Multiprotein Expression Strategy for Structural Biology of Eukaryotic Complexes

Daniel J. Fitzgerald,[1] Christiane Schaffitzel,[1] Philipp Berger,[2,4] Ralf Wellinger,[3] Christoph Bieniossek,[1] Timothy J. Richmond,[1] and Imre Berger[1,*]

[1] ETH Zürich, Institut für Molekularbiologie und Biophysik
[2] ETH Zürich, Institut für Zellbiologie, ETH-Hönggerberg, CH-8093 Zürich, Switzerland
[3] Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), 41092 Sevilla, Spain
[4] Present address: Laboratory of Biomolecular Research, Molecular Cell Biology, Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland.
*Correspondence: iberger@mol.biol.ethz.ch
DOI 10.1016/j.str.2007.01.016

SUMMARY

The concept of the cell as a collection of multisubunit protein machines is emerging as a cornerstone of modern biology, and molecular-level study of these machines in most cases will require recombinant production. Here, we present and validate a strategy to rapidly produce, permutate, and posttranslationally modify large, eukaryotic multiprotein complexes by using DNA recombination in a process that is fully automatable. Parallel production of 12 protein complex variants within a period of weeks resulted in specimens of sufficient quantity and homogeneity for structural biology applications.

INTRODUCTION

Proteomics approaches including affinity copurification and large-scale, two-hybrid experiments have revealed that multiprotein complexes are the rule rather than the exception in cells (Forler et al., 2003; Rual et al., 2005; Parrish et al., 2006). Molecular-level studies of the structure and function of most multiprotein complexes will require advanced recombinant protein production technologies, as many are not sufficiently homogenous or abundant to be purified from source. The majority of X-ray crystal structures of multiprotein complexes solved to date have in common that the specimens were purified from source in sufficient quantity and uniformity to yield well-diffracting crystals (Dutta and Berman, 2005). Crystal structures of multiprotein complexes that were recombinantly produced are few in number, while structures of recombinantly produced, large eukaryotic complexes are almost nonexistent. Likewise, for single-particle electron microscopy, resolution depends on preparing a uniform sample. This is technically a considerable challenge for scarce, heterogeneous multisubunit specimens, which would describe most of the eukaryotic proteome. Therefore, presently, a great fraction of the proteome is largely inaccessible for structural studies.

For eukaryotic multiprotein complex production, we recently introduced MultiBac, a system for protein coexpression using multigene baculoviral vectors as a superior alternative to coinfection by single-gene baculoviruses (Berger et al., 2004; Roy, 2004). For production in *E. coli*, the Duet system (Tolia and Joshua-Tor, 2006) and pST44 polycistronic vector system (Tan et al., 2005) exist. These three systems have in common a reliance on conventional cloning strategies utilizing restriction digestion, electrophoretic separation of DNA fragments, and DNA ligation. Already for single genes, this process is time and labor intensive and is essentially refractory to high-throughput methods. An alternative is combining DNA fragments by recombinases, and several systems for single-gene expression constructs have emerged (Liu et al., 1998; Muyrers et al., 2004; Benoit et al., 2006). For multigene expression vector generation, we describe here a general, simple, and rapid recombination-based approach using baculovirus that yields samples of sufficient quantity and uniformity for structural studies. The overall principle is readily adaptable to other expression systems.

