bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # Large protein complex production using the SmartBac System -
2 Strategies and Applications

3 Yujia Zhai [1, 2, *], Danyang Zhang [1, 2], Leiye Yu [1, 2], Fang Sun [1, 2], Fei Sun [1,2,3, *]

4 [1] National Key Laboratory of Biomacromolecules, CAS Center for Excellence in
5 Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences,
6 Beijing 100101, China.

7 [2] School of Life Sciences, University of Chinese Academy of Sciences, Beijing,
8 China.

9 [3] Center for Biological Imaging, Core Facilities for Protein Science, Institute of
10 Biophysics, CAS, Beijing, China.

11 [*] Correspondence should be addressed to Y.Z. (yujia@ibp.ac.cn) or F.S.
12 (feisun@ibp.ac.cn)

13 **Keywords**

14 Baculovirus system; Cre recombination; Gibson assembly; Vectors; Protein
15 complex production.

16

17

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **ABSTRACT**

2 Recent revolution of cryo-electron microscopy has opened a new door to
3 solve high-resolution structures of macromolecule complexes without
4 crystallization while how to efficiently obtain homogenous macromolecule
5 complex sample is therefore becoming a bottleneck. Here we report SmartBac, an
6 easy and versatile system for constructing large-sized transfer plasmids used to
7 generate recombinant baculoviruses that express large multiprotein complexes in
8 insect cells. The SmartBac system integrates the univector plasmid-fusion system,
9 Gibson assembly method and polyprotein strategy to construct the final transfer
10 plasmids. The fluorescent proteins are designed to be co-expressed with
11 recombinant proteins to monitor transfection and expression efficiencies. A
12 scheme of screening an optimal tagged subunit for effective purification is
13 provided. Six large multiprotein complexes including the human exocyst complex
14 and dynactin complex were successfully expressed, suggesting a great potential
15 of SmartBac for its wide application in the future structural biology study.

16

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **INTRODUCTION**

2 With the rapid development of single-particle cryo-electron microscopy
3 (cryo-EM), more and more macromolecular machineries structures, e.g.
4 spliceosome[1,2], ryanodine receptor[3-5], anaphase promoting complex[6], light
5 harvest complex[7] and mitochondrial respirasome[8,9], have been solved to
6 near-atomic resolution, which have been waiting for many years. Since cryo-EM
7 does not need crystals, many macromolecular complexes that are difficult to be
8 crystallized are now ready for structural studies; thus, more and more structural
9 biology laboratories are becoming interested in studying the structures of large
10 macromolecular complexes. However, how to obtain enough highly purified
11 specimen suitable for cryo-EM is therefore becoming a new bottleneck, which has
12 restricted the wide application of cryo-EM technology.

13 One method for obtaining large protein complexes is to extract them from
14 biological tissues. However, this method is not very suitable for low-abundance
15 samples. In addition to consuming large amounts of reagents, protein extraction
16 usually yields samples with low yield and purity. And even for high-abundance
17 samples, it is still difficult to prepare and purify mutant proteins for functional
18 studies. Thus, recombinant expression of protein complexes is more commonly
19 preferred.

20 The baculovirus expression system (BVES) is a powerful tool for recombinant
21 protein production[10] because it is safe, high expression levels can be achieved,
22 and post-translational modifications can be incorporated. The most common
23 baculovirus used for gene expression is AcMNPV (*Autographa californica* multiple
24 nuclear polyhedrosis virus), which has a large, circular double-stranded DNA
25 genome (about 130kb) that can accommodate very large exogenous DNA

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fragments [10]. However, it is not easy to introduce foreign genes by conventional molecular cloning methods due to the large size of AcMNPV genome. Therefore, researchers have modified the AcMNPV genome to allow effective foreign gene insertion by site–specific transposition or homogenous recombination[11,12]. The widely-used Bac to Bac baculovirus expression system (Invitrogen Co.) is an example of the successful use of this approach.

Three strategies are commonly used to overexpress multiprotein complexes in insect cells. In the first strategy insect cells are infected with multiple types of baculoviruses, each of which carries one or two gene expression cassettes (GECs). This strategy, which involves molecular cloning, is relatively simple and has been successfully applied to the Bac to Bac system by many research groups[13-18]. However, when multiple types of baculoviruses are used, the total number of infectious viruses added to the expression culture should be comparable to the total number added when the culture is infected with a single baculovirus. This will inevitably lead to a lower protein yield. In addition, the expression levels of the individual subunits are often imbalanced, which can result in improper complex assembly.

The second strategy used to express multiprotein complexes is to construct a transfer plasmid carrying multiple GECs. The commercial pFastbac-Dual vector features two promoters for expression of two proteins simultaneously. Similar triple or quadruple expression vectors have also been built using traditional molecular cloning methods [19]. The MultiBac system generates multi-GEC donor and acceptor vectors from junior plasmids carrying individual GECs by homing endonuclease-based multiplication module [20]. Then the final transfer plasmid is produced by Cre-mediated recombination between the donor and acceptor vectors [20]. Yet the problem of unequal subunit stoichiometry still exists.

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 The third strategy is the polyprotein strategy that has been used by
2 coronaviruses to produce multiple functional nonstructural proteins (nsps), which
3 are involved in the formation of the replicase-transcription complex (RTC) that
4 mediates viral replication and transcription[21]. The nsps are encoded in
5 open-reading frame 1a (ORF1a) and ORF1b and are synthesized initially as two
6 large polyproteins, pp1a and pp1ab. During or after synthesis, these poliproteins
7 are cleaved by virus-encoded proteases into 16 nsps[22]. These nsps, together with
8 other viral proteins and, possibly, cellular proteins, assemble into
9 membrane-bound RTC[23]. This strategy has been exploited to produce protein
10 complexes in the baculovirus system in recent years[24,25].By this strategy, idividual
11 subunits are separated by protease cleavage sites and expressed as a long
12 polyprotein. In vivo processing of the polyprotein allows the proper assembly of
13 the multi-subunit complex. This method is very good for balancing expression
14 levels and achieving the correct subunit stoichiometry[24]. But when expressing
15 high molecular weight multiprotein complexes, the DNA fragment encoding the
16 polyprotein is very long. It is usually not easy to build such large transfer plasmids
17 in the average laboratory. In addition, with increasing gene length, gene synthesis
18 becomes more expensive and time-consuming. Other potential problems include
19 instability of the recombinant baculovirus due to the large foreign gene insertion
20 and inefficient virus amplification.

