Baculovirus expression: old dog, new tricks

Imre Berger & Arnaud Poterszman

To cite this article: Imre Berger & Arnaud Poterszman (2015) Baculovirus expression: old dog, new tricks, Bioengineered, 6:6, 316-322, DOI: 10.1080/21655979.2015.1104433

To link to this article: http://dx.doi.org/10.1080/21655979.2015.1104433

© 2015 The Author(s). Published with license by Taylor & Francis Group, LLC

Accepted author version posted online: 21 Oct 2015.

Submit your article to this journal

Article views: 168

View related articles

View Crossmark data
Baculovirus expression: old dog, new tricks

Imre Berger1,2,* and Arnaud Poterszman3,4,5,6,*

1The European Molecular Biology Laboratory (EMBL); Grenoble Outstation and Unit of Virus Host-Cell Interactions (UVHCI); Université Grenoble Alpes-EMBL-CNRS; UMR 5233; Grenoble, France; 2The School of Biochemistry; University of Bristol; Clifton, United Kingdom; 3Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC); Center for Integrative Biology (CBI); Department of Integrated Structural Biology; Illkirch; France; 4Center National de la Recherche Scientifique (CNRS) UMR 7104; Illkirch, France; 5Institut National de la Santé et de la Recherche Médicale (INSERM) U964; Illkirch, France; 6Université de Strasbourg; Strasbourg, France

Since its inception more than 30 years ago, the baculovirus expression vector system (BEVS) has been used prolifically to produce heterologous proteins for research and development. In the cell, a cornerstone of biological activity are multiprotein complexes, catalyzing essential functions. BEVS has been uniquely successful to unlock such complex assemblies for high-resolution structural and functional analysis. Synthetic biology approaches have been implemented to optimize multigene assembly methods, accelerating upstream processes. Specialized baculoviral genomes are being created with functions tailored to enhance production of particular target protein classes. Here we comment on current and emerging developments in the field and their potential to accelerate protein complex research.

Introduction

In eukaryotes, protein complexes composed of many subunits carry out most essential physiological processes, and are intensely researched today. Some complexes are highly abundant in cells, and can be purified with relative ease from endogenous source material in the quality and quantity required for high resolution analysis of their structure and mechanism. Proteasomes, ribosomes, globins and RNA polymerases are impressive examples. However, this appears to the exception rather than the rule. Our insight into the overwhelming complexity of cellular biology increases rapidly, and it appears that the majority of the protein complexes which function in our body are characterized by low abundance and, moreover, by compositional heterogeneity, which thoroughly complicates or even rules out their extraction from tissue.

Recombinant production affords a convenient means to resolve this bottleneck. Consequently, recombinant protein production has become a dominant force for producing protein specimens today. A number of expression systems exist, utilizing prokaryotic or eukaryotic host organisms, or also cell-free systems, each with their own merit.1,2 Among these, heterologous expression in mammalian cells or insect cells, respectively, have emerged as methods of choice for eukaryotic protein specimens, exemplified by a rapidly growing number of entries in the Protein Data Bank (PDB) using these expression hosts. Mammalian expression has been determinant for the production of large, multidomain, sometimes highly glycosylated eukaryotic cell surface receptors and their supramolecular assemblies.3 For eukaryotic multiprotein complexes with many subunits, the baculovirus expression vector system (BEVS) has turned out to be particularly powerful, unlocking the structure and mechanism of many important complex assemblies that had remained inaccessible to detailed analysis beforehand.4,5 Today, baculovirus is prolifically used to accelerate a wide range of applications in the life sciences (Fig. 1).