RESULTS AND DISCUSSION

The strategy underlying the presented approach makes use of expression plasmids classified as either "Acceptors" and "Donors," which can be fused in vitro to yield multigene constructs (Figure 1A). A *loxP* sequence present on all vectors allows for in vitro fusing of Donors with Acceptors by Cre recombinase (Figure 1A, left). Acceptor plasmids contain standard replication origins, while Donor plasmids carry a R6Kγ conditional origin, rendering their propagation dependent on hosts expressing the *pir* gene (Experimental Procedures). Transforming in vitro fusion reactions of Donors and Acceptors into *pir⁻* bacterial strains eliminates nonfused Donors, while the unique antibiotic-resistance markers present on each Donor and Acceptor allow for selection of the desired Donor-Acceptor fusions. Donor-Acceptor fusions enter the MultiBac genome by Tn7 transposition in DH10MultiBac cells (Berger et al., 2004). A separate pathway to funnel multigene cassettes into MultiBac DNA is shown in Figure 1A (right). Individual Donors or Donor-Donor
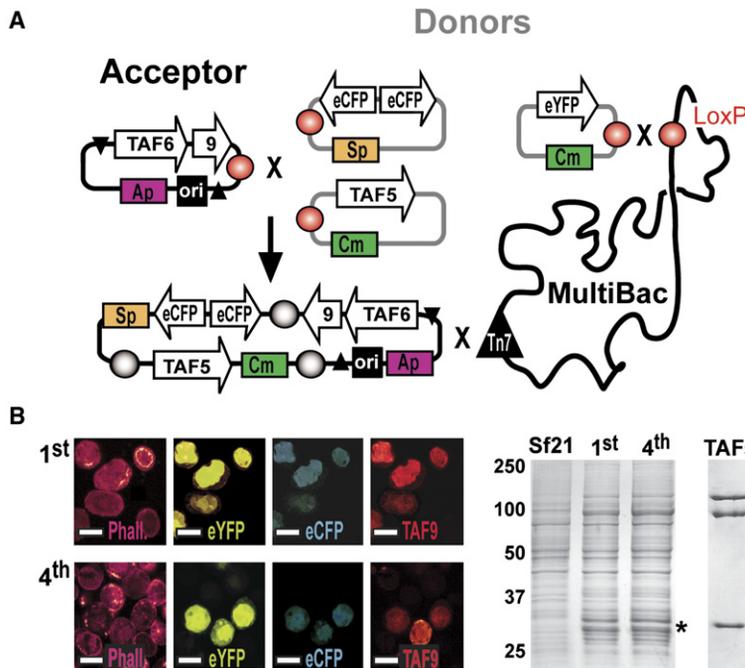
Figure 1. Rapid Assembly of Multiprotein Expression Vectors

(A) In vitro recombination (left). An Acceptor plasmid (far left, black line) and two Donor plasmids (center, gray line) are fused to an Acceptor-Donor-Donor triple fusion plasmid by a single Cre-*loxP* reaction. Recombinant genes encode for human TBP-associated factors TAF5, TAF6, and TAF9 and for fluorescent protein eCFP (two copies). Antibiotic markers (colored boxes), expression cassettes (open arrows), *loxP* (red circles) and Tn7 transposition sequences (black triangles), and nonconditional origins of replication (boxed) are indicated. In vivo recombination (right). Generation of a six-gene baculoviral expression vector is shown. Acceptor-Donor fusions access MultiBac DNA through Tn7 transposition (bottom center). Single Donors enter the *loxP* site through Cre-*loxP* recombination (top, right) in DH10MultiBac^Cre cells. Here, a Donor plasmid carrying eYFP first entered via *loxP*, and the other five genes on an Acceptor-Donor-Donor triple fusion were then added via Tn7.

(B) Viral stability. Fluorescence micrographs of cells infected with first- (top-left row) or fourth-generation (bottom-left row) virus from the MultiBac expression vector shown in (A). Micrographs were taken from the same field of Sf21 cells. Phalloidin stains actin in infected and uninfected cells (far left). Specific fluorescence signals for eYFP, eCFP, and TAF9 were detected concurrently from infected cells. Scale bars are 20 μm. Right: Coomassie-stained SDS-PAGE of uninfected Sf21 cells, cells infected with first- or fourth-generation virus, and purified TAFs. Molecular weights are indicated. Both eCFP and eYFP migrate in the region of the asterisk.

fusions are prepared in *pir*[+] cells, and these vectors enter MultiBac DNA by Cre-*loxP* recombination in DH10MultiBac^Cre cells.

