Crystal Structure of ORF12 from Lactococcus lactis phage p2 Identifies a Tape Measure Protein Chaperone

Marina Siponen¹, Giuliano Sciara¹, Manuela Villion²,⁴, Silvia Spinelli¹, Julie Lichière¹, Christian Cambillau¹, Sylvain Moineau²,³,⁴* and Valérie Campanacci¹*

¹Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I & II, Campus de Luminy, case 932, 13288 Marseille cedex 09, France;
²Groupe de Recherche en Écologie Buccale (GREB), Faculté de Médecine Dentaire;
³Félix d’Hérelle Reference Center for Bacterial Viruses;
⁴Département de Biochimie et de Microbiologie, Faculté des Sciences et de Génie, Université Laval, Québec City, Québec, Canada, G1K 7P4.

*Correspondence to VC: Fax; +33 491-266-720; E-mail: valerie.campanacci@afmb.univ-mrs.fr.
SM: Fax: (418) 656-2861; E-mail: Sylvain.Moineau@bcm.ulaval.ca.

Running title: Phage p2 ORF12 structure
Abstract

We report here the characterization of the non-structural protein ORF12 of the virulent lactococcal phage p2, which belong to the Siphoviridae family. ORF12 was produced as a soluble protein, which forms large oligomers (6-15 mers) in solution. Using anti-ORF12 antibodies, we have confirmed that ORF12 is not found in the virion structure but is detected in the second half of the lytic cycle, indicating that it is a late-expressed protein. The structure of ORF12, solved by SAD and refined at 2.9 Å resolution, revealed a never observed fold as well as the presence of an hydrophobic patch at its surface. Furthermore, crystal packing of ORF12 formed long spirals in which a hydrophobic, continuous, crevice was identified. This crevice exhibited a repeated motif of aromatic residues, which coincided with the same repeated motif usually found in tape-measure protein (TMP), predicted to form helices. A model of a complex between a repeated motif of the TMP of phage p2 (ORF14) was generated, in which the TMP helix fitted exquisitely in the crevice and the aromatic patches of ORF12. We suggest therefore that ORF12 might act as a chaperone for TMP hydrophobic repeats, maintaining TMP in solution during the tail assembly of the lactococcal siphophage p2.
Introduction

During industrial milk fermentation, *Lactococcus lactis* cells are added to transform milk into an array of fermented products such as cheese. However, this manufacturing process may be impaired by lytic phages present in the factory environment as well as in the milk itself (30). Due to the destructive effects of phage infections on bacterial fermentation, much effort has been undertaken to isolate and study the biodiversity of these bacteriophages. Lactococcal bacteriophages belong to at least 10 different genetically-distinct species of double stranded DNA viruses (9). Of them, three lactococcal phage species, all belonging to the *Siphoviridae* family, are the major source of problems in milk fermentation, namely the 936, P335 and c2 species (8, 28, 29). Furthermore, members of the 936 species are by far responsible for the larger part of infections (50-80%) (1, 24, 41). Numerous phages of the 936 species have been isolated and several have been characterized at the genome level (25). However, little is known concerning their molecular mechanisms of infection, although we recently solved the structure of the receptor-binding protein (RBP) of our model 936-like phage, namely the virulent phage p2 (38, 43), and of phages belonging to the P335 species (27, 34, 37, 38).

As for all viruses, bacteriophage genomes are quite compact, leaving little room for non-coding sequences (4). In fact, phage genes are disposed in an operon-type organization (4) and the order of genes corresponds to the different phases of the infection cycle. Moreover, genes are often in clusters (referred to as modules), with gene products from adjacent genes generally found to interact with each other. Interestingly, phage genome organization, including individual gene order, is often conserved within a given species, particularly within the *Siphoviridae* family. In the case of *L. lactis* virulent phages belonging to the 936 or P335 species, this principle applies.
particularly to the morphogenesis gene module, which includes all the genes coding for the phage structural protein genes. For the tail assembly, a module comprises a set of genes between the portal protein, which is connecting the tail to the capsid, and the RBP, which is located at the tip of the tail and is involved in host recognition (39, 43).