21 There are several other considerations that need to be taken into account
22 when using BVES to express large multiprotein complexes. First of all, the
23 transfer vectors carrying genes for multiple subunits need to be designed so that
24 molecular cloning can be easily accomplished. Because the difficulty of molecular
25 cloning increases with the size of the target construct, transfer vectors that allow
26 efficient selection for positive transformants are necessary. Second, one tagged

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 subunit protein is often used to purify entire protein complexes; however, due to a
2 lack of prior knowledge, sometimes we have to screen for the optimal tagged
3 subunit that allows the best purification of the entire complex. A smart
4 experimental scheme is needed to avoid exerting too much effort in building these
5 screening vectors. Third, it is also important to quickly determine whether protein
6 expression is sufficient and virus amplification is successful because insect
7 expression systems are more time-consuming compared with *Escherichia coli*
8 expression systems. The sooner these problems are identified, the faster they can
9 be solved.

10 To overcome these problems in expressing recombinant multisubunit proteins,
11 we have developed SmartBac, a simple and versatile vector system, which
12 combines the advantages of the three strategies described above. The SmartBac
13 vectors are optimized for building large constructs, and by using these vectors,
14 large protein complexes with an intricate subunit composition can be conveniently
15 expressed. The addition of a LacZ-alpha cassette allows positive recombinants
16 with large DNA inserts to be easily selected by blue-white screening. A univector
17 plasmid-fusion system (UPS) strategy has been incorporated in our vector
18 system[26], so that final construction of large transfer plasmids can be realized by
19 Cre-loxP site-specific recombination between donor and acceptor vectors. This
20 approach makes the preparation of large plasmids easier. To simplify vector
21 construction and obtain more homogenous samples, the polyprotein strategy and
22 Gibson assembly were used to construct the transfer plasmid. The co-expression
23 of EGFP and tagRFP in the SmartBac system provides real-time visualization of
24 transfection and expression. In addition, the SmartBac system can be used to
25 conveniently determine the best tagged subunit that provides the best purification
26 of the entire protein complex. Using the SmartBac system, we have expressed

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 many large multiprotein complexes, including the human exocyst complex and
2 dyncatin complex. This system can also be applied to bacteria, yeast, or
3 mammalian cells with minor modifications. We expect the SmartBac system will
4 aid structural and functional studies of large multiprotein complexes in the future.

5 **RESUTLS**

6 **SmartBac Vectors**

7 In order to overcome the difficulties in building large plasmids with
8 conventional cloning methods, we used the broadly applicable UPS strategy [26] to
9 set up our SmartBac vector system. Briefly, this strategy uses Cre–loxP
10 site-specific recombination to catalyze fusion between the univector (donor vector)
11 and host vector (acceptor vector). The kanamycin-resistant donor has a
12 conditional R6Kγ origin of replication that allows its propagation only in bacterial
13 hosts expressing the pir gene, which encodes the essential replication protein
14 π[27,28]. Selection for the UPS recombination products is achieved by selecting for
15 kanamycin resistance (Kan^R) after transformation into a pir- strain; the Kan^R gene
16 in the donor vector can be expressed in a pir- background only when covalently
17 linked to an acceptor vector that has a functional origin of replication (oriColE1) [26].
18 This strategy has been successfully used in the MultiBac system[20] and is
19 advantageous for the preparation of large plasmids and their mutants due to the
20 relatively small sizes of the donor and acceptor vectors.

21 We designed four acceptor plasmids (4V1G, 4V1R, 5V1TG and 5V1TR) and
22 two donor plasmids (4V2G and 4V2R) for use in the SmartBac system (**Figure 1,**
23 see also **Supplementary Materials and Methods S1**). The acceptors can
24 recombine with the donors via Cre-LoxP site-specific recombination. The
25 acceptors harbor a p15A origin of replication that allows propagation in common

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 cloning strains of *E. coli* at low copy number, which is better for the stability of
2 large plasmids. The acceptors also contain resistance markers for ampicillin and
3 gentamycin and flanking mini-Tn7 elements for the generation of recombinant
4 baculoviruses.

5 Transgene expression in infected insect cells is driven by the baculovirus late
6 p6.9 promoter [29]. Compared with the routinely used very late polyhedrin promoter,
7 the p6.9 promoter drives expression at earlier stages of infection when cells are
8 more likely to be in good condition and therefore the aggregation of expressed
9 foreign proteins may be avoided[30-32].

10 The 4V1 acceptor vectors (4V1G and 4V1R) carry an N-terminal HA-tagged
11 TEV protease coding sequence followed by the TEV protease cleavage site (TCS)
12 and a Twin-Strep tag coding sequence followed by a recognition site for
13 enterokinase. Between multiple cloning site (MCS) 1 and 2, there is a LacZ-alpha
14 expression cassette, which allows blue/white selection of recombinant clones.
15 Downstream of MCS2, there is another TCS and an EGFP (4V1G) or TagRFP
16 (4V1R) coding sequence. The fluorescent and target proteins can be expressed
17 as a single ORF. By observing the fluorescence of infected cells, we can easily
18 determine whether the target protein has been expressed.

19 In the 5V1T acceptor vectors (5V1TG and 5V1TR), different from 4V1
20 acceptor vectors, the TEV protease and EGFP (5V1TG) or tagRFP (5V1TR)
21 coding sequences are fused and expressed as a GP64 promoter-driven ORF.

22 The 4V2 donor vectors (4V2G and 4V2R) carry an N-terminal 10×His coding
23 sequence followed by a recognition site for enterokinase. Both vectors contain a
24 kanamycin resistance marker. The screening region is composed of a high-copy
25 PUC origin of replication and a LacZ-alpha expression cassette, flanked by MCS1
26 and MCS2. Downstream of MCS2, there is a TCS and a fluorescent protein

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 (EGFP in 4V2G and tagRFP in 4V2R) coding sequence. The expression of the
2 target protein is driven by the very late p10 promoter. The 4V2 vectors also
3 contain the conditional origin of replication, R6Kγ. Once the screening region is
4 replaced by a foreign gene, the donor vector only contains the R6Kγ origin and
5 can only be propagated in *E. coli* strains with the pir+ genotype.

6 There are several single restriction sites located on both sides of the p6.9 and
7 p10 promoter regions in the 4V1/5V1 acceptor and 4V2 donor vectors,
8 respectively, so that they can be replaced by other baculovirus promoters, if
9 needed.