Keywords: automation, baculovirus, DNA recombineering, insect cell culture, multiprotein complex, protein modification, recombinant expression, structural biology, synthetic biology

*Correspondence to: Imre Berger; Email: iberger@embl.fr; Arnaud Poterszman; Email: Arnaud.Poterzman@igbmc.fr
Submitted: 08/10/2015
Revised: 09/29/2015
Accepted: 09/30/2015
http://dx.doi.org/10.1080/21655979.2015.1104433

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.
baculovirus was published in 1983. Max Summers, Gale Smith and colleagues used Autographa californica multiple nucleopolyhedrovirus (AcMNPV) to express Autographa californica multiple nucleopolyhedrin (polh), with the human gene encoding for INF-β. They have substituted the gene of a viral protein, polyhedrin (polh), with the human gene encoding for INF-β. Since this hallmark study, the procedures were step-by-step improved to render this expression system user friendly, involving engineering of recombinant viruses and development of time and cost-effective protocols for protein production in insect cell cultures.

The original procedure to generate recombinant baculoviruses was based on homologous recombination into the polyh locus, in insect cells, between a viral genome and a transfer plasmid harboring the gene(s) of interest. This was not overly efficient, and required inspection of infected cells by eye to identify the polyh deficient phenotype which would lack granular shapes that were birefringent. This approach was much improved when suitable restriction enzymes were identified to cleave the baculovirus within the polyh locus, thus linearizing the baculoviral genome and eliminating non recombined genomes. Now, recircularization of the viral genome, and thus live and infectious baculovirus generation depended on the success of heterologous gene integration, which still occurred by homologous recombination in the insect cell which thereby invited their own destruction, after releasing recombinant baculovirions to infect further cells in the culture. This was already a much improved procedure, and in the hands of experienced users could be applied relatively straightforwardly in the research laboratory.

A breakthrough which propelled baculovirus expression and rendered it genuinely accessible to many laboratories was the construction of an artificial chromosome (BAC) made up of the baculoviral genome supplemented by DNA elements that enabled stable propagation of this BAC at very low copy number (1-2) in E.coli DH10 cells. This approach profoundly simplified the generation of recombinant baculovirus, removing laborious and time consuming interventions to clonally isolate producer viruses by plaque-purification (Fig. 2A). The first and to date most popular approach for inserting foreign genes into the BAC relies on Tn7-mediated transposition. This BAC/Tn7 entry system is used by several popular systems including the classical Bac-to-Bac™ system from Invitrogen / Life sciences, and also by the more recent MultiBac system which in recent years became the lead technology for BEVS-mediated production of multiprotein complexes for many applications in academia and industry. In the BAC/Tn7 approach, prior to the production of the recombinant virus, the gene of interest is cloned into a transfer plasmid harboring the expression cassette which is flanked by short DNA sequences representing the right and left arms of the Tn7 transposon. The sequenced transfer plasmid is then transformed into custom E. coli DH10 cells which contain the BAC and in addition a helper plasmid that supplies the Tn7 transposase needed to carry out the transposition into the Tn7 attachment site that is present on the BAC. The resulting recombinant BAC DNA (bacmid) is extracted and can be verified easily by a PCR reaction with appropriate primers. The recombinant bacmid DNA is then used to transfect insect cells using a lipidic transfection reagent in small cell cultures. Infected cells then realize the viral genome, giving rise to live baculovirions that can then be used to infect larger cell cultures to produce and purify the protein specimen of interest (Figs. 1, 2).

Numerous variations to the BAC approach have been implemented. A BAC was constructed carrying a disabling mutation in an essential gene (ORF1692) which impairs virus replication. This detrimental mutation is repaired when the mutant BAC is co-transfected with a transfer plasmid containing the expression cassette and a wild type copy of the mutated gene. Homologous recombination between the viral DNA and the transfer plasmid restores replication, eliminates the bacterial replicon in the polyhedrin locus, and integrates the gene of interest (Figs. 2A, B). As replication of not recombinant viral genomes is ruled out here, clonal selection procedures are not

