains have been identified for serogroups: I - MS2, II - GA, III - Qβ and IV - SP phages (Miyake et al. 1971).

In correlation with serological groups ssRNA bacteriophages are also grouped by their affinity to nitrocellulose membranes (filtration method), replicase template specificity, hybrid particle formation, physicochemical parameters, genogrouping and metaviromics (see Chapter 2 of (Pumpens 2020) for a review).

A special role in the development of genogrouping was played by the complete genome sequencing of phage MS2 (Fiers et al. 1976; Inokuchi et al. 1988; Adhin et al. 1989; Groeneveld et al. 1996). Subsequently, complete genomic sequences were obtained for coliphages Qβ, GA, JP34, KU1 and SP, as well as for non-coliphage PP7 (Olsthoorn et al. 1995) and multi-host phage PRR1 (Ruokoranta et al. 2006). Significantly more differences in genome organization were observed in Acinetobacter phage AP205 (Klovins et al. 2002), and Caulobacter phage Cb5 (Kazaks et al. 2011).

The characterization of known viruses so far has been possible because their host bacteria were known and could be cultured. On the other hand, the characterization of new viruses has been hindered by the fact that many host microorganisms do not grow under laboratory conditions, therefore it is very difficult to obtain and cultivate the corresponding live viruses. Often, among distantly related phages there is limited serological cross-reactivity of proteins or absent nucleic acid hybridization reactions (Delwart 2007). In general, the world of bacteriophages is recognized as a “dark matter” in the world of biology. It is estimated that the global very dynamical phage population contains about 10^{31} particles (Clokie et al. 2011; Pumpens 2020) - for comparison only about 10^{19} grains of sand are believed to exist on Earth. The long existence and variability of phages has created a population with fantastic genetic diversity (Hatfull 2015). The amount of phage particles can reach up to 10^{10} per liter of seawater and 10^{9} per gram of soil (Delwart 2007). The diversity of different bacteriophages is reflected by their significant impact on global biological, geological and chemical processes and depend on the number and diversity of host bacteria (Hatfull 2015; Breitbart et al. 2002).

Several molecular methods have been developed to genetically characterize emerging viruses without prior replication *in vitro* or the use of virus-specific reagents. A major

breakthrough in phage research was the development of a metagenomic sequencing approach. Breitbart with colleges (Breitbart et al. 2002) in their study found that in seawater samples, bacteriophage sequences can make up 75-90% of all viral sequences. At the beginning of this century, a group of scientists concluded that most genomes of marine microflora phages are found in the form of DNA (Weinbauer 2004; Culley et al. 2006). Even in more recent massive sequencing experiments, very few RNA phages were found (Zhang and Fang 2006; Carding et al. 2017). Nevertheless, the search for new RNA bacteriophages continued. In 2015, a study was published on the first genomes of marine RNA phages - EC and MB, which were found using a new sequencing method (Greninger and DeRisi 2015).

Krishnamurthy and his colleagues performed the first massive sequence analysis experiment. In this study, 122 new RNA bacteriophage phylotype sequences were identified that differ significantly from the RNA phages described and studied previously (Fig.2) (Krishnamurthy et al. 2016). Sequences were obtained from samples collected in various ecological niches around the world, demonstrating that ssRNA phages are far more widespread than previously recognized. Sequence analysis of new RNA bacteriophages indicated new previously unseen architectures of genomes. For example, the 5 kb genomes of the two newly discovered ssRNA phages (AVE000, AVE001) were much longer than all previously sequenced *Leviviridae* viral genomes, ranging in size from 3.73 to 4.27 kb. The transcriptome of one RNA phage Streptomyces bacteriophage ϕ0 was obtained from a pure culture of *Streptomices avermitilis*, which reveals a possible trophism of ssRNA phages to Gram-positive bacteria (Krishnamurthy et al. 2016). Very recently, another metagenome study identified more than 1000 near-complete genomes of ssRNA phages (Callanan et al. 2020).

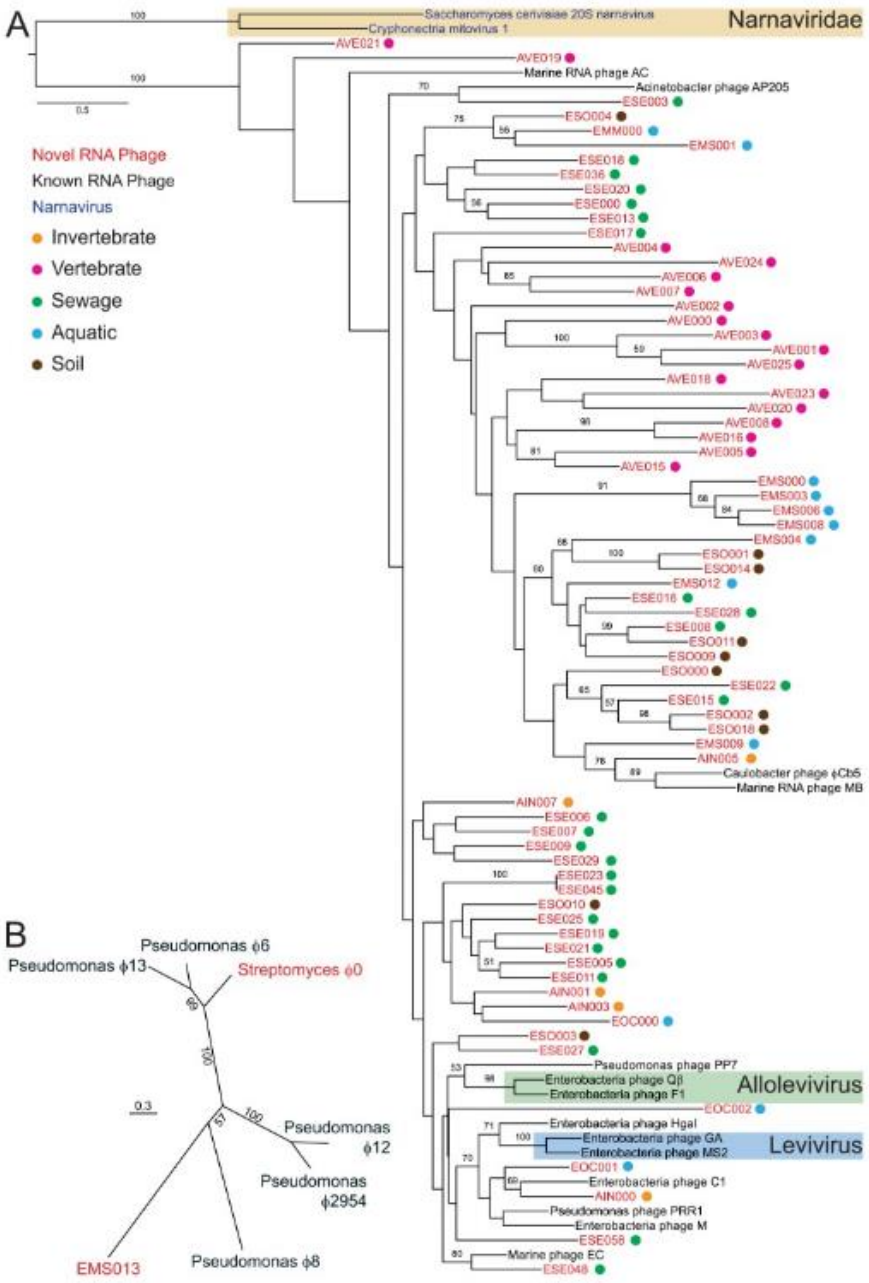

Fig.2. Phylogenetic analyses of novel RNA bacteriophage genomes discovered in metagenomic sequencing datasets (Krishnamurthy et al. 2016).

5.1.2. ssRNA phage hosts

Bacteria are infected by ssRNA phages via attachment to their fimbriae or pili. *Escherichia coli* infecting ssRNA phages use sex fimbriae or F-pili, to enter the host cell (Fig.3). The physiological function of F-pili is conjugation - binding of two bacteria to promote the exchange of genetic information. This process is achieved by shortening the F-pili after binding. F plasmid enters the bacterium through a conjugation pore formed by proteins encoded in F plasmid. It is thought that ssRNA phages have adopted this plasmid transfer strategy to enter cells (Paranchych et al. 1970). After attachment, and subsequent shortening of the pili, the phage maturation protein is cleaved, and the genomic RNA enters the bacterium (Danziger and Paranchych 1970). ssRNA bacteriophages, which infects

Pseudomonas and *Caulobacter*, do not use sex pili but enter the host using polar or somatic pili (Bradley and Robertson 1968; Schmidt 1966).

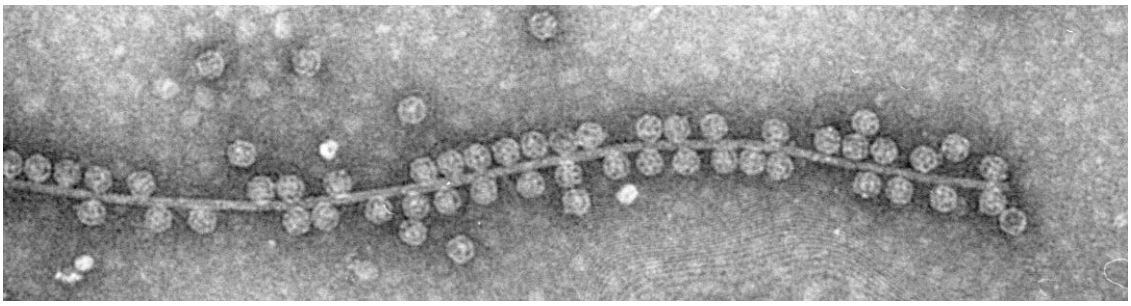

Fig.3. Electron micrograph of F-pili and ssRNA bacteriophage MS2 complexes (Author: Dr.biol. Velta Ose).

Studies with radiolabeled phages allowed to quantitatively analyze the attachment of phage particles to the pili. It was found that up to 1000 virions can be attached to F-pili, and this reaction is reversible. Mono- and bivalent cations can promote the adsorption reaction. The effectiveness of the attachment is affected also by the structure of the pili (Paranchych and Frost 1988).

The genus of bacteria and its pili types limit the range of possible hosts for each group of ssRNA phages, making these viruses highly specialized and host-specific (Bollback and Huelsenbeck 2001). However, certain species of ssRNA phages, such as the PRR1 phage, can infect bacteria from different genera, such as *Escherichia coli*, *Pseudomonas aeruginosa*, *Salmonella typhimurium*, and *Vibrio cholerae*. It uses a special pili whose genes are located in plasmids encoding antibiotic resistance, therefore phage is rather plasmid than host dependent. Other enterobacteria, such as *Shigella*, *Proteus* and *Salmonella*, may also be at risk for coliphage infections. This is possible if these bacteria obtain F-factor-containing plasmids from *E. coli*, the expression of which results in the formation of F-pili (Olsthoorn R. 2011).

ssRNA bacteriophages were previously thought to infect only gram-negative bacteria. These belong to the previously mentioned genera: *Escherichia*, *Pseudomonas*, *Caulobacter*, *Salmonella*, *Vibrio* (Olsthoorn R. 2011), bacteria of the genus *Acinetobacter* can be infected with phage AP205 (Klovins et al. 2002). As mentioned before, after metagenome sequencing analysis, the ssRNA bacteriophage genome was detected also in pure culture of Gram-positive bacteria (Krishnamurthy et al. 2016). This significantly expands the diversity of potential RNA phage hosts and allows for the possibility that ssRNA bacteriophage species specializing in infection of gram-positive bacteria also exist. ssRNA bacteriophages are widespread worldwide and are found in bacterial isolates from sewage and mammalian excrements (Bollback and Huelsenbeck 2001, Callanan et al. 2020).

5.1.3. Genome organization of ssRNA bacteriophages

The genome of ssRNA phages comprises a 3.3-4.3 kb long (+) strand ssRNA molecule. In general, in RNA bacteriophages of the genus *Levivirus* it is by about 600-700 nucleotides shorter than in members of the genus *Allelovivirus* (Callanan et al. 2018; Bollback and Huelsenbeck 2001). For currently known ssRNA bacteriophages, the RNA molecule encodes 4 genes (Fig.4): the maturation protein (known also as A or A2) gene, the CP gene, and the replicase (RP) gene. Maturation protein is responsible for attachment of virions to bacterial pili, while RP is the catalytic subunit of RNA-dependent RNA polymerase (RdRp). The fourth gene differs among ssRNA bacteriophage genera. The fourth gene from the

Levivirus genus encodes the lysis protein, responsible for destruction of cellular membrane. Genomes of *Allolevivirus* representatives do not possess a separate gene for lysis protein, but they encode the readthrough protein A1, which is a minor, C-terminally extended version of CP with largely unknown function, although it has been shown to be necessary for the infection process (Bollback and Huelsenbeck 2001). The readthrough protein sequence overlaps with the coding region of the CP and both have a common initiation codon. Extended minor structural protein A1 is formed by translation of the "weak" UGA stop codon. Tryptophan is inserted in place of the stop codon and translation continues to form a CP with an extended C-terminus. This is relatively rare event happening in only about 5% of cases. This type of translation of overlapping genes allows these proteins to be produced in the specified proportions (Bollback and Huelsenbeck 2001) and in the required amount - *Allolevivirus* bacteriophages contain about 12 molecules of extended CP (Weiner and Weber 1971; Weiner and Weber 1973). This is essential for the formation of viable Qβ particles (Skamel et al. 2014).

In the genomes of ssRNA bacteriophages, the genes are arranged in a certain order: first is a maturation protein gene, followed by a CP gene, and finally a RNA polymerase coding sequence. The location of the fourth gene in the genome depends on the species of RNA bacteriophage (Blaisdell et al. 1996). Generally, the lysis protein gene overlaps partially (Olsthoorn et al. 1995) or completely (Rumnieks and Tars 2012) with the CP and/or RP genes, albeit in a different reading frame, except in the phage AP205, in which lysis protein is encoded by a very short ORF at the very 5′ end of the genome (Klovins et al. 2002).

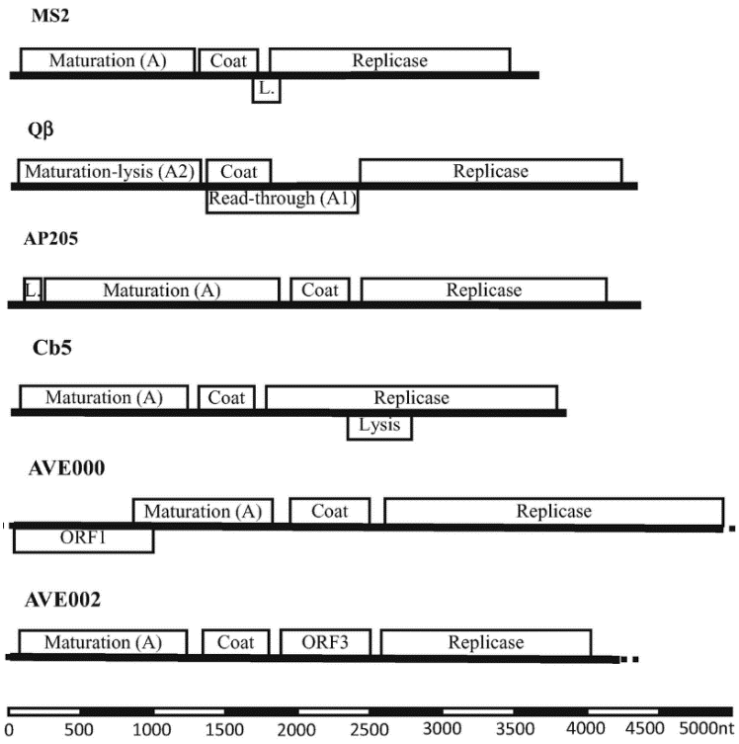

Fig.4. Genomes of ssRNA phages MS2, Qβ, AP205, Cb5, AVE000 and AVE002. Genes are shown as boxes. L. stands for lysis protein. In the incomplete sequences of phages AVE001 and AVE002, ORF1 and ORF3 of genes encoding two putative proteins are shown (Tars 2020).

In *Levivirus* genome the lysis protein gene overlaps with CP and polymerase coding sequences. The synthesis of these proteins depends on the secondary structure of RNA,

which affects ribosome binding during translation (Bollback and Huelsenbeck 2001). At the final stage of the CP translation, the ribosomes normally detach from the RNA upon reaching the termination codon. However, in minority of cases after the synthesis of the CP, the ribosome slides backwards and initiates the translation of lysis protein. Lysis protein is synthesized in small amounts at the end of infectious cycle, which ensures gradual accumulation of the protein and provides sufficient time for new virions to be made prior cell lysis (Callanan et al. 2018).

The genome of phages of the *Allolevivirus* genus does not code lysis protein, its function is performed by the bi-functional maturation protein (Olsthoorn 2011), albeit by a completely different mechanism.

Some of the newly discovered ssRNA phage genomes appear to have additional ORFs with yet unknown function. For example, AVE001 genome contains an ORF before maturation protein gene, while AVE002 contains ORF in between coat and replicase genes (Fig 4, (Krishnamurthy et al. 2016)). In both cases the putative ORFs appear to be much too long to account for lysis protein.

After recent discovery of numerous new ssRNA phage sequences it has become obvious that the classical *levivirus-allolevivirus* classification is outdated, since most of the new sequences do not seem to fit in any of these. However, by studying the organization of the ssRNA bacteriophage genomes, encoded proteins and capsid structures it may be possible to construct phage evolutionary pathways. *Leviviridae* members are not evolutionarily closely related to any other bacteriophages except for a very distant relationship to the only known dsRNA phages belonging to family *Cystoviridae* since it is believed that all RNA viruses share common ancestry (Hillman and Cai 2013). The most common gene or protein to look for relationship is the relatively conservative RNA-dependent RNA polymerase (RdRp). Some eukaryotic viruses belonging to *Narnaviridae*- and *Ourmia*-like families carry somewhat related RdRp genes to those of *Levivirividae* phages. Presumably in these cases RdRp genes have been "borrowed" from each other via horizontal gene transfer. Replicase is the only protein from the *Leviviridae* family that clearly presents relatives within other viruses, while other *Leviviridae* encoded proteins do not seem to have any relatives neither among other viruses nor other proteins in general (Tars 2020).

5.1.4. Coat proteins and capsid

Capsids of ssRNA phages are composed of 178 CP molecules and one maturation protein molecule (Dent et al. 2013). In the CP structure, the N-terminal hairpin structure is followed by a 5-stranded β-sheet and two c-terminal helices (Fig.6. A). Due to the T = 3 symmetry, the CP exists in three slightly different conformations - A, B and C. Two CP monomers form a very stable dimer. In dimeric form, two β-sheets are joined together to form a continuous 10-stranded β-sheet (Fig.6. B). Thus, the capsid is composed of 89 dimers of the CP. An asymmetric AB dimer is formed by two monomers in conformations A and B, and the symmetric CC dimer is formed by two monomers in conformation C (Valegard et al. 1990; Golmohammadi et al. 1996). In ssRNA phage particles, one CC dimer is replaced by a maturation protein (Dent et al. 2013; Koning et al. 2016; Dai et al. 2017).

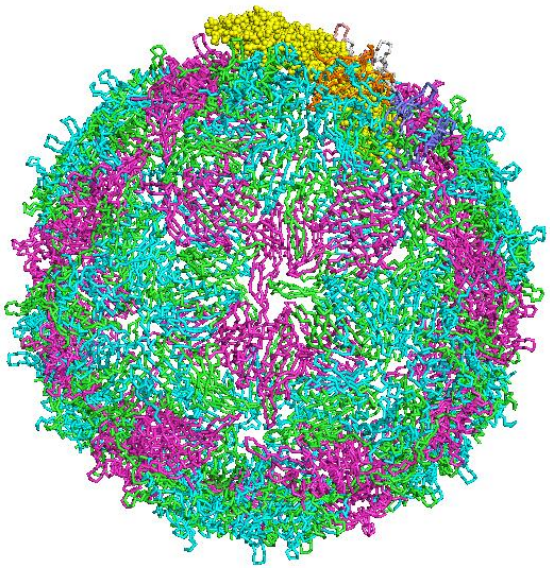

Fig.5. Particle of ssRNA bacteriophage MS2 of *Leviviridae* family. A, B and C subunits are shown in different colours, A protein is shown in yellow.

In CP monomer several flexible loops are connecting β-strands. The loop between the β-strands denoted F and G, known as FG-loop, makes important contacts required for virus assembly. In capsids of MS2-related phages the FG-loop is found in two apparently different conformations dependent on the contacts between subunits around 5-fold or 3-fold axes (Stonehouse et al. 1996). In some other phages the conformational differences among FG loops are less obvious (Tars et al. 2000; Plevka et al. 2009).

In a recent cryo-EM study it was found that in bacteriophage Qβ a single copy of an isolated CP dimer is located inside the particle, near the A1 protein and bound to the genomic RNA. Therefore, Qβ virion actually contains 180 CP subunits, but one CP dimer is not a part of the protein shell (Cui et al. 2017).

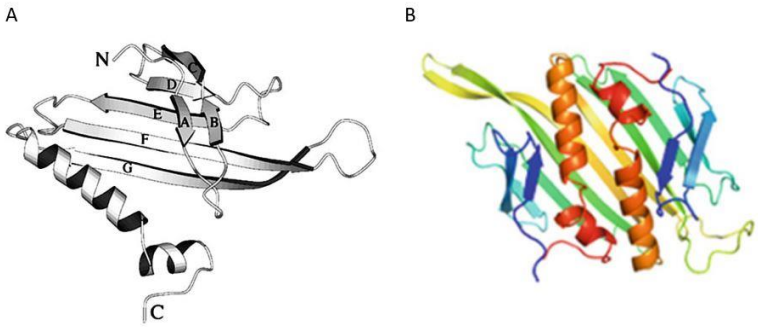

Fig.6. A - Schematic drawing of a B subunit of phage Qβ. The strands are denoted A–G (Golmohammadi et al. 1996). B – structure of CP dimers in phage MS2 (Tars 2020).

Eight structures of ssRNA phages – MS2 (Valegard et al. 1990), fr (Liljas et al. 1994), Qβ (Golmohammadi et al. 1996), GA (Tars et al. 1997), PP7 (Tars et al. 2000), PRR1 (Persson et al. 2008), Cb5 (Plevka et al. 2009) and AP205 (Shishovs et al. 2016) - have been determined by now. The obtained results revealed that the CP fold is very conserved despite high sequence variability. The only significant variation in CP structure so far has been observed in the *Acinetobacter* phage AP205, where a single β-strand has "travelled" from N- to C terminal part. As a consequence, in AP205 CP both termini are the most surface-exposed parts of the assembled particle, which explains its increased tolerance to long N- and C-terminal fusions (Shishovs et al. 2016).

The structural stability of the capsid can be ensured in several ways. Non-covalent protein-protein interactions are the most common among ssRNA phages. The capsid stability of some phages is enhanced by covalent disulfide bonds that bind dimers around the icosahedral symmetry axes in five or six bundles. Capsids, containing disulfide bonds are very stable. Like in several plant viruses, capsids of some ssRNA phage are stabilized by metal ions, for example in Cb5 phage three protein dimers are held together by Ca^{2+} ions around quasi-3-fold axis. Capsid stabilization can also be enhanced by RNA packed in particles as seen in the already mentioned Cb5 phage (Plevka et al. 2009).

5.1.5. Maturation protein

Maturation protein is essential for the infectivity as it is involved in attachment to bacterial pili (Verbraeken and Fiers 1972). Even though in virion it replaces a single CP dimer, the structure of A protein is not in any way similar to CP. Structure of A protein is composed of two parts - helical, which interacts with RNA and β part facing the capsid exterior and interacting with the pilus receptor (Rumnieks and Tars 2017). Presence of A protein also impacts the conformation and placement of neighboring CP dimers contributing to slight asymmetry of icosahedral particle (Gorzelnik et al. 2016). In *Allolevivirus* phages like Qβ maturation protein A1 also triggers the cell lysis. It targets host MurA enzyme involved in bacterial peptidoglycan biosynthesis from the pathway of cell wall construction (Bernhardt et al. 2001). There are notable structural differences in structure of A proteins among *Allolevivirus* and *Levivirus* phages. While the helical parts are somewhat conserved among MS2 and Qβ, the β parts are very different. The observed differences in structure reflects the functional needs - while β part MS2 A protein must interact only with F pili, the β part of the A1 protein of Qβ also interacts with *E.coli* MurA protein.

5.1.6. Replicase

Genome replication and transcription of ssRNA bacteriophages are performed by an RNA-dependent RNA polymerase (RdRp) called replicase. In ssRNA phages a fully functional RdRp is composed of four subunits. The catalytic subunit is encoded by phage genome, while translation elongation factors Ef-TU and Ef-TS and ribosomal protein S1 comprise the rest of subunits, which are recruited from the host cell.

5.1.7. Lysis protein

ssRNA phages utilize at least three very different strategies to accomplish the cell lysis. As discussed above, *Alloleviviruses* make use of bi-functional A1 protein. In MS2, lysis protein forms pores in membrane, which are permeable to ions (Goessens et al. 1988), leading to disruption of the electrostatic potential across the cellular membrane and subsequent activation of autolysins. Lysis protein of phage M also is inserted in membrane, but it blocks flippase activity of MurJ protein, necessary for downstream synthesis of peptidoglycane (Chamakura and Young 2020).

5.1.8. RNA

The positive-sense genome of ssRNA phages can act as messenger RNA and can be directly translated into viral proteins by the host cell's ribosomes. Replicase is used during replication of the genome to synthesize a negative-sense antigenome which is further used as a template to create a new positive-sense viral genome. More than 70% of the genome is

involved in short- and long-distance base-pair interactions forming stem-loops, pseudoknots and other regions that essentially consist of dsRNA (Fig.7) (Skripkin et al. 1990). Therefore, although each particle contains a single strand of RNA, most of the genome is actually double-stranded. The genome has a well-defined 3D structure that is more or less identical in all virions. The genome structure was first visualized by cryo-EM in phage MS2 (Koning et al. 2016; Dai et al. 2017) and later RNA structures were reported also for phage Qβ (Gorzelnik et al. 2016; Cui et al. 2017).

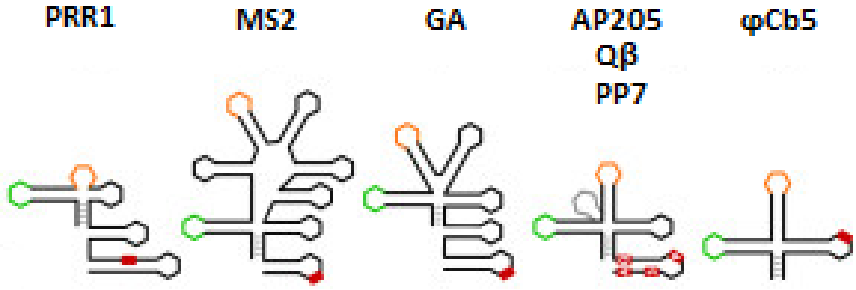

Fig.7. Schematic representations of 3′ untranslated regions from some phages based either on published data or RNA secondary structure predictions (Rumnieks and Tars 2012).

In MS2, Qβ and other related phages a genomic RNA fragment forms about 20 nucleotide long stem-loop structure known as translation repressor (TR) which is able to form complex with CP dimers. TR fragment locates around the replicase start codon (Valegard et al. 1994). The main role of this interaction seems to be the repression of the translation of the replicase gene in late stages of infection, when the presence of the replicase is no longer needed (Gralla et al. 1974). This interaction also provides a nucleation site for virion assembly and promotes the packaging of specific RNA inside the capsid. Similar interactions with the same biological purpose exist in some other ssRNA phages - PP7 (Chao et al. 2008), PRR1 (Persson et al. 2008) and Qβ (Fig.8) (Rumnieks and Tars 2014). Comparison of structures of all known complexes revealed similar interaction pattern for MS2, PRR1 and Qβ, but quite different for PP7. It is clear that the CP-TR interaction plays a rather minor role in the life cycle of phage, as TR-deficient mutants are viable (Peabody 1997). Some other phages - AP205 and Cb5, do not seem to have this specific interaction at all (see (Tars 2020) for the review).

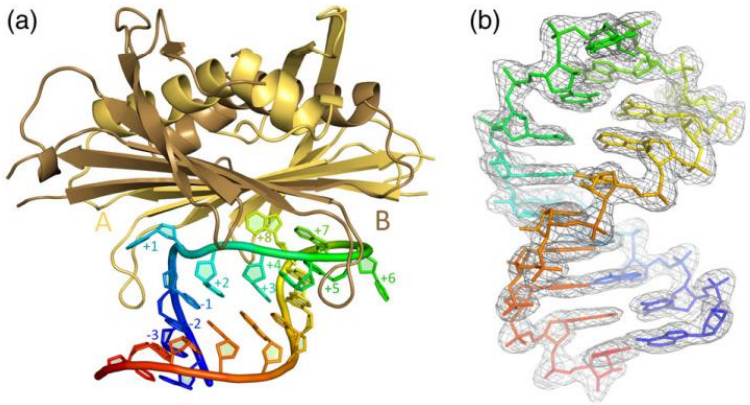

Fig.8. Three-dimensional structure of the Qβ CP–operator complex. (a) Overall structure of the complex. The CP dimer is represented in light orange (monomer A) and light brown (monomer B), and the RNA is rainbow-colored blue (5′ end) to red (3′ end). (b) A close-up view of the RNA hairpin. (Rumnieks and Tars 2014)

Cryo-EM virion structure of the bacteriophage MS2 revealed many other specific RNA-CP interactions. The fragments of the viral RNA in contact with CP subunits are distributed across the whole genome, consistent with a large number of discrete and similar contacts within each particle (Fig.9). Many of these sites match previous predictions of the locations of multiple RNA stem-loop sequences with cognate CP affinity termed packaging signals. Usually CP-RNA interactions occur involving various stem-loop structures from the genome (Rolfsson et al. 2016).

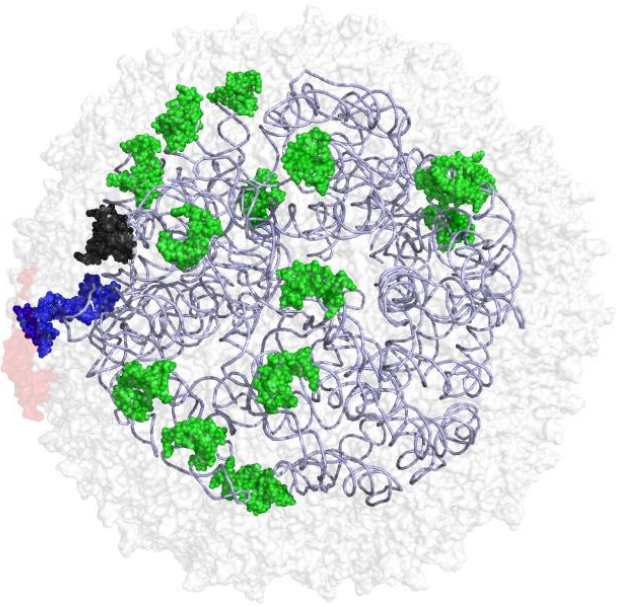

Fig.9. Asymmetric cryo-EM reconstruction of phage MS2. CP and AP are shown as semitransparent light gray and red surface models, respectively. High-resolution CP-binding genome stem loops are shown as green sphere models. TR is shown in black and AP interacting stem-loop in blue. A lower-resolution genome model is shown as a coil (Tars 2020).

5.1.9. Infection process of ssRNA bacteriophages

ssRNA bacteriophage maturation protein recognizes and binds to host surface receptors. This is followed by the injection of genetic material, complexed with the maturation protein, into the cytoplasm of the bacterium, leaving the empty capsid outside the cell (Zinder 1965). Once the genomic RNA enters the cytoplasm of the host cell, it can act directly as mRNA for protein synthesis. The amounts of proteins to be translated are tightly regulated by a variety of mechanisms, including the accessibility of ribosome binding sites, the translational coupling of genes, the formation of RNA secondary structure elements, protein-RNA interactions and read-through of translation stop codons (see (Tars 2020) for a review).

5.1.10. Studies of diversity of ssRNA bacteriophages

Along with the development of next-generation sequencing methods and bioinformatics techniques, it has become possible to perform efficient and fast analyses of metagenomes that are independent of microorganism cultivation (Delwart 2007). A metagenome is a collection of genetic material in a complex environmental sample. The sequences obtained from the metagenomic analysis can be used to identify the organisms present in the respective environment. For example, soil, wastewater, sludge, intestinal

microbiome samples can be analyzed. Initially, metagenomic analyses were restricted to double-stranded DNA genomes. However, further developments have made it possible to analyze all types of genomes, including single-stranded DNA and RNA, which has significantly influenced the discovery of new hypothetical viruses, among them ssRNA bacteriophages. Viral metagenomic analyses have shown that less than 1% of existing viral diversity has been studied to date. Even more surprisingly, 60 to 99% of the viral sequences found in the metagenome analysis may not be homologous to the already known viral genomes (Mokili et al. 2012; Greninger and DeRisi 2015; Krishnamurthy et al. 2016; Shi et al. 2016).

5.2. Virus like particles

Most proteins, including viral CPs, can be recombinantly produced in bacterial, yeast, insect and mammalian cell cultures. The produced CP subunits are able to self-assemble into so-called virus-like particles (VLPs), which structurally and morphologically resemble the structure of the respective viruses. The first expressed phage coat gene was from MS2 in 1998 by Fiers` team (Remaut et al. 1982). VLPs of different origins are currently successfully produced in bacterial systems such as *E.coli* (Rodriguez-Limas et al. 2013), yeast (Freivalds et al. 2014) and insect cell systems (Roldao et al. 2010), as well as in in mammalian cells such as Chinese hamster ovary cells (CHO) or human embryonic kidney cells (HEK) (Rodriguez-Limas et al. 2013). In recent years, methods have been developed to ensure the production of VLPs in plant cells or *in vitro* - a “cell-free” system (see (Kushnir et al. 2012) for a review). Bacterial expression system is the most suitable for production of ssRNA phage CPs, because bacteria are their natural hosts. Regardless of the expression system, usually the produced CPs of ssRNA phages are soluble, in the correct conformation and assembly competent.

Unlike viruses, VLPs cannot replicate because they do not contain the viral genome and other components necessary for infection. VLPs can be formed from multiple copies of a single CP or in case of more complex viruses from copies of several different structural proteins. For example, ssRNA phage, Hepatitis B core antigen HBcAg, human papillomavirus HPV VLPs consist of a single structural protein, while the VLPs of viruses of the *Reoviridae* family are more complex and built from two to four different CPs arranged in multiple layers (Fuenmayor et al. 2017).

The association of CPs of ssRNA phages in regular structures occurs spontaneously. The process is provided using the information "encoded" in the viral CP - the protein folds in the most favorable conformation after synthesis. This means that the formation of VLP occurs without the involvement of other components of the virus or host (Peabody 2013, Mohsen et al. 2017). Capsids of enveloped viruses, such as human immunodeficiency HIV, HCV, influenza viruses, are encircled by the lipid bilayer (Roy and Noad 2009). The production of VLPs of enveloped viruses can be considerably more complex since the genes of several coat and envelope proteins must be co-expressed. Enveloped VLPs also require a more complex host cell system (Kushnir et al. 2012).

5.2.1. Immunogenic properties of VLPs

Highly structured VLPs elicit a significant immune response: immunization with VLPs produces active and long-lasting antibodies (see (Jennings and Bachmann 2008) for a review). The effects of VLPs on immune system are observed at the humoral and cellular

levels. They stimulate B-cell-driven immune responses and have a significant effect on T lymphocytes (Roy and Noad 2009).

VLP diameters are similar to the dimensions of respective viruses and range from 20 to 200 nanometers, a size that allows free entry into the lymphatic vessels and subcapsular region of lymph nodes. Subsequently VLPS are up taken by professional antigen presenting cells (Manolova et al. 2008). Over the time of evolution, the vertebrate immune system has adapted to recognize and respond appropriately to structured particles of viral or other pathogens consisting of repeats of subunits with specific antigens. Viral particle epitopes bind to specific immunoglobulins that are present on B-cell receptors on the surface of B-lymphocytes, acting as a signal for B-cell activation and results in an immediate, T-cell independent immunoglobulin M (IgM) response. IgM is the largest antibody produced by the vertebrate and first appears in response to an antigen. Once VLPs enter the body, they are taken up by antigen-presenting cells by phagocytosis. VLPs are cleaved and displayed by both MHC-II and MHC-I complexes for CD4+ and CD8+ T cells, respectively. To induce a humoral immune response, B cells interact with the CD4+ T helper cell (Th), activating the formation of antibodies and memory cells. To induce cellular immune responses, immature CD8+ cytotoxic T lymphocytes (CTLs) proliferate and differentiate into effector and specific memory CTLs, leading to cytokine release (Fig.10) (Nooraei et al. 2021). VLPs with epitopes are recognized, associated with components of the complement system and multivalent antibodies, and degraded by phagocytosis. In addition to B-cell activation and IgM expression, VLPs cause a T-cell-dependent IgG response, the formation of embryonic (germinal) centers, and B-cell memory, which protects the body from recurrent infection. As mentioned before, due to their small size, VLPs are able to cross pores in the walls of lymphatics and enter the lymphatic system. Upon reaching the lymphoid organs, VLPs may interact directly with B cells and cause the formation of antibodies (see (Jennings and Bachmann 2008) for a review). VLP induces an even stronger T-cell response in the presence of an additional stimulus, such as an adjuvant.

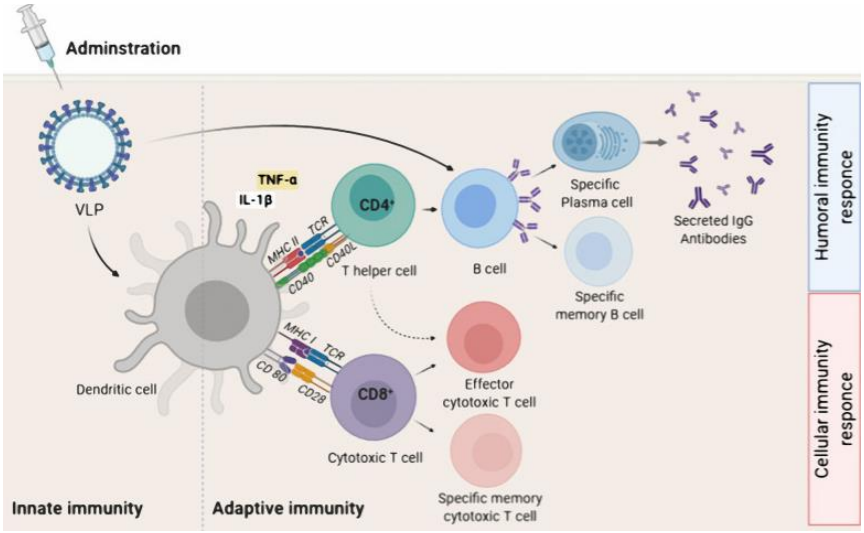

Fig.10. Adaptive immune activation induced by VLP-based vaccines. (Nooraei et al. 2021).