10 **Schemes for the expression of large multiprotein complexes using the**
11 **SmartBac system**

12 The SmartBac system was designed for easier and faster expression of large
13 multiprotein complexes in insect cells. A variety of experimental schemes could be
14 applied to produce the final transfer plasmids from the SmartBac vectors. Here we
15 just present two schemes to exemplify how to use the SmartBac vectors. In the
16 example shown in **Figure 2a**, a vector is designed to express a multiprotein
17 complex composed of eight different subunits (e.g. subunits A, B, C, D, E, F, G
18 and H) in insect cells. If the molecular weight of the multiprotein complex is less
19 than 600 kDa, we propose using Scheme 1 (**Figure 2a**). The eight subunits are
20 divided into two groups so that the sum of the lengths of the genes in one group is
21 as similar as possible to the other group. Then for each group of genes, a fusion
22 DNA fragment (ABCD and EFGH) with TCS coding sequences separating the
23 adjacent genes is designed.

24 Next, the long ABCD and EFGH fragments are further divided into two short
25 DNA fragments AB and CD and EF and GH, which can be obtained easily by

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

overlapping PCR (**Figure 2b**). To avoid unnecessary trouble in overlapping PCR, the TCS cleavage sites described above should be coded by multiple degenerate sequences (**Supplementary Materials and Methods S2**). Then, fragments AB and CD are assembled with a linearized SmartBac RFP-expressing acceptor plasmid (we use 4V1R in **Figure 2b**, but 5V1TR can also be used) utilizing a Gibson assembly reaction [33]. Fragments EF and GH are also assembled with a linearized SmartBac GFP-expressing donor plasmid (4V2G) using the same method. The positive recombinants can be easily selected by blue-white screening. Finally, the acceptor-4V1R-ABCD and donor-4V2G-EFGH vectors are recombined via Cre-LoxP site-specific recombination to generate the final transfer plasmid ABCD-EFGH. After transforming this plasmid into DH10Bac competent cells, recombinant bacmid will be obtained. This bacmid will be transfected into insect cells to produce high-titer baculovirus BV-ABCD-EFGH used to express the target eight-subunit complex.

If the molecular weight of the multiprotein complex is greater than 600 kDa, the size of the final transfer plasmid constructed using Scheme 1 will be larger than 25kb. It is usually not easy to build such a large plasmid without experience. And even if the construction is successful, the multiprotein complex may fail to be expressed in insect cells. This is because the recombinant bacmid generated from the large transfer plasmid is prone to display an intrinsic genetic instability due to the large foreign gene insertion [30,34]. Spontaneous deletion of the foreign gene insertion may occur during the amplification of P2 virus (our unpublished data). So, in this case Scheme 2 should be used. As shown in **Figure 2c**, Fragments A and B are assembled with linearized 5V1TR, and fragments C and D are fused with linearized 4V2 vector. A stop codon has been inserted upstream of the coding sequences by PCR so that the fluorescent protein in 4V2G and 4V2R

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 will not be expressed. The same method is used to clone fragments E, F, G and H.
2 Next, two different final transfer plasmids ABCD (RFP) and EFGH (GFP) are built.
3 These plasmids will produce two types of recombinant baculoviruses, one
4 expressing subunits A, B, C and D and RFP, and the other expressing subunits E,
5 F, G and H and GFP. Insect cells co-infected with both baculoviruses will produce
6 the entire protein complex, with the appearance of tagRFP and EGFP
7 fluorescence indicating successful expression of the target protein complex.

8 The SmartBac acceptors carry an optional N-terminal Twin-Step-tag
9 sequence, and the donors carry an optional N-terminal His-Tag sequence. Either
10 tag can be fused to a target subunit and used as a handle to purify the target
11 subunit along with its associated subunits. If the biochemical properties of the
12 protein complex are known, it is easy to determine which subunit is the most
13 suitable to fuse with the affinity tag. However, when previous knowledge is limited,
14 it may be hard to pick the appropriate subunit because different affinity-tagged
15 subunits often differ in their effectiveness in purifying the entire complex. Imagine
16 that we are expressing an eight-subunit complex and are not sure which subunit is
17 suitable for labeling. If we use a classical “Trial and Error” approach and construct
18 multiple large final transfer plasmids, the workload will be very high. In addition,
19 the complicated clone scheme is often confusing. To solve this problem, we
20 propose a simple and universal scheme (Scheme 3, see **Figure 3**). Two large
21 final transfer plasmids ABCD (RFP) and EFGH (GFP) are built according to
22 Scheme 2 but where none of the eight subunits are labeled with affinity tags
23 (Figure 3a). An additional eight smaller transfer plasmids (from V1-TSA to V1-TSH)
24 based on one acceptor vector (4V1G, 4V1R, 5V1TG or 5V1TR) are constructed,
25 each expressing an N-terminal Twin-Strep-tagged subunit. A total of ten
26 recombinant baculoviruses are obtained, including BV-ABCD (RFP), BV-EFGH

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 (GFP) and BV-TSn (where n ranges from A to H) (**Figure 3b**). The insect cells are
2 co-infected with three types of baculoviruses, BV-ABCD (RFP), BV-EFGH (GFP)
3 and one type of BV-TSn. The baculovirus combinations used for screening are
4 shown in **Figure 3c**. After purification, we know affinity-tagged subunit H results in
5 the best purification of the entire complex. To increase yield and obtain a more
6 homogenous sample, a new intermediate vector containing tagged subunit H,
7 EFG-TSH (GFP), is built (**Figure 3d**). The resulting new recombinant baculovirus,
8 BV-EFG-TSH (GFP), along with the existing recombinant baculovirus, BV-ABCD
9 (RFP), are used to express the multiprotein complex, and real-time infection and
10 expression is monitored by observing the fluorescence of co-expressed EGFP
11 and tagRFP.