We first constructed a six-gene baculoviral expression vector containing the human TBP-associated factors TAF5, TAF6, and TAF9 and the fluorescent proteins eCFP (two copies) and eYFP (Figure 1A). TAF5-Donor was generated by "seamless cloning" with BD In-Fusion recombinase (Benoit et al., 2006; Fitzgerald et al., 2006) and was fused with eCFP-eCFP-Donor and TAF6-TAF9-Acceptor, resulting in a triple fusion. eYFP-Donor entered MultiBac DNA by Cre-*loxP* recombination, and thereafter the triple fusion entered eYFP-MultiBac DNA via Tn7 transposition. Composite bacmid was then selected for initial virus generation. The procedure, starting with encoding DNAs to harvesting six-gene baculovirus, took 2 weeks.

Successful protein production with baculovirus depends on virus stability. Deletion genotypes relying on coinfection with wild-type specimens for propagation accumulate over generations (Simon et al., 2006), while selective pressure drives specific deletions of recombinant inserts (Pjilman et al., 2004). The numerous identical expression elements present in MultiBac viruses are potential homologous recombination sites and are thus of concern (Roy, 2004). To address this, we challenged the virus shown in Figure 1A by serial passaging four times. Only budded virus was harvested (Braunagel et al., 1998), and we maintained a multiplicity of infection of ~0.1 throughout. Heterologous protein expression of first- and fourth-generation viruses was assayed at the single-cell level by fluorescence microscopy. Both viruses produced all proteins assayed in >90% of infected cells (Figure 1B, left; Experimental Procedures). SDS-PAGE analysis further revealed constant expression levels of fluorescent proteins and TAFs (Figure 1B, right). Taken together, the data provide compelling evidence for virus stability in our experimental setup and validate this approach for automated applications in which lengthy analytical steps during virus generation are to be avoided.

TAFs 5, 6, and 9 are components of human general transcription factor TFIID (Thomas and Chiang, 2006). However, their stoichiometry within TFIID is not clearly established. Purification of TAFs 5, 6, and 9 (Experimental Procedures) resulted in a yield of 4 mg purified complex per liter cell culture. Characterization of this complex by size-exclusion chromatography (not shown) and analytical ultracentrifugation revealed a single 15S (402 kDa) species (Figure 2A), which is consistent with two copies each of TAFs 5, 6, and 9 (predicted 394 kDa). Electron micrographs showed uniform particles and thus confirmed sample quality (Figure 2B). Class averages revealed apparent two-fold symmetry consistent with a hexameric $(\alpha\beta\gamma)_2$ arrangement.

For high-resolution structural and functional studies, it is often necessary to produce numerous sequence variants of wild-type protein(s) by using bioinformatic and biochemical data. Chromatin-remodeling factors are multiprotein complexes that hydrolyze ATP to modify chromosome structure (Saha et al., 2006). In Figure 3A, a series
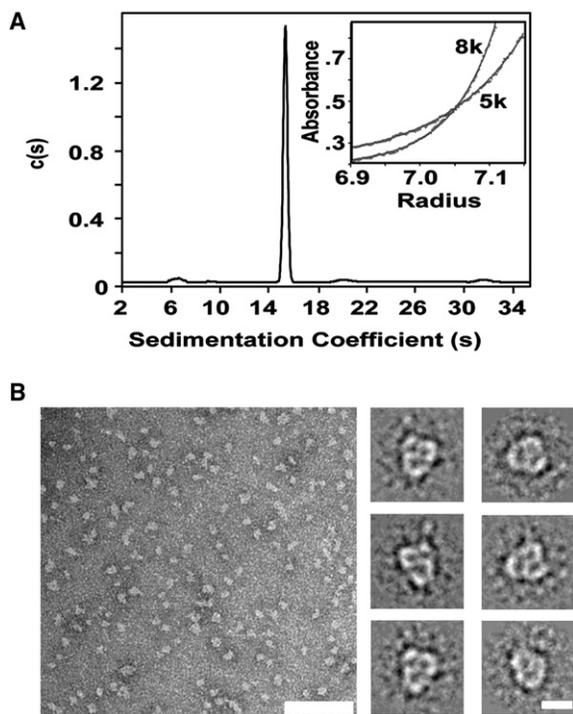
Figure 2. Characterization of Human TAF5,6,9 Transcription Factor Complex

(A) Results from analytical ultracentrifugation experiments of purified TAF complex composed of TBP-associated factors TAF5, TAF6, and TAF9 expressed from the baculovirus shown in Figure 1. Sedimentation velocity experiments revealed a single species migrating at 15S. The inset depicts representative curves from sedimentation equilibrium experiments at 5,000 rpm (5k) and 8,000 rpm (8k) together with the respective single-component fit.

