Journal of Structural Biology 175 (2011) 159–170

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: A benchmarking study

Didier Busso [a,1], Yoav Peleg [b,1], Tatjana Heidebrecht [c], Christophe Romier [a], Yossi Jacobovitch [b], Ada Dantes [b], Loubna Salim [a], Edouard Troesch [a], Anja Schuetz [d], Udo Heinemann [d,e], Gert E. Folkers [f], Arie Geerlof [g,2], Matthias Wilmanns [g], Andrea Polewacz [h], Claudia Quedenau [h], Konrad Büssow [h,i], Rachel Adamson [j], Elena Blagova [k], Julia Walton [k], Jared L. Cartwright [j], Louise E. Bird [l,3], Raymond J. Owens [l,3], Nick S. Berrow [l,4], Keith S. Wilson [k], Joel L. Sussman [b,m], Anastassis Perrakis [c], Patrick H.N. Celie [c,*]

[a] *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de Santé et de Recherche Médicale (Inserm), U964/Centre National deRecherche Scientifique (CNRS), UMR 7104, Université de Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France*
[b] *The Israel Structural Proteomics Center (ISPC), Faculty of Biochemistry, Weizmann Institute of Science, Rehovot 76100, Israel*
[c] *NKI Protein Facility and Division of Biochemistry, The Netherlands Cancer Institute (NKI), Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands*
[d] *Helmholtz Protein Sample Production Facility, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Str. 10, 13125 Berlin, Germany*
[e] *Institut für Biochemie, Freie Universität Berlin, Takustr. 6, 14095 Berlin, Germany*
[f] *Bijvoet Centre for Biomolecular Research, Utrecht University, NMR Spectroscopy Department, Padualaan 8, 3584 CH Utrecht, The Netherlands*
[g] *EMBL Hamburg Outstation, Notkestrasse 85, 22603 Hamburg, Germany*
[h] *Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany*
[i] *Department of Structural Biology, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany*
[j] *Protein Production Laboratory, Technology Facility, Department of Biology, University of York, Wentworth Way, York YO10 5DD, UK*
[k] *YSBL, Department of Chemistry, Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, UK*
[l] *OPPF, Division of Structural Biology, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK*
[m] *Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel*

ARTICLE INFO

Article history:
Available online 5 March 2011

Keywords:
Escherichia coli
Co-expression
Cloning strategies
Enzyme-free cloning
Gateway™
In-Fusion™
LIC
Restriction-free cloning

ABSTRACT

Escherichia coli (*E. coli*) remains the most commonly used host for recombinant protein expression. It is well known that a variety of experimental factors influence the protein production level as well as the solubility profile of over-expressed proteins. This becomes increasingly important for optimizing production of protein complexes using co-expression strategies. In this study, we focus on the effect of the choice of the expression vector system: by standardizing experimental factors including bacterial strain, cultivation temperature and growth medium composition, we compare the effectiveness of expression technologies used by the partners of the Structural Proteomics in Europe 2 (SPINE2-complexes) consortium. Four different protein complexes, including three binary and one ternary complex, all known to be produced in the soluble form in *E. coli*, are used as the benchmark targets. The respective genes were cloned by each partner into their preferred set of vectors. The resulting constructs were then used for comparative co-expression analysis done in parallel and under identical conditions at a single site. Our data show that multiple strategies can be applied for the expression of protein complexes in high yield. While there is no 'silver bullet' approach that was infallible even for this small test set, our observations are useful as a guideline to delineate co-expression strategies for particular protein complexes.

1. Introduction

Multi-protein complexes are often key-regulators in many cellular processes. These complexes can differ in size, varying from only two or three-components to large multimeric-complexes (Charbonnier et al., 2008; Doucet and Hetzer, 2010; Riccio, 2010). Systems biology data have generated many insights into the different pathways and protein networks at the cellular level (Charbonnier et al., 2008). Within the past decade, results from both *in vivo* and *in vitro* studies have illustrated the importance

* Corresponding author. Fax: +31 20 5121954.
E-mail address: p.celie@nki.nl (P.H.N. Celie).
[1] These authors contributed equally to this work and should be considered co-first authors.
[2] Present address: Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, Germany.
[3] Present address: OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxford OX11 0FA, UK.
[4] Present address: Protein Expression Unit, Institute for Research in Biomedicine, Parc Cientific de Barcelona, C/Josep Samitier, 1–5, 08028 Barcelona, Spain.

1047-8477/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2011.03.004

of analyzing the composition and mechanisms of protein assembly to unravel complex biological processes. To obtain the protein assemblies which are the subject of biochemical, biophysical and structural analyses necessary to achieve such mechanistic insight, one can isolate endogenous complexes, either by *in vitro* reconstitution from individually expressed protein components, or by heterologous expression of all components in the same host cell. A large effort has been made in technological developments allowing co-expression of recombinant proteins in both prokaryotes and eukaryotes. Expression in eukaryotic cells, such as *sf9* insect cells or mammalian cell lines, may be favored because of post-translational modifications that are essential for protein function and/or stability and because of the presence of particular chaperone systems that may improve protein folding. An example is the expression of a 400 kDa heterohexameric subcomplex of human TFIID containing two copies of each of the three TAF proteins, which was successfully expressed in insect cells using the baculovirus expression system (Fitzgerald et al., 2006). Despite the advantages of the eukaryotic systems, *Escherichia coli* remains the primary system of choice for expressing protein complexes (Bieniossek et al., 2009; Perrakis and Romier, 2008; Romier et al., 2006; Tan et al., 2005; Tolia and Joshua-Tor, 2006). Expression in *E. coli* has the benefit of obtaining large quantities at low cost and at short time, for either individual proteins or protein complexes. In addition, integration of both DNA-cloning and protein expression technologies in well-established high-throughput platforms allow parallel testing of multiple protein variants, as well as different strains and/or culture conditions (Berrow et al., 2006; Vijayyachandran et al., this issue). Moreover, the absence of particular post-translational modifications (e.g. glycosylation) within the *E. coli* system is sometimes an advantage for X-ray crystallography studies, where non-homogenous protein preparations are likely to have an adverse effect on the success-rate of finding crystallization hits. Co-expression in *E. coli* is a strategy that can often present advantages over *in vitro* reconstitution or re-folding of the individually expressed partners, allowing proper folding of the protein partners and formation of a soluble complex *in vivo*, thus overcoming solubility problems of the individually expressed components (Li et al., 1997; Romier et al., 2006).

