

Review for Advances in Virus Research [12/04/05]

Title: **Chlorella viruses**

Takashi Yamada¹, Hideki Onimatsu¹, and James L. Van Etten²

¹Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, Higashi-Hiroshima 739-8530, Japan; ²Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68583-0722, USA

Running title: chlorella viruses

Correspondence with: Takashi Yamada

Department of Molecular Biotechnology, Graduate School of
Advanced Sciences of Matter, Hiroshima University,
1-3-1 Kagamiyama, Higashi-Hiroshima 739-8530, Japan
Phone/Fax: +81-824-24-7752
E-mail: tayamad@hiroshima-u.ac.jp

Key Words Chlorella viruses, PBCV-1, virion structure, gene expression, genome rearrangements, polysaccharide synthesis and degradation

I. Introduction

II. Research history, classification, and specific features

- A. History and classification of viruses and their hosts
- B. Specific features of the chlorella viruses

III. Virion structure

IV. Virus life cycle

V. Virus gene expression: immediate early and late gene expression

VI. Virus protein synthesis, modification, and degradation

VII. Diversity of chlorella virus genomes

- A. Hairpin ends and inverted terminal repeats
- B. PBCV-1 genes
- C. Linearity of gene arrangements in the chlorella viruses
- D. Patterns of gene rearrangements
 - 1. Large deletions/insertions
 - 2. Small deletions/insertions
 - 3. Gene amplification
 - 4. Gene replacement

VIII. Cell wall digestion: chitinase, chitosanase, polysaccharide lyase and β -glucanase

IX. Glycosyltransferases and synthesis of polysaccharides and fucose

- A. Glycosylation of virus structural proteins
- B. Extracellular polysaccharide synthesis
- C. Fucose synthesis

X. Relationship of chlorella viruses to other viruses

XI. Perspectives

Acknowledgments

References

Legends to Figures

Abstract

Chlorella viruses or chloroviruses are large, icosahedral, plaque-forming, double-stranded-DNA-containing viruses that replicate in certain strains of the unicellular green alga *Chlorella*. DNA sequence analysis of the 330 kbp genome of *Paramecium bursaria* chlorella virus 1 (PBCV-1), the prototype of this virus family (*Phycodnaviridae*), predict ~366 protein-encoding genes and 11 tRNA genes. The predicted gene products of ~50% of these genes resemble proteins of known function, including many that are completely unexpected for a virus. In addition, the chlorella viruses have several features and encode many gene products that distinguish them from most viruses. These products include: (i) multiple DNA methyltransferases and DNA site-specific endonucleases, (ii) the enzymes required to glycosylate their proteins and synthesize polysaccharides such as hyaluronan and chitin, (iii) a virus-encoded K⁺ channel (called Kcv) located in the internal membrane of the virions, (iv) a SET domain containing protein (referred to as vSET) that dimethylates Lys27 in histone 3, and (v) PBCV-1 has three types of introns; a self-splicing intron, a spliceosomal processed intron and a small tRNA intron. Accumulating evidence indicates that the chlorella viruses have a very long evolutionary history. This review mainly deals with recent research on the virion structure, genome rearrangements, gene expression, cell wall degradation, polysaccharide synthesis, and evolution of PBCV-1 as well as other related viruses.

I. Introduction

Members, including the chlorella viruses, and prospective members of the family *Phycodnaviridae* constitute a genetically diverse, but morphologically similar, group of viruses with eukaryotic algal hosts from both fresh and marine waters. The family name derives from two distinguishing characteristics: i) "phyco" from their algal hosts, and ii) "dna" because all of these viruses have dsDNA genomes (Wilson *et al.*, 2005b). The phycodnaviruses have some of the largest virus genomes known, ranging in size from ~170 kb to 560 kb and contain several hundred protein encoding genes.

The phycodnaviruses are among the viroplankton recently recognized as important ecological elements in aqueous environments. They, along with other viruses, play important roles in the dynamics of algal blooms, nutrient cycling, algal community structure, and possibly gene transfer between organisms. The discovery phase of aquatic viruses, including

the phycodnaviruses, is just beginning with new viruses continually being discovered as more environmental samples are examined. Ongoing metagenomic studies involving massive DNA sequencing reveal a greater viral diversity than could have been imagined just a few years ago (e.g. Hambly and Suttle, 2005; Wommack and Colwell, 2000). The genetic diversity that exists in the phycodnaviruses, albeit with only a few genomes sequenced, is enormous. To illustrate this diversity, viruses in three genera of the phycodnaviruses have been sequenced and each of these viruses encode several hundred genes. However, only 14 of these genes are common to all three viruses (Dunigan *et al.*, in press). Thus, there are more than 1000 unique genes in just these three viruses.

Accumulating evidence also indicates that the phycodnaviruses are probably very old viruses. The phycodnaviruses together with the poxviruses, iridoviruses, asfarviruses, and the recently discovered 1.2 Mb Mimivirus, probably have a common evolutionary ancestor, perhaps arising at the time eukaryotes separated from prokaryotes, ca. 3 billion years ago (Raoult *et al.*, 2004; Villarreal, 2005; Villarreal and DeFilippis, 2000). All of these viruses share 9 gene products and 33 more gene products are present in at least two of these five viral families (Iyer *et al.*, 2001; Raoult *et al.*, 2004). Collectively, these viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDV) (Iyer *et al.*, 2001).

This review focuses on the chlorella viruses that constitute one genera in the family *Phycodnaviridae*. Phycodnaviruses are large (mean diameter of 160 +/- 60 nm) icosahedrons and, where known, the viruses have an internal membrane that is required for infection. Phylogenetic analyses of the δ -DNA polymerases from the phycodnaviruses indicate that they are more closely related to each other than to other dsDNA viruses and that they form a monophyletic group, consistent with a common ancestor (Wilson *et al.*, 2005b). However, the viruses fall into six clades which correlate with their hosts and each has been given genus status. Often the genera can be distinguished by additional properties; e.g. lytic versus lysogenic life styles or linear versus circular genomes (Wilson *et al.*, 2005b). Members of the genus *Chlorovirus* (chlorella viruses) infect fresh water algae, whereas members of the other five genera (*Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphidovirus*) infect marine algae.

The type chlorella virus is *Paramecium bursaria* chlorella virus (PBCV-1). Because there have been several recent reviews on the phycodnaviruses and the chlorella viruses (e.g.

Van Etten 2003; Van Etten *et al.*, 1991, 2002; Van Etten and Meints, 1999; Kang *et al.*, 2005; Dunigan *et al.*, in press), this review deals mainly with recent research on virion structure, infection cycle, genome rearrangements, gene expression, cell wall degradation and polysaccharide synthesis. Additional information, including a complete list of chlorella virus publications and additional images of the viruses, is available on the “World of Chlorella Viruses” web site at: <http://www.ianr.unl.edu/plantpath/facilities/Virology/index.htm/>. The general history of the algal viruses and ecological aspects of these fascinating viruses can be found in other reviews (e.g. Brussaard, 2004; Fuhrman, 1999; Lemke, 1976; Dodds, 1979; Müller *et al.*, 1998; Suttle, 2000, 2005; Wommack and Colwell, 2000).

II. Research history, classification and specific features

A. History and classification of viruses and their hosts

In 1978, Kawakami and Kawakami (1978) described the appearance of large (~180 nm in diameter) icosahedral, lytic viruses in zoochloellae of the protozoan *Paramecium bursaria* (designated zoochlorella cell virus, ZCV) after the algae were released from the paramecium. No virus particles were detected in zoochloellae growing symbiotically inside the paramecium cells although ZCV particles were present in the depressions of the pellicle, between the cilia, and in the food vacuole of the paramecium. ZCV infected the zoochlorella by adsorption and digestion of the cell wall and virus particles accumulated in the algal cytoplasm. After cell lysis, progeny viruses were released into the medium.

Independently, a few years latter, similar lytic viruses were described in zoochloellae isolated from the green coelenterate *Hydra viridis* (Meints *et al.*, 1981; Van Etten *et al.*, 1981) and also from *P. bursaria* (Van Etten *et al.*, 1982). A laboratory infection system for these viruses was developed using exsymbiotic zoochlorella strains as hosts, including *Chlorella* NC64A that was originally isolated from a *P. bursaria* (Van Etten *et al.*, 1983a). This system allows chlorella viruses to be produced in large quantities and the viruses can be assayed by plaque formation using standard bacteriophage techniques (Van Etten *et al.*, 1983a). Since these early studies literally hundreds of chlorella viruses have been isolated from natural sources.

Chlorella viruses included in the genus *Chlorovirus* (Wilson *et al.*, 2005b) currently consist of three species: (i) Viruses that infect *Chlorella* NC64A (NC64A viruses). (ii) Viruses that infect *Chlorella* Pbi (Pbi viruses). NC64A viruses neither infect nor attach to *Chlorella* Pbi, and vice versa. (iii) Viruses that infect symbiotic zoochlorella in the coelenterate *Hydra viridis*. *Hydra* zoochlorella have not been cultured free of virus so these viruses can only be isolated from chlorella cells freshly released from hydra. Recently, a virus that infects zoochlorella of the heliozoon *Acanthocystis turfacea* was described (Bubeck and Pfitzner, 2005). These viruses, designated ATCV-1 and ATCV-2, infect *Chlorella* SAG 3.83, a symbiont of *A. turfacea* but do not infect either *Chlorella* SAG 211-6, a host for the NC64A viruses, or *Chlorella* SAG 241-80, a host for the Pbi viruses.

NC64A viruses have been isolated from fresh water collected in the United States (Van Etten *et al.*, 1985a), South America, Japan (Yamada *et al.*, 1991), China (Zhang *et al.*, 1988), South Korea (Cho *et al.*, 2002), Australia, Israel, and Italy (Van Etten *et al.*, 2002). Pbi viruses initially were discovered in fresh water collected in Europe (Reisser *et al.*, 1988), and more recently in water collected in Australia, Canada, and the northern United States or at higher elevations in the western United States (J.L. Van Etten and M. Nelson, unpublished results). The most important factors influencing the distribution of NC64A and Pbi viruses are probably latitude and altitude. *Chlorella* NC64A and *Chlorella* Pbi were originally isolated from American and European isolates of *P. bursaria*, respectively. The component sugars in the cell walls of *Chlorella* NC64A and *Chlorella* Pbi differ considerably (Kapaun *et al.*, 1992). Because the viruses can distinguish the two chlorella isolates, it seemed likely that the host receptor for the viruses might also serve as the recognition factor for becoming a symbiont in the paramecia. However, *Chlorella* NC64A and *Chlorella* Pbi each established a stable symbiotic relationship with both American and European isolates of *P. bursaria* (Reisser *et al.*, 1991).

Recently, 18S rRNA sequence analyses of zoochlorella from American and European paramecia have been conducted (Hoshina *et al.*, 2004; 2005). Zoochlorella 18S rRNAs separate into two lineages; NC64A (USA), Syngen 2-3 (USA), Cs2 (China), MRBG1 (Australia), and strains from Japan belong to the American type, whereas PB-SW1 (Germany) and CCAP 1660/11 (UK) strains belong to the European type. The American type symbionts have three group-I introns in the 18S rRNA genes, whereas a single group-I intron, located at a

different position, exists in the European symbionts. Likewise, the 18S rRNA sequence distinguishes the two groups: (i) Cs2, MRBG1, and strains from Japan and (ii) PB-SW1 and CCAP 1660/11. The American type and European type of zoochlorellae correspond to the hosts for NC64A viruses and Pbi viruses, respectively. It will be interesting to determine where *Chlorella* SAG3.83, that is the host for new chlorella viruses, fits into this scheme (Bubeck and Pfitzner, 2005).

B. Specific features of the chlorella viruses.

The chlorella viruses have many interesting properties. Most of these studies have been conducted on PBCV-1 and its-related NC64A viruses such as chlorella virus Kyoto 2 (CVK2) (Van Etten *et al.*, 1991; Van Etten and Meints, 1999; Yamada *et al.*, 1991). Some of these features are summarized as follows: (i) The virus genomes are large linear dsDNAs (315-380 kbp) with cross-linked hairpin termini. (ii) The viruses infect host cells by a bacteriophage-like mechanism. (iii) Chlorella viruses often encode multiple type II DNA methyltransferases and DNA site-specific (restriction) endonucleases (Nelson *et al.*, 1993; 1998). (iv) Unlike other glycoprotein-containing viruses, chlorella viruses encode most, if not all, of the components required to glycosylate their proteins (Wang *et al.*, 1993; Graves *et al.*, 2001; Markine-Goriynoff *et al.*, 2004). (v) Chlorella viruses were the first viruses discovered to have more than one type of intron: PBCV-1 has a self-splicing intron in a transcription factor TFIIS-like gene (Yamada *et al.*, 1994; Li *et al.*, 1995), a spliceosomal-processed intron in its DNA polymerase gene (Grabherr *et al.*, 1992; Zhang *et al.*, 2001), and a small intron in one of its tRNA genes (Nishida *et al.*, 1999a). (vi) Many chlorella virus encoded proteins are either the smallest or among the smallest proteins of their class and some may represent the minimal catalytic unit. (vii) Some translational components such as tRNAs (Li *et al.*, 1997; Nishida *et al.*, 1999a) and elongation factor EF-3 (Yamada *et al.*, 1993) are virus encoded and expressed during the infection cycle. (viii) Chlorella viruses encode a functional K⁺ channel protein called Kcv. Kcv was the first virus encoded K⁺ channel to be discovered and it is the smallest protein known to form a functional K⁺ selective channel (Plugge *et al.*, 2000; Gazzarrini *et al.*, 2003; Kang *et al.*, 2003).

III. Virion structure

PBCV-1 virions have a sedimentation coefficient of about 2300 S in sucrose density gradients (Van Etten *et al.*, 1983b) and an estimated molecular mass of 1×10^9 Daltons (Yonker *et al.*, 1985). The virion contains 64% protein, 21% to 25% DNA, and 5% to 10% lipid (Skrdla *et al.*, 1984). The PBCV-1 virion contains more than 100 virion-encoded proteins (Skrdla *et al.*, 1984; D. D. Dunigan *et al.*, unpublished results) including two DNA restriction endonucleases, DNA binding proteins and protein kinases (Yamada *et al.*, 1996). The PBCV-1 54 kDa major capsid protein Vp54 is a glycoprotein and comprises ~40% of the virus protein. Vp54 has been crystallized and consists of two eight-stranded, antiparallel β -barrel, "jelly-roll" domains related by a pseudo-sixfold rotation (Nandhagopal *et al.*, 2002).