(B) Structure of heterohexameric TAF5,6,9 complex analyzed by negative-stain electron microscopy. Micrographs revealed particles of uniform size, ~13 nm in diameter (left). Class averages (right) show a rounded tripartite shape with apparent two-fold symmetry. Scale bars are 100 nm in the overview and 10 nm in the class averages.

of chromatin-remodeling complexes and variants of the wild-type complexes were created by PCR amplification and funneling the products by seamless cloning into two Donors and one Acceptor. Cre recombination was used to generate 8 three-gene and 12 two-gene permutations in a single-day parallel transformation experiment. In contrast, traditional subcloning of all of these permutations would be expected to require months. Expression, purification, and crystallization screening of ten permutations resulted in initial protein crystals (Figure 3A, bottom), and a subsequent round of revised expression experiments based on this construct resulted in crystals suitable for X-ray structure determination.

MALDI mass spectroscopy revealed that a wild-type three-member chromatin-remodeling complex from Figure 3A carried a single phosphate that could be removed in vitro with λ-Ppase, as determined by Pro-Q phosphostaining (Figure 3B) and MALDI analysis (not shown). We tested whether coexpression of λ-PPase with the complex would also remove the phosphate group. In vivo dephos-
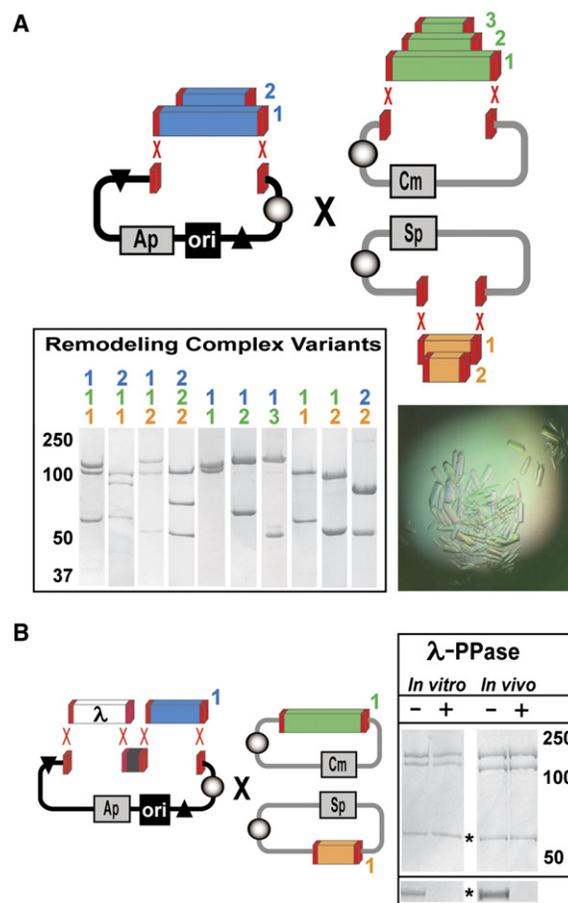

Figure 3. Rapid Multiprotein Permutation by Recombination

(A) In vitro shuffling of multigene expression cassettes. Genes encoding for chromatin-remodeling complex variants (blue, green, and orange) were PCR amplified and separately inserted into Donors and Acceptors by In-Fusion recombination (red crosses). Different gene combinations were then generated by in vitro Cre fusion (black cross). SDS gel segments (bottom left) of purified complex variants are defined by numbers above each lane. Molecular weights are indicated. One of ten purified complexes yielded protein crystals (bottom right), which, after a subsequent round of revised expression experiments based on the crystallizable construct, diffracted to better than 4 Å resolution.