The characterization of tail assembly genes of lactococcal phages has been more extensive for temperate siphophages belonging to the P335 species (27, 34, 37, 38). Because of the similarities in genome organization, the findings in this phage species can, in some cases, be used as clues towards understanding the morphology of 936-like phages. For the temperate phage Tuc2009 (P335 species), all structural proteins required for tail and baseplate assembly have been identified (27, 34, 37, 38). Genes located between those encoding for the tape measure protein (TMP) and the BppL (RBP) were identified as corresponding to components of the baseplate structure, located at the tail distal end. Furthermore, gene coding for the major tail protein (MTP) was also identified at a position upstream from the tmp. Between the genes coding for the MTP and the TMP in Tuc2009 are two gene products identified as gpG and gpGT, which are not present in the phage particle. These two proteins were named based on their likely role analogous to the tail assembly proteins present in coliphage lambda, a model virus belonging to the Siphoviridae family (21, 27, 47). gpGT has an essential role in lambda tail assembly, acting prior to tail shaft assembly, while the role of gpG in tail assembly is not known (21). Both gpG and gpGT are also absent from mature lambda virions (21). It has been argued that they may act as assembly chaperones (47).

A close examination of 936 genomes indicates the presence of two genes coding for gpG and gpGT-like proteins. Analysis of phage p2 genome, closely related to the one of lactococcal phage sk1 (7), revealed that the putative tail assembly proteins could correspond to gene products ORF12 and ORF13. These two genes are followed by the TMP gene corresponding to orf14,
other genes coding for other structural proteins and the RBP gene orf18. During our on-going investigation of the structure of phage p2, we are reporting here the cloning, expression and crystal structure of ORF12 in order to decipher its role in the tail assembly process.
MATERIALS AND METHODS

Bacterial strains and phage. *Lactococcus lactis* subsp. *cremoris* MG1363 (14) was grown at 30°C in M17 supplemented with 0.5 % glucose (GM17). In phage p2 (31) infection experiments, 10 mM CaCl$_2$ was added to plates or medium. Propagation of phages and the titers of the lysates were performed as described previously (12).

Intracellular detection of ORF12 during phage infection. *L. lactis* MG1363 strain was grown in GM17 until OD$_{600}$ reached 0.5, then it was infected with virulent phage p2 at multiplicity of infection of 5. Samples were taken at 5 min intervals and flash-frozen (-80°C). Cell pellets were resuspended in 10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 0.3% SDS and lysed with a Bead Beater. The cytoplasmic extracts were then dosed by a standard Bradford assay, and 5 µg of each sample were migrated on a 15% SDS-PAGE. The gel was electrotransferred (30 V) overnight at 4°C with a transblot apparatus (Bio-Rad) onto a PVDF membrane (Hybond P, GE-Healthcare) using Tris-glycine-methanol buffer (25 mM Tris pH 8.3, 192 mM glycine, 10% methanol). The intracellular production of ORF12 during the infection was subsequently detected with a protein A purified anti-ORF12 antibody (Davids Biotechnologie GmbH, Germany). Briefly, the membrane was first blocked with 5% skim milk in PBST (phosphate buffer supplemented with 0.1% Tween-20) for at least 1h on a rotational shaker. The membrane was then treated with a primary antibody anti-ORF12 diluted 1:100 000 (in blocking buffer) for 1h at room temperature. Following washes with PBST, the anti-ORF12 was detected after a 1h incubation with a secondary antibody HRP-labeled anti-rabbit IgG diluted 1:100 000 in blocking buffer (Rockland Immunochemicals). After other washes in PBST, the membrane was rinsed with PBS before the
final detection with the ECL Plus™ detection kit (GE-Healthcare) following the manufacturer instructions.