Ultrastructural studies have been conducted on both intact and disrupted chlorella virus virions using either negative staining (Becker *et al.*, 1993) or cryo-electron microscopy with three-dimensional image reconstructions (26 Å resolution) (Yan *et al.*, 2000). The latter study was complemented and extended by fitting the structure of the major capsid protein Vp54 to the cryo-electron microscopy density maps of PBCV-1 (Simpson *et al.*, 2003). The outer glycoprotein capsid is icosahedral and surrounds a lipid bilayer membrane. The membrane is connected to the outer shell by regularly spaced proteins. Disruption of the membrane destroys PBCV-1 infectivity (Skrdla *et al.*, 1984). The outer diameter of the viral capsid varies from a minimum of 1,650 Å along the two- and three-fold axes to a maximum of 1,900 Å along the five-fold axes. The capsid shell consists of 1,680 donut-shaped trimeric capsomers plus 12 pentameric capsomers at each icosahedral vertex. The trimeric capsomers are arranged into 20 trisymmetrons (each containing 66 trimers) and 12 pentasymmetrons (each containing 30 trimers and one pentamer at the icosahedral vertices) (Fig. 1). Assuming all the trimeric capsomers are identical, the outer capsid of the virus contains 5,040 copies of the major capsid protein Vp54. The triangulation number (T) for the virus is 169 ($h = 7, k = 8$) and the virus has a right handed, skew class of T lattice (Caspar and Klug, 1962).

Most of the trimeric capsomers have a central, concave depression surrounded by three protruding towers. The trimeric capsomers are 72 Å in diameter and ~75 Å high. The capsomers interconnect at their bases in a contiguous shell that is 20-25 Å thick. Twelve pentamer capsomers, each ~70 Å in diameter, exist at the virus five-fold vertices and probably consist of a different protein. Each pentamer has a cone-shaped, axial channel at its base. One or more proteins appear below the axial channel and outside the inner membrane (Fig. 1B).

This protein(s) may be responsible for digesting the host cell wall during infection. Presumably contact between the virus and its host receptor alters the channel sufficiently to release the wall-degrading enzyme(s).

Recently, complementary information about the PBCV-1 virion structure was obtained by atomic force microscopy (AFM) (Kuznetsov *et al.*, 2005). Since AFM is not dependent on symmetry averaging as is cryo-electron microscopy, it can reveal unique properties of individual particles. From the response of the AFM tip in contact with the particles, the virus particles appear somewhat soft and are readily deformed. The individual trimeric capsomers appear to have a small hole in their center and a distinctive triangular shape that is more angular and accentuated than the “doughnut” shape deduced from the cryo-electron microscopy (Yan *et al.*, 2000). The pentagonal vertices are formed by five copies of a different protein, and this has yet another unique protein in its center (Fig. 1D and E). The central protein exhibits some unusual behavior when subjected to AFM tip pressure; it disappears into the virion interior, leaving a distinct hole. When the AFM tip pressure is decreased, it returns to its original position. Interestingly, virion degradation is accompanied by the appearance of many small, uniform, spherical and virus like particles (VLP) consistent with T = 1 or 3 icosahedral products (Kuznetsov *et al.*, 2005).

IV. Virus life cycle

PBCV-1 infects its host by attaching rapidly, specifically, and irreversibly to the external surface of the algal cell wall (Meints *et al.*, 1984). Attachment always occurs at a virus vertex, possibly with hair-like appendages (Van Etten *et al.*, 1991), and is followed by degradation of the host wall at the attachment point. The determinants for host range are associated with attachment. Recently, Onimatsu *et al.* (2004) reported that a virion protein Vp130 from chlorovirus CVK2 binds specifically to the host *Chlorella* cell wall. Vp130 is a homolog of PBCV-1 protein A140/145R and consists of 1,126 amino acid residues (predicted mol wt, 121,257 and pI, 10.76). The Vp130 N-terminus is blocked by some unknown structure and the C-terminus consists of 23 tandem PAPK repeats. Internally, Vp130 contains seven repeats of 70-73 amino acids, each copy of which is separated by several PAPK sequences. This protein is well conserved among the NC64A viruses. Because externally added Vp130

competes with CVK2 in binding to host cells, Vp130 is most likely a host-recognizing protein in the virion (Onimatsu *et al.*, 2004). Immune electron microscopy using Vp130 antibody established that the protein is localized specifically at the vertices of the CVK2 virion (H. Onimatsu *et al.*, manuscript in preparation).

Attachment of the virion to the host cell wall probably alters the virion structure slightly, allowing release of a virion-packaged wall digesting enzyme(s). Following host cell wall degradation, the internal membrane of the virus probably fuses with the host membrane resulting in entry of the viral DNA and virion-associated proteins. An empty capsid is left on the cell surface. Infection results in rapid depolarization of the host membrane (Mehmel *et al.*, 2003; Frohns *et al.*, 2006) and we hypothesize that this rapid depolarization is caused by a virus-encoded K^+ channel (called Kcv) located in the internal membrane. This depolarization may aid in the release of DNA into the cell and/or limit subsequent infection by additional viruses.

Circumstantial evidence indicates that the PBCV-1 DNA and suspected virion-associated proteins quickly move to the nucleus where early transcription is detected within 5-10 min p.i. (Schuster *et al.*, 1986; Kawasaki *et al.*, 2004). Experimental results indicate that host chromosomal DNA begins to be degraded, possibly due to two restriction endonucleases that are packaged in the PBCV-1 virion, within min of infection (D. D. Dunigan *et al.*, unpublished results; I.V. Agarkova *et al.*, manuscript in preparation).

In the immediate-early phase of infection, the host is reprogrammed to transcribe viral RNAs. Very little is known about how this occurs, but chromatin remodeling may be involved. PBCV-1 encodes a 119 amino acid SET domain containing protein (referred to as vSET) that dimethylates Lys27 in histone 3 (Manzur *et al.*, 2003). vSET is packaged in the PBCV-1 virion and accumulating evidence indicates that vSET is involved in repression of host transcription following PBCV-1 infection (K. L. Manzur *et al.*, manuscript in preparation).

PBCV-1 DNA replication begins 60-90 min p.i. and is followed by transcription of late virus genes (Schuster *et al.*, 1986; Van Etten *et al.*, 1984). Ultrastructural studies of PBCV-1 infected chlorella suggest that the nuclear membrane remains intact, at least during the early stages of virus replication (Meints *et al.*, 1986). However, a functional host nucleus is not required for virus replication since PBCV-1 can replicate, albeit poorly and with a small burst size, in UV-irradiated cells (Van Etten *et al.*, 1986). Approximately 2-3 hr p.i., assembly of

virus capsids begins in localized regions in the cytoplasm, called virus assembly centers, which become prominent at 3-4 hr p.i. (Meints *et al.*, 1986). By five hr p.i., the cytoplasm is filled with infectious progeny virus particles (~1000 particles/cell) and by 6-8 hr p.i. localized lysis of the host cell releases progeny virions. Of the progeny released, 25 to 50% of the particles are infectious; i.e. each infected cell yields ~350 plaque-forming units (PFU) (Van Etten *et al.*, 1983b). Intact infectious PBCV-1 particles accumulate inside the host 30-40 min before release. Other chlorella viruses have longer replication cycles than PBCV-1. For example, NC64A virus NY-2A requires approximately 18 hr for replication and consequently forms smaller plaques.

Some virion proteins are processed by specific proteinase activities (Songsri *et al.*, 1997). One of them is a signal peptidase-like activity and removes the N-terminal 25-33 amino acids of target proteins which contain a highly hydrophobic sequence of 17 amino acid residues. Lysine with an acidic amino acid on the N-side always precedes the cleavage site. Proteins A168R, A203R, and A532L are targets of this activity. These results lead to the following questions: (i) What enzymes (either of viral or host origin) are responsible for processing the virion proteins? (ii) When, where, and how does processing occur in the course of chlorovirus replication? (iii) What are the biological effects of processing?

Thus progress has been made on characterization of the viral structural proteins as well as the whole virion architecture. However, many fundamental questions remain to be answered about the assembly of such a large, complex virus particles in the host cells. Like poxviruses, iridoviruses, and African swine fever virus (ASFV), chlorovirus assembly occurs in localized regions in the cytoplasm, referred to as virus assembly centers (Van Etten *et al.*, 1991). (i) Where and how is the virus assembly center determined? (ii) What is the origin of the virus internal membrane? Is it derived from the ER, like other membrane-containing viruses (Cobbold *et al.*, 1996; Wolf *et al.*, 1998)? (iii) What is the role of the membrane in assembling the virion (e.g. does it function as a scaffold)? (iv) How is genomic DNA packaged into an empty preformed capsid? (v) How are the more than 100 component proteins specifically assembled into a virion?

V. Virus gene expression: immediate early and late gene expression

The PBCV-1 genome does not encode either a recognizable RNA polymerase or RNA polymerase subunit (Van Etten and Meints, 1999). The lack of a virus-encoded RNA polymerase suggests that the infecting viral DNA is targeted to the cell nucleus and that a host RNA polymerase initiates viral transcription, possibly in conjunction with virus-packaged transcription factors. Consistent with this possibility, PBCV-1 encodes at least four transcription factor-like elements, TFIIB, TFIID, TFIIS, and VLTF-2. PBCV-1 also encodes two enzymes involved in forming the mRNA cap structure, an RNA triphosphatase (Ho *et al.*, 2001) and an RNA guanylyltransferase (Ho *et al.*, 1996). However, there is no evidence that any of these proteins are packaged in the virion. The size, amino acid sequence, and biochemical properties of the PBCV-1 capping enzymes resemble yeast capping enzymes more than the multifunctional poxvirus and ASFV RNA capping enzymes (Gong and Shuman, 2002). PBCV-1 also encodes an active RNase III that presumably is involved in processing either virus mRNAs and/or tRNAs (Zhang *et al.*, 2003). In addition, PBCV-1 encodes two proteins that contain sequence elements of superfamily II helicases. Superfamily II helicases are involved in transcription (Tanner and Linder, 2001).

Two studies (Schuster *et al.*, 1986; 1990) which examined PBCV-1 RNA synthesis reached the following conclusions: i) Viral infection rapidly inhibits host RNA synthesis. ii) Chloroplast rRNAs but not cytoplasmic rRNAs are degraded beginning at about 30 min p.i. iii) PBCV-1 transcription is temporally programmed, and the first transcripts appear within 5-10 min p.i. (immediate early transcription). A few, but not all, early virus transcripts are synthesized in the absence of *de novo* protein synthesis. The synthesis of late transcripts that occurs after 60 to 90 min p.i. (late transcription) requires translation of an early virus gene(s). iv) Early and late virus genes are interspersed throughout the PBCV-1 genome. v) Full length gene probes often hybridize to mRNA transcripts that are 40-60% larger than the gene itself. vi) Transcription mapping of seven PBCV-1 genes revealed that the virus has mRNAs with 5' untranslated regions as small as 14 nucleotides and as large as 149 nucleotides. Transcripts often extend beyond the translational stop codon (Schuster *et al.*, 1990; Graves and Meints, 1992; Hiramatsu *et al.*, 1999; Kawasaki *et al.*, 2004).

A recent study (Kawasaki *et al.*, 2004) has produced some new insights into immediate early expressed genes. They isolated and characterized 23 chlorovirus PBCV-1 and CVK2 genes expressed in host cells as early as 5-10 min p.i. Some of these immediate early gene

products resembled transcriptional factors and mRNA-processing proteins including, TFIIB, helicases, mRNA capping enzyme (RNA guanylyltransferase), nucleolin, and a bean transcription factor. Other immediate early genes encoded factors influencing translation such as possible aminoacyl tRNA synthetases, possible ribosomal proteins and unknown proteins. Enzymes involved in polysaccharide synthesis were also found. All of these gene transcripts had a poly(A) tail, which decreased in size by 20 min p.i., possibly caused by an exonuclease. A typical TATA-box and a common 5'-ATGACAA element were present in the promoter region of all 23 immediate early genes, which may be recognized by host RNA polymerase and transcription factors. As suggested by the presence of the poly(A) tail, all of the immediate early genes contain a typical poly(A) addition signal "5'-AATAAA-3'", 10-90 bp downstream of the translational stop codon.

At 40 min p.i., a dramatic change occurs in the transcription of the chlorovirus genes. The immediate early genes gradually decreased in size, suggesting some weakening or cessation of poly(A) polymerase activity. Concurrently, some larger transcripts began to appear. These larger transcripts are due to readthrough from an upstream ORF and/or into a downstream ORF. These results indicate that promoter selection changes around 40 min p.i. by some mechanism, possibly involving regulatory proteins encoded by some of the immediate early transcripts. In other organisms, transcription termination signals in a gene are recognized by a series of RNA-binding proteins and RNA-processing enzymes including CstF (cleavage stimulation factor F), CPSF (cleavage and polyadenylation specific factor), and poly(A) polymerase; but these factors may not function well after 40 min p.i., resulting in elongated or unprocessed transcripts. Once in the cytosol, the poly(A) tail of mRNA is gradually shortened by an exonuclease (deadenylation nuclease called DAN) that digests the tail in the 3' to 5' direction. Once the size of the poly A tail reaches a critical threshold, the mRNA 5' cap is removed (decapping) and the RNA is rapidly degraded. Therefore, a 3'-extension of each transcript by readthrough might serve as an alternate way to protect a coding region from degradation by 3'-exonuclease. This is a unique feature of chloroviruses contrasting to similar large viruses such as vaccinia and ASFV, both of which encode a functional poly(A) polymerase and mRNAs are polyadenylated (Johnson *et al.*, 1993; Yanez *et al.*, 1995).

A poly(A) tail is also involved in initiation of translation. Efficient translation requires the mRNA poly (A) tail to bind to poly(A)-binding proteins, which, in turn, interact with translation initiation factor eIF-4G. Poly(A)-deficient mRNAs formed after 40 min p.i., might require an IRES (internal ribosome entry site)-dependent mechanism in order to be translated (Sachs, 2000).

A conserved nucleotide sequence has been identified in the promoter region of genes expressed late in PBCV-1 infected cells. Kang *et al.* (2004a) reported that an AAAAATAnTT element or a subset of this sequence is located 6-30 nucleotides upstream of the ATG start codon of seven late-expressed PBCV-1 genes.

However, many fundamental questions regarding chlorovirus gene transcription remain to be answered including: (i) What kind of RNA polymerase and its related factors are involved? (ii) How are transcription initiation and termination regulated? (iii) What mechanism switches transcription from early to late? (iv) What DNA elements are responsible for the promoter function? (v) What trans-acting factors are involved in the regulation?