(B) Coexpression of a posttranslation modification enzyme. Monophosphorylated purified remodeling complex (lane 1 in [A]) was dephosphorylated with λ-PPase in vitro or by coexpression of λ-PPase in vivo. Acceptor-Donor-Donor fusions were created as in (A) to cointegrate λ-PPase in an Acceptor as schematically indicated (left). Coomassie-stained SDS gel lanes of purified complex after in vitro or in vivo dephosphorylation (right). Molecular weights are indicated. The region of the gels marked with an asterisk is shown (bottom right) after Pro-Q phosphoprotein gel staining.

phorylation was quantitative (Figure 3B). We anticipate that shuttling of a series of modifiying enzymes (e.g., kinases, glycosylases) or, by the same token, putative stable interaction partners in such a framework would likewise provide a rapid in vivo screening assay in a eukaryotic setting.

Our strategy combines homologous (In-Fusion) and site-specific (Cre-*loxP*) recombination in tandem for

rapid and flexible generation of multigene coexpression vectors. Revised expression experiments can be easily accommodated in this system given the pyramidal assembly of multigene expression vectors from single or double gene progenitors. All steps in the process, from cloning to protein expression, could be fully automated in a high-throughput setting in a format that generates structural biology grade samples. The general strategy can easily be adapted to existing prokaryotic, yeast, or mammalian expression vectors by providing the appropriate promoters and terminators, resistance markers, and replicons.

EXPERIMENTAL PROCEDURES

Acceptor and Donor Plasmids and Derivatives

Acceptor plasmids pFL and pKL are derivatives of pFBDM (Berger et al., 2004) with an additional *loxP* sequence. pFL contains a high-copy number ColE1 replicon and an ampicillin (Ap)-resistance marker, and pKL contains a pBR322-derived, low-copy number origin of replication and a kanamycin-resistance marker. Donor plasmids are pUCDM (Berger et al., 2004) and pSPL. pSPL is identical to pUCDM, except for the spectinomycin (Sp)-resistance marker (chloramphenicol in pUCDM). All vectors contain two expression cassettes and a multiplication module that allows for adding additional cassettes (Berger et al., 2004). All vector sequences and further details about expression cassettes, recombination elements, and restriction maps can be obtained from our website (http://www.mol.biol.ethz.ch/groups/berger_group/MultiBac). The components of the system presented can be obtained from the authors. Vector and insert PCR amplifications were carried out with ultra-high-fidelity Phusion polymerase (Finzymes). BD In-Fusion reactions were performed by following the procedures described (Benoit et al., 2006; Fitzgerald et al., 2006). In vitro Cre-*loxP* reactions were carried out as recommended by the manufacturer (NEB).

Fluorescence Microscopy

Cells infected with composite MultiBac viruses were assayed by fluorescence spectroscopy for phalloidin, eCFP, and eYFP as described (Berger et al., 2004), with the exception that Cy5-conjugated phalloidin (Molecular Probes) was used. Flourescence analyses were performed with a Leica TCS SP confocal microscope. TAF9 was detected with a polyclonal goat anti-TAF32 antibody (Santa Cruz Biotechnology) and a Cy3-conjugated donkey anti-goat antibody (Jackson Immuno-Research, Inc.).

TAF5,6,9 Complex Purification and Characterization

TAF heterohexamer was purified by following standard protocols by using TALON (BD Biosciences) and calmodulin affinity resin (Stratagene), followed by size-exclusion chromatography on a Superose 6 column (Amersham Biosciences). TAF heterohexamer was applied to a Superdex S200 gel-permeation column (Pharmacia) calibrated by using a high-molecular weight gel filtration kit (Amersham Biosciences).

Analytical ultracentrifugation studies were carried out in a Beckman Optima XL-I analytical ultracentrifuge at 4°C. Program Sednterp v1.08 was used to calculate partial specific volumes and solution density. Interference data from sedimentation equilibrium experiments were analyzed with the Ultrascan v6 program. Sedimentation velocity experiments were carried out at 30,000 rpm, and interference data were analyzed by using continuous c(s) distribution in Sedfit v8.52b (Fitzgerald et al., 2004).

