Journal of Structural Biology 175 (2011) 198–208

Contents lists available at ScienceDirect

Journal of Structural Biology

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

Robots, pipelines, polyproteins: Enabling multiprotein expression in prokaryotic and eukaryotic cells

Lakshmi Sumitra Vijayachandran [a], Cristina Viola [a], Frederic Garzoni [a], Simon Trowitzsch [a], Christoph Bieniossek [a], Maxime Chaillet [a], Christiane Schaffitzel [a], Didier Busso [b,c,d,e], Christophe Romier [b,c,d,e], Arnaud Poterszman [b,c,d,e], Timothy J. Richmond [f], Imre Berger [a,*]

[a] European Molecular Biology Laboratory (EMBL), Grenoble Outstation, and Unit of Virus Host Cell Interactions UVHCI, UJF-CNRS-EMBL Unite Mixte International UMI 3265, rue Jules Horowitz, 38042 Grenoble Cedex 9, France
[b] IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire), 1 rue Laurent Fries, BP10142, Illkirch F-67400, France
[c] Inserm, U964, Illkirch F-67400, France
[d] CNRS, UMR7104, Illkirch 67400, France
[e] Université de Strasbourg, Strasbourg F-67000, France
[f] Institute for Molecular Biology and Biophysics IMB, Swiss Federal Institute of Technology ETH Hönggerberg, Schafmattstrasse 20, CH 8093 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 15 December 2010
Received in revised form 11 March 2011
Accepted 11 March 2011
Available online 17 March 2011

Keywords:
ACEMBL
Automation
Baculovirus
BEVS
High-throughput
Insect cells
Mammalian host
MultiBac
pET-MCN
Polyprotein
Recombineering
Robotics
Structural biology

ABSTRACT

Multiprotein complexes catalyze vital biological functions in the cell. A paramount objective of the SPINE2 project was to address the structural molecular biology of these multiprotein complexes, by enlisting and developing enabling technologies for their study. An emerging key prerequisite for studying complex biological specimens is their recombinant overproduction. Novel reagents and streamlined protocols for rapidly assembling co-expression constructs for this purpose have been designed and validated. The high-throughput pipeline implemented at IGBMC Strasbourg and the ACEMBL platform at the EMBL Grenoble utilize recombinant overexpression systems for heterologous expression of proteins and their complexes. Extension of the ACEMBL platform technology to include eukaryotic hosts such as insect and mammalian cells has been achieved. Efficient production of large multicomponent protein complexes for structural studies using the baculovirus/insect cell system can be hampered by a stoichiometric imbalance of the subunits produced. A polyprotein strategy has been developed to overcome this bottleneck and has been successfully implemented in our MultiBac baculovirus expression system for producing multiprotein complexes.

1. Introduction

The structural proteomics initiative SPINE (Structural Proteomics IN Europe) was a highly successful European project aimed at production and structural characterization of mostly single proteins with important roles in human health. By building on these successes, SPINE2 (full name SPINE2-COMPLEXES) addressed more challenging biological systems which consists of many interlocking protein subunits assembled in complexes to exert their biological function. Multiprotein complexes are emerging as cornerstones of biological activity in the cell, and deciphering their structure and function is imperative for advancing research in health and disease (e.g. Alberts, 1998; Nie et al., 2009a). Many essential

Abbreviations: BEVS, baculovirus expression vector system; CFP, cyan florescent protein; CMV, cytomegalovirus; dpa, day of proliferation arrest; dsRed, red fluorescent protein; *E. coli*, *Escherichia coli*; EGFP, enhanced green fluorescent protein; EM, electron microscopy; FRET, fluorescence resonance energy transfer; HT, high throughput; kb, kilo base; kDa, kilo dalton; MOI, multiplicity of infection; NMR, nuclear magnetic resonance (spectroscopy); ORF, open reading frame; Ori, origin of replication; p10, p10 baculoviral late promoter; polh, polyhedrin baculoviral very late promoter; R6Kγ, bacteriophage R6Kγ; SDS–PAGE, sodium dodecylsulfate–polyacryamide gel electrophoresis; Sf9, Sf21, *Spodoptera frugiperda* cell lines 9 or 21, respectively; SLIC, sequence and ligation independent cloning; tcs, TEV protease cleavage site; TEV, tobacco etch virus, resp. a protease (N1A) from this virus; YFP, yellow fluorescent protein.