VI. Virus protein synthesis, modification, and degradation

Some early virus-encoded proteins appear within 15 min p.i. Since cycloheximide, but not chloramphenicol, inhibits viral replication, PBCV-1 proteins are synthesized on cytoplasmic ribosomes and not organellar ribosomes (Skrdla *et al.*, 1984). How the virus takes over the host translational machinery forcing it to translate virus mRNAs is unknown. However, some of the factors involved in this process are expected to be virus encoded. Therefore, studying the chlorella virus system may reveal new insights into the regulation of eukaryotic protein synthesis.

The chlorella viruses are the first known viruses to encode a translation elongation factor (EF) (Yamada *et al.*, 1993). The gene for a putative EF-3 is highly conserved in all chlorovirus isolates examined so far. The EF-3 proteins from CVK2 and PBCV-1 (94% amino acid identity) have ~45% amino acid identity to an EF-3 protein from fungi (Belfield and Tuite, 1993; Chakraburty, 2001). The fungal protein stimulates EF-1 α -dependent binding of aminoacyl-tRNA to the A site of the ribosome. Like fungal EF-3 proteins, the CVK2 and PBCV-1 proteins have an ABC transporter family signature and two ATP/GTP binding-site

motifs.

PBCV-1 and CVK2 codon usages are biased to codons ending in XXA/U (63%) over those ending in XXC/G (37%) (Schuster *et al.*, 1990; Nishida *et al.*, 1999a). This bias is expected because PBCV-1 DNA is 40% G+C (CVK2, 41% G+C), whereas host nuclear DNA is 67% G+C (Van Etten *et al.*, 1985b). Therefore, finding that PBCV-1 encodes 11 tRNA genes may not be surprising: 3 for Lys, 2 each for Asn and Leu, and 1 each for Ile, Tyr, Arg, and Val. Similarly CVK2 encodes 14 tRNA genes: 3 for Lys, 2 each for Asn and Leu, and 1 each for Arg, Asp, Gly, Gln, Ile, Tyr, and Val (Nishida *et al.*, 1999a). None of the tRNAs have a CCA sequence encoded at the 3' end of the acceptor stem. Typically these 3 nucleotides are added separately to tRNAs. Some chlorella viruses encode as many as 16 tRNAs (Nishida *et al.*, 1999a; Cho *et al.*, 2002). There is a strong correlation between the abundance of virus encoded tRNAs and the virus gene codon use (Nishida *et al.*, 1999a; Lee *et al.*, 2005).

The virus encoded tRNAs contain internal A and B boxes characteristic of RNA polymerase III promoter elements, suggesting the tRNAs might be transcribed individually by RNA polymerase III (Nishida *et al.*, 1999a). However, the tRNA genes are transcribed as a large precursor RNA and processed via intermediates to mature tRNAs at both early and late stages of virus replication. Some, if not all, of the tRNAs are aminoacylated *in vivo*, suggesting they probably function in viral protein synthesis (Nishida *et al.*, 1999a). Possibly, the virus-encoded EF-3 in combination with the virus encoded tRNAs alter the host protein synthetic machinery to preferentially translate viral mRNAs.

PBCV-1 has several genes encoding proteins that are involved in post-translational modification. In addition to putative glycosyltransferases (see below), PBCV-1 encodes 7 Ser/Thr-protein kinases (P. Valbuzzi *et al.*, unpublished results), one putative Tyr-protein kinase and a putative Tyr phosphatase. Three protein kinases are packaged in the chlorella virus CVK2 virion (Yamada *et al.*, 1996). PBCV-1 also encodes several enzymes involved in post-translational modification such as an ERV/ALR protein, which functions as a protein thiol oxidoreductase (Senkevich *et al.*, 2000), a putative protein disulfide isomerase and a prolyl 4-hydroxylase that converts Pro-containing peptides into hydroxyl-Pro-containing peptides (Ericksson *et al.*, 1999). Moreover, PBCV-1 encodes two putative proteins that interact with ubiquitin, a ubiquitin C-terminal hydrolase and a Skp1 protein. Skp1 proteins belong to the SCF-E3 ubiquitin ligase family that targets cell cycle proteins and other regulatory factors for

degradation (Deshaies, 1999). Finally PBCV-1 encodes at least one putative serine proteinase.

VII. Diversity of chlorella virus genomes

The PBCV-1 genome is a linear 330,744-bp, nonpermuted dsDNA molecule with covalently closed hairpin termini (Rohozinski *et al.*, 1989). The termini consist of 35-nucleotide-long, covalently closed hairpin loops that exist in one of two forms; the two forms are complementary when the 35-nucleotide sequences are inverted (flip-flop) (Zhang *et al.*, 1994). Identical 2221-bp inverted repeats are adjacent to each hairpin end (Strasser *et al.*, 1991). The remainder of the PBCV-1 genome contains primarily single-copy DNA (Girton and Van Etten, 1987). These features can be compared with those of other chloroviruses because the genomic sequences of three more chlorella viruses have been completed recently - viruses NY-2A and AR158 which, like PBCV-1, infect *Chlorella* NC64A and virus MT325 which infects *Chlorella* Pbi. These sequences are available on the website <http://greengene.uml.edu>, which represents work in progress; the sequences have not yet been published or deposited in the public databases. Some comparative data are listed in Table 1. PBCV-1 and all other NC64A virus genomes are ~40% G+C, which is significantly lower than the 67% G+C of the host *Chlorella* NC64A nuclear DNA (Van Etten *et al.*, 1985b). Chlorovirus MT325 that is a Pbi virus has a slightly higher G+C content (~45%). The newly sequenced genomes vary from 315 kbp (MT325) to 369 kbp (NY-2A), reflecting a difference in the number of genes (~330 to ~400).

Some of the additional proteins encoded by NY-2A genes include ubiquitin, chitin synthase, N-acetylglucosaminyl transferase, 6 transposases and 43 homing endonucleases. Furthermore, inteins exist in two of the NY-2A gene products, the α -subunit of ribonucleotide reductase and a putative helicase (L. A. Fitzgerald *et al.*, manuscript in preparation). Not all chlorovirus genes are required for virus replication in the laboratory. For example, extended deletions can occur in the chlorovirus genomes; 27-kb to 37-kbp deletions in PBCV-1 (Landstein *et al.*, 1995) and 30-kbp to 42-kbp deletions in CVK1 (Songsri *et al.*, 1995) are located in the left terminus of the genome. A detailed comparison of the gene contents between viruses should identify a set of highly conserved genes and variable or dispensable genes.

A. Hairpin ends and inverted terminal repeats

The 35-nucleotide sequence of the PBCV-1 hairpin end differs considerably from the hairpin loop sequence (43 nucleotides) of chlorovirus CVK2 (S. Hiramatsu and T. Yamada, unpublished result) (Fig. 2). It is interesting that a single base change in the most distal position of the CVK2 sequence reduces the loop to 33 nucleotides. A region of 15-16 bp immediately adjacent to the hairpin end in the inverted terminal repeats also differs between PBCV-1 and CVK2; however beyond this region, the inverted repeat sequences are nearly identical between the two viruses.

The terminal inverted repeats are 2.2-2.4 kbp in size (Table 2). Although the inverted repeats of PBCV-1 and CVK2 are similar, this is not true for all viruses. The PBCV-1 inverted repeat region was hybridized to 36 other NC64A viruses. Twenty-eight hybridized very well to the probe, however, eight did not hybridize at all, indicating sequence differences. Such rearrangements may be mediated by numerous short repeated sequences that frequently occur at the junctions between the inverted repeats and the single copy region (Yamada and Higashiyama, 1993). Comparison of the nucleotide sequence of the inverted repeats among chloroviruses PBCV-1, CVK2, NY-2A, and AR158 indicated no significant identity in the terminal inverted repeats sequences except for occasional homology islands, a ~600-bp region next to the hairpin loop and a 400~600 bp region immediately adjacent to the single copy region (Nishida *et al.*, 1999b).

Yamada and Higashiyama (1993) detected a site-specific nick in the inverted repeat region of chlorovirus, CVK1. This nick might serve as the initiation point for DNA replication.

B. PBCV-1 genes

In the initial report describing the sequence of the PBCV-1 genome, 702 ORFs of 65 codons or larger were identified and 377 of them were predicted to encode proteins (Li *et al.*, 1995; 1996; Lu *et al.*, 1995; 1996; Kutish *et al.*, 1996). However, mistakes are being detected in the original sequence; often two adjacent ORFs consist of a single ORF. Currently we believe that PBCV has 691 ORFs of 65 codons or larger, of which 366 are protein-encoding. PBCV-1 protein-encoding genes were identified initially by the following criteria: i) A minimal size of 65 codons initiated by an ATG codon. ii) The largest ORF was chosen when competing ORFs overlapped. iii) ORFs with AT-rich (>70%) sequences in the 50 nucleotides

upstream of the putative initiation codons. To date, most of the protein-encoding genes have met these criteria.

Unlike the poxviruses, in which genes near the terminal regions are transcribed toward the termini (Moss, 1996), the 366 PBCV-1 putative protein-encoding genes are evenly distributed on both strands and, with one exception, intergenic space is minimal. In fact, 275 ORFs are separated by less than 100 nucleotides. The exception is a 1788 nucleotide sequence near the middle of the genome. This DNA region, which contains many stop codons in all reading frames, encodes 11 tRNA genes. The 2.2 kb inverted terminal repeat region of the PBCV-1 genome contains 4 ORFs, which are duplicated (Lu *et al.*, 1995). Approximately 50% of the 366 PBCV-1 gene products have been tentatively identified, including some that seem irrelevant to virus replication. Some PBCV-1 genes are closely related to genes of bacteria and their viruses, whereas other PBCV-1 genes appear eukaryotic in origin. Consequently, the chlorella virus genomes contain an interesting mosaic of prokaryotic- and eukaryotic-genes. Some of the PBCV-1 gene products are listed in Table 2. Additional comments on these genes can be found in other reviews (Van Etten and Meints, 1999; Van Etten *et al.*, 2002; Van Etten, 2003).

C. Linearity of gene arrangements in the chlorella viruses

To understand the fundamental organization of the chlorovirus genomes and to identify essential genes for viral replication, it will be necessary to separate highly conserved regions from variable and/or dispensable regions. Nishida *et al.* (1999b) compared the gene arrangement between PBCV-1 and CVK2. Four major variations were detected: (i) insertion of an approximately 20-kbp sequence near the left end of CVK2, (ii) a duplication of the gene for the major capsid protein in CVK2, (iii) deletions/insertions of some ORFs, and (iv) a divergence in the terminal inverted repeat sequences. Despite these changes, colinearity was maintained for most of the PBCV-1 and CVK2 genes.

The recent sequencing of three more chlorella viruses make it possible to compare gene arrangements over their entire genomes. Figure 3 compares the positions of 22 randomly chosen genes among four chlorella viruses. Although a few minor rearrangements occur (mentioned in section D 2), in general, colinearity exists among the NC64A viruses (Fig. 3A). In contrast, there is almost no co-linearity between the genomes of PBCV-1 and MT325, a Pbi

virus (Fig. 3B). Some PBCV-1 genes that are absent in NY-2A and AR158, such as A245R and A646L, are present in MT325, whereas the genes for A544R and A604L that are conserved in NC64A-viruses are missing in MT325. A detailed comparison of the gene contents and the genome architecture between the two major groups of chloroviruses is in progress (L. A. Fitzgerald *et al.*, manuscript in preparation).

D. Patterns of gene rearrangement

1. Large deletions/insertions

The CVK2 genome is approximately 20 kbp larger than the PBCV-1 genome, primarily because of an extra 22.2 kbp sequence close to the left terminus (Fig. 3A) (Chuchird *et al.*, 2002). This 22.2 kbp region has five gene copies of the Vp260-like protein, a possible viral-surface glycoprotein. Although none of these copies occur in the corresponding region in the PBCV-1 genome, four and three copies of the Vp260-like protein encoding genes are also located together at almost the same position in the NY-2A and AR158 genomes. These rearrangements do not appear to be mere insertions/deletions but more complicated gene replacement events (Chuchird *et al.*, 2002).

Four PBCV-1 spontaneously derived antigenic variants were isolated that contain 27- to 37-kbp deletions at the left end of the 330-kbp genome (Landstein *et al.*, 1995). Two of these mutants had deletions that began at nucleotide positions 4.9 kb or 16 kb and ended at position 42 kb. The two deletions probably resulted from recombination at a repeated sequence. The other two mutants, which probably arose from nonhomologous recombination, lacked the entire left-terminal 37-kb of the PBCV-1 genome, including the 2.2-kbp terminal inverted repeats. The deleted left terminus was replaced by the transposition of an inverted 7.7- or 18.5-kbp copy from the right end of the PBCV-1 genome. Similar 30-45 kbp deletions were also obtained with NC64A virus CVK1 after exposure of CVK1-infected cells to UV radiation (Yamada and Higashiyama, 1993; Songsri *et al.*, 1995). These deletions also occurred in the left terminal portion of the virus genome, possibly by homologous recombination. These experiments indicate that 40 to 45 kbp (12-13% of the total genome) at the left end of PBCV-1 and CVK2 genomes are unnecessary for viral replication in the laboratory. Similar deletion mutants occur in the poxviruses (for a review, see Turner and Moyer, 1990) and ASFV (Blasco *et al.*, 1989a; 1989b); these virus genomes also have inverted terminal repeats and hairpin ends

like the chloroviruses. The generation of chlorovirus deletions may be explained by the models proposed for deletions/transpositions in the poxvirus genomes (Turner and Moyer, 1990; Shchelkunov and Totmenin, 1995).

2. Small deletions/insertions

Small DNA deletions/insertions also often occur in chlorella virus genomes. Typically these events consist of one or two genes. An example is seen at the ORF A430L(Vp54)-A431L-432L region in the PBCV-1 genome. The entire A431L-like sequence is missing from this gene cluster in the CVK2 genome, leaving a two gene cluster A430L-432L (Nishida *et al.* 1999b). In addition a duplicate copy of A430L (A430L') is inserted in the CVK2 genome at the region corresponding to PBCV-1 ORF A452L-A454L, resulting in a cluster of A452L-A430L'-A454L (Nishida *et al.*, 1999b). In this case, short 5'GTTTT or 5'CAAAA sequences are located at the rearrangement points and are assumed to be involved in the rearrangements. Similarly, a region surrounding the PBCV-1 A250R (Kcv) gene serves as an example of such rearrangements; this region was sequenced in 17 NC64A viruses. In PBCV-1 the genes for A251R and A252R are located immediately downstream of A250R; however, these genes are absent in 12 of the 17 NC64A viruses (Kang *et al.*, 2004a). In addition, 2 of the viruses had an extra ORF inserted between A248R and A250R. Many other examples of small insertions/deletions occur in the chlorella viruses. For example, virus NY-2A encodes 18 DNA methyltransferases, whereas PBCV-1 has 7 (L. A. Fitzgerald *et al.*, manuscript in preparation).

