


InteBac: An integrated bacterial and baculovirus expression vector suite

Veronika Altmannova | Andreas Blaha | Susanne Astrinidis | Heidi Reichle | John R. Weir 

Friedrich Miescher Laboratory of the Max-Planck-Society, Max-Planck-Ring 9, Tübingen, Germany

Correspondence

John R. Weir, Friedrich Miescher Laboratory of the Max-Planck-Society, Max-Planck-Ring 9, Tübingen, Germany. Email: john.weir@tuebingen.mpg.de

Present address

Andreas Blaha, Research Institute of Molecular Pathology (IMP), Campus-Vienna-Biocenter 1, Vienna, Austria.

Funding information

Max-Planck-Gesellschaft; Max Planck Society

Abstract

The successful production of recombinant protein for biochemical, biophysical, and structural biological studies critically depends on the correct expression organism. Currently, the most commonly used expression organisms for structural studies are *Escherichia coli* (~70% of all PDB structures) and the baculovirus/ insect cell expression system (~5% of all PDB structures). While insect cell expression is frequently successful for large eukaryotic proteins, it is relatively expensive and time-consuming compared to *E. coli* expression. Frequently the decision to carry out a baculovirus project means restarting cloning from scratch. Here we describe an integrated system that allows simultaneous cloning into *E. coli* and baculovirus expression vectors using the same PCR products. The system offers a flexible array of N- and C-terminal affinity, solubilization and utility tags, and the speed allows expression screening to be completed in *E. coli*, before carrying out time and cost-intensive experiments in baculovirus. Importantly, we describe a means of rapidly generating polycistronic bacterial constructs based on the hugely successful biGBac system, making InteBac of particular interest for researchers working on recombinant protein complexes.

KEYWORDS

bacterial expression, cloning system, Gibson assembly, insect cell expression, polycistronic

1 | INTRODUCTION

Obtaining recombinant protein of interest can be a challenging multi-parametric problem. The parameters to consider include codon usage, vector, fusion tag, expression organism, expression conditions, and purification strategy.¹ Previous work has described the use of universal vectors compatible with *Escherichia coli*, baculovirus, and mammalian expression systems for example the pOPIN system.² However, for insect cell

expression the excellent MultiBac system³ has set the standard. A recent baculovirus expression method has combined MultiBac with Gibson assembly⁴ to yield the biGBac system.^{5,6} Using biGBac one can assemble a five open reading frame polycistronic vector in a single step, and combine five of these in a second step to assemble up to 25 ORFs in a single vector. Our overwhelming satisfaction with biGBac leads us to further develop the system while creating a parallel, compatible, system for *E. coli*.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

We have previously used the pST44 polycistronic expression system for bacterial expression.⁷ Not only does pST44 provide a rapid means of generating multicistronic constructs, but the pST44 family pTRC50 vectors are excellent expression vectors in their own right. However, the pST44 method uses restriction enzyme based approaches which have been superseded by ligation independent approaches, particularly Gibson assembly, in the last 10 years.⁴

Until recently in our lab, one would decide to pursue either an *E. coli* approach or an insect cell approach to obtaining a recombinant protein of interest. Working in parallel was of course always possible, but would require creating different PCR products, preparing different vectors, and being limited by the fusion proteins available for each system. To address this limitation we set out to create a unified and integrated bacterial/insect cell expression system. Our goal was to be able to take a single PCR product and clone this into numerous *E. coli* and insect cell expression vectors. One would process both sets of vectors in parallel, and have the result for *E. coli* expression before one even transfected insect cells. The ultimate result being that one has more time and resources to explore the parametric space of recombinant protein expression (summarized in Figure 1).

When pursuing the expression of multisubunit complexes, the creation of a polycistronic construct, for either insect cell or bacterial expression, is a useful means of both ensuring appropriate complex stoichiometry but also reducing the complexity of a biochemical reconstitution. However, the establishment of the appropriate

conditions by screening subunits and fusion tags is easier when combining multiple monocistronic constructs in co-expression experiments. In insect cells, this can be achieved by co-infecting with multiple different viruses, albeit with some limitations (described in Reference 10). We seldom co-infect with more than two viruses, and it is important to screen different ratios of the virus. In our lab, this is usually done volumetrically combining different ratios of virus A with virus B. In *E. coli*, the situation is complicated by both antibiotic and origin of replication usage and subsequent plasmid incompatibility. To this end, we modified our *E. coli* vector set with origins of replication and antibiotic resistance compatible with co-expression.

