

ARTICLE

Received 20 Dec 2013 | Accepted 3 Jun 2014 | Published 2 Jul 2014

DOI: 10.1038/ncomms5278

OPEN

# Protruding knob-like proteins violate local symmetries in an icosahedral marine virus

Preeti Gipson<sup>1</sup>, Matthew L. Baker<sup>1</sup>, Desislava Raytcheva<sup>2,3</sup>, Cameron Haase-Pettingell<sup>3</sup>, Jacqueline Piret<sup>2</sup>, Jonathan A. King<sup>3</sup> & Wah Chiu<sup>1</sup>

Marine viruses play crucial roles in shaping the dynamics of oceanic microbial communities and in the carbon cycle on Earth. Here we report a 4.7-Å structure of a cyanobacterial virus, Syn5, by electron cryo-microscopy and modelling. A C $\alpha$  backbone trace of the major capsid protein (gp39) reveals a classic phage protein fold. In addition, two knob-like proteins protruding from the capsid surface are also observed. Using bioinformatics and structure analysis tools, these proteins are identified to correspond to gp55 and gp58 (each with two copies per asymmetric unit). The non 1:1 stoichiometric distribution of gp55/58 to gp39 breaks all expected local symmetries and leads to non-quasi-equivalence of the capsid subunits, suggesting a role in capsid stabilization. Such a structural arrangement has not yet been observed in any known virus structures.

<sup>1</sup>Verna and Marris McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>2</sup>Department of Microbiology, Northeastern University, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Correspondence and requests for materials should be addressed to W.C. (email: wah@bcm.edu).

Marine viruses are the most abundant and diverse life forms in the oceans. They constitute >90% of the nucleic acid containing material in the oceans<sup>1</sup>. It has been estimated that, based on their population ( $\sim 10^{30}$ ), if they were stretched out end to end, they could span sixty galaxies<sup>1</sup>. Only in the past decade have we started understanding the complexity of oceanic microbial ecosystems and their impact on global ecosystems<sup>2</sup>. Marine viruses are major biomass contributors to bio-geochemical cycles on earth, being responsible for 20% of the biomass cycled in the oceans everyday<sup>1</sup>. *Synechococcus* and *Prochlorococcus* are the most abundant cyanobacteria in the oceans, fixing  $\sim 30\%$  of CO<sub>2</sub> of the atmosphere through photosynthesis. The cyanophages, or phages infecting cyanobacteria, are key players in host genetic diversity and microbial community variability<sup>2</sup>. Their modes of infection and horizontal gene transfer introduce population selection pressure, which drives host–virus co-evolution<sup>3</sup>. Also, lateral gene transfer<sup>4</sup> during evolution is probably responsible for the strong phylogenetic similarity found between the cyanophages and the phages of enteric bacteria<sup>5</sup>. Not surprisingly, cyanophages are efficient reservoirs of both genetic diversity<sup>1</sup> and novel genes<sup>6</sup>.

Despite their importance, studies of marine viruses/phages are both recent and limited. This is especially true in terms of understanding their capsid structure and function, limiting our understanding of their efficiency as infection agents. Capsid subunits have to be capable of assembling into a closed icosahedral procapsid to package double-stranded (ds)DNA, and then transform to the mature capsid lattice stable enough to contain and protect the highly compressed genome. To date, only the mature capsids of cyanophage P-SSP7, infecting *Prochlorococcus*, have been structurally determined at near atomic resolution<sup>7</sup>.

Here we present the near atomic resolution structure of cyanophage Syn5, which infects *Synechococcus*, the dominant cyanobacteria in both the rich coastal and oligotrophic waters of the ocean. Syn5 is a dsDNA virus belonging to the *Podoviridae* family with a T7 bacteriophage-like genome organization. In an earlier study on the genomic characterization of Syn5 (ref. 6), a low-resolution electron cryo-microscopy (cryo-EM) analysis reported ‘knob’-like features in the icosahedral capsid, along with a short tail and unique horn-like structure. The knob-like proteins display a unique structural arrangement in the mature capsid, but are absent in the immature virion structure, also reported here. We show here that these knob-like proteins break all local symmetry in an overall icosahedral capsid shell of the mature virion. Our structural and bioinformatic analyses assign two candidate gene products to the knob-like densities. Together, the structures provide significant insight into the assembly and maturation of marine viruses.

## Results

**Structure of the mature virion.** The mature Syn5 cyanophage was imaged using a JEM3200FSC electron cryo-microscope (300 keV) at liquid nitrogen temperature, images were recorded on a Gatan 10K  $\times$  10K CCD (charge-coupled device) camera. Figure 1a shows a typical image of Syn5. The power spectrum of Syn5 particles in an individual CCD frame<sup>8</sup> is shown in Supplementary Fig. 1a, indicating visible signal beyond 5 Å resolution. An *ab initio* featureless initial model (Supplementary Fig. 1b) was generated using a small set ( $\sim 1,000$ ) of particles by Fourier cross-common lines principle<sup>9</sup> implemented in multi-path simulated annealing three-dimensional (3D) reconstruction routine<sup>10</sup>. A final icosahedral reconstruction was obtained from  $\sim 12,000$  individual particle images (Fig. 1b). The resolution of

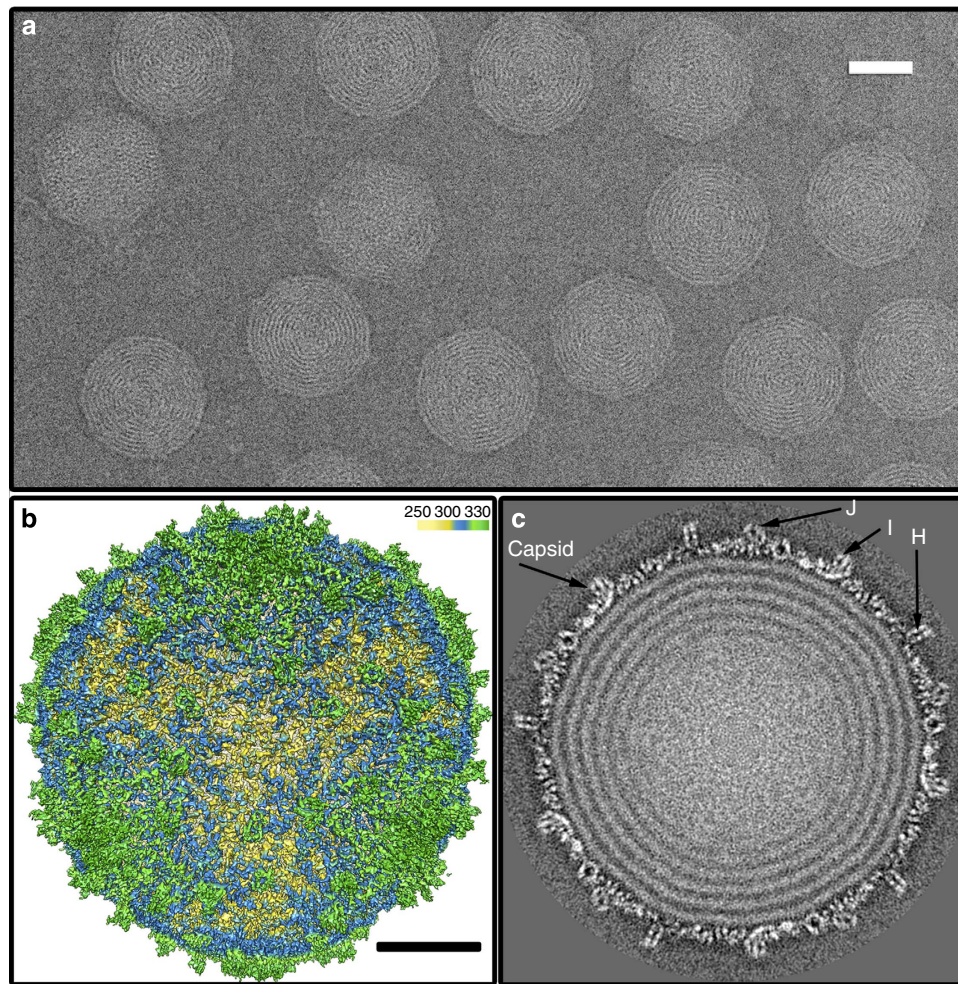
the map was estimated and validated by using the high-resolution (HR) noise substitution method<sup>11</sup>. A Fourier shell correlation (FSC<sub>true</sub>) was calculated as described previously<sup>11</sup> estimating the resolution of the map to be 4.7 Å at 0.143 FSC cut-off (Supplementary Fig. 1c).