3. Gene amplification

Eighty-four of the PBCV-1 ORFs resemble one or more other PBCV-1 ORFs forming 26 families. Thirteen families have two members, 8 families have three members, 3 families have six members and 2 families have eight members. One six-member family contains multiple ankyrin-like repeats (Peters and Lux, 1993). Five members in another family encode proteins that resemble the PBCV-1 major capsid protein Vp54, although these genes do not hybridize to one another on Southern blots. These observations suggest some gene amplification mechanisms probably exist in the PBCV-1 genome. Additional examples of gene amplification in other chloroviruses, such as Vp260, are described above. It will be

interesting to identify the regions in the 370 kbp NY-2A genome that are expanded as compared to the 330 kbp PBCV-1 genome. However, it is known that the region corresponding to A292R-A330R in PBCV-1 contains genes that are amplified extensively in NY-2A; A328L (two times), A154L (5 times), A354R (10 times), and A315L (12 times). These amplified genes are at least one reason that the NY-2A genome is larger than PBCV-1.

4. Gene replacements

Comparison of the gene arrangements between chloroviruses has also revealed gene replacements. In the CVK2 genome, a single ORF (corresponding to PBCV-1 A330R) is replaced with a 5 kbp sequence containing the genes encoding chitin synthase (*chs*), UDP-glucose dehydrogenase (*ugdh*), and two other ORFs (Ali *et al.*, 2005). In PBCV-1, the hyaluronan synthase gene (*has*, A98R) is located at nucleotide positions 51-53 kbp in a cluster of A93L-A94L (β -1,3-glucanase)-A98R (hyaluronan synthase)-A100R (glucosamine synthase) (Li *et al.*, 1995). A similar gene cluster occurs at the corresponding region in the CVK2 genome; however, the two internal ORFs are replaced with different genes, *bgl2* encoding another β -1,3-glucanase and *chs2* encoding another chitin synthase. The latter arrangement is also found in the NY-2A and AR158 genomes (Ali *et al.*, 2005).

Often these rearrangement events include a set of genes and it is certainly possible that the set may have genes whose products are functionally related, e.g. like restriction endonucleases and their companion DNA methyltransferases. In fact, the ORF immediately adjacent to *chs2* described above resembles a chitin deacetylase (Ali *et al.*, 2005).

To summarize, the different sizes of the chlorovirus genomes as well as their large and small deletions and insertions suggest that dynamic and frequent rearrangements of virus genomes occur in natural environments. These variations probably result from several mechanisms. The fact that the left end of the chlorovirus genome is tolerant to deletions/insertions/rearrangements suggests that a recombinational “hotspot(s)” in this region allows viruses to exchange genes among themselves and possibly with their host(s). However, despite these differences, the location of most PBCV-1 genes, many of which are probably housekeeping genes, are nearly co-linear in the 330-370 kb NC64A viruses, suggesting similar overall genome organization between these chlorovirus isolates. Given the fact that the chlorella viruses often encode multiple transposases and homing endonucleases, one might

expect the virus genomes to be unstable. However, this does not appear to be the case, at least in the laboratory.

VIII. Cell wall digestion: chitinase, chitosanase, polysaccharide lyase, and β -glucanase

In a normal lytic cycle, PBCV-1 attaches to the surface of host chlorella cells and degrades the cell wall at the point of attachment; the viral core is then released into the host cytoplasm, leaving an empty capsid on the cell wall (Meints *et al.*, 1984). Within 6 to 8 hr p.i., nascent viruses exit the cells after cell lysis. Thus, both the initial and final stages of the PBCV-1 replication cycle require cell wall-digesting activities. A common characteristic of virus-sensitive chlorella strains is a rigid cell wall containing glucosamine in addition to other sugars; glucosamine comprises 7-17% of the total sugars in the cell wall (Meints *et al.*, 1988; Kapaun *et al.*, 1992; Kapaun and Reisser, 1995; Takeda, 1988; 1995). The presence of glucosamine suggests that enzymes degrading polymers of glucosamine, like chitin (β -1,4-linked polymer of N-acetyl-D-glucosamine) and chitosan (β -1,4-linked polymer of D-glucosamine with various degrees of N-acetylation), might be involved in the viral infection process. In fact, when chlorovirus CVK2 proteins are separated into the capsid and core particle fractions (Yamada *et al.*, 1997; Songsri *et al.*, 1997), and assayed by SDS-PAGE with chitosan or chitin as a substrate in the gel matrix (Yamada *et al.*, 1997), several enzymatically active proteins with molecular masses ranging from 35 to 70 kDa are detected in the core fraction. Of these, a 65-kDa protein has the most chitosanase activity and a few 50- to 60-kDa proteins have chitinase activities. Moreover, three PBCV-1 ORFs have significant amino acid sequence identities with bacterial chitinases and chitosanases (Lu *et al.*, 1996). Yamada *et al.* (1997) characterized a chitosanase (*vChta-1*) gene from virus CVK2. This gene, which corresponds to PBCV-1 ORF A292L, encodes two functional chitosanase proteins with apparently different roles in virus replication. The larger 65-kDa chitosanase is packaged in the virion and presumably functions during infection. In contrast, the smaller 37-kDa enzyme remains in the host cytoplasm, where it most likely aids in cell wall digestion during viral release. The enzymatic characterization of the PBCV-1 A292L protein has also been described (Sun *et al.*, 1999).

As for the chitinase genes, a PBCV-1 ORF (A181/182R) encodes an active chitinase with two catalytic domains; the enzyme belongs to the family 18 glycosyl hydrolases (Sun *et al.*, 1999). A A181/182R homolog is present in CVK2 (Hiramatsu *et al.*, 1999; 2000). The first catalytic domain resembles a catalytic sequence from *Saccharopolyspora (Streptomyces) erythraeus* (30% identity) chitinase, whereas the second domain resembles a chitinase from *Ewingella americana* (34.7% identity). The two catalytic domains are linked by a short, proline rich, sequence. This structure suggests that the two vChti-1 chitinase domains might have independent origins (Hiramatsu *et al.*, 1999). A C-terminal-truncated derivative of vChti-1, containing the first catalytic domain, produced chitobiose from either chitotetraose, chitohexaose, or high-molecular mass chitin; this product is typical of an exochitinase. In contrast, an N-terminal-truncated derivative of vChti-1 produced N-acetylglucosamine from chitobiose as well as chitooligosaccharides. Therefore, this domain possessed N-acetylglucosaminidase activity as well as endochitinase activity. The presence of two catalytic domains with different enzymatic properties in the viral enzyme may allow natural substrates to be hydrolyzed in a cooperative fashion (Hiramatsu *et al.*, 2000).

The CVK2 chitinase gene (*vChti-1*) is expressed in virus-infected cells beginning at 120 min p.i. However, the 94 kDa protein product is not incorporated into virions but remains in the medium after cell lysis (Hiramatsu *et al.*, 1999). Similar results were reported for the PBCV-1 A181R/A182R protein (Sun *et al.*, 1999). Another PBCV-1 ORF, A260R, also resembles a chitinase (Sun *et al.*, 1999). This gene is expressed late in infection, and the 55 kDa protein is enzymatically active and packaged in PBCV-1 virions. All of these chitinase and chitosanase genes are widely conserved in the chlorella viruses, suggesting their importance in viral replication.

However, *Chlorella* NC64A cells do not exhibit any drastic morphological changes when treated with vChti-1 and/or vChta-1 chitosanase (Hiramatsu *et al.*, 1999; Yamada *et al.*, 1999). Therefore, additional virus encoded enzymes are probably involved in host cell wall digestion. To detect such cell wall-degrading activities, *E. coli* lysates expressing virus CVK2 genes were assayed by monitoring halo-forming activity using *Chlorella* NC64A cells as a substrate. These experiments revealed two algal-lytic activities (vAL-1 and vAL-2) (Sugimoto *et al.*, 2000). *Val-1* transcription and translation products appear at 60 min p.i. and 90 min p.i., respectively. The vAL-1 protein is not incorporated into the viral particles but remains in the

cell lysate, suggesting it is involved in cell wall digestion during viral release. *Chlorella* NC64A cell walls are digested by vAL-1 under physiological conditions. TLC and MALDI-TOF mass spectrometric analyses of the degradation products (oligosaccharides) revealed that the major oligosaccharides had unsaturated D-glucuronic acid (GlcA) (C4=C5) at the reducing terminus, and a side chain attached at C2 or C3 of GlcA(C4=C5). The side chain consisted primarily of Ara, GlcNAc and Gal. These results indicate that vAL-1 is a novel polysaccharide lyase, cleaving chains of either β - or α -1,4-linked GlcAs (Sugimoto *et al.*, 2004). In addition to *Chlorella* NC64A, vAL-1 lysed cells of four *C. vulgaris* isolates as well as *Chlorella* SAG-241-80 (Sugimoto *et al.*, 2000; Chuchird *et al.*, 2001). A PBCV-1 ORF (A94L) encodes a protein with 26 to 30% amino acid identity with family 16 endo- β 1,3-glucanases from several bacteria. A94L also has the critical amino acids in the catalytic site of family 16 endo- β 1,3- and endo- β 1,3-1,4-glucanases. The A94L recombinant protein hydrolyzed the β 1,3-glucan laminarin and had slightly less hydrolytic activity on β 1,3-1,4-glucan lichenan and barley β -glucan (Sun *et al.*, 2000). The *a94l* gene is expressed early in PBCV-1 infection. Furthermore, the A94R protein appears early during PBCV-1 replication but then disappears by 120 min p.i. The biological function of this β 1,3 glucanase is unknown as approximately 50% of chlorella viruses lack this gene. However, some chlorella viruses lacking A94L homologs contain another gene encoding an endo- β 1,3-glucanase (*bgl2*) (Ali *et al.*, 2005). The functions of these genes and their products are unknown.

The chlorella viruses may be a good source of polysaccharide digesting enzymes because cell wall polysaccharides can differ significantly among chlorella isolates (Yamada and Sakaguchi, 1982). Thus one predicts that some viral enzymes may have some unique activities.

At the same time, the origin and evolution of the genes encoding the polysaccharide-degrading enzymes in the chloroviruses are intriguing. Recently, a gene encoding a chitinase, that shares significant amino acid sequence identity with vChti-1, was discovered in host *Chlorella* NC64A (A. M. Ali *et al.*, manuscript in preparation). Comparison of the structure, nature and function of this gene product with vChti-1 may reveal information on the origin and evolution of the viral genes.

IX. Glycosyltransferases and synthesis of polysaccharides and fucose

A. Glycosylation of the virus structural proteins

Virus PBCV-1 contains six possible glycosyltransferase encoding genes, *a64r*, *a111/114r*, *a222/226r*, *a328l*, *a473l* and *a546l* (Van Etten *et al.*, 2002; Van Etten, 2003). None of these putative glycosyltransferases have an identifiable signal peptide that would target them to the endoplasmic reticulum (ER) or Golgi. Furthermore, the cellular protein localization program PSORT predicts that four of these proteins are located in the cytoplasm and two have a transmembrane domain (Van Etten, 2003). The *a64r* gene encodes a 638 amino acid protein that has four motifs conserved in 'Fringe type' glycosyltransferases. Analysis of 13 PBCV-1 antigenic variants, with differences in the major capsid protein Vp54 glycans, have mutations in *a64r* that correlated with specific antigenic variations. Dual infection experiments with different antigenic variants established that wild-type PBCV-1 could be formed by complementation and recombination of the variants. These results led to the conclusion that *a64r* encodes a glycosyltransferase associated with Vp54 glycan synthesis (Graves *et al.*, 2001). Typically, viral proteins are glycosylated by host-encoded glycosyltransferases located in the ER and Golgi (Doms *et al.*, 1993; Knipe, 1996; Olofsson and Hansen, 1998). Consequently, the glycan portion of virus glycoproteins is host-specific. Therefore, glycosylation of PBCV-1 major capsid protein Vp54 differs from this paradigm. Additional experiments to support this statement are provided in a recent review (Markine-Goriaynoff *et al.*, 2004).

These findings lead to several questions including: Are the Vp54 glycan precursors attached to a lipid carrier such as undecaprenolphosphate, which serves as an intermediate in bacterial peptidoglycan synthesis (Raetz and Whitfield, 2002), or dolichol diphosphate which serves the same function in eukaryotic cells (Reuter and Gabius, 1999)? Could Vp54 glycosylation reflect an ancestral pathway that existed prior to the ER and Golgi formation?

B. Extracellular polysaccharide synthesis

The chloroviruses are also unusual because they encode enzymes involved in the biosynthesis of the linear polysaccharides hyaluronan (also called hyaluronic acid) and/or chitin. Typically, hyaluronan is only found in the extracellular matrix of vertebrates and capsules of a few pathogenic bacteria (DeAngelis, 1999; 2002). It is composed of ~20,000

alternating β -1,4-glucuronic acid and β -1,3-N-acetylglucosamine residues (DeAngelis, 1999). Unexpectedly, virus PBCV-1 contains genes encoding hyaluronan synthase (HAS) (DeAngelis et al., 1997) and two other enzymes involved in the synthesis of hyaluronan precursors, glutamine:fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase (Landstein et al., 1998). All three genes are expressed early during PBCV-1 infection. These results led to the discovery that hyaluronan lyase-sensitive, hair-like fibers begin to accumulate on the surface of PBCV-1 infected host cells by 15 min p.i. (Graves et al., 1999). By 4 hr p.i., the infected cells are covered with a dense fibrous hyaluronan network.

The *has* gene is present in many, but not all, chloroviruses isolated from diverse geographical regions (Graves *et al.*, 1999), suggesting that not all chloroviruses encode hyaluronan. Surprisingly, many chloroviruses that lack a *has* gene, have a gene encoding a functional chitin synthase (CHS). Furthermore, cells infected with these viruses produce chitin fibers on their external surface (Kawasaki *et al.*, 2002). Chitin, an insoluble linear homopolymer of β -1,4-linked N-acetylglucosamine residues, is a common component of insect exoskeletons, shells of crustaceans, and fungal cell walls (Gooday *et al.*, 1986).

Some chloroviruses contain both *has* and *chs* genes and form both hyaluronan and chitin on the surface of the infected cells (Kawasaki *et al.*, 2002; Yamada and Kawasaki, 2005). Finally, a few chloroviruses probably lack both genes because no extracellular polysaccharides are formed on the host surface of cells infected with these viruses (Graves *et al.*, 1999).

The fact that many chloroviruses encode enzymes involved in extracellular polysaccharide biosynthesis suggests that the polysaccharides are important in the virus life cycle. At present this function is unknown; however, we have considered three possibilities.