Many factors can influence the expression of proteins in *E. coli*, including the bacterial strain used for expression, expression system, growth medium and temperature of induction (Berrow et al., 2006; Graslund et al., 2008). In addition to the factors that may influence expression of individual proteins, the experimental results of protein co-expression are affected by several specific factors. These include the choice of partner, position of the affinity-tag (C- or N-terminal) used for co-purification (Diebold et al., this issue; Fribourg et al., 2001; Romier et al., 2006) and the selection of the protein domains used in the co-expression study (Fribourg et al., 2001).

The selected strategy used for protein co-expression may also have an additional impact. Co-expression can be conducted using either single or multiple constructs. In the case of a single plasmid, this can be either poly-cistronic, (i.e. having a single promoter for multiple genes that are transcribed in the same mRNA) or, alternatively, the plasmid can contain multiple genes, each controlled by a separate promoter (transcribed each in a distinct mRNA). When two or more constructs are co-transformed into a single cell, each vector should at least comprise a different antibiotic selection marker (Perrakis and Romier, 2008; Zeng et al., 2010) and each vector could harbor compatible (i.e. distinct) or incompatible (i.e. similar) replicons (Johnston et al., 2000; Perrakis and Romier, 2008; Velappan et al., 2007; Yang et al., 2001).

In the present study, we conducted a systematic benchmarking study exploring the effect of different co-expression strategies, as reflected by the choice of expression vectors, on the production and solubility of different complexes. Within the SPINE2-complexes consortium, each partner has its own set of preferred, often customized, vectors that are suited for protein co-expression. Therefore, we aimed to perform a systematic analysis of different vectors, which were commonly used at eight SPINE2-complexes consortium partner sites (Division of Biochemistry, The Netherlands Cancer Institute (NKI), Amsterdam; Helmholz Protein Sample Production Facility (PSPF), Berlin/Braunschweig; Structural Biology Unit, EMBL-Hamburg Outstation, Hamburg; Oxford Protein Production Facility (OPPF), Division of Structural Biology, Oxford; The Israel Structural Proteomics Center (ISPC), Weizmann Institute of Science, Rehovot; Integrative Structural Biology Program, Institute of Genetic, Molecular and Cellular biology (IGBMC), Strasbourg; NMR spectroscopy research group, Bijvoet center for Biomolecular Research, Utrecht and the Protein Production Laboratory, Department of Biology, University of York York). To compare the different co-expression systems, four protein complexes (three binary and one ternary) were selected, of which only one protein per complex contained an N-terminal Histidine tag for purification purposes (see Section 2.). Most expression vectors tested were based on the T7 promoter system for transcriptional regulation in combination with *E. coli* strain harboring the DE3 prophage (Studier et al., 1990). DNA cloning into the different expression vectors was performed at each individual partner site and protein co-expression was subsequently performed at one site (NKI, Amsterdam), under standardized experimental parameters and to minimize random variations. Our data show that multiple strategies can be applied for expression of complexes in high yield; there does not appear to be a preferred strategy yielding systematically optimal results for all four tested complexes. This emphasizes the importance of efficient high-throughput expression and purification methods also as the means to explore different strategies for a given problem to efficiently choose the best approach by trial and error.

2. Materials and methods

2.1. Selected complexes

Four protein complexes were selected for benchmarking the different co-expression vectors: (1) human Geminin:Cdt1, a 76.6 kDa trimeric complex with 2:1 stroichiometry (De Marco et al., 2009); (2) human TFIIEα:TFIIEβ, a 82.5 kDa dimeric complex (Jawhari et al., 2006); (3) viral influenza Importin-α5:PB2, a 58.6 kDa dimeric complex (Tarendeau et al., 2007); and (4) human NFYC:NFYB:NFYA, a 32.3 kDa trimeric complex (Romier et al., 2006). Details of the proteins and the selected domains thereof are presented in Table 1.

Original DNA constructs containing the respective genes were gathered and amplified at the NKI and subsequently distributed among SPINE2-complexes partners to be used as a template for re-cloning into the expression vectors of choice. All vectors used by each partner are described below and schematic diagrams with the details for all vectors are presented in Fig. 1.

The co-expression trials were categorized in four groups depending on the expression strategy. Group 1 and 2 comprises those trials for which proteins are expressed from multiple plasmids with either incompatible or compatible origin of replications, respectively. Expression trials from constructs that contain multiple genes under control of a single promoter (poly-cistronic transcript) or under control of separate promoters comprise groups 3 and 4, respectively (Table 2). In some expression trials of the ternary his-NFYC:NFYB:NFYA complex, a combination of strategies is used, e.g. when two plasmids with compatible origin of replications are used and one of these contains two genes for bi-cistronic expression (strategy 2 and 3 combined).