VLPs pack nucleic acids from the host cell and these usually are nearby mRNA molecules or other cellular RNAs, such as rRNA fragments. These molecules can stimulate the immune system using TLR7 (Hemmi et al. 2002) or TLR13 (Shi et al. 2011) receptors. Unspecific nucleic acids can be replaced by agents with desired functions. Loading VLPs with adjuvants can occur during *in vivo* expression by diffusion through the pores of the

particles, or *in vitro* after production of the CP by disrupting the VLPs and then allowing them to reassemble in the presence of adjuvant molecules. For vaccines the presence of adjuvants produces a more effective and stronger immune response. They activate antigen-presenting cells, mainly dendritic cells. Dendritic cell activators include double-stranded RNA (Alexopoulou et al. 2001) and single-stranded RNA (Heil et al. 2004) and DNA fragments rich in CpG sequences (Storni et al. 2004). These substances refer to toll-like receptor ligands. Packaged dsRNA stimulates TLR3, while ssRNA stimulates TLR7/8/13 on the B-cell surface, enhancing the B-cell response to the antigen. CpG-containing DNA fragments act on TLR9 (Jennings and Bachmann 2008). Surface modification and packaging with functional substances creates VLPs as excellent tools for targeted bioimaging and drug delivery (see chapter 23 in (Pumpens 2020) for a review). ssRNA phage CPs and VLPs have also found a number of applications as molecular biology research tools, particularly in the identification of protein-RNA interactions (Dong et al. 2015; Frietze et al. 2016), the creation of peptide display libraries (Chackerian et al. 2011), and the real-time imaging of RNA molecules in living cells (Crossey et al. 2015).

5.2.2. Applications of the ssRNA phage VLPs

Vaccine design

The best-known application of VLPs is their use in vaccine development. Some VLP vaccines are already commercially available, such as vaccines against human papillomavirus - *Cervarix* and *Gardasil*, against hepatitis B virus - *Engerix-B* and *Recombivax* (see (Mohsen et al. 2017)(Roldao et al. 2010) for a review) and hepatitis E virus - *Hecolin* (Huang, 2017). Several other VLP-based vaccines are currently in clinical trials or preclinical evaluation, such as against influenza virus, parvovirus, Norwalk virus, malaria, HIV. In total, more than 30 different VLPs of human and animal viruses were established by 2010 (Roldao et al. 2010), in 2013 this number exceeded 110 different types of VLPs (Zeltins 2013), and in 2017 – 174 (Huang et al. 2017).

VLPs are also used in the development of animal vaccines, such as vaccine against bluetongue, rotavirus, parvovirus (Jennings and Bachmann 2008), chicken anemia virus. The porcine circovirus-2 vaccine has been approved and commercially available since 2009 (Kushnir et al. 2012).

However, the use of VLPs in vaccine development is not restricted to induction of immune response against viruses of origin of the respective VLPs. Instead, VLPs can serve as tools for inducing strong immune response against immunological epitopes derived from other pathogens, undesired self-antigens or small molecules. These modified multimeric protein complexes - VLPs carrying foreign epitopes - are called chimeric VLPs. Over time, VLPs have proved to be very efficient molecular carriers for a variety of epitopes (see chapter 22 in (Pumpens 2020) for a review). VLPs are able to expose foreign epitopes in high density and provide them in certain conformation. The VLPs of ssRNA phages have excellent and well-established scaffold properties and structural tolerance to decoration by foreign immunogenic sequences. Today, many VLPs derived from ssRNA phages are used in vaccine design: AP205, fr, GA, MS2, PP7 and Qβ (Table 1).

Table 1. A list of various vaccines constructed on the ssRNA phage VLPs

VLP	Vaccine target	Method of modification, position of insertion	References
Qβ	Foot-and-mouth disease virus	Genetical fusion in CP C-terminus, stop codon suppression (A1 protein gene)	(Skamel et al. 2014)
	Hepatitis B virus	Genetical fusion in CP C-terminus, stop codon suppression (A1 protein gene)	(Vasiljeva et al. 1998)
	Human immunodeficiency virus (HIV)	C-terminal	(Purwar et al. 2018)
	Paroxysmal nocturnal hemoglobinuria	C-terminal	(Zhang et al. 2017)
	HIV	Chemical coupling	(Huber et al. 2008) (Sommerfelt 2009) (Mogus et al. 2020)
	Allergy	Chemical coupling	(Kundig et al. 2006)
	Allergy, asthma	VLPs mixed with CpG	(Casale et al. 2015)
	Alzheimer`s disease	Chemical coupling	(Chackerian et al. 2006) (Li et al. 2010) (Vandenberghe et al. 2017)
	Cancer	Chemical coupling	(Yin et al. 2016)
	Type 2 diabetes mellitus	Chemical coupling	(Spohn et al. 2017)
	Hypertension, diabetic neuropathy, atherosclerosis, hyperlipidemia	Chemical coupling	(Tissot et al. 2008) (Pan et al. 2019) (Wu et al. 2021)
	Influenza virus	Chemical coupling	(Skibinski et al. 2013)
	Lyme disease	Chemical coupling	(Marcinkiewicz et al. 2020)
	Lymphocytic choriomeningitis virus	Chemical coupling	(Mohsen et al. 2017)
	Malaria	Chemical coupling	(Phares et al. 2017) (Atcheson et al. 2020)
	Parkinson`s disease	Chemical coupling	(Doucet et al. 2017)
	Tauopathy	Chemical coupling	(Maphis et al. 2019)
	Chikungunya virus	Chemical coupling	(Basu et al. 2020)
fr	Hamster Polyomavirus	N-terminus	(Voronkova et al. 2002)
MS2	Cancer	AB loop	(Bolli et al. 2018) (Rolih et al. 2020)
	FMDV	AB loop	(Dong et al. 2016) (Wang et al. 2018)
	HIV	AB loop	(Peabody et al. 2008)

	Human papillomavirus (HPV)	N-terminus of single-chain dimer and AB loop	(Peabody et al. 2017) (Zhai et al. 2017)
	Hypercholesteremia	N-teminus of single-chain dimer	(Crossey, 2015)
	Influenza	AB loop of single-chain dimer	(Lagoutte et al. 2018)
	Malaria	AB loop of single-chain dimer	(Crossey et al. 2015)
PP7	HIV	AB loop of single-chain dimer	(Hunter et al. 2011) (Tyler et al. 2014)
	Influenza	N-terminus	(Lu et al. 2018)
	Pregnancy	AB loop of single-chain dimer	(Caldeira et al. 2015)
	Cancer	AB loop of single-chain dimer	(Sun and Sun 2016)
AP205	HIV	C-terminus N-terminus	(Tissot et al. 2010)
	Cancer	C-terminus N-terminus	(Tissot et al. 2010)
	Hypertension	C-terminus N-terminus	(Tissot et al. 2010)
	Influenza	C-terminus Chemical coupling	(Kirsteina et al. 2020)
	West Nile virus	C-terminus	(Cielens et al. 2014)
	Asthma, allergy	SpyTag-SpyCatcher	(Thrane et al. 2016) (Soongrung et al. 2020)
	Cancer	SpyTag-SpyCatcher	(Palladini et al. 2018)
	Cardiovascular disease	SpyTag-SpyCatcher	(Thrane et al. 2016)
	HIV	SpyTag-SpyCatcher	(Escolano et al. 2019)
	Hypertension	Chemical coupling	(Hu et al. 2017)
	Malaria	SpyTag-SpyCatcher	(Brune et al. 2016)
	Tuberculosis	SpyTag-SpyCatcher	(Thrane et al. 2016)
	WNV	Chemical coupling	(Spohn et al. 2010)
	SARS-CoV-2	SpyTag-SpyCatcher	(Dalvie et al. 2021) – preprint (Fougeroux et al. 2021)

As can be seen in the summarized table, many vaccine candidates against diseases of various origins - viral, bacterial and parasitic infections, systemic diseases, autoimmune diseases, etc., are developed using VLP technology. The broad variety of vaccines indicates that VLP technology can be used in prevention, therapeutic vaccination and treatment. The three most widely used VLPs of ssRNA phages are also highlighted here - Qβ, MS2 and AP205.

Although phage Qβ naturally forms an extended version of the C-terminus of the CP (A1 protein), in general these particles are not suitable for insertion of long foreign epitopes at the N-terminus of the CP or in the loops between the β-strands for robust genetic modification. The technology for developing candidates for the extended CP vaccines is based on stop codon suppression, but this does not achieve the required density of foreign epitope presentation. Qβ VLPs are best suited for chemical modifications. Although most of VLPs have suitable amino acid residues (mostly lysines) on the surface for chemical modification, by some not fully understood reason chemical coupling typically works much better with Qβ VLPs compared to VLPs of other ssRNA phages.

Phage MS2 VLPs are widely used for genetic modification in both the C-terminus and the AB loop. Another interesting modification is frequently used - a single-chain dimer,

where two CP monomers are covalently linked in tandem (Crossey et al. 2015). This technology by twofold reduces the number of exposed antigens, resulting in fewer steric clashes which can expand the spectrum of proteins to be inserted. However, in general MS2 VLPs are not suitable for exposure of large genetically fused epitopes even in the tandem dimer mode.

VLPs derived from phage AP205 have found the widest application in vaccine design. The CP is tolerant of insertions at both the N- and C-termini because the two terminal ends are located on the outer surface of the particles (Shishovs et al. 2016). An interesting method of linking foreign epitopes to AP205 VLPs makes use of SpyTag/SpyCatcher technology (Brune et al. 2016, Thrane et al. 2016) which works very efficiently and now is widely used. AP205 VLPs are tolerant to chemical conjugation as well. It is possible to combine genetic fusion and chemical coupling techniques to expose several different foreign epitopes on the surface of one AP205 VLP (Kirsteina et al. 2020).

Modification of VLPs

VLPs can be modified both externally and internally using genetic fusion or chemical coupling techniques. Binding an antigen to VLP increases its immunogenicity. Modification of VLPs with epitopes that elicit an immune response or with target cell-specific protein "addresses" for drug delivery purposes expands their applications (Pushko et al. 2013).

Chemical modification

Chemical modification methods involve changes in the functional groups on the surface of the VLPs. Amino-, hydroxyl-, carboxyl- or sulfhydryl- groups of amino acid residues can be used. Cysteine, lysine, glutamic acid, aspartic acid and tyrosine side chains found on the surface of VLP can be used in the formation of covalent bonds with antigen molecules (see (Zeltins 2013) for a review). For example, amino group lysine reacts with N-hydroxysuccinimide (NHS) esters to form a stable amide bond. Free cysteine residues readily react with maleimide moiety. By utilizing a linker with NHS ester at one end and a maleimide group at the other end, it is possible to couple an antigen with free cysteine on the surface of the VLP. Lysine and cysteine residues are most commonly used in the chemical modification of VLPs (Mohsen et al. 2017).

The target antigen must be placed in high density on the surface of the particles in order for the VLPs to be as effective as possible in development of immune response. However, often after chemical modification, the density of the antigen on the surface of the particles is insufficient. It is possible that residues of reactive amino acids are hidden, making it difficult to bind antigen. However even in cases when three-dimensional structures of both VLP and antigen is known and presence of reactive groups apparent, for somewhat obscure reasons chemical coupling can still fail to produce appreciable amounts of conjugate. Reactive amino acids, such as lysines or cysteines, can be introduced or removed by altering the viral CP sequence to enhance antigen binding to VLPs (Smith et al. 2013; Peabody 2013). It should be noted that introduced amino acids may affect VLP assembly and stability (Peabody 2013). However, chemical modification of VLPs has several advantages. Prior to conjugation to VLP, the antigen is produced separately and folds independently, meaning that both VLPs and antigen will likely have the correct conformations. Unobstructed separate folding makes it possible to introduce long proteins (Frietze et al. 2016) which often fail to fold correctly as genetical fusions to CPs of respective VLPs. Also, by chemical coupling it is possible to attach virtually any antigen, while genetic fusion is restricted to use of protein or peptide antigens.

Genetic modification

By genetic fusion the desired antigen coding sequence is added to the coding sequence of the viral CP. The resulting sequence is expressed and the antigen is synthesized together with VLP forming proteins (Smith, 2013). As a result, a large number of epitopes are located on the surface of the particles in a specific location and each VLP-forming CP molecule contains an antigen (Jennings and Bachmann 2008). With genetic modification techniques the size of antigens to be added to VLPs is limited (Frietze et al. 2016). If the antigen is significantly longer than the CP, it can affect the assembly of regular particles due to steric clashes (Smith, 2013). Genetical merging of two CP sequences into one chain can make the VLPs more stable, allowing for longer antigen introduction, as one foreign antigen will have to be exposed per two CP monomers (Frietze et al. 2016).

Semi-genetic modification method is utilized in the SpyTag and SpyCather technique which is based on the attachment of antigens to the surface of VLPs with isopeptide bonds. Sequence of SpyCather protein is genetically linked to the coding sequence of CP of AP205. Exposed SpyCatcher moiety spontaneously forms a covalent bond with an antigen, containing SpyTag peptide sequence. This system is called nano-glue or bioglue (Brune et al. 2016, Thrane et al. 2016).

Applications of ssRNA phage CPs and VLPs in imaging systems

RNA-protein interactions in ssRNA phages can be used for the tracking mRNA in living cells. For this purpose, phage CP is modified by making it assembly-deficient and a fluorescent tag, such as green fluorescence protein is attached. The mRNA of interest can be modified by attachment of translational repressor stem-loop sequences of the respective phages. In this way, the mRNA of interest gets labelled and can be tracked in a living cell (Park et al. 2020). Such systems have been developed for MS2 (Bertrand et al. 1998) and PP7 (Larson et al. 2011).

The use of imaging agents in combination with VLPs has contributed to the high-resolution and non-invasive visualization of labelled particles, as well as for the potential treatment of diseases (see Chapter 23 in (Pumpens 2020) for a review). A very recent study showed that VLPs derived from bacteriophages Qβ and PP7 encapsulate small-ultrared fluorescent protein (smURFP) using a versatile capsid disassemble-reassemble approach. Encapsulated smURFP produces bright *in vitro* images following phagocytosis by macrophages. *In vivo* imaging allows evaluation of the distribution of particles at different time points. The promising results marks the potential application of these constructs as non-invasive *in vivo* imaging agents (Herbert et al. 2020).

Targeted delivery

The VLPs of small RNA phages can be used as a tool for the targeted delivery of various diagnostic or therapeutic agents (see Chapter 23 in (Pumpens 2020) for a review). The surface of VLPs can be modified to display an ”address”, able to recognize a particular cell type. This concept assumes that the therapeutic or diagnostic agent packaged in VLPs will label or kill only the specific target cells. For example, peptide able to recognize hepatocellular carcinoma cells can be conjugated to MS2 VLPs, loaded with vastly different types of cargo, including low molecular mass chemotherapeutic drugs, siRNAs, or toxins resulting in selective killing of target cells (Ashley et al. 2011). Some other examples of the addresses used for the targeting of VLPs include cancer cell-targeting proteins (ElSohly et al. 2017), glycans (Rhee et al. 2012), aptamers (Cohen and Bergkvist 2013), antibodies (ElSohly et al. 2015), proteins with tags comprising anionic amino acids or DNA and gold

nanoparticles (Aanei et al. 2018), cell penetrating peptides and apolipoprotein E peptide (Pang et al. 2019).

VLPs also can package miRNAs, which are small, 20-24 nucleotide long, non-coding RNA molecules that bind to the 3'-untranslated regions of the corresponding mRNAs. miRNA binding blocks translation or induces mRNA degradation. This gene-regulating miRNA property is used in therapeutic strategies to regulate the translation of several target proteins, including in the regulation of tumor cells (Fu and Li 2016).

Peptide display libraries

Filamentous phage M13 is the most widely used tool for phage display and provides efficient means for epitope identification. However, a significant drawback of M13 display system is fact that the exposed peptides are not very immunogenic and phage particles cannot be directly used as vaccine candidates. For propriate B-cell activation high density epitope exposure is required. D. Peabody`s group has developed a novel peptide display platform based on VLPs of phage MS2, which are known to package the mRNA of CP, enabling easy sequence determination of selected particles. Thus, the method combines the high immunogenicity of MS2 VLPs with the affinity selection capabilities of other phage display systems (Chackerian et al. 2011).

6. METHODS

6.1.1. Construction of expression plasmids

The CP-encoding sequences from novel phages were ordered from General Biosystems and provided by the manufacturer cloned in pET24a (Novagen) as expression-vector.

Two expression vectors, pET24-M2ex3 and pET28-M2ex3, contain sequences from three distinct variants of influenza A M2e, and NcoI and BamHI sites for cloning were introduced via gene synthesis (General Biosystems). CP encoding sequences were PCR amplified from their respective expression plasmids using oligonucleotides that introduce NcoI and BamHI sites at both ends of the fragment. The PCR products were then cloned into NcoI/BamHI-digested pET24-M2ex3 and pET28-M2ex3 vectors to create expression plasmids for N- and C-terminal CP fusion proteins, respectively.

6.1.2. CP production

For small-scale expression and solubility analysis, *E. coli* BL21(DE3) cells were transformed with CP-encoding plasmids, individual colonies were inoculated in 5 ml of LB media supplemented with 30 μg/ml kanamycin and incubated at 37 °C overnight without shaking. The overnight cultures were transferred into 50 ml of 2xTY medium and the cells were grown at 37 °C or 15 °C with aeration until OD_{600} reached 0.6 to 0.8. IPTG was then added to a final concentration of 1 mM and the cultures were incubated for additional 4 h at 37 °C or 20 h at 15 °C, after which aliquots were harvested by centrifugation for assessment of expression level and solubility by SDS-PAGE.

Large-scale production for VLP purification purposes followed the same protocol using 2 l of 2xTY medium. To determine the solubility of the produced CPs, the aliquoted cells were suspended in lysis buffer (50 mM tris–HCl pH 8.0, 150 mM NaCl, 0.1% Triton X100, 1 mM PMSF) in a wet cell weight/buffer volume ratio of 1:4, lysed by sonication, centrifuged for 30 min at 13000 g and the supernatant and pellet analysed by SDS-PAGE. The same protocol was used for preparation of lysates for VLP purification.

6.1.3. Production and solubility assessment of fusion proteins

E. coli strain BL21(DE3) cells were transformed with plasmids encoding CP fusions, and individual colonies were inoculated in 10 ml of LB media supplemented with 30 μg/ml kanamycin and incubated overnight at 37 °C without agitation. The overnight cultures were transferred to 100 ml of 2xTY medium, and the cells were grown at 37 °C until an OD_{600} of 0.6, at which point IPTG (isopropyl-b-D-thiogalactoside) was added to a final concentration of 1 mM, and the cells were further incubated for 16 h at 20 °C. Cells were harvested by centrifugation. A small aliquot of cells was suspended in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Triton X-100) at a wet weight/volume ratio of 1:25; then, cells were sonicated for 1 min with 5 s on/off pulses (Ultrasonic Processor UP200S, Germany), and the lysate was clarified by centrifugation. The protein production level and soluble and insoluble proteins were analyzed by SDS-PAGE. For the subsequent VLP purification, the expression protocol was upscaled to 2 L.

6.1.4. VLP purification

For small-scale purification, 1 ml of clarified bacterial lysate was loaded on a 12 ml, 6.6 × 400 mm Sepharose 4 FF column (GE Healthcare) equilibrated with PBS. Chromatography was performed on an Acta Prime Plus system (GE Healthcare) with the flow rate set to 0.3 ml/min and fraction size to 1 ml. VLP-containing fractions were detected in SDS-PAGE and those of the highest purity were pooled and applied to a 0.7 ml, 6.6 × 50 mm Fractogel DEAE (M) ion exchange column (GE Healthcare). The flow-through was collected and the column further washed with 2 ml of PBS. Column-bound proteins were eluted with a linear 10 column volume gradient of PBS containing 1 M NaCl using a flow rate of 1 ml/min and fraction size of 1 ml on an Akta Pure 25 system (GE Healthcare). The VLPs were usually found in most of the fractions while the contaminating proteins only in some. The fractions of the highest purity were pooled, dialyzed against 20 mM tris–HCl pH 8.0, supplemented with glycerol to a final concentration of 50% and stored at − 20 °C for downstream experiments. The purification protocol was accordingly upscaled if a larger quantity of VLPs was required.

6.1.5. Purification of chimeric VLPs

Two grams of wet cells were disrupted by sonication, and the clarified cell lysate was precipitated overnight with 40% saturated ammonium sulphate. The precipitate was collected by centrifugation and dissolved in 2 ml of PBS supplemented with 0.2% Tween 20, 0.5 M urea and 1 mM PMSF (phenylmethylsulfonyl fluoride). The concentrated sample was loaded on a 30 ml Sepharose 4FF column (Omnifit 10 mm ID/ 330 mm) attached to an Akta PrimePlus chromatography system (GE Healthcare), and 2 ml fractions were collected in PBS with the flow rate set to 0.5 ml/min. Fractions were analyzed by 1% agarose gel electrophoresis and SDS-PAGE, and those containing VLPs were pooled and applied to a 5 ml Fracto-DEAE ion exchange column (Omnifit 10 mm ID/ 100 mm) pre-equilibrated with PBS. Bound proteins were eluted with a 25 ml 0.15 – 1 M NaCl gradient in PBS with the flow rate set to 2 ml/min and collected in 2 ml fractions. The purest fractions were pooled, dialyzed against PBS and stored frozen at -20 °C for further downstream experiments.

6.1.6. Electron microscopy

For transmission electron microscopy, VLP samples after gel-filtration were adsorbed on carbon-Formvar-coated copper grids and negatively stained with a 1% aqueous solution of uranyl acetate. The grids were examined in a JEM-1230 electron microscope (JEOL Ltd., Tokyo, Japan) operated at 100 kV. Electron micrographs were recorded with iTEM software (version 3.2, Soft Imaging System GmbH) using a side-mounted Morada digital camera (Olympus-Soft Imaging System GmbH, Munster, Germany).

6.1.7. Dynamic light scattering

DLS measurements were performed using Malvern Zetasizer Nano ZS and quartz cuvette ZEN2112, 173 Backscatter according to manufacturer's instructions.

6.1.8. Detection of disulfides in VLPs

Aliquots of purified cysteine-containing VLPs in storage buffer (20 mM Tris–HCl pH 8.0, 50% glycerol) were mixed with an equal volume of Laemmli buffer (0.125 M Tris–HCl

pH 6.8, 20% glycerol) with or without added 5% 2-mercaptoethanol. The samples were heated for 10 min at 95 °C and analyzed on a 15% polyacrylamide gel using a standard Tris–glycine SDS electrophoresis system.

6.1.9. Determination of VLP thermal stability

VLP samples at a concentration of 1 mg/ml in 20 mM tris–HCl, pH 8.0 or PBS were heated for 15 min in a Veriti thermal cycler (Applied Biosystems) in a 5 °C-increment step gradient and then loaded on a 1% agarose gel. After electrophoresis in the TAE buffer, the RNA was visualized with ethidium bromide and protein with Coomassie blue. The thermal stability was defined as the highest temperature at which the VLP band was still visible. For detection of impact on stability by bivalent metal ions or disulfide bonds, capsids before the heating were treated with 20 mM EDTA or 10 mM DTT, respectively.

6.1.10. ELISA

Protein concentrations were determined with the Bradford protein assay kit (Pierce). One hundred microliters of chimeric VLPs or M2ex3 protein (10 mg/mL) were adsorbed on a 96-well ELISA microplate (Sarstedt, Germany) overnight at 4 °C. Plates were blocked with 1% BSA blocking buffer for 1 h at 37 °C. Mouse anti-influenza A M2 monoclonal antibody 14C2 (Invitrogen) diluted at 1:250 in blocking buffer was added in three-fold dilutions to triplicate wells and incubated for 1 h at 37 °C. The plates were washed three times in PBS/0.05% Tween 20 in between different steps. Horseradish peroxidase (HRP)-conjugated polyclonal rabbit anti-mouse IgG (Sigma–Aldrich, USA) was diluted 1:1000 in blocking buffer and incubated for 1 h. Finally, color reactions were developed for 20 min after addition of the o-phenylenediamine substrate. The HRP enzymatic reaction was terminated by the addition of 2.0 M H_2SO_4, and the optical density was measured at 492 nm using an ELISA plate reader (BDSL Immunoskan MS, Finland). Antigenic significance of differences in antibody titers relative to negative control, i.e. the baseline was determined by measuring the OD429 of the wells covered only with BSA. For evaluating the antigenicity, the results were calculated in Microsoft Excel (version 10.0). By connecting the measurement result points obtained for each sample, a linear function intersecting the baseline was graphically represented. The hypothetical cross point was calculated and defined as the end point for evaluating the antigenicity.

Protein X-ray crystallography experiments were performed by Dr. Jānis Rūmnieks and therefore are not described in this section.

7. RESULTS

7.1. "Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data"

7.1.1. Production of the CP genes of novel ssRNA phages

To test the expression of the novel hypothetical ssRNA phage CP genes, 110 published sequences were selected (Greninger and DeRisi 2015; Krishnamurthy et al. 2016, Shi et al. 2016) covering different CP similarity groups and representing maximum diversity both in sequence and length. In some of the genomic sequences, there were two or three ORFs of similar lengths found in between the maturation and replicase genes, making it difficult to identify the CP ORF. Therefore, in these cases 2 or 3 sequences from a single phage were included in the study.

All of the CP ORFs were expressed in *Escherichia coli* using a T7 promoter-driven system in standard conditions. Most of the CPs were produced in the expected high amounts and only in nine cases no expression was detected. A subsequent solubility analysis revealed that about 60% of the CPs are at least partially soluble while the rest were found in inclusion bodies. Since some of the metagenomic sequences were obtained from sites below ambient temperature 20 °C, the synthesis of all insoluble protein was tested at 15 °C. Indeed, the results were not disappointing, 85% of previously insoluble CPs were at least partially soluble at 15 °C and only six remained in inclusion bodies.

7.1.2. Purification of proteins and characterization of VLP morphology

In the following process, all soluble proteins were purified by the classical chromatography method for VLP purification - gel filtration by sepharose 4FF. A total of 80 CPs were assembled in VLPs, as confirmed by electron microscopy. In the majority of cases, the VLP morphology resembled that of the previously characterized ssRNA phage VLPs with an apparent spherical shape 28 to 30 nm in diameter, corresponding to T = 3 icosahedral particles. The particles in some cases were slightly larger, possibly representing T = 4 symmetry, but in these cases, particles appeared somewhat irregular or incomplete (AVE000, GA-492). In two cases smaller particles presumably harboring T = 1 symmetry (NF-391, NF-443) were observed, and in AVE007 - a mixture of T = 1 and T = 3 particles. Several AVE015-like VLPs (AVE015, AVE016 and GQ-112) revealed a mixture of spherical and elongated particles.

7.1.3. Stabilizing disulfide bonds in the novel VLPs

In addition to non-covalent protein-protein interactions, capsid stability can be mediated by protein-RNA interactions, metal ions and disulfide bonds. The disulfides markedly increase the particle stability and have been a fundamental factor in the application of Qβ and AP205 VLPs as the most successful ssRNA phage-derived carriers.

In the previously studied Cys containing phages, disulfide bonds are formed between threefold and fivefold symmetry axes that results in covalently linked CP pentamers and hexamers in the capsid. Seventeen new VLPs containing at least two cysteine residues were selected for the assay and tested in non-reducing SDS-PAGE. Only EMS011, ESE001, Hubei10, GA-492 and GA-879 produced a pair of bands corresponding to the expected

pentameric and hexameric species. In the rest of cases, several lower molecular weight bands could be observed, indicating that all possible disulfide bridges have not been formed or are not able to form within VLPs. T = 1 particles of NF-443 VLPs involved only pentameric interactions, since hexameric contacts are absent in T=1 lattice. AVE018 VLPs apparently formed only a disulfide-linked CP dimer.

7.1.4. VLP thermal stability

Another important characteristic for VLPs is their thermal stability. It is necessary for high performance in downstream applications such as vaccine development, drug delivery, bioimaging, etc. To determine the thermal stability of the novel ssRNA phage VLPs, particles were subjected to increasing temperatures and their integrity visualized in native agarose gel electrophoresis. The thermal stability Tm was defined as the highest tested temperature at which the VLP band can no longer be observed.

More than 75% of the tested VLPs disassembled between 50 °C and 70 °C, a few were extremely unstable and were destroyed even at 35 °C, while some others had an unusually high Tm up to 95 °C. None of the newly characterized VLPs with the top three melting temperatures (AVE016 - 95 °C, ESE007 - 90 °C, Beihai14 - 85 °C) had disulfide bonds between the subunits while the five VLPs with experimentally detected inter-subunit disulfides displayed rather average thermal stability (55-70 °C). These results significantly alter the previously established view that inter-subunit disulfide bonds are necessary for formation of thermally stable ssRNA phage VLPs. The results convincingly demonstrate that stable particles can also be achieved by only non-covalent interactions.

Liekniņa et al. J Nanobiotechnol (2019) 17:61
https://doi.org/10.1186/s12951-019-0497-8

Journal of Nanobiotechnology

RESEARCH Open Access

Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data

Ilva Liekniņa, Gints Kalniņš, Ināra Akopjana, Jānis Bogans, Mihails Šišovs, Juris Jansons, Jānis Rūmnieks and Kaspars Tārs*

Abstract

Background: Protein shells assembled from viral coat proteins are an attractive platform for development of new vaccines and other tools such as targeted bioimaging and drug delivery agents. Virus-like particles (VLPs) derived from the single-stranded RNA (ssRNA) bacteriophage coat proteins (CPs) have been important and successful contenders in the area due to their simplicity and robustness. However, only a few different VLP types are available that put certain limitations on continued developments and expanded adaptation of ssRNA phage VLP technology. Metagenomic studies have been a rich source for discovering novel viral sequences, and in recent years have unraveled numerous ssRNA phage genomes significantly different from those known before. Here, we describe the use of ssRNA CP sequences found in metagenomic data to experimentally produce and characterize novel VLPs.

Results: Approximately 150 ssRNA phage CP sequences were sourced from metagenomic sequence data and grouped into 14 different clusters based on CP sequence similarity analysis. 110 CP-encoding sequences were obtained by gene synthesis and expressed in bacteria which in 80 cases resulted in VLP assembly. Production and purification of the VLPs was straightforward and compatible with established protocols, with the only exception that a considerable proportion of the CPs had to be produced at a lower temperature to ensure VLP assembly. The VLP morphology was similar to that of the previously studied phages, although a few deviations such as elongated or smaller particles were noted in certain cases. In addition, stabilizing inter-subunit disulfide bonds were detected in six VLPs and several possible candidate RNA structures in the phage genomes were identified that might bind to the coat protein and ensure specific RNA packaging.

Conclusions: Compared to the few types of ssRNA phage VLPs that were used before, several dozens of new particles representing ten distinct similarity groups are now available with a notable potential for biotechnological applications. It is believed that the novel VLPs described in this paper will provide the groundwork for future development of new vaccines and other applications based on ssRNA bacteriophage VLPs.

Background

The single-stranded RNA (ssRNA) bacteriophages of the Levivirdae family are small viruses that infect a variety of Gram-negative bacteria. Their virions consist of a compact, approximately 3500 to 4200 nucleotide-long genome packaged in a small, spherical-looking protein shell about 28 nm in diameter with an underlying T = 3 quasi-equivalent icosahedral symmetry. The capsid is constituted of the major coat protein (CP) and one or two species of minor structural proteins that are involved in recognition and packaging of the genome and are required to adsorb the virion to the bacterial receptor and convey the RNA genome into the cell. At least for the currently studied ssRNA phages, the minor virion proteins are not essential either for assembly or for structural integrity of the protein shell, and recombinant expression

*Correspondence: kaspars@biomed.lu.lv
Latvian Biomedical Research and Study Center, Rātsupītes 1, Riga LV1067, Latvia

of a cloned coat protein gene results in the appearance of virus-like particles (VLPs) that are morphologically very similar to native virions but have spontaneously packaged bacterial RNA inside the particles instead of the genome [1–4].

The ssRNA phage VLPs have found a variety of applications, mostly in the field of vaccine development where various antigens are presented onto the capsid surface to invoke a strong immune response. Phage Qβ VLPs conjugated with various peptide and small-molecule moieties have reached clinical trials against conditions such as hypertension [5], asthma [6] or smoking addiction [7], phage MS2 VLPs have been successfully used as carriers for epitopes from the human papilloma virus [8], while modified phage AP205 VLPs have shown promising results as vaccine candidates against West Nile virus [9]. The ssRNA phage VLP technology has been further extended for encapsulation of both macromolecular and small-molecule substances of interest inside the particles, which in combination with VLP surface modification allows for the development of targeted bioimaging and drug delivery agents (see [10, 11] for comprehensive reviews). ssRNA phage CPs and VLPs have also found a number of applications as tools for molecular biology research, notably in generation of peptide display libraries [12], identification of protein–RNA interactions [13, 14] and real-time imaging of RNA molecules in living cells [15].

Up to recently, the number of known ssRNA phages has been rather small. All of the currently identified phages use various Proteobacteria as their hosts; the great majority of these infect *Escherichia coli* and related Enterobacteria, while the remaining few target bacteria of the *Pseudomonas*, *Acinetobacter* or *Caulobacter* genera. The CPs of the Enterobacteria- and *Pseudomonas*-specific phages share very low, yet still detectable sequence similarity, while those from the *Acinetobacter* and *Caulobacter* phages, of which only a single representative of each has been sequenced, have no sequence similarity to other CPs [16, 17]. The different CPs vary considerably both in their overall amenability for modification and for tolerance of particular foreign sequences, as well as in their capability to recognize specific RNA for encapsulation and the stability of the assembled VLPs. Oftentimes, for a particular antigen a number of different VLP carriers and modification strategies have to be screened until a suitable one, if any, is found. In vaccine development and related areas, the immune response against the carrier coat protein has also to be taken into account, and a narrow range of available CPs adversely limits the number of potential vaccines that could be produced using the ssRNA phage platform. Discovery and characterization of novel ssRNA phage CPs and VLPs is therefore of considerable interest for continued developments in the area.

While no novel ssRNA phages have been isolated lately, the increasing metagenomic sequencing efforts in recent years have uncovered a previously unknown diversity of these viruses in nature. In 2015, genomes of two novel Leviviridae phages EC and MB were assembled from San Francisco wastewater [18], and soon a much wider study revealed over 150 partial ssRNA phage sequences in different RNA metagenomes [19]. A survey of RNA virus sequences from invertebrates resulted in more than 60 additional ssRNA phage genome sequences [20]. In the majority of the partial genomes, an open reading frame (ORF) between the conserved maturation and replicase genes can be identified that putatively encodes a coat protein, although the ORFs show great variation in length and sequence and in numerous cases no similarity to the known ssRNA phage CPs or any other proteins.

While the metagenomic studies have greatly expanded the known ssRNA phage diversity, infectious phages cannot be resurrected from the partial genome sequences, and their host bacteria, along with many other aspects of their biology, remain unknown. However, the CPs of the previously studied ssRNA phages have been able to assemble into VLPs in absence of other phage components, which provides an opportunity to obtain and study ssRNA phage VLPs even if the CP sequence is the only available information. In this study, we acquired 110 putative ssRNA phage CP-encoding ORFs from the metagenomic data using gene synthesis, and here we report the expression, purification and characterization of 80 novel ssRNA phage VLPs.

Results

CP similarity analysis

Based on multiple sequence alignment, the previously known ssRNA phage CPs can be divided into three broad similarity groups represented by the Enterobacteria- and *Pseudomonas*-infecting phages, the *Acinetobacter* phage AP205 [16] and the *Caulobacter* phage Cb5 [17], respectively. To reassess the ssRNA phage CP diversity in light of the new data, we compiled for comparison a set of all of the published CP sequences together with some additional ones that could be located in NCBI's nucleotide databases. However, for the CP sequences from the metagenomic data, a multiple sequence alignment was deemed unreliable due to the often very weak sequence similarity and broadly variant protein length that ranged from 105 to 208 residues compared to only 122 to 132 in the previously known phages. All available ssRNA phage CP sequences were therefore subjected to a BLAST similarity analysis, followed by UPGMA clustering based

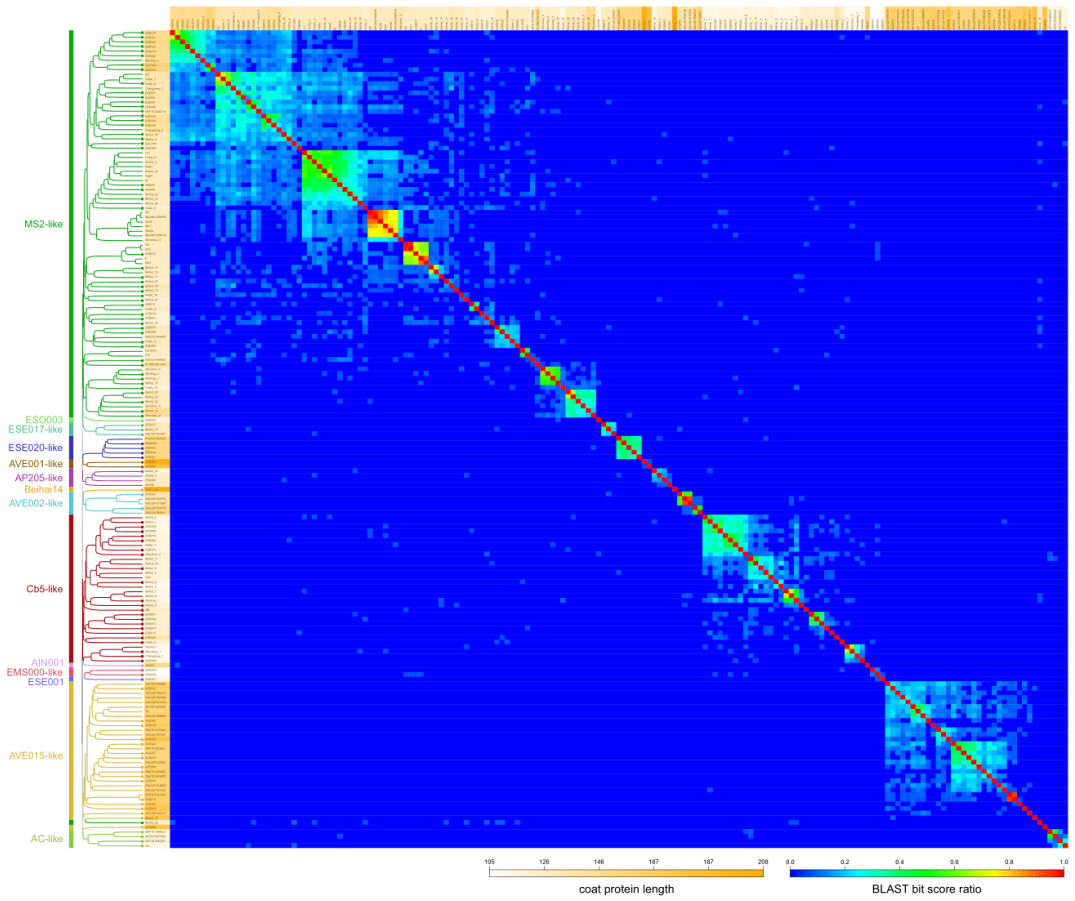

Fig. 1 ssRNA phage CP similarity groups. All available CP sequences were compared in BLAST analysis followed by UPGMA clustering based on BLAST bit score ratios (hit score/self score). The resulting dendrogram and heat map representation of CP diversity are presented. The orange shading of CP labels corresponds to their length distribution. The CPs that were studied experimentally in this paper are indicated with dots at branch tips

on BLAST bit score ratios (hit score/self score). The resulting clustering analysis and the resulting tree representation (Fig. 1), while not a proper phylogenetic reconstruction, provides useful information regarding the diversity of the novel CPs and their relatedness to the previously known phages.