A characteristic feature of the map is the presence of 60 copies of hexameric capsomeres and 12 copies of pentameric capsomeres ( $T=7$ ). One striking feature of the hexameric capsomere, which is different from any known bacteriophage structure, is the presence of protruding densities (Fig. 1b) referred hereby as ‘knob-like proteins’<sup>6</sup>. Figure 1c shows a slice view of the map with three such knob-like densities protruding at different heights from the capsid wall (labelled as H, I and J).

**Major capsid protein (gp39) of the mature virion.** At the reported resolution, the capsid density map clearly reveals the secondary structural elements (SSEs) of the protein subunits, such as long  $\alpha$ -helices and large  $\beta$ -sheets<sup>12,13</sup>. On the basis of location, the presence of SSEs and the expected structural similarity to other known bacteriophage major capsid proteins (Supplementary Fig. 2a,b), such as HK97 (gp5)<sup>14</sup>,  $\epsilon 15$  (gp7)<sup>15</sup> and T7 (gp10A)<sup>16</sup>, we segmented, averaged and constructed *de novo* C $\alpha$  backbone models for each gp39 subunit in the asymmetric unit using Gorgon<sup>17</sup>. Figure 2a shows a model of one gp39 subunit superimposed on the density map; the major domains—A, P, E-loop and N-arm domains—are clearly evident, while model of one asymmetric unit with seven gp39 subunits (Chain A–G) is seen in Fig. 2b. To validate the model, an analysis of the uniqueness of the solution obtained for the C $\alpha$  trace was carried out using an independent *de novo* modelling tool, Pathwalker (discussed in Methods section).

The major capsid protein of Syn5 (gp39) shows only  $\sim 16\%$  sequence identity when compared with the major capsid proteins of HK97 (gp5)<sup>14</sup> and  $\epsilon 15$  (gp7)<sup>15</sup>, whereas a higher sequence identity of  $\sim 44$  and  $\sim 26\%$  is observed with the coat proteins of P-SSP7 (gp10)<sup>7</sup> and T7 phage (gp10)<sup>16</sup>, respectively. In terms of structural domain arrangement, gp39 (332 aa) is topologically most similar to gp10A (345 aa), a coat protein of T7, and gp5 (282 aa), a coat protein of HK97. A C $\alpha$  root mean squared deviation of  $\sim 2.3$  Å is obtained from a pairwise topology comparison between gp39 and gp10A or gp5 with an overall  $\sim 115$  matched residues in each case<sup>18</sup>. A couple of significant differences are found in the A-domain region of the above proteins. In Syn5, the coat protein gp39 shows the presence of an ‘extra’ loop ( $\sim 30$  aa, coloured yellow, Fig. 2a and Supplementary Fig. 3a) when compared with gp5 in HK97. This loop region of gp39 subunits (chains C and D) are seen bound to protruding knob-like proteins (green densities, Fig. 3a). Second, in gp39 the loop region ( $\sim 26$  aa) forming the opening at the local six-fold axis of the hexamer (Fig. 2b) is wider and orthogonal to that observed in case of gp5 (HK97), where this loop is elevated straight towards the centre of the hexamer (Supplementary Fig. 3a). A similar difference as above is observed on comparison of the hexameric gp39 subunits (chains A–F) with the pentameric gp39 subunit (chain G), where the A-domain loop of the hexameric subunits around the opening at the six-fold axis is tilted by a  $\sim 90^\circ$  angle to that of the pentameric gp39 subunits lying around the five-fold axis (Supplementary Fig. 3b).

A pairwise FSC analysis between the seven gp39 subunits of an asymmetric unit shows a higher correlation at lower frequencies between four hexameric subunits (chains B:E and C:F; green curves, Supplementary Fig. 3c). These two FSC curves (green) show a higher than average FSC curve (solid black) when compared with the other subunit in a pairwise comparison (blue and red curves). These four gp39 subunits (chains B, C, E and F)



**Figure 1 | Cryo-EM map of Syn5.** (a) Syn5 virion particles observed in vitreous ice in various orientations. Scale bar, 300 Å. (b) Icosahedral map of Syn5 at 4.7 Å coloured radially, shows the arrangement of pentamers (green), hexamers (blue) and protruding densities (green) forming the capsid structure. Scale bar, 150 Å. (c) A 2D slice of the map showing the capsid wall and the protruding densities labelled as H, I and J.

are seen bound to the knob-like proteins, discussed later. Overall, their structural similarity is measured to be  $\sim 6.5$  Å as measured by the  $FSC = 0.33$ . The primary structure differences among the gp39 subunits lie in the A-domain and E-loop regions (red oval, Supplementary Fig. 3b). For instance, the E-loop of the pentameric gp39 subunit is tilted by  $\sim 45^\circ$  in comparison to the hexameric gp39 subunits, showing the poorest correlation.

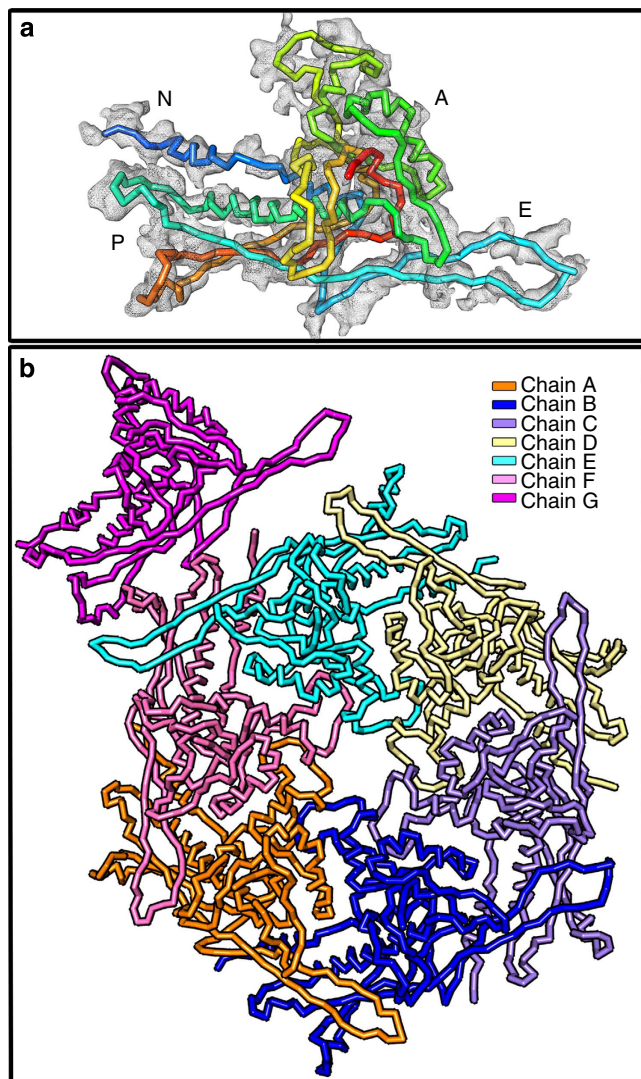
**Protruding knob-like capsid proteins identified.** The mature capsid of Syn5 contains several knob-like major densities protruding from the capsid surface (Fig. 3a). Here the knob-like proteins are labelled as I/H/J based on their positioning along a diagonal across the hexameric capsomere (Fig. 3b). The density H is located at the centre of the hexameric capsomere. Both I and J are present at the two opposite ends of the diagonal such that protein I always faces the pentameric vertex, while J faces the neighbouring hexameric capsomere. As seen in Fig. 3b,c, these knob-like proteins follow the strict icosahedral two/three-fold symmetries as expected from an icosahedral reconstruction.