ORF12 cloning, expression, and purification. The orf12 of phage p2 was cloned into the Gateway™ destination vector pETG-20A (Dr. Arie Geerlof, EMBL Hamburg) for protein production according to the standard Gateway™ protocols and using the following primers for the initial PCR: forward primer, 5’-

\[ GGGGACAAGTTTGTAACAAAAAGCAAGGCTTTAAGAAAACCTGTACTTCCAGGGTGCAAA \]

ACAATTTGAGTACAGCACG-3’, reverse primer, 5’-

\[ GGGGACCACCTTTGTACAAGAAAGCTGGGTATTTTATTTTTTCTCGCCACAATTCG-3’ \]

The att sequences are in italic, TEV recognition site coding sequence is in bold, stop codons are underlined. The final construct encoded a thioredoxin (Trx) fusion protein containing a N-terminal hexahistidine tag followed by a tobacco etch virus (TEV) protease recognition site. Protein expression was done in the *Escherichia coli* Rosetta(DE3)pLysS strain (Novagen). Production of the seleno-methionine labeled protein was performed by blocking the methionine biosynthesis pathway (11). Briefly, cells were grown at 37°C in M9 broth supplemented with 0.1 mM CaCl$_2$, 4 mM MgSO$_4$, 1.2% glycerol, 100 mg/l lysine, 100 mg/l phenylalanine, 100 mg/l threonine, 50 mg/l isoleucine, 50 mg/l leucine, 50 mg/l valine, 50 mg/l seleno-methionine, 1 ml/l oligo-elements. When the optical density reached 0.5, protein expression was induced with 0.5 mM isopropyl-β-thio-galactoside (IPTG) and cells were left overnight at 25°C. Bacterial cells were then harvested by centrifugation at 3,300 xg for 10 minutes, resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.25 mg/ml lysozyme, and EDTA-free antiproteases cocktail (Roche) and frozen at -80°C.
Pellets were quickly thawed at 37°C followed by an incubation with shaking at 4°C in the presence of 20 mM MgSO$_4$ and 10 µg/ml of DNase. Cells were then sonicated and cleared by centrifugation at 21,400 xg. After filtration (0.45 µm filter), the supernatant was loaded on a 5 ml HiTrap™ nickel affinity column (GE Healthcare) pre-equilibrated with buffer (10 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole). Proteins were eluted using the same buffer containing 50 mM and 250 mM imidazole. Prior to TEV protease cleavage, buffer was exchanged to 10 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole on a HiPrep™ 26/10 Desalting (GE Healthcare) column. The cleavage was performed at 4°C overnight using a 1:10 (w:w) ratio of TEV protease to target protein. After cleavage, protein was recovered in the flow-through fraction of a 5 ml HiTrap™ nickel affinity column. The eluted protein was further purified on a HiLoad 26/60 Superdex™ 200 (GE Healthcare) gel filtration column in a buffer containing 10 mM Tris pH 8.0 and 300 mM NaCl. Purified material was visualized on a 15% SDS-PAGE and concentrated to appropriate crystallization concentrations on an Amicon Ultra-15 Centrifugal filter unit with a cut-off size of 5 kDa.

**ORF12 biochemical and biophysical characterization.**

Purified protein was first analyzed for size and incorporation of seleno-methionine by matrix-assisted laser desorption ionisation time-of-flight (MALDI-TOF) mass spectrometry and trypsin peptide mass fingerprint (Brüker autoflex2, Daltonics Bremen, Germany). Then, analytical size-exclusion chromatography (SEC) was carried out with on-line multi-angle static light-scattering on a Alliance 2695 HPLC system (Waters) on a silica gel KW804 column (Shodex) in 10 mM Tris pH 8.0, 300 mM NaCl at a flow of 0.5 ml/min. The protein was loaded at a concentration of 6 mg/ml or 8 mg/ml. Detection was performed using a UV/VIS absorbance photodiode array.
detector 2996 (Waters), a triple-angle static light-scattering detector (MiniDAWNTM TREOS, Wyatt Technology), a quasi-elastic light-scattering instrument (DynaproTM, Wyatt Technology) and, a differential refractometer (Optilab® rEX, Wyatt Technology) (37). Molecular weight and hydrodynamic radius determination were performed by the ASTRA V software (Wyatt Technology) using a $dn/dc$ value of 0.175 ml/g.