Here we describe the implementation of a parallel system for the rapid screening of *E. coli* co-expression vectors that are compatible with simultaneous cloning into insect cell expression vectors. Furthermore, we describe a Gibson assembly-based approach for the rapid generation of polycistronic *E. coli* expression constructs. This system has greatly improved the workflow in our lab, and we hope other labs will benefit from our efforts.

2 | RESULTS

2.1 | Generation of N- and C-terminal fusion vectors

We started with creating an insect cell expression vector based on the pLIB vector^{5,6} with a variety of N-terminal

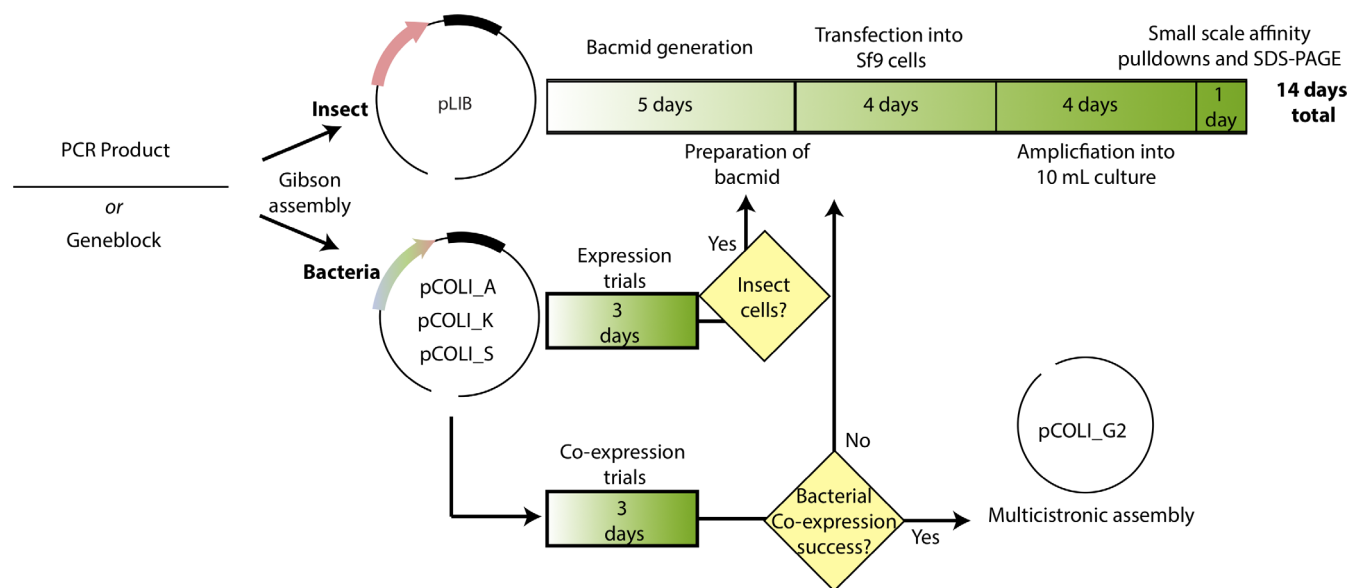


FIGURE 1 Workflow within the InteBac and biGBac systems. Using a single PCR product one clones simultaneously into pLIB (for insect cells) and pCOLI (for bacteria). Depending upon the results of the *Escherichia coli* expression trial, one can decide whether to proceed with insect cell work

fusion proteins. We divided the tags into different categories; affinity, solubilization, and utility (see Table 1). In order to have universal overhangs for both the insect cells and bacterial expression systems, these were designed to correspond to a rhinovirus-3C site between the fusion protein and the protein of interest (see Table 2 for all primer overhangs). We chose 3C cleavage site over the more frequently used TEV site, due to the 3C protease's higher catalytic activity at lower temperatures.⁸ Next we created a more limited set of C-terminal fusion proteins, placing a serine-glycine linker between the protein of interest and the C-terminal fusion protein. This linker ensured that the C-terminal overhang would be universal for all C-terminal fusions. We transferred these fusion protein ORFs from the pLIB backbone into the pTRC50 backbone (Ampr/pBR232 origin⁷), for bacterial expression (from now on referred to as pCOLI_A). Finally, in order to give us the greatest flexibility in *E. coli*, we

transferred all the expression cassettes into two additional backbones pCOLI-S (Strepr/RSF1030 origin⁹) and pCOLI_K (Kanr/CloDF13 origin¹¹). This combination of resistances and origins of replication gives the user the ability to co-express three proteins simultaneously (see Table S1 for an exhaustive list of all expression vectors).