These additional protruding densities were segmented after fitting the gp39 models to the density map. The segmented H density has a clip-like dimeric structure, labelled as H1/H2 in Fig. 4a–c. It shows elongated rod-like densities at the base near the six-fold (capsid-binding domain), while the top part flares

outwards from the capsid surface (protruding domain) (Fig. 4b). Automatic segmentation of H using Segger19 (ref. 19) and a rotational symmetry analysis revealed that it is a dimer, with two monomeric subunits related by a two-fold symmetry normal to the capsid surface (Fig. 4c and Supplementary Fig. 4a). The capsid-binding domain of H1 and H2 further extends into densities running parallel to the capsid surface in four directions (grey densities, Fig. 4a). The densities I and J appear to be anchored at the opposing ends of the diagonal formed by these elongating densities.

The segmented densities for I and J appear globular, exhibiting similar size and shape (Fig. 4d,e). Superposition of the I and J, and a difference map analysis revealed only minor structural differences; a structural similarity of 7 Å was observed between the segmented densities of I and J from the FSC analysis (Supplementary Fig. 4b). From the above analyses, we conclude that the densities observed at the I/J positions along each hexameric capsomere of the map are the same, which in turn suggests that they are made of the same protein. Both proteins I and J show three equivalent attachment sites (labelled with a circle, square and triangle in magenta, Fig. 4d,e) to two gp39 subunits lying at the opposite ends of the diagonal (that is, chains E/F and B/C, respectively). Density I is attached at two sites of the same gp39 subunit (chain E), namely the loop region immediately after the long helix (circle) and at the end of the E-loop (square)





**Figure 2 | Major capsid protein gp39.** (a) *De novo* backbone model of gp39 fit in the corresponding map density (grey) shows an HK-97-like fold; the residues are coloured blue to red from the N- to the C-termini. (b) Backbone model of the seven chains A-G of gp39 in one asymmetric unit forming the  $T=7$  icosahedral capsid.

(Fig. 4d). While at the other end, protein I extends further, slightly elevated, attaching to the protruding loop of the A-domain of the neighbouring gp39 subunit (triangle, chain F). Similarly, three equivalent attachment sites are observed for diagonally opposite protein J at two corresponding gp39 subunits (chain B/C) (Fig. 4e). This suggests that each of the I and J subunits spans across two adjacent gp39 subunits within a capsomere to stabilize the hexon.

**Gene product assignments to the knob-like proteins.** While assignment of the gp39 to the map density was straightforward because of expected phage capsid fold, the determination of corresponding gene products for I, J and H (H1/H2) was more challenging. Three late gene products, gp55 (156 aa), gp57 (131 aa) and gp58 (169 aa), were potential candidates<sup>20</sup> for the above densities. We performed several computational analyses on these candidates, both on their sequences and the map densities for I, J and H (H1/H2), including secondary structure prediction<sup>21,22</sup>, protein stability, amino-acid composition<sup>23,24</sup> and density-based secondary structure analysis with SSEHunter<sup>25</sup>. Sequence

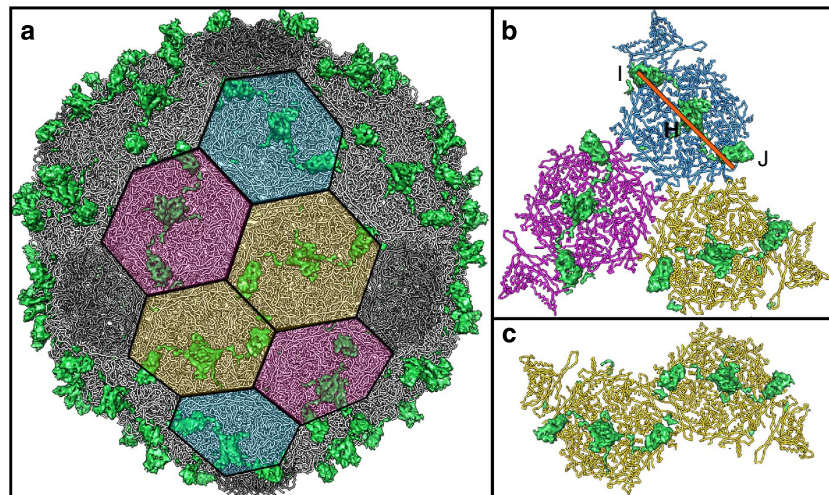
analyses predict<sup>23,24</sup> gp55 and gp58 to be stable proteins with consensus secondary structure predictions<sup>21,22</sup>, while gp57 is predicted to be an unstable protein with no consensus secondary structure prediction.

Secondary structure element analysis with SSEHunter of densities I/J identified major  $\beta$ -sheet regions (blue sheets, Fig. 5a), while density H1 or H2 showed two major helices in its capsid-binding domain (green cylinder, Fig. 5b). The secondary structure prediction of gp55 (156 aa) revealed mostly  $\beta$ -strands and loops, while the gp58 showed three major helices (at N-terminus) along with strands/loops (Fig. 5a,b). On the basis of converging results from the above density and sequence-based structure predictions, a correspondence was established between gp55 and I/J densities, as well as gp58 and the H1/H2 density. Also, the density and sequence analysis together hint that gp58 (~169 aa) forms a dimer consisting of two polypeptide chains. Hence, we conclude that each hexameric capsomere of Syn5 has two copies of gp55 at respective I/J positions and two copies of dimeric gp58 at positions H1 and H2. Here we were able to locate the SSEs such as helices/sheets (Fig. 5a,b); however, we were not able to build a model due to insufficient resolvability of these protruding regions and lack of homologous structures in the PDB for gp55/58.

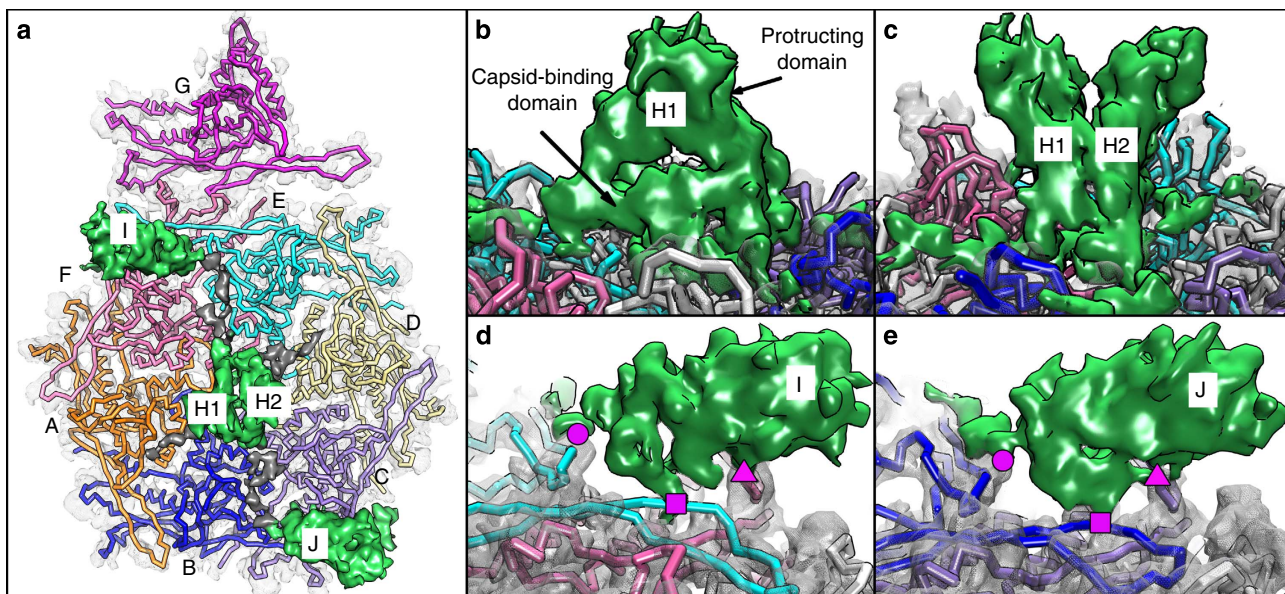
A BlastP sequence analysis<sup>26</sup> of gp55 returns TonB-dependent receptors as top hits with 28% sequence identity. A multiple sequence alignment between gp55 and the top Blast hit result showed similarities with the region of TonB receptor belonging to the porin superfamily (aa 211–385) (Supplementary Fig. 5). The TonB receptors play a role in sensing and signalling in the outer membrane of the Gram-negative bacteria and share a  $\beta$ -barrel-like structure<sup>27,28</sup>. The host of cyanophage Syn5 is also a Gram-negative cyanobacteria *Synechococcus*. Both gp55 sequence secondary structure prediction and density analysis hint towards a mostly  $\beta$ -stranded structure of gp55 (Fig. 5a), which might explain the observed sequence similarities with the TonB-dependent receptors (mostly  $\beta$ -stranded). Also, it is known that viruses can mimic both ligands and cell surface receptors of host cells, also known as the molecular mimicry mechanism<sup>29</sup>. Such a mechanism is used to parasitize the host cell surface receptors to hijack and affect certain cellular processes. It is possible that gp55 plays a role in weak host-cell surface recognition or increases the host-cell nutrient intake in a nutrient-deficient environment by mimicking the siderophore/TonB-dependent cell surface receptors and hence, increasing the efficiency of virus infection<sup>29</sup>.

