Protein Expression and Purification 70 (2010) 116–121

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Two-chaperone assisted soluble expression and purification of the bacteriophage T4 long tail fibre protein gp37

Sergio Galan Bartual [a], Carmela Garcia-Doval [a], Jana Alonso [b], Guy Schoehn [c,d], Mark J. van Raaij [a,e,*]

[a] Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain
[b] Laboratorio de Proteómica, Instituto de Investigación Sanitaria de Santiago, Complexo Hospitalario Universitario de Santiago, Spain
[c] Unit for Virus Host Cell Interaction, UMI 3265, Université Joseph Fourier, EMBL, CNRS, Grenoble, France
[d] Institut de Biologie Structurale Jean-Pierre Ebel, UMR5075 CEA-CNRS-UJF, Grenoble, France
[e] Instituto de Biologia Molecular de Barcelona (CSIC-Parc Cientific), c/Josep-Samitier 1-5, E-08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 11 August 2009
and in revised form 2 November 2009
Available online 12 November 2009

Keywords:
Bacteriophage
Fibre
Trimer
Chaperones
Electron microscopy

ABSTRACT

Bacteriophage T4 recognises its host cells through its long tail fibre protein gene product (gp) 37. Gp37 is a protein containing 1026 amino acids per monomer, forming a fibrous parallel homotrimer at the distal end of the long tail fibres. The other distal half-fibre protein, gp36, is much smaller, forming a trimer of 221 amino acids per monomer. Functional and structural studies of gp37 have been hampered by the inability to produce suitable amounts of it. We produced soluble gp37 by co-expression with two bacteriophage T4-encoded chaperones in a two-vector system; co-expression with each chaperone separately did not lead to good amounts of correctly folded, trimeric protein. An expression vector for the bacteriophage T4 fibrous protein chaperone gp57 was co-transformed into bacteria with a compatible bi-cistronic expression vector containing bacteriophage T4 genes 37 and 38. A six-histidine tag is encoded amino-terminal to the gp37 gene. Recombinant trimeric gp37, containing the histidine tag and residues 12–1026 of gp37, was purified from lysed bacteria by subsequent nickel-affinity, size exclusion and strong anion exchange column chromatography. Yields of approximately 4 mg of purified protein per litre of bacterial culture were achieved. Electron microscopy confirmed the protein to form fibres around 63 nm long, presumably gp36 makes up the remaining 11 nm in the intact distal half-fibre. Purified, correctly folded, gp37 will be useful for receptor-binding studies, high-resolution structural studies and for specific binding and detection of bacteria.

Introduction

Bacteriophages are complex one-time nanomachines that infect bacteria [1], outnumbering their hosts in some environments [2]. Their infection efficiency and high levels of specificity has led to their use in applications such as identification of bacteria [3] and phage therapy [4]. Bacteriophage T4 is a tailed phage that belongs to the *Myoviridae* family of the order *Caudovirales*, which includes 95% of all bacteriophages [5]. Bacteriophage T4 has a double-stranded DNA genome of 168,903 nucleotides that codifies for more than 40 structural proteins [6]. The structural proteins make up the head (containing the genomic DNA), the neck and the tail; the tail includes the base-plate and long and short tail fibres.

In the first step of infection, the long tail fibres recognise and bind reversibly to a surface receptor of the *Escherichia coli* host [7]. The C-terminal region of the long tail fibre protein gp37 mediates this specific interaction. When at least three of the six long tail fibres have correctly recognised their receptor, the base-plate changes conformation from the hexagon to the star shape [8], causing the short tail fibres, which are parallel homotrimers of gp 12 to extend and bind irreversibly to the host cell lipopolysaccharide inner core [9]. In this way, they function as immobile stays, and when the outer tail sheath contracts, the inner tail tube punctures the cell wall and injects the viral DNA into the cell [8,10].