**Crystallization, data collection, structure determination, analysis and topological modeling.**

Crystallization trials were performed using a sitting drop vapour diffusion technique implemented on a nanodrop-dispensing robot (PixSys or Honeybee-X8, Cartesian Inc.) in Greiner 288 well plates (10, 19, 40). The protein was initially screened using commercially available crystallization screens: Crystal screens I and II (Molecular Dimensions Ltd), Stura footprint screens (Molecular Dimensions Ltd), and NeXtal SM1 suite (Qiagen). The initial crystal hit was obtained in the Stura II footprinting screen condition number 2 (0.1M HEPES pH 7.5, 18% PEG600). Optimization screens were prepared by varying buffer pH values and precipitant concentrations based on this initial crystallization buffer hit (10, 19, 40). All crystallization plates were stored in a thermo-regulated room at 291K. After 48 h, diffracting crystals grew to a dimension of 50 x 50 x 150 µm in 0.1M HEPES pH 6.8, 15.9% PEG600 using 0.1 µl of protein mixed with 0.2 µl of reservoir solution. Crystals were quick-frozen in a liquid nitrogen flux with 10% glycerol as a cryo-protectant. A complete data set was collected at the European synchrotron radiation facility (ESRF, Grenoble, France) on beamline ID14-EH4. A total of 360 images were collected with an oscillation range of 1° at a wavelength of 0.9785 Å.

The data set was integrated and reduced using MOSFLM and SCALA from the CCP4 suite (5). Phases were calculated with SHELX (36). Phase extension to 2.9 Å was performed and a partial
model was built using RESOLVE (42). Cycles of manual model rebuilding were carried out using Coot (13). Refinement was performed with REFMAC (32) using TLS (Translation/Libration/Screw) segments defined by the TLS Motion Determination server (http://skuld.bmsc.washington.edu/~tlsmd/). Three-fold non-crystallographic symmetry (NCS) restraints were applied throughout refinement to the homotrimer found in the asymmetric unit. The final structure was analyzed using PROCHECK (20) and MolProbity (22). Final coordinates and structure factors were deposited in the Protein Data Bank (PDB) (http://www.rcsb.org/PDB) with code 3D8L. A summary of the structure determination and refinement statistics are presented in Table 1. Interaction models were generated using Coot (13) and refined using Turbo-Frodo (35). An all-canonical alpha-helix was built with Turbo-Frodo. The helix was subsequently slightly bent at the display to adapt to the groove shape, and geometrically refined with Turbo-Frodo option refine. The helix was further docked in ORF12 groove manually at the display with Turbo-Frodo option FBRT. Figures were generated with Pymol (http://pymol.sourceforge.net/) and Turbo-Frodo (35).
Results and discussion

Production and biochemical characterization of ORF12 from the virulent lactococcal phage p2 of the 936 species. Using well-established laboratory screening procedures (44-46), the cloning and subsequent over-production of a seleno-methionine labeled ORF12 full-length protein were successfully carried-out. Optimal conditions gave a yield of approximately 8 mg of soluble and purified protein per liter of E. coli culture in minimal medium. Furthermore, the protein was shown to be quite soluble, up to >10 mg/ml. The combination of good protein yield and high level of protein solubility allowed us to carry out both biophysical characterization and to perform crystallization trials for ORF12.

