Protein Expression and Purification 62 (2008) 1–8

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Effective high-throughput overproduction of membrane proteins in *Escherichia coli*

Euan Gordon [a,1], Rob Horsefield [b], Herman G.P. Swarts [c], Jan Joep H.H.M. de Pont [c], Richard Neutze [b], Arjan Snijder [a,*,1]

[a] *Department of Chemical and Biological Engineering, Chalmers University of Technology, Box 462, 40530 Gothenburg, Sweden*
[b] *Department of Chemistry, Biochemistry and Biophysics, University of Gothenburg, Box 462, 40530 Gothenburg, Sweden*
[c] *Department of Biochemistry, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands*

ARTICLE INFO

Article history:
Received 1 March 2008
and in revised form 1 July 2008
Available online 24 July 2008

Keywords:
Membrane proteins
His-tag
Structural genomics
Escherichia coli
Auto-induction
Protein production
Legionella pneumophila
Legionnaires' disease

ABSTRACT

Structural biology is increasingly reliant on elevated throughput methods for protein production. In particular, development of efficient methods of heterologous production of membrane proteins is essential. Here, we describe the heterologous overproduction of 24 membrane proteins from the human pathogen *Legionella pneumophila* in *Escherichia coli*. Protein production was performed in 0.5 ml cultures in standard 24-well plates, allowing increased throughput with minimal effort. The effect of the location of a histidine purification tag was analyzed, and the effect of decreasing the length of the N- and C-terminal extensions introduced by the Gateway cloning strategy is presented. We observed that the location and length of the purification tag significantly affected protein production levels. In addition, an auto-induction protocol for membrane protein expression was designed to enhance the overproduction efficiency such that, regardless of the construct used, much higher expression was achieved when compared with standard induction approaches such as isopropyl-β-D-thiogalactopyranoside (IPTG). All 24 targets were produced at levels exceeding 2 mg/l, with 18 targets producing at levels of 5 mg/l or higher. In summary, we have designed a fast and efficient process for the production of medically relevant membrane proteins with a minimum number of screening parameters.

Membrane proteins are fundamental components of most basic cellular processes such as energy generation, nutrient uptake, waste product removal and cell-signaling. They typically comprise up to 20–30% of the protein coding regions of all sequenced genomes [1]. Membrane proteins are also of tremendous economical importance as many drugs and drug candidates target membrane proteins. Despite this, there are relatively few membrane proteins with known 3D structure: less than two percent of the ~50,000 structures in the Protein Data Bank (PDB)[2] [2] are of independent integral membrane proteins. This under-representation is a consequence of the challenges associated with membrane proteins: low endogenous expression, high hydrophobicity requiring the presences of detergents and poor stability once solubilized all hinder attempts to obtain protein crystals suitable for X-ray crystallographic analysis.

Typically milligram quantities of pure protein are required for structural investigations, which has resulted in a bias of structure determination efforts towards membrane proteins which are naturally abundant, such as photosynthetic reaction centers [3] or aquaporins [4]. A more pain-staking approach is to circumvent the problem of obtaining recombinant proteins by growing kilograms of cells, for example see [5,6], to make up for low protein quantities. However, for both economic and practical reasons it is uncommon that this is a realistic solution for many academic laboratories. Optimization of membrane protein overproduction is therefore central to improved productivity.

With respect to this problem, purification tags, such as MBP, Flag and Streptavidin tags [7], have facilitated simple and successful purification. However, the Histidine tag (His-tag) is the most extensively used tag, where a stretch of six to ten histidine residues at the N- or C-terminus, or even within the protein, preferentially interact with divalent metal ions immobilized on chromatographic beads. A survey of the PDB shows that 17% of the deposited structures contain a six histidine sequence. Purification tags, however, can dramatically influence the expression level and hence the yield of purified target proteins [8]. In the case of AqpZ, an *Escherichia coli* aquaporin, the location of the tag did not influence protein production but the use of a 10 histidine tag instead of a six histidine

* Corresponding author.
E-mail address: arjan.snijder@astrazeneca.com (A. Snijder).
[1] Present address: AstraZeneca R&D Mölndal, Department of Cell, Protein and Structural Sciences, SE-431 83 Mölndal, Sweden.
[2] *Abbreviations used:* PDB, protein data bank; βOG, β-octylglucopyranoside; DDM, dodecylmaltoside; IMAC, immobilized metal affinity chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside.