Almost a half, or over 80, of the known ssRNA phage CP sequences cluster into an MS2-like supergroup with approximately ten recognizable subgroups. Two of those represent the currently recognized *Levivirus* and *Allolevivirus* F-pili specific phage genera, with an adjacent cluster containing the other conjugative-pili specific phages M [21], C-1 [22], PRR1 [23] and Hgal1 [22]. Two additional neighboring CP clusters are comprised solely of metagenomic sequences, with increasing CP length and decreasing CP similarity to the known phages. The supergroup includes some additional smaller clusters with low similarity to MS2 [24] or Qβ [4], one of which contains the previously known *Pseudomonas* phage PP7 [25]. A Cb5-like supergroup emerges as the second largest that includes the *Caulobacter* phage Cb5 and more than 30 related CP sequences from the metagenomic data. These further cluster into five subgroups, all formed by short and rather uniform-length proteins (most around 115–120 residues). One more supergroup with some 30 sequences emerges which is comprised entirely

of metagenomic sequences and contains two major recognizable subgroups. The CPs in this "AVE015-like" supergroup, named after a representative phage from the group, are considerably bigger than those of MS2- or Cb5-like phages with an average length around 165 residues, with some of them spanning 180 residues. The three supergroups comprise approximately 80% of all of the known ssRNA phage CP sequences; the rest cluster into several considerably smaller clades ranging from two to five CPs, designated here the ESE017-like, ESE020-like, AP205-like, AVE001-like, AVE002-like, EMS000-like and AC-like groups. Several CP sequences (AIN001, ESE001, ESO003 and Beihai14) have no reliable assignment to any of the clusters and were considered orphans. (For brevity, phage names such as "Beihai levi-like virus 14" from [20] are referred to as "Beihai14" in this paper.)

Expression of the novel CPs

Our BLAST analysis allowed to recognize approximately 14 distinct ssRNA phage coat protein types, which is a noticeable increase from the three CP types known before. We selected 110 CP sequences from the metagenomic data to cover all CP groups and represent maximum diversity both in sequence and in length, and obtained the sequences using gene synthesis to study them experimentally. Interestingly, in a few of the genome sequences, there were two or three predicted ORFs of similar length between the maturation and replicase genes. The predicted coat protein ORF was always the one immediately following the maturation gene, however, the other ORFs were also included for experimental characterization. All protein sequences used in the study are available in Additional file 1: Table S1.

All of the CP ORFs were initially expressed in *Escherichia coli* using a T7 promoter-driven system in standard conditions. The vast majority of the CPs were produced in the expected high levels and only in very few cases no expression was detected (Additional file 2: Figure S1). A subsequent solubility analysis (Additional file 2: Figure S2) however revealed that only about 60% of the CPs are at least partially soluble while the rest were found in inclusion bodies. In an effort to mitigate the issue, we expressed the insoluble proteins at 15 °C that indeed rendered 85% of the previously insoluble CPs at least partially soluble and only six remained in inclusion bodies also at the lower temperature. It can be noted that the few non-CP-encoding ORFs included in the analysis were either not expressed at all or were insoluble. Expression and solubility data are summarized in Table 1.

Purification and characterization of VLP morphology

After CP production, the crude *E. coli* lysates were separated by gel filtration and the fractions analyzed for CP presence in the expected molecular weight range for VLPs. In total 80, or approximately 72%, of the soluble CPs assembled into VLPs as confirmed by electron microscopy (Fig. 2). In the majority of cases, the VLP morphology resembled that of the previously characterized ssRNA phage VLPs with an apparent spherical shape 28 to 30 nm in diameter that corresponds to a T = 3 icosahedral particle. However, notable deviations from the standard particle size and shape were not uncommon. The VLPs formed by the AVE000 CP were noticeably bigger, reaching 35 to 40 nm in diameter which could correspond to a T = 4 icosahedral particle, but the preparation was rather heterogeneous with many elongated, squashed or incomplete particles. A somewhat similar view was observed also in GALT01000492 VLP preparations where the T = 3 particles were present in minority while the field of view was dominated by bigger VLP-resembling irregular objects. NFYT01000391 and NFZC01007443 CPs assembled into small particles approximately 18 nm in diameter with a presumed T = 1 icosahedral symmetry. In some other cases, two distinct VLP morphologies were present in the sample: a sizeable proportion of AVE016 VLPs appeared to have an elongated shape, while AVE007 and Beihai17 VLP preparations contained a mixture of T = 3 and T = 1 particles.

To further characterize the VLPs, we used dynamic light scattering (DLS) to determine the average particle size in solution. In a homogeneous sample, the particle size measured using DLS is in good agreement with values determined using other methods such as electron microscopy or X-ray crystallography, while significant deviations are an indicator of particle aggregation. For the majority of VLPs, the determined average particle diameter values (Table 1) lie within a range of 25–30 nm which is in good agreement with the size observed in EM. Using DLS, the NFYT01000391 and NFZC01007443 VLPs measured 18 to 19 nm in diameter which corresponds to the apparent T = 1 particles detected in EM, and likewise the bigger AVE000 VLPs measured 38 nm in diameter and the apparently heterogenous GALT01000492 preparation had an average particle diameter of 48 nm. VLPs with significant discrepancies between the EM and DLS data, such as NFYT01000214 with an observed diameter of 28–30 nm in EM but a measured size of 42 nm using DLS, or VLPs for which no reasonable estimate could not be obtained, likely indicate significant amount of aggregation in the samples.

Stabilizing disulfide bonds in the novel VLPs

Coat protein modifications introduced for vaccine development and related applications have a tendency to destabilize the assembled VLPs, and the experimental

Table 1 Properties of experimentally studied ssRNA phage coat proteins

CP	Similarity group	Length	Cysteines		TR	Production	Solubility		VLPs			
			Positions	S–S			37 °C	15 °C	EM	Morphology	d, nm	Tm, °C
MS2	*MS2*	*129*			++	+++	+++	*n.d.*	+++	*T=3*	28	70
Qβ	*MS2*	*132*	*74, 80*	*56*	++	+++	+++	*n.d.*	+++	*T=3*	30	90
PRR1	*MS2*	*131*			++	+++	+++	*n.d.*	+++	*T=3*	n.d.	75
PP7	*MS2*	*127*	*67, 72*	*56*	++	+++	+++	*n.d.*	+++	*T=3*	27	90
AIN000	MS2	131			+	++	–	++	++	T=3	n.d.	45
AIN002	MS2	132				+++	–	++	++	T=3	n.d.	55
AIN003	MS2	140				+++	–	++	++	T=3	28	55
AIN010	MS2	125				±	+	n.d.	–	n/a	n/a	n/a
AVE017	MS2	129			+	+++	+++	n.d.	++	T=3	26	65
AVE019	MS2	123			+	+++	+++	n.d.	+	T=3	27	50
AVE021	MS2	123			+	+++	++	n.d.	++	T=3	30	60
Beihai16	MS2	128			+	+++	+++	n.d.	–	n/a	n/a	n/a
Beihai17	MS2	129				+++	+++	n.d.	++	T=3, T=1	26	55
Beihai18	MS2	131				+++	+++	n.d.	+	T=3	29	50
Beihai19	MS2	134			+	+++	+++	n.d.	+	T=3	28	55
Beihai21	MS2	126			+	+++	++	n.d.	+	T=3	n.d.	40
Beihai23	MS2	130				+++	+	n.d.	+	T=3	28	35
Beihai26	MS2	123			+	+++	+++	n.d.	+	T=3	n.d.	50
Beihai27	MS2	122				+++	+++	n.d.	–	n/a	n/a	n/a
Beihai28	MS2	132			+	+++	+++	n.d.	++	T=3	30	65
Beihai30	MS2	149				+++	++	n.d.	++	T=3	n.d.	65
Beihai32	MS2	130			+	+++	++	n.d.	++	T=3	31	70
Beihai33	MS2	126			+	+++	+++	+++	++	T=3	28	n.d.
Beihai34	MS2	121				+++	+++	n.d.	+	T=3	25	50
EMS014	MS2	156			+	+++	+	n.d.	+	T=3	n.d.	n.d.
EOC000	MS2	129			+	+++	–	++	++	T=3	n.d.	n.d.
EOC005	MS2	159				+++	–	++	+	T=3	n.d.	60
ESE005	MS2	149			+	+++	–	–	n.d.	n/a	n/a	n/a
ESE006	MS2	129			+	+++	–	+++	++	T=3	n.d.	n.d.
ESE007	MS2	137	93, 115	M		+++	–	+++	++	T=3	30	90
ESE009	MS2	130				+++	–	+	–	n/a	n/a	n/a
ESE010	MS2	132				+++	+	n.d.	–	n/a	n/a	n/a
ESE012	MS2	143			+	+++	+	n.d.	+	T=3	n.d.	n.d.
ESE019	MS2	150				+++	–	++	++	T=3	32	50
ESE021	MS2	150			+	+++	–	+	+	T=3	n.d.	50
ESE024	MS2	149				+++	–	++	+	T=3	29	50
ESE025	MS2	152				+++	–	++	+++	T=3	36	40
ESE029	MS2	132				+++	+++	n.d.	+++	T=3	22	70
ESE030	MS2	142				+++	–	++	+++	T=3	n.d.	35
ESE037	MS2	140			+	+++	–	+++	++	T=3	n.d.	70
ESE046	MS2	137				+++	–	+++	++	T=3	31	36
ESE058	MS2	146			+	+++	–	++	+++	T=3	n.d.	55
ESO010	MS2	149			+	+++	–	+	–	n/a	n/a	n/a
Hubei2	MS2	124				+++	+	n.d.	–	n/a	n/a	n/a
Hubei3	MS2	128				+++	–	++	+	T=3	n.d.	60
Hubei6	MS2	125			+	+++	+	n.d.	++	T=3	24	60
Hubei8	MS2	137				+++	–	++	+	T=3	26	40
Hubei10	MS2	129	76, 77	56		+++	+++	n.d.	++	T=3	30	70

Table 1 (continued)

CP	Similarity group	Length	Cysteines		TR	Production	Solubility		VLPs			
			Positions	S–S			37 °C	15 °C	EM	Morphology	d, nm	Tm, °C
Hubei14	MS2	117			+	+++	+++	n.d.	+++	T=3	28	70
NFYT01000214	MS2	135	78,79,95			+++	++	n.d.	++	T=3	42	65
NFZC01009824	MS2	122	68, 70	n.d.		+++	+	n.d.	++	T=3	n.d.	n.d.
Shahe3	MS2	136			+	+++	–	+	++	T=3	29	50
Wenling2	MS2	128			+	+++	+++	n.d.	++	T=3	28	70
Wenling3	MS2	126			+	+++	++	n.d.	+	T=3	26	60
Wenzhou4	MS2	146			+	+++	–	++	+++	T=3	30	50
AVE004	AVE015	159	107,108	n.d.	+	+++	+++	n.d.	+	T=3	n.d.	n.d.
AVE006	AVE015	180				+++	++	n.d.	+	T=3	n.d.	n.d.
AVE007	AVE015	156	46.134	N		+++	++	n.d.	++	T=3, T=1	30	n.d.
AVE022	AVE015	158	59, 136		+	+++	+	n.d.	–	n/a	n/a	n/a
AVE024	AVE015	156	39, 58	N	+	+++	++	n.d.	+	T=3	n.d.	n.d.
AVE039	AVE015	158	56,89, 130	N	+	+++	++	n.d.	++	T=3	n.d.	60
GALT01000492	AVE015	165	103,104, 105	56		+++	++	n.d.	+++	Heterogenous	48	70
GALT01093879	AVE015	163	101,102,103	56		+++	++	n.d.	++	T=3	30	75
AVE000	AVE015	167	61,98,99	N		+++	–	+++	+++	T=4?	38	n.d.
AVE005	AVE015	164	94, 112	N	+	+++	–	+++	+	T=3	40	70
AVE015	AVE015	167			+	+++	–	++	++	T=3	n.d.	65
AVE016	AVE015	166	113, 158	N		+++	+++	n.d.	++	T=3, some elongated	30	95
AVE018	AVE015	169	9, 43, 92	D	+	+++	+++	n.d.	+	T=3	28	70
AVE020	AVE015	180	38, 146			+++	++	n.d.	–	n/a	n/a	n/a
AVE023	AVE015	177			+	+++	–	–	n.d.	n/a	n/a	n/a
Beihai12	AVE015	178			+	+++	+++	n.d.	–	n/a	n/a	n/a
GALQ01044112	AVE015	164	8,59,113	N		+++	+++	n.d.	++	T=3	32	80
AVE002	AVE002	140			+	+	+	n.d.	+++	T=3	31	65
GALQ01040378	AVE002	151				–	–	–	n.d.	n/a	n/a	n/a
Beihai13	ESE017	132				+++	–	+	++	T=3	n.d.	n.d.
ESE017	ESE017	132				+++	+	n.d.	+	T=3	27	70
GALQ01034907	ESE017	131				++	+	n.d.	+	T=3	27	60
AP205	*AP205*	*130*	*64, 68*	*56*		+++	+++	*n.d.*	+++	*T=3, T=1*	28	75
Beihai20	AP205	117				+++	++	n.d.	–	n/a	n/a	n/a
Cb5	*Cb5*	*122*				+++	+++	*n.d.*	+++	*T=3*	28	70
EMS002	Cb5	123	64, 68		+	+++	–	++	–	n/a	n/a	n/a
EMS011	Cb5	123	64.68	56		+++	–	++	++	T=3	33	55
EMS017	Cb5	123	64, 68			–	–	n.d.	n.d.	n/a	n/a	n/a
ESE016	Cb5	123				+++	++	n.d.	–	n/a	n/a	n/a
ESO001	Cb5	124			+	+++	–	++	++	T=3	40	n.d.
Beihai3	Cb5	119				+++	±	n.d.	+	T=3	24	60
Beihai6	Cb5	119			+	+++	+++	n.d.	–	n/a	n/a	n/a
MB	Cb5	127				++	–	++	–	n/a	n/a	n/a
Beihai1	Cb5	116			+	++	+	n.d.	+	T=3	n.d.	n.d.
EMS007	Cb5	116			+	+++	±	n.d.	–	n/a	n/a	n/a
ESE008	Cb5	138			+	+++	–	+++	–	n/a	n/a	n/a
ESE015	Cb5	119				+++	+	n.d.	+	T=3	31	55
ESE022	Cb5	116			+	+	–	++	–	n/a	n/a	n/a
ESE026	Cb5	119			+	+++	+++	n.d.	+	T=3	n.d.	50
ESO000	Cb5	115			+	+++	–	++	–	n/a	n/a	n/a
Wenzhou2	Cb5	122				+++	±	n.d.	++	T=3	26	60

Table 1 (continued)

CP	Similarity group	Length	Cysteines		TR	Production	Solubility		VLPs			
			Positions	S–S			37 °C	15 °C	EM	Morphology	d, nm	Tm, °C
Beihai9	Cb5	123			+	+++	+	n.d.	++	T = 3	29	50
Wenling1	Cb5	116				+++	+	n.d.	–	n/a	n/a	n/a
Changjiang1	Cb5	112			+	+++	+	n.d.	+	T = 3	n.d.	n.d.
EMS005	Cb5	112			+	+++	++	n.d.	–	n/a	n/a	n/a
Wenzhou1	Cb5	113			+	+++	+	n.d.	++	n/a	n/a	n/a
EMM000	ESE020	155			+	+++	–	++	+	T = 3	23	50
EMS001	ESE020	154				+++	+	n.d.	++	T = 3	n.d.	65
ESE020	ESE020	153			+	+++	+++	n.d.	+	T = 3	36	65
ESE041	ESE020	155			+	+++	–	+++	+	T = 3	34	50
AC	AC	115				+++	–	+++	++	T = 3	n.d.	50
NFYT01000391	AC	123	66,67,75	N		+++	+	n.d.	++	T = 1	18	75
NFZC01007443	AC	119	64, 66	5		+++	++	n.d.	++	T = 1	19	60
EMS000	EMS000	105	45, 48		+	+	++	n.d.	–	n/a	n/a	n/a
EMS003	EMS000	106				+++	–	+	–	n/a	n/a	n/a
AVE001	AVE001	202	82,83,84		+	+++	–	–	n.d.	n/a	n/a	n/a
AVE003	AVE001	183	88,90,176		+	+++	–	–	n.d.	n/a	n/a	n/a
Beihai14	Beihai14	208			+	+++	–	++	++	T = 3	31	85
ESE001	ESE001	118	61, 66	56		+++	±	n.d.	++	T = 3	30	75
ESO003	ESO003	113			+	+++	+	n.d.	++	T = 3	29	60
AIN001	AIN001	155			+	++	–	–	n.d.	n/a	n/a	n/a

Some previously studied CPs are included for reference and shown in italic. The listed properties include the assigned CP similarity group, length, presence of a translational repressor stem-loop (TR) in the genome (+, a putative hairpin structure predicted; ++, an experimentally confirmed TR), positions of cysteine residues in the protein if more than one is present, disulfide bonds in VLPs (56, covalently linked pentamers and hexamers; 5, pentamers; D, dimers; N, no disulfides detected), production level (+++, high; ++, average; +, low; ±, very low; –, not detected), solubility at 37 °C and 15 °C (+++, highly soluble; ++, at least 50% soluble; +, less than 50% soluble; –, completely insoluble), VLP formation by EM: (+++, highly efficient VLP formation; ++, reasonably good VLP formation; +, some detectable VLPs; –, no VLPs observed), characterization of VLP morphology, particle diameter from DLS measurements, and their "melting" temperature (thermal stability). n/a: not applicable due to lack of VLPs, n.d.: not determined

success rate appears to positively correlate with the stability of the starting unmodified particles. While in some of the studied ssRNA phages inter-subunit contacts are mediated solely by non-covalent protein–protein interactions, in others coordinated metal ions [26, 27] and protein–RNA interactions [27] have been found that contribute to particle stability. In yet other phages such as Qβ [28], PP7 [29] and AP205 [30] the CP subunits are covalently linked together with disulfide bonds. The disulfides markedly increase the particle stability and have been a substantial factor in the advancement of Qβ and AP205 VLPs as the most successful ssRNA phage-derived carriers. Screening for stabilizing disulfide bonds in the novel VLPs is therefore of interest for selecting the best candidates for future VLP carriers.

In all of the previously studied ssRNA phage particles where disulfide bonds exist, they are formed between CP loops positioned around the icosahedral threefold and fivefold symmetry axes that results in covalently linked CP pentamers and hexamers in the capsid. It cannot be excluded, however, that in other phages stabilizing disulfide bonds might occur also in other positions. We therefore selected all experimentally available CPs that were able to assemble into VLPs, could be purified to near homogeneity and which contained at least two cysteine residues, and subjected the VLPs to denaturing but non-reducing conditions. In such conditions the disulfide-containing Qβ and AP205 VLPs disassemble into pentameric and hexameric CP species that can be tracked in SDS–polyacrylamide gel electrophoresis (Fig. 3). From the 17 tested novel VLPs, only EMS011, ESE001, Hubei10, GALT01000492 and GALT01093879 produced a pair of bands corresponding to the expected pentameric and hexameric species (Fig. 3); in most cases, a number of lower molecular weight complexes could also be discerned, suggesting that not all of the possible disulfide bridges have been formed in the VLPs. All of these five CPs contain cysteine residues located similarly to Qβ or AP205 approximately in the middle of the sequence; the EMS011 and ESE001 CPs contain two cysteine residues five or six positions apart, while the Hubei10 CP has two and GALT01000492 and GALT01093879 have three consecutive cysteine residues. In the latter two CPs, apparently only two of the residues

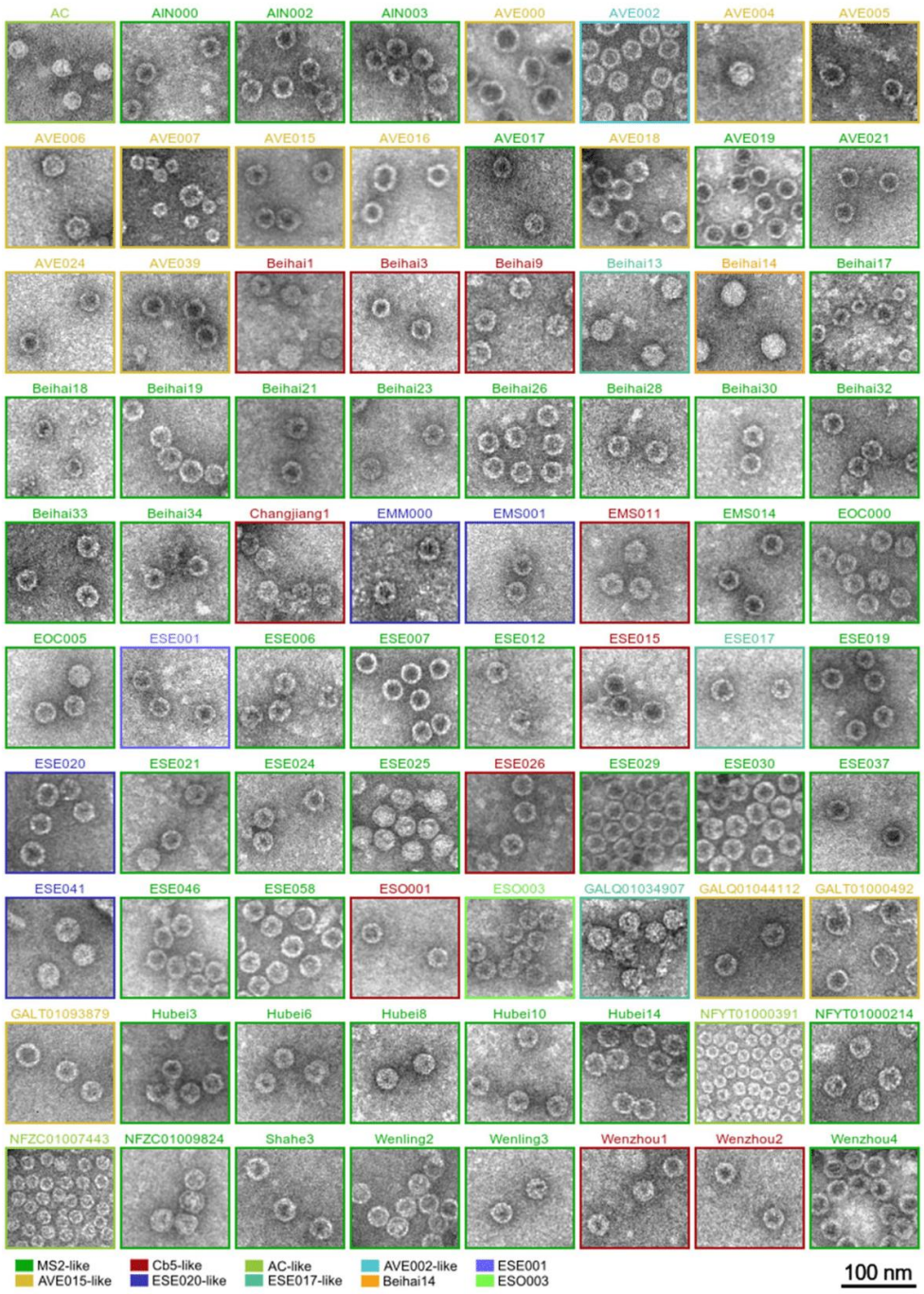

Fig. 2 Representative electron micrographs of the produced VLPs. The constituent CP similarity groups as of Fig. 1 are indicated

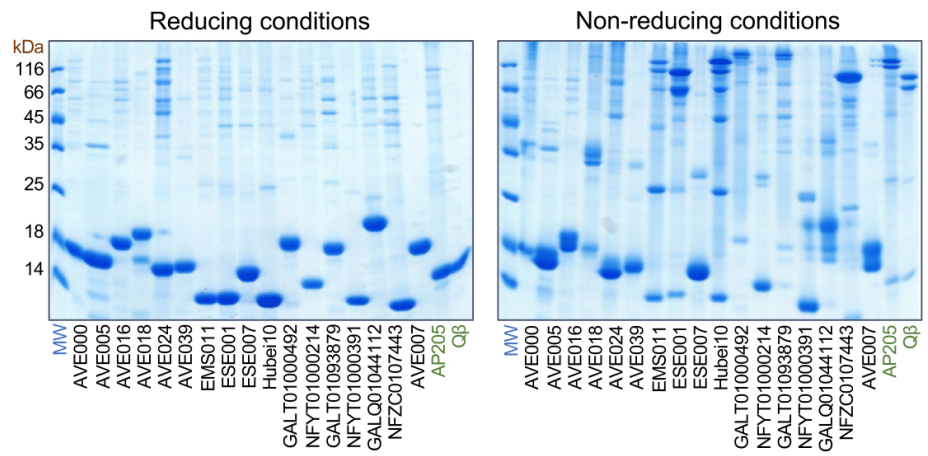

Fig. 3 Disulfide bonds in the produced VLPs. The VLPs were subjected to denaturing and reducing or non-reducing conditions and analyzed in SDS polyacrylamide gel electrophoresis

are involved in inter-subunit contacts, although different pairs of cysteine side chains might be involved in making pentameric and hexameric contacts. From the tested proteins also the NFZC0107443 CP contains two similarly located cysteine residues three positions apart, but in non-reducing conditions the VLPs resolved into only a single higher molecular weight species. This is however consistent with the assumed T = 1 icosahedral structure of the NFZC0107443 VLPs from EM data as T = 1 particles involve only pentameric but no hexameric interactions. The rest of the VLPs did not produce apparent hexameric or pentameric species, however, AVE018 VLPs appeared to contain another kind of higher molecular weight covalent species putatively corresponding to a disulfide-linked CP dimer.

VLP thermal stability

The thermal stability of a VLP is an important characteristic that positively correlates with the overall robustness of the particle and its performance in downstream applications such as vaccine development or drug delivery. To determine the thermal stability of the novel ssRNA phage VLPs, we subjected the particles to increasing temperatures and visualized their disassembly in native agarose gel electrophoresis. Intact ssRNA phage VLPs migrate as a distinct band in the gel which is detectable both when staining for RNA and for protein, but as the VLPs are heated and gradually disassemble, the VLP band accordingly becomes weaker until it completely disappears. The thermal stability Tm is defined as the lowest tested temperature at which the VLP band can no longer be observed.

The novel VLPs have a broad range of thermal stability (Table 1). While the majority (~ 77%) of the tested VLPs disassembled between 50 °C and 70 °C, a few were extremely unstable and were destroyed even at 35 °C, and some others had an unusually high Tm of up to 95 °C. VLPs of the previously studied ssRNA phages without inter-subunit covalent bonds typically disassemble at 60 to 70 °C [26, 27, 31] while those containing disulfides have a notably higher melting temperature of 75 to 95 °C. Interestingly, none of the newly characterized VLPs with the highest melting temperatures (AVE016, 95 °C, ESE007, 90 °C, Beihai14, 85 °C) have disulfide bonds between the subunits while the five VLPs with experimentally detected inter-subunit disulfides exhibit a relatively modest thermal stability between 55 and 75 °C. These results undermine the prior belief that inter-subunit disulfide bonds are a necessity for robust ssRNA phage VLPs and demonstrate that very stable particles can be built solely by non-covalent interactions. Further investigations are underway to determine the functional basis for the unusual stability of these VLPs.

Potential CP–RNA interactions

In a number of the previously studied ssRNA phages, the coat protein recognizes and binds a genomic RNA hairpin at the beginning of the replicase gene which regulates the synthesis of the replicase enzyme and contributes to specific packaging of phage genome into the virions. The hairpin is a stem-loop structure comprised of an approximately eight base pair-long stem with an unpaired adenosine residue and a three- to six-nucleotide-long loop,

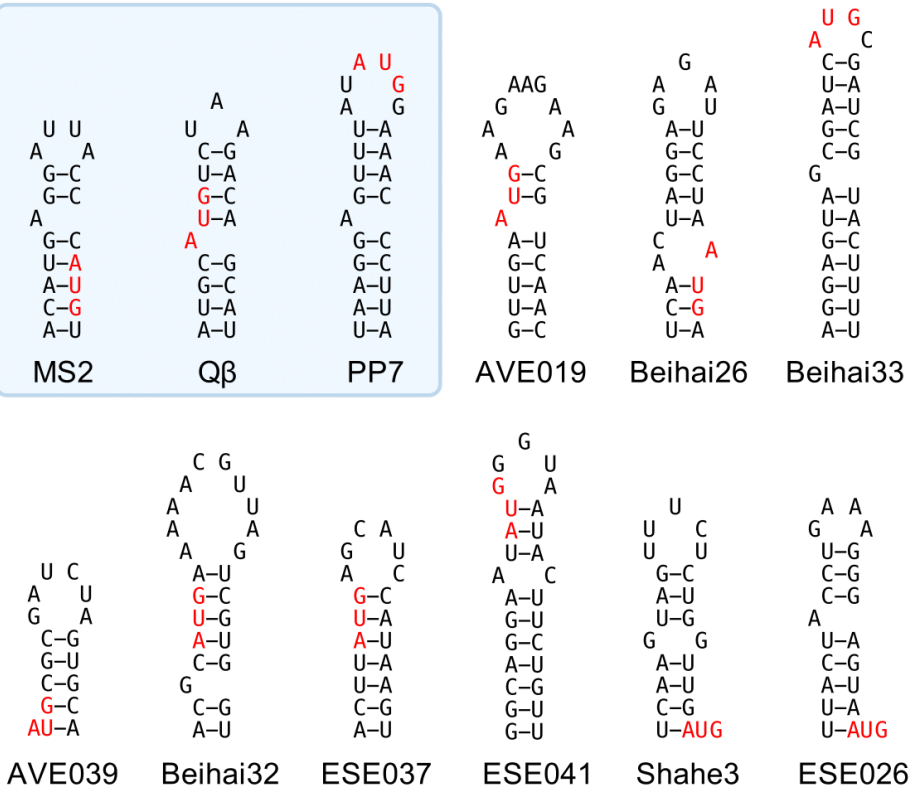

Fig. 4 Putative TR hairpins found in the metagenomic ssRNA phage sequences for some of the CPs used in this study. The known TR hairpins of MS2, Qβ and PP7 phages are shown for reference. The initiation codon of the replicase gene is shown in red

and is often designated the translational operator (TR) of the replicase gene (see [32] for a review). The TR can be appended to an RNA molecule of choice as a tag where it can direct packaging of specific RNA molecules inside VLPs or serve for identifying protein–RNA interactions or tracking of RNA molecules in a living cell (see [10] for a review). Currently two distinct CP RNA binding modes are known for the ssRNA phages, the first shared by the conjugative pili-specific phages MS2 [33], PRR1 [34] and Qβ [35], and the other one found in the *Pseudomonas* phage PP7 [36]. No CP-TR binding has been detected in the more distantly related phages AP205 and Cb5 despite considerable effort, suggesting that the interaction is not universally conserved among the ssRNA phages.

The CP ability to specifically bind RNA is a certain advantage for VLP and other potential applications, therefore for our subset of experimentally available CPs we surveyed the corresponding genome sequences for possible TR hairpins at the beginning of the replicase ORFs. In the majority of cases, a putative stem-loop structure around the replicase initiation codon could indeed be detected. A number of examples are compiled in Fig. 4; all predictions are provided in Additional file 2: Figure S3 and are summarized in Table 1. Within the MS2-like CP supergroup there appears to be a trend that phages with CP sequences relatively more similar to those of MS2, PRR1 or Qβ also contain a TR-resembling hairpin in the genome, while for more distant phages the TR structures look increasingly dubious. A notable exception is a small cluster of Beihai33, Wenling2 and Wenling3 CPs which have very weak similarity to either the Ms2, Qβ or PP7 CPs, yet all of them have a prominent hairpin with a tetranucleotide sequence AUGC in the loop. In addition, despite differences in sequence, the base pairing in the stem has been preserved, suggesting that the hairpins are evolutionary conserved and might function as TRs through a possibly novel RNA binding mechanism to the two already known. No analogous structural conservation is observed among related phages in other CP similarity groups, which renders the function

of the predicted hairpins as TRs somewhat questionable. However, affinities of the predicted TR hairpins for the respective CPs have to be experimentally determined, and discovery of additional protein–RNA interactions is clearly possible.

Discussion

In the current study we have analyzed over 100 novel ssRNA phage CPs with the main objective to find candidates for development of future VLP carriers. A number of properties for the CPs and VLPs are desirable for these purposes: (1) high-level CP expression in bacteria; (2) efficient assembly of the CP into VLPs; (3) high stability of the assembled VLPs; (4) simple and effective means of VLP purification and (5) VLP tolerance for chemical and genetic modification. An ability to package substances of interest inside the VLPs is further preferable. In our study we have addressed a number of these points, and the sample size allows for some general conclusions to be made.

For all of the previously studied ssRNA phages, the CPs could be cloned and expressed in *E. coli* in standard conditions which resulted in highly efficient assembly of VLPs. We adopted a similar strategy for the novel ssRNA phage CPs, and the high-level expression part indeed did not present any problems with only a few of the proteins not being produced. Instead, the solubility of the produced CPs proved to be the first roadblock as in the initial conditions approximately 40% of the tested proteins were found in inclusion bodies. In this respect, it can be noted that the reported sources for the metagenomic datasets are extremely diverse, ranging from intestinal contents of warm-blooded animals to deep sea microbial sediments and arctic soils in Svalbard. It is therefore conceivable that at least for some of the proteins, high-level expression at 37 °C is such a dramatic departure from the conditions in their native host bacteria that it adversely affects their folding, stability or assembly into VLPs. This assumption appeared to be correct as lowering the CP expression temperature indeed largely resolved the solubility issues.

As the study progressed, it was further established that the solubility of the CP does not necessarily lead to VLP formation. Generally, CPs of the MS2-like supergroup readily formed VLPs although there were some exceptions in particular subclusters. CPs of the AVE015-like supergroup were also generally able to assemble into VLPs, although a higher prevalence of aberrant particles was observed in these samples. In contrast, in the Cb5-like supergroup VLP formation was detected for less than a half of the examined CPs. It can be noted that the Cb5 VLPs are very sensitive to salt [27, 37] which might translate also to related CPs, and it is possible that the presence of salt in buffer solutions or during the EM staining procedure might have triggered VLP disassembly. While the issue could perhaps be alleviated by taking extra care not to expose these VLPs to salt at any point, such experiments were not attempted as particles this unstable would not be of much interest for subsequent biotechnological developments. From the smaller CP similarity clusters, all expressed CPs from the AVE002-like, ESE017-like, ESE020-like and AC-like CP groups also formed VLPs. From the remaining CPs, only those from the, Beihai14, ESE001 and ESO003 phages were able to assemble into particles.

The rather high proportion of assembly-deficient CPs was somewhat unexpected from our prior experience and, besides the possible VLP stability issues discussed above, likely has several additional reasons. The metagenomic data vary significantly in quality and in some cases the failure of a CP to be expressed or to assemble into VLPs might result from an incorrect sequence caused by sequencing or sequence assembly errors. In other cases, the high-level heterologous expression in *Escherichia coli* might cause issues for CPs from phages with markedly different original hosts that might be more complex than growth temperature alone. In most of the cases, however, the failure to form VLPs is presumably caused by the absence of other phage components during the assembly process. The assumption that the presence of unspecific RNA is sufficient to promote particle assembly has been built on a small subset of previously studied coat proteins, and there is no particular reason to expect similar properties for all ssRNA phages. Contrary to the large DNA phages, the ssRNA phages do not package their genome into a preformed empty capsid but instead the CP subunits condense around the genomic RNA molecule to form an enclosing protein shell. In the process, the phage maturation protein specifically binds both the genome and the coat protein, and the genomic RNA itself has a highly complex three-dimensional shape that is thought to actively promote its encapsidation [38–41]. Considering that the biological function of the CP is to build virions and not VLPs, it is conceivable that some phages might have evolved to rely on the maturation protein and/or the full-length genome for assembly more than others. This would in turn manifest in the observed incapability of the CP to assemble into VLPs when expressed separately from the other phage components. Also, in other RNA viruses, specific RNA packaging signals have been described (see [42] and references therein), and it cannot be excluded that in some ssRNA phage genomes yet unidentified RNA structures exist that are crucial for assembly. Experimental verification of such possibilities is however difficult or impossible in absence of the actual phage that can be studied in the laboratory.