---

**Figure 1. Baculovirus Expression Vector System BEVS.** The BEVS can be conceptualized on 3 levels. DNA elements comprising genes of interest, promoters, terminators, transcriptional enhancers and others are combined with plasmid backbone modules into (multi)gene expression cassettes on transfer plasmids (left). Transfer plasmids are then integrated into a baculoviral genome by using transposition or recombination methods to yield a composite baculoviral genome containing all heterologous DNA elements (center). Recombinant baculovirus containing the composite genome is then used to infect insect cell cultures to produce materials for a multitude of downstream applications (right). All three levels, transfer plasmid assembly, the baculoviral genome and the expression host cell, can be engineered by exploiting powerful synthetic biology tools, for optimizing and maximizing system performance.
Figure 2. Baculovirus multigene expression tools. (A) Flowcharts for production of recombinant viruses and protein expression protein production using either Tn7-mediated transposition (top) or homologous recombination (bottom) are shown. (B) Construction of recombinant baculovirus genomes are shown in a schematic representation. Transposition of an expression cassette embedded between the Tn7L and Tn7R DNA sequence into the Tn7 attachment site of the baculoviral genome is illustrated on the left (from Ref. 8). Homologous recombination combining a linearized baculoviral DNA genome and a transfer plasmid for inserting the gene(s) of interest is shown on the right (from Ref. 13). Most currently used viral genomes contain deletions of the genes encoding for chiA and V-cath, a chitinase and protease, respectively, shown to be detrimental for protein production and stability. This locus can be used to insert foreign genes including fluorescent makers such as YFP or the DsRed protein to monitor virus performance and protein production. A LoxP site-specific recombination site, inserted at this position, allows to generate composite viral genomes by in vivo Cre-LoxP fusion in E. coli with a customized transfer plasmids containing a conditional replication origin.9 (C) Multigene assembly DNA tools used in tandem recombineering (TR) are shown. So-called Acceptor and Donor plasmids contain multiple cloning sites (MCS) to insert genes of interest by conventional or samless cloning methods of choice. Genes are placed typically under the control of late baculoviral promoters (PH or p10) and eukaryotic polyadenylation signals (PH ter, SV40 or HSVtk). Acceptors and Donors possess multiplications modules (gray rectangles) located on both sides of the expression cassette as well a loxP site (circle filled in orange). (D) For multi-gene construct generation using multiplication modules, individual expression cassettes are excised by digestion with a pair of endonucleases and inserted via compatible restriction sites into the multiplication module of a progenitor plasmid. Following ligation, the restriction sites used for integration are eliminated and multiplication can be repeated iteratively using the intact multiplication module in the inserted cassette.6 Plasmids thus charged with several genes of interest can be fused by Cre mediated recombination via the LoxP sites. Acceptors have a regular origin of replication (ori ColE1), whereas donors have a conditional origin derived from R6Kγ phage (ori R6Kγ), facilitating multigene assembly.21,30,39
required. This effectively generates a time-saving one-step procedure. Ready-to-use linearized DNA Baclovirus DNAs are available commercially (e.g., OET’s FlashBac) but can also be prepared with reasonable cost and effort from appropriate bacterial strains. This approach lends itself to automation and thus to high-throughput applications when a large number of constructs are to be screened in expression tests. Automation of the BAC/Tn7 approach is also possible, but far more laborious as many more steps to prepare bacmid DNA have to be considered. Recently, a set of transfer plasmids, called OmniBac, combines the options to use both the Tn7/BAC system and classical heterologous recombination approach using the same entry reagents, thus allowing exploitation of the advantages of either system without having to reclone the heterologous genes of interest.  

Protein expression has drawn vast benefit from advances in synthetic biology and recombinering technologies that allow targeted modification of the baculovirus genome. Most platforms for virus production now rely on engineered viral genomes in which genes considered to be detrimental for protein production or stability have been deleted. Among the first genes that were removed following the original polh deletion are v-cath and chiA, that encode a cathepsin protease and a chitinase, respectively, which are involved in the liquefaction of infected larvae. This was followed up by elimination of further genes (p10, p26, p73). The loci for v-cath and chiA are juxtaposed in the viral genome, and this region has been utilized to insert genes encoding maker proteins such as yellow fluorescent protein (YFP) or sea anemone red fluorescent protein (dsRed), as well as chaperones or post-translational modification enzymes by site specific recombinolation, thus generating novel functionalized genomes. Additions can include kinases, phosphates as well as enzymes that allow to synthetize complex carbohydrates found on many secreted proteins of mammalian origin.  