2.2 | Untagged vectors

Many proteins are not amenable to N- or C-terminal tagging, but can be purified through the affinity tag on a binding partner. In order to facilitate the co-expression of several proteins, we required untagged vectors. Despite our efforts, we were unable to generate a generic N-terminal overhang that would work for both untagged *E. coli* and insect cell vectors. As such, there is a generic overhang for untagged insect cell and untagged *E. coli*

TABLE 1 Summary of fusion proteins used in the InteBac system

Fusion name	Description	Type	Mw N-term fusion
6xHis	IMAC purification	Affinity	3 kDa
12xHis	IMAC from insect cells		3.8 kDa
STREP	Twin strep-II tag		4.9 kDa
CBP	Calmodulin binding peptide		4.3 kDa
GST	Glutathione-transferase	Solubilization/affinity	26.7 kDa
MBP	6xHis plus maltose binding protein		42.7 kD
SUMO	6xHis plus SUMO	Solubilization	13.5 kDa
Trx	6xHis plus Thioredoxin		14 kDa
SNAP	6xHis plus SNAP tag	Utility	20 kDa
HA	6xHis plus 3 x HA	Identification	6.3 kDa
Myc	6xHis plus 6 x Myc		11 kDa

Primer name	Sequence (5' → 3')	Description
Tag_Fwd	CTGTTCCAGGGGCCCGATCC [ORF]	For cloning into all N-terminal fusion expression vectors
Rev	TCCTCTAGTACTTCTC GACAAGCTTTTA [rev_comp_ORF]	For cloning into all vectors with no C-terminal fusion
LIB_Fwd	CCACCATCGGGCGCGGATCC [ORF]	Cloning into pLIB vectors with no N-terminal fusion
Tag_Rev	TCCAGATCCAGATCCGCTTCCACT [rev_comp_ORF]	Cloning into all vectors with C-terminal fusion protein
COLI_Fwd	TTTGTTTAACTTTAAGAAG GAGACTGGATC[ORF]	Cloning into all pCOLI vectors with no N-term fusion

TABLE 2 Primers used to clone the gene of interest into the pLIB and pCOLI vectors as untagged or C- or N-terminal fusion constructs

vectors, rather a specific forward primer for each (pLIB_fwd and COLI_fwd respectively, see Table 2). To facilitate co-expression in *E. coli* we also created untagged pCOLI_S and pCOLI_K vectors. These vectors contain compatible origins of replication and resistances to facilitate co-transformation into bacteria.

2.3 | Multicistronic vectors

Our pLIB derived vectors remain fully compatible with the pBIG multicistronic vectors from the biGBac system.^{5,6} To create a multicistronic bacterial vector we took the pST44 vector backbone and added a gentamycin cassette (from now on referred to pCOLI_G2). We designed a set of PCR primers for amplifying the entire ORF from the pCOLI family of vectors (including the RBS, but not the promoter or terminator). The Gibson overhangs described in the biGBac system were thoroughly tested, both in silico and in vitro, to give the greatest assembly efficiency. As such we use the same principle, and indeed the same overhang sequences as in biGBac, to create a multicistronic pST44 vector, in addition to the use of the *Swa*I enzyme (summarised in Figure 2).

2.4 | Proof of concept—RPA complex

Our interest in homologous recombination has led us to look at several protein complexes involved, one of which is RPA (Replication Protein A, reviewed in References 12,13), a heterotrimer consisting of Rfa1, Rfa2, and Rfa3 in budding yeast.¹⁴ Expression and purification of yeast RPA in *E. coli* has been previously described in Reference 15. We cloned Rfa1, Rfa2, and Rfa3 into Strep-pCOLI-A, HispCOLI-S, and His-pCOLI-K respectively. We initially demonstrated that the complex could be expressed and partially purified through a co-expression of all three proteins in *E. coli* C41(DE3), and purification via the twin Strep-II tag, followed by confirmation of protein identity via western blotting (Figure 3, lanes 1–4). We amplified the expression cassettes for each of the three RPA subunits, and Gibson assembled into the linearized pCOLI-2G backbone. Gentamycin resistant transformants were confirmed by sequencing and subsequently transformed into the BL21 cells.