* Corresponding author. Address: European Molecular Biology Laboratory (EMBL Grenoble), BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France. Fax: +33 (0)476207199.
E-mail address: iberger@embl.fr (I. Berger).

1047-8477/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2011.03.007

protein complexes, particularly in human cells are comparatively hard to come by, thus complicating their study. Low endogenous levels and sample heterogeneity often impede purification from native sources in the quality and quantity required for structural studies, especially by X-ray diffraction methods requiring highly purified material that can crystallize. Recombinant techniques can overcome several of the bottlenecks encountered, and can further provide the means to modify complexes at the gene level resulting in alteration of the protein complex subunits to meet the high quality requirements for structural analysis. Consequently, SPINE2-COMPLEXES placed considerable emphasis on developing and implementing technologies, protocols and reagents to facilitate protein complex production for structural studies, in prokaryotic and eukaryotic host organisms. In this contribution, we will review the high throughput platforms at Strasbourg and Grenoble as well as the practical considerations concerning setting up and running a eukaryotic expression facility. New approaches for eukaryotic protein production, including the benefits of polyprotein design, will also be discussed.

Owing to a large part to structural genomics efforts such as the SPINE project, affordable methods and equipment have been developed to automate molecular cloning of expression constructs, recombinant expression screening and purification. There are obvious advantages of automation in the molecular biology laboratory. Procedures can be carried out in parallel and scaled-up accordingly to process large amounts of samples at reasonable cost. Automation puts constraints on the robustness of protocols to be implemented which are certainly more stringent than regular manual laboratory intervention requires. In our experience, a good protocol yielding reproducible results, be it for cloning, transformation or expression screening, by no means guarantees that it can be scripted without further ado into a robotics routine. Rather, seemingly robust protocols for manual experimentation often have to be optimized further to work robustly in a parallelized manner on a robot. This is beneficial to the laboratory implementing automation in several ways. Not all experiments will be carried out on a robot even in the most well-equipped laboratory, and the general success rate of manual experimentation, at least in our laborato-