Sequence analysis of gp58 (169 aa) revealed 25% sequence identity with the Hoc protein<sup>30</sup> from T4. However, most of the observed sequence identity is randomly distributed over the four domains of the Hoc protein (400 aa). Both gp58 and Hoc proteins are observed at the six-fold opening of the hexamers in Syn5 and T4 capsids, respectively. The Hoc proteins exist as monomers, consisting of three of the four domains with antigenic Ig-like structure<sup>31</sup>, while gp58 is present as a dimer with no predicted Ig-like domains. From the sequence analysis of gp58 the N-terminus region is predicted to have major helices (16–18 residues long). In our map we also observe two ~30-Å long rod-like helical densities (Fig. 5b) at the capsid-binding domain of each monomer of gp58, anchored at the six-fold depression of the capsid surface. This would suggest that the N-terminus region of gp58 is most likely the capsid-binding domain, which in turn implies that the C-terminus (predicted to be mostly loops and strands) possibly forms the protruding domain.

**Symmetry breaks observed at all local interaction sites.** In the mature Syn5 virion, the major capsid protein gp39 has an icosahedral packing, but the presence of protruding knob-like



**Figure 3 | Protruding knob-like densities.** (a) An overall arrangement of the knob-like densities (green) protruding from the capsid surface (grey model) formed by the major capsid protein gp39, where the pentameric vertices are shown in dark grey. Two triangular (T) faces are annotated as blue, pink and yellow hexagons per T-face to show strict icosahedral three- and two-fold symmetry. (b) A view of the strict icosahedral three-fold showing the arrangement of the protruding densities (green). The three knob-like densities (green) are labelled such that I faces the vertex, while J faces the neighbouring hexamer and H lies at the centre of the hexamer. (c) A view of the strict icosahedral two-fold showing the relative arrangement of I/H/J.



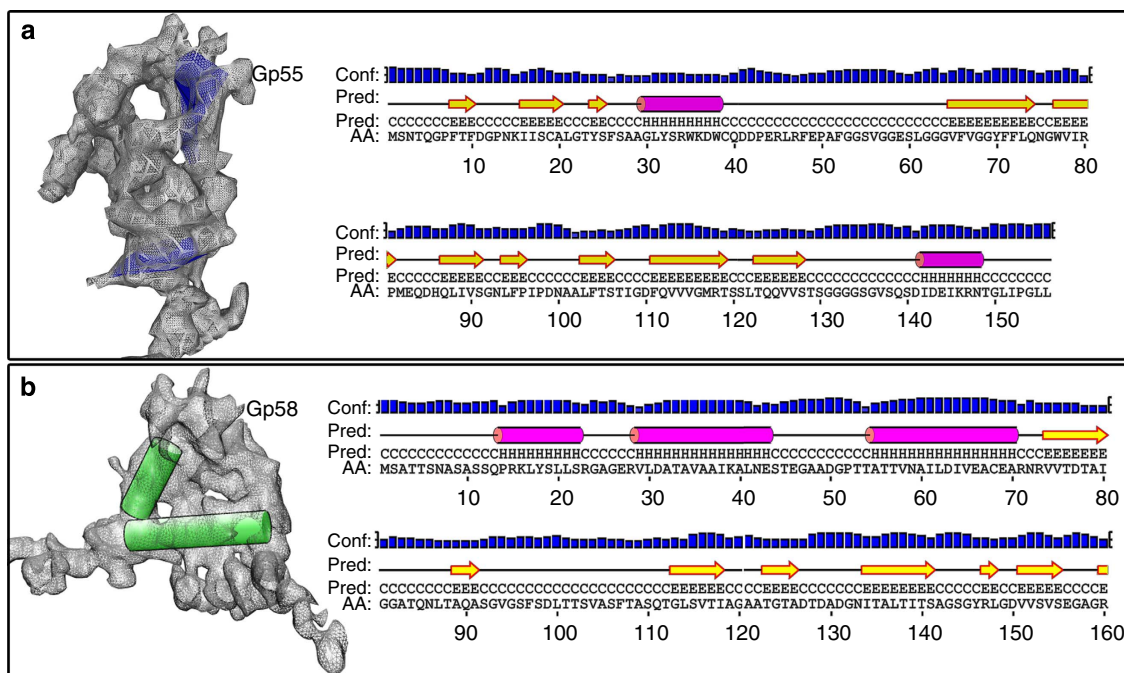
**Figure 4 | Binding sites of I/H/J.** (a) One asymmetric unit of Syn5 consisting of 11 polypeptide chains, labelled as A-J, where the two-polypeptide chains of H are labelled as H1/H2. The seven major capsid protein gp39 models (A-G) follow the colour scheme of Fig. 2b. Here the knob-like proteins (green) are seen connected by grey densities. (b) The binding site of the density H to gp39 (a side view such that only one monomer H1 is seen). (c) A close up of H in a 90° rotated view of (b) (a side view such that both monomers H1/H2 are seen). (d,e) A close up of the equivalent binding sites of densities I and J to gp39, annotated by magenta circle, square and rectangle symbols.

proteins gp55/58 introduces asymmetric local interfaces among the neighbouring capsomere gp39 subunits. Such a distribution of the knob-like proteins across the icosahedral capsid is not observed in other known phage/virus structures. Figure 6 and Supplementary Movie 1 show a range of all such interfaces observed at both the strict and local two/three-fold symmetry interactions between the capsomere subunits.

In Fig. 6a, the complete Syn5 capsid is presented in a two-dimensional (2D) lattice form, showing all the quasi-equivalent sites for a  $T=7$  capsid (red oval/triangle symbols for strict icosahedral two/three-fold, respectively, while yellow symbols for

local two/three-fold sites). In Fig. 6b is shown a close up of two neighbouring triangular faces where the icosahedral strict and local symmetry axes are labelled as above. Four types of two-fold interfaces are observed between the gp39 subunits of neighbouring capsomeres (Fig. 6c–f). Here in addition to the strict icosahedral two-fold symmetry interface (Fig. 6c), three additional local two-fold interfaces are present between the gp39 subunits of neighbouring hexameric and hexameric/pentameric capsomeres (Fig. 6d–f). However, these local two-fold symmetries are broken due to the unique diagonal positioning of gp55/58 (I/H/J positions) in the asymmetric unit.





**Figure 5 | SSEs in I/J and H1/H2.** The figure shows (a) gp55 (I/J) and (b) gp58 (H1/H2). Both (a,b) on the left show the density-based localization of the SSEs (helices as green cylinders and  $\beta$ -sheets as blue sheets) for I/J and H1/H2 densities, respectively. Two helices annotated by green cylinders are predicted in the capsid-binding domain of H. On the right is seen a sequence-based secondary structure prediction for gp55 and gp58, the corresponding gene candidates for densities I/J and H1/H2, respectively. Here 'Conf' refers to confidence in the secondary structure prediction, where the height of the histogram relative to a scale (on the left and right ends) represents the confidence of prediction on a scale of 0-1 (low = 0 and high = 1) and 'Pred' refers to the predicted secondary structure for a region.

Similarly, Fig. 6g-i shows the three types of three-fold interface observed between the gp39 subunits of neighbouring capsomeres. In Fig. 6f is shown the three-fold interface observed at the icosahedral strict three-fold axis. Two local three-fold interactions are present between the gp39 subunits of neighbouring hexameric and hexameric/pentameric capsomeres (Fig. 6h,i, respectively), but the local three-fold symmetry is again broken due to the gp55/58 binding.