Our Gibson assembly of polycistronic RPA was just as successful, if not more so than the co-expression of all three RPA subunits. Furthermore, there is the advantage of carrying out transformations with a single plasmid, using a single selection antibiotic, and the possibility of carrying out further co-expression with pCOLI-K and pCOLI-S (Figure 3).

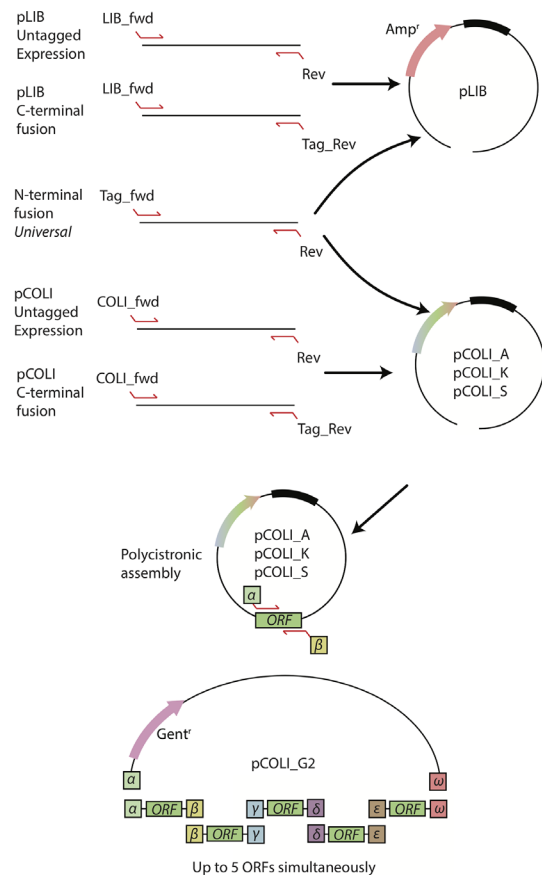


FIGURE 2 Cloning in the InteBac system. Primers, or overhangs on geneblocks, are chosen to match the vector. The N-terminal fusion vectors are truly universal, allowing for cloning into either the pLIB or the pCOLI backbones. From the pCOLI backbones one can generate a polycistronic construct with up to five insertions. Fewer insertions can be used, but the alpha and omega overhangs must be present

3 | DISCUSSION AND CONCLUSIONS

Our system is also well suited to the use of “oven-ready” synthetic dsDNA (Geneblocks), allowing the incorporation of the Tag_fwd and Rev overhangs into the synthetic DNA, before assembly into the pLIB or pCOLI vectors. Currently such synthetic dsDNA fragments are available up to ~3,000 bp in length, though this usually requires screening multiple colonies to find a transformant with the correct sequence. Sequence verified dsDNA fragments are also available, but at a higher price. Additionally we would recommend users to explore the use of codon optimization. In our experience, constructs optimized for expression in *E. coli* K12 also work well in either *Spodoptera frugiperda* or *Trichoplusia ni* (SF9 and Hi5 cells).