The purification of the novel VLPs was generally straightforward and in most cases, a previously

established two-step VLP purification procedure using gel filtration and ion exchange chromatography was able to yield an at least 90% pure preparation. In some cases, however, only about 50% pure material was obtained, presumably due to low VLP stability and/or co-aggregation with bacterial proteins. Still, for the great majority of the novel VLPs, purification does not pose any problems and is suitable for biotechnological processes.

To further characterize the VLPs and enable rational structure-guided modification of the capsid, efforts are currently underway to determine their high-resolution three-dimensional structures by X-ray crystallography. The tolerance for foreign antigens by chemical and genetic modification is also being tested in our laboratory for a number of VLPs, and preliminary data indicate that several could be of comparable or superior performance to the Ms2, Qβ and AP205 VLPs (to be published).

Conclusions

In this study, we have demonstrated that environmental viral sequences uncovered in metagenomic studies can be useful not only for comprehending the diversity of viruses in nature but can also be successfully utilized to reconstruct virus-like particles in a laboratory setting. In this way we have for the first time experimentally characterized 11 new ssRNA phage coat protein types and their ability to assemble into VLPs. The 80 novel ssRNA bacteriophage VLPs that we have obtained and characterized here will be important for development of new vaccines and related applications using the ssRNA phage VLP platform. The results also provide a rich ground for further fundamental studies of ssRNA bacteriophage biology such as their structure and protein–RNA interactions.

Methods

CP similarity analysis and clustering

The metagenomic ssRNA phage genome sequences were fetched from GenBank using accession numbers reported in [18–20] or sourced from supplementary data from [19]. An additional search for new ssRNA phage sequences was performed in January 2018 by querying the NCBI's nucleotide (nt) and environmental nucleotide (env_nt) sequence databases with all available ssRNA phage protein sequences using the tblastn program from the BLAST + package [43]. The additional ssRNA phage protein sequences extracted from the hits were iteratively used in repeated queries until no new sequences were detected. A total of 31 additional CP sequences were recovered.

All available ssRNA phage CP sequences were used to generate a BLAST database against which each sequence was individually queried using the blastp program. The results in CSV format were imported into a Google Sheets document for calculation of BLAST bit score ratios (BSRs; the BLAST bit score of the hit divided by the bit score of the query sequence matched against itself) and creation of a distance matrix using values of 1-BSR as the distance measure. The matrix was used to cluster the sequences with the UPGMA algorithm using the program neighbor from the Phylip package [44]. Figtree v 1.4.3. [45] was used for visualization of the resulting tree. The data from the clustering analysis together with the BSR values were used for generating a heat map of CP variation in Google Sheets using Google Apps Script scripting facilities.

CP expression

The CP-encoding sequences were synthesized by General Biosystems and provided by the manufacturer cloned in pET24a (Novagen) as expression-ready constructs.

For small-scale expression and solubility analysis, *E. coli* BL21(DE3) cells were transformed with CP-encoding plasmids, individual colonies were inoculated in 5 ml of LB media supplemented with 30 μg/ml kanamycin and incubated at 37 °C overnight without shaking. The overnight cultures were transferred into 50 ml of 2xTY medium and the cells were grown at 37 °C or 15 °C with aeration until OD_{600} reached 0.6 to 0.8. IPTG was then added to a final concentration of 1 mM and the cultures were incubated for additional 4 h at 37 °C or 20 h at 15 °C, after which aliquots were harvested by centrifugation for assessment of expression level and solubility by SDS-PAGE. Large-scale production for VLP purification purposes followed the same protocol using 2 l of 2xTY medium. To determine the solubility of the produced CPs, the aliquoted cells were suspended in lysis buffer (50 mM tris–HCl pH 8.0, 150 mM NaCl, 0.1% Triton X100, 1 mM PMSF) in a wet cell weight/buffer volume ratio of 1:4, lysed by sonication, centrifuged for 30 min at 13000 g and the supernatant and pellet analyzed in SDS-PAGE. The same protocol was used for preparation of lysates for VLP purification.

VLP purification

For small-scale purification, 1 ml of clarified bacterial lysate was loaded onto a 12 ml, 6.6 × 400 mm Sepharose 4 FF column (GE Healthcare) equilibrated with PBS. Chromatography was done on an Acta Prime Plus system (GE Healthcare) with the flow rate set to 0.3 ml/min and fraction size to 1 ml. VLP-containing fractions were detected in SDS-PAGE and those of the highest purity were pooled and applied to a 0.7 ml, 6.6 × 50 mm Fractogel DEAE (M) ion exchange column (GE Healthcare). The flow-through was collected and the column further washed with 2 ml of PBS. Column-bound

proteins were eluted with a linear 10 column volume gradient to PBS containing 1 M NaCl using a flow rate of 1 ml/min and fraction size of 1 ml on an Akta Pure 25 system (GE Healthcare). The VLPs were usually found in most of the fractions while the contaminating proteins only in some. The fractions of the highest purity were pooled, dialyzed against 20 mM tris–HCl pH 8.0, supplemented with glycerol to a final concentration of 50% and stored at − 20 °C for downstream experiments. The purification protocol was accordingly upscaled if a larger quantity of VLPs was required.

Electron microscopy

For transmission electron microscopy, VLP samples after gel-filtration were adsorbed on carbon-Formvar-coated copper grids and negatively stained with a 1% aqueous solution of uranyl acetate. The grids were examined in a JEM-1230 electron microscope (JEOL Ltd., Tokyo, Japan) operated at 100 kV. Electron micrographs were recorded with iTEM software (version 3.2, Soft Imaging System GmbH) using a side-mounted Morada digital camera (Olympus-Soft Imaging System GmbH, Munster, Germany).

Dynamic light scattering

DLS measurements were performed using Malvern Zetasizer Nano ZS and quartz cuvette ZEN2112, 173 Backscatter according to manufacturer's instructions.

Detection of disulfides in VLPs

Aliquots of purified cysteine-containing VLPs in storage buffer (20 mM Tris–HCl pH 8.0, 50% glycerol) were mixed with an equal volume of Laemmli buffer (0.125 M Tris–HCl pH 6.8, 20% glycerol) with or without added 5% 2-mercaptoethanol. The samples were heated for 10 min at 95 °C and run on a 15% polyacrylamide gel using a standard Tris–glycine SDS electrophoresis system.

Determination of VLP thermal stability

Assessment of thermal stability was done essentially as described before [27]. VLP samples at a concentration of 1 mg/ml in 20 mM tris–HCl, pH 8.0 were heated for 15 min in a Veriti thermal cycler (Applied Biosystems) in a 5 °C-increment step gradient and then loaded on a 1% agarose gel. After electrophoresis in TAE buffer, the RNA was visualized with ethidium bromide and protein with Coomassie blue.

Prediction of TR hairpins

The starting positions of replicase ORFs in the metagenomic ssRNA phage genome sequences were located based either on the available genome annotations or found by examining the sequences manually. A region flanking 20 nucleotides in both directions from the first nucleotide of the replicase ORF was extracted from each genome and the sequences were input to rnafold using the default parameters.

Additional files

Additional file 1: Table S1. Protein sequences used in the study. Accession numbers are provided for the respective genome sequences. Entries without an accession number were sourced from Dataset S1 from Krishnamurthy et al. [19]. In cases where more than one open reading frame was found between the maturation and replicase genes, the coat protein corresponds to ORF2.

Additional file 2: Figure S1. SDS-PAGE analysis of the production level of the metagenomic ssRNA phage CPs. The samples represent total cellular protein content four hours after the CP expression was induced at 37 °C. 1 and M2—protein molecular weight markers (M1: bands of 10, 15, 25, 35, 40, 55, 70, 100, 130 and 180 kDa; M2: bands of 14, 18, 25, 35, 45, 66 and 116 kDa). **Figure S2.** Solubility of the metagenomic ssRNA phage CPs at 37°C and 15°C. The samples represent the soluble (s) and insoluble (d) cellular protein fractions after the CPs were produced at the indicated temperature. The proteins were produced at 15 °C only if they were completely insoluble at 37 °C. M2—lanes with MW marker—14, 18, 25, 35, 45, 66 and 116 kDa. M1 and M2—protein molecular weight markers (M1: bands of 10, 15, 25, 35, 40, 55, 70, 100, 130 and 180 kDa; M2: bands of 14, 18, 25, 35, 45, 66 and 116 kDa). **Figure S3.** Predicted RNA hairpin structures at the beginning of the replicase gene in the metagenomic ssRNA phage genome sequences. A region flanking 20 nucleotides in each direction from the first nucleotide of the replicase gene was used for the prediction.

Authors' contributions

IL supervised the study, performed VLP expression and solubility analysis, purified VLPs, analyzed disulfide bonds in VLPs and wrote the manuscript; GK participated in VLP purification; IA performed CP expression in bacteria; JB purified VLPs and performed DLS experiments; JJ did electron microscopy; MS performed thermal stability experiments; JR designed the study, extracted CP sequences, designed expression constructs, performed CP similarity analysis and clustering, did RNA structure prediction and wrote the manuscript; KT designed and supervised the study and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by Grant 1.1.1.1/16/A/104 from the European Research and Development Fund.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 21 January 2019 Accepted: 4 May 2019
Published online: 13 May 2019

References

1. Kastelein RA, Berkhout B, Overbeek GP, van Duin J. Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein. Gene. 1983;23:245–54.
2. Peabody DS. Translational repression by bacteriophage MS2 coat protein expressed from a plasmid. A system for genetic analysis of a protein–RNA interaction. J Biol Chem. 1990;265:5684–9.
3. Kozlovskaya TM, Pumpen PP, Dreilina DE, Tsimanis AJ, Ose VP, Tsibinogin VV, Gren EJ. Formation of capsid-like structures as a result of expression of coat protein gene of RNA phage fr. Dokl Akad Nauk SSSR. 1986;287:452–5.
4. Kozlovska TM, Cielens I, Dreilinna D, Dislers A, Baumanis V, Ose V, Pumpens P. Recombinant RNA phage Q beta capsid particles synthesized and self-assembled in *Escherichia coli*. Gene. 1993;137:133–7.
5. Tissot AC, Maurer P, Nussberger J, Sabat R, Pfister T, Ignatenko S, Volk HD, Stocker H, Muller P, Jennings GT, et al. Effect of immunisation against angiotensin II with CYT006-AngQb on ambulatory blood pressure: a double-blind, randomised, placebo-controlled phase IIa study. Lancet. 2008;371:821–7.
6. Beeh KM, Kanniess F, Wagner F, Schilder C, Naudts I, Hammann-Haenni A, Willers J, Stocker H, Mueller P, Bachmann MF, Renner WA. The novel TLR-9 agonist QbG10 shows clinical efficacy in persistent allergic asthma. J Allergy Clin Immunol. 2013;131:866–74.
7. Cornuz J, Zwahlen S, Jungi WF, Osterwalder J, Klingler K, van Melle G, Bangala Y, Guessous I, Muller P, Willers J, et al. A vaccine against nicotine for smoking cessation: a randomized controlled trial. PLoS ONE. 2008;3:e2547.
8. Tumban E, Muttil P, Escobar CA, Peabody J, Wafula D, Peabody DS, Chackerian B. Preclinical refinements of a broadly protective VLP-based HPV vaccine targeting the minor capsid protein, L2. Vaccine. 2015;33:3346–53.
9. Spohn G, Jennings GT, Martina BE, Keller I, Beck M, Pumpens P, Osterhaus AD, Bachmann MF. A VLP-based vaccine targeting domain III of the West Nile virus E protein protects from lethal infection in mice. Virol J. 2010;7:146.
10. Pumpens P, Renhofa R, Dishlers A, Kozlovska T, Ose V, Pushko P, Tars K, Grens E, Bachmann MF. The True story and advantages of RNA phage capsids as nanotools. Intervirology. 2016;59:74–110.
11. Rohovie MJ, Nagasawa M, Swartz JR. Virus-like particles: next-generation nanoparticles for targeted therapeutic delivery. Bioeng Transl Med. 2017;2:43–57.
12. Peabody DS, Manifold-Wheeler B, Medford A, Jordan SK. do Carmo Caldeira J, Chackerian B: immunogenic display of diverse peptides on virus-like particles of RNA phage MS2. J Mol Biol. 2008;380:252–63.
13. Bardwell VJ, Wickens M. Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. Nucleic Acids Res. 1990;18:6587–94.
14. Tsai BP, Wang X, Huang L, Waterman ML. Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. Mol Cell Proteomics. 2011;10(M110):007385.
15. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. Mol Cell. 1998;2:437–45.
16. Klovins J, Overbeek GP, van den Worm SH, Ackermann HW, van Duin J. Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. J Gen Virol. 2002;83:1523–33.
17. Kazaks A, Voronkova T, Rumnieks J, Dishlers A, Tars K. Genome structure of *Caulobacter* phage phiCb5. J Virol. 2011;85:4628–31.
18. Greninger AL, DeRisi JL. Draft genome sequences of Leviviridae RNA phages EC and MB recovered from San Francisco wastewater. Genome Announc. 2015;3:e00652-15.
19. Krishnamurthy SR, Janowski AB, Zhao G, Barouch D, Wang D. Hyperexpansion of RNA bacteriophage diversity. PLoS Biol. 2016;14:e1002409.
20. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, et al. Redefining the invertebrate RNA virosphere. Nature. 2016;540:539.
21. Rumnieks J, Tars K. Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. BMC Microbiol. 2012;12:277.
22. Kannoly S, Shao Y, Wang IN. Rethinking the evolution of single-stranded RNA (ssRNA) bacteriophages based on genomic sequences and characterizations of two R-plasmid-dependent ssRNA phages, C-1 and Hgal1. J Bacteriol. 2012;194:5073–9.
23. Ruokoranta TM, Grahn AM, Ravantti JJ, Poranen MM, Bamford DH. Complete genome sequence of the broad host range single-stranded RNA phage PRR1 places it in the *Levivirus* genus with characteristics shared with Allolevivirus. J Virol. 2006;80:9326–30.
24. Min Jou W, Haegeman G, Ysebaert M, Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature. 1972;237:82–8.
25. Olsthoorn RCL, Garde G, Dayhuff T, Atkins JF, van Duin J. Nucleotide sequences of a single-stranded RNA phage from *Pseudomonas aeruginosa*: kinship to coliphages and conservation of regulatory RNA structures. Virology. 1995;206:611–25.
26. Persson M, Tars K, Liljas L. The capsid of the small RNA phage PRR1 is stabilized by metal ions. J Mol Biol. 2008;383:914–22.
27. Plevka P, Kazaks A, Voronkova T, Kotelovica S, Dishlers A, Liljas L, Tars K. The structure of bacteriophage phiCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. J Mol Biol. 2009;391:635–47.
28. Golmohammadi R, Fridborg K, Bundule M, Valegård K, Liljas L. The crystal structure of bacteriophage Qb at 3.5 Å resolution. Structure. 1996;4:543–54.
29. Tars K, Fridborg K, Bundule M, Liljas L. Crystal structure of phage PP7 from *Pseudomonas aeruginosa* at 3.7 Å resolution. Virology. 2000;272:331–7.
30. Shishovs M, Rumnieks J, Diebolder C, Jaudzems K, Andreas LB, Stanek J, Kazaks A, Kotelovica S, Akopjana I, Pintacuda G, et al. Structure of AP205 coat protein reveals circular permutation in ssRNA bacteriophages. J Mol Biol. 2016;428:4267–79.
31. Axblom C, Tars K, Fridborg K, Orna L, Bundule M, Liljas L. Structure of phage fr capsids with a deletion in the FG loop: implications for viral assembly. Virology. 1998;249:80–8.
32. Rumnieks J, Tars K. Protein–RNA interactions in the single-stranded RNA bacteriophages. Subcell Biochem. 2018;88:281–303.
33. Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature. 1994;371:623–6.
34. Persson M, Tars K, Liljas L. PRR1 coat protein binding to its RNA translational operator. Acta Crystallogr D Biol Crystallogr. 2013;69:367–72.
35. Rumnieks J, Tars K. Crystal structure of the bacteriophage qbeta coat protein in complex with the RNA operator of the replicase gene. J Mol Biol. 2014;426:1039–49.
36. Chao JA, Patskovsky Y, Almo SC, Singer RH. Structural basis for the coevolution of a viral RNA-protein complex. Nat Struct Mol Biol. 2008;15:103–5.
37. Bendis I, Shapiro L. Properties of *Caulobacter* ribonucleic acid bacteriophage phi Cb5. J Virol. 1970;6:847–54.
38. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ. Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. Nat Commun. 2016;7:12524.
39. Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature. 2017;541:112–6.
40. Gorzelnik KV, Cui Z, Reed CA, Jakana J, Young R, Zhang J. Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qbeta reveals a single maturation protein and the genomic ssRNA in situ. Proc Natl Acad Sci U S A. 2016;113:11519–24.
41. Rumnieks J, Tars K. Crystal structure of the maturation protein from bacteriophage Qbeta. J Mol Biol. 2017;429:688–96.
42. Twarock R, Bingham RJ, Dykeman EC, Stockley PG. A modelling paradigm for RNA virus assembly. Curr Opin Virol. 2018;31:74–81.
43. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
44. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6.: Department of Genome Sciences, University of Washington, Seattle.; 2005.
45. Rambaut A. FigTree, version 1.4.3. http://tree.bio.ed.ac.uk/software/figtree. 2009.

Publisher's Note

7.2. "Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes"

During our previous findings we had established a collection of more than 80 different VLPs, derived from hypothetic ssRNA CPs. Most of them can be obtained in sufficient quantity and purity for crystallographic studies. Laboratory co-workers performed crystallization trials of our entire ~80 VLP library and were able to obtain crystals for more than half of the VLPs. In 22 cases the obtained crystals diffracted to high resolution and the corresponding structures were solved by Dr. Jānis Rūmnieks.

7.2.1. Capsid shape and size

When comparing the length of the CPs with the size of the VLPs, no strong correlation was observed. However, the longest CP Beihai14 also produced the largest VLPs and shortest Wenzhou1 protein produced the smallest VLPs.

Like the previously studied capsids of ssRNA phages, the majority of the novel VLPs had a morphologically spherical shape, and only the AC, AVE015, GQ-112, and Beihai14 capsids had a pronounced polyhedral appearance.

In the previous study we found several CPs which form VLPs of T = 1 symmetry. In nature, the existence of functional phage with such small particles is not possible, because there would be not enough room for 3-4 kb long viral genome. Formation of small VLPs can be assumed to be artifacts of the recombinant expression system. Possibly, in these cases formation of larger particles is mediated by nucleic acid or A protein.

During previous examination of the particles by electron microscopy, several CPs were found to form two types of particles having spherical and elongated shapes. Crystals of AVE016 VLPs turned out to contain prolate particles 28.9 nm in width and 34.6 nm in length. Geometrically, these particles correspond to a prolate icosahedron of a $T = 3$, $Q = 4$ architecture and consist of 210 CP monomers or 105 CP dimers. AVE016 particle is composed of 21 distinct CP monomers in eight major conformations with their structural differences almost entirely limited to the EF and FG loops. Compared to other phages, AVE016 has a rather long genome, therefore it is possible that the prolate shape is not a mere artefact but adaptation of phage to accommodate longer RNA.

7.2.2. Intersubunit interactions

During crystallographic studies, metal ions were observed in the AVE015, AVE019, Beihai19, Beihai32, ESE007, ESE021, and ESE058 VLPs. Therefore, capsids were tested for the impact of the metal ions to VLP stability by measuring their thermal denaturation in the presence of EDTA. With the exception of the AVE015 VLPs for which the results were somewhat obscure, the addition of chelating agent indeed reduced thermal stability by at least 5 °C.

7.2.3. Potential carriers

The widely represented MS2-like group (Beihai21, ESE007, ESE058, NT-214, and AVE019) revealed structures similar to the phages of this group previously studied, and it can be assumed that these VLPs would also possess similar capacity for genetical modification.

Two of the determined structures, PQ-465 and ESE001 were previously described as AP-205-like phages. PQ-465 had no significant differences in the structure of CP compared to AP205. The ESE001 CP is much shorter (118 aa) and its sequence could not be safely reconciled with other CPs. The three-dimensional structure of the ESE001 VLP also revealed a CP fold with AP205-like features, but the αB helix was completely absent. The solved structures point out the potential of these particles as carriers.

STRUCTURAL BIOLOGY

Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes

Jānis Rūmnieks, Ilva Liekniņa, Gints Kalniņš, Mihails Šišovs, Ināra Akopjana, Jānis Bogans, Kaspars Tārs*

The single-stranded RNA (ssRNA) bacteriophages are among the simplest known viruses with small genomes and exceptionally high mutation rates. The number of ssRNA phage isolates has remained very low, but recent metagenomic studies have uncovered an immense variety of distinct uncultured ssRNA phages. The coat proteins (CPs) in these genomes are particularly diverse, with notable variation in length and often no recognizable similarity to previously known viruses. We recombinantly expressed metagenome-derived ssRNA phage CPs to produce virus-like particles and determined the three-dimensional structure of 22 previously uncharacterized ssRNA phage capsids covering nine distinct CP types. The structures revealed substantial deviations from the previously known ssRNA phage CP fold, uncovered an unusual prolate particle shape, and revealed a previously unseen dsRNA binding mode. These data expand our knowledge of the evolution of viral structural proteins and are of relevance for applications such as ssRNA phage–based vaccine design.

INTRODUCTION

The single-stranded RNA (ssRNA) bacteriophages are a group of small bacterial viruses known to infect different Proteobacteria. The ssRNA phages have short 3.5- to 4.5-kb positive-sense RNA genomes that encode only three common proteins: the maturation protein (MP), a minor structural protein that functions to adsorb the phage to the bacterial receptor and deliver the genome into the host cell; the coat protein (CP), the major structural component of the capsid; and the catalytic subunit of an RNA-dependent RNA polymerase (RdRp), necessary for replicating the viral genome (*1*). The ssRNA phages are believed to represent the oldest extant RNA virus lineage (*2*), and only their RdRp has identifiable homologs in other viruses, while the MP and CP are unique to the ssRNA phages.

The ssRNA phages have played an important role in studies of genome structure and replication, translational control mechanisms, protein-RNA interactions, and other fundamental problems in biology, and they have been convenient models for investigating virus structure, assembly, and evolution. The ssRNA phage MS2 was among the first viruses with a determined atomic resolution capsid structure (*3*) and with the entire virion, including the packaged genome in a well-defined conformation, reconstructed at near-atomic resolution (*4*). The ssRNA phage CPs adopt a fold not found in any other viruses, with an N-terminal β hairpin, a central five-stranded antiparallel β sheet, and two C-terminal α helices, and they are further unusual in that two CP molecules very tightly interact to form dimers with a common hydrophobic core. Consequently, the de facto subunits of ssRNA phage particles are CP dimers, composed of a single 10-stranded β sheet lining the interior of the particle and the α helices and β hairpins exposed to the capsid exterior. The complete virion is built of 89 CP dimers and a single copy of a genome-bound MP in place of another CP dimer. The assembled particle is approximately 28 nm in diameter and, disregarding the slight irregularity introduced by the MP, follows $T = 3$ quasi-equivalent icosahedral symmetry.

To date, high-resolution structures of eight ssRNA phages infecting *Escherichia* (*3*, *5*–*7*), *Pseudomonas* (*8*, *9*), *Caulobacter* (*10*), and *Acinetobacter* (*11*) have been determined, which have shown that the CP fold is conserved despite high sequence variability. Thus far, the only notable variation in CP structure has been observed in the *Acinetobacter* phage AP205, in which the N-terminal strand of the β-hairpin has been translocated to the C terminus; however, the close vicinity of monomer N and C termini in the dimer results in the formation of a two-stranded β structure at a position analogous to the β-hairpin in the other phages. Variations in the CP fold are expected given the exceptionally high mutation rates of RNA viruses, but for reasons not entirely clear, the number and diversity of isolated ssRNA phages have remained very low, which has, in turn, limited continued structural studies on these viruses. There appears to be no reason to presume a genuine scarcity of ssRNA bacteriophages in nature, and the extent to which their structural diversity has been sampled and explored remains largely unknown.

The ever-increasing metagenomic sequencing efforts within the last decade have documented an expanse of highly diverse uncultured microbial and viral life forms in every environment that has been examined. In 2016, two studies uncovered more than 200 novel ssRNA phage sequences in various RNA metagenomes (*12*, *13*), and very recently, thousands of additional ssRNA phage genomes have been found (*14*, *15*), demonstrating that the true ubiquity and diversity of these viruses in nature have been greatly underestimated and underexplored. Still, beyond comparative analyses with experimentally well-characterized species, the metagenomic studies are limited in their capacity for providing deeper insight into the biology of the newly found life forms, and in the case of highly diverged bacteriophage genomes, often provide little or no information about the structure, host bacteria, lysis strategies, and other essential characteristics of these viruses. For the previously studied ssRNA phages, expression of a cloned CP gene results in assembly of virus-like particles (VLPs), which lack the MP and package random

Latvian Biomedical Research and Study Center, Rātsupītes 1, LV1067, Riga, Latvia.
*Corresponding author. Email: kaspars@biomed.lu.lv

bacterial RNA instead of the viral genome, but that are otherwise morphologically indistinguishable from native virions. In context of metagenome data, this gives an opportunity to obtain previously unexplored CP-encoding sequences via chemical gene synthesis and recombinantly express them in bacteria to resurrect phage-like particles in the laboratory. Following this approach, we previously acquired more than 100 novel ssRNA phage CP sequences, which in many cases had no recognizable similarity to the previously known phages or to each other, and we were able to obtain 80 different VLPs of 11 distinct CP types, 8 of which were experimentally characterized (*16*).

With CP sequences diverged beyond recognition, determination of their three-dimensional structure is instrumental in uncovering distant relationships among the ssRNA phages and can offer valuable clues for better understanding the evolution of these viruses in general. In addition, the ssRNA phage VLPs have found applications as carriers for foreign antigens in vaccine development (*17*), for which detailed knowledge of the location of CP terminal and loop regions is essential to enable structure-guided design of new medications. To address these questions, we have determined crystal structures of 22 metagenome-derived ssRNA phage VLPs, which have revealed substantial deviations from the previously known CP fold and uncovered previously unseen ssRNA phage particle shapes.

RESULTS AND DISCUSSION

Structure determination

Our laboratory currently holds a collection of more than 80 different metagenome-derived ssRNA bacteriophage VLPs, nearly all of which could be produced in sufficient quantity, homogeneity, and purity for crystallographic studies. We thereby set out to perform crystallization trials of our entire ~80 VLP library and were able to obtain crystals for more than half of the VLPs. Diffraction data to a resolution sufficient for model building (<4 Å) were obtained for 22 VLPs covering nine distinct CP types (table S1). The resolution of the determined VLP structures ranges from 2.6 to 4.0 Å, with an average of 3.4 Å, which in all cases allowed us to build all-atom models. In addition, one of the VLPs in our study, EMS014, had disassembled at a particular crystallization condition, and a crystal of CP dimers had grown instead, which in this case allowed for a further 1.25-Å subunit structure to be solved. The majority of the models could be built for all residues except for the AC, Beihai14, and PQ-465 CPs, which had unstructured termini, and the AVE019, Beihai19, ESE014, ESE020, and ESE058 CPs, in which certain flexible loops were present. X-ray data collection, reduction and refinement statistics, and model quality indicators are presented in table S2.

CP fold

Within our sample set of 22 VLPs, only in 10 cases was the CP structure consistent with the canonical MS2 CP fold with the N-terminal hairpin, five-stranded β sheet, and two C-terminal α helices, while notable deviations are observed for the other CPs (Fig. 1). Only the four central β strands and a single C-terminal α helix are retained in all structures, while the terminal regions, particularly the N termini, are variable between different CP types. Pairwise superposition of all available ssRNA phage CP monomers further illustrates that approximately half of the newly determined structures do not have substantial similarity to those of the previously studied phages

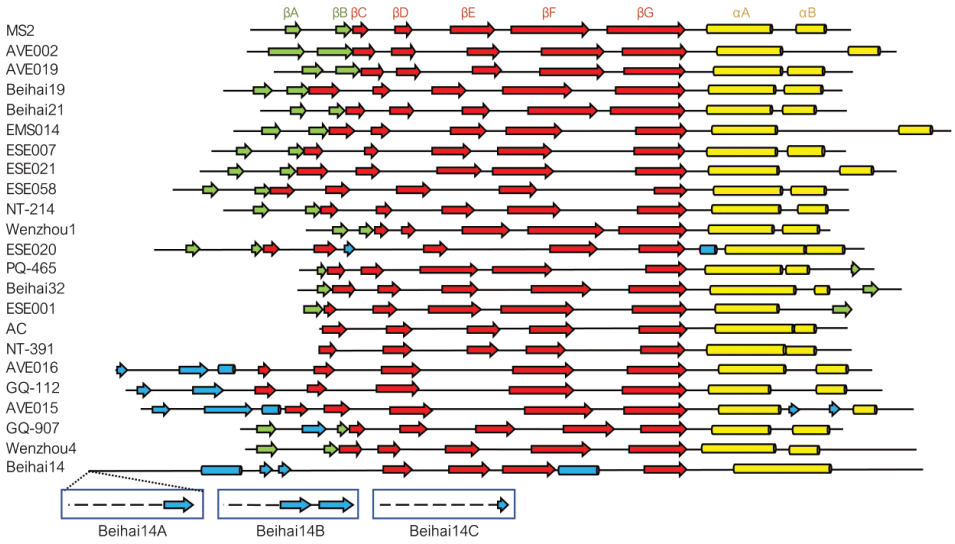

Fig. 1. Coat protein fold of the novel ssRNA bacteriophages. Secondary structure elements are represented as arrows (β strands) and cylinders (α helices). The canonical fold of the bacteriophage MS2 CP with annotated secondary structure elements is included for comparison; for consistency, this nomenclature has been used throughout the text regardless of the actual position of the elements in the particular novel CPs. β strands that constitute the central sheet are shown in red, C-terminal α helices in yellow, β strands that form an N-terminal β-hairpin or an analogous structure in green, and additional secondary structure elements in blue. The different N-terminal structures of Beihai14 quasi-equivalent CP monomers are additionally indicated.

(fig. S1). Structures of the nine distinct CP types reveal both unsuspected similarities in distant ssRNA phage lineages and unexpected differences in close relatives.

MS2-like CPs

The majority of the currently available metagenomic ssRNA phage CP sequences fall into a broad MS2-like supergroup, and also the greatest percentage of the newly solved VLP structures belong to this similarity cluster. Of the newly characterized MS2-like phages, the Beihai19 and Beihai21 CP structure is similar to that of the previously studied phages of this group, such as MS2 and Qβ; the structures from phages ESE007, ESE058, NT-214, and AVE019 are slightly more diverged, yet still very similar. Although the length of these proteins varies from 123 to 146 residues, the size and position of secondary structure elements remain virtually unchanged, and the structural differences almost entirely arise from differently sized loops and their relative packing. The EMS014 and ESE021 CPs are the most diverged MS2-like CPs in our study and also the longest (156 and 150 residues, respectively), which for the most part can be attributed to markedly extended loops between the two C-terminal α helices. Both CPs are relatively similar with a 27% sequence identity, and both have secondary structure elements in essentially the same positions as in the canonical MS2 fold, but in the ESE021 CP, the αB helices have swapped positions between the two monomers (Fig. 2), which has never been observed in ssRNA phages before. Together with the N-terminal β-hairpins, the extensive interhelix loops in the EMS014 and ESE021 CPs completely wrap the helices and almost entirely isolate them from the outside environment.

Cb5-like CPs

The AVE002 and Wenzhou1 CPs in our study are most closely related to a CP type first described in the previously studied bacteriophage Cb5 (*10*). Despite undetectable sequence similarity to the Cb5 CP or to each other, the two proteins do not show significant deviations in their structural organization and follow the canonical ssRNA phage CP fold. The AVE002 CP is somewhat longer (140 residues) than that of Cb5 (122 residues), with the difference mostly attributed to an extensive surface-exposed loop between helices αA and αB, but contrary to the MS2-like EMS014 and ESE021 CPs, the AVE002 interhelix loops bend to the opposite direction over the β-hairpins. The Wenzhou1 CP (Fig. 3) spans only 113 residues, but the overall size of the CP dimer is similar to that of the other Cb5-like proteins. The savings in the Wenzhou1 CP arise from a significantly smaller N-terminal β-hairpin and a very short βD strand, which almost reduces the central β sheet to only three strands. The space vacated by the shrunken βD strand is partially occupied by the C terminus, which as a result becomes buried underneath the capsid surface.

AP205-like CPs

The remaining group of CPs with previously determined related structures, the AP205-like CPs (*11*), is represented by phages PQ-465 and ESE001 in our study. The PQ-465 CP shares 30% sequence identity with that of AP205, and there are no noteworthy differences in the three-dimensional structure of the two proteins. The ESE001 CP (Fig. 3) is notably more distinct and was previously recognized as a separate CP type (*16*) as it could not be reliably aligned to any other CPs. However, the three-dimensional structure of ESE001 VLPs revealed a CP fold with distinctive AP205-like features such as a β-hairpin–like structure formed by the N and C termini, a notable gap between α helices of the two monomers, and stabilizing disulfide bonds between CP subunits. The AP205-like CPs also characteristically have a very short αB helix that lies roughly perpendicular to the αA helix and the β sheet, but in the ESE001 CP, the αB helix is completely eliminated and the protein only has a single C-terminal helix.

Beihai32-like CPs

The remaining VLPs in our study represent six previously uncharacterized ssRNA bacteriophage CP types, and the first of those, putatively named here Beihai32-like CPs, includes proteins from bacteriophages Beihai32 and Wenzhou4, which have lengths of 130 and 146 residues, respectively. Previously, these CPs were included in the MS2-like CP supergroup due to borderline sequence similarity (*16*), but their three-dimensional structure suggests that these proteins should be recognized as a separate type. Structure superposition–based analysis hints at a remote similarity between the Beihai32 CP and AP205-like CPs (fig. S1), and the N- and C-termini of the Beihai32 CP appear to be arranged similarly to the AP205-like proteins (Fig. 3). The Wenzhou4 CP is closely related to the Beihai32 protein as evidenced both by sequence identity and structure superposition (fig. S2) but does not show notable similarity to the AP205-like proteins. Instead, the Wenzhou4 CP has a rather distinct and unique arrangement of its terminal regions, with the N terminus stretching over the αA helices to the other edge of the dimer and the proline-rich C terminus folding into an unusual Λ-shaped structure that extends some 15 Å above the rest of the protein (Fig. 3). While the C termini lack any secondary structure, the N termini form a pair of β-hairpin–resembling structures at positions analogous to the MS2-like phages.

ESE020

The CP of bacteriophage ESE020 belongs to a distinct cluster of approximately 155-residue-long proteins with no detectable sequence similarity to other CPs, and the ESE020 VLP crystal structure reveals notable differences in the CP architecture (Fig. 3). The ESE020 CP dimer is marked by massively elongated loops connecting β strands E and F (the EF loops), which in the particle extend to the

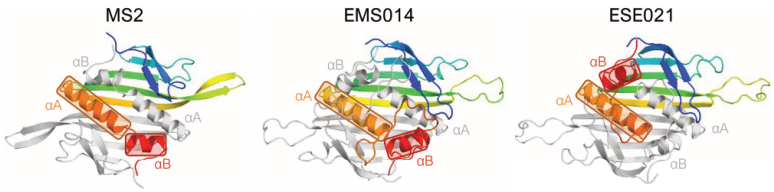

Fig. 2. Swapped C-terminal α helices in the ESE021 coat protein. In the MS2 coat protein dimer, the C-terminal α helices αA and αB of one monomer (rainbow-colored) are positioned roughly end to end to each other in a groove formed by the same helices in the other monomer (light gray). The MS2-like EMS014 and ESE021 coat proteins are closely related, but while the fold of the EMS014 CP closely follows that of MS2, in the ESE021 CP, the αB helices are swapped between the two monomers.

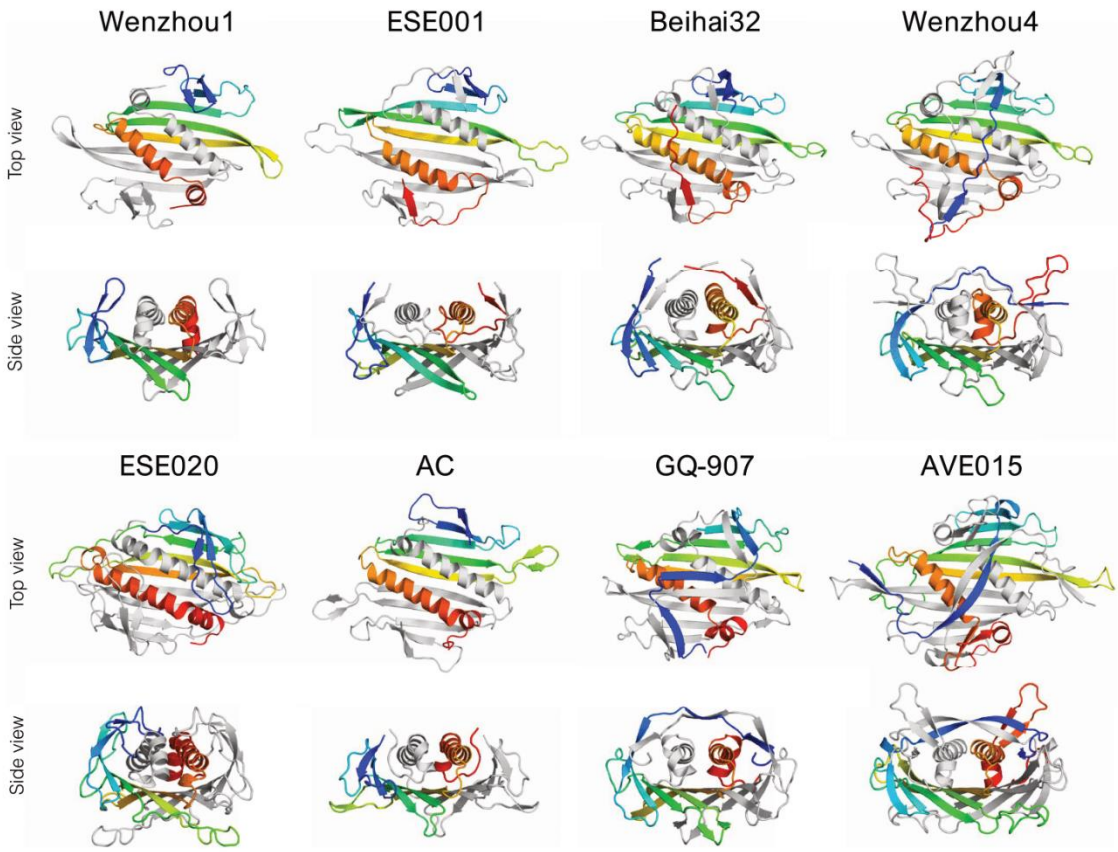

Fig. 3. Three-dimensional structure of coat protein dimers. Proteins representing different variations of the CP fold are shown with one of the monomers rainbow-colored blue (N terminus) to red (C terminus) and the other monomer shown in light gray. All CP dimers are shown as seen from outside of the particle (top view) and in a roughly perpendicular view looking lengthwise down the C-terminal α helices (side view).

icosahedral threefold symmetry axes where they mediate contacts with neighboring CP molecules. Apparently, because of steric restraints, the EF loops are disordered and not visible around the fivefold axes, and here, analogous contacts are accomplished by the shorter loops between β strands F and G (the FG loops). In other ssRNA phages, interactions around both threefold and fivefold symmetry axes are mediated solely by the FG loops, which often adopt different conformations around each in response to the different spatial environments, but the solution to use a different loop in each case is, so far, unique to the ESE020 CP. Among other distinctive features, the ESE020 CP has a short single-turn α helix between βG and αA, which makes it the only currently known ssRNA phage CP with three C-terminal helices. The N-terminal region of the protein is also different from the other CPs and instead of a β-hairpin contains a small three-stranded β sheet in which the extra strand originates from a curled-up extension of the βD strand.