12 **Multiprotein complexes expressed using the SmartBac system**

13 To test the SmartBac system, we expressed the human exocyst complex in
14 insect cells. The exocyst complex is responsible for tethering secretory vesicles to
15 the plasma membrane in preparation for soluble N-ethylmaleimide-sensitive factor
16 (NSF) attachment protein receptor (SNARE) mediated membrane fusion [35]. The
17 human exocyst complex contains eight evolutionary conserved subunits—EXOC1
18 (102 kDa), EXOC2 (104 kDa), EXOC3 (86 kDa), EXOC4 (110 kDa), EXOC5 (82
19 kDa), EXOC6 (94 kDa), EXOC7 (78 kDa) and EXOC8 (82 kDa). Because the
20 published literature does not provide information about which subunit is the most
21 suitable for complex purification, we used Scheme 3 to screen the target subunits.
22 All of the vectors we built for exocyst expression are shown in **Table 1** (see also
23 **Supplementary Materials and Methods S3**). First, we constructed two types of
24 recombinant baculoviruses, BV-E1547 and BV-E2863, to express the eight
25 subunits without any tags according to Scheme 2. These baculoviruses also
26 expressed tagRFP and EGFP, respectively, which allowed us to conveniently

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 determine whether virus infection and protein expression was successful (**Figure**
2 **4a**). We also produced eight additional types of recombinant baculoviruses, each
3 expressing an individual subunit with an N-terminal Twin-Strep tag (BV-SE1 to
4 BV-SE8). Then we co-infected insect cells with BV-E1547, BV-E2863 and a
5 baculovirus expressing a single affinity tagged-subunit (BV-SE1 to BV-SE8). The
6 best purification of the entire exocyst complex was achieved using tagged EXOC5
7 (BV-SE5) (**Figure 4b**). Then we constructed a new donor vector 4V2-E1S5 which
8 contains EXOC1 and N-terminal Twin Strep-tagged EXOC5. Recombination
9 between 4V2-E1S5 and 5V1TR-E47 (contains EXOC4 and EXOC7) produced a
10 new final transfer plasmid E1S547 (**Table 1**), from which recombinant baculovirus
11 BV-E1S547 was obtained. Insect cells were co-infected with BV-E1S547 and
12 BV-2863. After one-step strep-affinity purification, the entire exocyst complex with
13 high purity was obtained (**Figure 4c**). The tethering activity of this purified exocyst
14 complex was determined via In vitro liposome reconstruction assay (data is not
15 shown here). Negative-staining electron microscopy (nsEM) of the sample
16 revealed homogenous rod-like particles (**Figure 4d**). Preliminary 2D classification
17 of nsEM images (**Figure 4e**) and 3D reconstruction (**Figure 4f**) indicate that the
18 human exocyst complex exhibits a similar dimension and shape with the extracted
19 exocyst complex from yeast [36]. The detailed information about primer design,
20 molecular cloning, cell transfection, protein expression and purification, and
21 electron microscopy is described in **Supplementary Materials and Methods S4**.

22 We also reconstituted the human dynactin complex using the SmartBac
23 system. Dynactin is a multiprotein complex that works with cytoplasmic dynein to
24 transport cargo along microtubules. It is a large complex of approximately 1.2
25 MDa composed of 23 subunits corresponding to 11 different types of proteins [37].
26 Dynactin is built around a short actin-like filament composed of Arp1 (43 kDa, 8

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 copies) and β-actin (42 kDa, 1 copy). The barbed end and the pointed end of this
2 filament are capped by the CapZα-CapZβ complex (33 kDa, 31 kDa) and the
3 Arp11-p25-p27-p62 complex (46 kDa, 20kDa, 21 kDa, 52 kDa) respectively. The
4 shoulder complex, which is made up of p50 (45 kDa, 4 copies), p24 (21 kDa, 2
5 copies) and p150/p135 (142 kDa/127 kDa, 2 copies), is positioned toward the
6 barbed end of the Arp1 filament [38]. As shown in **Table 2**, three types of vectors
7 were used to express the 11 subunits of the dynactin complex, and the N-terminal
8 Twin-Strep tag on p135 was used to purify the whole complex. The shoulder
9 complex proteins, p135, p50 and p24. were expressed by BV-M5, which was
10 generated from the plasmid 5V1TG-M5. The final transfer plasmid AB was
11 obtained through recombination of the acceptor plasmid 5V1TR-B and donor
12 plasmid 4V2-A. Plasmid AB was then used to produce the recombinant
13 baculovirus BV-AB expressing the other eight dynactin subunits. Insect cells were
14 co-infected with BV-M5 and BV-AB to express the entire dynactin complex. After
15 one-step strep-affinity purification, the dynactin complex was purified well with
16 rational stoichiometry of its subunits (**Figure 4g**). After glycerol density gradient
17 ultracentrifugation, the further purified dynactin complex exhibited a single visible
18 band on a native gel (**Figure 4h**), suggesting a high homogeneity of the specimen.
19 This band was excised and subjected to mass spectrometry and all 11 human
20 dynactin complex subunits were identified (our unpublished data). This recombant
21 human dynactin complex sample was further investigated by nsEM and
22 subseqeuent 2D image classification (**Figure 4i**), showing a rod-like particle with
23 a shoulder at one end, which is similar to that of the endogenous dynactin
24 complex purified from pig brains [38].

25 Besides the human exocyst and dynactin complexes, we also successfully
26 expressed many other protein complexes using the SmartBac system (**Figure 4j**).

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These include the human COPI complex (7 subunits, 558 kDa) [39], cytoplasmic Dynein complex (12 subunits, 1380 kDa) [40], CSN complex (8 subunits, 343 kDa) [41] and SCF complex (5 subunits, 180 kDa) [42]. The recombinant COPI complex sample has been used to study the structure of coatomer in its soluble form [39]. Our results indicate that the SmartBac system can be used to express a wide range of large multiprotein complexes.

DISCUSSIONS

Obtaining large multiprotein complexes through recombinant expression has always been challenging for researchers who need a sufficient quantity of high-purity protein for structural or biochemical studies. The key to successful protein production using the baculovirus expression system is the construction of the final transfer plasmid containing multiple protein subunits. The MultiBac system uses polycistronic vectors carrying multiple GECs for the expression of multiprotein complexes [20]. This expression strategy requires junior plasmids containing only one GEC to be constructed first. Then several rounds of plasmid construction have to be done to obtain multi-GEC-contained donor and acceptor vectors. The final transfer plasmid carrying all GECs is produced by Cre-LoxP recombination between the donor and acceptor vectors. MultiBac has been proven powerful in generating multiprotein complexes [24,43], especially when robotic support is available. But in ordinary laboratories without robotics, the first two procedures require a great deal of time. And as more GECs are added to a single vector, plasmid construction becomes more difficult, due to the increasing size of the plasmid and the lack of an effective screening method for large positive recombinants.

