Crystal structure of Bacillus subtilis SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP

David Veesler, Stéphanie Blangy, Silvia Spinelli, Paulo Tavares, Valérie Campanacci, Christian Cambillau

PROTEIN STRUCTURE REPORT

Crystal structure of *Bacillus subtilis* SPP1 phage gp22 shares fold similarity with a domain of lactococcal phage p2 RBP

David Veesler,[1] Stéphanie Blangy,[1] Silvia Spinelli,[1] Paulo Tavares,[2] Valérie Campanacci,[1] and Christian Cambillau[1]*

[1]Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I and II, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France

[2]Unité de Virologie Moléculaire et Structurale, CNRS UPR3296 and IFR 115, Bâtiment 14B, CNRS, Gif-sur-Yvette, France

Received 20 April 2010; Revised 28 April 2010; Accepted 28 April 2010
DOI: 10.1002/pro.416
Published online 11 May 2010 proteinscience.org

Abstract: SPP1 is a siphophage infecting the gram-positive bacterium *Bacillus subtilis*. It is constituted by an icosahedric head and a long non-contractile tail formed by gene products (gp) 17–21. A group of 5 small genes (gp 22–24.1) follows in the genome those coding for the main tail components. However, the belonging of the corresponding gp to the tail or to other parts of the phage is not documented. Among these, gp22 lacks sequence identity to any known protein. We report here the gp22 structure solved by X-ray crystallography at 2.35 Å resolution. We found that gp22 is a monomer in solution and possesses a significant structural similarity with lactococcal phage p2 ORF 18 N-terminal "shoulder" domain.

Keywords: phage SPP1; Siphoviridae; *Bacillus subtilis*; crystal structure

Introduction

Phages represent a highly diverse group of viruses infecting bacteria and are the most populated biological entity on earth.[1] The vast majority of them belong to the Caudovirales order and is composed by a double-stranded DNA enclosed in an icosahedral capsid to which is attached a tail. More than 60% of known phages are member of the Siphoviridae family characterized by the presence of a long non-contractile tail, as is the case of SPP1. SPP1 is a virulent *Bacillus subtilis* phage encapsulating its genome in a 60-nm wide isometric capsid connected to a 160-nm long tail.[2,3] Host infection is initiated by the binding of its tail-tip to YueB, a membrane protein with a large ectodomain protruding out of the thick peptidoglycan layer.[4,5] This specific and irreversible interaction triggers a cascade of events resulting in DNA ejection into the *B. subtilis* cytoplasm.

The tail components of SPP1 have been thoroughly identified, and the tail overall structure was studied by electron microscopy, making it possible to assign most of its components (gp17–gp21) to

Abbreviations: r.m.s.d.; root mean square deviation; RBP, receptor binding protein; MALS/QELS/UV/RI, on-line multiangle laser light scattering, quasi-elastic light scattering, UV absorbance and refractive index detectors; EM, electron microscopy

Grant sponsor: ANR; Grant numbers: ANR BLAN07-1_191968, ANR-07-BLAN-0095; Grant sponsor: Ministère français de l'Enseignement Supérieur et de la Recherche; Grant number: 22976-2006.

*Correspondence to: Dr. Christian Cambillau, Architecture et Fonction des Macromolécules Biologiques, UMR 6098 CNRS and Universités d'Aix-Marseille I and II, Campus de Luminy, Case 932, 13288 Marseille Cedex 09, France. E-mail: cambillau@afmb.univ-mrs.fr

Table I. *Data Collection and Refinement Statistics*

SPP1 gp22	
Data collection[a]	
Beamline	BM-14 (ESRF)
Space group	$P2_12_12_1$
Unit cell dimensions (Å)	$a = 42.9$, $b = 64.1$, $c = 101.05$
Wavelenght (Å)	0.97872
Resolution (Å)	54.138–2.35 (2.48–2.35)
R_{sym}[b] (%)	9.0 (46.3)
Mn(I)/σI	27.3 (6.3)
Completeness (%)	100 (100)
Redundancy[a]	14.2 (14.6)
Refinement[a]	
Resolution (Å)	39.68–2.35 (2.57–2.35)
R_{work}/R_{free} (%)	20.7/24.4 (21.9/27.6)
No. of reflections	12,092 (2846)
Mean isotropic B-factor (Å^2)	33.0
r.m.s.d. bond lengths (Å)/angles (°)	0.015/1.53

[a] Values in parentheses refer to the highest resolution shell.
[b] $R_{sym} = \sum (|I(h,i) - I(h)|)/\sum (I(h,i))$

electron densities.[2] However, a large volume of electron density remained unassigned at the tail tip distal end. This fact and the identification of five gene products coded downstream of gp21 (gp22, gp23, gp23.1, gp24, gp24.1), the most distally assigned tail component, led to hypothesize that these gps might form the tip of the tail.[2] To bring further structural insights to prove or disprove this hypothesis, we have undertaken the structural study of these orphan gps.