1046-5928/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2008.07.005

tag did [9], a specific example where tag location was unimportant. While other examples of the influence of tags are available, including an extensive analysis of a *multi-construct approach* [10], there is no systematic survey of the effect of tag location with a range of different membrane proteins.

Here, we describe the heterologous expression of 24 membrane proteins, our bench-mark set, from the human pathogen *Legionella pneumophila*—the causative agent of Legionnaires' disease (see Supplementary material). The placement of the His-tag demonstrates that the tags location greatly influences expression levels. In addition, we show a positive effect when reducing the length of the extensions at the N- and C-termini that originate from the Gateway cloning system. We have also fine-tuned the induction method in order to optimize protein expression; the auto-induction system presented here was superior to IPTG induction for successful overproduction of membrane proteins. We describe a flow-scheme for the simple and rapid successful overproduction of membranes protein, which may be easily adopted by both large 'structural genomics consortia' as well as more modest laboratories devoid of robotics. By using three constructs together with one *E. coli* strain and one growth and induction protocol, all 24 proteins of our bench-mark set are produced to levels exceeding 2 mg/l.

Materials and methods

Chemicals, bacterial strains and media

Chemicals were obtained from Sigma Aldrich, Merck and Amersham Biosciences. All cloning and transformation steps, unless otherwise stated, were carried out using DH5α, and DB3.1 cells, including the preparation of pDNR221 DNA and the destination vectors carrying the *ccdB* gene. Repetitive propagation of *ccdB* carrying vectors in DB3.1 cells was avoided in order to prevent mutations to the *ccdB* gene. For protein production, the *E. coli* strains Rosetta (λDE3), C41 (λDE3), C43 (λDE3), and BL21*(λDE3) were used. For the production of plasmid DNA, cells were grown in LB media (1% (w/v) peptone, 0.5% (w/v) yeast extract and 0.5 (w/v) NaCl), whereas protein overproduction cells were grown in a rich media, 32Y (3.2% (w/v) yeast extract, 0.8% (w/v) peptone and 100 mM NaCl in 10 mM Tris–HCl pH 7.6). In addition, a minimal media containing 1× NPS [24] (25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM $NaHPO_4$, 1 mM $MgSO_4$, 50 mg/l thiamine, 30 mM glucose, 0.2 g/l casamino acids, 0.02 mM $FeSO_4$, 8.2 mg/l $MgCl_2{\cdot}6H_2O$, $MnCl_2{\cdot}4H_2O$, 1.0 mg/l $CaCl_2{\cdot}6H_2O$) was assessed. Other media such as 4× TY (2.0% (w/v) yeast extract, 3.2% (w/v) peptone, 100 mM NaCl) and M9 minimal medium (30 g/l Na_2HPO_4, 15 g/l KH_2PO_4, 5 g/l NH_4Cl, 2 g/l glucose, 500 μg/l thiamine) were also used. The media were supplemented with antibiotics at the following concentrations where necessary: kanamycin, 100 μg/ml; ampicillin, 100 μg/ml; chloramphenicol, 34 μg/ml. For auto-induction of the T7 promoter, 0.5% (v/v) glycerol, 0.05% (w/v) glucose and 0.02–0.2% (w/v) α-lactose was added aseptically after autoclaving.