Basic characterization on a 15% SDS-PAGE identified a single 10 kDa sized band (expected size 10,552 Da). The seleno-methionine protein was also subjected to MALDI-TOF mass spectroscopy (MS), which identified a 10.7 kDa protein. Analysis of trypsin digests by MS also confirmed the full incorporation of 3 seleno-methionines in the protein. A first indication of the oligomerization state of ORF12 arose from protein elution off the HiLoad 26/60 Superdex™ 200 gel filtration column. Based on the calibration curve established for this particular column, ORF12 was identified by SDS-PAGE in fractions eluting between 55 and 65 kDa (data not shown). This suggested that the protein is likely present as an hexamer in solution. ORF12 was then subjected to weight and size analysis using the MALS/UV/RI spectroscopy (37), which confirmed a higher oligomerization state in solution (Fig. 1). The native protein (blue line, Fig. 1) was injected at a concentration of 6 mg/ml. The curve exhibited a slow decrease from the main peak, which corresponded to a mass of 58 ± 3 kDa, to be compared to a theoretical mass of 64.1 kDa for an hexamer. The SeMet labeled ORF12 was injected at 8 mg/ml and exhibited a
comparable behavior, but with the main peak at a mass of 158 ± 8 kDa (red line, Figure 1), which corresponded to a 15-mer polymerization (theoretical: 160.3 kDa). In both cases, however, all forms between the oligomer and the monomer were observed. The hydrophobic effect of the SeMet labeling explains the higher oligomerization state of ORF12 (2).

**Structure of ORF12.** Crystal structure determination yielded a solution with a well-defined electron density map relative to the 2.9 Å resolution. The resulting refined structure had an R/R\textsubscript{free} of 20.3/25.3 (for a complete list of data collection and refinement statistics see Table 1). The asymmetric unit of these R32 ($a = b = 158.3$, $c = 99.5$; $\alpha = \beta = 90.0$, $\gamma = 120.0$) crystals contained three monomers of ORF12 (Fig. 2a). The electron density allowed the reconstruction of all 91 amino acids in chains A and B (Fig. 2b), while the first three amino acids of chain C could not be modeled. These three molecules could be superimposed with main-chain root-mean-square deviation (r.m.s.d.) values of 0.32-0.45 Å, as determined by the SuperPose server (26), within the experimental error at this resolution. The coordinates of ORF12 were submitted to different structure analysis servers (DALI, ProFunc) (16) with in view the identification of closely related structures, and hence a putative function. No structural neighbors were identified, meaning that ORF12 has an original, never observed, fold.

ORF12 is an entirely $\alpha$-helical protein, composed of five $\alpha$-helices total (H1-H5) (Fig. 1). All helices, with the exception of helix H2, are perfectly amphiphilic helices oriented in such a way to form a hydrophobic cleft on the inner surface of the protein. H2 on the other hand has few amino acids involved in the cleft and a majority of outside surface residues. The fifth helix (H5) serves both to form the non-polar cleft and in the interaction between two monomers. Two types of interfaces can be observed between the three monomers of the asymmetric unit in the crystals.
The first two monomers (Chains A and B) are side-by-side interacting through a single helix (helix H5) with the third monomer (Chain C) sitting in a head-to-tail position with the second (Chain B). The side-by-side helices are also oriented in a head-to-tail manner with Glu 82 of one chain H-bonding with Lys 89 of the other, and vice versa (Fig. 3c). The interface formed by the head-to–tail association of two monomers creates an extended hydrophobic cleft whose surface is formed in large part by the non-polar face of helix H3. The interface is mainly formed by non-polar interactions of residues present in the loops and strands between the defined helices. Notably, one monomer has a 10-amino acid loop between H4 and H5 and a 6-amino acid loop between H2 and H3, and the first two residues of H3 at the surface. The other presents both the N and C termini, including the last residues of H5, plus the first two residues of H2. This later helix includes Phe 20, which forms with the C-terminal residue Trp 87 an aromatic residue pocket situated in the hydrophobic cleft (Supp. Figure 1).

The three monomers of the asymmetric unit have very few buried residues in their core (probe radius 1.6 Å, no cut-off). Chains A, B and C have respectively, 87, 86, 84 surface exposed amino acid residues (with a special note concerning chain C, which has 3 residues missing), indicating that virtually all residues are accessible either to solvent, or available for interactions among themselves in the crystal packing.