Electron Microscopy

Protein complex (50 μg/ml) was crosslinked with 0.1% formaldehyde and adsorbed to glow-discharged carbon film for 30 s, followed by staining with a 0.75% (w/v) uranyl formate solution for 30 s (Ohi et al., 2004). Images of the complex were recorded with a Fei Morgagni 268(D) electron microscope (80 kV, magnification 80,000×). Images were taken at 1–4 μm defocus with a SIS Megaview III CCD camera (1300 × 1024 pixels). IMAGIC-5 (Image Science) was used for image processing of 4,000 individual molecular images of the complex. After a "reference-free" alignment procedure, images were subjected to multivariate statistical analysis and classification (van Heel and Frank, 1981). The resulting class averages were then used as reference images in subsequent rounds of alignment.

REFERENCES

Benoit, R.M., Wilhelm, R.N., Scherer-Becker, D., and Ostermeier, C. (2006). An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids. Protein Expr. Purif. *45*, 66–71.

Braunagel, S.C., Parr, R., Belyavskyi, M., and Summers, M.D. (1998). *Autographa californica* nucleopolyhedrovirus infection results in Sf9 cell cycle arrest at G2/M phase. Virology *244*, 195–211.

Berger, I., Fitzgerald, D.J., and Richmond, T.J. (2004). Baculovirus expression system for heterologous multiprotein complexes. Nat. Biotechnol. *22*, 1583–1587.

Dutta, S., and Berman, H.M. (2005). Large macromolecular complexes in the Protein Data Bank: a status report. Structure *13*, 381–388.

Fitzgerald, D.J., DeLuca, C., Berger, I., Gaillard, H., Sigrist, R., Schimmele, K., and Richmond, T.J. (2004). Reaction cycle of the yeast Isw2 chromatin remodeling complex. EMBO J. *23*, 3836–3843.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods *3*, 1021–1032.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izzaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat. Biotechnol. *21*, 89–92.

Liu, Q., Li, M.Z., Leibham, D., Cortez, D., and Elledge, S. (1998). The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. Curr. Biol. *8*, 1300–1309.

Muyrers, J.P., Zhang, Y., Benes, V., Testa, G., Rientjes, J.M., and Stewart, F. (2004). ET recombination: DNA engineering using homologous recombination in *E. coli*. Methods Mol. Biol. *256*, 123–139.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative staining and image classification - powerful tools in modern electron microscopy. Biol. Proced. Online *6*, 23–34.

Parrish, J.R., Gulyas, K.D., and Finley, R.L., Jr. (2006). Yeast two-hybrid contributions to interactome mapping. Curr. Opin. Biotechnol. *17*, 387–393.

Pjilman, G.B., de Vrij, J., van den End, F.J., Vlak, J.M., and Martens, D.E. (2004). Evaluation of baculovirus expression vectors with

enhanced stability in continuous cascaded insect-cell bioreactors. Biotechnol. Bioeng. *87*, 743–753.

Roy, P. (2004). Baculovirus solves a complex problem. Nat. Biotechnol. *22*, 1527–1528.

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. Nature *437*, 1173–1178.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodeling: the industrial revolution of DNA around histones. Nat. Rev. Mol. Cell Biol. *7*, 437–447.

Simon, O., Williams, T., Caballero, P., and Lopez-Ferber, M. (2006). Dynamics of deletion genotypes in an experimental insect virus population. Proc. Biol. Sci. *7*, 783–790.

Tan, S., Kern, R.C., and Selleck, W. (2005). The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. Protein Expr. Purif. *40*, 385–395.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Tolia, N.H., and Joshua-Tor, L. (2006). Strategies for protein coexpression in *Escherichia coli*. Nat. Methods *3*, 55–64.

van Heel, M., and Frank, J. (1981). Use of multivariate statistics in analysing the images of biological macromolecules. Ultramicroscopy *6*, 187–194.