AC-like CPs

Bacteriophage AC was among the first ssRNA bacteriophage genomes found in RNA metagenome data, and we have determined the three-dimensional structure of AC VLPs, as well as that of another AC-like phage NT-391. The CPs within this group are relatively short (115 and 123 residues for AC and NT-391, respectively) and do not show sequence similarity or clear structural relatedness to any other CP type. For the AC CP, secondary structure-based superposition reveals faint similarity to the MS2-like phages (fig. S1), and from the βC strand onward, the fold of the AC-like CPs closely resembles that of MS2 (Figs. 1 and 3). However, the N-terminal β-hairpins are completely missing in the AC-like phages, which is probably the most minimalistic approach we have observed in the current study; the missing hairpin is, however, partially compensated by a rolled-up corner of the central β sheet at the respective position. Also, contrary to most other ssRNA phages, the C termini of AC-like CPs are positioned relatively far from the N termini, are not involved in any interactions, and extend away from the particle.

GQ-907

The CP of the GQ-907 phage belongs to a small group of distinct sequences previously recognized by us as the ESE017-like CP type

(*16*). There is no detectable sequence similarity between these CPs and other CP types, but by structure superposition, the GQ-907 shows weak resemblance to the Beihai32 CP and to some MS2-like CPs (fig. S1). From the βC strand to the C terminus, the GQ-907 CP fold closely resembles that of MS2 (Fig. 1), but the N termini extend over the α helices in a somewhat similar manner to the Wenzhou4 CP (Fig. 3). The GQ-907 N termini are, however, organized in a distinctive zigzag pattern of three consecutive two-stranded β sheets, with the lateral structures located at the same positions as the N-terminal β-hairpins in MS2-like phages and the central one positioned directly above the α helices.

AVE015-like CPs

The metagenomic data revealed a major previously unknown ssRNA phage lineage with relatively large genomes and distinct CPs approximately 165 residues in length, which we previously designated the AVE015-like CPs (*16*). We have determined VLP structures of phages AVE015, AVE016, and GQ-112, and while the three CPs do not share more than 20% sequence identity, their three-dimensional structure is similar. The core fold of these proteins still consists of a five-stranded β sheet and two C-terminal α helices, but their N-terminal region is once again completely reorganized (Fig. 3). In a remotely similar manner to the GQ-907 CP, the N termini of AVE015-like CPs intertwine over the αA helices to form a two-stranded β sheet, but a structure analogous to the N-terminal β-hairpin does not exist in these proteins. The AVE015 and AVE016 proteins contain a short α helix in a position roughly analogous to the βB strand, and in all AVE015-like CPs, the N termini complement another short strand to the central β sheet near the FG loop. Other remarkable features of this CP type include notably long loops between β strands E and F and between the C-terminal α helices; the interhelix loop is particularly extensive in the AVE015 CP and forms a prominent surface-exposed β-hairpin.

Beihai14

By far, the most unusual CP in our study is that of the bacteriophage Beihai14, which spans 208 residues and is currently the longest known ssRNA phage CP. The Beihai14 CP is a singleton with no similar sequences found to date, and its relation to other ssRNA phages remains obscure. The familiar CP structure in the Beihai14 protein is reduced to four central β strands and a single C-terminal α helix, but an additional 10-residue α helix has been inserted into the FG loop, which is the first observation in the ssRNA phages (Fig. 4A). The 85 residues preceding the β sheet almost entirely lack defined secondary structure but, apart from the very N termini, adopt a well-defined conformation. Of these, residues 45 to 80 form a giant loop distantly similar to an oversized β-hairpin of the MS2-like phages, which contains a 10-residue α helix that packs alongside the C-terminal αA helix. The 40 N-terminal residues pass under the VLP surface and extend along the intersubunit interface to the fivefold and threefold icosahedral symmetry axes, where they form five-stranded β-barrels or threefold assemblies of β-hairpin–like structures, respectively (Fig. 4B), whereas the very N termini point toward the center of the particle and are not resolved in the structure due to disorder. The N termini in the Beihai14 CP, in a way, resemble the flexible N-terminal arms in several plant ssRNA viruses, where they also mediate intersubunit contacts and are ordered in some but disordered in other CP monomers (*18*–*20*). Last, the 20 C-terminal residues of the Beihai14 CP are also unusually arranged and constitute a short surface-exposed loop immediately following the αA helix, after which the backbone threads underneath the N-terminal segment and again resurfaces at the icosahedral quasi-threefold symmetry axis.

Capsid shape and size

With two exceptions, the newly characterized ssRNA phage capsids are of $T = 3$ quasi-equivalent icosahedral symmetry and range from 28.3 to 32.5 nm in diameter (table S1). The shortest CP in our study, Wenzhou1, forms the smallest VLPs, and the long AVE015-like and Beihai14 CPs are the largest, but overall, the association between the CP length and the size of the assembled particles is not particularly strong (fig. S3). However, the CP length generally correlates with the length of their FG loops that, in turn, determine pore sizes around the icosahedral threefold and fivefold symmetry axes. The size of these pores show great variation, from none at all in Beihai14 VLPs to approximately 25-Å-wide openings around the icosahedral threefold symmetry axes in GQ-907 VLPs; this raises a question of how well the RNA genome is protected inside these particles, as a molecule of ribonuclease would appear to be able to pass through this pore, but apparently, this does not cause issues for the phage in its natural environment. Like the previously studied ssRNA phages, the majority of the novel VLPs are of a roughly spherical shape, and only the AC, AVE015, GQ-112, and Beihai14 capsids are of notably polyhedral appearance (fig. S4).

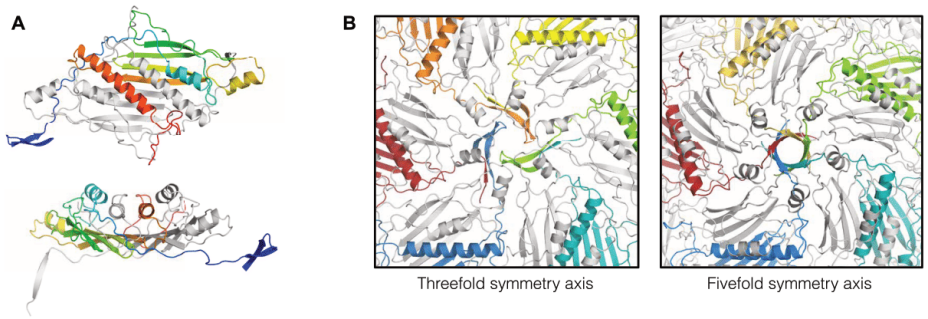

Fig. 4. The unusual coat protein of bacteriophage Beihai14. (**A**) Three-dimensional structure of the Beihai14 CP. The CP dimer is colored and shown in two different orientations as in Fig. 3. (**B**) Interactions around the threefold and fivefold icosahedral symmetry axes in the assembled particle. The CP monomers with their N-terminal extensions involved in the interactions are shown in color.

One of the AC-like CPs in our study, NT-391, is assembled into small ~18.3-nm particles of $T = 1$ symmetry. A $T = 1$ particle has over five times smaller volume than a $T = 3$ capsid, which would leave too little space for packaging the viral genome; hence, the small VLPs can be safely assumed to be artifacts of the recombinant expression system. While uncommon, $T = 1$ particles have been previously observed in preparations of recombinant AP205 and some mutant MS2 VLPs (*21*). In our previous studies, we found that NC-443, another CP of the AC-like group with 24% sequence identity to NT-391, forms $T = 1$ particles as judged by electron microscopy (*16*); at the same time, the AC CP, which is more similar to the NT-391 CP (35% sequence identity), formed normal $T = 3$ capsids. It can be noted that both the NT-391 and NC-443, but not AC, CPs make use of intersubunit disulfide bonds, and it cannot be excluded that in the absence of the viral genome, the formation of covalent bonds between CP dimers is somehow involved in triggering assembly of the smaller particles.

Electron microscopy of AVE016 and, to some extent, several other AVE015-like VLPs revealed a mixture of spherical and elongated particles; the elongated particles, however, appeared to be a minority fraction in all cases. While the AVE015 and GQ-112 VLP crystal structures revealed normal $T = 3$ capsids, unexpectedly, the AVE016 VLP crystals turned out to correspond to prolate particles 28.9 nm in width and 34.6 nm in length. Geometrically, these particles correspond to a fivefold prolate icosahedron of a $T = 3$, $Q = 4$ architecture (*22*) and consist of 210 CP monomers or 105 CP dimers. In a $T = 3$ particle, the CP molecules are present as three slightly different conformers, whereas the asymmetric unit of the $T = 3$, $Q = 4$ AVE016 particle is composed of 21 CP monomers in eight major conformations with their structural differences almost entirely limited to the EF and FG loops (Fig. 5). There have been previous reports of elongated rod-like structures assembled from recombinant ssRNA phage CPs (*23*, *24*), and it cannot be excluded that the prolate AVE016 VLPs likewise are artifacts of the recombinant expression system; however, such homogeneous and well-defined aberrant structures are uncommon and might indicate biological relevance. An AVE016-like phage will have to be isolated and examined in the laboratory to determine the natural virion morphology, but it can be noted that a number of the large tailed double-stranded DNA phages have prolate icosahedral heads (*25*), and several plant viruses with multipartite genomes form bacilliform particles of variable lengths that package genome segments of different sizes (*26*). The AVE015-like viruses appear to have large genomes that might approach 5 kb in some cases, and emergence of elongated capsids would be consistent with a requirement for increased genome packaging capacity. It could then be speculated that assembly of the elongated recombinant AVE016 VLPs is triggered by a subpopulation of longer bacterial RNA molecules that do not optimally fit into $T = 3$ capsids.

Intersubunit interactions

Contacts between CP dimers within the assembled particles include a wide variety of hydrophobic, stacking, polar, and electrostatic interactions, but the number and type of contacts significantly vary among different VLPs. The solvent-buried surface areas of CP dimers within the capsid differ several fold, with the smallest observed in the AC-like and Beihai32-like VLPs and the largest in capsids formed by the Beihai14 and some of the longer MS2-like CPs (table S1). As with the particle size, there is some, but not a very significant, correlation between the CP length and the surface area they bury in the capsid. Somewhat more unexpectedly, there is also not a very strong relation between subunit interface areas and the previously determined thermal stability of the VLPs (*16*), suggesting that monomer-monomer interactions, additional RNA-mediated contacts,

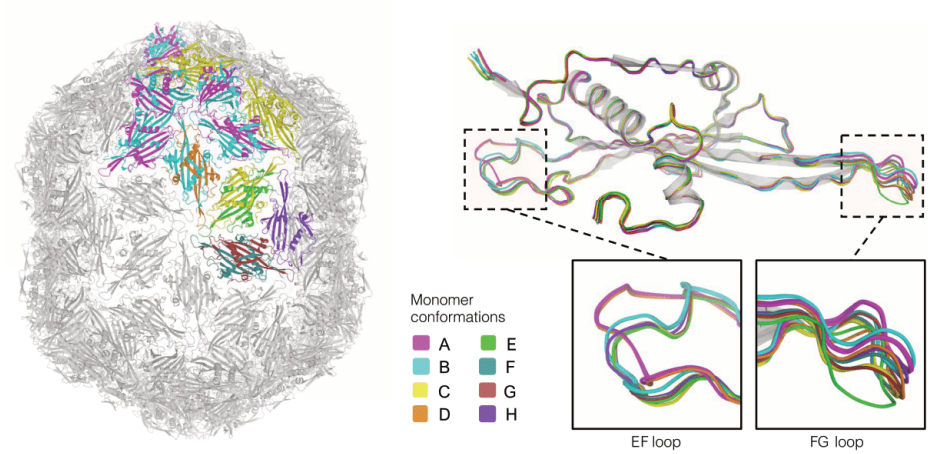

Fig. 5. The elongated virus-like particle of bacteriophage AVE016. The assembled fivefold-prolate capsid (left) is shown in light gray with the asymmetric unit of the particle, composed of 21 coat protein monomers, shown in color. When backbone atoms of individual CP monomers are superimposed (top right), eight distinct conformations are evident, which mainly differ in positioning of the EF and FG loops (bottom right). The conformers are denoted A to H, with their coloring consistent throughout the figure.

or other factors besides intersubunit contacts can have a significant influence on the overall particle stability.

In several viruses, capsid-bound metal ions have been shown to bridge negatively charged residues in neighboring subunits and participate in other intersubunit contacts. Bound calcium ions have been previously observed in ssRNA phages PRR1 (*9*) and Cb5 (*10*), where they were found to significantly enhance the particle stability. In the newly determined structures, metal ions were observed in the AVE015, AVE019, Beihai19, Beihai32, ESE007, ESE021, and ESE058 VLPs. Except for ESE021, no metal ions were present in the crystallization solution, and the bound ions were modeled as calcium; the ESE021 was crystallized in presence of zinc acetate, and accordingly, zinc ions were modeled in this case. In all cases, the metal ions were located at the quasi-threefold symmetry axes like in the previous Cb5 and PRR1 structures. We also tested the contribution of the metal ions to VLP stability by measuring their thermal denaturation in the presence of EDTA, and with the exception of the AVE015 VLPs where the results were somewhat ambiguous, the particle disintegration temperature of the metal ion–containing particles was reduced by at least 5°C when the chelating agent was added (fig. S5).

A number of ssRNA phages such as Qβ (*6*), PP7 (*8*), and AP205 (*11*) use disulfide bonds between subunits for additional particle stabilization. In all of the known cases, the disulfides are formed between pairs of cysteine residues in CP FG loops, resulting in covalently linked subunits around the icosahedral threefold and fivefold symmetry axes. In our study, disulfide bonds were present in VLPs formed by the AP205-like ESE001 and PQ-465 CPs, the AC-like NT-391 CP, and the MS2-like NT-214 CP. The disulfides are conserved in most of the known AP205-like CPs and, to some extent, also in the AC-like phages but are not found in close relatives of the NT-214 CP, suggesting that the disulfide bonds are fairly easily gained and lost during evolution. The S—S bonds have been shown to considerably increase the overall stability of the particle, but apparently, in nature, this might only be necessary as an adaptation to certain harsh environments and not as a general requirement for phage survival.

RNA binding

The CPs of ssRNA phages MS2 (*27*), PRR1 (*28*), Qβ (*29*), and PP7 (*30*) bind a specific genomic RNA hairpin at the beginning of the replicase gene which serves to repress translation of the enzyme late in infection. The RNA recognition mode of the PP7 CP is very different from that of the other three phages, and no specific CP-RNA binding has been observed in the distantly related phages Cb5 and AP205, which suggests that the interaction is not universally present in the ssRNA bacteriophages. The interaction appears not to be conserved even within the MS2-like group: For the respective phages in our study, RNA secondary structure predictions did not reveal convincing hairpins around the replicase initiation codon (*16*), and despite the otherwise significant structural similarity, the residues known to be involved in specific RNA binding were not conserved in these CPs. Sequence-specific RNA interactions are evidently not required when building the virion, and protein-RNA cross-linking studies and asymmetric cryogenic electron microscopy reconstructions of the MS2 and Qβ bacteriophages point to a model in which the capsid assembly is instead nucleated by many surface-exposed RNA hairpins in the folded genome that bind CP with moderate affinity (*4*, *31*, *32*). This model is also consistent with packaging of unspecific bacterial RNA into recombinant VLPs, largely achieved through electrostatic interactions between the RNA backbone and positively charged CP residues facing the capsid interior.

In the VLP structures, we generally observed fragmented and uninterpretable RNA density below the capsid surface, with a single notable exception of the Wenzhou1 VLPs, which revealed prominent electron density of double-stranded RNA (dsRNA) at the icosahedral twofold symmetry axes (Fig. 6A). The electron density in this case was interpretable as an A-form RNA double helix, which for simplicity was modeled as 10 A-U base pairs, although due to crystal averaging, no particular RNA bases could be distinguished. The resolution of the Wenzhou1 VLP structure was not sufficiently high for a detailed analysis of the CP-RNA interaction, although the CP side chains involved in binding the RNA backbone could be recognized (Fig. 6B). The structure does not indicate any potential sequence-specific interactions and suggests that the Wenzhou1 CP dimer is adapted for nonspecifically binding a double-stranded region of RNA. The RNA density was observed only under CP "CC" dimers, where both monomers are in the quasi-equivalent C conformation, but not around the fivefold axes, where the they adopt the asymmetric "AB" conformation. The structural basis for this discrimination is not entirely clear as the RNA binding surface in both CP dimer conformations is virtually identical, but superposition of an RNA-bound CC dimer on top of an RNA-free AB dimer inside the particle suggests that spatial constraints around the icosahedral fivefold symmetry axes force the EF loops to be positioned relatively closer to each other, which in turn makes RNA binding to AB dimers unfavorable due to a steric clash between the RNA backbone and the EF loop of a neighboring dimer (Fig. 6C). The observed RNA binding pattern in the Wenzou1 VLPs could be of biological significance and might hint that the native genome contains many dsRNA segments positioned at favorable places for binding to CP CC dimers, which would accomplish a similar function in recognition and packaging of the viral genome as RNA hairpins do in the MS2 and Qβ phages. Wenzhou1 is distantly related to the bacteriophage Cb5, the structure of which revealed intercalated RNA bases between CP subunits with a presumed role in particle stability (*10*). While no directly analogous interactions were detected in the Wenzhou1 VLPs, it can be envisioned that branched RNA secondary structures could bind to two or more adjacent CP CC dimers in the assembled particle and serve a functionally similar role. The observation of bound RNA in two different Cb5-like phages might suggest that in this group of viruses, RNA generally plays a more important structural role than in the other ssRNA phage lineages. It might also explain the stability issues of many of the Cb5-like VLPs that we have observed, as random bacterial RNA might not always be a sufficiently good substitute for the complex folded genome for holding the particle together.

In the context of RNA binding, it is interesting to also note the Beihai14 CP, which has its N termini exposed to the interior of the particle. The N termini, which were disordered and not visible in the electron density, contain a string of positively charged and polar residues, which is consistent with their role in binding RNA. It can be envisioned that in the Beihai14 virus, the CP N termini are extended into the folded genome in a functionally similar manner as several ribosomal proteins have long extensions that penetrate into the organelle to organize and stabilize the RNA structure (*33*). It can be noted that several plant and animal ssRNA viruses have positively charged N-terminal CP extensions, which are involved

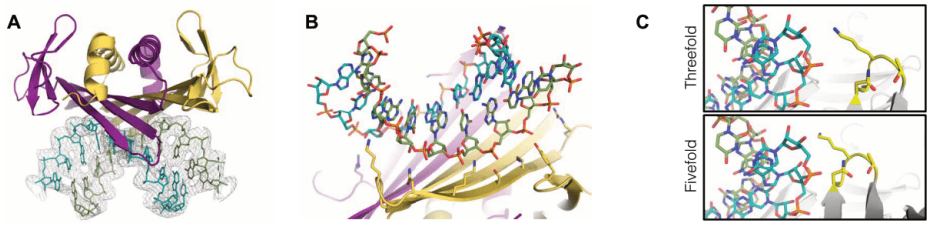

Fig. 6. dsRNA binding of the Wenzhou1 coat protein. (**A**) Overview of the protein-RNA interaction. A Wenzhou1 CP dimer is colored yellow-orange and purple as per each monomer and shown bound to a 10–base pair dsRNA fragment colored in teal and pale green as per each strand. An icosahedrally averaged F_o-F_c omit map contoured at 3 σ is presented to illustrate the observed RNA density inside the particle. (**B**) Detailed view of the protein-RNA interface. The side chains involved in RNA binding are shown in stick representation. (**C**) Discrimination of dsRNA binding between CP AB and CC dimers. The EF loop (yellow) of a neighboring CP CC dimer around an icosahedral threefold symmetry axis allows unrestricted binding of the RNA helix (top), while the same loop at the fivefold symmetry axis appears to cause a steric clash that prevents RNA binding (bottom).

in RNA binding (*34*); hence, the Beihai14 phage represents an interesting case of convergent evolution between two unrelated protein families.

Evolution of the CP fold

RNA viruses are the fastest-evolving life forms on Earth, which makes untangling their evolutionary past a nontrivial task. All RNA viruses are believed to be monophyletic and descended from an ancestral RdRp (*2*), which is also the only universally conserved protein in all RNA virus lineages. Compared to most other viral proteins, the capsid proteins are also fairly well conserved, and those of the single jelly roll architecture are ubiquitously found in plant, animal, and fungal RNA viruses (*35*). The ssRNA phage lineage, however, is believed to have separated very early from the other RNA viruses and evolved their own unique type of CP not found in any other modern viruses. The 22 novel ssRNA phage VLP structures now provide a window to the results of probably billions-of-years-long independent evolution of these proteins.

The present study has revealed that the structural diversity of the ssRNA phages extends significantly beyond the canonical MS2-like CP fold and redefines the core ssRNA phage CP architecture as only a four-stranded central β sheet and a single C-terminal α helix. Both elements make up the hydrophobic core of the CP dimer; the central β sheet additionally forms an RNA binding surface, and the α helices are important for dimerization. Evidently, any significant modification to these structures is so detrimental to the virus that none are observed even in the most highly diverged CPs. These constraints are not nearly as strict in the CP terminal regions, particularly in the N termini, where the high mutation rate of the ssRNA phages can be observed in full strength. The canonical MS2 CP fold topologically corresponds to an α-β sandwich, but a repeatedly observed theme in the novel VLP structures is a three-layered configuration where the C-terminal α helices are covered either by the N termini, like in the Wenzhou4, GQ-907, and AVE015-like CPs, or by long loops, as evidenced in some of the MS2-like proteins. However, these proteins otherwise do not appear to be closely related, which hints of convergence to this particular architecture from different starting points. Conversely, sequence alignment and structural superposition point to a recently shared evolutionary history of the EMS014 and ESE021 as well as the Beihai32 and Wenzhou4 CPs, and the prominent changes in the CP fold (swapped αB helices and rearranged N and C termini, respectively) appear to be relatively new acquisitions. It can be presumed that similar CP reorganizations with a certain regularity happen in all ssRNA phage lineages, which in turn suggests for some caution when trying to infer the evolutionary past of the ssRNA phages based solely on their capsid structure.

A recent phylogenetic reconstruction of ssRNA phages suggests of an ancient split into two lineages, of which the first is represented by the extant Cb5-like and AVE015-like viruses, while the other comprises the rest of the currently known ssRNA phages including MS2 (*15*). The phylogenetic analysis places the MS2-like and Cb5-like phage lineages at the opposite ends of the tree, but somewhat strikingly, their CPs follow the same canonical fold and their structures can be reasonably well superimposed (fig. S1). At the same time, some supposedly more closely related (e.g., ESE020-like and MS2-like, or AVE015-like and Cb5-like) phages present considerably larger structural differences in their CPs. A possible model to explain the current results would be to assume that before the ancient split into the two lineages, the ancestral ssRNA phage CP had a structure resembling the "canonical" CP fold with the N-terminal β-hairpin and that this architecture has survived relatively unchanged in the MS2-like, ESE020-like, and Cb5-like lineages. In the lineage leading to the GQ-907 phage, the AB loop of the ancestral CP can be envisioned to have increasingly extended to a point where it allowed the βA strand of one CP molecule to pair with the βB strand of the other monomer while maintaining essentially the same β-hairpin–like structure. Conversely, the AC-like phages might have emerged when the CP N terminus in a particular lineage started to become shorter, first losing the βA and then the βB strand, while the central β sheet simultaneously compensated the loss by gradually extending upward. In the lineages leading to the modern Beihai32-like and AP205-like phages, the CP C terminus might have started to extend and at some point could have substituted the βA strand, which became redundant and got lost, resulting in the observed circular permutation. The Beihai14 CP has diverged the most from others, but the long loop preceding the central β sheet still occupies broadly the same position as the N-terminal β-hairpin in MS2-like phages, suggesting that it too evolved from an ancestral MS2-like fold. Last, the AVE015-like CPs are the only proteins without a structure analogous to the N-terminal β-hairpin, but their N-terminal configuration might have emerged when the original β-hairpins unfolded and then rearranged with the βA strand switched to pair with the nearby βF strand.

MATERIALS AND METHODS

VLP production and purification

CP expression and VLP production and purification were done essentially as previously described (*16*). Briefly, CP coding sequences in pET24 bacterial expression vectors were obtained by gene synthesis, the constructs were transformed in *Escherichia coli* strain BL21(DE3), and the CP expression was done for 4 hours at 37°C or for 20 hours at 15°C. Cells were harvested by centrifugation, resuspended in lysis buffer [50 mM tris-HCl (pH 8.0), 150 mM NaCl, 0.1% Triton X-100, and 1 mM phenylmethylsulfonyl fluoride] in a wet cell weight/buffer volume ratio of 1:4, lysed by sonication, and centrifuged for 30 min at 13,000*g*. The clarified lysate was loaded on a Sepharose 4 FF column (GE Healthcare) equilibrated with phosphate-buffered saline (PBS), the VLP-containing fractions were pooled and applied to a Fractogel DEAE (M) (Merck Millipore) ion exchange column, and the bound proteins were eluted with a linear 10 column volume gradient to PBS containing 1 M NaCl. The purest VLP-containing fractions were pooled and used for crystallization.

Crystallization and structure determination

Using Amicon Ultra 100K centrifugal filter units (Merck Millipore), the purified VLPs were transferred to a 20 mM tris-HCl (pH 8.0) buffer and concentrated to approximately 10 mg/ml with an assumption that A_{260} (absorbance at 260 nm) of 8.0 corresponds to a VLP concentration of 1 mg/ml. Initial crystallization trials were done with commercial or in-house formulated screens in 0.4-μl sitting drops using a Tecan EVO 75 liquid handling robot, with further optimization of crystal growth conditions as necessary. The final crystallization conditions of all VLPs are provided in table S2. Crystals were flash frozen in liquid nitrogen, cryoprotected with 30% glycerol if necessary, and x-ray diffraction data were collected at MAX-lab (Lund, Sweden) beamline I911-3, BESSY II (Berlin, Germany) beamline 14.1, or MAX IV (Lund, Sweden) beamline BioMAX. We were unable to solve the Beihai14 VLP structure using frozen crystals, and in this case, several datasets were collected from a single ~1.5-mm crystal mounted inside a capillary at room temperature and subsequently merged together. Diffraction data were scaled and merged using Mosflm (*36*) and Scala (*37*) from the CCP4 software suite (*38*) or using XDS (*39*) through the XDSAPP graphical user interface (*40*). VLP orientation in the unit cell with respect to the standard icosahedral orientation was determined using the locked self-rotation function in GLRF (*41*). The crystallographic asymmetric unit was deduced from the crystal symmetry and unit cell parameters and prepared from a model of bacteriophage MS2 [Protein Data Bank (PDB) ID: 2MS2]. The VLP position in the unit cell was either obvious from crystal symmetry or was determined by additional translation search. In case of a one-dimensional search, the VLP model was systematically translated along the axis of interest, an initial map calculated (see below) at each location and selected for the position with the lowest R factor. Two- and three-dimensional translation searches were performed in Phaser (*42*). A correctly oriented and positioned MS2 model was placed in the unit cell and examined for crystal contacts, followed by adjustment of the VLP diameter, if necessary. The placed MS2 model was then used for calculating F_{calc} to 10-Å resolution with the program SFALL from the CCP4 software suite, and a 5-Å mask around the icosahedral asymmetric unit was generated using the program MAMA (*43*). The CCP4 program SIGMAA (*44*) was used to calculate weighted Fourier coefficients, which were used for calculating a map with the CCP4 program FFT. The map was averaged in the program AVE (*45*) using noncrystallographic symmetry and the previously prepared mask, and F_{calc} structure factors from the averaged map were calculated with SFALL. The new F_{calc} set was again used as an input for SIGMAA, and the averaging procedure was repeated for 10 more cycles. Fit of the averaged map to experimental data was monitored with the CCP4 program RSTATS at the end of each cycle. The final set of F_{calc} structure factors was used as an input for phase extension. The procedure followed the same 10-cycle averaging protocol as described above, but after the end of the last cycle, higher-resolution data corresponding to a shell of one Miller index along the longest cell axis was added to the input data, followed by another 10 cycles of averaging; the procedure was repeated until all data to the resolution of the crystal had been included. The final icosahedrally averaged map was used for manual model building in COOT (*46*). The EMS014 subunit structure was solved by molecular replacement using the program PHASER with a polyalanine model of the ESE021 CP dimer as the search model. Refinement was done in PHENIX (*47*), and the model was validated using the tools provided in COOT and the MolProbity server (*48*).

Determination of VLP thermal stability

VLP samples at a concentration of 1 mg/ml in PBS buffer with or without 20 mM EDTA were heated for 15 min in a Veriti thermal cycler (Applied Biosystems) in a 5°C-increment step gradient and then loaded on an ethidium bromide containing 1% agarose gel. Electrophoresis was performed in 1× TAE buffer, after which the RNA was visualized under ultraviolet light and protein with subsequent staining with Coomassie blue. The thermal stability was defined as the highest temperature at which the VLP band was still detectable.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/36/eabc0023/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. P. Pumpens, *Single-Stranded RNA Phages: From Molecular Biology to Nanotechnology* (CRC Press, 2020).
2. Y. I. Wolf, D. Kazlauskas, J. Iranzo, A. Lucía-Sanz, J. H. Kuhn, M. Krupovic, V. V. Dolja, E. V. Koonin, Origins and evolution of the global RNA virome. *mBio* **9**, e02329-18 (2018).
3. K. Valegård, L. Liljas, K. Fridborg, T. Unge, The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41 (1990).
4. X. Dai, Z. Li, M. Lai, S. Shu, Y. Du, Z. H. Zhou, R. Sun, In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116 (2017).
5. L. Liljas, K. Fridborg, K. Valegård, M. Bundule, P. Pumpens, Crystal structure of bacteriophage fr capsids at 3.5 Å resolution. *J. Mol. Biol.* **244**, 279–290 (1994).
6. R. Golmohammadi, K. Fridborg, M. Bundule, K. Valegård, L. Liljas, The crystal structure of bacteriophage Qβ at 3.5 Å resolution. *Structure* **4**, 543–554 (1996).
7. K. Tars, M. Bundule, K. Fridborg, L. Liljas, The crystal structure of bacteriophage GA and a comparison of bacteriophages belonging to the major groups of *Escherichia coli* leviviruses. *J. Mol. Biol.* **271**, 759–773 (1997).
8. K. Tars, K. Fridborg, M. Bundule, L. Liljas, The three-dimensional structure of bacteriophage PP7 from *Pseudomonas aeruginosa* at 3.7-Å resolution. *Virology* **272**, 331–337 (2000).
9. M. Persson, K. Tars, L. Liljas, The capsid of the small RNA phage PRR1 is stabilized by metal ions. *J. Mol. Biol.* **383**, 914–922 (2008).
10. P. Plevka, A. Kazaks, T. Voronkova, S. Kotelovica, A. Dishlers, L. Liljas, K. Tars, The structure of bacteriophage φCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. *J. Mol. Biol.* **391**, 635–647 (2009).

11. M. Shishovs, J. Rumnieks, C. Diebolder, K. Jaudzems, L. B. Andreas, J. Stanek, A. Kazaks, S. Kotelovica, I. Akopjana, G. Pintacuda, R. I. Koning, K. Tars, Structure of AP205 coat protein reveals circular permutation in ssRNA bacteriophages. *J. Mol. Biol.* **428**, 4267–4279 (2016).
12. S. R. Krishnamurthy, A. B. Janowski, G. Zhao, D. Barouch, D. Wang, Hyperexpansion of RNA bacteriophage diversity. *PLOS Biol.* **14**, e1002409 (2016).
13. M. Shi, X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen, C.-X. Li, X.-C. Qin, J. Li, J.-P. Cao, J.-S. Eden, J. Buchmann, W. Wang, J. Xu, E. C. Holmes, Y.-Z. Zhang, Redefining the invertebrate RNA virosphere. *Nature* **540**, 539–543 (2016).
14. E. P. Starr, E. E. Nuccio, J. Pett-Ridge, J. F. Banfield, M. K. Firestone, Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25900–25908 (2019).
15. J. Callanan, S. R. Stockdale, A. Shkoporov, L. A. Draper, R. P. Ross, C. Hill, Expansion of known ssRNA phage genomes: From tens to over a thousand. *Sci. Adv.* **6**, eaay5981 (2020).
16. I. Liekniņa, G. Kalniņš, I. Akopjana, J. Bogans, M. Šišovs, J. Jansons, J. Rūmnieks, K. Tārs, Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data. *J. Nanobiotech.* **17**, 61 (2019).
17. P. Pumpens, R. Renhofa, A. Dishlers, T. Kozlovska, V. Ose, P. Pushko, K. Tars, E. Grens, M. F. Bachmann, The true story and advantages of RNA phage capsids as nanotools. *Intervirology* **59**, 74–110 (2016).
18. S. C. Harrison, A. J. Olson, C. E. Schutt, F. K. Winkler, G. Bricogne, Tomato bushy stunt virus at 2.9 Å resolution. *Nature* **276**, 368–373 (1978).
19. K. Tars, A. Zeltins, L. Liljas, The three-dimensional structure of cocksfoot mottle virus at 2.7 Å resolution. *Virology* **310**, 287–297 (2003).
20. P. Plevka, K. Tars, A. Zeltins, I. Balke, E. Truve, L. Liljas, The three-dimensional structure of ryegrass mottle virus at 2.9 A resolution. *Virology* **369**, 364–374 (2007).
21. M. A. Asensio, N. M. Morella, C. M. Jakobson, E. C. Hartman, J. E. Glasgow, B. Sankaran, P. H. Zwart, D. Tullman-Ercek, A selection for assembly reveals that a single amino acid mutant of the bacteriophage MS2 coat protein forms a smaller virus-like particle. *Nano Lett.* **16**, 5944–5950 (2016).
22. A. Luque, D. Reguera, The structure of elongated viral capsids. *Biophys. J.* **98**, 2993–3003 (2010).
23. I. Cielens, V. Ose, I. Petrovskis, A. Strelnikova, R. Renhofa, T. Kozlovska, P. Pumpens, Mutilation of RNA phage Qβ virus-like particles: From icosahedrons to rods. *FEBS Lett.* **482**, 261–264 (2000).
24. J. Rumnieks, V. Ose, K. Tars, A. Dislers, A. Strods, I. Cielens, R. Renhofa, Assembly of mixed rod-like and spherical particles from group I and II RNA bacteriophage coat proteins. *Virology* **391**, 187–194 (2009).
25. M. F. Moody, The shape of the T-even bacteriophage head. *Virology* **26**, 567–576 (1965).
26. A. Sicard, Y. Michalakis, S. Gutiérrez, S. Blanc, The strange lifestyle of multipartite viruses. *PLOS Pathog.* **12**, e1005819 (2016).
27. K. Valegård, J. B. Murray, P. G. Stockley, N. J. Stonehouse, L. Liljas, Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature* **371**, 623–626 (1994).
28. M. Persson, K. Tars, L. Liljas, PRR1 coat protein binding to its RNA translational operator. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 367–372 (2013).
29. J. Rumnieks, K. Tars, Crystal structure of the bacteriophage Qβ coat protein in complex with the RNA operator of the replicase gene. *J. Mol. Biol.* **426**, 1039–1049 (2014).
30. J. A. Chao, Y. Patskovsky, S. C. Almo, R. H. Singer, Structural basis for the coevolution of a viral RNA–protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
31. K. V. Gorzelnik, Z. Cui, C. A. Reed, J. Jakana, R. Young, J. Zhang, Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qβ reveals a single maturation protein and the genomic ssRNA in situ. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11519–11524 (2016).
32. Ó. Rolfsson, S. Middleton, I. W. Manfield, S. J. White, B. Fan, R. Vaughan, N. A. Ranson, E. Dykeman, R. Twarock, J. Ford, C. C. Kao, P. G. Stockley, Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* **428**, 431–448 (2016).
33. D. J. Klein, P. B. Moore, T. A. Steitz, The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**, 141–177 (2004).
34. A. L. N. Rao, Genome packaging by spherical plant RNA viruses. *Annu. Rev. Phytopathol.* **44**, 61–87 (2006).
35. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2401–E2410 (2017).
36. T. G. G. Battye, L. Kontogiannis, O. Johnson, H. R. Powell, A. G. W. Leslie, *iMOSFLM*: A new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281 (2011).
37. P. R. Evans, Scala. Joint CCP4 + ESF-EAMCB. *Newslett. Protein Crystallogr.* **33**, 22–24 (1997).
38. M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, Overview of the *CCP4* suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
39. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
40. K. M. Sparta, M. Krug, U. Heinemann, U. Mueller, M. S. Weiss, XDSAPP2.0. *J. Appl. Crystallogr.* **49**, 1085–1092 (2016).
41. L. Tong, M. G. Rossmann, Rotation function calculations with GLRF program. *Methods Enzymol.* **276**, 594–611 (1997).
42. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *Phaser* crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
43. G. J. Kleywegt, T. A. Jones, Halloween ... masks and bones, in *From First Map to Final model. Proceedings of the CCP4 Study Weekend*, S. Bailey, R. Hubbard, D. Waller, Eds. (SERC Daresbury Laboratory, Daresbury, 1994), pp. 59–66.
44. R. J. Read, Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
45. G. J. Kleywegt, R. J. Read, Not your average density. *Structure* **5**, 1557–1569 (1997).
46. P. Emsley, K. Cowtan, *Coot*: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in *Phenix*. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
48. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity*: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
49. E. Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
50. L. Holm, Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326–5327 (2019).
51. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Acknowledgments: We thank the staff at the MAX-lab, MAX IV, and BESSY II synchrotrons for their assistance during our many data collection visits. **Funding:** This work was supported by the European Regional Development Fund through grant 1.1.1.1/16/A/104. **Author contributions:** J.R. designed the study, solved the crystal structures, analyzed and interpreted the results, and wrote the paper. I.L. supervised VLP production and purification, purified VLPs, and determined their thermal stability. G.K. and M.Š. crystallized VLPs. I.A. performed CP expression in bacteria. J.B. purified VLPs. K.T. designed and supervised the study, analyzed and interpreted the results, and wrote the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper, Supplementary Materials, or available in public databases. The three-dimensional structures have been deposited in the Protein Data Bank under accession codes 6YF7 (AC), 6YF9 (AVE002), 6YFA (AVE015), 6YFB (AVE016), 6YFC (AVE019), 6YFD (Beihai14), 6YFE (Beihai19), 6YFF (Beihai21), 6YFG (Beihai32), 6YFH (EMS014 VLP), 6YFI (EMS014 subunit), 6YFJ (ESE001), 6YFK (ESE007), 6YFL (ESE020), 6YFM (ESE021), 6YFN (ESE058), 6YFO (GQ-112), 6YFP (GQ-907), 6YFQ (NT-214), 6YFR (NT-391), 6YFS (PQ-465), 6YFT (Wenzhou1), and 6YFU (Wenzhou4). Additional data or materials related to this paper may be requested from the authors.