We report in this contribution the crystal structure of SPP1 gp22 at 2.35 Å resolution. We describe the overall structure and discuss its oligomerization state based on the crystal lattice organization and on static plus dynamic light scattering measurements performed in solution. The gp22 fold is similar to lactococcal phage p2 ORF18 (p2 RBP) shoulder domain, suggesting a possible common origin for these two protein domains.

Results and Discussion

Gp22 crystallization and structure determination

Gp22 crystallized in the orthorombic space group $P2_12_12_1$ with unit cell parameters $a = 42.9$ Å, $b = 64.1$ Å, and $c = 101.05$ Å (Table I). The diffraction limit was 2.35 Å and data quality allowed us to find the six Se sites present in the two molecules of the asymmetric unit and then to phase and build a model. The final model was refined to R_{work} and R_{free} values of 20.66 and 24.36%, respectively. The two monomers of one asymmetric unit are very similar with a r.m.s.d value of 0.5 Å over the first 120 Cα positions [Fig. 1(A)].

Overall structure

Gp22 is a 146 residue-long all-β protein formed by two small three-stranded antiparallel β sheets (βB and βC) and one large five-stranded mixed β sheet (βA). The three β sheets are arranged in two superimposed layers, one of these being formed by βA and the other by βB and βC [Fig. 1(A–C)]. The gp22 fold can thus be described as a pseudo-β-sandwich. The electron density map allowed us to build all residues from the first glycine of our construct to residue 121/120 for both molecules (A/B) of the asymmetric unit. Only a small part of each C-terminus was ordered and thus modeled in the final structure: residues 126–137 and 126–132 in molecule C (a gp22 monomer related to the molecule A by a crystallographic symmetry operation) and B, respectively. Indeed, only a few residues are visible in a region that is stabilized both by hydrogen bonds and van der Waals contacts established with a neighboring monomer in the crystal lattice and involving strand 7 in βC, strand 9 in βA and the loop connecting strands 9–10 in βA. It should be noted that the relative orientation between each monomer of the asymmetric unit and their C-terminal part is different, probably reflecting a high flexibility of the gp22 C-terminus. An intramolecular disulfide bridge is observed in each monomer between cysteine residues 76 and 86 located in strand 9 of β-sheet A and in the turn following strand 9, respectively.

Oligomeric state

In the crystal, each asymmetric unit is composed by a gp22 dimer. The interface buried surface area between the two corresponding monomers is of 1515.7 Å^2 with 703.8 Å^2 and 811.9 Å^2 contributed by monomer A and B, respectively. This interface is comparable in size to typical values found in protein–protein complexes[6] and involves salt bridges, hydrogen bonds and van der Waals contacts. However, SEC/MALS/QELS/RI measurements revealed that gp22 is monomeric in solution with a measured mass of 15.3 kDa (theoretical monomer M_W= 16.7 kDa) and a hydrodynamic radius of 2.3 nm. This strongly suggests that gp22 is monomeric and that the observed dimer is only crystallographic. Other phage proteins were shown to oligomerize and adopt their correct stoichiometry only upon partner binding, as is the case of SPP1 phage adaptor (gp15) and stopper (gp16) proteins.[7] We thus cannot rule out the possibility that gp22 is oligomeric in an *in vivo* context.