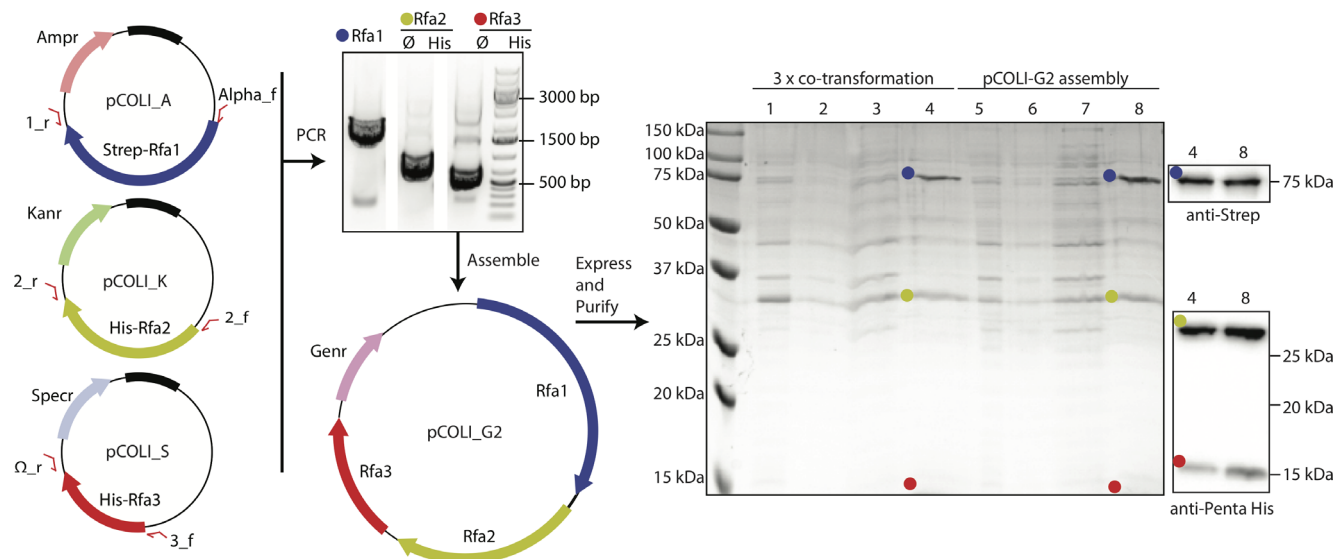


FIGURE 3 Cloning and expression of the trimeric RPA complex from yeast. Each RPA subunit was cloned into a different pCOLI backbone, which were then used for co-expression. Additionally, we generated a multicistronic assembly of RPA into the pCOLI_G2 backbone. We compared the expression of the co-transformation versus pCOLI_G2. SDS-PAGE was run of crude lysate (lanes 1 and 5), clear lysate (lanes 2 and 6), flow through from the resin (lanes 3 and 7), and elution from the beads (4 and 8)

The implementation of our InteBac system has greatly streamlined the work processes in our laboratory. It has allowed us to explore additional experimental space in terms of finding suitable expression conditions for our protein of interest. Previous systems, including the “pCoofy”¹⁶ have also made use of single PCR products to be integrated into a range of different expression vectors; including those for insect cell expression. InteBac differs from pCoofy in several key areas. Firstly, InteBac has been designed to work as an “add on” to the biGBac system, allowing the generated insect cell expression vectors to be turned into multicistronic vectors through a single Gibson assembly step. Secondly, InteBac makes use of three different bacterial expression backbones facilitating rapid co-expression of up to three proteins in *E. coli* before moving to a polycistronic construct. Finally, we offer the possibility of producing a multicistronic *E. coli* expression vector (based on the principles outlined in biGBac^{5,6}) in pCOLIG2. As such we consider InteBac of particular interest to those researchers working on protein complexes.

4 | MATERIALS AND METHODS

4.1 | Vector construction

All cloning and plasmid manipulation steps were first carried out in silico using the SnapGene software (GSL Bio-Tech LLC). The pTRC50 and pLIB vectors were gifts from Song Tan (Penn State) and Jan Michael Peters (IMP Vienna), respectively. PCR amplifications were carried out

using 2× Q5 Master Mix (NEB), with cycling times and temperatures according to the manufacturer’s instructions. The Kanamycin and Streptomycin resistance/origin of replication modules were synthetic dsDNA constructs (IDT). The gentamycin cassette was amplified from the pBIG2 vectors. Since the gentamycin cassette also contains one restriction site for BglII, we introduced a silent mutation into its sequence. All affinity tags insertions and plasmid manipulation was carried out using a combination of synthetic dsDNA (IDT) and Gibson assembly. Successful assemblies were verified by Sanger sequencing.

4.2 | Gibson master mix

For all Gibson assemblies we used our own master mix. Briefly, a 5x isothermal reaction buffer was prepared (25% PEG-8000, 500 mM Tris-HCl pH 7.5, 50 mM MgCl₂, 50 mM DTT, 1 mM each of the 4 dNTPs, and 5 mM NAD) and pre-aliquoted. To prepare Gibson master mix we combined 320 μl 5X ISO buffer with 0.64 μl of 10 U/μl T5 exonuclease, 20 μl of 2 U/μl Phusion polymerase and 160 μl of 40 U/μl Taq ligase (all enzymes from NEB). ddH₂O was then added to a final volume of 1.2 ml.