Synthetic biology is emerging as a profound game-changer in the life sciences and synthetic biology approaches are opening up entirely new avenues that may transform recombinant protein production in the future. We are only beginning to scratch at the surface of what is possible. The design and synthesis of the entire one megabase Mycoplasma mycoides bacterial genome has been achieved. We now have all tools in hand to entirely rewrite the 130 kilobases baculovirus genome and create synthetic versions comprising only DNA elements essential for laboratory applications. Synthetic baculoviral genomes with improved genetic stability and customized for optimal protein production of specific classes of proteins at high levels, will benefit for applications ranging from drug discovery to industrial protein therapeutics production.  

### Strategies for Protein Complex Production

Expression levels and solubility of recombinant proteins expressed by BEVS can range from a few micrograms to hundreds of milligrams per liter, and cannot be predicted. Typically, constructs need to be modified by truncation, mutation or deletion of low complexity regions. The use of adapted affinity tags is often the only means to obtain pure protein. Numerous pipelines for expression screening and protein production at medium to large scale (250 ml-6L) have been established. Preparations of multiprotein complexes in particular for structural biology applications has its own challenges and specifics. Sometimes, functional subunits of a given complex can be produced in isolation. Then, protein complexes may be assembled in vitro from these individually expressed subunits. Reconstitution methods are particularly useful when the complex is short-lived and thus cannot be purified intact. One or several components may not be proteinaceous, such as in protein-DNA or protein-RNA complexes or enzymatic assemblies with coenzymes and prosthetic factors, or when a small-molecular ligand is required to stabilize a complex. Here, components are expressed individually, and revised expression experiments can be easily accommodated to modify the composition of the complex by replacing one or more subunits by mutant versions for example to define protein functional domains, optimize solubility or to reduce conformational heterogeneity by eliminating disordered regions that often interfere with high-resolution structural studies. In many cases, however, in vitro reconstitution is not applicable as individual subunits of a complex often cannot be expressed and manipulated in absence of their protein partners. Then, co-expression is a vital prerequisite. In BEVS, this can be achieved by co-infection with multiple viruses, each expressing a single polypeptide, or, alternatively, by infection with a single virus driving the expression of all foreign genes required. If a large number of proteins need to be co-expressed, a combination of both approaches can be applied. This strategy has been particularly successfully for the reconstitution, structural and functional analysis of large multi-protein complexes including large human general transcription factors TFIIH and TFIIID and two essential multiprotein components of the gene regulation machinery. Co-expression of many gene products from several viruses each encoding for a single or limited number of heterologous proteins exploits the capacity of insect cells to be simultaneously infected by multiple viruses. This approach affords a certain flexibility to adjust protein expression ratios by controlling the multiplicity of infection (MOI) of the individual viruses which are administrated, and has been used successfully for mapping protein-protein interactions, and for providing additional factors for post-translational modifications and chaperones for enhancing proper folding. Co-infection can work reasonably well for small-scale experiments when a small number (2-3) baculoviruses are added, but scale-up is inefficient and logistically demanding, in particular for the production of complex assemblies that would require co-infection of a large number of viruses, and viral titers for all of these would need to be carefully gauged. The uptake of the individual viruses by the cells in an infected culture will follow a Poisson distribution, resulting in subpopulations of cells
infected by subsets of viruses.\textsuperscript{35} Furthermore, it was shown that excessive viral load can inhibit protein expression.\textsuperscript{36} Using a single viral genome that contains all heterologous genes circumvents these issues, thus ensuring co-expression of all heterologous genes in every cell in an infected culture. Several early baculovirus transfer plasmids were designed for accepting 2-4 genes (pFastBac\textsuperscript{TM} Dual from Invitrogen / Life Sciences, and pAcUW51, pAcAB, pAcAB4).\textsuperscript{37,38}