Biological role

Gp22 does not have any significant sequence similarity to any other protein hindering thus the possibility to postulate a function based on sequence conservation. However, a DALI search for structurally similar proteins provided a very interesting result: the gp22 structure superimposes onto the shoulder domain of the receptor-binding protein (RBP, ORF 18) of the lactococcal phage p2.[8,9] The structural superimposition between gp22 and the chain A of

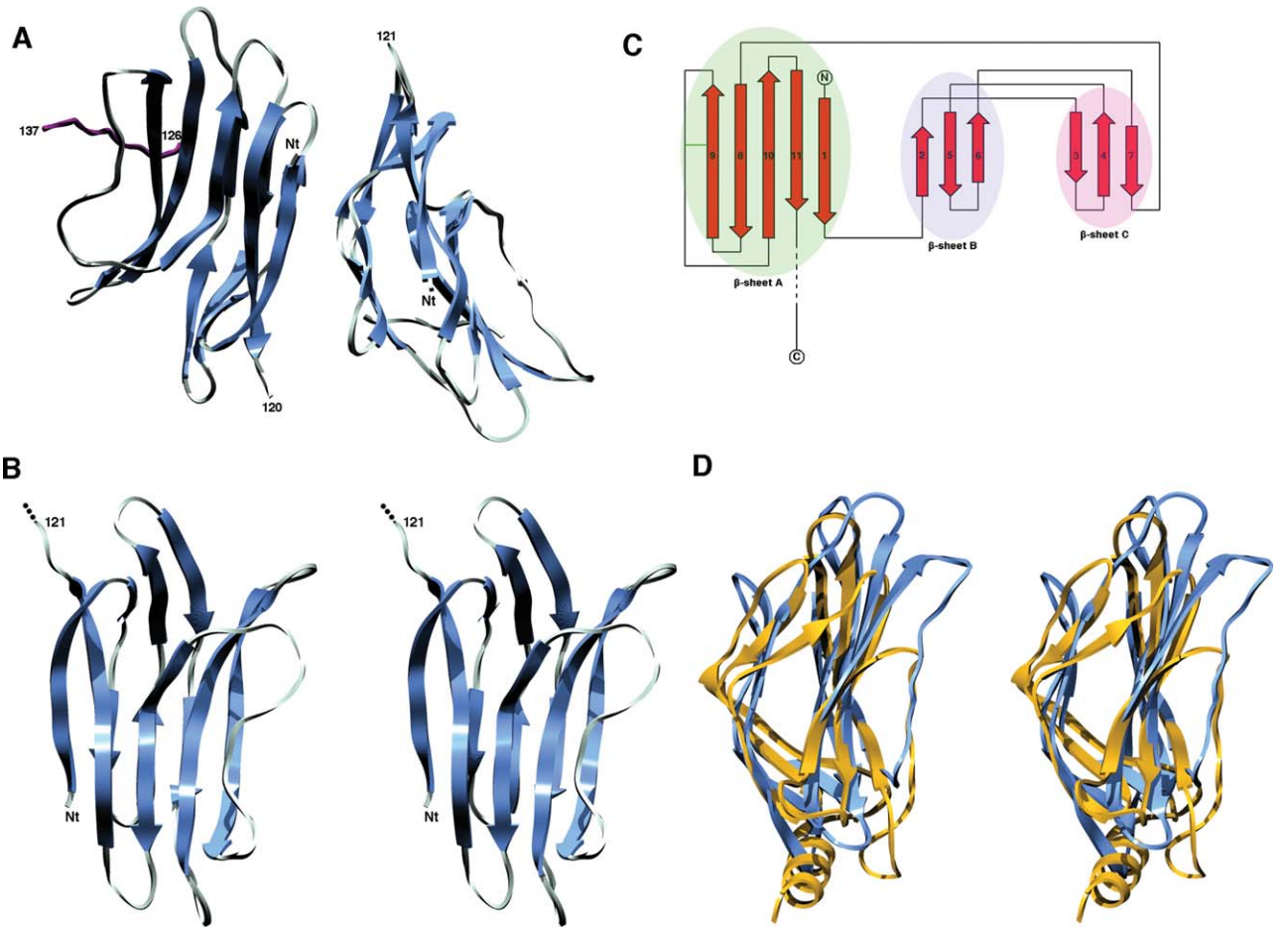

Figure 1. Overall phage SPP1 gp22 structure. A: Ribbon representation of the gp22 dimer in the asymmetric unit. The C-terminal part of monomer C (a symmetry related gp22 monomer) is depicted in purple. B: Stereo ribbon representation of the gp22 monomer. C: Gp22 topology diagram. The disulfide bridge is indicated by a green bar. D: Stereo ribbon representation of the superimposition between SPP1 gp22 and phage p2 ORF18 shoulder domain. Gp22 is depicted in blue and ORF 18 shoulder domain in gold. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