Submitted 31 March 2020
Accepted 20 July 2020
Published 2 September 2020
10.1126/sciadv.abc0023

Citation: J. Rūmnieks, I. Liekniņa, G. Kalniņš, M. Šišovs, I. Akopjana, J. Bogans, K. Tārs, Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes. *Sci. Adv.* **6**, eabc0023 (2020).

7.3. "Novel ssRNA phage VLP platform for displaying foreign epitopes by genetic fusion"

7.3.1. CP selection for genetic modification

The ability of the CPs of the new ssRNA phages to form stable VLPs is not the only criterion for use in vaccine platform design. In order to introduce foreign epitope insertions into the CP, knowledge of the location of the N- and C-terminal ends in the particle would be highly desirable, but despite our previous structural studies in most cases such data were not available. We selected 43 perspective sequences from our available CP library for a systematic study of genetic fusions with a model peptide M2ex3 derived from three different A type influenza M2 extracellular domains. The particles selected for the study represent the 11 previously described CP groups. Overall, in the study we included 17 members from the MS2-like supergroup, six AP205-like, five AVE015-like, three Cb5-like, three ESE020-like, two AC-like and Beihai32-like and a single CP from each of the following groups: AVE002, Beihai14, ESE001, ESE017 and ESO003.

7.3.2. Ability of chimeric proteins to form VLPs

The plasmid constructs with the M2ex3 peptide fused to either the N- or C-terminus of the CPs under investigation were created, and the chimeric CPs were expressed in *E.coli* using a standard T7 promoter-driven system. The production of chimeric proteins was heterogeneous compared to unmodified CP. Most chimeric CPs were produced in sufficient quantities. Subsequent analysis revealed that 61 of the chimeric CPs produced were at least partially soluble. All soluble proteins were purified by the classical chromatography method developed for VLP purification. Altogether, 16 of the soluble chimeric proteins were detected as assembled into VLPs. Of these, five CPs - ESE002, PQ-465, PQ-061, PQ-789 and Shahe1 - were able to form VLPs with M2ex3 fused at either terminus; for another five CPs - Beihai14, Beihai28, Beihai32, GA-879 and NF-391 - the peptide could only be fused to the C-terminus. According to EM data in all cases the observed morphology of the chimeric VLPs closely resembled those of respective unmodified T = 3 particles with a characteristic spherical shape approximately 30 nm in diameter; the only exception is the NF-391-C protein, which was similar to the unmodified CP and assembled into smaller ~ 18 nm T = 1 particles. After analysis of production data of chimeric VLPs, tested particles were sufficiently robust to yield at least moderate amounts of highly purified particles with one exception of Beihai28-C VLPs.

7.3.3. Stability of chimeric VLPs

The genetically inserted foreign sequence can have a negative effect on the stability of the particles. The stability of the chimeric VLPs was investigated by thermal denaturation measurements where purified chimeric VLPs were subjected to increasing temperatures. The measured thermal stability of all of the chimeric particles was in the range of 60 to 80 °C, only slightly lower than in cases of respective unmodified VLPs.

7.3.4. Antigenicity of the chimeric particles

For the appropriate immune response, the foreign antigen must be properly folded and freely exposed on the surface of the VLP. The location of the CP termini is of primary

importance in the context of chimeric fusion proteins. The antigenicity of the chimeric VLPs were tested in ELISA experiments using the commercially available monoclonal antibody 14C2 that is specific against the type A influenza M2 protein. The results showed that the M2ex3 peptide is located on the exterior of all of the chimeric VLPs.

Vaccine 38 (2020) 6019–6026

Contents lists available at ScienceDirect

Vaccine

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

Novel ssRNA phage VLP platform for displaying foreign epitopes by genetic fusion

Ilva Liekniņa, Darja Černova, Jānis Rūmnieks, Kaspars Tārs *

Latvian Biomedical Research and Study Center, Ratsupites 1 k-1, LV1067 Riga, Latvia

ARTICLE INFO

Article history:
Received 30 March 2020
Received in revised form 30 June 2020
Accepted 7 July 2020
Available online 24 July 2020

Keywords:
Virus-like particles
ssRNA bacteriophages
VLP vaccines

ABSTRACT

Virus-like particles (VLPs) can be used as efficient carriers of various antigens and therefore serve as attractive tools in vaccine development. Although VLPs of different viruses can be used, VLPs of ssRNA phages have convincing advantages due to their unique properties, including efficient protein production in bacterial and yeast expression systems, low production cost and easy and fast purification. Currently, the range of ssRNA phage VLPs is limited. In particular, this is true for VLPs that tolerate insertions at the N- and C-termini of the coat protein. It is therefore necessary to find new alternatives within the known ssRNA phage VLP range. From previous studies, we found approximately 80 new VLPs forming ssRNA phage coat proteins. In the current study, we attached a model peptide to the N- and C-termini of coat proteins. As a model peptide, we used a triple repeat of 23 N-terminal residues of the ectodomain of the influenza M2 protein, used previously in the development of the flu vaccine. Examining 43 novel phage coat proteins for the ability to form chimeric VLPs, we found ten new promising candidates for further vaccine design, five of which were tolerant to insertions at both the N- and C-termini. Furthermore, we demonstrate that most of the chimeric VLPs have good antigenic properties as judged from their reactivity with anti-M2 antibodies.

1. Introduction

Virus-like particles (VLPs) are protein shells composed of viral structural proteins that lack an encapsulated virus genome. VLPs morphologically closely resemble viruses of their origin and display similar immunogenic properties, yet VLPs are noninfectious and hence highly attractive for the development of safe vaccines. Furthermore, the efficacy of VLP vaccines is not restricted against viruses, from which the particular VLP has originated. Instead, viral coat proteins can be genetically or chemically modified to produce VLPs that display heterologous epitopes at a high density on their surface. Antigen presentation in a repetitive order is known to evoke stronger humoral immune responses when compared to the antigen alone [1], and high-density display onto the VLP surface provides efficient means to significantly enhance the generated immune response [2]. Because of their shape and size, VLPs can effectively enter lymph nodes [3,4], and their highly repetitive structure allows them to efficiently crosslink with B-cell receptors [5,6]. Moreover, VLPs can enhance antigen uptake by antigen presenting cells (APCs), which further stimulates the immune system and produces long-lived, high-titer antibody responses at low doses even in the absence of adjuvants [7,8]. VLPs also may contain cellular RNAs, thereby enhancing the immune response through the action of the receptors TLR3 and TLR7. Importantly, tolerance to self-antigens may be overcome by displaying them on the surface of VLPs, enabling the development of therapeutic vaccines against certain noninfectious diseases and conditions. Due to their immunological properties and excellent safety profile, VLPs are thus among the most promising carriers for the development of new vaccines [9,10].

Single-stranded RNA (ssRNA) bacteriophages are small viruses infecting a number of gram-negative bacteria. ssRNA phages are among the simplest known viruses and consist of an RNA genome packaged inside an icosahedral protein shell that is approximately 28 nm in diameter and composed of 178 copies of the coat protein (CP) and a single molecule of the receptor-binding maturation protein. Recombinant expression of ssRNA phage coat proteins in bacteria or yeast typically results in efficient assembly of VLPs without the need for any other viral components. The coat proteins can be easily genetically manipulated to produce fusion proteins with antigens of interest that assemble in chimeric VLPs; alternatively, antigens can be chemically coupled onto the surface of preassembled unmodified VLPs. Altogether, the ssRNA phage platform is an

* Corresponding author.
E-mail address: kaspars@biomed.lu.lv (K. Tārs).

https://doi.org/10.1016/j.vaccine.2020.07.016
0264-410X/

attractive option for developing VLP-based vaccine prototypes [11]. However, CPs from different phages significantly vary in their ability to carry particular foreign sequences. The majority of successfully modified ssRNA phage VLPs have been created using only a few ssRNA phage CPs, and these include the bacteriophages PP7, MS2 and AP205, which have demonstrated much higher tolerance for fusing with different antigens than all others. There are numerous prophylactic and therapeutic vaccine candidates that have been constructed based on bacteriophage VLPs employing a genetic fusion methodology, for example, AP205-based flu vaccine [12], MS2-based foot-and-mouth disease vaccine [13], MS2-based cancer therapeutic [14], MS2-based cholesterol-lowering vaccine [15], PP7-based HPV vaccine [16], and MS2 anti-hCG contraceptive [17]. Chemical coupling of foreign epitopes on the surface of VLPs was developed as an alternative method to genetic fusion, for example, Qβ-based atherosclerosis therapy [18], Qβ-based Alzheimer's disease therapy [19–21], AP205-based HIV vaccine [22] and Qβ-based diabetes therapy [23]. Despite the many advantages, the limited supply of suitable CPs for modification puts certain limitations for designing new vaccines that need to be addressed for continued development of the ssRNA phage VLP platform.

Recent metagenomic studies have uncovered many novel ssRNA phage sequences with only very distant similarities to previously studied sequences. In a previous study, we characterized more than 100 new ssRNA phage coat protein sequences obtained by gene synthesis from metagenomic data, in which 80 cases were able to assemble into VLPs [24]. Many of these CPs have no detectable sequence similarity to those of the previously studied phages, and their potential for carrying antigens remains unknown.

Type A influenza viruses are a significant cause of severe seasonal respiratory tract infections in humans. Up to 650 000 deaths annually are associated with respiratory diseases from seasonal influenza, according to new estimates by the United States Centers for Disease Control and Prevention (CDC), the World Health Organization and global health partners (WHO, 2018). Existing vaccines are based on inducing a protective neutralizing antibody response against the globular head of the viral surface protein hemagglutinin (HA) [4]. However, the high mutation rate of influenza viruses leads to rapid accumulation of point mutations in the HA protein, with the consequence that vaccines have to be reformulated each year based on predictions of which strains will be circulating during the next influenza season [25]. Considering the additional unstable efficacy of the current vaccines and their rather complicated and time-consuming production in fertilized chicken eggs, the development of a new generation of universal, effective and easy-to-manufacture influenza vaccines emerges as a global public health priority.

As integral components of the viral envelope, particles of influenza viruses contain several proton-selective pores formed by the viral protein M2. The 23 N-terminal residues of the M2 protein form an ectodomain (M2e) that faces the exterior of the virion and is highly conserved among different influenza A strains. M2e represents an attractive target for prophylactic influenza vaccines [26,27], and M2e-based vaccination has indeed resulted in broad-range antibody-mediated protection in influenza A animal models [28–31]. However, the small size of the isolated M2 ectodomain makes it a poor immunogen, and approaches such as presentation of M2e peptides on the surface of large macromolecular assemblies are required for generating a robust immune response [32].

To find the best candidates from our VLP library for the insertion of foreign amino acid sequences, we characterized 43 novel ssRNA bacteriophage coat proteins for their tolerance of N- and C-terminal fusions with three consecutive copies of M2e (M2ex3) derived from avian, human and swine influenza A virusesas a model peptide.

2. Materials and methods

2.1. Construction of expression plasmids

Two expression vectors, pET24-M2ex3 and pET28-M2ex3, contain sequences from three distinct variants of influenza A M2e [33] (Supplementary Fig. S2), and NcoI and BamHI sites for cloning were obtained via gene synthesis (General Biosystems). Coat protein-encoding sequences (Supplementary Table S1) were PCR-amplified from their respective expression plasmids [24] using oligonucleotides that introduce NcoI and BamHI sites at both ends of the fragment . The PCR products were then cloned into NcoI-BamHI-digested pET24-M2ex3 and pET28-M2ex3 vectors to create expression plasmids for N- and C-terminal CP fusion proteins, respectively.

2.2. Production and solubility assessment of fusion proteins

E. coli strain BL21(DE3) cells were transformed with plasmids encoding CP fusions, and individual colonies were inoculated in 10 ml of LB media supplemented with 30 μg/ml kanamycin and incubated overnight at 37 °C without agitation. The overnight cultures were transferred to 100 ml of 2xTY medium, and the cells were grown at 37 °C until an OD_{600} of 0.6, at which point IPTG (isopropyl-β-D-thiogalactoside) was added to a final concentration of 1 mM, and the cells were further incubated for 16 h at 20 °C. Cells were harvested by centrifugation. A small aliquot of cells were suspended in lysis buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% Triton X-100) at a wet weight/volume proportion of 1:25; then, cells were sonicated for 1 min with 5 s on/off pulses, and the lysate was clarified by centrifugation. The protein production level (Supplementary Fig. S3) and soluble and insoluble proteins were analyzed by SDS-PAGE. For the subsequent VLP purification, the expression protocol was upscaled to 2 L.

2.3. Purification of chimeric VLPs

Two grams of wet cells were disrupted by sonication, and the clarified cell lysate was precipitated overnight with 40% saturated ammonium sulfate. The precipitate was collected by centrifugation and dissolved in 2 ml of phosphate-buffered saline supplemented with 0.2% Tween 20, 0.5 M urea and 1 mM PMSF (phenylmethylsulfonyl fluoride). The concentrated preparation was loaded onto a 30 ml Sepharose 4FF column (Omnifit 10 mm ID/ 330 mm) attached to an Akta PrimePlus chromatography system (GE Healthcare), and 2 ml fractions were collected in PBS with the flow rate set to 0.5 ml/min. Fractions were analyzed by 1% agarose gel electrophoresis and SDS-PAGE, and those containing VLPs were pooled and applied to a 5 ml Fracto-DEAE ion exchange column (Omnifit 10 mm ID/ 100 mm) pre-equilibrated with PBS. Bound proteins were eluted with an 25 ml 0.15 – 1 M NaCl gradient in PBS with the flow rate set to 2 ml/min and collected in 2 ml fractions. The purest fractions were pooled, dialyzed against PBS and stored frozen at −20 °C for further downstream experiments.

2.4. Electron microscopy

Purified VLP samples (1 mg/ml) were adsorbed on carbon-Formvar-coated copper grids and negatively stained with a 1% aqueous solution of uranyl acetate. The grids were examined in a JEM-1230 electron microscope (JEOL Ltd., Tokyo, Japan) operated at 100 kV. Electron micrographs were recorded with iTEM software (version 3.2, Soft Imaging System GmbH) using a side-mounted Morada digital camera (Olympus-Soft Imaging System GmbH, Munster, Germany).

2.5. Thermal stability assay

Thermal stability experiments were performed essentially as described previously [34]. Chimeric VLP samples at a concentration of 1 mg/ml in PBS with or without 10 mM DTT were heated for 15 min in a Veriti thermal cycler (Applied Biosystems, Singapore) in a 5 °C increment step gradient and then loaded on a 1% agarose gel. After electrophoresis in TAE buffer, the RNA was visualized with ethidium bromide, and the protein was visualized with Coomassie blue.

2.6. Elisa

Protein concentrations were determined with the Bradford protein assay kit (Pierce). One hundred microliters of chimeric VLPs or M2ex3 protein (10 μg/mL) (Supplementary Fig. S4) were adsorbed on a 96-well ELISA microplate (Sarstedt, Germany) overnight at 4 °C. Plates were blocked with 1% BSA blocking buffer for 1 h at 37 °C. Mouse anti-influenza A M2 monoclonal antibody 14C2 (Invitrogen) diluted at 1:250 in blocking buffer was added in three-fold dilutions to triplicate wells and incubated for 1 h at 37 °C. The plates were washed three times in PBS/0.05% Tween 20 in between different steps. Horseradish peroxidase (HRP)-conjugated polyclonal rabbit anti-mouse IgG (Sigma–Aldrich, USA) was diluted 1:1000 in blocking buffer and incubated for 1 h. Finally, color reactions were developed for 20 min by adding the o-phenylenediamine substrate. The HRP enzymatic reaction was terminated by the addition of 2.0 M H_2SO_4, and the optical density was measured at 492 nm using an ELISA plate reader (BDSL Immunoskan MS, Finland). Antigenicity significance of differences in antibody titers relative to negative control, i.e. the baseline was determined by measuring the OD_{429} of the wells covered only with BSA. For evaluating the antigenicity, the results were calculated in Microsoft Excel (version 10.0). By connecting the measurement result points obtained for each sample, a linear function intersecting the baseline was graphically represented. The hypothetical cross point was calculated and defined as the end point for evaluating the antigenicity. The individual titration curves can be found in the Supplementary Fig. S5.

3. Results

3.1. Stable VLPs for genetic fusion experiments

While the majority of the previously characterized metagenome-derived ssRNA phage CPs were able to assemble into VLPs, there were significant variations in their production levels, homogeneity and stability. Successful use in vaccine development requires that not only the modified CPs assemble into VLPs but also the chimeric particles can be inexpensively produced in high yields and are sufficiently stable to withstand downstream purification procedures and long-term storage. Guided by these criteria, we selected 38 perspective sequences from the available CP library for a systematic study of genetic fusions with a model peptide. A further tBLASTn search through the NCBI nucleotide databases uncovered several additional novel ssRNA phage CPs, and five more sequences located under GenBank accession codes PQDQ01000061 (further denoted as PQ-061), PQDQ01001338 (PQ-338), PQDQ01001357 (PQ-357), PQDQ01001465 (PQ-465) and PQDQ01001789 (PQ-789) were extracted, acquired via gene synthesis and included in the CP modification study. The additional CP sequences were chosen due to the presence of disulfide bridges, which are known to significantly increase the stability of VLPs. The resulting panel of 43 CPs includes representatives from 11 of the previously recognized 14 CP similarity groups and contains 17 representatives from the MS2-like supergroup, six AP205-like, five AVE015-like, three Cb5-like, three ESE020-like, two AC-like and Beihai32-like and a single CP from each of the following groups: AVE002, Beihai14, ESE001, ESE017 and ESO003 [24]. For comparison, we additionally included the Qβ and AP205 CPs in the study as the current "gold standard" ssRNA phage VLP vectors. All new sequences used in the study are available in Supplementary Table S1.

3.2. 16 Of obtained fusion proteins assemble into VLPs

The plasmid constructs with the M2ex3 peptide fused to either the N- or C-terminus of the CPs under investigation were created, and the chimeric CPs were expressed in *Escherichia coli* using a standard T7 promoter-driven system. Most of the chimeric CPs were produced in high amounts; however, in 18 cases, no expression was detected (Table 1). In the majority of cases, either both the N- and C-termini or neither of the fusion proteins were produced, and in the few remaining instances, predominantly the C-terminal variants failed to be produced. Subsequent analysis revealed that 61 of the chimeric CPs that were produced were also at least partially soluble (Fig. 1, Table 1). Clarified *E. coli* lysates of the soluble chimeric CPs were then applied to a gel filtration column, and fractions were analyzed by agarose gel electrophoresis and SDS-PAGE to detect VLPs. Altogether, 16 of the soluble chimeric proteins were detected as assembled into VLPs, as judged from their elution profile from the gel filtration column (Table 1). Of these, five CPs (ESE002, PQ-465, PQ-061, PQ-789 and Shahe1) were able to form VLPs with M2ex3 fused at either terminus; for another five CPs (Beihai14, Beihai28, Beihai32, GA-879 (GALT01093879) and NF-391 (NFYT01000391)), the peptide could only be fused to the C-terminus to retain the ability to assemble into particles. For further characterization, the VLPs were further purified using ion exchange chromatography and examined by electron microscopy (Fig. 2). In all cases, the observed morphology of the chimeric VLPs closely resembled those of respective unmodified T = 3 virus-like particles with a characteristic spherical shape approximately 30 nm in diameter; the only exception is the NF-391-C protein, which was similar to the unmodified CP and assembled into smaller ~ 18 nm T = 1 particles. Most VLP-forming chimeric proteins displayed high production and solubility levels, except for ESE002-C, Shahe1-N and PQ-061-N (Fig. 1). It should be noted that several of the VLP-forming fusion CPs were found partially in insoluble fraction, indicating some folding issues with this particular fusion antigen. Overall, the chimeric VLPs were sufficiently robust to yield at least moderate amounts of highly purified particles with the sole exception of Beihai28-C VLPs, which were partially lost during purification due to their tendency to adhere to the ion exchange matrix and ultrafiltration membranes, suggesting misfolding and/or aggregation issues in this particular case.

3.3. Chimeric VLPs are thermostable at 60–80 °C

Despite successful assembly into VLPs, the addition of heterologous sequences to the coat protein can have a negative impact on the overall stability of the particles, leading to gradual dissociation of the VLPs into CP subunits and consequent inactivation of the potential vaccine. Therefore, the stability of the chimeric VLPs was investigated by thermal denaturation measurements in which aliquots of purified VLPs were subjected to increasing temperatures and then analyzed by native agarose gel electrophoresis. The thermal stability was then defined as the highest temperature at which the VLP band could still be visualized. The measured thermal stability of all of the chimeric particles was in the range of 60 to 80 °C and only slightly lower than that of unmodified VLPs

Table 1
Properties of genetically modified ssRNA phage coat proteins.

Coat protein	Similarity group	Expression in *E.Coli*		Protein solubility		Formation of VLPs	
		M2ex3 in N-terminus	M2ex3 in C-terminus	M2ex3 in N-terminus	M2ex3 in C-terminus	M2ex3 in N-terminus	M2ex3 in C-terminus
AP205	AP205	–	+	n/d	+	n/d	+
Qβ	Qβ	–	–	n/d	n/d	n/d	n/d
AVE002	AVE002	+	–	+	–	–	n/d
AVE016	AVE015	+	+	–	+	n/d	–
AVE039	AVE015	+	–	+	n/d	–	n/d
Beihai14	Beihai14	+	+	+	+	–	+
Beihai18	MS2	+	+	+	+	–	–
Beihai19	MS2	+	+	+	+	–	–
Beihai23	MS2	+	+	+	+	–	–
Beihai26	MS2	+	+	+	+	–	–
Beihai28	Beihai32	+	+	+	+	n/d	+
Beihai30	MS2	+	–	+	n/d	–	n/d
Beihai32	Beihai32	–	+	–	+	n/d	+
Beihai33	MS2	+	+	+	+	–	–
Beihai34	MS2	–	–	n/d	n/d	n/d	n/d
Beihai9	Cb5	+	+	+	+	–	–
EMS001	ESE020	+	+	+	+	–	–
EMS011	Cb5	+	+	+	+	–	–
ESE001	ESE001	–	–	n/d	n/d	n/d	n/d
ESE002	AP205	+	+	+	+	+	+
ESE007	MS2	+	+	+	+	–	–
ESE017	ESE017	+	+	+	+	–	–
ESE020	ESE020	+	–	+	n/d	–	n/d
ESE029	MS2	+	+	+	+	–	–
ESE030	MS2	+	+	+	+	–	–
ESE041	ESE020	+	+	+	–	–	n/d
ESO003	ESO003	+	+	+	+	–	–
GA-112	AVE015	+	+	–	–	n/d	n/d
GA-492	AVE015	+	+	+	+	–	–
GA-879	AVE015	+	+	+	+	–	+
Hubei10	MS2	–	–	n/d	n/d	n/d	n/d
Hubei14	MS2	+	+	–	+	n/d	–
NF-214	MS2	+	+	+	+	–	–
NF-391	AC	+	+	+	+	–	+
NF-443	AC	+	+	–	–	n/d	n/d
PQ-061	AP205	+	+	+	+	+	+
PQ-338	AP205	–	–	n/d	n/d	n/d	n/d
PQ-357	MS2	+	+	–	–	n/d	n/d
PQ-465	AP205	+	+	+	+	+	+
PQ-789	AP205	+	+	+	+	+	+
Shahe1	AP205	+	+	+	+	+	+
Shahe3	MS2	+	+	–	–	n/d	n/d
Wenling2	MS2	–	–	n/d	n/d	n/d	n/d
Wenzhou2	Cb5	+	+	+	+	–	–
Wenzhou4	MS2	+	+	+	+	–	–

The listed properties include the CP similarity group, production level (+, detected; –, not detected), solubility (+, soluble; –, insoluble), VLP formation confirmed by EM: (+, efficient VLP formation; –, no VLPs observed), n.d.: not determined.

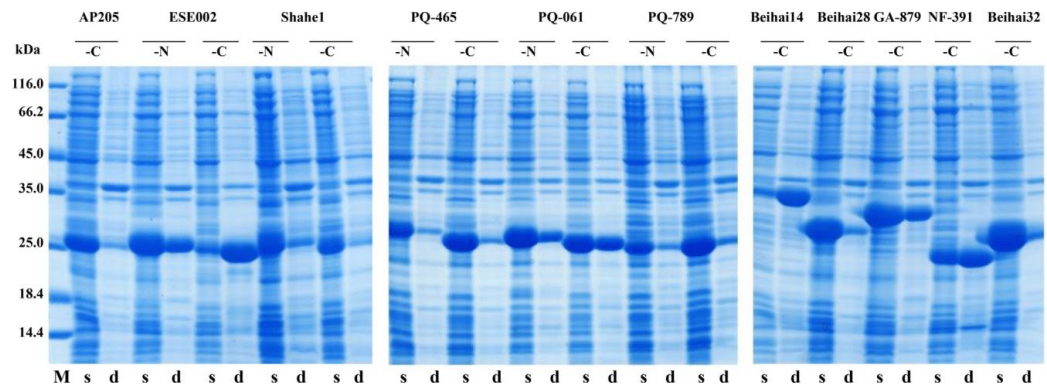

Fig. 1. Solubility of chimeric capsid proteins forming VLPs was analyzed by SDS polyacrylamide gel electrophoresis and stained with Coomassie blue. M: molecular weight marker (Thermo Scientific, #26610); s: supernatant of cell lysate after centrifugation; d: pellet of cell lysate after centrifugation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

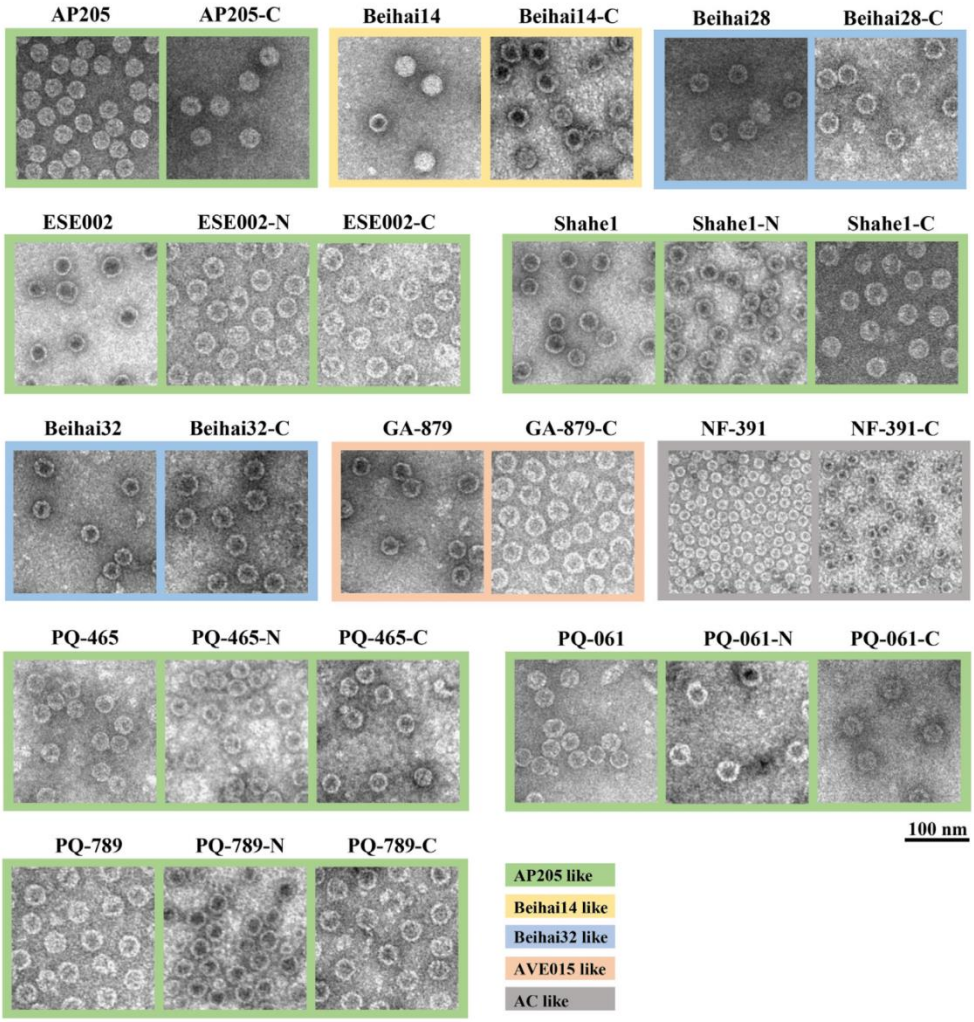

Fig. 2. Representative electron micrographs of the native and chimeric VLPs. The CP similarity groups are indicated in the colored background.

(Table 2). The majority of the chimeric CPs that were able to assemble into VLPs contain cysteine residues at positions that would permit the formation of intersubunit disulfide bonds. To assess the contribution of the potential disulfides to the overall particle stability, thermal denaturation experiments were repeated in the presence of the reducing agent dithiothreitol. A reduction in particle stability was indeed observed for all cysteine-containing VLPs, ranging from a modest 10 °C decrease in the cases of ESE002, GA-879 and NF-391 to a drastic 35 °C decrease for the N-terminally modified PQ-789 coat protein VLPs.

3.4. M2ex3 peptide is located on the exterior of all of the chimeric VLPs

For a strong and appropriate immune response, the antigen must be properly folded and freely exposed onto the surface of the virus-like particle. The available three-dimensional structures of ssRNA phage particles have demonstrated that the localization of the CP termini is evolutionarily not well conserved, and only in the case of phage AP205 are both termini freely available on the surface of the particle. The location of the CP termini is of primary importance in the context of chimeric fusion proteins; however, unfortunately, the very weak sequence identity makes predictions for the location of CP termini unreliable for many of the novel ssRNA phage CPs. Additionally, structural information per se does not ensure efficient antigenicity, which should be tested experimentally. Therefore, the antigenicities of the chimeric VLPs were tested in ELISA experiments using the commercially available monoclonal antibody 14C2 that is specific against the type A influenza M2 protein. The results indicated that the M2ex3 peptide is located on the exterior of all of the chimeric VLPs (Fig. 3); however, the antibody dilution at the detection limit ranged quite substantially within one order of magnitude from moderate (6500) in case NF-391-C to very high (greater than50000) for Beihai32-C.

Table 2
Properties of promising VLP-forming vaccine candidates and chimeric VLPs.

Chimeric VLP	Total Cys in CP	Cys positions in CP	Tm, °C	Tm, °C + DTT
AP205	2	64, 68	75	50
AP205-C	2	64, 68	70	50
Beihai14	0	–	85	n/d
Beihai14-C	0	–	80	n/d
Beihai28	0	–	65	n/d
Beihai28-C	0	–	80	n/d
Beihai32	0	–	70	n/d
Beihai32-C	0	–	70	n/d
ESE002	2	63, 71	65	55
ESE002-C	2	63,71	60	50
ESE002-N	2	63, 71	60	50
GA-879	3	101, 102, 103	75	55
GA-879-C	3	101, 102, 103	65	55
NF-391	3	66, 67, 75	75	65
NF-391-C	3	66, 67, 75	70	60
PQ-465	2	63, 67	80	55
PQ-465-C	2	63,67	80	55
PQ-465-N	2	63, 67	75	50
PQ-338	2	61,64	65	40
PQ-338-C	2	61,64	60	35
PQ-338-N	2	61,64	65	RT
PQ-789	2	57, 61	70	50
PQ-789-C	2	57, 61	70	45
PQ-789-N	2	57, 61	75	40
Shahe1	3	64, 68, 97	75	55
Shahe1-C	3	64, 68, 97	80	50
Shahe1-N	3	64, 68, 97	75	45

The listed properties include the presence and positions of cysteine residues in the capsid protein and the VLP "melting" temperature (thermal stability) without or with 10 mM DTT. n/a: not applicable due to the lack of VLPs, n.d.: not determined.

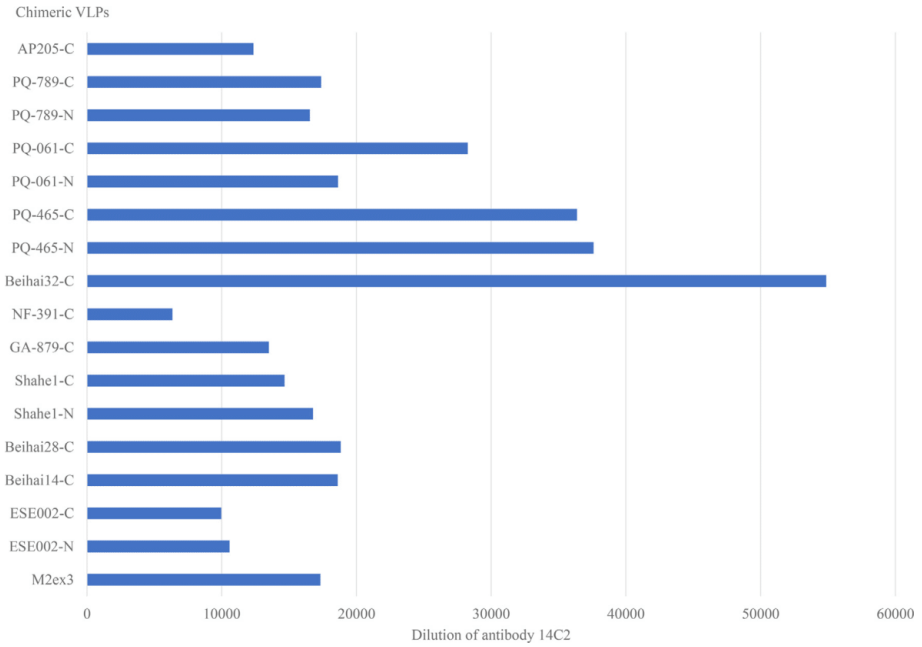

Fig. 3. Antigenicity test of chimeric VLPs. Level of antigenicity was determined by direct ELISA. Unmodified and chimeric VLPs were tested for their ability to bind to mouse anti-influenza A M2 monoclonal antibody 14C2 (Invitrogen) and were determined by calculating the hypothetical (possible) cross-point with the baseline. The results of native VLPs were below the established baseline and are not shown.

4. Discussion

While VLP-based vaccine technology has several advantages and good prospects for widespread application against a variety of infectious diseases, only a limited number of suitable VLP platforms are currently available. Here, we have characterized 43 novel ssRNA bacteriophage CPs for their ability to form chimeric VLPs when fused with a model peptide. Although the exact fusion partner definitely might influence the CP folding and VLP forming abilities our results provide a good platform for initial choice of new antigen carriers in VLP-based vaccine development.

The majority of ssRNA phage CP sequences currently discovered in metagenomic sequencing data fall into a broad similarity group that includes the CP of the widely studied ssRNA phage MS2. The MS2 CP itself as well as that of its close relative bacteriophage fr [35,36] have been among the first carrier proteins explored for the creation of chimeric VLPs. However, MS2 and fr have had a relatively low tolerance to long insertions and have therefore been of limited use. In MS2, short insertions can be made in the so-called AB loops connecting beta strands A and B that are located on the capsid surface. In the case of further stabilization of MS2 VLPs, by creating covalent coat protein dimers, it was demonstrated that essentially any insertions up to 10 residues can be tolerated [37,38]. The coat protein of the bacteriophage Qβ, which has approximately 20% sequence identity to MS2, has performed notably better, although the fusion strategy is very different from that used in MS2. Bacteriophage Qβ is unique among the other ssRNA phages in that it contains several copies of a C-terminally elongated coat protein variant A1 incorporated in the particle, and the CP has therefore a naturally evolved ability to tolerate C-terminal extensions in up to 14–48% of CP monomers, depending on the length of the insert [39]. In our current study, approximately 40% of the investigated CPs fall into the MS2-like CP similarity group; however, their performance as VLP carriers has been underwhelming. While the vast majority of the fusion proteins were produced in high amounts and were soluble, none of them were able to assemble into VLPs. es Therefore, our results reinforce the general idea that MS2-like sequences have limited potential as carriers of bulky foreign antigens in vaccine development. The main two reasons why MS2-like sequences are underperforming are evident from crystal structures of MS2-like VLPs; first, N- and C-termini are marginally located on the surface, and second, these termini are clustered together around symmetry axes, creating steric problems for accommodating bulky antigens [40].