## Discussion

Our structure of the mature virion of Syn5 presents for the first time a direct structural insight of a marine virus, Syn5, which infects the dominant cyanobacteria *Synechococcus* in the oceans. Surprisingly, in spite of being relatively primitive on an evolutionary scale, the structure of Syn5 reveals a unique and complex arrangement of capsid subunits not observed in other virus structures (Supplementary Fig. 6). Here each asymmetric unit has four more knob-like capsid subunits (two copies of gp55 and two copies of gp58), in addition to the regular seven major capsid subunits (gp39) in a  $T=7$  arrangement. Consequently, each asymmetric unit in Syn5 is made up of 11 polypeptide chains with a stoichiometric ratio of 7:2:2 for gp39:gp55:gp58. Such a non 1:1 distribution of gp55/58 to gp39 breaks all expected local symmetries in an overall icosahedral capsid shell. This in turn leads to non-quasi-equivalence of the capsid subunits, making the structural arrangement of Syn5 an exception to the theory of quasi-equivalence<sup>32</sup>. The studies of marine viruses are both recent and limited; here our structural analysis of Syn5 elucidates an understanding of their capsid structure and function.

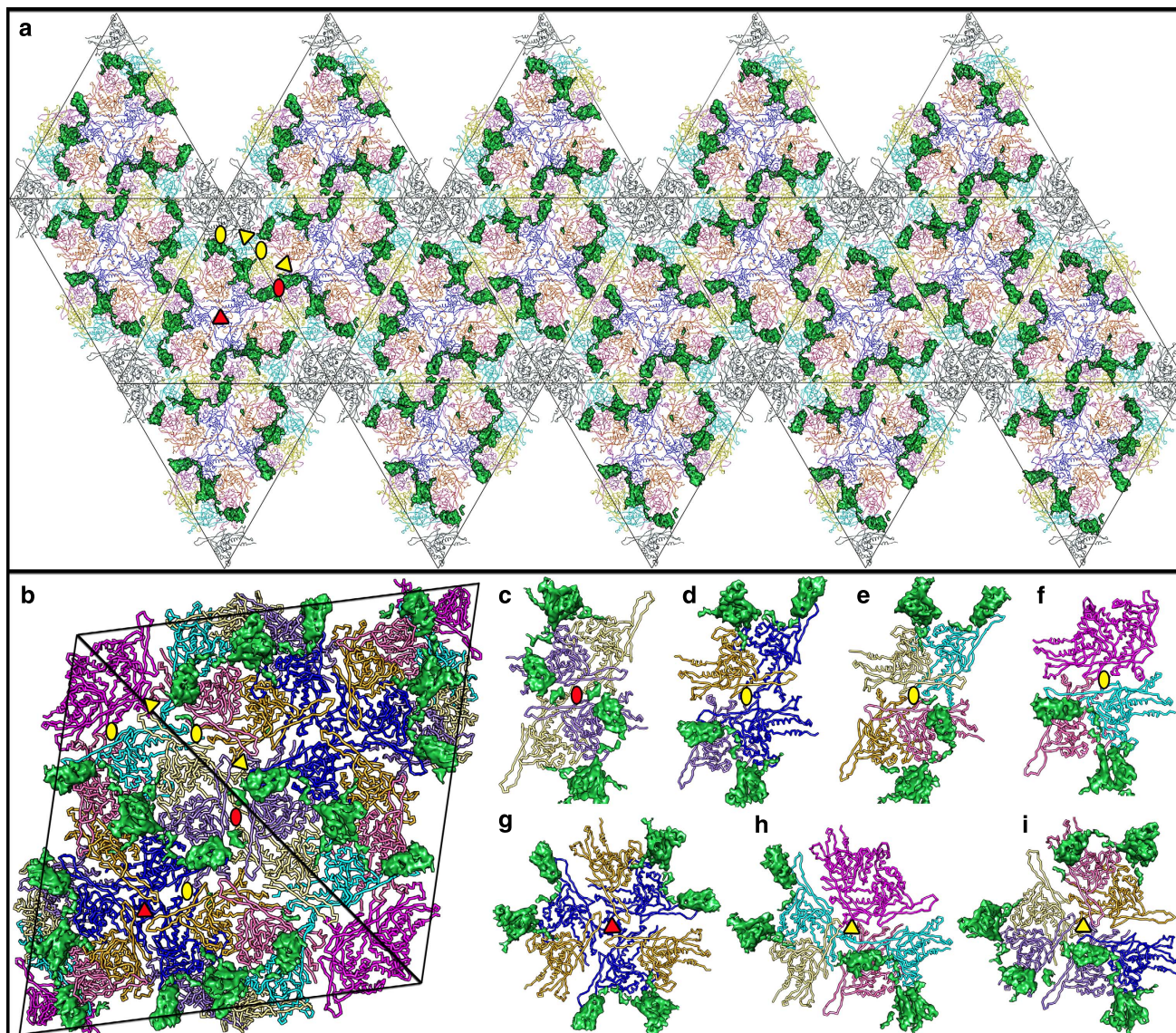
The mature capsid of dsDNA viruses need to be stable enough to resist the pressure for highly condensed genome<sup>33</sup>. In other phage/virus structures, the outer capsid proteins (also known as

decoration/stabilizing/stapling proteins) are usually found at the three- or two-fold regions (dotted lines, Supplementary Fig. 6), with the three-fold known as the weakest site for icosahedral capsids<sup>33,34</sup>. In HK97, covalent bonding stabilizes the three-fold sites, although many phages/viruses recruit decoration proteins to stabilize this region. Phages lambda, L and T4 stabilize the three-fold region by incorporating trimers (Supplementary Fig. 7) of stabilizing proteins, gpD<sup>35</sup>, Dec<sup>36</sup> and Soc<sup>37</sup>, respectively, while in adenovirus, minor capsid protein IX trimers are incorporated in this region<sup>38</sup>. In the case of  $\epsilon$ 15, the stapling protein gp10 is present as dimers, stabilizing the two-fold interactions between the neighbouring capsomeric subunits<sup>15</sup> (Supplementary Fig. 7). The presence of penton base-associated fibre trimers in adenovirus<sup>39,40</sup> cause a symmetry break at the five-fold; however, unlike in Syn5, the symmetry at the quasi-equivalent local two/three-fold sites is maintained.

While Syn5 contains a major capsid protein (gp39) similar to other bacteriophages, the two other knob-like outer capsid proteins (gp55/58) are novel proteins. Unlike the outer capsid proteins observed in viruses/phages mentioned above, these knob-like proteins (gp55/58) in Syn5 do not bind at the inter-capsomere interfaces located at the strict icosahedral or local two/three-fold symmetry axes (Supplementary Fig. 7). Instead, both gp55/58 are bound to the major capsid (gp39) subunits in unique diagonal positions within a hexameric capsomere presumably, stabilizing the intra-capsomere hexameric subunit interactions (Fig. 4d,e). Furthermore, none of the pentameric subunits has any of these associated proteins. Again, such a structural arrangement of capsid proteins has never been observed in any icosahedral virus structure.