4.3 | Linear vector preparation

All vectors (except for pCOLI_G2) are designed to be linearized with the same combination of restriction enzymes (BamHI and HindIII). Proper vector

linearization and subsequent purification are critical to the success of downstream cloning. Briefly, we took 1 μ g of plasmid (typically from a midi-prep [Qiagen]) and digested in a 20 μ l reaction with 20 units of BamHI-HF (NEB) and 20 units of HindIII-HF (NEB) in CutSmart buffer (NEB) for 3 hr. Each reaction was then gel purified using the Wizard SV kit (Promega) according to the manufacturer's instructions, with the exception that 30 μ l of ddH₂O was used for elution from the column. The final linearized plasmid product had a typical concentration of 60–100 ng/ μ l. To linearize pCOLI_G2 vector the restriction enzymes BglIII and XhoI were used and the plasmid was further processed as described before.

4.4 | Insert preparation

All inserts were amplified using Q5 polymerase (NEB). PCR reactions were gel purified using Wizard SV gel and PCR cleanup. In case of amplification from a plasmid source, we paid attention to the size of the insert versus the template. In case of any potential overlap we treated our PCR reactions with 1 μ l DpnI (NEB) to eliminate the donor plasmid. DpnI was then heat-inactivated (15 min 65°C) before gel purification.

4.5 | Gibson cloning and verification

Routinely we mixed 4.5 μ l of purified insert with 0.5 μ l of vector and added this 5 μ l to one 15 μ l aliquot of Gibson master mix. Our Gibson reactions were then incubated for 1 hr at 50°C, and then transformed directly into chemically competent XL1-Blue. From numerous colonies we would typically grow two, and prep one for sequencing, keeping the other as a backup. Typically, with a well-prepared vector (see above) our cloning success rate is >95%, so we considered it wasteful to “prescreen” with analytical digests. All agarose gels shown are 0.8% agarose, stained with GelGreen (Biotium Inc), and imaged with a ChemiDoc MP imaging system (BioRad).

4.5.1 | Cloning of RPA

The RPA ORFs (Rfa1, Rfa2, and Rfa3) were amplified from *Saccharomyces cerevisiae* genomic DNA (SK1 strain), using the following primers Rfa1_Tag_Fwd.

(CTGTTCCAGGGGCCCGGATCC ATGAGCAGTGT TCAACTTTTCGAGGGGCGAT), Rfa1_rev (TCCTCTAG TACTTCTCGACAAGCTTTTATTAAGCTAACAAAGCC-TTGGATAACTCATCGGCAAG), Rfa2_Tag_Fwd (CTG TTCCAGGGGCCCGGATCCATGGCAACCTATCAACC-ATATAACGAATATTC), Rfa2_rev (TCCTCTAGTACTTC

TCGACAAGCTTTTATCATAGGGCAAAGAAGTTATT-GTCATCAAAAG), Rfa3_Tag_Fwd (CTGTTCCAGGGGCCCGGATCCATGGCCAGCGAAACACCAAGAGTTGACCCC), Rfa3_rev (TCCTCTAGTACTTCTCGACAAGCT TTTACTAGTATATTTCTGGGTATTTCTTACATAG). Rfa1 was cloned into pCOLI_A_Strep, Rfa2 into pCOLI_K_His, and Rfa3 into pCOLI_S_His. For the multicistronic assembly the Rfa1 RBS/ORF was amplified using the Alpha_Fwd and CasI_rev primers; Rfa2 with the CasII_fwd and CasII_rev primers and Rfa3 with CasIII_fwd and Omega_rev primers (Table S2). The PCR amplified RBS/ORFs for each of the three RPA subunits were then assembled into linearized pCOLI_G2, with a 3–5 fold molar excess over the plasmid backbone, as previously described for pBIG assembly.^{5,6} Gibson reactions were transformed directly into chemically competent XL1-blue *E. coli*, and selected on gentamycin LB agar plates. Minipreps of eight positive colonies were prepared, and subject to SmaI digest to release the individual RBS/ORF cassettes. Digests were then subject to agarose gel electrophoresis, and those clones that had bands of the appropriate molecular weight were sequence verified.