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 To simplify the process for constructing vectors that express multiple protein
2 subunits, we developed the SmartBac system. We mimicked the polyprotein
3 production strategy of coronavirus to realize the expression of multiple subunits.
4 Although this strategy has been discussed and applied in some laboratories
5 [25,44,45], a specialized vector system and standardized procedures are not available.
6 By generating one GEC expressing a long polyprotein composed of multiple
7 subunits in donor and acceptor vectors, the SmartBac system does not require
8 the construction of numerous junior vectors. Large numbers of gene fragments
9 with overlapping sequences can be produced rapidly by PCR and directly used for
10 Gibson assembly with linearized vectors. Positive recombinants can be easily
11 selected by blue-white screening and then donor and acceptor vectors carrying a
12 long GEC can be combined by Cre-LoxP recombination to produce the final
13 transfer plasmid.

14 To ensure a high success rate, we fused multiple gene fragments together by
15 overlapping PCR so no more than three DNA fragments (including the linearized
16 vector) were included in a single Gibson assembly reaction. The size of the final
17 transfer plasmid produced using the schemes we have provided is usually less
18 than 25 kb, so either chemical transformation (Trans2blue or XL10-Gold
19 ultracompetent cells) or electroporation can be used.

20 To increase the stability of the large final transfer plasmid propagated in *E. coli*,
21 we included a p15A replication origin of low copy number in the acceptor vectors
22 and cultured the bacteria at 30℃. The restriction sites flanking the promoter
23 region allow for promoter exchange if needed (eg. To optimize expression levels).

24 To monitor the expression of target proteins, we added the most commonly
25 used fluorescent proteins EGFP and tagRFP genes to the SmartBac vectors, so

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 that observation can be easily performed with a basic fluorescence microscope.
2 4V1 and 5V1 vectors provide two different expression modes for these protein
3 markers. 4V1 vectors use a single GEC to express these fluorescent proteins,
4 TEV protease and the target protein subunits. However, for some constructions,
5 the fluorescent protein was insufficiently cleaved from its upstream expressed
6 subunit (our unpublished data). It had a bad effect on the assembly of the entire
7 complex. In this case, 5V1 vectors can be selected because they use one GEC to
8 express the target subunits, and another GEC to express the fluorescent protein
9 and TEV protease. We have provided two easy-to-use schemes to guide the
10 design of large transfer plasmids containing multiple genes as well as a verified
11 scheme to screen an optimal tagged subunit for complex purification.

12 With the development of cryo-EM, more and more research groups are
13 carrying out structural and functional studies of multiprotein complexes.
14 Recombinant expression is a promising method for obtaining sufficient quantities
15 of high-purity samples, and, we expect that the SmartBac system will allow more
16 researchers to successfully express and purify large multiprotein complexes.

17

18 **COMPETING INTERESTS**

19 Parts of this study (SmartBac system) has been submitted to apply a Chinese
20 patent for invention with the application number of 201610248592.8.

21 **AUTHORS' CONTRIBUTIONS**

22 Fei Sun initiated and supervised the project. YZ designed all the SmartBac
23 systems including vectors and application strategies. YZ performed all the
24 experiments of molecular cloning and expression constructs production. YZ, DZ,

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LY and Fang Sun performed protein complex purification and preliminary electron microscopic characterization. YZ and Fei Sun wrote the manuscript.

ACKNOWLEDMENTS

We would be grateful to Ping Shan and Ruigang Su (Fei Sun's lab) for their help on lab maintenance. We would like to thank Xiang Ding and Mengmeing Zhang from Laboratory of Proteomics, Core Facilities for Protein Science, at the Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS), for their help with mass spectrometry analysis. We would like to thank Center for Biological Imaging (CBI, http://cbi.ibp.ac.cn), IBP, CAS for the electron microscopy work.

This work was supported by grants from the Strategic Priority Research Program of Chinese Academy of Sciences (XDB08030202) to Fei Sun, the National Basic Research Program (973 Program) of Ministry of Science and Technology of China (2014CB910700) to Fei Sun, and National Natural Science Foundation of China (31771566) to YZ.

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **MATERIALS AND METHODS**

2 **Vector construction**

3 A portion of the 4V1 vectors was derived from pFastbacDUAL (Invitrogen,
4 USA). 4V2 vectors and all other portions of the 4V1 vectors were synthesized by
5 Genewiz, China. The DNA fragments were fused together via Gibson assembly
6 (E2611, NEB, England) to generate the 4V1G, 4V1R, 4V2G and 4V2R vectors.
7 Then, using 4V1 vectors, 5V1 vectors were generated by Gibson assembly and
8 other classical molecular cloning methods. The sequences of the SmartBac
9 vectors are shown in **Supplementary Materials and Methods S1**. We
10 recommend using SnapGene Viewer (http://www.snapgene.com/) to view the
11 plasmid maps.

12 **Gibson assembly reactions**

13 The linearized plasmid fragments were obtained by restriction enzyme
14 digestion or PCR using Q5 High-fidelity DNA Polymerase (M0492, NEB, England).
15 DNA fragments to be inserted into the SmartBac vectors were amplified by PCR
16 to produce the appropriate overlaps. The overlapping primers were designed
17 according to NEB instruction manual for E2611. Assembly was done in a 15-20 μl
18 reaction volume with 0.2-0.3 pmols each DNA fragment. Samples were incubated
19 in a thermocycler at 50°C for 60 minutes.

20 **Blue-white selection of positive recombinants**

21 To perform blue-white selection, 2.5 μl assembled product was added to 100 μl
22 chemically competent cells. The 4V1- and 5V1-based constructs were
23 transformed into chemically competent Mach1™-T1R (Invitrogen) or DH5alpha or
24 Trans2-blue (TransGen Biotech, China) cells. After 1 h recovery in SOC medium
25 at 37°C, cells were plated onto LB agar plates containing 100 μg/ml ampicillin, 40

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 μg/ml IPTG and 100 μg/ml Bluo-gal. The 4V2-based constructs were transformed
2 into chemically competent GT115 cells (InvivoGen, USA), and then cells were
3 plated onto LB agar plates containing 50 μg/ml kanamycin, 40 μg/ml IPTG and
4 100 μg/ml Bluo-gal. Single white colonies were picked and grown in 5 ml LB
5 medium with the proper antibiotics for further plasmid extraction and PCR
6 analysis. The positive recombinants were sequenced at BioSune, China.