Beihai28 and Beihai32, which share 40% identity were previously included in MS2-similarity group due to very weak sequence similarities. However, our recent data suggests that they form a distinct group on their own – Beihai32-like –, structurally displaying remote similarity to AP205. In both cases, C-terminal fusion variant, was able to assemble into VLPs. Furthermore, C-terminal fusion of Beihai32 displayed the highest antigenicity of all tested constructs. Interestingly, fusion of related Beihai28 had significantly lower antigenicity. To some extent this can be explained by structure of Beihai32 (J.R., to be published separately), in which C-terminus extends far from VLP surface, making any attached antigen very accessible, while C-terminus of Beihai28 is by 10 residues shorter, potentially resulting in accessibility problems.

In our previous study, a Cb5-like coat protein supergroup emerged as the second largest; however, only four sequences from this group made it in our CP modification study. The underlying reason for such underrepresentation is that less than half of the previously tested Cb5-like CPs were able to assemble into VLPs in the first place, and those that did were generally not very stable and cumbersome to purify. Of the three Cb5-like proteins investigated in the current study (Beihai9, EMS011, and Wenzhou2), none were able to form chimeric VLPs. Of the other CP similarity groups, none of the fusion CPs from the AVE002-, ESE020-, ESE017- and ESO003-like groups were able to assemble into particles, which implies that CPs from these groups are generally not very prospective for future developments.

The last of the major previously recognized CP groups, the AVE015-like sequences, was represented by five different sequences in our study. Of these, only one, the GA-879 CP with a C-terminally fused M2ex3 peptide, was able to assemble into VLPs, which makes this group not a particularly promising candidate for further exploration for new VLP carriers. One of the two tested AC-like CPs, the C-terminally fused NF-391 coat protein, was also able to assemble into VLPs, albeit only in the smaller T = 1 particles. Additionally, the Beihai14 CP retained the ability to assemble into chimeric particles when fused with the M2ex3 peptide at the C-terminus. The Beihai14 CP is the only sequence of its kind and is unique among others for its extreme length (208 residues; for comparison, the MS2 CP is 129 residues long) and the very high stability of VLPs, despite the lack of stabilizing disulfide bonds. The Beihai14 CP and related sequences should be discovered in the future and thus remain interesting for further research in VLP vaccine development.

Clearly, however, the AP205-like coat proteins stand out among all others as the best carriers for foreign antigens. Since the beginning of this century, AP205 VLPs have been very attractive objects for vaccine design due to their ability to tolerate genetically fused foreign sequences and self-assemble even when the epitopes are long (up to 55 amino acids) and have multiple cysteines. The ability to fuse to both ends of AP205 VLPs allows better display of N- and/or C-terminal epitopes, which can be problematic for other VLPs [41,42]. The suitability of AP205 VLPs for genetic fusions can be easily explained from the structural point of view; both the C- and N-termini are located on the surface of VLPs and well separated in the neighboring CP subunits [40].

Of the six AP205-like CPs tested, five were able to form chimeric VLPs, and all tolerated both N- and C-terminal fusions equally well. While the excellent properties of the AP205 coat protein in this respect have been well documented, our study suggests that such qualities are almost a universal feature of the whole group: the AP205-like CPs used in our study share only 15–30% sequence identity to the AP205 CP and less than 37% identity to each other. It therefore seems reasonable to assume that in addition to the five novel CPs described in this study, there might be a significant amount of further sufficiently distinct yet undiscovered CPs with similarly good properties for the development of chimeric VLPs.

It can be noted that most of the VLPs that were able to carry the M2ex3 peptide contain intersubunit disulfide bonds. For the AP205-like group, the cysteine residues appear to be highly conserved and apparently significantly contribute to the overall stability of the particles, as demonstrated by the at least 20 °C reduction in the thermal stability in the presence of a reducing agent. Disulfide bonds were also present in the AC-like NF-391-C VLPs, and out of the five tested AVE015-like CPs, only one (GA-879-C) was able to form particles and was also the only one that contained disulfides. On the other hand, the AC-like NF-443 CP could not form VLPs despite having disulfides, and similarly, these did not help in the case of MS2-like Hubei10 and NF-214 chimeric CPs; however, the chimeric Beihai14 VLPs were among the most stable irrespective of lacking any disulfide bonds. Therefore, although beneficial, the disulfides do not appear to be sufficient or absolutely required for being a robust and stable carrier.

The differences in observed antigenicities among different VLPs presumably reflect accessibility of exposed M2 peptide to antibodies. We expect that VLPs with higher antigenicity test

should also display better immunogenicity, for which efficient display on VLP surface is also required.

In conclusion, in this study, we characterized several dozens of novel ssRNA bacteriophage CPs for their ability to form chimeric VLPs decorated with a model peptide from the influenza A virus. Our data suggest that the AP205-like coat proteins are the most prospective ssRNA phage-derived VLP vectors, especially PQ-465, C- and N-terminal fusions of which displayed high antigenicities. Our results pinpoint also several other previously unknown CP groups with good potential for VLP-based vaccine development applications worth further exploration. In particular, Beihai32 VLPs seem to have very good potential for C-terminal fusions due to the highest observed antigenicity. The M2ex3-carrying VLPs developed during the current study will be further investigated as prototypes for a universal influenza vaccine.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the European Research and Development Foundation grant 1.1.1.1/16/A/104.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2020.07.016.

References

[1] Bachmann MF, Jennings GT. Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. Nat RevImmunol 2010;10:787–96.
[2] Pumpens P, Renhofa R, Dishlers A, Kozlovska T, Ose V, Pushko P, et al. The true story and advantages of RNA phage capsids as nanotools. Intervirology 2016;59:74–110.
[3] Manolova V, Flace A, Bauer M, Schwarz K, Saudan P, Bachmann MF. Nanoparticles target distinct dendritic cell populations according to their size. Eur J Immunol 2008;38:1404–13.
[4] Cubas R, Zhang S, Kwon S, Sevick-Muraca EM, Li M, Chen C, et al. Virus-like particle (VLP) lymphatic trafficking and immune response generation after immunization by different routes. J Immunother 2009;32:118–28.
[5] Lopez-Sagaseta J, Malito E, Rappuoli R, M.J. B. Self-assembling protein nanoparticles in the design of vaccines. Comput Struct Biotechnol J 2016;14:58-68.
[6] Marini A, Zhou Y, Li Y, Taylor IJ, Leneghan DB, Jin J, et al. A universal plug-and-display vaccine carrier based on HBsAg VLP to maximize effective antibody response. Front Immunol 2019;10:2931.
[7] Duan H, Chen X, Boyington JC, Cheng C, Zhang Y, Jafari AJ, et al. Glycan masking focuses immune responses to the HIV-1 CD4-binding site and enhances elicitation of VRC01-class precursor antibodies. Immunity 2018;49(301–11):e5.
[8] Kanekiyo M, Bu W, Joyce MG, Meng G, Whittle JR, Baxa U, et al. Rational design of an epstein-barr virus vaccine targeting the receptor-binding site. Cell 2015;162:1090–100.
[9] Perotti M, Perez L. Virus-like particles and nanoparticles for vaccine development against HCMV. Viruses 2019;12.
[10] Karch CP, Burkhard P. Vaccine technologies: From whole organisms to rationally designed protein assemblies. Biochem Pharmacol 2016;120:1–14.
[11] Pumpens P. Single-stranded RNA phages: From molecular biology to nanotechnology. United States: CRC Press Taylor & Francis Group; 2020.
[12] Schmitz N, Beerli RR, Bauer M, Jegerlehner A, Dietmeier K, Maudrich M, et al. Universal vaccine against influenza virus: linking TLR signaling to anti-viral protection. Eur J Immunol 2012;42:863–9.
[13] Dong YM, Zhang GG, Huang XJ, Chen L, Chen HT. Promising MS2 mediated virus-like particle vaccine against foot-and-mouth disease. Antiviral Res 2015;117:39–43.
[14] Frietze KM, Roden RB, Lee JH, Shi Y, Peabody DS, Chackerian B. Identification of anti-CA125 antibody responses in ovarian cancer patients by a novel deep sequence-coupled biopanning platform. CancerImmunol Res 2016;4:157–64.
[15] Crossey E, Amar MJA, Sampson M, Peabody J, Schiller JT, Chackerian B, et al. A cholesterol-lowering VLP vaccine that targets PCSK9. Vaccine 2015;33:5747–55.
[16] Caldeira Jdo C, Medford A, Kines RC, Lino CA, Schiller JT, Chackerian B, et al. Immunogenic display of diverse peptides, including a broadly cross-type neutralizing human papillomavirus L2 epitope, on virus-like particles of the RNA bacteriophage PP7. Vaccine 2010;28:4384–93.
[17] Caldeira J, Bustos J, Peabody J, Chackerian B, Peabody DS. Epitope-specific anti-hCG vaccines on a virus like particle platform. PLoS ONE 2015;10:e0141407.
[18] Tissot AC, Spohn G, Jennings GT, Shamshiev A, Kurrer MO, Windak R, et al. A VLP-based vaccine against interleukin-1alpha protects mice from atherosclerosis. Eur J Immunol 2013;43:716–22.
[19] Chackerian B, Rangel M, Hunter Z, Peabody DS. Virus and virus-like particle-based immunogens for Alzheimer's disease induce antibody responses against amyloid-beta without concomitant T cell responses. Vaccine 2006;24:6321–31.
[20] Li QY, Gordon MN, Chackerian B, Alamed J, Ugen KE, Morgan D. Virus-like peptide vaccines against Abeta N-terminal or C-terminal domains reduce amyloid deposition in APP transgenic mice without addition of adjuvant. J Neuroimmune Pharmacol 2010;5:133–42.
[21] Wiessner C, Wiederhold KH, Tissot AC, Frey P, Danner S, Jacobson LH, et al. The second-generation active Abeta immunotherapy CAD106 reduces amyloid accumulation in APP transgenic mice while minimizing potential side effects. J Neurosci 2011;31:9323–31.
[22] Pastori C, Tudor D, Diomede L, Drillet AS, Jegerlehner A, Rohn TA, et al. Virus like particle based strategy to elicit HIV-protective antibodies to the alpha-helic regions of gp41. Virology 2012;431:1–11.
[23] Spohn G, Schori C, Keller I, Sladko K, Sina C, Guler R, et al. Preclinical efficacy and safety of an anti-IL-1beta vaccine for the treatment of type 2 diabetes. Mol Ther Methods Clin Dev 2014;1:14048.
[24] Lieknina I, Kalnins G, Akopjana I, Bogans J, Sisovs M, Jansons J, et al. Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data. J Nanobiotechnology 2019;17:61.
[25] Paules CI, Fauci AS. Influenza Vaccines: Good, but We Can Do Better. J Infect Dis 2019;219:S1–4.
[26] Saelens X. The role of matrix protein 2 ectodomain in the development of universal influenza vaccines. J Infect Dis 2019;219:S68–74.
[27] Pielak RM, Chou JJ. Influenza M2 proton channels. Biochim Biophys Acta 2011;1808:522–9.
[28] Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W. A universal influenza A vaccine based on the extracellular domain of the M2 protein. Nat Med 1999;5:1157–63.
[29] Fan J, Liang X, Horton MS, Perry HC, Citron MP, Heidecker GJ, et al. Preclinical study of influenza virus A M2 peptide conjugate vaccines in mice, ferrets, and rhesus monkeys. Vaccine 2004;22:2993–3003.
[30] Turley CB, Rupp RE, Johnson C, Taylor DN, Wolfson J, Tussey L, et al. Safety and immunogenicity of a recombinant M2e-flagellin influenza vaccine (STF2.4xM2e) in healthy adults. Vaccine 2011;29:5145–52.
[31] De Vlieger D, Hoffmann K, Van Molle I, Nerinckx W, Van Hoecke L, Ballegeer M, et al. Selective engagement of fcgammaRIV by a M2e-specific single domain antibody construct protects against influenza a virus infection. Front Immunol 2019;10:2920.
[32] Kim MC, Song JM, O E, Kwon YM, Lee YJ, Compans RW, et al. Virus-like particles containing multiple M2 extracellular domains confer improved cross-protection against various subtypes of influenza virus. Mol Ther. 2013;21:485-92.
[33] Ramirez A, Morris S, Maucourant S, D'Ascanio I, Crescente V, Lu IN, et al. A virus-like particle vaccine candidate for influenza A virus based on multiple conserved antigens presented on hepatitis B tandem core particles. Vaccine 2018;36:873–80.
[34] Plevka P, Kazaks A, Voronkova T, Kotelovica S, Dishlers A, Liljas L, et al. The structure of bacteriophage phiCb5 reveals a role of the RNA genome and metal ions in particle stability and assembly. J Mol Biol 2009;391:635–47.
[35] Mastico RA, Talbot SJ, Stockley PG. Multiple presentation of foreign peptides on the surface of an RNA-free spherical bacteriophage capsid. J Gen Virol 1993;74(Pt 4):541–8.
[36] Pushko P, Kozlovskaya T, Sominskaya I, Brede A, Stankevica E, Ose V, et al. Analysis of RNA phage fr coat protein assembly by insertion, deletion and substitution mutagenesis. Protein Eng 1993;6:883–91.
[37] Peabody DS. Subunit fusion confers tolerance to peptide insertions in a virus coat protein. Arch Biochem Biophys 1997;347:85–92.
[38] Peabody DS et al. Immunogenic display of diverse peptides on virus-like particles of RNA phage MS2. J Mol Biol 2008;380:252–63.
[39] Vasiljeva I, Kozlovska T, Cielens I, Strelnikova A, Kazaks A, Ose V, et al. Mosaic Qbeta coats as a new presentation model. FEBS Lett 1998;431:7–11.
[40] Shishovs M, Rumnieks J, Diebolder C, Jaudzems K, Andreas LB, Stanek J, et al. Structure of AP205 Coat Protein Reveals Circular Permutation in ssRNA Bacteriophages. J Mol Biol 2016;428:4267–79.
[41] Tissot AC, Renhofa R, Schmitz N, Cielens I, Meijerink E, Ose V, et al. Versatile virus-like particle carrier for epitope based vaccines. PLoS ONE 2010;5:e9809.
[42] Thrane S, Janitzek CM, Matondo S, Resende M, Gustavsson T, de Jongh WA, et al. Bacterial superglue enables easy development of efficient virus-like particle based vaccines. J Nanobiotechnology 2016;14:30.

8. DISCUSSION

The discovery and subsequent studies of ssRNA phages have provided an invaluable contribution to understanding the phage biology, and to the development of molecular biology in general. However, so far, search for new viable ssRNA phages has resulted in relatively few cultivated specimens. Only a few of the vast numbers that actually exist in nature have been discovered, isolated and studied. A significant obstacle in isolation of novel living ssRNA phages might be unavailability of culturable hosts. However, owing to modern sequencing capabilities, huge databases of metagenomic sequences are being created, which, almost like the Leeuwenhoek's microscope in the 17th century, have opened up a new world. Therefore, one possibility to study yet unavailable phages is to obtain their morphological twins - VLPs. The simplicity of the genome structure of ssRNA phages, as well as the universal presence of conservative replicase gene allows the desired CP sequences to be extracted from metagenome databases.

Although the produced VLPs lack the natural components of actual phages – genomic RNA and maturation protein, these artificial structures still provide a valuable information about diversity of CP folds, particle shapes and RNA binding properties. Most importantly, the obtained CPs and VLPs can be further used as valuable tools in vaccine development, drug delivery, imaging, and other applications.

In this study, more than 120 new ssRNA phage CPs were analyzed with the main aim of finding new candidates for the development of VLPs as carriers. In order to become a versatile carrier, candidate VLP must possess a number of essential properties:

1. Stable and high expression of the CP gene in an easy scalable expression system - preferably *E.coli*.
2. The CP should self-assemble into VLPs of well-defined size and shape.
3. VLPs should be able to pack the nucleic acids inside particles.
4. The VLPs must be stable, resistant to robust purification steps and long-term storage.
5. VLPs should be tolerant to chemical and genetic modifications.

VLPs particles from several non-coliphage CPs have been previously successfully produced in *E. coli* (Kazaks et al. 2011; Shishovs et al. 2016), as well as coliphage-derived VLPs have been produced in yeast (Freivalds et al. 2006; Freivalds et al. 2014) and other expression systems (Arevalo et al. 2016). In our case the synthesis of new putative phage CPs was performed using T7 promoter regulated pET plasmid in *E.coli*, as this system is well characterized, robust, tolerant for different temperatures and easy upscaled. The chosen strategy has justified itself, since almost all of the selected genes were expressed and CPs produced. However, the solubility of the synthesized CPs proved to be the first setback, as approximately 40% of the tested proteins were insoluble in the primary screening. Because the sources of metagenomic datasets vary enormously, from the intestinal contents of warm-blooded animals to cold deep-sea microbial sediments, it is conceivable that, for at least some proteins, synthesis at 37 °C interferes with the correct conformation of the CP. This assumption seemed to be correct, because lowering the temperature for CP synthesis did indeed largely solve the solubility problems.

In nature, in many simple viruses capsid formation occurs without an external energy source or special assistance from the host cell or dedicated scaffolding proteins. Under the right conditions, capsids may be spontaneously formed by CPs only due to the CP-CP and CP-nucleic acid interactions. However, as our study progressed, it was further established that the solubility of the CP does not necessarily lead to VLP formation. The formation or

stability of these particles may be affected by some external conditions, such as appropriate salt concentration or pH. It cannot be ruled out that in some cases the chosen gene was not a CP gene or contained sequencing errors, leading to assembly-deficient CPs. In the natural environment, the formation of infectious virions is a more complex process involving other mechanisms (Twarock et al. 2018). The expression system can affect the result as well as some unknown host-factor might be involved in the formation of the particles absent in *E.coli*. Also, in some cases presence of other phage components, such as genome or A protein might be required for particle formation. The absence of these components can in some cases affect the formation of artificial particles. Although each potential VLP might have some interesting characteristics that would make it worthwhile to improve the methods for obtaining them, they are not particularly suitable for biotechnological purposes due to the lack of a number of functional features that an excellent candidate for a vaccine carrier must have.

VLPs, suitable for downstream biotechnological processing should be easily purified using the classic methods - precipitation with PEG or salts, gel filtration and ion exchange chromatography. In most cases purification of new VLPs was straightforward and was able to yield at least 90% pure preparation. Due to the positive charge of the particle surface, several VLPs did not adsorb on the anion exchange matrix but passed through. The purity of such particles was the highest. Purification by anion exchange chromatography can in some way characterize the particles. Owing to this method, one of obtained VLPs (AVE015) was discovered, where there were two types of particles in the population - empty and filled with nucleic acid, and they were easy to separate. Such empty particles could be of interest as carriers for some substances that do not compete with nucleic acids. Nucleic acids are usually packaged in artificial particles on a charge basis - negatively charged nucleic acids bind to positively charged amino acid residues that form the inner surface of the particles (Nooraei et al. 2021). The charge differences between the various VLPs could be judged from the primary screening on the native agarose gel, where in some cases particles migrated towards the anode. In previous studies with Qβ VLP I have observed a similar observation when particles with a point mutation in the CP were detected (unpublished data). These particles also migrated towards the anode. Positively charged particles could be promising biotechnological tools worth further investigation.

VLPs still are considered promising candidates for vaccine technology. By the end of March this year, in the National Library of Medicine (PubMed) more than a hundred scientific publications on the use of VLPs in vaccine technology can be found. In 2020, there were more than 300 such publications. This year, several scientific articles have been published about use AP205 and Qβ VLPs for development anti-SARS-CoV-2 vaccines and diagnosis (Chan et al. 2021, Dalvie et al. 2021, Fougeroux et al. 2021).

VLP-based vaccine technology has several advantages and good prospects for its widespread use against various infectious or systemic diseases, with only a limited number of suitable VLP platforms currently available. VLPs derived from ssRNA phages are the most applicable candidates for the presentation of foreign epitopes. In this study, 43 novel ssRNA bacteriophage CPs were characterized by their ability to form chimeric VLPs. The study was performed with a model peptide that had already proven to be a successful candidate for insertions in other VLPs. It should be noted that the obtained model will not be universal, because proteins or their parts of other origins can affect both the synthesis of the chimeric protein and the ability to assemble in regular structures. Also, it could happen that a particular antigen can be easily presented on the surface of one subset of VLPs, while other VLPs might be better suited for presentation of other antigens. In this scenario a set of

available different VLPs would be advantageous, allowing for individual tailoring of best VLPs for each antigen.

As mentioned before, when comparing the CPs of the new putative phages with known ssRNA phages, most are similar to MS2. Also, the majority (17) of CPs selected for genetic insertions represented this MS2-like group. MS2 itself and its close relative fr VLPs were among the first particles to be used in the construction of chimeric VLPs (Mastico et al. 1993) (Pushko et al. 1993). Unfortunately, these particles did not tolerate long foreign epitope insertions and their use was limited. In MS2 VLPs, insertions have been performed in the AB loop. Although the loop is flexible and located on the surface, it is tolerant only of inserts up to 10 amino acids. To improve the stability and capacity for foreign inserts of the MS2 VLP chimeric capsid, a single-chain covalent dimer has been constructed (Peabody 1997; Peabody et al. 2008). In such a more stable structure, it became possible to modify the N-terminal end. In this model, epitopes can be inserted at both the loop and the end (Peabody et al. 2017). However, testing of new CPs from MS2-like group revealed that their performance as VLP carriers is insufficient. None of the CP of this group was able to form chimeric VLPs, although the production and solubility of the chimeric proteins were adequate.

Qβ VLPs is the most widely used carrier in vaccine design. However, the biggest advantages of Qβ are apparent by use of chemical coupling. The phage Qβ is unique among the other ssRNA phages in that it naturally contains several copies of a C-terminally elongated CP A1 incorporated in the particle, and therefore the CP has a naturally evolved ability to tolerate C-terminal extensions (Vasiljeva et al. 1998). To create such chimeric-mosaic particles the stop codon suppression approach is used. The main disadvantage of this method are resulting heterogeneous particles with irregular arrangement of foreign epitopes. Furthermore, efficient incorporation rate of long antigens is problematic. Therefore, our results strengthen the overall idea that MS2-like sequences have limited potential as carriers in vaccine development. Reasons of why MS2-like CP sequences are likely unsuitable for genetic modification in N- and C-termini can be seen in the structural analysis of unmodified VLPs from MS2-like group. Both termini of these particles only border the surface and are clustered together around symmetry axes, creating steric problems for long antigen exposure (Shishovs et al. 2016).

Modification of the CPs of the Beihai32-like group indicates a distant similarity of this group to the AP205-like group. Both CPs chosen from this group were tolerant of insertion at the C-terminus with subsequent formation of chimeric particles. In addition, Beihai32 C-terminal fusion displayed the highest antigenicity of all tested constructs. Comparing the antigenicity with the other chimeric particle of this group, Beihai28 showed the weakest reactivity with specific antibodies. Structural analysis of Beihai32 revealed that the C-terminus of this phage extends far from the surface of the VLP, making any inserted antigen highly accessible, while the C-terminus of Beihai28 is 10 residues shorter, possibly affecting sufficient surface exposure of the introduced foreign antigen. Also, the chimeric Beihai28 VLP is unstable, which could be due to model peptide interference with the integrity of the capsid.

AP205 and AP205-like phage CPs and their VLPs stand out as the best candidates in the development of a platform as vaccine carriers. Since the introduction of the characterized phage AP205 VLPs, it has been evaluated as an excellent candidate for genetic modification at both the N- and C-termini. These particles also stand out due to their structural stability because they are stabilized by disulfide bonds. Such tolerance to genetic fusions had not been observed for any of the other known VLPs of ssRNA phages (Tissot et al. 2010, Thrane

et al. 2016). The tolerance of AP205 VLPs to genetic fusions can be easily explained from a structural point of view; both the C- and N-termini are located on the surface of the VLP and well separated in neighboring CP subunits (Shishovs et al. 2016). Five of the six modified AP205-like group phage CPs were able to form chimeric VLPs and all five were tolerant to both N- and C-terminal fusions. Although the outstanding properties of the AP205 CP in this regard are well documented, this study confirms that such properties are a universal feature of almost the whole group. The AP205-like CPs used in the study have only 15-30% sequence identity with the AP205 CP and are less than 37% identical to each other. Therefore, it seems reasonable to assume that in addition to the five new CPs described in this study, there could be a significant amount of sufficiently distinct but as yet undiscovered CPs with equally good properties to develop chimeric VLPs.

Five CP sequences were selected from the AVE015-like group for genetic modification. Only one of them, GA-879 CP with an insert at the C-terminus, formed chimeric VLPs. These results suggest that AVE015-like VLPs are not very promising candidates for vaccine design using terminal modification. However, the analysis of the crystal structure of the phage AVE015 of this group revealed one interesting feature: especially long loops between β-strands E and F and between the C-terminal α helices. The interhelix loop is particularly pronounced and forms a prominent surface-exposed β-hairpin. This feature could be used to introduce genetic insertions in surface exposed loops.

In the first study, the Cb5-like CP group was revealed as the second largest; however, only four CP sequences were used in the modification study. The main reason why such a small number of candidates was selected from the large potential range of VLPs is that less than half of the previously tested Cb5-like CPs were able to form VLPs. Some of the native particles were unstable and technologically unpromising due to external impurities. Of the three Cb5-like proteins used in this study (Beihai9, EMS011, and Wenzhou2), none was able to form chimeric VLPs. Of the remaining CP similarity groups, none of the fusion CPs of the AVE002-, ESE020-, ESE017-, and ESO003-like groups were able to assemble into particles, which means that the CPs in these groups are generally not very promising for future development as epitope carriers either.

One of the two tested AC-like CPs was able to form chimeric VLPs. NF-391 CP was tolerant only for C-terminal modifications. This case is special and worth highlighting, because both the unmodified CP and the chimeric ones assembled into the smallest possible T=1 symmetry particles.

The Beihai14 CP is the only one of its kind and is unique among others due to its extraordinary length - 208 amino acids. On average, the CP lengths of ssRNA phages are about 130 amino acids. This phage also has an exceptionally high VLP thermal stability (95 °C), despite the lack of stabilizing disulfide bonds. This CP is tolerant to C-terminal modifications with subsequent formation of chimeric particles. Chimeric particles also displayed extraordinary thermostability, which is an important property for potential vaccine candidates. The crystal structure was also solved for this phage, which later revealed the unique properties of this VLP. The fold of Beihai CP was by far the most extraordinary. For this phage, the classical CP structure has been reduced by one β-strand and only one C-terminal α helix was present, while an additional 10-residue α helix was inserted in the FG loop. Apart from these peculiarities, several other atypical elements were observed, like very long unstructured N-terminal part, adopting different conformation in A, B and C subunits and harboring many positively charged residues – features, typical for some completely unrelated plant viruses. Undoubtedly, such structural properties can affect the stability of the capsid.

The results of the study indicated that most chimeric VLPs with M2ex3 peptide contain Cys residues, which can form inter-subunit disulfide bonds. For the AP205-like group, cysteine residues are a conservative feature and obviously have a significant effect on the overall stability of the particles, as evidenced by a decrease in thermal stability at a temperature of at least 20 °C in the presence of a reducing agent. Disulfide bonds are also present in two other groups of CPs: AC-like NF-391-C VLP and AVE015-like GA-879-C. Thermal stability data suggest that the stability of these particles is only slightly affected by the reducing agent. Apparently, for these particles, as well as for Beihai14, stability is provided by non-covalent subunit interactions. Therefore, while useful, disulfides do not appear to be sufficient or even necessary for formation a stable and robust carrier.

The observed differences in antigenicity between different chimeric VLPs reflect the availability of the exposed M2ex3 peptide for antibodies. In the case of chimeric VLPs, antigenicity is expected to positively correlate with immunogenicity, because adequate stimulation of the immune system also requires effective exposure of foreign epitopes on the surface of the particle.

Combining all results obtained by discovering numerous new VLPs, identification of previously unseen structures of CPs, genetic modifications to obtain chimeric VLPs that expand the range of carriers used in biotechnology, it can be argued that this massive study has made a significant contribution to field on VLP-based recombinants.

9. CONCLUSIONS

1. Most sequences of the hypothetical ssRNA bacteriophage CPs obtained from metagenomic databases can be efficiently expressed in *E. coli*.
2. The cultivation temperature of *E.coli* affects the solubility of CPs sourced from more extreme environments.
3. Out of 110 trials, 80 CPs were able to assemble in VLPs, as confirmed by electron microscopy.
4. Most of the new VLPs have apparent spherical shape 28 to 30 nm in diameter, corresponding to T = 3 icosahedral particles.
5. Not all VLPs containing Cys residues were involved in classical pentameric and hexameric interactions.
6. Inter-subunit disulfide bonds are neither sufficient nor necessary for formation of thermally stable ssRNA phage VLPs.
7. The widely represented MS2-like group and AP205-like group revealed structures similar to the phages of this group previously studied. Particles from the AP205 group has a good potential as foreign epitope carriers.
8. Five CPs - ESE002, PQ-465, PQ-061, PQ-789 and Shahe1 - were able to form VLPs with M2ex3 fused at either terminus; five CPs - Beihai14, Beihai28, Beihai32, GA-879 and NF-391 - the peptide could be fused to the C-terminus.
9. All chimeric VLPs have required antigenicity.
10. The solved crystal structures revealed several previously unseen features of ssRNA phages such as additional secondary structure elements and elongated particles.

10. MAIN THESIS OF DEFENSE

1. The CP genes of unknown ssRNA bacteriophages can be found in metagenomic sequence databases.
2. It is possible to study properties and potential applications of putative ssRNA bacteriophages by obtaining VLPs formed from recombinant CPs.
3. Some VLPs of putative ssRNA bacteriophage possess unseen secondary structure elements of CPs and particles of different shapes.
4. VLPs of new ssRNA bacteriophages open up new and vast possibilities in vaccine design and biotechnology.

11. PUBLICATIONS

1. **“Production and characterization of novel ssRNA bacteriophage virus-like particles from metagenomic sequencing data”**

Ilva Liekniņa, Gints Kalniņš, Ināra Akopjana, Jānis Bogans, Mihails Šišovs, Juris Jansons, Jānis Rūmnieks, Kaspars Tārs. *J Nanobiotechnology*, 2019 May 13;17(1):61. PMID: 31084612. PMCID: PMC6513524. DOI: 10.1186/s12951-019-0497-8

2. **“Three-dimensional structure of 22 uncultured ssRNA bacteriophages: Flexibility of the coat protein fold and variations in particle shapes”**

Jānis Rūmnieks, Ilva Liekniņa, Gints Kalniņš, Mihails Šišovs, Ināra Akopjana, Jānis Bogans, Kaspars Tārs. *Science Advances*, 2020 Sep 2;6(36):eabc0023. PMID: 32917600. PMCID: PMC7467689. DOI: 10.1126/sciadv.abc0023

3. **“Novel ssRNA phage VLP platform for displaying foreign epitopes by genetic fusion”**

Ilva Liekniņa, Darja Černova, Jānis Rūmnieks, Kaspars Tārs. *Vaccine*, 2020 Aug 27;38(38):6019-6026. PMID: 32713683. DOI: 10.1016/j.vaccine.2020.07.016

12. APPROBATION OF THE RESEARCH

1. I.Liekniņa, D.Černova, J.Rūmnieks, K.Tārs. FEBS 3+ Starptautiskā zinātniskā konference 2019.g. "Novel VLPs of ssRNA phages for vaccine development". Stenda referāts.
2. I. Liekniņa. 2019. "Novel VLPs of ssRNA phages for vaccine development", BMC un OSI starpinstitucionālais seminārs.
3. I. Liekniņa, D. Černova, J. Rūmnieks, K. Tārs. 2020. g. LU 78. konference. "Novel ssRNA phage VLP platform for displaying foreign epitopes by genetic fusion" 2020

13. ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the group leader Kaspars Tārs for the almost endless and patient support of my work and the opportunity to be employed in his research group. I thank all my colleagues from the BMC Structural Biology group, especially Jānis Rūmnieks for his ideas and response, Zane Alsberga and Darja Černova for selfless work in search of vaccine candidates. I thank to all co-workers, who helped work-out this research: Gints Kalniņš, Mihails Šišovs, Jānis Bogans, Ināra Akopjana, Juris Jansons.

Many thanks to my first teachers – Ivars Petrovskis and Andris Dishlers, under whose guidance I have learned many methods and gained a sense of VLPs. Many thanks also to Professor Pauls Pumpens, our "father" of VLPs.

The research was carried out within the project "Obtaining and characterization of virus-like particles of novel RNA phages" (ERDF 1.1.1.1/16/A/104).

14. REFERENCES

Aanei, I. L., T. Huynh, Y. Seo, and M. B. Francis. "Vascular Cell Adhesion Molecule-Targeted Ms2 Viral Capsids for the Detection of Early-Stage Atherosclerotic Plaques." *Bioconjug Chem* 29, no. 8 (Aug 15 2018): 2526-30.

Adhin, M. R., A. Hirashima, and J. van Duin. "Nucleotide Sequence from the Ssrna Bacteriophage Jp34 Resolves the Discrepancy between Serological and Biophysical Classification." *Virology* 170, no. 1 (May 1989): 238-42.

Alexopoulou, L., A. C. Holt, R. Medzhitov, and R. A. Flavell. "Recognition of Double-Stranded Rna and Activation of Nf-Kappab by Toll-Like Receptor 3." *Nature* 413, no. 6857 (Oct 18 2001): 732-8.

Arevalo, M. T., T. M. Wong, and T. M. Ross. "Expression and Purification of Virus-Like Particles for Vaccination." *J Vis Exp*, no. 112 (Jun 2 2016).

Ashley, C. E., E. C. Carnes, G. K. Phillips, P. N. Durfee, M. D. Buley, C. A. Lino, D. P. Padilla, *et al.* "Cell-Specific Delivery of Diverse Cargos by Bacteriophage Ms2 Virus-Like Particles." *ACS Nano* 5, no. 7 (Jul 26 2011): 5729-45.

Atcheson, E., G. Cabral-Miranda, A. M. Salman, and A. Reyes-Sandoval. "Discovery of Four New B-Cell Protective Epitopes for Malaria Using Q Beta Virus-Like Particle as Platform." *NPJ Vaccines* 5 (2020): 92.

Basu, R., L. Zhai, B. Rosso, and E. Tumban. "Bacteriophage Qbeta Virus-Like Particles Displaying Chikungunya Virus B-Cell Epitopes Elicit High-Titer E2 Protein Antibodies but Fail to Neutralize a Thailand Strain of Chikungunya Virus." *Vaccine* 38, no. 11 (Mar 4 2020): 2542-50.

Bernhardt, T. G., I. N. Wang, D. K. Struck, and R. Young. "A Protein Antibiotic in the Phage Qbeta Virion: Diversity in Lysis Targets." *Science* 292, no. 5525 (Jun 22 2001): 2326-9.

Bertrand, E., P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, and R. M. Long. "Localization of Ash1 Mrna Particles in Living Yeast." *Mol Cell* 2, no. 4 (Oct 1998): 437-45.

Blaisdell, B. E., A. M. Campbell, and S. Karlin. "Similarities and Dissimilarities of Phage Genomes." *Proc Natl Acad Sci U S A* 93, no. 12 (Jun 11 1996): 5854-9.

Bollback, J. P., and J. P. Huelsenbeck. "Phylogeny, Genome Evolution, and Host Specificity of Single-Stranded Rna Bacteriophage (Family Leviviridae)." *J Mol Evol* 52, no. 2 (Feb 2001): 117-28.

Bolli, E., J. P. O'Rourke, L. Conti, S. Lanzardo, V. Rolih, J. M. Christen, G. Barutello, *et al.* "A Virus-Like-Particle Immunotherapy Targeting Epitope-Specific Anti-Xct Expressed on Cancer Stem Cell Inhibits the Progression of Metastatic Cancer in Vivo." *Oncoimmunology* 7, no. 3 (2018): e1408746.

Bradley, D. E., and D. Robertson. "The Structure and Infective Process of a Contractile Pseudomonas Aeruginosa Bacteriophage." *J Gen Virol* 3, no. 2 (Sep 1968): 247-54.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer. "Genomic Analysis of Uncultured Marine Viral Communities." *Proc Natl Acad Sci U S A* 99, no. 22 (Oct 29 2002): 14250-5.

Brune, K. D., D. B. Leneghan, I. J. Brian, A. S. Ishizuka, M. F. Bachmann, S. J. Draper, S. Biswas, and M. Howarth. "Plug-and-Display: Decoration of Virus-Like Particles Via Isopeptide Bonds for Modular Immunization." *Sci Rep* 6 (Jan 19 2016): 19234.

Caldeira, J., J. Bustos, J. Peabody, B. Chackerian, and D. S. Peabody. "Epitope-Specific Anti-Hcg Vaccines on a Virus Like Particle Platform." *PLoS One* 10, no. 10 (2015): e0141407.

Callanan, J., S. R. Stockdale, A. Shkoporov, L. A. Draper, R. P. Ross, and C. Hill. "Expansion of Known Ssrna Phage Genomes: From Tens to over a Thousand." *Sci Adv* 6, no. 6 (Feb 2020): eaay5981.

Callanan, J., S. R. "Rna Phage Biology in a Metagenomic Era." *Viruses* 10, no. 7 (Jul 21 2018).

Carding, S. R., N. Davis, and L. Hoyles. "Review Article: The Human Intestinal Virome in Health and Disease." *Aliment Pharmacol Ther* 46, no. 9 (Nov 2017): 800-15.

Casale, T. B., J. Cole, E. Beck, C. F. Vogelmeier, J. Willers, C. Lassen, A. Hammann-Haenni, *et al.* "Cyt003, a Tlr9 Agonist, in Persistent Allergic Asthma - a Randomized Placebo-Controlled Phase 2b Study." *Allergy* 70, no. 9 (Sep 2015): 1160-8.

Chackerian, B., C. Caldeira Jdo, J. Peabody, and D. S. Peabody. "Peptide Epitope Identification by Affinity Selection on Bacteriophage Ms2 Virus-Like Particles." *J Mol Biol* 409, no. 2 (Jun 3 2011): 225-37.

Chackerian, B., M. Rangel, Z. Hunter, and D. S. Peabody. "Virus and Virus-Like Particle-Based Immunogens for Alzheimer's Disease Induce Antibody Responses against Amyloid-Beta without Concomitant T Cell Responses." *Vaccine* 24, no. 37-39 (Sep 11 2006): 6321-31.

Chamakura, K. R., and R. Young. "Single-Gene Lysis in the Metagenomic Era." *Curr Opin Microbiol* 56 (Aug 2020): 109-17.

Chan, S. K., P. Du, C. Ignacio, S. Mehta, I. G. Newton, and N. F. Steinmetz. "Biomimetic Virus-Like Particles as Severe Acute Respiratory Syndrome Coronavirus 2 Diagnostic Tools." *ACS Nano* 15, no. 1 (Jan 26 2021): 1259-72.