An insight into the functional implications of the unique arrangement of outer capsid proteins observed in Syn5, is gained by a comparative analysis of the hexameric capsomeres, observed





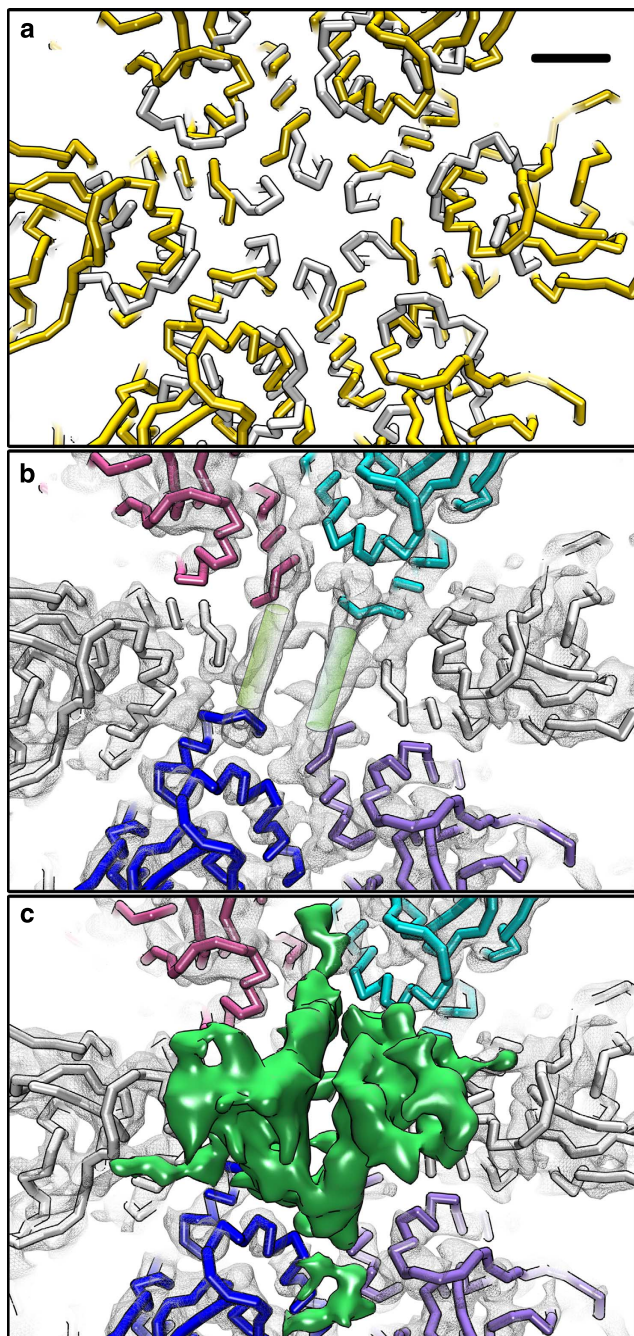
**Figure 6 | Symmetry breaks in Syn5 icosahedral lattice.** (a) A 2D representation of the  $T=7$  lattice in Syn5, showing all the 20 triangular faces with an arrangement of the knob-like densities (green); the hexameric gp39 models are coloured as per colour scheme of Fig. 2b, while the pentameric gp39 is coloured grey. The strict icosahedral three- and two-fold symmetry axes between one set of two triangular (T) faces are marked by red triangle and oval symbols, respectively, while the local three- and two-fold axes are marked by corresponding yellow symbols. (b) A close up of two T-faces from (a) is shown enclosed in a rhombus shape (solid black lines). (c-f) An array of all the two-fold interfaces observed in Syn5, where (c) shows the strict icosahedral two-fold interface (axis, red oval) and (d-f) show the local two-fold interfaces (axes, yellow oval). (g-i) An array of all the three-fold interfaces, with (g) showing the strict icosahedral three-fold interface (axis, red triangle) and (h,i) showing the local three-fold interfaces (axes, yellow triangle).

in known  $T=7$  virus structures. In Syn5, the opening at the six-fold, composed of six gp39 proteins, measures  $\sim 28\text{--}30\text{ \AA}$  in diameter, while the opening measures  $\sim 12\text{--}14\text{ \AA}$  in other phages such as HK97 (ref. 14),  $\epsilon 15$  (ref. 15), P22 (ref. 41) and P-SSP7 (ref. 7) (Fig. 7a and Supplementary Fig. 2b). This is due to a loop in the A-domain, which is orientated differently than the corresponding loop in other known phage structures (Supplementary Fig. 3a). Such a significantly wider opening at the six-fold in Syn5 would likely not provide the necessary protection of the viral genome. The positioning of the gp58 protein dimer (H1 and H2) atop the six-fold opening, together with its size relative to the six-fold axis opening, suggests that it is a plug that seals the wide opening, protecting the genome and enhancing capsid stability (Fig. 7b,c). Owing to the size and geometry of the pentameric opening, the gp58 dimer would not

be able to fit the dimensions. A similar arrangement has been observed in T4 phages, where the outer capsid protein, Hoc, does not bind to the mutant hexamer opening when it is made up of only five major capsid (gp23) subunits<sup>31</sup>.

Interestingly, two gp55s are always bound to the E-loop region of two gp39 subunits, which are also bound to the gp58 molecule at their A-domains. Such a specific binding explains the co-occurrence of two gp58 and two gp55 molecules always along one specific diagonal of a hexameric capsomere. This also hints that the incorporation of gp55 molecules is not solely guided by the curvature of the hexamer. Possibly the incorporation of gp58 dimer to seal the six-fold opening causes some domain movements, which in turn exposes binding sites for the incorporation of two gp55 molecules. This would mean that gp55 incorporation compensates for the conformational instability caused by gp58





**Figure 7 | Gp58 (H1/H2) incorporation.** (a) Difference between the opening at the centre of the hexamer in Syn5 (gold) and that seen in HK97 (white). Scale bar, 13 Å. (b) The same region of (a) in Syn5 map (slice view, grey mesh) with fitted Syn5 model (white). Here two helical densities (green cylinders) are observed connecting four gp39 subunits (coloured). (c) A non-slice view of (b) to illustrate that the helical densities in (b) correspond to the capsid-binding domain of gp58 (green) contributed by its H1 and H2 monomers.

binding. Such domain-level conformational changes induced by the binding of small proteins has been observed in other macromolecular complexes such as ribosomes, where the binding of ribosome modulation factor induces a conformational change in the 30S head domain of the 100S ribosome, exposing new interaction sites<sup>42</sup>.

Our cryo-EM analysis of the procapsids of Syn5 show the absence of protruding densities corresponding to gp55/58 in the

immature prohead particles, which instead have a thicker, less angular and smaller cage-like structure (Supplementary Fig. 8a,b). This hints at the incorporation of these outer capsid proteins in the later stage of maturation. The absence of protruding proteins in the procapsids may facilitate scaffolding protein release through the openings at the hexameric capsomeres<sup>41</sup>. It is known that the filling of DNA during the maturation process of the viruses can produce extreme pressures ( $\sim 60$  atm) causing capsid expansion, which in turn lead to structural rearrangements<sup>33</sup>. It is possible that such events in Syn5 lead to a wider opening at the six-fold axis of hexameric capsomeres, pushing the pentameric capsomeres upwards, as observed from the difference analysis between the procapsid and mature capsid maps (Supplementary Fig. 8c). As a result, gp58 may be added during maturation to seal the openings at the hexameric capsomeres and protect the viral genome. In turn, gp55 may also be added concurrently to help lock in the gp58 dimer, as discussed above. The expansion and angularization of the capsid may contribute to the availability of the binding sites along gp39 for both gp55 and gp58, explaining their incorporation along the same diagonal of the hexameric capsomeres.

As such, both gp55 and gp58 appear to play the role of stabilizing proteins in the mature capsid of Syn5. Also, the sequence analysis of gp55 hints that it might play a role in weak cell surface recognition or mimic host cell surface receptors. These cyanophage–host systems are found in harsh oligotrophic environments of the oceans<sup>43</sup>; such surface proteins might help in binding to non-host cells as well<sup>30</sup> to aid in travelling to their widely separated host cells.

Considering virus–host co-evolution<sup>44</sup>, cyanophages such as Syn5 are likely as ancient as their host cyanobacteria ( $\sim 2.8$  billion years), presenting an ancient lineage to the present day viruses. It is known that cyanophages such as Syn5 and P-SSP7 show synteny and homology to enteric phages<sup>45</sup>. Unlike Syn5, the marine virus P-SSP7 does not have accessory proteins to enhance capsid stability. However, some relatively more recent enteric phages and complex animal viruses have been reported to show the presence of capsid-stabilizing proteins. It appears that during the course of evolution, viruses diverged to adopt various efficient ways for capsid stabilization, such as covalent bonding and the incorporation of decoration/stabilizing proteins<sup>33</sup>. The observation of protruding capsid proteins in Syn5 hints that such genes were likely acquired very early on for roles such as capsid stabilization, weak host cell surface recognition and host cell surface receptor mimicking. It has been suggested that phage/viral genes can travel laterally by several recombination events across wide phylogenetic distances—with different genes in the same phage often having different ancestry<sup>46</sup>. The sequence identities observed between knob-like proteins of Syn5 and the equivalence in other enteric phages, as well as some bacterial proteins, hint towards such lateral gene recombination events.