7 **Production of the final transfer plasmid by Cre-LoxP Recombination**

8 The donor and acceptor vectors (0.1 pmols each) were mixed with 1 μl Cre
9 recombinase (M0298, NEB) in a 20 μl reaction and incubated at 30°C for 1 h. Ten
10 microliters of the reaction mixture were added to 100 μl chemically competent
11 Trans2-blue cells. After heat-shock at 30°C for 30 s, 500 μl SOC medium was
12 added, and the suspension was incubated at 37°C for 1 h with shaking (if the size
13 of the recombined vector was larger than 15 kb, the suspension was incubated at
14 30°C for 4 hrs). The cell suspension was plated on LB agar plates containing 50
15 μg/ml kanamycin and 100 μg/ml ampicillin. The plates were incubated at 37°C
16 overnight (or 30°C for 24 hrs). Positive colonies were verified by PCR using the
17 primers Loxp-F (5'-CCACTGCGCCGTTACCAC-3') and Loxp-R
18 (5'-GCCGGTATGTACAGGAAG-3'). A 375 bp PCR product was amplified from
19 positive clones. The final transfer plasmids were extracted from the positive
20 clones.

21 **Production of Recombinant Baculovirus**

22 Chemically competent DH10Bac cells were transformed with the final transfer
23 plasmid according to the Bac to Bac manual instructions (Invitrogen). For
24 transformation of large plasmids, the transformation mixture was incubated at
25 30°C with shaking for 8-12 hrs and plates were incubated at 30°C for more than

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48 hrs. Single white colonies (3-4) were innoculated into 5 ml LB medium containing 50 μg/ml kanamycin, 14 μg/ml gentamicin, and 10 μg/ml tetracycline. Recombinant bacmids were extracted and verified by PCR amplification with three pairs of primers (Tn7R:5'- GTTTTCCCAGTCACGAC-3' and 5'-AAGTTTGAGCAGCCGCGTAG-3'; Tn7L:5'- 5'-CAGGAAACAGCTATGAC-3' and 5'-ACCTCCCCCTGAACCTGAAA-3'; Empty: 5'-GTTTTCCCAGTCACGAC-3' and M13 Reverse :5'-CAGGAAACAGCTATGAC-3'). Using the "Tn7R" and "Tn7L" primer pairs, PCR products of 661 bp and 521 bp, respectively, are amplified from recombinant bacmids. If the recombinant bacmid is contaminated with wild-type bacmid, a PCR product of 300 bp will produced using the "Empty" primer pairs. It is recommended to verify the existence of all of the subunit genes in the recombinant bacmids by PCR if the size of final transfer plasmid is larger than 20kb.

Transfection and Virus Production in Insect cells

Transfection and Baculovirus production were done according to the Bac to Bac manual (Invitrogen, USA). Successful transfection was determined by the expression of EGFP and/or tagRFP fluorescent proteins. P2 virus was used for expression.

REFERENCES

1 Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science (New York, N.Y* **349**, 1182-1191, doi:10.1126/science.aac7629 (2015).

2 Nguyen, T. H. D. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. *Nature* **530**, 298-302, doi:10.1038/nature16940 (2016).

3 Yan, Z. *et al.* Structure of the rabbit ryanodine receptor RyR1 at near-atomic

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 resolution. *Nature* **517**, 50-55, doi:10.1038/nature14063 (2015).
2 4 des Georges, A. *et al.* Structural Basis for Gating and Activation of RyR1. *Cell*
3 **167**, 145-157 e117, doi:10.1016/j.cell.2016.08.075 (2016).
4 5 Wei, R. *et al.* Structural insights into Ca(2+)-activated long-range allosteric
5 channel gating of RyR1. *Cell Res* **26**, 977-994, doi:10.1038/cr.2016.99 (2016).
6 6 Chang, L., Zhang, Z., Yang, J., McLaughlin, S. H. & Barford, D. Atomic structure
7 of the APC/C and its mechanism of protein ubiquitination. *Nature* **522**,
8 450-454, doi:10.1038/nature14471 (2015).
9 7 Wei, X. *et al.* Structure of spinach photosystem II-LHCII supercomplex at 3.2 A
10 resolution. *Nature* **534**, 69-74, doi:10.1038/nature18020 (2016).
11 8 Gu, J. *et al.* The architecture of the mammalian respirasome. *Nature* **537**,
12 639-643, doi:10.1038/nature19359 (2016).
13 9 Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory
14 supercomplexes. *Nature* **537**, 644-648, doi:10.1038/nature19774 (2016).
15 10 Jarvis, D. L. Baculovirus-insect cell expression systems. *Methods Enzymol* **463**,
16 191-222, doi:10.1016/S0076-6879(09)63014-7 (2009).
17 11 Luckow, V. A., Lee, S. C., Barry, G. F. & Olins, P. O. Efficient generation of
18 infectious recombinant baculoviruses by site-specific transposon-mediated
19 insertion of foreign genes into a baculovirus genome propagated in
20 Escherichia coli. *J Virol* **67**, 4566-4579 (1993).
21 12 Hitchman, R. B., Possee, R. D. & King, L. A. High-throughput baculovirus
22 expression in insect cells. *Methods Mol Biol* **824**, 609-627,
23 doi:10.1007/978-1-61779-433-9_33 (2012).
24 13 Birnbaum, M. E. *et al.* Molecular architecture of the alphabeta T cell
25 receptor-CD3 complex. *Proc Natl Acad Sci U S A* **111**, 17576-17581,
26 doi:10.1073/pnas.1420936111 (2014).
27 14 Chung, Y. C. *et al.* Expression, purification and characterization of
28 enterovirus-71 virus-like particles. *World J Gastroenterol* **12**, 921-927 (2006).
29 15 Hu, Y. C., Hsu, J. T., Huang, J. H., Ho, M. S. & Ho, Y. C. Formation of
30 enterovirus-like particle aggregates by recombinant baculoviruses
31 co-expressing P1 and 3CD in insect cells. *Biotechnol Lett* **25**, 919-925 (2003).
32 16 Kee, Y. *et al.* WDR20 regulates activity of the USP12 x UAF1 deubiquitinating
33 enzyme complex. *J Biol Chem* **285**, 11252-11257,
34 doi:10.1074/jbc.M109.095141 (2010).
35 17 Wan, L. C. *et al.* Proteomic analysis of the human KEOPS complex identifies
36 C14ORF142 as a core subunit homologous to yeast Gon7. *Nucleic acids*
37 *research* **45**, 805-817, doi:10.1093/nar/gkw1181 (2017).
38 18 Yoo, H. Y., Kumagai, A., Shevchenko, A., Shevchenko, A. & Dunphy, W. G. The
39 Mre11-Rad50-Nbs1 complex mediates activation of TopBP1 by ATM. *Mol Biol*