PDB 2WZP yields a Z score of 5.7 with a r.m.s.d value of 3.7 Å over 91 equivalent Cα positions. The two proteins exhibit very close pseudo β-sandwich folding pattern with only subtle differences [Fig. 1(D)]. Nevertheless, the role of the p2 ORF 18 N-terminus being only structural (anchoring the RBP to the baseplate),[8] it is difficult to propose a function based on this observation. It should be noted that the RBP is trimeric in its native state[8,9] and that the shoulder domain oligomerization interface is formed by a three-helix bundle (each monomer contributing an α-helix). As this α-helix is absent in gp22, this oligomerization strategy should not be conserved for this protein *in vivo*. This illustrates a frequently observed phenomenon in phages that is the reuse of a single protein module to fulfill different functions. Finally, we did not find any unambiguous way to fit the gp22 structure into the distal part of the tail tip electron microscopy reconstruction, due to the low resolution (~20 Å) of this contrast EM density.

Materials and Methods

Gp22 cloning, expression, and purification

The SPP1 gp22 nucleotidic sequence was PCR amplified and cloned by Gateway recombination into the pETG-20A vector.[10] The resulting construct encoded a N-terminal fusion with a His_6-tagged thioredoxine followed by a TEV (tobacco etch virus) protease cleavage site. The plasmid was transformed in Escherichia coli T7 Express I^q pLysS strain (New England Biolabs) and expression was carried out at 25°C overnight with 0.5 m*M* IPTG in a minimal medium containing 50 mg/L seleno-methionine for production of seleno-methionine-labeled protein.[11] After harvesting, cell lysis was done by addition of 0.25 mg/mL lysozyme, a freezing/thawing cycle and sonication. Gp22 purification was performed via a first Ni^{2+}-affinity step using 50 m*M* imidazole for elution. After desalting and TEV protease cleavage [10:1 (w/w) protein:TEV protease ratio, 4°C overnight], a

second Ni^{2+}-affinity step was performed followed by a gel-filtration on a preparative Superdex 200 26/60 column. MALDI-TOF mass spectrometry analysis was used to check protein integrity as well as Se-Met incorporation.

Crystallization and structure determination

We performed an initial nanocrystallization screening[12] in 96-well Greiner plates with a protein concentration of 7.5 mg/mL. We optimized a promising hit, obtained in condition 73 of the Structure screen (0.01*M* $CoCl_2$, 0.1*M* MES pH 6.5, 1.8*M* $(NH_4^+)_2SO_4$, Molecular Dimensions Limited), by varying precipitant concentration and pH. The gp22 crystal used for data collection was obtained in 0.1*M* MES pH 7.2, 0.01*M* $CoCl_2$, 2.05*M* $(NH_4^+)_2SO_4$ and flash frozen in mother liquor supplemented with 14% glycerol. Data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on BM14 (SeMet peak data set at the Se K edge, λ = 0.97872 Å). Data processing and scaling were done with XDS,[13] POINTLESS[14] and SCALA.[14] A random selection of 7.5% of the data (test set) was assigned for calculation of the free *R* factor[15] and was not included in the refinement. Phasing was performed by the single anomalous diffraction method using the *phenix.autosol* wizard.[16,17] Resulting phases were the starting point for automatic model building with ArpWarp.[18] Model building and refinement were done with Coot,[19] *phenix.refine*,[20] and BUSTER-TNT.[21] TLS groups definition was assisted by the TLSMD server.[22] Final refinement statistics and model quality are summarized in Table I. Structure analysis was helped by the protein interfaces, surfaces and assemblies server (PISA),[23] the protein–protein interface server,[24] the DALI server[25] as well as with promotif2.[26] Figures were generated with Chimera[27] and topdraw.[28] The coordinates have been deposited to the PDB with accession code 2XC8.

Light scattering measurements

The oligomerization state and size of gp22 in solution were studied by MALS/QELS/UV/RI coupled online with an analytical SEC column, as described.[29–31] MALS, QELS, UV, and RI measurements were achieved with a MiniDawn Treos (Wyatt technology), a Dynapro (Wyatt technology), a Photo Diode Array 2996 (Waters) and an Optilab rEX (Wyatt technology), respectively. The SEC column was a 15-mL KW-803 column (Shodex) run at 0.5 mL/min on an Alliance HPLC 2695 system (Waters) in a buffer containing 10 m*M* Hepes pH 7.5, 150 m*M* NaCl.