Chao, J. A., Y. Patskovsky, S. C. Almo, and R. H. Singer. "Structural Basis for the Coevolution of a Viral Rna-Protein Complex." *Nat Struct Mol Biol* 15, no. 1 (Jan 2008): 103-5.

Cielens, I., L. Jackevica, A. Strods, A. Kazaks, V. Ose, J. Bogans, P. Pumpens, and R. Renhofa. "Mosaic Rna Phage Vlps Carrying Domain Iii of the West Nile Virus E Protein." *Mol Biotechnol* 56, no. 5 (May 2014): 459-69.

Clokie, M. R., A. D. Millard, A. V. Letarov, and S. Heaphy. "Phages in Nature." *Bacteriophage* 1, no. 1 (Jan 2011): 31-45.

Cohen, B. A., and M. Bergkvist. "Targeted in Vitro Photodynamic Therapy Via Aptamer-Labeled, Porphyrin-Loaded Virus Capsids." *J Photochem Photobiol B* 121 (Apr 5 2013): 67-74.

Crossey, E., M. J. A. Amar, M. Sampson, J. Peabody, J. T. Schiller, B. Chackerian, and A. T. Remaley. "A Cholesterol-Lowering Vlp Vaccine That Targets Pcsk9." *Vaccine* 33, no. 43 (Oct 26 2015): 5747-55.

Crossey, E., K. Frietze, D. L. Narum, D. S. Peabody, and B. Chackerian. "Identification of an Immunogenic Mimic of a Conserved Epitope on the Plasmodium Falciparum Blood Stage Antigen Ama1 Using Virus-Like Particle (Vlp) Peptide Display." *PLoS One* 10, no. 7 (2015): e0132560.

Cui, Z., K. V. Gorzelnik, J. Y. Chang, C. Langlais, J. Jakana, R. Young, and J. Zhang. "Structures of Qbeta Virions, Virus-Like Particles, and the Qbeta-Mura Complex Reveal Internal Coat Proteins and the Mechanism of Host Lysis." *Proc Natl Acad Sci U S A* 114, no. 44 (Oct 31 2017): 11697-702.

Culley, A. I., A. S. Lang, and C. A. Suttle. "Metagenomic Analysis of Coastal Rna Virus Communities." *Science* 312, no. 5781 (Jun 23 2006): 1795-8.

Dai, X., Z. Li, M. Lai, S. Shu, Y. Du, Z. H. Zhou, and R. Sun. "In Situ Structures of the Genome and Genome-Delivery Apparatus in a Single-Stranded Rna Virus." *Nature* 541, no. 7635 (Jan 5 2017): 112-16.

Dalvie, N. C., S. A. Rodriguez-Aponte, B. L. Hartwell, L. H. Tostanoski, A. M. Biedermann, L. E. Crowell, K. Kaur, *et al.* "Engineered Sars-Cov-2 Receptor Binding Domain Improves Immunogenicity in Mice and Elicits Protective Immunity in Hamsters." *bioRxiv* (Mar 4 2021).

Danziger, R. E., and W. Paranchych. "Stages in Phage R17 Infection. Ii. Ionic Requirements for Phage R17 Attachment to F-Pili." *Virology* 40, no. 3 (Mar 1970): 547-53.

Delwart, E. L. "Viral Metagenomics." *Rev Med Virol* 17, no. 2 (Mar-Apr 2007): 115-31.

Dent, K. C., R. Thompson, A. M. Barker, J. A. Hiscox, J. N. Barr, P. G. Stockley, and N. A. Ranson. "The Asymmetric Structure of an Icosahedral Virus Bound to Its Receptor Suggests a Mechanism for Genome Release." *Structure* 21, no. 7 (Jul 2 2013): 1225-34.

Dong, Y. M., J. C. Cai, H. T. Chen, and L. Chen. "Protection of a Novel Epitope-Rna Vlp Double-Effective Vlp Vaccine for Foot-and-Mouth Disease." *Antiviral Res* 134 (Oct 2016): 108-16.

Dong, Y. M., G. G. Zhang, X. J. Huang, L. Chen, and H. T. Chen. "Promising Ms2 Mediated Virus-Like Particle Vaccine against Foot-and-Mouth Disease." *Antiviral Res* 117 (May 2015): 39-43.

Doucet, M., A. El-Turabi, F. Zabel, B. H. M. Hunn, N. Bengoa-Vergniory, M. Cioroch, M. Ramm, *et al.* "Preclinical Development of a Vaccine against Oligomeric Alpha-Synuclein Based on Virus-Like Particles." *PLoS One* 12, no. 8 (2017): e0181844.

ElSohly, A. M., J. I. MacDonald, N. B. Hentzen, I. L. Aanei, K. M. El Muslemany, and M. B. Francis. "Ortho-Methoxyphenols as Convenient Oxidative Bioconjugation Reagents with Application to Site-Selective Heterobifunctional Cross-Linkers." *J Am Chem Soc* 139, no. 10 (Mar 15 2017): 3767-73.

ElSohly, A. M., C. Netirojjanakul, I. L. Aanei, A. Jager, S. C. Bendall, M. E. Farkas, G. P. Nolan, and M. B. Francis. "Synthetically Modified Viral Capsids as Versatile Carriers for Use in Antibody-Based Cell Targeting." *Bioconjug Chem* 26, no. 8 (Aug 19 2015): 1590-6.

Escolano, A., H. B. Gristick, M. E. Abernathy, J. Merkenschlager, R. Gautam, T. Y. Oliveira, J. Pai, *et al.* "Immunization Expands B Cells Specific to Hiv-1 V3 Glycan in Mice and Macaques." *Nature* 570, no. 7762 (Jun 2019): 468-73.

Fiers, W., R. Contreras, F. Duerinck, G. Haegeman, D. Iserentant, J. Merregaert, W. Min Jou, *et al.* "Complete Nucleotide Sequence of Bacteriophage Ms2 Rna: Primary and Secondary Structure of the Replicase Gene." *Nature* 260, no. 5551 (Apr 8 1976): 500-7.

Fougeroux, C., L. Goksoyr, M. Idorn, V. Soroka, S. K. Myeni, R. Dagil, C. M. Janitzek, *et al.* "Capsid-Like Particles Decorated with the Sars-Cov-2 Receptor-Binding Domain Elicit Strong Virus Neutralization Activity." *Nat Commun* 12, no. 1 (Jan 12 2021): 324.

Freivalds, J., A. Dislers, V. Ose, D. Skrastina, I. Cielens, P. Pumpens, K. Sasnauskas, and A. Kazaks. "Assembly of Bacteriophage Qbeta Virus-Like Particles in Yeast Saccharomyces Cerevisiae and Pichia Pastoris." *J Biotechnol* 123, no. 3 (May 29 2006): 297-303.

Freivalds, J., S. Kotelovica, T. Voronkova, V. Ose, K. Tars, and A. Kazaks. "Yeast-Expressed Bacteriophage-Like Particles for the Packaging of Nanomaterials." *Mol Biotechnol* 56, no. 2 (Feb 2014): 102-10.

Frietze, K. M., D. S. Peabody, and B. Chackerian. "Engineering Virus-Like Particles as Vaccine Platforms." *Curr Opin Virol* 18 (Jun 2016): 44-9.

Frietze, K. M., R. B. Roden, J. H. Lee, Y. Shi, D. S. Peabody, and B. Chackerian. "Identification of Anti-Ca125 Antibody Responses in Ovarian Cancer Patients by a Novel Deep Sequence-Coupled Biopanning Platform." *Cancer Immunol Res* 4, no. 2 (Feb 2016): 157-64.

Fu, Y., and J. Li. "A Novel Delivery Platform Based on Bacteriophage Ms2 Virus-Like Particles." *Virus Res* 211 (Jan 4 2016): 9-16.

Fuenmayor, J., F. Godia, and L. Cervera. "Production of Virus-Like Particles for Vaccines." *N Biotechnol* 39, no. Pt B (Oct 25 2017): 174-80.

Goessens, W. H., A. J. Driessen, J. Wilschut, and J. van Duin. "A Synthetic Peptide Corresponding to the C-Terminal 25 Residues of Phage Ms2 Coded Lysis Protein Dissipates the Protonmotive Force in Escherichia Coli Membrane Vesicles by Generating Hydrophilic Pores." *EMBO J* 7, no. 3 (Mar 1988): 867-73.

Golmohammadi, R., K. Fridborg, M. Bundule, K. Valegard, and L. Liljas. "The Crystal Structure of Bacteriophage Q Beta at 3.5 a Resolution." *Structure* 4, no. 5 (May 15 1996): 543-54.

Gorzelnik, K. V., Z. Cui, C. A. Reed, J. Jakana, R. Young, and J. Zhang. "Asymmetric Cryo-Em Structure of the Canonical Allolevivirus Qbeta Reveals a Single Maturation Protein and the Genomic Ssrna in Situ." *Proc Natl Acad Sci U S A* 113, no. 41 (Oct 11 2016): 11519-24.

Gralla, J., J. A. Steitz, and D. M. Crothers. "Direct Physical Evidence for Secondary Structure in an Isolated Fragment of R17 Bacteriophage Mrna." *Nature* 248, no. 445 (Mar 15 1974): 204-8.

Greninger, A. L., and J. L. DeRisi. "Draft Genome Sequences of Leviviridae Rna Phages Ec and Mb Recovered from San Francisco Wastewater." *Genome Announc* 3, no. 3 (Jun 25 2015).

Groeneveld, H., F. Oudot, and J. V. van Duin. "Rna Phage Ku1 Has an Insertion of 18 Nucleotides in the Start Codon of Its Lysis Gene." *Virology* 218, no. 1 (Apr 1 1996): 141-7.

Hatfull, G. F. "Dark Matter of the Biosphere: The Amazing World of Bacteriophage Diversity." *J Virol* 89, no. 16 (Aug 2015): 8107-10.

Heil, F., H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, and S. Bauer. "Species-Specific Recognition of Single-Stranded Rna Via Toll-Like Receptor 7 and 8." *Science* 303, no. 5663 (Mar 5 2004): 1526-9.

Hemmi, H., T. Kaisho, O. Takeuchi, S. Sato, H. Sanjo, K. Hoshino, T. Horiuchi, *et al.* "Small Anti-Viral Compounds Activate Immune Cells Via the Tlr7 Myd88-Dependent Signaling Pathway." *Nat Immunol* 3, no. 2 (Feb 2002): 196-200.

Herbert, F. C., O. R. Brohlin, T. Galbraith, C. Benjamin, C. A. Reyes, M. A. Luzuriaga, A. Shahrivarkevishahi, and J. J. Gassensmith. "Supramolecular Encapsulation of Small-Ultrared Fluorescent Proteins in Virus-Like Nanoparticles for Noninvasive in Vivo Imaging Agents." *Bioconjug Chem* 31, no. 5 (May 20 2020): 1529-36.

Hillman, B. I., and G. Cai. "The Family Narnaviridae: Simplest of Rna Viruses." *Adv Virus Res* 86 (2013): 149-76.

Hu, X., Y. Deng, X. Chen, Y. Zhou, H. Zhang, H. Wu, S. Yang, *et al.* "Immune Response of a Novel Atr-Ap205-001 Conjugate Anti-Hypertensive Vaccine." *Sci Rep* 7, no. 1 (Oct 3 2017): 12580.

Huang, X., X. Wang, J. Zhang, N. Xia, and Q. Zhao. "Escherichia Coli-Derived Virus-Like Particles in Vaccine Development." *NPJ Vaccines* 2 (2017): 3.

Huber, M., W. C. Olson, and A. Trkola. "Antibodies for Hiv Treatment and Prevention: Window of Opportunity?". *Curr Top Microbiol Immunol* 317 (2008): 39-66.

Hunter, Z., E. Tumban, A. Dziduszko, and B. Chackerian. "Aerosol Delivery of Virus-Like Particles to the Genital Tract Induces Local and Systemic Antibody Responses." *Vaccine* 29, no. 28 (Jun 20 2011): 4584-92.

Inokuchi, Y., A. B. Jacobson, T. Hirose, S. Inayama, and A. Hirashima. "Analysis of the Complete Nucleotide Sequence of the Group Iv Rna Coliphage Sp." *Nucleic Acids Res* 16, no. 13 (Jul 11 1988): 6205-21.

Yin, Z., S. Dulaney, C. S. McKay, C. Baniel, K. Kaczanowska, S. Ramadan, M. G. Finn, and X. Huang. "Chemical Synthesis of Gm2 Glycans, Bioconjugation with Bacteriophage Qbeta, and the Induction of Anticancer Antibodies." *Chembiochem* 17, no. 2 (Jan 2016): 174-80.

Jennings, G. T., and M. F. Bachmann. "The Coming of Age of Virus-Like Particle Vaccines." *Biol Chem* 389, no. 5 (May 2008): 521-36.

Kazaks, A., T. Voronkova, J. Rumnieks, A. Dishlers, and K. Tars. "Genome Structure of Caulobacter Phage Phicb5." *J Virol* 85, no. 9 (May 2011): 4628-31.

Kirsteina, A., I. Akopjana, J. Bogans, I. Lieknina, J. Jansons, D. Skrastina, T. Kazaka, *et al.* "Construction and Immunogenicity of a Novel Multivalent Vaccine Prototype Based on Conserved Influenza Virus Antigens." *Vaccines (Basel)* 8, no. 2 (Apr 24 2020).

Klovins, J., G. P. Overbeek, S. H. E. van den Worm, H. W. Ackermann, and J. van Duin. "Nucleotide Sequence of a Ssrna Phage from Acinetobacter: Kinship to Coliphages." *J Gen Virol* 83, no. Pt 6 (Jun 2002): 1523-33.

Koning, R. I., J. Gomez-Blanco, I. Akopjana, J. Vargas, A. Kazaks, K. Tars, J. M. Carazo, and A. J. Koster. "Asymmetric Cryo-Em Reconstruction of Phage Ms2 Reveals Genome Structure in Situ." *Nat Commun* 7 (Aug 26 2016): 12524.

Krishnamurthy, S. R., A. B. Janowski, G. Zhao, D. Barouch, and D. Wang. "Hyperexpansion of Rna Bacteriophage Diversity." *PLoS Biol* 14, no. 3 (Mar 2016): e1002409.

Kundig, T. M., G. Senti, G. Schnetzler, C. Wolf, B. M. Prinz Vavricka, A. Fulurija, F. Hennecke, *et al.* "Der P 1 Peptide on Virus-Like Particles Is Safe and Highly Immunogenic in Healthy Adults." *J Allergy Clin Immunol* 117, no. 6 (Jun 2006): 1470-6.

Kushnir, N., S. J. Streatfield, and V. Yusibov. "Virus-Like Particles as a Highly Efficient Vaccine Platform: Diversity of Targets and Production Systems and Advances in Clinical Development." *Vaccine* 31, no. 1 (Dec 17 2012): 58-83.

Lagoutte, P., C. Mignon, G. Stadthagen, S. Potisopon, S. Donnat, J. Mast, A. Lugari, and B. Werle. "Simultaneous Surface Display and Cargo Loading of Encapsulin Nanocompartments and Their Use for Rational Vaccine Design." *Vaccine* 36, no. 25 (Jun 14 2018): 3622-28.

Larson, D. R., D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer. "Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene." *Science* 332, no. 6028 (Apr 22 2011): 475-8.

Li, Q. Y., M. N. Gordon, B. Chackerian, J. Alamed, K. E. Ugen, and D. Morgan. "Virus-Like Peptide Vaccines against Abeta N-Terminal or C-Terminal Domains Reduce Amyloid Deposition in App Transgenic Mice without Addition of Adjuvant." *J Neuroimmune Pharmacol* 5, no. 1 (Mar 2010): 133-42.

Liljas, L., K. Fridborg, K. Valegard, M. Bundule, and P. Pumpens. "Crystal Structure of Bacteriophage Fr Capsids at 3.5 a Resolution." *J Mol Biol* 244, no. 3 (Dec 2 1994): 279-90.

Lu, I. N., A. Kirsteina, S. Farinelle, S. Willieme, K. Tars, C. P. Muller, and A. Kazaks. "Structure and Applications of Novel Influenza Ha Tri-Stalk Protein for Evaluation of Ha Stem-Specific Immunity." *PLoS One* 13, no. 9 (2018): e0204776.

Manolova, V., A. Flace, M. Bauer, K. Schwarz, P. Saudan, and M. F. Bachmann. "Nanoparticles Target Distinct Dendritic Cell Populations According to Their Size." *Eur J Immunol* 38, no. 5 (May 2008): 1404-13.

Maphis, N. M., J. Peabody, E. Crossey, S. Jiang, F. A. Jamaleddin Ahmad, M. Alvarez, S. K. Mansoor, *et al.* "Qss Virus-Like Particle-Based Vaccine Induces Robust Immunity and Protects against Tauopathy." *NPJ Vaccines* 4 (2019): 26.

Marcinkiewicz, A. L., I. Lieknina, X. Yang, P. L. Lederman, T. M. Hart, J. Yates, W. H. Chen, *et al.* "The Factor H-Binding Site of Cspz as a Protective Target against Multistrain, Tick-Transmitted Lyme Disease." *Infect Immun* 88, no. 5 (Apr 20 2020).

Mastico, R. A., S. J. Talbot, and P. G. Stockley. "Multiple Presentation of Foreign Peptides on the Surface of an Rna-Free Spherical Bacteriophage Capsid." *J Gen Virol* 74 (Pt 4) (Apr 1993): 541-8.

Miyake, T., I. Haruna, T. Shiba, Y. H. Ito, and K. Yamane. "Grouping of Rna Phages Based on the Template Specificity of Their Rna Replicases." *Proc Natl Acad Sci U S A* 68, no. 9 (Sep 1971): 2022-4.

Mogus, A. T., L. Liu, M. Jia, D. T. Ajayi, K. Xu, R. Kong, J. Huang, *et al.* "Virus-Like Particle Based Vaccines Elicit Neutralizing Antibodies against the Hiv-1 Fusion Peptide." *Vaccines (Basel)* 8, no. 4 (Dec 15 2020).

Mohsen, M. O., A. C. Gomes, G. Cabral-Miranda, C. C. Krueger, F. M. Leoratti, J. V. Stein, and M. F. Bachmann. "Delivering Adjuvants and Antigens in Separate Nanoparticles Eliminates the Need of Physical Linkage for Effective Vaccination." *J Control Release* 251 (Apr 10 2017): 92-100.

Mohsen, M. O., L. Zha, G. Cabral-Miranda, and M. F. Bachmann. "Major Findings and Recent Advances in Virus-Like Particle (Vlp)-Based Vaccines." *Semin Immunol* 34 (Dec 2017): 123-32.

Mokili, J. L., F. Rohwer, and B. E. Dutilh. "Metagenomics and Future Perspectives in Virus Discovery." *Curr Opin Virol* 2, no. 1 (Feb 2012): 63-77.

Nooraei, S., H. Bahrulolum, Z. S. Hoseini, C. Katalani, A. Hajizade, A. J. Easton, and G. Ahmadian. "Virus-Like Particles: Preparation, Immunogenicity and Their Roles as Nanovaccines and Drug Nanocarriers." *J Nanobiotechnology* 19, no. 1 (Feb 25 2021): 59.

Olsthoorn, R. C., G. Garde, T. Dayhuff, J. F. Atkins, and J. Van Duin. "Nucleotide Sequence of a Single-Stranded Rna Phage from Pseudomonas Aeruginosa: Kinship to Coliphages and Conservation of Regulatory Rna Structures." *Virology* 206, no. 1 (Jan 10 1995): 611-25.

Olsthoorn R., van Duin J. *Bacteriophages with Ssrna*. John Wiley & Sons, 2011.

Overby, L. R., G. H. Barlow, R. H. Doi, M. Jacob, and S. Spiegelman. "Comparison of Two Serologically Distinct Ribonucleic Acid Bacteriophages. Ii. Properties of the Nucleic Acids and Coat Proteins." *J Bacteriol* 92, no. 3 (Sep 1966): 739-45.

Palladini, A., S. Thrane, C. M. Janitzek, J. Pihl, S. B. Clemmensen, W. A. de Jongh, T. M. Clausen, *et al.* "Virus-Like Particle Display of Her2 Induces Potent Anti-Cancer Responses." *Oncoimmunology* 7, no. 3 (2018): e1408749.

Pan, Y., Z. Zhou, H. Zhang, Y. Zhou, Y. Li, C. Li, X. Chen, *et al.* "The Atrqbeta-001 Vaccine Improves Cardiac Function and Prevents Postinfarction Cardiac Remodeling in Mice." *Hypertens Res* 42, no. 3 (Mar 2019): 329-40.
Pang, H. H., C. Y. Huang, Y. W. Chou, C. J. Lin, Z. L. Zhou, Y. L. Shiue, K. C. Wei, and H. W. Yang. "Bioengineering Fluorescent Virus-Like Particle/Rnai Nanocomplexes Act Synergistically with Temozolomide to Eradicate Brain Tumors." *Nanoscale* 11, no. 17 (Apr 25 2019): 8102-09.
Paranchych, W., and L. S. Frost. "The Physiology and Biochemistry of Pili." *Adv Microb Physiol* 29 (1988): 53-114.
Paranchych, W., P. M. Krahn, and R. D. Bradley. "Stages in Phage R17 Infection." *Virology* 41, no. 3 (Jul 1970): 465-73.
Park, S. Y., H. C. Moon, and H. Y. Park. "Live-Cell Imaging of Single Mrna Dynamics Using Split Superfolder Green Fluorescent Proteins with Minimal Background." *RNA* 26, no. 1 (Jan 2020): 101-09.
Peabody, D. S. *Rna Bacteriophages*. Brenner`S Encyclopedia of Genetics (Second Edition). Elsevier, 2013. doi:https://doi.org/10.1016/B978-0-12-374984-0.01349-8.
Peabody, D. S. "Role of the Coat Protein-Rna Interaction in the Life Cycle of Bacteriophage Ms2." *Mol Gen Genet* 254, no. 4 (Apr 28 1997): 358-64.
Peabody, D. S., B. Manifold-Wheeler, A. Medford, S. K. Jordan, J. do Carmo Caldeira, and B. Chackerian. "Immunogenic Display of Diverse Peptides on Virus-Like Particles of Rna Phage Ms2." *J Mol Biol* 380, no. 1 (Jun 27 2008): 252-63.
Peabody, J., P. Muttil, B. Chackerian, and E. Tumban. "Characterization of a Spray-Dried Candidate Hpv L2-Vlp Vaccine Stored for Multiple Years at Room Temperature." *Papillomavirus Res* 3 (Jun 2017): 116-20.
Persson, M., K. Tars, and L. Liljas. "The Capsid of the Small Rna Phage Prr1 Is Stabilized by Metal Ions." *J Mol Biol* 383, no. 4 (Nov 21 2008): 914-22.
Phares, T. W., A. D. May, C. J. Genito, N. A. Hoyt, F. A. Khan, M. D. Porter, M. DeBot, *et al.* "Rhesus Macaque and Mouse Models for Down-Selecting Circumsporozoite Protein Based Malaria Vaccines Differ Significantly in Immunogenicity and Functional Outcomes." *Malar J* 16, no. 1 (Mar 13 2017): 115.
Plevka, P., A. Kazaks, T. Voronkova, S. Kotelovica, A. Dishlers, L. Liljas, and K. Tars. "The Structure of Bacteriophage Phicb5 Reveals a Role of the Rna Genome and Metal Ions in Particle Stability and Assembly." *J Mol Biol* 391, no. 3 (Aug 21 2009): 635-47.
Plotkin, S. "History of Vaccination." *Proc Natl Acad Sci U S A* 111, no. 34 (Aug 26 2014): 12283-7.
Pumpens, P. *Single-Stranded Rna Phages: From Molecular Biology to Nanotechnology*. United States: CRC Press Taylor & Francis Group, 2020.
Purwar, M., J. K. Pokorski, P. Singh, S. Bhattacharyya, H. Arendt, J. DeStefano, C. C. La Branche, *et al.* "Design, Display and Immunogenicity of Hiv1 Gp120 Fragment Immunogens on Virus-Like Particles." *Vaccine* 36, no. 42 (Oct 8 2018): 6345-53.
Pushko, P., T. Kozlovskaya, I. Sominskaya, A. Brede, E. Stankevica, V. Ose, P. Pumpens, and E. Grens. "Analysis of Rna Phage Fr Coat Protein Assembly by Insertion, Deletion and Substitution Mutagenesis." *Protein Eng* 6, no. 8 (Nov 1993): 883-91.
Pushko, P., P. Pumpens, and E. Grens. "Development of Virus-Like Particle Technology from Small Highly Symmetric to Large Complex Virus-Like Particle Structures." *Intervirology* 56, no. 3 (2013): 141-65.
Remaut, E., P. D. Waele, A. Marmenout, P. Stanssens, and W. Fiers. "Functional Expression of Individual Plasmid-Coded Rna Bacteriophage Ms2 Genes." *EMBO J* 1, no. 2 (1982): 205-9.
Rhee, J. K., M. Baksh, C. Nycholat, J. C. Paulson, H. Kitagishi, and M. G. Finn. "Glycan-Targeted Virus-Like Nanoparticles for Photodynamic Therapy." *Biomacromolecules* 13, no. 8 (Aug 13 2012): 2333-8.
Rodriguez-Limas, W. A., K. Sekar, and K. E. Tyo. "Virus-Like Particles: The Future of Microbial Factories and Cell-Free Systems as Platforms for Vaccine Development." *Curr Opin Biotechnol* 24, no. 6 (Dec 2013): 1089-93.
Roy, P., and R. Noad. "Virus-Like Particles as a Vaccine Delivery System: Myths and Facts." *Adv Exp Med Biol* 655 (2009): 145-58.
Roldao, A., M. C. Mellado, L. R. Castilho, M. J. Carrondo, and P. M. Alves. "Virus-Like Particles in Vaccine Development." *Expert Rev Vaccines* 9, no. 10 (Oct 2010): 1149-76.
Rolfsson, O., S. Middleton, I. W. Manfield, S. J. White, B. Fan, R. Vaughan, N. A. Ranson, *et al.* "Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage Ms2." *J Mol Biol* 428, no. 2 Pt B (Jan 29 2016): 431-48.
Rolih, V., J. Caldeira, E. Bolli, A. Salameh, L. Conti, G. Barutello, F. Riccardo, *et al.* "Development of a Vlp-Based Vaccine Displaying an Xct Extracellular Domain for the Treatment of Metastatic Breast Cancer." *Cancers (Basel)* 12, no. 6 (Jun 8 2020).

Rumnieks, J., and K. Tars. "Crystal Structure of the Bacteriophage Qbeta Coat Protein in Complex with the Rna Operator of the Replicase Gene." *J Mol Biol* 426, no. 5 (Mar 6 2014): 1039-49.
Rumnieks, J., and K. Tars. "Crystal Structure of the Maturation Protein from Bacteriophage Qbeta." *J Mol Biol* 429, no. 5 (Mar 10 2017): 688-96.
Ruokoranta, T. M., A. M. Grahn, J. J. Ravantti, M. M. Poranen, and D. H. Bamford. "Complete Genome Sequence of the Broad Host Range Single-Stranded Rna Phage Prr1 Places It in the Levivirus Genus with Characteristics Shared with Alloleviviruses." *J Virol* 80, no. 18 (Sep 2006): 9326-30.
Sakurai T., Watanabe I., Ohno T. "Isolation and Serological Grouping of Rna Phages." *Uirusu* 17, no. 4 (1967).
Schmidt, J. M. "Observations on the Adsorption of Caulobacter Bacteriophages Containing Ribonucleic Acid." *J Gen Microbiol* 45, no. 2 (Nov 1966): 347-53.
Scott, D. W. "Serological Cross Reactions among the Rna-Containing Coliphages." *Virology* 26 (May 1965): 85-8.
Shi, M., X. D. Lin, J. H. Tian, L. J. Chen, X. Chen, C. X. Li, X. C. Qin, *et al.* "Redefining the Invertebrate Rna Virosphere." *Nature* 540, no. 7634 (Dec 22 2016): 539-43.
Shi, Z., Z. Cai, A. Sanchez, T. Zhang, S. Wen, J. Wang, J. Yang, S. Fu, and D. Zhang. "A Novel Toll-Like Receptor That Recognizes Vesicular Stomatitis Virus." *J Biol Chem* 286, no. 6 (Feb 11 2011): 4517-24.
Shishovs, M., J. Rumnieks, C. Diebolder, K. Jaudzems, L. B. Andreas, J. Stanek, A. Kazaks, *et al.* "Structure of Ap205 Coat Protein Reveals Circular Permutation in Ssrna Bacteriophages." *J Mol Biol* 428, no. 21 (Oct 23 2016): 4267-79.
Skamel, C., S. G. Aller, and A. Bopda Waffo. "In Vitro Evolution and Affinity-Maturation with Coliphage Qbeta Display." *PLoS One* 9, no. 11 (2014): e113069.
Skibinski, D. A., B. J. Hanson, Y. Lin, V. von Messling, A. Jegerlehner, J. B. Tee, H. Chye de, *et al.* "Enhanced Neutralizing Antibody Titers and Th1 Polarization from a Novel Escherichia Coli Derived Pandemic Influenza Vaccine." *PLoS One* 8, no. 10 (2013): e76571.
Skripkin, E. A., M. R. Adhin, M. H. de Smit, and J. van Duin. "Secondary Structure of the Central Region of Bacteriophage Ms2 Rna. Conservation and Biological Significance." *J Mol Biol* 211, no. 2 (Jan 20 1990): 447-63.
Smith, M. T., A. K. Hawes, and B. C. Bundy. "Reengineering Viruses and Virus-Like Particles through Chemical Functionalization Strategies." *Curr Opin Biotechnol* 24, no. 4 (Aug 2013): 620-6.
Sommerfelt, M. A. "Circular Ccr5 Peptide Conjugates and Uses Thereof (Wo2008074895)." *Expert Opin Ther Pat* 19, no. 9 (Sep 2009): 1323-8.
Soongrung, T., K. Mongkorntanyatip, T. Peepim, S. Jitthamstaporn, P. Pitakpolrat, T. Kaewamatawong, C. M. Janitzek, *et al.* "Virus-Like Particles Displaying Major House Dust Mite Allergen Der P 2 for Prophylactic Allergen Immunotherapy." *Allergy* 75, no. 5 (May 2020): 1232-36.
Spohn, G., N. Arenas-Ramirez, G. Bouchaud, and O. Boyman. "Endogenous Polyclonal Anti-Il-1 Antibody Responses Potentiate Il-1 Activity During Pathogenic Inflammation." *J Allergy Clin Immunol* 139, no. 6 (Jun 2017): 1957-65 e3.
Spohn, G., G. T. Jennings, B. E. Martina, I. Keller, M. Beck, P. Pumpens, A. D. Osterhaus, and M. F. Bachmann. "A Vlp-Based Vaccine Targeting Domain Iii of the West Nile Virus E Protein Protects from Lethal Infection in Mice." *Virol J* 7 (Jul 6 2010): 146.
Stonehouse, N. J., K. Valegard, R. Golmohammadi, S. van den Worm, C. Walton, P. G. Stockley, and L. Liljas. "Crystal Structures of Ms2 Capsids with Mutations in the Subunit Fg Loop." *J Mol Biol* 256, no. 2 (Feb 23 1996): 330-9.
Storni, T., C. Ruedl, K. Schwarz, R. A. Schwendener, W. A. Renner, and M. F. Bachmann. "Nonmethylated Cg Motifs Packaged into Virus-Like Particles Induce Protective Cytotoxic T Cell Responses in the Absence of Systemic Side Effects." *J Immunol* 172, no. 3 (Feb 1 2004): 1777-85.
Sun, Y., and Y. Sun. "[Preparation and Activity Validation of Pp7 Bacteriophage-Like Particles Displaying Pap114-128 Peptide]." *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* 32, no. 10 (Oct 2016): 1356-61.
Tars, K. *Biocommunication of Phages*. Ssrna Phages: Life Cycle, Structure and Applications. Springer, Cham, 2020. doi:https://doi.org/10.1007/978-3-030-45885-0.
Tars, K., M. Bundule, K. Fridborg, and L. Liljas. "The Crystal Structure of Bacteriophage Ga and a Comparison of Bacteriophages Belonging to the Major Groups of Escherichia Coli Leviviruses." *J Mol Biol* 271, no. 5 (Sep 5 1997): 759-73.
Tars, K., K. Fridborg, M. Bundule, and L. Liljas. "The Three-Dimensional Structure of Bacteriophage Pp7 from Pseudomonas Aeruginosa at 3.7-a Resolution." *Virology* 272, no. 2 (Jul 5 2000): 331-7.
Thrane, S., C. M. Janitzek, S. Matondo, M. Resende, T. Gustavsson, W. A. de Jongh, S. Clemmensen, *et al.* "Bacterial Superglue Enables Easy Development of Efficient Virus-Like Particle Based Vaccines." *J Nanobiotechnology* 14 (Apr 27 2016): 30.
Tissot, A. C., P. Maurer, J. Nussberger, R. Sabat, T. Pfister, S. Ignatenko, H. D. Volk, *et al.* "Effect of Immunisation against Angiotensin Ii with Cyt006-Angqb on Ambulatory Blood Pressure: A

Double-Blind, Randomised, Placebo-Controlled Phase Iia Study." *Lancet* 371, no. 9615 (Mar 8 2008): 821-7.
Tissot, A. C., R. Renhofa, N. Schmitz, I. Cielens, E. Meijerink, V. Ose, G. T. Jennings, *et al.* "Versatile Virus-Like Particle Carrier for Epitope Based Vaccines." *PLoS One* 5, no. 3 (Mar 23 2010): e9809.
Tyler, M., E. Tumban, D. S. Peabody, and B. Chackerian. "The Use of Hybrid Virus-Like Particles to Enhance the Immunogenicity of a Broadly Protective Hpv Vaccine." *Biotechnol Bioeng* 111, no. 12 (Dec 2014): 2398-406.
Twarock, R., R. J. Bingham, E. C. Dykeman, and P. G. Stockley. "A Modelling Paradigm for Rna Virus Assembly." *Curr Opin Virol* 31 (Aug 2018): 74-81.
Valegard, K., L. Liljas, K. Fridborg, and T. Unge. "The Three-Dimensional Structure of the Bacterial Virus Ms2." *Nature* 345, no. 6270 (May 3 1990): 36-41.
Valegard, K., J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas. "Crystal Structure of an Rna Bacteriophage Coat Protein-Operator Complex." *Nature* 371, no. 6498 (Oct 13 1994): 623-6.
Vandenberghe, R., M. E. Riviere, A. Caputo, J. Sovago, R. P. Maguire, M. Farlow, G. Marotta, *et al.* "Active Abeta Immunotherapy Cad106 in Alzheimer's Disease: A Phase 2b Study." *Alzheimers Dement (N Y)* 3, no. 1 (Jan 2017): 10-22.
Vasiljeva, I., T. Kozlovska, I. Cielens, A. Strelnikova, A. Kazaks, V. Ose, and P. Pumpens. "Mosaic Qbeta Coats as a New Presentation Model." *FEBS Lett* 431, no. 1 (Jul 10 1998): 7-11.
Vasquez, C., N. Granboulan, and R. M. Franklin. "Structure of the Ribonucleic Acid Bacteriophage R17." *J Bacteriol* 92, no. 6 (Dec 1966): 1779-86.
Verbraeken, E., and W. Fiers. "Further Evidence on the Role of the a Protein in Bacteriophage Ms2 Particles." *FEBS Lett* 28, no. 1 (Nov 15 1972): 89-92.
Voronkova, T., A. Grosch, A. Kazaks, V. Ose, D. Skrastina, K. Sasnauskas, B. Jandrig, *et al.* "Chimeric Bacteriophage Fr Virus-Like Particles Harboring the Immunodominant C-Terminal Region of Hamster Polyomavirus Vp1 Induce a Strong Vp1-Specific Antibody Response in Rabbits and Mice." *Viral Immunol* 15, no. 4 (2002): 627-43.
Wang, G., Y. Liu, H. Feng, Y. Chen, S. Yang, Q. Wei, J. Wang, D. Liu, and G. Zhang. "Immunogenicity Evaluation of Ms2 Phage-Mediated Chimeric Nanoparticle Displaying an Immunodominant B Cell Epitope of Foot-and-Mouth Disease Virus." *PeerJ* 6 (2018): e4823.
Weinbauer, M. G. "Ecology of Prokaryotic Viruses." *FEMS Microbiol Rev* 28, no. 2 (May 2004): 127-81.
Weiner, A. M., and K. Weber. "Natural Read-through at the Uga Termination Signal of Q-Beta Coat Protein Cistron." *Nat New Biol* 234, no. 50 (Sep 15 1971): 206-9.
Wu, D., Y. Pan, S. Yang, C. Li, Y. Zhou, Y. Wang, X. Chen, *et al.* "Pcsk9qbeta-003 Vaccine Attenuates Atherosclerosis in Apolipoprotein E-Deficient Mice." *Cardiovasc Drugs Ther* 35, no. 1 (Feb 2021): 141-51.
Zeltins, A. "Construction and Characterization of Virus-Like Particles: A Review." *Mol Biotechnol* 53, no. 1 (Jan 2013): 92-107.
Zhai, L., J. Peabody, Y. S. Pang, J. Schiller, B. Chackerian, and E. Tumban. "A Novel Candidate Hpv Vaccine: Ms2 Phage Vlp Displaying a Tandem Hpv L2 Peptide Offers Similar Protection in Mice to Gardasil-9." *Antiviral Res* 147 (Nov 2017): 116-23.
Zhang, L., W. Qiu, S. Crooke, Y. Li, A. Abid, B. Xu, M. G. Finn, and F. Lin. "Development of Autologous C5 Vaccine Nanoparticles to Reduce Intravascular Hemolysis in Vivo." *ACS Chem Biol* 12, no. 2 (Feb 17 2017): 539-47.
Zhang, T., and H. H. Fang. "Applications of Real-Time Polymerase Chain Reaction for Quantification of Microorganisms in Environmental Samples." *Appl Microbiol Biotechnol* 70, no. 3 (Apr 2006): 281-9.
Zinder, N. D. "Portraits of Viruses: Rna Phage." *Intervirology* 13, no. 5 (1980): 257-70.
Zinder, N. D. "Rna Phages." *Annu Rev Microbiol* 19 (1965): 455-72.