A theoretically unlimited number of heterologous genes can be co-expressed by the more recent MultiBac system, which exploits a method called tandem-recombineering (TR) based on seamless gene integration by sequence and ligation independent cloning techniques into an array of small transfer plasmid modules which are then conjoined by site-specific recombination using the Cre-Lox fusion reaction.\textsuperscript{39} Multigene expression cassettes are then inserted into the MultiBac baculoviral genome by Tn7 transposition into a Tn7 attachment site (in the polh locus) and by site-specific recombination into a LoxP site on the viral backbone (in the v-cath-chiA locus)\textsuperscript{9,21} A series of transfer plasmids were developed to facilitate multigene cassette assembly (Figs. 2C, D). A large and growing set of tools now exist to assemble DNA elements into large functional multigene circuits, including homing endonuclease based fragment assembly, one-step isothermal in vitro recombination (Gibson cloning), or restriction free (Rf) cloning to name a few\textsuperscript{39,40,41} all of which can be used for multigene expression with the MultiBac system.

In case of complexes with many different subunits or containing large proteins, the size of the insert and clustering of identical promoter and terminator sequences may lead to genetic instability of the recombinant baculoviral genome, resulting in deletions notably of the inserted foreign DNA. This can be avoided by rigorously adhering to virus amplification protocols designed to minimize such damage.\textsuperscript{21} Alternatively, heterologous genes can be spread over multiple insertion sites in the baculoviral genome by iterative homologous recombination steps carried out on a BAC in \textit{E.coli} cells.\textsuperscript{42,43} Applying these strategies can significantly increase the number of genes that can be co-expressed while preserving the stability of the viral genome.

### Stoichiometry of Complex Subunits

Assembly and activity of cellular multienzyme complexes critically depends on correct stoichiometry of the subunits present in the complex. In the cell, elaborate mechanisms involving chaperones, assembly factors and transport between cellular compartments exist that assist and regulate complex assembly. A recombinant expression experiment typically aims at substantial overexpression in a heterologous host cell, in which corresponding mechanisms may or may not exist, but will likely fail to cope with very high level expression of the foreign proteins. This challenge often needs to be addressed during expression and during purification by tight control of the proper stoichiometry of the individual subunits that make up the complex.

Both for complexes reconstituted \textit{in vitro} and for complexes assembled \textit{in situ} within the expression host cell, the desired physiological complex will need to be separated from excess subunits and partially assembled sub-complexes, that would otherwise give rise to heterogenous sample preparations impeding down-stream applications, notably structural and mechanistic analyses that require ultra-pure sample. Purification schemes typically combine orthogonal techniques in a sequential manner that has to be customized with care to yield physiological assemblies that are homogenous in composition.\textsuperscript{44} On the other hand, it may be essential to tightly control the amounts of subunits already when they are expressed for successful complex assembly, prior to extraction. This sometimes imposing challenge was recently addressed in an elegant fashion by adapting a strategy used by certain viruses in nature to realize their proteome. For example Coronavirus expresses its gene product repertoire from 2 large open reading frames that give rise to polyproteins that are co-translationally processed by highly specific proteases, also encoded in the ORFs, into the individual protein components. A recombinant approach which recapitulates this strategy was implemented in the MultiBac system recently, exploiting the action of tobacco etch virus (TEV) protease to process a synthetic polyprotein that was inserted into the MultiBac baculoviral genome by means of a customized transfer plasmid.\textsuperscript{11,45} This approach, ComplexLink, was successfully applied to produce, for the first time, influenza polymerase that could be crystallized unlocking the atomic structure of this complex which has remained elusive for decades.\textsuperscript{46}

