FEBS Letters 584 (2010) 669–674

FEBS Letters

journal homepage: www.FEBSLetters.org

Stoichiometric protein complex formation and over-expression using the prokaryotic native operon structure

Christian Poulsen, Simon Holton, Arie Geerlof, Matthias Wilmanns, Young-Hwa Song *

European Molecular Biology Laboratory (EMBL), Hamburg Outstation, c/o DESY, Notkestrasse 85, 22607 Hamburg, Germany

ARTICLE INFO

Article history:
Received 20 July 2009
Revised 22 December 2009
Accepted 22 December 2009
Available online 18 January 2010

Edited by Gianni Cesareni

Keywords:
Operon
Protein–protein interaction
Protein complex
Structural biology
Crystallization of protein complexes
Mycobacterium tuberculosis

ABSTRACT

In prokaryotes, operon encoded proteins often form protein–protein complexes. Here, we show that the native structure of operons can be used to efficiently overexpress protein complexes. This study focuses on operons from mycobacteria and the use of *Mycobacterium smegmatis* as an expression host. We demonstrate robust and correct stoichiometric expression of dimers to higher oligomers. The expression efficacy was found to be largely independent of the intergenic distances. The strategy was successfully extended to express mycobacterial protein complexes in *Escherichia coli*, showing that the operon structure of gram-positive bacteria is also functional in gram-negative bacteria. The presented strategy could become a general tool for the expression of large quantities of pure prokaryotic protein complexes for biochemical and structural studies.

Structured summary:
MINT-7542207: *ESAT-6* (uniprotkb:Q50206) and *CFP-10* (uniprotkb:O33084) *bind* (MI:0407) *by blue native page* (MI:0276)
MINT-7542534: *ESAT-6* (uniprotkb:P0A564) and *CFP-10* (uniprotkb:P0A566) *bind* (MI:0407) *by X-ray crystallography* (MI:0114)
MINT-7542187: *CFP-10* (uniprotkb:P0A566) and *ESAT-6* (uniprotkb:P0A564) *bind* (MI:0407) *by blue native page* (MI:0276)
MINT-7542652: *CFP-10* (uniprotkb:P0A566) and *ESAT-6* (uniprotkb:P0A564) *bind* (MI:0407) *by molecular sieving* (MI:0071)
MINT-7542474, MINT-7542303: *CFP-10* (uniprotkb:P0A566) *physically interacts* (MI:0915) with *ESAT-6* (uniprotkb:P0A564) *by pull down* (MI:0096)

1. Introduction

The structural and functional characterization of multi-component protein complexes is a major area of study in the field of systems biology. In the past decade of the post genomic era, several protein interaction networks of diverse organisms have been reported [1]. These studies have revealed that proteins often do not function in isolation within cells, but form functional complexes by associating to binding partners. Although it has been known for a long time that functionally related genes are organized in operons in prokaryotes and that they are transcribed in a coordinated manner [2], systematic computational studies for prediction of operons have only been carried out recently [3]. To study protein complexes at the structural or biochemical level, generally milligrams of purified complex are required. Recombinant technology has been routinely used to produce single proteins using heterologous expression systems. Although several co-expression strategies are available for both eukaryotic and prokaryotic hosts, it often remains challenging to obtain large amounts of native protein complexes with defined stoichiometric functional compositions [4]. Here we take advantage of the fact that functionally related genes in prokaryotes are organized in operons and that they are co-transcriptionally and co-translationally regulated [2,5,6]. Currently about 1110 bacterial genomes have been sequenced [7]. Systematic computational tools for the annotation of genes are being developed and several studies are available for predicting prokaryotic operons [3,5,8]. Studies aimed at predicting *Mycobacterium tuberculosis* (*M. tuberculosis*) operons have recently suggested that there are more than 1000 operons consisting of two or more open reading frames (ORFs) [3,9,10]. These data also indicate that more than half of the 4000 ORFs of *M. tuberculosis* are organized in operons [3,11] (Supplementary Table 1).

* Corresponding author. Fax: +49 4089902149.
E-mail address: song@embl-hamburg.de (Y.-H. Song).

0014-5793/$36.00
doi:10.1016/j.febslet.2009.12.057

In this study we have investigated the possibility of establishing a generic, effective and reliable expression and purification method for protein complexes by exploiting the native operon structures in *M. tuberculosis*. Initially we explored whether native operons are functional in a homologous expression host of the same taxon, in this case *M. smegmatis*. As proof of principle, we explored three paralogous and two orthologous operons of the WXG-100 protein family. To further demonstrate the applicability of our method, we extended the study to the operon encoding the proteasome, a large protein oligomer. Furthermore, we show that the operon structures of a gram-positive bacterium (*M. tuberculosis)* are also recognized heterologously by a gram-negative bacterium (*Escherichia coli*). This study represents a substantial advance in that (i) it reduces the design of multi-gene expression constructs to a single step and (ii) it permits the expression of milligrams of protein complex in a native and functional state using inducible promoters.

2. Materials and methods

2.1. Construction, expression and purification of recombinant protein complexes

For the construction of the expression vectors, the studied operons were amplified from genomic DNA of *M. tuberculosis* (strain H37Rv), *M. leprae and M. smegmatis* (strain mc2 155) using PCR. The employed expression vectors were pMyNT, a pSD26 derivative, for over-expression in *M. smegmatis* [12] and pETM-Z (a kind gift from Günter Stier) or pETM-11 [13] for over-expression in *E. coli*. The primers were designed such that the PCR products could be ligated into the expression vectors, which were linearized using the restriction enzymes, NcoI, BamHI and HindIII, respectively. Primers are listed in Supplementary Table 2.