Molecular biology

DNA manipulations were carried out as described by Sambrook et al. [11]. The GeneElute plasmid purification kit was used for plasmid DNA preparation for sequencing and for site specific recombination reactions (BP and LR reactions). Restriction and modifying enzymes were from New England Biolabs, Fermentas or Invitrogen. DNA sequencing was by MWG Biotech. Selected target genes were amplified by PCR from the genomic DNA from *Legionella pneumophila* sub sp. *pneumophila* strain Philadelphia 1 (ATCC 33152, American Type Culture Collection). The Gateway system [12,13] was used for cloning and transfer of open reading frames and any modifications to the manufactures protocols are noted.

Primers (Invitrogen) contained extensions in order to incorporate the *att*B1 and *att*B2 recombination sites; GGGGACAAGTTTGTACAA AAAAGCAGGCTCT and GGGGACCACTTTGTA CAAGAAAGCTGGGTT for the forward and reverse primer, respectively. Platinum Pfx DNA polymerase (Invitrogen) was used to synthesize PCR products which were then separated by gel electrophoresis and purified (Qiagen kit). Entry clones were created by recombination of the purified PCR products with the entry vector pDONR221 (Invitrogen). Two microlitre of BP clonase buffer (Invitrogen) was mixed with 1 μl (100 ng) pDONR221, 1 μl (50–150 ng) *att*B-PCR product, 5.6 μl TE-buffer pH 8.0 (10 mM Tris–HCl pH 8.0, 1 mM EDTA) and 0.4 μl BP clonase mix (Invitrogen). After incubation overnight at room temperature, the reactions were inactivated at 65 °C for 10 min. Then 1 μl of the BP reaction was transformed into chemically competent *E. coli* DH5α cells and selected on LB-plates containing kanamycin. Plasmid minipreps were made from 3 ml overnight LB cultures supplemented with kanamycin. After verification of the BP products using a TatI restriction assay, the DNA sequence of the entry clones was confirmed by sequencing (MWG Biotech). For plasmid construction please refer to Supplementary material.

SDS–PAGE analysis

Electrophoresis was carried out in 0.75 mm thick gels, using a BioRad MiniGel setup. Coomassie straining was as previously described [14], except that microwave heating was used at each staining step to reduce the total staining and destaining procedure time (30 min) [15].

Small-scale protein production

Small-scale analytical protein production was performed in 0.5 ml cultures in standard 24 well culture plates (TPP). Expression experiments were carried out at 18 °C, 30 °C, 37 °C and 37 °C/18 °C with shaking at a rate of 225 rpm and 25 mm orbital throw. Various expression media were evaluated, as detailed above: LB, 4×TY, 32Y medium, M9 minimal medium and NPS medium. The media were supplemented with ampicillin and chloramphenicol. For auto induction, the media also contained 0.05% (w/v) glucose, 0.5% (v/v) glycerol and 0.02% (w/v) α-lactose. Growth and induction proceeded for 14–18 h. For comparison purposes, cells were induced for 3 h by 0.1 mM IPTG after 12 h of growth.

Dot blot analysis

Cell samples were diluted with an equal volume of 10 M urea. Subsequently, 1–5 μl of this mixture was spotted onto polyvinylideneфluoride membrane together with a standard. The blots were developed using a mouse anti-His6 antibody (BD) and a goat anti-mouse antibody conjugated to alkaline phosphatase (Sigma). Production levels were determined by quantification of the alkaline phosphatase activity using a chromogenic assay with BCIP and NBT (Melford Laboratory). Spot intensity was quantified using ImageJ [16] using the SpotFinder plugin [17]. Duplicate dots were spotted and the correlation coefficient between duplicates was 0.97. Correlation between different colonies was 0.89, and finally the correlation between separate growth experiments was 0.75.

Large-scale preparation

Large-scale membrane protein production was performed in 500 ml of 32Y media or NPS media. The media, in non-baffled 5 L flasks, was supplemented with ampicillin, chloramphenicol and an auto-induction cocktail to give a final concentration of 0.05% (w/v) glucose, 0.5% (v/v) glycerol and 0.02% (w/v) α-lactose. The media was inoculated with 1 ml of overnight LB culture and then incubated at