- (i) The polysaccharides prevent uptake of virus-infected chlorella by the paramecium. Presumably, such infected algae would lyse inside the paramecium and the released virions would be digested by the protozoan. This scenario would be detrimental to virus survival.
- (ii) The viruses have another host that acquires the virus by taking up the polysaccharide-covered algae.
- (iii) Virus-infected cells aggregate, presumably due to the extracellular polysaccharide. This aggregation, which could trap uninfected cells, might aid the virus in finding its next host. However, a complicating factor in understanding the biological role of these two polysaccharides is that some viruses apparently lack both genes.

Ali *et al.* (2005) investigated the genetic relationship between “hyaluronan-synthesizing” and “chitin-synthesizing” viruses by characterizing two genomic regions in the chitin-synthesizing virus CVK2 and comparing them to PBCV-1. One region surrounds the CVK2 *chs* region and the other corresponds to the region containing PBCV-1 *has*. A single PBCV-1 ORF (A330R) is replaced in the CVK2 genome with a 5 kb region containing *chs*, *ugd2* (a second gene encoding UDP-glucose dehydrogenase) and two other ORFs. In CVK2 the location of the PBCV-1 *has* gene is replaced with another *chs* gene (described above in section VII C 4). Some chloroviruses lack *ugd2*. These results suggest that chloroviruses change from “*has* viruses” to “*chs* viruses” or from “*chs* viruses” to “*has* viruses” by exchanging genes (Fig. 4).

These observations also indicate that there is no functional incompatibility between the two genes or their gene products, i.e. hyaluronan and chitin. These conclusions are interesting because it has been suggested that the *has* gene in vertebrates evolved from chitin synthase or cellulose synthase through the addition of a β 1-3 glycosyltransferase activity to a pre-existing β 1-4 glycosyltransferase enzyme, and that the ability to synthesize hyaluronan occurred relatively recently in metazoan evolution (Lee and Spicer, 2000). Recently another *chs*-like gene was reported to be encoded by the 336 kbp *Ectocarpus siliculosus* virus EsV-1 (Delaroque *et al.*, 2001). EsV-1 also belongs to the *Phycodnaviridae*, but is lysogenic, in contrast to the lytic chlorella viruses. No information is available on the expression or function of the EsV-1 *chs*-like gene.

The discovery of viral encoded enzymes involved in polysaccharide synthesis leads to many questions, such as: (i) What is the function of these extracellular polysaccharides? (ii) Why do infected cells expend huge amounts of energy on these processes when they are going to lyse in a few hours? (iii) How were these genes acquired by the viruses? (iv) Which was the original form of viral-encoded polysaccharides, hyaluronan or chitin? (v) How are hyaluronan and chitin polymers synthesized in the host cells and translocated through the cell wall?

C. Fucose synthesis

PBCV-1 also encodes enzymes involved in nucleotide sugar metabolism. Two enzymes encoded by PBCV-1, GDP-D-mannose 4,6 dehydratase (GMD) and the bifunctional GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER) comprise the highly conserved

pathway that converts GDP-D-mannose to GDP-L-fucose. In vitro reconstruction of the biosynthetic pathway using recombinant PBCV-1 GMD and GMER produced in *E. coli* synthesized GDP-L-fucose (Tonetti *et al.*, 2003). Unexpectedly, however, the PBCV-1 GMD, which has been crystallized recently (Rosano *et al.*, 2005), also catalyzes the NADPH-dependent reduction of the intermediate GDP-4-keto-6-deoxy-D-mannose, forming GDP-D-rhamnose. Similar results were obtained with GMD and GMER encoded by virus CVK2 (K. Isono *et al.*, unpublished data). Both fucose and rhamnose are constituents of the glycans attached to the PBCV-1 major capsid protein Vp54; therefore, the virus might encode the enzymes to meet this need. However, both sugars are also components of the uninfected host cell wall (Meints *et al.*, 1988).

X. Relationship of chloroviruses to other viruses

As mentioned in the introduction, the phycodnaviruses are members of the superfamily of viruses referred to as NCLDV and that accumulating evidence indicates that the phycodnaviruses have a long evolutionary history, possibly beginning at the time eukaryotes separated from prokaryotes (over three billion years ago). The evidence includes: (i) Phylogenetic analyses of DNA polymerases place the algal virus enzymes near the root of all eukaryotic δ -DNA polymerases (Villarreal and DeFilippis, 2000). (ii) Phylogenetic analyses of other PBCV-1 encoded proteins often place the proteins near the root of their eukaryotic counterparts, e.g. the K⁺ channel protein Kcv (Plugge *et al.*, 2000) and ornithine decarboxylase (Shah *et al.*, 2004). (iii) Many chlorovirus encoded proteins are either the smallest or among the smallest proteins in their class. Examples include the type II DNA topoisomerase (Dickey *et al.*, 2005), Kcv (Plugge *et al.*, 2000), ornithine decarboxylase (Shah *et al.*, 2004) and lysine di-methyltransferase (Manzur *et al.*, 2003). These proteins could represent progenitors of their more complex relatives. (iv) Some PBCV-1-encoded enzymes are more flexible than those from higher eukaryotic organisms. For example, some virus enzymes carry out two functions whereas more 'advanced' organisms require two separate enzymes to accomplish the same tasks. One interpretation of this observation is that these virus proteins may be progenitor enzymes and thus more precocious than their highly evolved counterparts in higher eukaryotes, where two separate enzymes carry out the function of one virus enzyme. This dual functionality in the PBCV-1 enzymes does not result from gene fusion. Examples

include: a) ornithine decarboxylase which decarboxylates arginine more efficiently than ornithine (Shah *et al.*, 2004). b) dCMP deaminase which also deaminates dCTP and dCDP, as well as the expected dCMP (Y. Zhang *et al.*, manuscript in preparation). c) GDP-D-mannose 4,6 dehydratase not only catalyzes the formation of GDP-4-keto-6-deoxy-D-mannose, which is an intermediate in the synthesis of GDP-L-fucose, the enzyme also reduces the same intermediate to GDP-D-rhamnose (Tonetti *et al.*, 2003). (v) Even though PBCV-1 encodes both prokaryotic- and eukaryotic-like proteins, the 40% G+C content is fairly uniform throughout the genome. This pattern suggests that most of the genes have existed together in the virus for a long time. (vi) The major coat protein from several dsDNA viruses that infect all three domains of life, including bacteriophage PRD1, human adenoviruses, and a virus STIV infecting the Archaea, *Sulfolobus solfataricus*, are structurally similar to that of PBCV-1. This finding led Benson *et al.*, (2004) to suggest that all these viruses may have a common evolutionary ancestor, even though there is no significant amino acid sequence similarity among their proteins. (vii) Finally, one of the earliest eukaryotic cells could have resembled a single celled alga (Yoon *et al.*, 2004).

Some evolutionary biologists have suggested that NCLDVs may be the origin of the nucleus in eukaryotic cells (e.g. Bell, 2001; Pennisi, 2004; Villarreal, 2005), whereas other biologists (Raoult *et al.*, 2004) have suggested that Mimivirus, and by inference the NCLDVs, may represent a fourth domain of life. This later hypothesis resulted from a phylogenetic tree of life derived from the concatenated sequences of seven universally conserved proteins sequences: arginyl-tRNA synthetase, methionyl-tRNA synthetase, tyrosyl-tRNA synthetase, RNA polymerase II largest subunit, RNA polymerase II second largest subunit, PCNA, and 5'-3' exonuclease. Mimivirus formed a branch near the origin of the Eukaryotic domain.

Continuing with this hypothesis, maybe the NCLDVs originally arose from a more complex ancestor and this progenitor ancestor (virus) became associated with different organisms that evolved into various eukaryotic lineages. These lineages placed various demands on the evolving NCLDVs resulting in the loss of some genes, either by donating genes to their hosts or the genes were no longer required for replication of the evolving viruses. The net result is that the NCLDVs now only have nine common genes. Of course over 3 billion years of evolution the viruses also acquired selected genes from their hosts that were necessary for survival.

Eight of the ten viruses with the largest sequenced genomes are NCLDVs and six of them are members of the *Phycodnaviridae* (<http://giantvirus.org>). The largest algal virus genome sequenced to date infects *Emiliana huxleyi* and has a genome of 407 kb that contains ~470 protein encoding genes (Wilson *et al.*, 2005a). However, larger algal virus genomes exist; examples include lytic viruses specific for *Phaeocystis pouchetii* (PpV) (Jacobsen *et al.*, 1996), *Chrysochromulina ericina* (CeV-01B) (Sandaa *et al.*, 2001), and *Pyramimonas orientalis* (PoV-01B) (Sandaa *et al.*, 2001), which have genomes of approximately 485, 510, and 560 kb, respectively. To put the size of these viral genomes into perspective, the smallest bacterium, *Mycoplasma genitalium* has a genome of 580 kb that contains ~470 protein-encoding genes (Fraser *et al.*, 1995) and the smallest Archaea, *Nanoarchaeum equitans*, has a 490 kb genome that contains ~550 protein encoding genes (Waters *et al.*, 2003). Estimates of the minimum genome size required to support life are ~250 protein-encoding genes (e.g. Itaya, 1995; Mushegian and Koonin, 1996). The Mimivirus genome is actually larger than 25 microbial genomes currently in the databases (<http://giantvirus.org>). Thus a big question is: what differentiates large complex DNA viruses from small obligate intracellular microorganisms? It is also obvious that metagenomic sequencing projects are identifying many more genes associated with large dsDNA viruses (e.g. Breitbart *et al.*, 2002; Venter *et al.*, 2004; Tyson *et al.*, 2004). Without doubt many of these will be algal infecting viruses (e.g. Suttle, 2005).

The presumed long evolutionary history of the NCLDVs can also explain why the phycodnaviruses are so diverse. Of the six genera in the family *Phycodnaviridae*, members of two genera in addition to the chlorella viruses have been sequenced, the circular *Emiliana huxleyi* virus (EhV-1) 407 kb genome (Wilson *et al.*, 2005a) and the linear *Ectocarpus siliculosus* virus (EsV-1) 335 kb genome (Delaroque *et al.*, 2001). Including PBCV-1, each of these three viruses codes for several hundred proteins (Dunigan *et al.*, in press). However, it is truly amazing that only 14 protein-encoding genes are common to all three viruses. This means that just these three viruses contain more than a 1000 unique protein encoding genes. Despite the large genetic diversity in these three sequenced phycodnaviruses, phylogenetic analyses of their δ -DNA polymerases (Chen and Suttle., 1996; Wilson *et al.*, 2005a) and a superfamily of archeo-eukaryotic primases (Iyer *et al.*, 2005) indicate that the phycodnaviruses fall into a monophyletic clade within the NCLDVs.

The long evolutionary history can also explain the diversity among viruses that infect *Chlorella* NC64A or *Chlorella* Pbi. The amino acid sequence of protein homologs between *Chlorella* NC64A infecting viruses can differ by as much as 25%. This difference can be ~50% between homologs from NC64A and Pbi viruses. This diversity can be used to identify amino acid substitutions that alter the functional properties of proteins, e.g. as has been done with the chlorovirus encoded K⁺ channels (Kang *et al.*, 2004b; Gazzarrini *et al.*, 2004).

In addition to the genomic sequence of PBCV-1, sequences of several other chlorella virus genomes are now available, and they make it possible to determine a pan-genome of chlorella virus, which consists of a core genome shared by all isolates and a dispensable genome consisting of partially shared and isolate-specific genes. The core genes, which are highly conserved among different viral isolates, reflect their importance in the basic infection cycle. The variable genes are related to recent growth history and the range of infection of individual virus isolates. One outcome of this diversity is that the total number of chlorovirus protein-encoding genes is much larger than that from any one virus.

XI. Perspectives

It has been ~25 years since the first chloroviruses were described. Sequencing the 330 kb PBCV-1 genome about 10 years ago revealed, and is continuing to reveal, many unexpected genes such as genes encoding proteins involved in polysaccharide biosynthesis, polyamine biosynthesis, and ion transport proteins. The discovery of these genes, and the fact that many of their gene products are easy to work with, has led to a recent expansion in the number of investigators who are studying chloroviruses and their genes. However, many fundamental events in chlorovirus replication have only begun to be studied as indicated by some of the questions posed in this review. The biggest handicap facing studies on the chloroviruses, as well as all of the phycodnaviruses, is the lack of a functional molecular genetic system. However, new tools are coming into play with the chlorella viruses. For example, microarrays containing all of the putative PBCV-1 protein encoding genes are now being used to study PBCV-1 gene expression, proteomic studies are identifying all of the virion-encoded proteins packaged in the virion, and the USA Department of Energy has recently agreed to sequence the PBCV-1 host *Chlorella* NC64A. These new tools should lead to a greater understanding of virus-host relationships in the near future.

Acknowledgments

Research in the Van Etten laboratory has been supported in part by Public Health Service grant GM32441 from the National Institute of General Medical Sciences, the National Science Foundation grant EF-0333197, and grant P20-RR15635 from the COBRE program of the National Center for Research Resources.

REFERENCES

- Ali, M. M. A., Kawasaki, T., and Yamada, T. (2005). Genetic rearrangements on the chlorovirus genome that switch between hyaluronan synthesis and chitin synthesis. *Virology*, **342**, 102-110.
- Becker, B., Lesenmann, D. E., and Reisser, W. (1993). Ultrastructural studies on a chlorella virus from Germany. *Arch. Virol.* **130**, 145-155.
- Belfield, G. P. and Tuite, M. F. (1993). Translation elongation factor 3: a fungus-specific translation factor? *Mol. Microbiol.* **9**, 411-418.
- Bell, P.J. (2001). Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus. *J. Mol. Evol.* **53**, 251-256.
- Benson, S. D., Bamford, J. K. H., Bamford, D. H., and Burnett, R. M. (2004). Does common architecture reveal a virus lineage spanning all three domains of life? *Mol. Cell* **16**, 673-685.
- Blasco, R., Agüero, M., Almendral, J. M., and Vinuela, E. (1989a). Variable and constant regions in African swine fever virus DNA. *Virology* **168**, 330-338.
- Blasco, R., De La Dega, I., Almazan, F., Agüero, M., and Vinuela, E. (1989b). Genetic variation of African swine fever virus: variable regions near the ends of the viral DNA. *Virology* **173**, 251-257.
- Breibart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., and Rohwer, F. (2002). Genomic analysis of uncultured marine communities. *Proc. Natl. Acad. Sci. USA* **99**, 14250-14255.
- Brussaard, C.P.D. (2004). Viral control of phytoplankton populations - a review. *J. Eukarot. Microbiol.* **51**, 125-138.