4.5.2 | Bacterial test expressions

Chemically competent BL21(DE3) *E. coli* were transformed with either a combination of pCOLI_A_Strep_Rfa1, pCOLI_K_His_Rfa2, and pCOLI_S_Rfa3 (co-transformation) OR pCOLI_G2_RPA (multicistronic assembly). Then, 25 mL LB shake cultures of *E. coli* were grown in the presence of all appropriate antibiotics at 37°C. As the culture reached an OD₆₀₀ of 0.6 IPTG was added to a final concentration of 500 μ M, for a 3-hr induction. Cells were harvested, and resuspended in lysis buffer (50 mM Na-HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM MgCl₂, 2 mM BME, 1 mM AEBSF, 2.5 units/ml benzonase). Resuspended cells were then broken using sonication, and the lysate was cleared by ultracentrifugation. The cleared lysate was subject to affinity purification using Strep-Tactin XT resin (IBA), according to the manufacturer's instructions. The resin was subject to several washes with ice-cold lysis buffer, before elution with lysis buffer supplemented with biotin. Fractions from the expression/purification were analyzed using SDS-PAGE stained with InstantBlue (Sigma). Western blotting was carried out using standard laboratory protocols, using the anti-PentaHis (Qiagen) and anti-Strep II (Abcam ab76949) as primary antibodies and HRP conjugated anti-mouse or anti-rabbit secondary antibodies (Merck).

ACKNOWLEDGMENTS

We thank the members of the Weir Lab for extensive testing of the system, and for comments on the manuscript.

Work in the Weir Lab is funded by the Max Planck Society. The plasmids will be available to academic labs via AddGene.

AUTHOR CONTRIBUTIONS

Veronika Altmannova: Data curation; investigation. **Andreas Blaha:** Data curation; investigation. **Susanne Astrinidis:** Data curation; investigation. **Heidi Reichle:** Data curation; investigation. **John Weir:** Conceptualization; formal analysis; funding acquisition; supervision; writing-original draft.

ORCID

John R. Weir  <https://orcid.org/0000-0002-6904-0284>

REFERENCES

- Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front Microbiol.* 2014;5:172.
- Berrow NS, Alderton D, Sainsbury S, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* 2007;35:e45.
- Bieniossek C, Imasaki T, Takagi Y, Berger I. MultiBac: Expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci.* 2012;37:49–57.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods.* 2009;6(5):343–345. <http://dx.doi.org/10.1038/nmeth.1318>
- Weissmann F, Petzold G, VanderLinden R, et al. biGBac enables rapid gene assembly for the expression of large multi-subunit protein complexes. *Proc Natl Acad Sci U S A.* 2016;113: E2564–E2569.
- Weissmann F, Peters J-M. Expressing multi-subunit complexes using biGBac. *Methods Mol Biol.* 2018;1764: 329–343.
- Tan S, Kern RC, Selleck W. The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expr Purif.* 2005;40:385–395.
- Raran-Kurussi S, Tzs J, Cherry S, Tropea JE, Waugh DS. Differential temperature dependence of tobacco etch virus and rhinovirus 3C proteases. *Anal Biochem.* 2013;436:142–144.
- Som T, Tomizawa J. Origin of replication of *Escherichia coli* plasmid RSF 1030. *Mol Gen Genet.* 1982;187:375–383.
- Vijayachandran LS, Viola C, Garzoni F, et al. Robots, pipelines, polyproteins: Enabling multiprotein expression in prokaryotic and eukaryotic cells. *J Struct Biol.* 2011;175:198–208.
- Nijkamp HJJ, de Lang R, Stuitje AR, van den Elzen PJ, Veltkamp E, van Putten AJ. The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13. *Plasmid.* 1986;16: 135–160.
- Wold MS. Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem.* 1997;66:61–92.
- Chen R, Wold MS. Replication protein a: Singlestranded DNA's first responder: Dynamic DNA-interactions allow replication protein a to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays.* 2014;36: 1156–1161.
- Brill SJ, Stillman B. Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase. *Genes & Development.* 1991;5(9): 1589–1600. <http://dx.doi.org/10.1101/gad.5.9.1589>
- Sibenaller ZA, Sorensen BR, Wold MS. The 32- and 14-Kilodalton Subunits of Replication Protein A Are Responsible for Species-Specific Interactions with Single-Stranded DNA†. *Biochemistry.* 1998;37(36):12496–12506. <http://dx.doi.org/10.1021/bi981110+>
- Scholz J, Besir H, Strasser C, Suppmann S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnology.* 2013;13(1):12. <http://dx.doi.org/10.1186/1472-6750-13-12>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Altmannova V, Blaha A, Astrinidis S, Reichle H, Weir JR. InteBac: An integrated bacterial and baculovirus expression vector suite. *Protein Science.* 2021;30:108–114. <https://doi.org/10.1002/pro.3957>