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 *Cell* **20**, 2351-2360, doi:10.1091/mbc.E08-12-1190 (2009).
2 19 Belyaev, A. S. & Roy, P. Development of baculovirus triple and quadruple
3 expression vectors: co-expression of three or four bluetongue virus proteins
4 and the synthesis of bluetongue virus-like particles in insect cells. *Nucleic*
5 *acids research* **21**, 1219-1223 (1993).
6 20 Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for
7 heterologous multiprotein complexes. *Nat Biotechnol* **22**, 1583-1587,
8 doi:10.1038/nbt1036 (2004).
9 21 Bartlam, M., Yang, H. & Rao, Z. Structural insights into SARS coronavirus
10 proteins. *Curr Opin Struct Biol* **15**, 664-672, doi:10.1016/j.sbi.2005.10.004
11 (2005).
12 22 Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. Virus-encoded proteinases and
13 proteolytic processing in the Nidovirales. *J Gen Virol* **81**, 853-879,
14 doi:10.1099/0022-1317-81-4-853 (2000).
15 23 Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. A contemporary view of
16 coronavirus transcription. *J Virol* **81**, 20-29, doi:10.1128/JVI.01358-06
17 (2007).
18 24 Vijayachandran, L. S. *et al.* Robots, pipelines, polyproteins: enabling
19 multiprotein expression in prokaryotic and eukaryotic cells. *J Struct Biol* **175**,
20 198-208, doi:10.1016/j.jsb.2011.03.007 (2011).
21 25 Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I.
22 Multiprotein complex production in insect cells by using polyproteins.
23 *Methods Mol Biol* **1091**, 131-141, doi:10.1007/978-1-62703-691-7_8 (2014).
24 26 Liu, Q., Li, M. Z., Leibham, D., Cortez, D. & Elledge, S. J. The univector
25 plasmid-fusion system, a method for rapid construction of recombinant DNA
26 without restriction enzymes. *Curr Biol* **8**, 1300-1309 (1998).
27 27 Metcalf, W. W., Jiang, W. & Wanner, B. L. Use of the rep technique for allele
28 replacement to construct new Escherichia coli hosts for maintenance of R6K
29 gamma origin plasmids at different copy numbers. *Gene* **138**, 1-7 (1994).
30 28 Filutowicz, M. *et al.* Role of the pi initiation protein and direct nucleotide
31 sequence repeats in the regulation of plasmid R6K replication. *Basic Life Sci*
32 **30**, 125-140 (1985).
33 29 Hill-Perkins, M. S. & Possee, R. D. A baculovirus expression vector derived
34 from the basic protein promoter of Autographa californica nuclear
35 polyhedrosis virus. *J Gen Virol* **71 (Pt 4)**, 971-976,
36 doi:10.1099/0022-1317-71-4-971 (1990).
37 30 van Oers, M. M. Opportunities and challenges for the baculovirus expression
38 system. *J Invertebr Pathol* **107 Suppl**, S3-15, doi:10.1016/j.jip.2011.05.001
39 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 31 Ishiyama, S. & Ikeda, M. High-level expression and improved folding of
2 proteins by using the vp39 late promoter enhanced with homologous DNA
3 regions. *Biotechnol Lett* **32**, 1637-1647, doi:10.1007/s10529-010-0340-7
4 (2010).
5 32 Li, S. F., Wang, H. L., Hu, Z. H. & Deng, F. Genetic modification of baculovirus
6 expression vectors. *Virol Sin* **27**, 71-82, doi:10.1007/s12250-012-3236-y
7 (2012).
8 33 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several
9 hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318
10 (2009).
11 34 Pijlman, G. P., van Schijndel, J. E. & Vlak, J. M. Spontaneous excision of BAC
12 vector sequences from bacmid-derived baculovirus expression vectors upon
13 passage in insect cells. *J Gen Virol* **84**, 2669-2678, doi:10.1099/vir.0.19438-0
14 (2003).
15 35 Wu, B. & Guo, W. The Exocyst at a Glance. *J Cell Sci* **128**, 2957-2964,
16 doi:10.1242/jcs.156398 (2015).
17 36 Heider, M. R. *et al.* Subunit connectivity, assembly determinants and
18 architecture of the yeast exocyst complex. *Nat Struct Mol Biol* **23**, 59-66,
19 doi:10.1038/nsmb.3146 (2016).
20 37 Reck-Peterson, S. L. Dynactin revealed. *Nat Struct Mol Biol* **22**, 359-360,
21 doi:10.1038/nsmb.3022 (2015).
22 38 Urnavicius, L. *et al.* The structure of the dynactin complex and its interaction
23 with dynein. *Science (New York, N.Y* **347**, 1441-1446,
24 doi:10.1126/science.aaa4080 (2015).
25 39 Wang, S. *et al.* Structural characterization of coatomer in its cytosolic state.
26 *Protein Cell* **7**, 586-600, doi:10.1007/s13238-016-0296-z (2016).
27 40 Zhang, K. *et al.* Cryo-EM Reveals How Human Cytoplasmic Dynein Is
28 Auto-inhibited and Activated. *Cell* **169**, 1303-1314 e1318,
29 doi:10.1016/j.cell.2017.05.025 (2017).
30 41 Mosadeghi, R. *et al.* Structural and kinetic analysis of the COP9-Signalosome
31 activation and the cullin-RING ubiquitin ligase deneddylation cycle. *Elife* **5**,
32 doi:10.7554/eLife.12102 (2016).
33 42 Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin
34 ligase complex. *Nature* **416**, 703-709, doi:10.1038/416703a (2002).
35 43 Berger, I. *et al.* The multiBac protein complex production platform at the
36 EMBL. *J Vis Exp*, e50159, doi:10.3791/50159 (2013).
37 44 Chen, X., Pham, E. & Truong, K. TEV protease-facilitated stoichiometric
38 delivery of multiple genes using a single expression vector. *Protein Sci* **19**,
39 2379-2388, doi:10.1002/pro.518 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 45 Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the
2 research toolbox for multiprotein complexes. *Trends Biochem Sci* **37**, 49-57,
3 doi:10.1016/j.tibs.2011.10.005 (2012).