- Bubeck, J. A. and Pfitzner, A. J. (2005). Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *J. Gen. Virol.* **86**, 2871-2877.
- Caspar, D.L.D., and Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1-24.
- Chakraborty, K. (2001). Translational regulation by ABC systems. *Res. Microbiol.* **152**, 391-399.
- Chen, F. and Suttle, C. A. (1996). Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* **219**, 170-178.
- Cho, H. H., Park, H. H., Kim, J. O., and Choi, T. J. (2002). Isolation and characterization of chlorella viruses from freshwater sources in Korea. *Mol. Cells* **14**, 168-176.
- Chuchird, N., Hiramatsu, S., Sugimoto, I., Fujie, M., Usami, S., and Yamada, T. (2001). Digestion of *Chlorella* cells by chlorovirus-encoded polysaccharide degrading enzymes. *Microbes and Environment*, **16**, 197-205.
- Chuchird, N., Nishida, K., Kawasaki, T., Fujie, M., Usami, S., and Yamada, T. (2002). A variable region on the chlorovirus CVK2 genome contains five copies of the gene for Vp260, a viral-surface glycoprotein. *Virology* **295**, 289-298.
- Cobbold, C., Whittle, J. T., and Wileman, T. (1996). Involvement of the endoplasmic reticulum in the assembly and envelopment of African swine fever virus. *J. Virol.* **70**, 8382-8390.
- DeAngelis, P. L. (1999). Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell Mol. Life Sci.* **56**, 670-682.
- DeAngelis, P. L. (2002). Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat. Rec.* **268**, 317-326.
- DeAngelis, P. L., Jing, W., Graves, M. V., Burbank, D. E., and Van Etten, J. L. (1997). Hyaluronan synthase of chlorella virus PBCV-1. *Science* **278**, 1800-1803.
- Delaroque, N., Müller, D. G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001). The complete DNA sequence of the *Ectocarpus siliculosus* virus genome. *Virology* **287**, 112-132.
- Deshais, R. J. (1999). SCF and Cullin/RING H2-based ubiquitin ligases. *Annu. Rev. Cell. Dev. Biol.* **15**, 435-467.
- Dickey, J.S., Choi, T.-J., Van Etten, J.L., and Osheroff, N. (2005). Chlorella virus Marburg topoisomerase II: high DNA cleavage activity as a characteristic of chlorella virus type II

enzymes. *Biochemistry* **44**, 3899-3908.

Dodds, J. A. (1979). Viruses of marine algae. *Experientia* **35**, 440-442.

Doms, R. W., Lamb, R. A., Rose, J. K., and Helenius, A. (1993). Folding and assembly of viral membrane proteins. *Virology* **193**, 545-562.

Dunigan, D.D., Fitzgerald, L.A., and Van Etten, J.L. (200x) Phycodnaviruses: a peak at genetic diversity. *Virus Research* (in press).

Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999). Evidence for 4-hydroxyproline in viral proteins: characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J. Biol. Chem.* **274**, 22131-22134.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., et al. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403.

Frohns, F., Käsmann, A., Schäfer, B. Kramer, D., Mehmel, M., Kang, M., Van Etten, J.L., Gazzarrini, S., Moroni, A., and Thiel, G. (2006). Potassium ion channels of chlorella viruses cause rapid depolarization of host cells during infection. *J. Virol.* (in press).

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541-548.

Gazzarrini S, Kang M, Van Etten JL, Tayefeh S, Kast SM, DiFrancesco D, Thiel G, and Moroni A. (2004). Long distance interactions within the potassium channel pore are revealed by molecular diversity of viral proteins. *J. Biol. Chem.* **279**, 28443-28449.

Gazzarrini, S., Severino, M., Morandi, M., Lombardi, M., DiFrancesco, D., et al. 2003. The viral K⁺ channel Kcv: Structural and functional features. *FEBS Lett.* **552**, 12-16.

Girton, L., and Van Etten, J. L. (1987). Restriction site map of the chlorella virus PBCV-1 genome. *Plant Mol. Biol.* **9** 247-257.

Gong, C. and Shuman, S. (2002). Chlorella virus RNA triphosphatase: mutational analysis and mechanism of inhibition by tripolyphosphate. *J. Biol. Chem.* **277**, 15317-15324.

Gooday, G.W., Humphreys, A.M., and McIntosh, W.H. (1986). Role of chitinases in fungal growth. In "Chitin in Nature and Technology" (R. Muzzarelli, C. Jeuniaux, C., and G. W. Gooday, Eds.) p 83-91. Plenum Press, New York.

Grabherr, R., Strasser, P., and Van Etten, J. L. (1992). The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. *Virology* **188**, 721-731.

Graves, M. V., Bernadt, C. T., Cerny, R., and Van Etten, J. L. (2001). Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. *Virology* **285**, 332-345.

Graves, M.V., Burbank, D. E., Roth, R., Heuser, J., DeAngelis, P. L., and Van Etten, J. L. (1999). Hyaluronan synthesis in virus PBCV-1 infected chlorella-like green algae. *Virology* **257**, 15-23.

Graves, M V., and Meints, R. H. (1992). Characterization of the gene encoding the most abundant in vitro translation product from virus-infected chlorella-like algae. *Gene* **113**,149-155.

Hambly, E. and Suttle, C.A. (2005). The virosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* **8**, 444-450.

Hiramatsu, S., Fujie, M., Usami, S., Sakai, K., and Yamada, T. (2000). Two catalytic domains of chlorella virus CVK2 chitinase. *J. Biosci. Bioeng.* **89**, 252-257.

Hiramatsu, S., Ishihara, M., Fujie, M., Usami, S., and Yamada, T. (1999). Expression of a chitinase gene and lysis of the host cell wall during chlorella virus CVK2 infection. *Virology* **260**, 308-315.

Ho, C. K , Gong, C., and Shuman, S. (2001). RNA triphosphatase component of the mRNA capping apparatus of *Paramecium bursaria* Chlorella virus 1. *J. Virol.* **75**, 1744-1750.

Ho, C. K , Van Etten, J. L., and Shuman, S. (1996). Expression and characterization of an RNA capping enzyme encoded by chlorella virus PBCV-1. *J. Virol.* **70**, 6658-6664.

Hoshina, R., Kamako, S., and Imamura, N. (2004). Phylogenetic position of endosymbiotic green algae in *Paramecium bursaria* Ehrenberg from Japan. *Plant Biol.* **6**, 447-453.

Hoshina, R., Kato, Y., Kamako, S., and Imamura, N. (2005). Genetic evidence of “American” and “European” type symbiotic algae of *Paramecium bursaria* Ehrenberg. *Plant Biol.* **7**, 526-532.

Itaya, M. (1995). An estimation of minimal genome size required for life. *FEBS Lett* **362**, 257-260.

Iyer, L. M., Aravind, L., and Koonin, E. V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* **75**, 11720-11734.

Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic Acids Res.* **33**, 3875-3896.

Jacobsen, A., Bratbak, G., and Haldal, M. (1996). Isolation and characterization of a virus infecting *Phaeocystis pouchetii* Prymnesiophyceae. *J. Phycol.* **32**, 923-927.

Johnson, G. P., Goebel, S. J., and Paoletti, E. (1993). An update on the vaccinia virus genome. *Virology* **196**, 381-401.

Kang, M., Dunigan, D.D., and Van Etten, J. L. (2005). Chlorovirus: a genus of Phycodnaviridae that infects certain chlorella-like green algae. *Mol Plant Pathol*, **6**, 212-224.

Kang, M., Graves, M., Mehmel, M., Moroni, A., Gazzarrini, S., Thiel, G., Gurnon, J. R., and Van Etten, J. L. (2004a) Genetic diversity in chlorella viruses flanking *kcv*, a gene that encodes a potassium ion channel protein. *Virology* **326**, 150-159.

Kang, M., Moroni, A., Gazzarrini, S., DiFrancesco, D., Thiel, G., Severino, M., and Van Etten, J.L. (2004b). Small potassium ion channel proteins encoded by chlorella viruses. *Proc. Natl. Acad. Sci. USA* **101**, 5318-5324.

Kang, M., Moroni, A., Gazzarrini, S., and Van Etten, J. L. (2003). Are algal viruses a rich source of ion channel genes? *FEBS Lett.* **18**, 2-6.

Kapaun, E., Loos, E., and Reisser, W. (1992). Cell wall composition of virus-sensitive symbiotic chlorella species. *Phytochemistry* **31**, 3103-3104.

Kapaun, E., and Reisser, W. (1995). A chitin-like glycan in the cell wall of a *Chlorella* sp. Chlorococcales, Chlorophyceae. *Planta* **197**, 577-582.

Kawakami, H., and Kawakami, N. (1978). Behavior of a virus in a symbiotic system. *Paramecium bursaria-zoochlorella*. *J. Protozool.* **25**, 217-225.

Kawasaki, T., Tanaka, M., Fujie, M., Usami, S., Sakai, K., and Yamada, T. (2002). Chitin synthesis in chlorovirus CVK2-infected *Chlorella* cells. *Virology* **302**, 123-131.

Kawasaki, T., Tanaka, M., Fujie, M., Usami, S., and Yamada, T. (2004). Immediate early genes expressed in chlorovirus infection. *Virology* **318**, 214-223.

Knipe, D. M. (1996). Virus-host cell interactions. In "Fields Virology, 3rd ed." (B.N. Fields, D.M. Knipe, P.M. Howley, R.M. Chanock, J.L. Melnick, T.P. Monath, B. Roizman, and S.E. Straus, Eds.) pp. 273-299. Lippincott-Raven Publ, Philadelphia.

Kutish, G. F., Li, Y., Lu, Z., Furuta, M., Rock, D. L., and Van Etten, J. L. (1996). Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map positions 182 to 258. *Virology* **223**,303-317.

Kuznetsov, Y. G., Gurnon, J. R., Van Etten, J. L., and McPherson, A. (2005) Atomic force microscopy investigation of a chlorella virus, PBCV-1. *J. Str. Biol.* **149**, 256-263.

- Landstein, D., Burbank, D. E., Nietfeldt, J. W., and Van Etten, J. L. (1995). Large deletions in antigenic variants of the chlorella virus PBCV-1. *Virology* **214**, 413-420.
- Landstein, D., Graves, M. V., Burbank, D. E., DeAngelis, P., and Van Etten, J. L. (1998). Chlorella virus PBCV-1 encodes functional glutamine: fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase enzymes. *Virology* **250**, 388-396.
- Lee, D.A., Graves, M.V., Van Etten, J.L. and Choi, T-J. (2005). Functional implication of the tRNA genes encoded in the chlorella virus PBCV-1 genome. *Plant Pathol. J.* (in press).
- Lee, J. Y. and Spicer, A. P. (2000). Hyaluronan: A multifunctional megaDalton, stealth molecule. *Curr. Opin. Cell Biol.* **12**, 581-586.
- Lemke, P. A. (1976). Viruses of eucaryotic microorganisms. *Annu. Rev. Microbiol.* **30**, 105-145
- Li, Y., Lu, Z., Burbank, D. E., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995). Analysis of 43 kb of the chlorella virus PBCV-1 330 kb genome: map position 45 to 88. *Virology* **212**, 134-150.
- Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., et al. (1997). Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology* **237**, 360-377
- Lu, Z., Li, Y., Que, Q., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1996). Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map positions 88 to 182. *Virology* **216**, 102-123.
- Lu, Z., Li, Y., Zhang, Y., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995). Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome. *Virology* **206**, 339-352.
- Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A. W., et al. (2003). A dimeric viral SET domain methyltransferase specific to Lys²⁷ of histone H3. *Nat. Struct. Biol.* **10**, 187-196.
- Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., and Vanderplasschen, A. (2004). Glycosyltransferases encoded by viruses. *J. Gen. Virol.* **85**, 2741-2754.
- Mehmel, M., Rothermel, M., Meckel, T., Van Etten, J. L, Moroni, A., and Thiel, G. (2003). Possible function for virus encoded K⁺ channel Kcv in the replication of chlorella virus PBCV-1 *FEBS Lett.* **552**, 7-11.
- Meints, R. H., Burbank, D. E., Van Etten, J. L., and Lamport, D. T. A. (1988). Properties of the chlorella receptor for the virus PBCV-1. *Virology* **164**, 15-21.

- Meints, R. H., Lee, K., Burbank, D. E., and Van Etten, J. L. (1984). Infection of a chlorella-like alga with the virus, PBCV-1: ultrastructural studies. *Virology* **138**, 341-346.
- Meints, R. H., Lee, K., and Van Etten, J. L. (1986). Assembly site of the virus PBCV-1 in a chlorella-like green alga: ultrastructural studies. *Virology* **154**, 240-245.
- Meints, R. H., Van Etten, J. L., Kuczmariski, D., Lee, K., and Ang, B. (1981). Viral infection of the symbiotic chlorella-like alga present in *Hydra viridis*. *Virology* **113**, 698-703.
- Moss, B. (1996). Poxviridae: The viruses and their replication. In "Fields Virology" (B. N. Fields, D. M. Knipe, P. M. Howley, R. M. Chanock, J. L. Melnick, T. P. Monath, B. Roizman, and S. E. Straus, Eds.) pp. 2637-2671. Lippincott-Raven, Philadelphia.
- Müller, D.G., Kapp, M., and Knippers, R. (1998). Viruses in marine brown algae. *Adv. Virus Res.* **50**, 49-67.
- Mushegian, A. R. and Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA* **93**, 10268-10273.
- Nandhagopal, N., Simpson, A., Gurnon, J. R., Yan, X., Baker, T. S., et al. (2002). The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc. Natl. Acad. Sci. USA* **99**, 14758-14763.
- Nelson, M., Burbank, D. E., and Van Etten, J. L. (1998). Chlorella viruses encode multiple DNA methyltransferases. *Biol. Chem.* **379**, 423-428.
- Nelson, M., Zhang, Y., and Van Etten, J. L. (1993). DNA methyltransferases and DNA site-specific endonucleases encoded by chlorella viruses. In "DNA Methylation: Molecular Biology and Biological Significance" (J.P. Jost, H.P. Saluz, Eds.) pp. 186-211. Birkhauser Verlag, Basel
- Nishida, K., Kawasaki, T., Fujie, M., Usami, S., and Yamada, T. (1999a). Aminoacylation of tRNAs encoded by chlorella virus CVK2. *Virology* **263**, 220-229.
- Nishida, K., Kimura, Y., Kawasaki, T., Fujie, M., and Yamada, T. (1999b). Genetic variation of chlorella viruses: variable regions localized on the CVK2 genomic DNA. *Virology* **255**, 376-384.
- Olofsson, S., and Hansen, J. E. S. (1998). Host cell glycosylation of viral glycoproteins ? a battlefield for host defence and viral resistance. *Scand. J. Infect Dis.* **30**, 435-440.
- Onimatsu, H., Sugimoto, I., Fujie, M., Usami, S., and Yamada, T. (2004). Vp130, a chloroviral surface protein that interacts with the host *Chlorella* cell wall. *Virology* **319**, 71-80.
- Pennisi, E. (2004). Evolutionary biology. The birth of the nucleus. *Science* **305**, 1641-1644.