All methods relating to protein expression in *M. smegmatis* were carried out following the reported method [12]. In brief, the *M. smegmatis* mc2 155 competent cells were transformed with the positive recombinant plasmids using electroporation and plated out on 7H10 agar plates supplemented with 50 μg/mL hygromycin. Single colonies were inoculated to express the protein complex of interest. The culture medium consisted of Middlebrook 7H9 medium supplemented with 0.2% (v/v) glycerol, 0.2 % (w/v) glucose, 0.05% (v/v) tween 80 and 50 μg/mL hygromycin. The culturing condition was 37 °C with a 200 rpm shaking speed. The protein expression was started at 2.0 OD^{600nm} by adding acetamide to a final concentration of 35 mM.

Expression in *E. coli* was carried out by following the reported method [32] using BL21(DE3) Codon Plus RIL strain (Stratagene). In both expression systems the cells were incubated over night after induction of protein expression.

The purification of the expressed protein complexes was standardized. The crude extracts were obtained by centrifugation (30.000 rcf) after sonication of the cells in the extraction buffer, 50 mM Tris–HCl, pH 8.0, 300 mM NaCl and 20 mM imidazol. The protein complexes were further purified in three steps. First, His-affinity chromatography was performed using 50 mM Tris–HCl, pH 8.0, 300 mM NaCl, 500 mM imidazol as elution buffer. Second, the His-tag was removed by TEV protease treatment in extraction buffer supplemented with 10 mM β-mercapto-ethanol, and as the polishing step size-exclusion chromatography was applied using the buffer, 50 mM Tris–HCl, pH 8.0, 300 mM NaCl.

2.2. Mass spectrometry analysis

The purified complexes were identified by mass spectroscopic analysis. The samples were prepared according to the manufacturer's instruction. In brief, the excised bands from the SDS and native gels were destained, dehydrated, vacuum dried and incubated overnight with methylated porcine trypsin (trypsin gold, Promega). Peptides, as well as full-length complex, were analyzed with MALDI-TOF using a Voyager-DE-STR mass spectrometer.

2.3. Sample preparation for electron microscopy (EM)

The specimens for electron microscopy were prepared according to published methods [14]. In brief, about 1–2 μL proteasome solution at a concentration of 0.5 μM were applied to a carbon coated, glow discharged 600-mesh EM grid. The specimens were negatively contrasted by overlaying with 2% uranyl acetate. The electron micrographs were recorded on a Philips CM12 microscope in low dose mode.

3. Results

3.1. Native operon structures are recognized by bacteria within the same taxon

Comparative genomic analyses of *M. tuberculosis* and the avirulent vaccine strain *BCG* have revealed 16 *Regions of Difference*, designated RD1-RD16, respectively [11,15,16]. Among them, RD1 is the only region which is present in all virulent strains and absent in all avirulent ones [17,18]. The potential virulence factors CFP-10 and ESAT-6 have been studied intensively due to their property to generate strong T cell immune responses [19]. They are encoded in an operon that is an integral part of RD1 [20]. CFP-10 and ESAT-6 (Rv3874, Rv3875) form a 1:1 heterodimeric complex [21,22]. The two proteins are part of a large protein family, which is generally present among mycobacteria irrespective of their virulence [23]. The genome of *M. tuberculosis* contains 11 CFP-10/ESAT-6-like pairs, five pairs of which are situated within their own gene clusters encoding a type-VII secretion system as in RD1. The NMR structure of the CFP-10/ESAT-6 complex was determined from proteins that were separately expressed in *E. coli* and co-refolded, post purification, to produce a heterodimeric complex [21].

Here, we have cloned the entire coding region of the CFP-10/ESAT-6 operon, including their intergenic base pairs into the expression vector pMyNT (Fig. 1A), to be used in the homologous expression host *M. smegmatis*. The CFP-10/ESAT-6 complex could be expressed and purified in an equimolar ratio of the two protein components (Fig. 2A, lanes 1 and 4). The yield of purified protein complex was about 8 mg per litre of cell culture.

We further tested the method for other CFP-10/ESAT-6-like coding operons, all of which possess different intergenic distances ranging from 10 to 32 base pairs (Fig. 1B). The sequence identity between the different paralogs and orthologs is in the range of 15–65%. We were able to successfully express three CFP-10/ESAT-6 paralogous complexes from *M. tuberculosis* and the orthologous complexes from *M. leprae* and *M. smegmatis* (Fig. 4). All pairs were expressed in equimolar ratios and therefore migrated as single bands with native PAGE (polyacrylamide gel electrophoresis) (Fig. 4B). This confirms that the native operon structures can be used for homologous recombinant expression of targets consisting of two ORFs, irrespective of the different intergenic distances and base pairs.

To generalize and to demonstrate the advantages of our method in the context of potentially challenging large protein assemblies, we selected a high molecular weight test system from *M. tuberculosis*, the proteasome complex, for expression. The proteasome is encoded by an operon containing the ORFs Rv2109c (α-subunit) and Rv2110c (β-subunit), forming an $\alpha_7\beta_7\beta_7\alpha_7$ complex with a molecular weight of 730 kDa [24]. With our protocol, this