As mentioned above, the monomers in the asymmetric unit reveal two types of interactions, which were analyzed for the individual monomer positions in the asymmetric unit trimer using the EMBL-EBI Protein Interfaces, Surfaces and Assemblies (PISA) server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver). The largest surface interaction area was found to involve monomers B and C in the trimer, with an interface of 642 Å². A smaller surface was observed between the interactions described above for H5 of chains A and B (149 Å²), while
A and C had virtually no interactions (Fig. 2a). It is the larger surface area involved in the B-C type interaction, which forms the aromatic pocket at the interface of both monomers.

Generating symmetry mates of the ORF12 trimer leads to the formation of two large spirals of compactly packed monomers (Fig. 3). Cross-section of the spirals revealed an inner solvent channel of 36 Å with a complete diameter of 80 Å. The two spirals display weak contacts between them, namely those described above between monomers A and B. The contacts within the spiral are those as described for monomers B and C, with interaction surface areas of 709 Å² for monomer B and its interacting partner, monomer A in a symmetry mate. The B-C interface corresponds to 642 Å², while C with its symmetry mate A have an interface of 754 Å². Therefore, the average value for this type of interaction is 702 Å², corresponding to 11 % of the total surface area of ORF12 monomer. This kind of surface area is comparable to that observed for Fabs/proteins interactions (23) and is above the threshold of surface area being significant for true biological interactions (17). We therefore propose that a crystallographic spiral corresponds to the elongation of the oligomers (15 mers in the case of SeMet ORF12) observed in solution, triggered by the decrease in solubility resulting from the increase of the concentration of precipitant occurring during the crystallization process.

**ORF12 is a non-structural protein, but is expressed during phage infection.** Phage p2 was purified through CsCl gradient to obtain highly concentrated preparations (10¹¹-10¹² PFU/ml), which was used to infect *L. lactis* MG1363. Then, a time-course infection was performed and samples were taken at time intervals. Intracellular extracts were tested for the production of ORF12 (Fig. 4). The phage protein ORF12 was first detected at time 15 min after the beginning of the infection and at the expected size of 10 kDa, confirming that it is a late expressed phage protein (6). The production of ORF12 then peaked at 30 min and its concentration started to
decrease coinciding with lysis of the host culture. This is in agreement with previous data that estimated the latency period of phage p2 (the time from infection to release of new phage progeny) at up to 30 minutes (15). As expected, we could not find ORF12 in the structure of phage p2, thus its role in the phage infection process was still unknown.

Putative function of ORF12: model of the TMP repeat segment and its binding to ORF12.

Each spiral formed by the crystal packing of the non-structural phage p2 protein ORF12 displays several noteworthy features. Firstly, non-polar crevices are located regularly at the inner face of the spiral, within a continuous cleft. In contrast, the external face of the spiral exhibits polar residues facing the solvent. Secondly, within the non-polar cleft, a highly repetitive motif of aromatic residues can be identified (Fig. 5a).

Based on its position in the genome, and by comparison with other Siphoviridae, we hypothesized that ORF12 may play a role in tail assembly. Due to the peculiar characteristics of the tape measure protein (TMP/ORF14) of phage p2, with two hydrophilic domains at each sequence extremity, but with a long hydrophobic helix, likely insoluble by itself (Fig. 5d), we examined the possibility that ORF12 might be a chaperone complexing to the hydrophobic central part of TMP and maintaining it in solution before its assembly with the major tail protein (MTP) to form the phage tail. The TMP acts as a ruler or template that measures length during tail assembly (18). Besides, the N- (1-522 aa) and C-termini (898-999 aa) of the TMP, which likely be globular and interact with the portal protein on the N-terminal side and with the baseplate on the C-terminal side, the middle part exhibits a 40-amino acid repeat of evenly spaced aromatic residues (Fig. 5d). Furthermore, this type of repeat (although varying in length from phage to phage) appears to be a common feature of TMP proteins in Siphoviridae (3, 7, 27).
Since this repeat region of the TMP molecule is predicted to be an $\alpha$-helix, we generated a small 42-mer helical model of this repeat to verify if it could bind within the twisted hydrophobic crevice identified in the large twisted spiral structure of ORF12 (Fig. 3). This structure-structure docking revealed that the amphiphilic helix has correctly spaced aromatic residues to fit into the aromatic residue binding pockets of ORF12 spiral (Fig. 3b,c and Supplemental Fig. 2). These residues are observed on the non-polar surface of the hydrophobic crevice (Fig. 3a), while the other side of the spiral displays a solvent exposed polar-side bearing many charged residues.