- Peters, L. L., and Lux, S. E. (1993). Ankyrins: structure and function in normal cells and hereditary spherocytes. *Semin Hematol.* **30**, 85-118.
- Plugge, B., Gazzarrini, S., Nelson, M., Cerana, R., Van Etten, J. L., et al. (2000). A potassium channel protein encoded by chlorella virus PBCV-1. *Science* **287**, 1641-1644.
- Raetz, C. R., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635-700.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Clavarie J-M. The 1.2-megabase genome sequence of mimivirus. *Science* **306**, 1344-1350.
- Reisser, W., Burbank, D. E., Meints, R. H., Becker, B., and Van Etten, J. L. (1991). Viruses distinguish symbiotic *Chlorella* spp. of *Paramecium bursaria*. *Endocytobiosis Cell Res.* **7**, 245-251.
- Reisser, W., Burbank, D. E., Meints, S. M., Meints, R. H., Becker, B., and Van Etten, J. L. (1988). A comparison of viruses infecting two different chlorella-like green algae. *Virology* **167**, 143-149.
- Reuter, G., and Gabius, H. J. (1999). Eukaryotic glycosylation: whim of nature or multipurpose tool. *Cell Mol. Life Sci.* **55**, 368-422.
- Rohozinski, J., Girton, L. E., and Van Etten, J. L. (1989). Chlorella viruses contain linear nonpermuted double stranded DNA genomes with covalently closed hairpin ends. *Virology* **168**, 363-369.
- Rosano, C., Zuccotti, S., Sturla, L., Fruscione, F., Tonetti, M., and Bolognesi, M. (2005). Quaternary assembly and crystal structure of GDP-D-mannose 4,6 dehydratase from *Paramecium bursaria* chlorella virus. *Biochem. Biophys. Res. Commun.* (in press).
- Sachs, A. (2000). Cell cycle-dependent translation initiation: IRES elements prevail. *Cell* **101**, 243-245.
- Sandaa, R. A., Heldal, M., Catsberg, T., Thyraug, R., and Bratbak, G. (2001). Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* **290**, 272-280.
- Schuster, A.M., Girton, L., Burbank, D. E., and Van Etten, J. L. (1986). Infection of a chlorella-like alga with the virus PBCV-1: transcriptional studies. *Virology* **148**, 181-189.
- Schuster, A. M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., et al. (1990). Transcription and sequence studies of a 4.3 kbp fragment from a dsDNA eukaryotic algal virus. *Virology* **176**, 515-523.

Senkevich, T. G., White, C.L., Koonin, E. V., and Moss, B. (2000). A viral member of the ERV1/ALR protein family participates in a cytoplasmic pathway of disulfide bond formation. *Proc. Natl. Acad. Sci. USA* **97**,12068-12073.

Shah, R., Coleman, C.S., Mir, K., Baldwin, J., Van Etten, J.L. Grishin, N.V., Pegg, A.E., Stanley, B.A., and Phillips, M.A. (2004). *Paramecium bursaria* chlorella virus-1 encodes an unusual arginine decarboxylase that is a close homolog of eukaryotic ornithine decarboxylases. *J. Biol. Chem.* **279**, 35760-35767.

Shchelkunov, S. N., and Totmenin, A. V. (1995). Two types of deletions in orthopoxvirus genomes. *Virus Genes* **9**, 231-245.

Simpson, A. A., Nandhagopal, N., et al., (2003). Structural analysis of Phycodnaviridae and Iridoviridae. *Acta Crystallogr* **D59**, 2053-2959.

Skrdla, M. P., Burbank, D. E., Xia, Y., Meints, R. H., and Van Etten, J. L. (1984). Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. *Virology* **135**, 308-315.

Songsri, P., Hamazaki, T., Ishikawa, Y., and Yamada, T. (1995). Large deletions in the genome of chlorella virus CVK1. *Virology* **214**, 405-412.

Songsri, P., Hiramatsu, M., Fujie, M., and Yamada, T. (1997). Proteolytic processing of chlorella virus CVK2 capsid proteins. *Virology* **227**, 252-254.

Strasser, P., Zhang, Y. P., Rohozinski, J., and Van Etten, J. L. (1991). The termini of the chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology* **180**, 763-769.

Suda, K., Tanji, Y., Hori, K., and Unno, H. (1999). Evidence for a novel chlorella virus encoded alginase. *FEMS Microbiol. Lett.* **180**, 45-53.

Sugimoto, I., Hiramatsu, S., Murakami, D., Fujie, M., Usami, S., and Yamada, T. (2000). Algal-lytic activities encoded by chlorella virus CVK2. *Virology* **277**, 119-126.

Sugimoto, I., Onimatsu, H., Fujie, M., Usami, S., and Yamada, T. (2004). vAL-1, a novel polysaccharide lyase encoded by chlorovirus CVK2. *FEBS Lett.* **559**, 51-56.

Sun, L., Adams, B., Gurnon, J. R., Ye, Y., and Van Etten, J. L. (1999). Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. *Virology* **263**, 376-387.

Sun, L., Gurnon, J. R., Adams, B. J., Graves, M. V., and Van Etten, J. L. (2000). Characterization of a β -1,3-glucanase encoded by chlorella virus PBCV-1. *Virology* **276**, 27-36.

Suttle, C. A. (2000). Ecological, evolutionary, and geochemical consequences of viral infection of cyanobacteria and eukaryotic algae. In "Viral Ecology" (C. Hurst, Ed.) pp. 247-296. Academic Press, New York.

Suttle, C. A. (2005). Viruses in the sea. *Nature* **437**, 356-361.

Takeda, H. (1988). Classification of *Chlorella* strains by cell wall sugar composition. *Phytochemistry* **27**, 3822-3826.

Takeda, H. (1995). Cell wall composition and taxonomy of symbiotic chlorella from paramecium and acanthocystis. *Phytochemistry* **40**, 457-459.

Tanner, N. K. and Linder, P. (2001). DexD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol. Cell* **8**, 251-262.

Tonetti, M., Zanardi, D., Gurnon, J. R., Fruscione, F., Armirotti, A., et al. (2003). Chlorella virus PBCV-1 encodes two enzymes involved in the biosynthesis of GDP-L-fructose and GDP-D-rhamnose. *J. Biol. Chem.* **278**, 21559-21565.

Turner, P. C., and Moyer, R. W. (1990). The molecular pathogenesis of poxviruses. *Curr. Top. Microbiol. Immunol.* **163**, 125-151.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R.J., Richardson, R. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

Van Etten, J. L. (2003). Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.*, **37**, 153-195.

Van Etten, J.L., Burbank, D.E., Kuczarski, D., and Meints, R.H. (1983a). Virus infection of culturable chlorella-like algae and development of a plaque assay. *Science* **219**, 994-996.

Van Etten, J.L., Burbank, D. E., Joshi, J., and Meints, R. H. (1984). DNA synthesis in a chlorella-like alga following infection with the virus PBCV-1. *Virology* **134**, 443-449.

Van Etten, J. L., Burbank, D. E., and Meints, R. H. (1986). Replication of the algal virus PBCV-1 in UV-irradiated chlorella. *Intervirology* **26**, 115-120.

Van Etten, J. L., Burbank, D.E., Schuster, A. M., and Meints, R. H. (1985a). Lytic viruses infecting a chlorella-like alga. *Virology* **140**, 135-143.

Van Etten, J. L., Burbank, D. E., Xia, Y., and Meints, R. H. (1983b). Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. *Virology* **126**, 117-125.

- Van Etten, J.L., Graves, M. V., Muller, D. G., Boland, W., and Delaroque, N. (2002). *Phycodnaviridae* - large DNA algal viruses. *Arch. Virology* **147**, 1479-1516.
- Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991). Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* **55**, 586-620.
- Van Etten, J. L., and Meints, R. H. (1999). Giant viruses infecting algae. *Ann. Rev. Microbiol.* **53**, 447-494.
- Van Etten, J.L., Meints, R. H., Burbank, D. E., Kuczmarski, D., Cuppels, D. A., and Lane, L. C. (1981). Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. *Virology* **113**, 704-711.
- Van Etten, J. L., Meints, R. H., Kuczmarski, D., Burbank, D. E., and Lee, K. (1982). Viruses of symbiotic chlorella-like algae isolated from *Paramecium bursaria* and *Hydra viridis*. *Proc. Natl. Acad. Sci. USA* **79**, 3867-3871.
- Van Etten, J.L., Schuster, A.M., Girton, L., Burbank, D.E., Swinton, D., and Hattman, S. (1985b). DNA methylation of viruses infecting a eukaryotic chlorella-like green alga. *Nucleic Acids Res.* **13**, 3471-3478.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D. et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.
- Villarreal, L. P., and DeFilippis, V. R. (2000). A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* **74**, 7079-7084.
- Villarreal, L.P. (2005). *Viruses and the Evolution of Life*. American Society of Microbiology Press. pp. 395. Washington D.C.
- Wang, I.-N., Li, Y., Que, Q., Bhattacharya, M., Lane, L. C., et al. (1993). Evidence for virus-encoded glycosylation specificity. *Proc. Natl. Acad. Sci. USA* **90**, 3840-3844.
- Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton GG, Simon M, Soll D, Stetter KO, Short JM, and Noordewier M. *et al.* (2003). The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A* **100**, 12984-12988.
- Wilson, W.H., Schroeder, D.C., Allen, M.J., Holden M.T., Parkhill, J., Barrell, B.G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M.A. Price, C., Rabinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. *et al.* (2005a). Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* **309**, 1090-1092.
- Wilson, W. H., Van Etten, J. L., Schroeder, D. S., Nagasaki, K., Burssaard, C., Delaroque, N., Bratbak, G., and Suttle, C. (2005b). *Phycodnaviridae*. In "Virus Taxonomy, VIIIth Report of

the ICTV” (C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger and L. A. Ball Eds.), pp. 163-175. Elsevier/Academis Press, London.

Wolf, S., Maier, I., Katsaros, C., and Müller, D. G. (1998). Virus assembly in *Hincksia hincksiae* (Ecocarpales, Phaeophyceae). An electron and fluorescence microscopic study. *Protoplasma* **203**, 153-167.

Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Revs.* **64**, 69-114.

Yamada, T., Chuchird, N., Kawasaki, T., Nishida, K., and Hiramatsu, S. (1999). *Chlorella* virus as a source of novel enzymes. *J. Biosci. Bioeng.* **88**, 353-361.

Yamada, T., Fukuda, T., Tamura, K., Furukawa, S., and Songsri, P. (1993). Expression of the gene encoding a translational elongation factor 3 homolog of chlorella virus CVK2. *Virology* **97**, 742-750.

Yamada, T., Furukawa, S., Hamazaki, T., and Songsri, P. (1996). Characterization of DNA-binding proteins and protein kinase activities in chlorella virus CVK2. *Virology* **219**, 395-406.

Yamada, T., and Higashiyama, T. (1993). Characterization of the terminal inverted repeats and their neighboring tandem repeats in the chlorella CVK1 virus genome. *Mol. Gen. Genet.* **241**, 554-563.

Yamada, T., Higashiyama, T., and Fukuda, T. (1991). Screening of natural waters for viruses which infect chlorella cells. *Appl. Environ. Microbiol.* **57**, 3433-3437.

Yamada, T., Hiramatsu, S., Songsri, P., and Fujie, M. (1997). Alternative expression of a chitosanase gene produces two different proteins in cells infected with chlorella virus CVK2. *Virology* **230**, 361-368.

Yamada, T., and Kawasaki, T., (2005). Microbial synthesis of hyaluronan and chitin: new approaches. *J. Biosci. Bioeng.* **99**, 521-528.

Yamada, T., and Sakaguchi, K. (1982). Comparative studies on chlorella cell walls: induction of protoplast formation. *Arch. Microbiol.* **132**, 10-13.

Yamada, T., Tamura, K., Aimi, T., and Songsri, P. (1994). Self-splicing group I introns in eukaryotic viruses. *Nucleic Acids Res.* **22**, 2532-2537.

Yan, X., Olson, N. H., Van Etten, J. L., Bergoin, M., Rossmann, M. G., and Baker, T. S. (2000). Structure and assembly of large, lipid-containing, dsDNA viruses. *Nat. Struct. Biol.* **7**, 101-103.

Yanez, R. J., Rodriguez, J. M., Nogal, M. L., Yuste, L., Enriquez, C., et al. (1995). Analysis of the complete nucleotide sequence of African swine fever virus. *Virology* **208**, 249-278.

- Yonker, C. R., Caldwell, K. D., Giddings, J. C., Van Etten, J. L. (1985). Physical characterization of PBCV virus by sedimentation field flow fractionation. *J. Virol. Methods* **11**, 145-160.
- Yoon, H.S., Hackett, J.D., Cinigliz, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809-818.
- Zhang, Y., Adams, B., Sun, L., Burbank, D. E., and Van Etten, J. L. (2001). Intron conservation in the DNA polymerase gene encoded by chlorella viruses. *Virology* **285**, 313-321.
- Zhang, Y., Burbank, D. E., and Van Etten, J. L. (1988). Chlorella viruses isolated in China. *Appl. Environ. Microbiol.* **54**, 2170-2173.
- Zhang, Y., Calin-Jageman, I., Gurnon, J. R., Choi, T.-J., Adams, B., Nicholson, A. W., and Van Etten, J. L. (2003). Characterization of a chlorella virus PBCV-1 encoded ribonuclease III. *Virology* **317**, 73-83.
- Zhang, Y., Strasser, P., Grabherr, R., and Van Etten, J. L. (1994). Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. *Virology* **202**, 1079-1082.

Legend to figures.

Figure 1. Three-dimensional image reconstruction of chlorella virus PBCV-1 from cryo-electron micrographs and atomic force micrographs. **A.** The virion capsid consists of 12 pentasymmetrons and 20 trisymmetrons. Five trisymmetrons are highlighted in the reconstruction (blue) and a single pentasymmetron is colored yellow. A pentavalent capsomer (white) lies at the center of each pentasymmetron. Each pentasymmetron consists of one pentamer plus 30 trimers. Eleven capsomers form the edge of each trisymmetron (black dots) and therefore each trisymmetron has 66 trimers (Yan *et al.*, 2000). **B.** Dense material (blue arrow) [cell wall digesting enzyme(s)?] is present at each vertex (red arrow) between the vertex and the membrane. **C-F.** Atomic force microscopy images of the surfaces of PBCV-1 virions showing the pentameric arrangements of proteins around the fivefold vertices, with a unique protein on the vertex (Kuznetsov *et al.*, 2005).