The observation of capsid stabilizing proteins in Syn5 suggests the evolutionary significance of capsid stability/efficiency, where such genes were either acquired quite early on or more recently during virus evolution by means of lateral gene transfer. As the evolutionary age of marine viruses predates that of the enteric phages and animal viruses, it is possible that these structural features were acquired from the former during the course of evolution—although it may be a more recent phenomenon if these genes were acquired from the latter.

## Methods

**Electron cryo-microscopy.** A sample of mature Syn5 virions was isolated and purified as described<sup>5</sup>. Briefly, *Synechococcus* WH8109 was grown to mid-log (in artificial sea water under constant light at 28 °C) and infected with a multiplicity of infection = 0.001 phage per cell. On clearing, 1% CHCl<sub>3</sub>, 0.1% Triton X-100 and



0.01 mg ml<sup>-1</sup> of lysozyme were added to complete lysis. The lysate of cell debris was removed by centrifugation and filtration. The phage was precipitated with 0.5 M NaCl and 10% PEG (8 K) stirring overnight in the cold. The precipitated phage was collected by centrifugation and resuspended in 50 mM Tris pH 7.5, 100 mM NaCl and 100 mM MgCl<sub>2</sub>. The suspension was loaded onto a CsCl step density gradient, the phage particles sedimented to the interface between  $\rho$ 1.4 and  $\rho$ 1.5. The resulting phage was concentrated by Vivaspin MWCO 100K (Sartorius).

Aliquots of 2.7  $\mu$ l of the purified phage sample were applied to glow-discharged (Gatan Plasma Cleaner) 400 mesh Quantifoil R1.2/1.3 copper grids (hole size 1.2  $\mu$ m, Quantifoil Inc.), which were vitrified in liquid ethane by a FEI Vitrobot (MARK IV). Images of the frozen, hydrated sample were collected at a JEM3200FSC electron cryo-microscope (JEOL, Tokyo, Japan) operated at 300 keV at liquid nitrogen specimen temperature. The microscope is equipped with a field emission gun and an in-column omega energy filter (a slit width of 20 eV was used for data collection). The microscope settings include condenser aperture = 50  $\mu$ m, objective aperture = 120  $\mu$ m and spot size = 1. The images were recorded on a Gatan 10K  $\times$  10K CCD camera, where 1,000 CCD frames were recorded at a nominal magnification of 80,000 (0.66  $\text{\AA}$  per pixel sampling rate), with a defocus range of 0.7–3.0  $\mu$ m. The micrographs were computationally binned (2X) to obtain a final sampling of 1.32  $\text{\AA}$  in the images.

**Image processing and map validation.** Particles in various orientations were selected automatically using the swarm module in EMAN2 (ref. 47); the false-positive particles were deleted manually. This produced an initial data set of 18,000 particles. The contrast transfer function parameters for each CCD image were manually determined using ctfit in EMAN1 (ref. 8). An initial model was built from a small data set of 1,000 particles by assigning random orientations in multi-path simulated annealing<sup>10</sup>. The particle orientations were refined at an increasing resolution limit starting from 50  $\text{\AA}$  up to 10  $\text{\AA}$ . An iterative refinement was done until convergence to obtain the final map from ~12,000 particles. An FSC plot was obtained between the two maps generated from randomly split even/odd data sets. This FSC plot is called FSC<sub>data</sub>.

To validate the map resolution and assess any noise overfitting during refinement, the method of HR noise substitution was used<sup>11</sup>, here the results are shown in Supplementary Fig. 1c. For this, a second stack from the original experimental data set was generated, where data beyond 10  $\text{\AA}$  was removed by randomizing the phases<sup>11</sup>. These HR noise-substituted data were then subjected to the identical protocol of 3D reconstruction as mentioned above for the experimental data. An FSC plot was obtained between the two maps generated from the randomly split even/odd HR noise data sets. This FSC plot is called FSC<sub>noise</sub>. In the HR noise-substituted data, the FSC drops significantly to zero past 10  $\text{\AA}$ , beyond which the data were substituted with noise, showing no significant noise overfitting (shaded blue area). An FSC<sub>true</sub> (black solid line) was plotted by calculating the relative error between the FSC<sub>data</sub> (pink dotted curve) and FSC<sub>noise</sub> (blue dotted curve), as described previously<sup>10</sup>. The true data with no overfitting are shaded pink in Supplementary Fig. 1c. The FSC<sub>true</sub> plot was used to estimate the resolution of the final map to be 4.7  $\text{\AA}$  at FSC = 0.143. We applied experimentally determined structure factors<sup>47</sup> to the map for sharpening, limited to the reported resolution limit of 4.7  $\text{\AA}$ .

**Map visualization and analysis.** UCSF Chimera<sup>48</sup> was used for map visualization, analysis and generation of the molecular graphics images. The segmentation of the densities corresponding to the major capsid protein and the outer capsid proteins were done using Chimera and Avizo (<http://www.vsg3d.com/avizo/overview>). To generate an average of the six-hexameric subunits in one asymmetric unit for model building purposes, their corresponding densities were aligned in Foldhunter program<sup>49</sup>, while an average was calculated by proc3d in EMAN1. A pairwise FSC was calculated between the computationally segmented seven subunits in an asymmetric unit of the icosahedral map, where no symmetry is applied, to measure the correlation among the gp39 subunits within one asymmetric unit<sup>50</sup>.

**Sequence analysis and secondary structure prediction.** Various bioinformatic tools were used to analyse the sequences of gp39, gp55, gp57 and gp58 proteins. For the multiple sequence alignment and secondary structure prediction, PSIPRED<sup>21</sup> and Jpred<sup>22</sup> servers were used, while the physical and chemical parameters such as molecular weight, amino-acid composition, instability index, hydrophobicity and so on were calculated using ProtParam<sup>23</sup> and PredictProtein<sup>24</sup> servers.

The knob proteins gp55/58 being farthest from the centre (highest alignment errors) are poorly resolved compared to the major capsid proteins, hence we have not built model for these proteins. Moreover, the capsid surface of Syn5 is thin and smoother as seen in Fig. 1a compared with other known phage structures such as  $\epsilon$ 15 and P22, hence fewer features to align at the extreme radius of the capsid shell. However, we were able to localize major SSEs using SSEHunter<sup>25</sup> in the map densities of gp55/gp58. Also, our analysis hints that the protruding density gp58 found at the opening of the hexamer is composed of two polypeptide chains.

**Model building and refinement for gp39.** For model building, each of the seven individual gp39 subunits from one asymmetric subunit were cropped out of the full map using UCSF's Chimera<sup>48</sup>. Individual gp39s were aligned with Foldhunter<sup>49</sup>

and then averaged using proc3d, both of which are available in EMAN1 (ref. 8). Using the initial averaged gp39 density as a template, a second round of segmentation, alignment and averaging resulted in a final average gp39 subunit.

SSE identification was then performed on the averaged gp39 subunit using SSEHunter in Gorgon<sup>51</sup>. Five helices and two  $\beta$ -sheets were identified and corresponded to those found in capsid proteins of other tailed dsDNA bacteriophages, such as gp5 in HK97 (ref. 14). In addition, a density skeleton was computed that revealed the topological linkages between the observed SSEs. Jpred 3.0 (ref. 22) was then used to predict the secondary structure from the sequence, also revealing five helices and several beta strands.

Using Gorgon, an initial topology for gp39 was constructed by establishing a sequence to structure correspondence between the predicted and observed SSEs using the density skeleton as a constraint. From this topology, a C $\alpha$  backbone model was then constructed using Gorgon's semi-automated model building tools. Briefly, C $\alpha$  backbone  $\alpha$ -helices were first constructed in the density at the positions found by SSEHunter using the Helix editor function in the 'semi-automatic atom placement' utility in Gorgon. Loops between the  $\alpha$ -helices were then built using Atom editor and Position editor functions in the 'semi-automatic atom placement' utility in Gorgon, which allows the user to place individual C $\alpha$  backbone atoms along the density skeleton at a given spacing (~3.8  $\text{\AA}$  for C $\alpha$ -C $\alpha$  distances). Model building proceeded until the entire sequence of gp39 was placed within the density. Manual refinement of atom position was done interactively in Gorgon to remove any potential clashes and correct bad C $\alpha$ -C $\alpha$  distances. The final model was saved as a PDB file.