### Baculovirus Expression: Stay tuned!

Baculovirus-mediated production of heterologous proteins in insect cells has made come a long way since its inception in the Summers lab many years ago.\textsuperscript{47,48} Proteins for a wide range of applications, in basic and applied research, from structural biology to vaccinology have benefitted substantially from this powerful and versatile system. Many proteins and protein complexes were produced, often for the first time, by BEVS, in the quality and quantity to enable high resolution structural ad mechanistic analysis, advancing fundamental insight and accelerating drug discovery. Much has been done over the years to develop improved reagents and robust protocols, which already allow efficient application of the system also by non-specialist users, and BEVS is firmly established a work-horse for eukaryotic protein production in many laboratories, in academia and industry. Improved viral backbones with customized functionalities for particular protein classes and applications are being developed, compellingly underscoring the enormous potential of this system at the forefront of life science research.

Conceptually, BEVS can be understood as a system operating at 3 distinct levels, all of which are amenable to optimization by engineering (Fig. 1). The first level is the transfer reagents which should enable rapid and flexible assembly of multicomponent gene expression circuits, ideally in a highly parallelized, automated manner suited in a robotics envirornment. Promoters,
terminators, gene boundaries and affinity tags should be freely exchangeable at this level, with minimal investment for the scientists carrying out the experiment. Methods from synthetic biology are essential for achieving this objective, and are being successfully implemented in BEVS, substantially increasing throughput. Moreover, gene synthesis has become thoroughly affordable, effectively multiplying the options to design and lead to success multiparameter expression experiments.

The baculoviral genome constitutes the second level in this concept. The benefit of modifying the baluloviral genome was already evident very early on, when the polh gene was removed from the viral backbone. Before, recombinant baculovirus had to be identified by a specialist user observing the loss of occluded virus phenotype in infected cells, which occurred with very low efficiency. This improved considerably when the polh locus was disrupted, and dramatically when the baculoviral genome was linearized by endonuclease cleavage in the polh locus. The polh deletion was soon followed by further knockouts of genes non-essential in laboratory culture, which exhibited a negative effect on virus production, notably v-cath and chiA, and others. Today, with the advent of powerful technologies to manipulate the industrial biotechnology community to the improvement of the recombinant expression in this prokaryotic host. Undoubtedly, it is much more demanding to engineer a eukaryotic cell. The advent of powerful genome engineering tools, notably CRISPR/Cas has revolutionized eukaryotic genome engineering with transformative potential. We anticipate that these technologies will be progressively exploited in the near future to extend optimization to the host cell genome, to create specialized cell strains with defined functionalities. Coupled to proteomics and metabolic flux analyses during heterologous expression, genome engineering could be likewise applied to eliminate negative factors in the host cell genome, by targeted deletions ideally in high-throughput. Moreover, helper functionalities such as chaperones and modifiers, but also selected components of a protein complex under investigation or even complete complexes could then be outsourced conveniently from the baculovirus to the host cell genome. This could open entirely new possibilities for drug discovery, greatly accelerating research and progress. For instance, in the pharmaceutical industry, hundreds of kinases are produced, often with baculoviral systems, and many remain problematic and rely on accessory factors. Multiprotein membrane protein targets are increasingly in focus, and are basically not accessible with available methods. Moreover, a major focus in the industrial biotechnology community is on assembling and testing synthetic metabolic circuits and signaling cascades, for which the here described technologies may be uniquely useful. We expect that engineering on all levels will ever improve and optimize this versatile and powerful expression system, enabling researchers to address the most complex question in the biology of health and disease states in the years to come.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgment
We thank all members of our laboratories for helpful discussion.

Funding
This work was funded by the CNRS, the INSERM, the Université de Strasbourg (UdS), the Alsace Region and the French Infrastructure for Integrated Structural Biology (FRISBI), Institute, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions. AP acknowledges support from the Agence Nationale de la Recherche (ANR), the Institut National du Cancer (INCA), the Fondation pour la Recherche sur le Cancer (ARC), the Fondation pour la Recherche Médicale (FRM) and La Ligue contre le Cancer. IB acknowledges support from ANR, ARC and the European Commission (EC) Framework Program (FP) 7 project ComplexINC (279039).

References