Figure 2. Comparison of the nucleotide sequences at the PBCV-1 and CVK2 terminal hairpin ends. The most distal portions of the terminal inverted repeats (shown by arrows) are relatively similar between PBCV-1 and CVK2 DNAs (unidentical bases are shown in red), whereas the sequences of the covalently closed hairpin loops (35 nucleotides for PBCV-1 and 43 nucleotides for CVK2) are completely different. The hairpin loops exist in one of two forms (shown in blue and green); the two forms are complementary when the sequences are inverted (flip-flop) (Zhang *et al.*, 1994).

Figure 3. Comparison of the location of 22 genes in four chlorovirus genomes: between PBCV-1 and CVK2 (A), and PBCV-1 and MT325 (Pbi-virus) (B). Corresponding gene positions are connected by lines. PBCV-1 genes used as landmarks are (1) A3R, (2) A35L, (3) A50L, (4) A100R, (5) A107L, (6) A125L, (7) A181R, (8) A185R, (9) A245R, (10) A292L, (11) tRNA genes, (12) A357L, (13) A413L, (14) A430L, (15) A448L, (16) A464R, (17) A533R, (18) A575L, (19) A577L, (20) A604L, (21) A646L, and (22) A666L. Some PBCV-1 genes are duplicated or missing in the other virus genomes. Although a few minor rearrangements occur (ex. a large insertion between (2) and (3) in CVK2), in general, colinearity exists among the NC64A viruses (A). In contrast, there is almost no co-linearity

between the genomes of PBCV-1 and MT325 (B). Unpublished sequence data for NY-2A and MT325 (J. L. Van Etten and M. V. Graves) are used in the comparison.

Figure 4. A diagram illustrating the genetic differences between HA and/or chitin synthesizing chloroviruses (Ali *et al.*, 2005). Line 1 shows the location of three PBCV-1 genes involved in the synthesis of HA. Viruses CVIK1 and CVHA1 that synthesize both HA and chitin (Line 2) have a chitin synthesizing gene inserted in the position of the PBCV-1 A330R ORF. Viruses that only synthesize chitin have two *chs* genes. These viruses may still have a *ugdh* gene (viruses CVK2, NY-2A, and AR158 in Line 3) or lose it (virus CVSA1 in Line 4). It is equally likely that the events depicted could occur in reverse order, i.e. from Line 4 to Line 1.

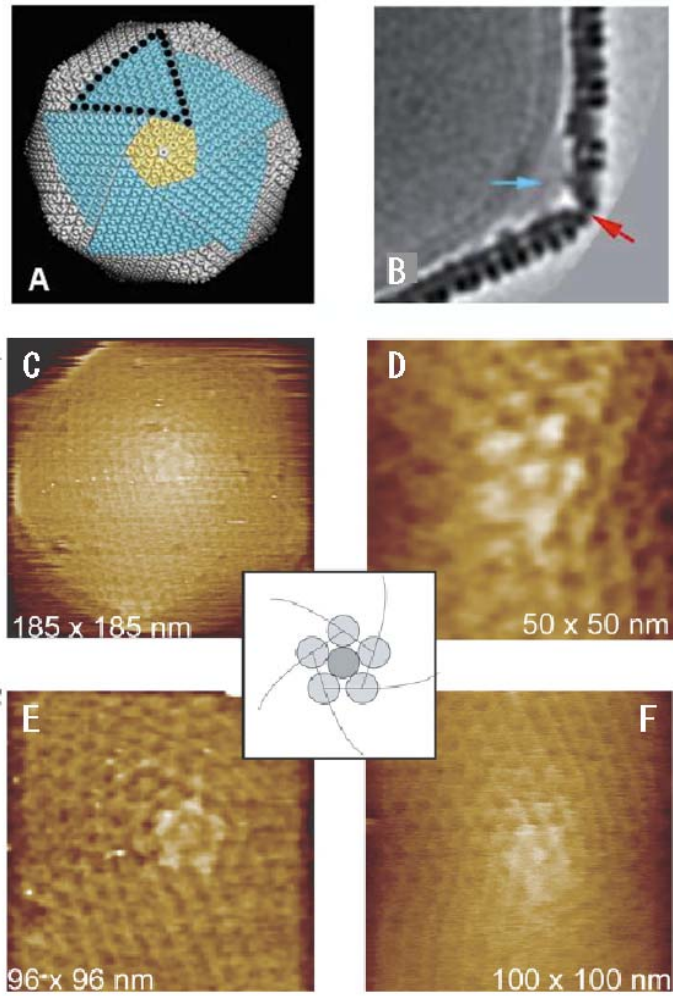


Fig. 1.

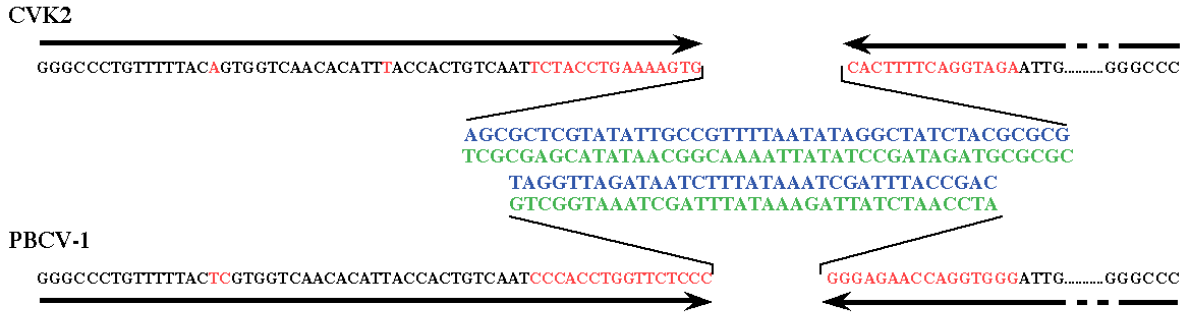


Fig. 2.

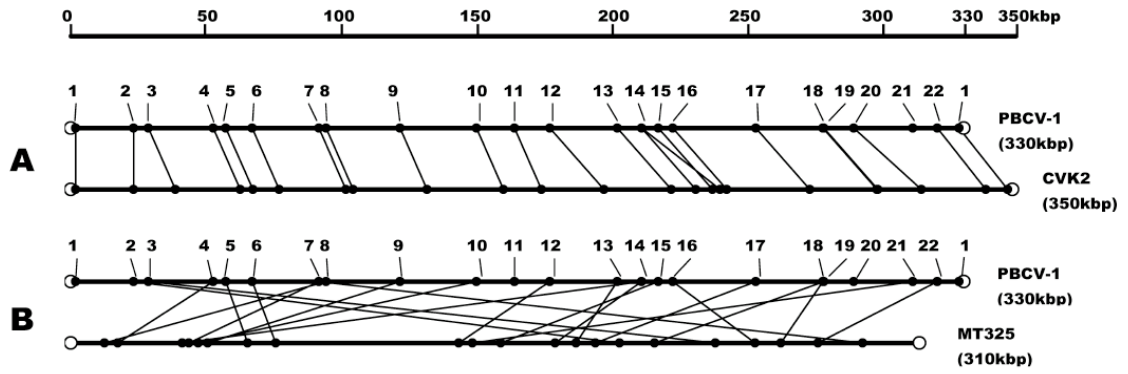


Fig. 3.

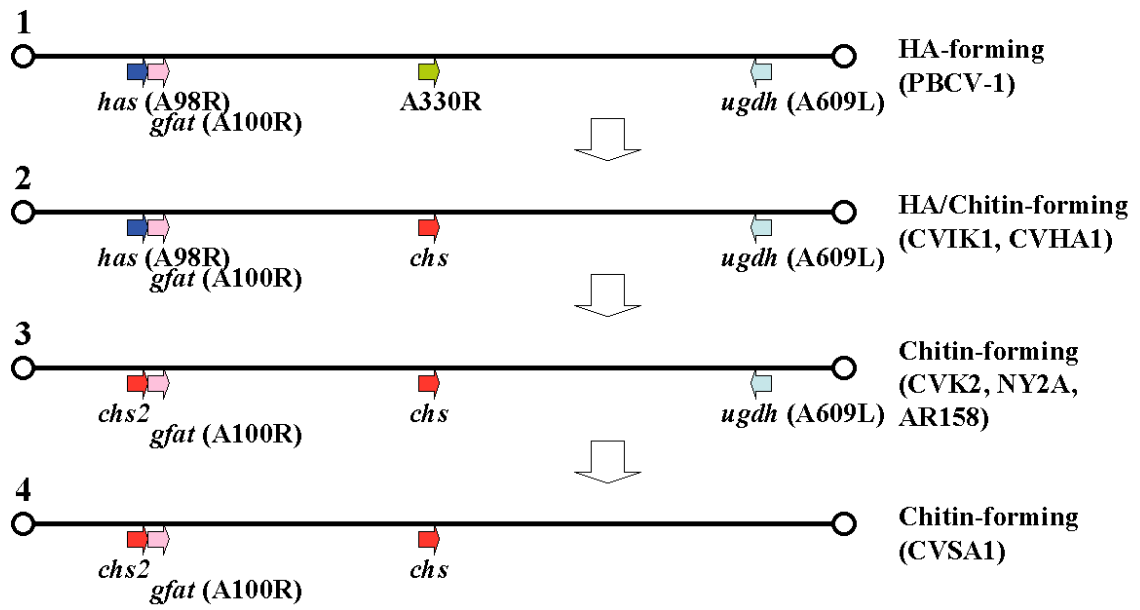


Fig. 4.

Table 1. Comparison of the Genome Organization among Several *Chlorella* Viruses.

Viruses	Genome Size (bp)	G+C Content (%)	Size of Terminal Inverted Repeats (kbp)	No. of Protein Encoding Genes	No. of tRNA Encoding Genes
PBCV-1	330,744	39.6	2.2	367	11
NY-2A	368,683 ^a	40.3	2.4	~420	7
A158R	344,690 ^a	40.4	0.27 ^b	~400	6
MT325	314,335 ^a	44.9	ND	~330	10
CVK2	~350,000 ^c	ND	2.4	ND	14

^a Unpublished results can be seen at website <http://greengene.uml.edu>.

^b Tentative size to be verified.

^c Tentative size based on **pulsed-field gel electrophoresis** (PFGE)-estimation.

ND, Not determined.

Table 2. Representative ORFs encoded by chlorella virus PBCV-1^a

DNA Replication, Recombination and Repair		Signaling, Ion channels and Protein Kinases	
DNA polymerase (δ)	A185R (913)	Potassium ion channel protein	A250R (94)*
DNA primase	A468R (443)	Ligand-gated channel protein	A162L (411)
ATP-dependent DNA ligase	A544R (298)*		A163R (433)
DNA topoisomerase II	A583L (1061)*	Ser/Thr protein kinase	A34R (308)*
PCNA	A193L (262)		A248R (308)*
	A574L (264)		A277L (303)*
RNase H	A399R (194)		A278L (610)*
Replication factor C (Archae large subunit)			A282L (569)*
	A417L (429)		A289 L (283)*
Helicase (superfamily III)	A456L (654)		A614L (577)*
Exonuclease	A166R (268)*	Tyr-protein kinase	A617R (321)
Pyrimidine dimer-specific glycosylase	A50L (141)*	Tyr-protein phosphatase	A305L (204)
DNA packaging ATPase	A392R (258)		
DNA Methyltransferase and Site-specific Endonucleases		Nucleotide Metabolism	
Adenine DNA methylase	A251R (326)*	Aspartate transcarbamylase	A169R (323)*
	A581R (265)*	Ribonucleotide reductase large subunit	A629R (771)
Cytosine DNA methylase	A517L (344)*	Ribonucleotide reductase small subunit	A476 (324)
	A530R (335)*	Thioredoxin	A427L (119)
	A683L (367)	Glutaredoxin	A438 (78)*
DNA site-specific endonuclease	A252R (342)*	dUTP pyrophosphatase	A551L (141)*
	A579L (183)*	dCMP deaminase	A596R (142)*
Insertion and Transposition		Thymidylate kinase	A416R (188)
Transposase	A625R (433)	Nucleotide triphosphatase	A326L (209)
Homing endonuclease (GIY-YIG)	A163L (165)	Cytidine deaminase	A200R (118)
	A287R (251)	Thymidylate synthase X	A674R (216)*
	A315L (246)	ATPase	A392R (258)
	A351L (358)	ATPase	A554L (271)
	A495R (221)		
	A539R (173)	Sugar Manipulation Enzymes	
	A651L (230)	Glucose dehydrogenase	A609L (389)*
Homing endonuclease (HNH)	A87R (456)	Glucosaminasesynthase	A100R (595)*
	A267L (314)	Hyaluronan synthase	A98R (568) *
	A422R (342)	Fucose synthase	A295L (317)*
	A478L (310)	GDP-D-mannose dehydratase	A118R (345)*
	A490L (310)	Mannosyltransferase	A64R (638)
		Fucosyltransferase	A114R (485)
Transcription		Glycosyltransferase	A111R (389)
Transcription factor IIB	A107L (290)		A222/226R (432)
Transcription factor IID	A552R (270)		A328L (355)
Transcription factor IIS	A125L (180)		A473L (517)
VLTF2-type transcription factor	A482R (215)		A546L (321)
RNA triphosphatase	A449R (193)*		
RNA guanylyltransferase	A103R (330)*		
Helicase (superfamily II)	A153R (459)		
	A363R (811)	Cell Wall Digestion	
SWI/SNF helicase	A548L (458)	Chitinase	A181/182R (830)*
SkiI helicase	A241R (725)		A260R (484)*
RNase III	A464R (245)*	Chitosanase	A292L (328)*
Histone H3, Lys 27 dimethylase	A612L *	β -1,3-glucanase	A94L (364)*
		Polysaccharide lyase	A215L (321)
Protein Synthesis, Modification, and Degradation			
Translation elongation factor III	A666L (918)	Lipid Manipulation Enzymes	
Prolyl-4-hydroxylase	A85R (242)*	Glycerophosphoryl diesterase	A49L (219)
Protein disulfide isomerase	A448L (106)	2-Hydroxyacid dehydrogenase	A53R (363)
Thiol oxidoreductase	A465R (118)	Lysophospholipase	A271L (159)
Ubiquitin C-terminal hydrolyase	A105L (284)	N-acetyltransferase	A654L (197)
SKP-1 protein	A39L (151)		
Zn metalloproteinase	A604L (134)		

^a ORFs for structural proteins including the major capsid protein Vp54 (A430L, 437 aa) and its related proteins are not listed. The number in brackets refers to the number of codons in the ORF. An asterisk means that the recombinant protein has been produced and shown to have the expected activity.