The bacteriophage T4 long tail fibres are made up of four different proteins, encoded by genes 34, 35, 36 and 37 [11,12] and called gp34, gp35, gp36 and gp37, respectively. Scanning electron microscopy studies showed them to form a rod-like structure divided in a proximal half-fibre of about 70 nm long and a distal half-fibre of about 74 nm long, connected at an angle of around 160° [13]. The proximal-half fibres consist of homotrimers of gp34 (forming the "thigh") and a monomer of gp35 (the "knee"), while the distal half-fibres contain a homotrimer of gp36 (near the "knee") and a homotrimer of gp37, forming the "shin" and "foot". Receptor-binding is mediated by gp37; both lipopolysaccharide and the outer

* Corresponding author. Address: Instituto de Biologia Molecular de Barcelona (CSIC-Parc Cientific), Departamento de Biologia Estructural, c/Josep-Samitier 1-5, E-08028 Barcelona, Spain.
E-mail address: mark.vanraaij@ibmb.csic.es (M.J. van Raaij).

1046-5928/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2009.11.005

membrane protein C (OmpC) have been reported as primary T4 receptors [14]. Gp37 contains 1026 amino acids per monomer.

Bacteriophage T4 employs three known chaperone proteins, gp31, gp38 and gp57. Gp31 substitutes for GroES in the folding of the bacteriophage head protein gp23, enlarging the folding cage [15]. Gp57 appears to be a "general" T4 tail fibre chaperone and is needed for the correct trimeric assembly of the trimeric short and long tail fibre proteins gp12, gp34, gp36 and gp37 [16]. Gp57 is a small protein of 79-residues (8613 Da) that lacks aromatic amino acids, cysteines and prolines. *In vitro*, it adopts different oligomeric states [17]. For the correct trimeric assembly of gp37, the specific chaperone gp38 must also be present [16]. The molecular bases for the mechanisms of the chaperone activities of gp38 and gp57 are as yet unclear.

The atomic structure of a large part of the short tail fibre protein gp12 has become available thanks to co-expression in *E. coli* of gp12 with gp57, which led to correctly folded protein [18,19], followed by limited proteolysis, crystallisation and structure solution [20,21]. Up to now, no structures have been reported for any of the long tail fibre proteins, and they only have limited and local sequence similarity with other beta-structured fibrous proteins [22]. Here, we present a double-plasmid *E. coli* expression system for the co-expression of bacteriophage T4 gp37 (amino acids 12–1026) with the viral chaperones gp57 and gp38. The expressed protein gp37 contains an amino-terminal six-histidine expression tag for convenient purification. The protein has been purified and shown to fold into 63 nm long fibres by electron microscopy.

Materials and methods

Purification of bacteriophage T4 DNA

One ml from an overnight culture of *E. coli* K12 was inoculated with bacteriophage T4 obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany) and incubated 30 min at 37 °C with vigorous shaking. Infected cells where transferred to a 50 ml flask with 4 ml of fresh LB broth and incubated as above until lysis was observed (4–6 h). Remaining *E. coli* cells were lysed by adding four drops of chloroform followed by 5 min incubation at 37 °C. Cultures were then centrifuged at 3900 *g* at 4 °C for 20 min and viral DNA was extracted from the supernatant using a bacteriophage DNA extraction kit (Lambda Mini Kit, Qiagen, Hilden, Germany).

Construction of expression plasmids

The gp57 expression plasmid pET(Ap)g57 (Fig. 1) was used as source for the T4 gp57 chaperone. This vector is based on pET21a(+) (Novagen, Nottingham, UK) and contains a gene for conferring ampicillin resistance and a ColE1 (pBR322) replicon. The nucleotide sequences coding for gp38 and gp37 were obtained by polymerase chain reactions [23] using T4 DNA as template. Primers used for amplification of the gp37 gene were 5′-GGCCGGATCC**AAG**-CAAAAYCGCAGGWRCACGTCC-3′ (forward primer) and 5′-CCGCTC-GAGCTTATGCTAAACGAACGATATAG-3′ (reverse primer); Y indicates a mixture of C and T; W of A and T. BamHI and XhoI restriction sites are underlined and the first cloned codon of gp37 (amino-acid number 12) is marked in bold. Primers used for amplification of the gp38 gene were 5′-GGCCCATATGTCGGCCCTTCTAAAT**ATG**AAAA-TAT-3′ (forward primer) and 5′-GGCCCTCGAGTTATACGTAAT GCTTGAATAAACGC-3′ (reverse primer). NdeI and XhoI restriction sites are underlined and the start codon of gp38 is marked in bold. Amplified DNA was purified using a QIAquick PCR Purification Kit (Qiagen GmbH, Hilden, Germany). The CDF replicon based expression plasmid pCDF-Duet (Streptomycin resistance) (Novagen, Nottingham, UK) was used for the cloning of the gp37 and gp38-encoding genes, yielding pCDF(Sm)g37 and pCDF(Sm)g37g38, respectively, (Fig. 1). The gp37-encoding gene was cloned into the first multi-cloning site (MCS1) of pCDF, yielding pCDF(Sm)g37. Then, the gp38-encoding sequence was cloned into MCS2 of pCDF(Sm)g37, yielding pCDF(Sm)g37g38 (Fig. 1). Coding sequences for gp37 and gp38 were analysed by DNA sequence analysis to verify the integrity and correctness of the inserts (Sistemas Genomicos, Valencia, Spain). The vectors described express gp37 residues 12–1026 with an N-terminal six-histidine tag, both chaperones gp38 and gp57 are expressed without any purification tags.