A possible function of ORF12 might therefore be to cover the central segment of p2 TMP composed of repeated hydrophobic motifs, to keep them in solution. The N-terminus of TMP (1-521 aa) and its C-terminus (898-999 aa) sequences resemble more globular proteins and are probably soluble by themselves. We hypothesize that maintaining the TMP in solution, would help the assembly of MTPs around the TMP hydrophobic helix, a scheme seen in the virulent Bacillus siphophage SPP1 tail structure (33).

**Conclusion.**

Using structural genomic approaches, it is expected that the knowledge of the 3D structure of a protein of unknown function may reveal its function if its fold resembles that of a protein of known function. The structure of p2 ORF12 being novel, such an approach was not possible. However, because the number of orfs in phages is limited and gene organization may be conserved, genomic comparisons may be the source of functional hypotheses, which can then be checked in silico or experimentally. Here, we suggest, based on strong topological observations, that the non-structural ORF12 of phage p2 (and possibly other siphophages) may serve as a chaperone of TMP central domain to maintain it in solution and present it to MTPs to facilitate
the tail assembly process. The perfect match of ORF12 spiral (observed as an oligomer in solution) hydrophobic patches with the aromatic residues of TMP repeats leads us to postulate that a complex between the TMP residues 522-898 and a spiral of ~70-mer might be stable long enough in solution to allow MTP to approach TMP and the tail to be formed.

Acknowledgements

We thank Dr. Arie Geerlof, who kindly provided the Gateway plasmid pETG-20A for His-Trx fusion. We are grateful to Denise Tremblay and Hélène Deveau for helpful discussion. This work was supported in part by the Marseille-Nice Genopole®, by the company BioXtal and by a grant from the Agence Nationale de la Recherche (BLAN07-1_191968) to CC and VC and by a strategic grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to SM.
References


   Prisant, J. S. Richardson, and D. C. Richardson. 2003. Structure validation by Calpha


   to processing conditions determine the fate of Lactococcus lactis bacteriophages in

25. Mahony, J., H. Deveau, S. Mc Grath, M. Ventura, C. Canchaya, S. Moineau, G. F.
   Fitzgerald, and D. van Sinderen. 2006. Sequence and comparative genomic analysis of
   lactococcal bacteriophages jj50, 712 and P008: evolutionary insights into the 936 phage


28. Moineau, S., M. Borkaev, B. J. Holler, S. A. Walker, J. K. Kondo, E. R. Vedamuthu,
   and P. A. Vandenbergh. 1996. Isolation and characterization of lactococcal
   bacteriophages from cultured buttermilk plants in the United States. J Dairy Sci 79:2104-
   2111.


Table 1: Data collection and refinement statistics for p2 ORF12. All values represented in brackets belong to the last shell.