To validate the model, we then used our Pathwalking protocol<sup>17</sup> to determine whether the solution found in Gorgon was unique. The initial C $\alpha$  positions were iteratively perturbed ( $\sigma = 1$ ) using e2pathwalker.py such that 100 potential model paths were computed with Pathwalker. For calculating these paths, the LKH TSP<sup>17</sup> solver was used. Results were examined and compared in UCSF's Chimera<sup>48</sup>. A small amount of noise was added ( $\sigma = 1$ ) to the positions of the initial C $\alpha$  model using e2pathwalker.py from EMAN2. One hundred potential model paths were then computed using e2pathwalker.py and then compared in UCSF's Chimera. In each case, the pathwalking model resulted in a continuous chain trace through the density map without any visible density crossovers. Topologically, all the models appeared similar with some differences occurring in the first ~25 amino acids. For the purposes of the remaining modelling, the first 25 amino acids were truncated from the model. Manual refinement of C $\alpha$  positions was done interactively in Gorgon to correct bad C $\alpha$ -C $\alpha$  distances. In addition, COOT was used to remove clashes within and between subunits in the asymmetric unit. The final model was saved as a PDB file.

**Accession numbers.** A C $\alpha$  backbone model of the major capsid protein gp39 of the mature virion of Syn5 has been deposited in the RCSB Protein Data Bank under accession code 4BML. The original 3D cryo-EM density map has been deposited in the EMDDataBank under accession code EMD-5954.

## References

- Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
- Rohwer, F. & Thurber, R. V. Viruses manipulate the marine environment. *Nature* **459**, 207–212 (2009).
- Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474**, 604–608 (2011).
- Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
- Raycheva, D. A., Haase-Pettingell, C., Piret, J. M. & King, J. A. Intracellular assembly of cyanophage Syn5 proceeds through a scaffold-containing procapsid. *J. Virol.* **85**, 2406–2415 (2011).
- Pope, W. H. *et al.* Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a “horned” bacteriophage of marine synechococcus. *J. Mol. Biol.* **368**, 966–981 (2007).
- Liu, X. *et al.* Structural changes in a marine podovirus associated with release of its genome into *Prochlorococcus*. *Nat. Struct. Mol. Biol.* **17**, 830–836 (2010).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Crowther, R. A. Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Phil. Trans. Roy. Soc.* **261**, 221–230 (1971).
- Liu, X., Jiang, W., Jakana, J. & Chiu, W. Averaging tens to hundreds of icosahedral particle images to resolve protein secondary structure elements using a multi-path simulated annealing optimization algorithm. *J. Struct. Biol.* **160**, 11–27 (2007).
- Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
- Chiu, W., Baker, M. L., Jiang, W., Dougherty, M. & Schmid, M. F. Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* **13**, 363–372 (2005).

13. Baker, M. L., Baker, M. R., Hryc, C. F. & Dimaio, F. Analyses of subnanometer resolution cryo-EM density maps. *Methods Enzymol.* **483**, 1–29 (2010).
14. Wikoff, W. R. *et al.* Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* **289**, 2129–2133 (2000).
15. Baker, M. L. *et al.* Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proc. Natl Acad. Sci. USA* **110**, 12301–12306 (2013).
16. Agirrezabala, X. *et al.* Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J.* **24**, 3820–3829 (2005).
17. Baker, M. L., Baker, M. R., Hryc, C. F., Ju, T. & Chiu, W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* **97**, 655–668 (2012).
18. Nguyen, M. N., Tan, K. P. & Madhusudhan, M. S. CLICK—topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res.* **39**, W24–W28 (2011).
19. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
20. Raytcheva, D. A., Haase-Pettingell, C., Piret, J. & King, J. A. Two novel proteins of Cyanophage Syn5 compose its unusual horn structure. *J. Virol.* **88**, 2047–2055 (2014).
21. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
22. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).
23. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**, 531–552 (1999).
24. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Res.* **32**, W321–W326 (2004).
25. Baker, M. L., Ju, T. & Chiu, W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* **15**, 7–19 (2007).
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
27. Noinaj, N., Guillier, M., Barnard, T. J. & Buchanan, S. K. TonB-dependent transporters: regulation, structure, and function. *Annu. Rev. Microbiol.* **64**, 43–60 (2010).
28. Mirus, O., Strauss, S., Nicolaisen, K., von Haeseler, A. & Schleiff, E. TonB-dependent transporters and their occurrence in cyanobacteria. *BMC Biol.* **7**, 68 (2009).
29. Alcamí, A. Viral mimicry of cytokines, chemokines and their receptors. *Nat. Rev. Immunol.* **3**, 36–50 (2003).
30. Sathaliyawala, T. *et al.* Functional analysis of the highly antigenic outer capsid protein, Hoc, a virus decoration protein from T4-like bacteriophages. *Mol. Microbiol.* **77**, 444–455 (2010).
31. Fokine, A. *et al.* Structure of the three N-terminal immunoglobulin domains of the highly immunogenic outer capsid protein from a T4-like bacteriophage. *J. Virol.* **85**, 8141–8148 (2011).
32. Caspar, D. L. D. & Klug, A. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24 (1962).
33. Mateu, M. G. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys.* **531**, 65–79 (2012).
34. Zandi, R. & Reguera, D. Mechanical properties of viral capsids. *Phys. Rev. E* **72**, 021917 (2005).
35. Yang, F. *et al.* Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat. Struct. Biol.* **7**, 230–237 (2000).
36. Tang, L., Gilcrease, E. B., Casjens, S. R. & Johnson, J. E. Highly discriminatory binding of capsid-cementing proteins in bacteriophage L. *Structure* **14**, 837–845 (2006).
37. Iwasaki, K. *et al.* Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein. *Soc. Virol.* **271**, 321–333 (2000).
38. San Martín, C. Latest insights on adenovirus structure and assembly. *Viruses* **4**, 847–877 (2012).
39. Liu, H. *et al.* Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* **329**, 1038–1043 (2010).
40. Reddy, V. S., Natchiar, S. K., Stewart, P. L. & Nemerow, G. R. Crystal structure of human adenovirus at 3.5 Å resolution. *Science* **329**, 1071–1075 (2010).
41. Chen, D. H. *et al.* Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proc. Natl Acad. Sci. USA* **108**, 1355–1360 (2011).
42. Polikanov, Y. S., Blaha, G. M. & Steitz, T. A. How hibernation factors RMF, HPF, and YfiA turn off protein synthesis. *Science* **336**, 915 (2012).
43. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
44. Lindell, D. E. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).
45. Clokie, M. R., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31–45 (2011).
46. Hendrix, R. W. Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* **61**, 471–480 (2002).
47. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
48. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
49. Jiang, W., Baker, M. L., Ludtke, S. J. & Chiu, W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* **308**, 1033–1044 (2001).
50. Zhang, Q. *et al.* Cryo-EM structure of a molluscan hemocyanin suggests its allosteric mechanism. *Structure* **21**, 604–613 (2013).
51. Baker, M. L. *et al.* Modeling protein structure at near atomic resolutions with Gorgon. *J. Struct. Biol.* **174**, 360–373 (2011).

## Acknowledgements

This research has been supported by NIH grants (R56AI075208, R01GM079429 and P41GM103832) and Robert Welch Foundation (Q1242). P.G. is supported by NIH training grant 5T32AI055413. We thank Joanita Jakana for assistance with the data collection, Dr Xiangnan Liu for advice on image processing and Matt Dougherty for generating the movie animations. The authors acknowledge the Texas Advanced Computing Center (TACC, <http://www.tacc.utexas.edu>) at The University of Texas in Austin for providing HPC resources that have contributed to the research results reported within this paper.

## Author contributions

P.G., J.A.K. and W.C. designed the research; P.G. and M.L.B. performed the research; D.R., C.H.-P. and J.P. contributed new reagents/analytical tools; P.G., M.L.B. and W.C. analysed the data; P.G. and W.C. wrote the paper with inputs from other authors.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Gipson, P. *et al.* Protruding knob-like proteins violate local symmetries in an icosahedral marine virus. *Nat. Commun.* **5**:4278 doi: 10.1038/ncomms5278 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>