Expression, solubility and quaternary structure of gp37 with and without chaperones

To test soluble expression of gp37 in the absence and presence of gp57 and/or gp38, cultures of JM109(DE3) transformed with pCDF(Sm)g37 or pCDF(Sm)g37g38 and the same plasmids co-transformed with pET(Ap)gp57 were prepared to test all four combinations: gp37 alone, gp37 with gp57, gp37 with gp38 and gp37 with both gp57 and gp38. Transformed bacteria were grown at 37 °C in 200 ml of LB medium (10 g/l Bacto-tryptone, 5 g/l yeast extract, 10 g/l sodium chloride), supplemented with ampicillin (50 mg/l) and streptomycin (50 mg/l) as needed (Sigma–Aldrich, Steinheim, Germany), to an optical density of approximately 0.7 units measured at 600 nm. After cooling on ice, expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG)[1] to a final concentration of 0.1 mM and growth was continued overnight at 16 °C. Bacteria were harvested by centrifugation for 20 min at 4000 *g* at 4 °C and pellets were re-dissolved in 5 ml lysis buffer (50 mM Tris–HCl pH 7.4, 50 mM ammonium chloride and 4% (w/v) glycerol).

Cells were lysed by 10 rounds of 15 s sonications alternated with incubation on ice using an UP 200S Ultraschall Prozessor (Dr. Hielscher Gmbh, Teltow, Germany) and extracts were centrifuged for 20 min at 20,000 *g* at 10 °C. Samples from the supernatant and the pellet re-suspended in 5 ml of lysis buffer were checked by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS–PAGE) [24] using 10% polyacrylamide gels. To detect the presence of SDS-resistant trimeric protein, samples without heating and heated for 4 min at 98 °C prior to loading were analysed [25].

Expression and purification of soluble, trimeric gp37

One litre cultures of *E. coli* JM109 (DE3) freshly transformed with pCDF(Sm)g37g38 and pET(Ap)g57 were grown, induced and harvested as described above, redissolving pellets in 20 ml lysis buffer. Protease inhibitor cocktail (Sigma–Aldrich, Steinheim, Germany) was added and re-suspended pellets were stored at −20 °C. Bacteria were lysed by a triple pass through an emulsifyer (Avestin emulsifyer C5, Avestin Europe GmbH, Mannheim, Germany). Magnesium sulphate was added to lysates to a final concentration of 10 mM and nucleases (ribonuclease and desoxyribonuclease) (Sigma–Aldrich, Steinheim, Germany) were added to a final concentration of 0.01 mg/ml each. Extracts were incubated 30 min on ice and clarified by centrifugation (1 h 39,000*g* at 10 °C). Supernatants were supplemented with 50 mM imidazole and loaded onto a nickel–nitrilotriacetic acid agarose (Qiagen, Hilden, Germany) column (1 ml of agarose suspension per 4 ml of extract). The column was washed twice with buffer A (50 mM Tris–HCl pH 7.4, 300 mM sodium chloride, 1% (w/v) glycerol), supplemented with 50 mM imidazole. Elution was done with four passes of 4 ml of buffer A

[1] *Abbreviations used:* gp, gene product; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IPTG, isopropyl-β-d-thiogalactoside; SDS–PAGE, polyacrylamide gel electrophoresis.