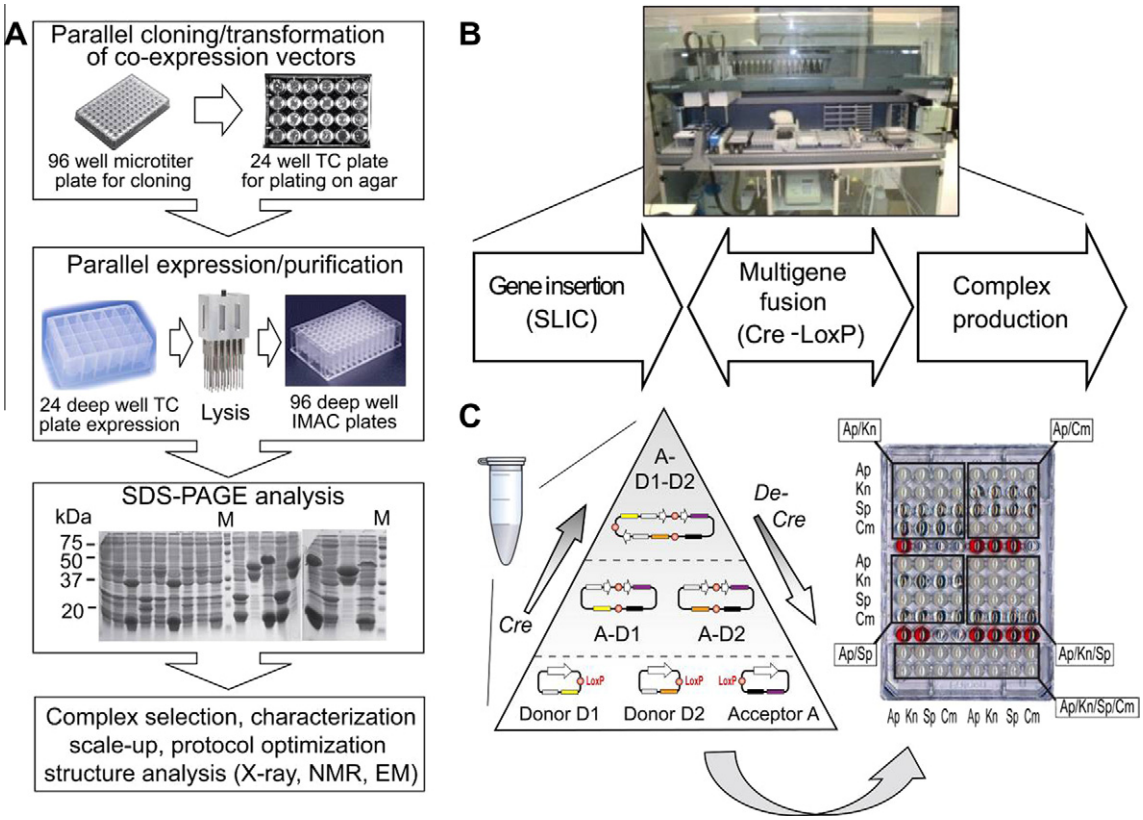

Fig.1. SPINE2: automation for HT protein complex production. (A) The high throughput pipeline for screening co-expression constructs at IGBMC Strasbourg is shown in a schematic representation. Relying on semi-automated procedures, co-expression constructs are prepared in a parallel fashion by the cloning method of choice. This can be carried out in small volumes for example on a 96 well micro-titer plate. Constructs are transformed and plated on agar provided on 24 well tissue culture (TC) plates (top). Parallel expression of single colonies is performed in 2–4 ml miniculture volumes in deep well TC plates screening a variety of parameters (media, temperature, induction, etc.). Cells are pelleted and lysed by a multitip sonicator. Overproduced soluble protein complexes are purified from the cleared lysate by metal affinity purification on deep well plates using automated procedures (middle). The complexes retained on the affinity resin are analyzed by SDS–PAGE (below). Promising complex candidates are prioritized for scale-up, biochemical and biophysical characterization and structure determination (bottom). (B) The multigene cloning and expression pipeline ACEMBL at the EMBL Grenoble relies on fully automated protocols scripted into robotic routines that are implemented on a liquid handling workstation (a Tecan Evo II LHW is shown). One or several genes are inserted in specifically designed synthetic plasmids called Donors and Acceptors (Table 1) by sequence and ligation independent cloning (SLIC). These plasmids are next concatamerized by Cre–LoxP fusion into multigene constructs for protein complex production. The combination of SLIC and Cre–LoxP reactions for multigene vector construction is termed tandem recombineering (TR). The Cre-reaction is an equilibrium reaction and can also be exploited for disassembling constructs. (C) The Cre–LoxP fusion reaction is detailed in a schematic representation (left). In a single Eppendorf tube, Donors and one Acceptor (Table 1), each containing a LoxP sequence, are combined with Cre recombinase. Cre–LoxP fusion generates in this tube all possible combinations in an equilibrium reaction (arrows marked by *Cre* and *De-Cre*). After transformation and plating, desired Donor–Acceptor fusions are selected by the specific antibiotic combination conferred by the resistance markers on the plasmids. Identification of the desired recombinants can be carried out on a 96 well microtiter plate (right). D is Donor, A is Acceptor. A–D and A–D–D denote Donor–Acceptor fusions. LoxP sites are shown as colored balls. Ap, Cm, Kn, Sp stand for antibiotics ampicillin, chloramphenicol, kanamycin and spectinomycin. Antibiotics combinations are boxed. Red dye is added in wells on the microtiter plate for orientation. ACEMBL system Donors and Acceptors are listed in Table 1. C is adapted from Bieniossek et al. (2009) with permission of the publisher.