References

1. Brussow H, Hendrix RW (2002) Phage genomics: small is beautiful. Cell 108:13–16.
2. Plisson C, White HE, Auzat I, Zafarani A, Sao-Jose C, Lhuillier S, Tavares P, Orlova EV (2007) Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. EMBO J 26:3720–3728.
3. Tavares P, Santos MA, Lurz R, Morelli G, de Lencastre H, Trautner TA (1992) Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. J Mol Biol 225:81–92.
4. Sao-Jose C, Baptista C, Santos MA (2004) Bacillus subtilis operon encoding a membrane receptor for bacteriophage SPP1. J Bacteriol 186:8337–8346.
5. Sao-Jose C, Lhuillier S, Lurz R, Melki R, Lepault J, Santos MA, Tavares P (2006) The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. J Biol Chem 281:11464–11470.
6. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci USA 93:13–20.
7. Lhuillier S, Gallopin M, Gilquin B, Brasiles S, Lancelot N, Letellier G, Gilles M, Dethan G, Orlova EV, Couprie J, Tavares P, Zinn-Justin S (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci USA 106:8507–8512.
8. Sciara G, Bebeacua C, Bron P, Tremblay D, Ortiz-Lombardia M, Lichiere J, van Heel M, Campanacci V, Moineau S, Cambillau C (2010) Structure of lactococcal phage p2 baseplate and its mechanism of activation. Proc Natl Acad Sci USA 107:6852–6857.
9. Spinelli S, Desmyter A, Verrips CT, de Haard HJ, Moineau S, Cambillau C (2006) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. Nat Struct Mol Biol 13:85–89.
10. Vincentelli R, Bignon C, Gruez A, Canaan S, Sulzenbacher G, Tegoni M, Campanacci V, Cambillau C (2003) Medium-scale structural genomics: strategies for protein expression and crystallization. Acc Chem Res 36: 165–172.
11. Doublie S (1997) Preparation of selenomethionyl proteins for phase determination. Methods Enzymol 276: 523–530.
12. Sulzenbacher G, Gruez A, Roig-Zamboni V, Spinelli S, Valencia C, Pagot F, Vincentelli R, Bignon C, Salomoni A, Grisel S, Maurin D, Huyghe C, Johansson K, Grassic A, Roussel A, Bourne Y, Perrier S, Miallau L, Cantau P, Blanc E, Genevois M, Grossi A, Zenatti A, Campanacci V, Cambillau C (2002) A medium-throughput crystallization approach. Acta Cryst D58: 2109–2115.
13. Kabsch W (1988) Automatic indexing of rotation diffraction patterns. J Appl Cryst 21:67–72.
14. Evans P (2006) Scaling and assessment of data quality. Acta Cryst D62:72–82.
15. Brunger AT (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472–475.
16. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung LW (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Cryst D65:582–601.
17. Zwart PH, Afonine PV, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, McKee E, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Storoni LC, Terwilliger TC, Adams PD (2008) Automated structure solution with the PHENIX suite. Methods Mol Biol 426: 419–435.
18. Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, Kirillova O, Vonrhein C, Perrakis A, Lamzin VS

(2004) Breaking good resolutions with ARP/wARP. J Synchrotron Radiat 11:56–59.
19. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Cryst D66: 486–501.
20. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Cryst D64:61–69.
21. Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Cryst D60:2210–2221.
22. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Cryst D62:439–450.
23. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.
24. Reynolds C, Damerell D, Jones S (2009) ProtorP: a protein-protein interaction analysis server. Bioinformatics 25:413–414.
25. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. Bioinformatics 24:2780–2781.
26. Hutchinson EG, Thornton JM (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. Protein Sci 5:212–220.
27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.
28. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19:311–312.
29. Veesler D, Blangy S, Siponen M, Vincentelli R, Cambillau C, Sciara G (2009) Production and biophysical characterization of the CorA transporter from Methanosarcina mazei. Anal Biochem 388:115–121.
30. Campanacci V, Veesler D, Lichiere J, Blangy S, Sciara G, Moineau S, van Sinderen D, Bron P, Cambillau C (2010) Solution and electron microscopy characterization of lactococcal phage baseplates expressed in *Escherichia coli*. J Struct Biol [Epub ahead of print].
31. Veesler D, Dreier B, Blangy S, Lichiere J, Tremblay D, Moineau S, Spinelli S, Tegoni M, Pluckthun A, Campanacci V, Cambillau C (2009) Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARPin and camelid VHH binding mode. J Biol Chem 284:30718–30726.