<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB access code</td>
<td>3D8L</td>
</tr>
<tr>
<td>Space group</td>
<td>R 3 2</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>158.3 158.3 99.5 90.0 90.0 120.0</td>
</tr>
<tr>
<td>Beamline</td>
<td>ID14-EH4</td>
</tr>
<tr>
<td>Detector</td>
<td>ADSC Q315</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9785</td>
</tr>
<tr>
<td>Rotation range (°)</td>
<td>360</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>35.0-2.90 (3.06-2.90)</td>
</tr>
<tr>
<td>Nr. observations</td>
<td>230325 (31122)</td>
</tr>
<tr>
<td>Nr. unique reflections</td>
<td>10739 (1556)</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.9 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>21.4 (20.0)</td>
</tr>
<tr>
<td>I / sI</td>
<td>29.1 (6.4)</td>
</tr>
<tr>
<td>R_{sym} (%)</td>
<td>9.2 (44.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Refinement</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>30.0-2.90 (2.98-2.90)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>10218</td>
</tr>
<tr>
<td>No. atoms protein</td>
<td>2211</td>
</tr>
<tr>
<td>water/buffer</td>
<td>4</td>
</tr>
<tr>
<td>R/R_{free} all last shell</td>
<td>25.3/20.3 (33.5/29.4)</td>
</tr>
<tr>
<td>r.m.s.d. bonds (Å)</td>
<td>0.015</td>
</tr>
<tr>
<td>r.m.s.d. angles (°)</td>
<td>1.591</td>
</tr>
<tr>
<td>Mean B-value (Å²)</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>64.43</td>
</tr>
<tr>
<td>Water</td>
<td>51.33</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
<td></td>
</tr>
<tr>
<td>Favored region</td>
<td>92.8</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>5.7</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>1.5</td>
</tr>
</tbody>
</table>
**Figures legends**

**Figure 1.** On-line multi-angle laser light-scattering, absorbance, and refractive index analysis of ORF12 in solution. The abscissa indicates the time of elution from the HPLC column, the left ordinate indicates the molar mass in g/mol (Da). The absorption peaks are in blue (native) and red (SeMet), and the dash line indicates the molar mass. The experimental mass is given in black digits (kDa). The native ORF12 has been injected at a concentration of 6 mg/ml and the SeMet derivative at 8 mg/ml.

**Figure 2.** Structure of p2 ORF12: A) The crystallographic trimer ribbon representation with monomers identified by their letter in the PDB (3D8L). The ribbon is colored blue to red from N- to C-terminus. The surface interaction area between each monomer are given. B) Stereo view of ORF12 monomer, same coloring as in (A). The helices are numbered 1 to 5.

**Figure 3.** View of the spirals formed by ORF12 crystal packing. The crystal is formed from such spirals packed side by side. A) Side view of the 2 spirals side by side. Each monomer is colored brown, violet and yellow repeatedly. B) View of the spirals rotated by 90° around the vertical axis. The gross dimensions of the spiral are displayed. C) Sticks representation of a stretch of the two side-by-side positioned spirals illustrating the dense network of charges interactions, but loose packing, between the two spirals. The color code is blue for positively charged residues, red for negatively charged residues, pale green for semi-polar residues, green for methionines, yellow for aliphatics and violet for aromatics.
Figure 4. Intracellular detection of ORF12 during phage p2 infection of *L. lactis* MG1363. Samples were taken at various time intervals after phage infection and ORF12 was detected by Western blot analysis. ORF12, 5 ng of purified protein; p2, $1 \times 10^{10}$ PFU of CsCl-purified phages were loaded, which corresponded to $7.8 \pm 1.5 \mu g$.

Figure 5. Model of complex between the TMP hydrophobic helix and ORF12 spiral. A) Spheres representation of a segment of four ORF12 modules in the spiral. The hydrophobic patches are visible in the center of the twisted spiral, in a twisted crevice, formed of aromatic residues. The color code is blue for positively charged residues, red for negatively charged residues, pale green for semi-polar residues, green for methionines, yellow for aliphatics and violet for aromatics. B) View of the TMP segment 777-818, modeled as a curved $\alpha$-helix. All atoms are colored orange, and the aromatic side chains are colored pink. The periodicity of the aromatic residues of TMP coincides with that observed in the ORF12 spiral, as outlined by the red arrows. C) Model of a complex between the TMP segment 777-818 and four ORF12 modules in the spiral. Color coding is the same as in A) and B). D) Sequence of p2 TMP (1-999). The repeats area (522-898) is underlined. The segment chosen in the above modelisation is identified in a yellow filled box with aromatic residues colored in violet.
Time of infection (min)

Non infect

0  5  10  15  20  25  30  35  40  Orf12  p2