4

5

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 # FIGURE LEGENDS

2 **Figure 1. SmartBac vector maps.** The SmartBac system includes four acceptor
3 plasmids (4V1G, 4V1R, 5V1TG and 5V1TR) and two donor plasmids (4V2G and
4 4V2R). Vector maps were produced by SnapGene Software
5 (http://www.snapgene.com/).

6 **Figure 2. Schemes for the expression of large multiprotein complexes. (a)**
7 The eight-subunit protein complex to be expressed. The eight genes are divided
8 into two groups according to their sizes. Two long polyproteins are designed with
9 TEV cleavage sites separating the adjacent genes. ∑> represents the TEV
10 cleavage site. **(b)** Schematic representation of Scheme 1 for the expression of
11 multiprotein complexes with a molecular weight less than 600 kDa. Here the
12 acceptor vector 4V1R is used, but 5V1TR can also be used. **(c)** Schematic
13 representation of Scheme 2 for the expression of multiprotein complexes with a
14 molecular weight greater than 600 kDa. The fluorescent protein in the 4V2G/4V2R
15 donor vector is not expressed because a stop codon has been inserted at the end
16 of the fusion gene, which is located at the upstream of the coding sequence of the
17 fluorescent protein. The coding sequences of EGFP and tagRFP can also be
18 removed by restriction enzyme digestion.

19 **Figure 3. Screening for the best affinity-tagged subunit through co-infection**
20 **of insect cells (Scheme 3).** (**a**) Diagrams of the ten types of transfer plasmids.
21 The final transfer plasmids, ABCD (RFP) and EFGH (GFP), are generated using
22 Scheme 2, and each will express four protein subunits without affinity labels. Each
23 of the other eight transfer plasmids will express one subunit with an N-terminal
24 Twin-Strep (TS) tag. Either the 4V1 or 5V1 vector can be used here. (**b**)
25 Production of ten types of recombinant baculoviruses (RBVs). Transformation of

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 the ten types of plasmids into DH10Bac competent cells generates 10 types of
2 RBVs. (**c**) Screening baculovirus combinations to find the subunit that results in
3 the best purification. The ten types of RBVs are divided into eight groups, and
4 each group contains BV-ABCD (RFP), BV-EFGH (GFP) and one BV-TSn (where
5 n corresponds to the subunit, A to H). Insect cells are co-infected with eight
6 groups of RBVs and strep-affinity resin is used to pull down proteins bound to the
7 TS-tagged subunit. The tagged subunit that allows the best purification of the
8 whole complex is selected. In this example, subunit H is the best. (**d**) Production
9 of the multiprotein complex. Based on the screening result in (c), a new final
10 transfer plasmid EFG-TSH (GFP) is constructed, which expresses an N-terminal
11 TS-tagged subunit H. The whole protein complex will be purified from insect cells
12 co-infected with BV-ABCD (RFP) and BV-EFG-TSH (GFP).

13 **Figure 4 Examples of multiprotein complexes expressed using the**
14 **SmartBac system. (a)** Fluorescence signals for tagRFP (top) and EGFP (bottom)
15 detected from Sf9 cells transfected with BVE1S547 and BV2863 (see **Table 1**). **(b)**
16 Coomassie-stained SDS-PAGE gel of human exocyst complex purified using
17 eight different Twin-Strep tagged subunits (BV-SE1 to BV-SE8, see **Table 1**). **(c)**
18 Coomassie-stained SDS-PAGE gel of human exocyst complex purified from
19 insect cells co-infected with BV-2863 and BV-E1S547 (see **Table 1**). The exocyst
20 complex was purified using Twin-Strep-tagged subunit EXOC5. **(d)** Electron
21 micrograph of negative-stained recombinant human exocyst complex. The bar
22 represents 100 nm. **(e)** Representative classes from 2D classification of
23 recombinant human exocyst complex particles. **(f)** 3D reconstruction of
24 recombinant human exocyst complex based nsEM data. **(g)** Coomassie-stained
25 SDS-PAGE gel of the human dynactin complex purified by one-step strep-affinity
26 purification. **(h)** Coomassie-stained 3-8% Native-PAGE gel of purified human

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 dynactin complex after glycerol density gradient centrifugation purification. **(i)**
2 Single-particle nsEM analysis of recombinant human dynactin complex with the
3 representative raw micrograph (top) and 2D class averages (bottom). Scale bar,
4 50 nm. **(j)** Coomassie-stained SDS-PAGE gel of purified recombinant human
5 COPI complex, human dynein complex, human CSN complex and human SCF
6 complex.
7

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Recombination of human Exocyst complex using SmartBac system

Subunit	**Intermediate plasmid**	**Final transfer plasmid**	**Recombinant baculovirus**
TS-tagged EXOCn		5V1TG-SEn	BV-SEn
EXOC2, EXOC8	4V2-E28	E2863	BV-E2863
EXOC6, EXOC3	5V1TG-E63		
EXOC1, EXOC5	4V2-E15	E1547	BV-E1547
EXOC4, EXOC7	5V1TR-E47		
EXOC1, TS-tagged EXOC5	4V2-E1SE5	E1S547	BV-E1S547

bioRxiv preprint doi: https://doi.org/10.1101/219246; this version posted November 14, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Recombination of human Dynactin complex using SmartBac system

<table>
<tr><th>Subunit</th><th>Intermediate plasmid</th><th>Final transfer plasmid</th><th>Recombinant baculovirus</th></tr>
<tr><td>TS-tagged p135</td><td rowspan="3">5V1TG-M5</td><td rowspan="3">5V1TG-M5</td><td rowspan="3">BV-M5</td></tr>
<tr><td>p50</td></tr>
<tr><td>p24</td></tr>
<tr><td>Arp1</td><td rowspan="2">4V2-A</td><td rowspan="9">AB</td><td rowspan="9">BV-AB</td></tr>
<tr><td>Beta-actin</td></tr>
<tr><td>CapZ alpha</td><td rowspan="7">5V1TR-B</td></tr>
<tr><td>CapZ beta</td></tr>
<tr><td>P25(DCTN5)</td></tr>
<tr><td>P27(DCTN6)</td></tr>
<tr><td>Arp11</td></tr>
<tr><td>P62(DCTN4)</td></tr>
</table>

Figure 1

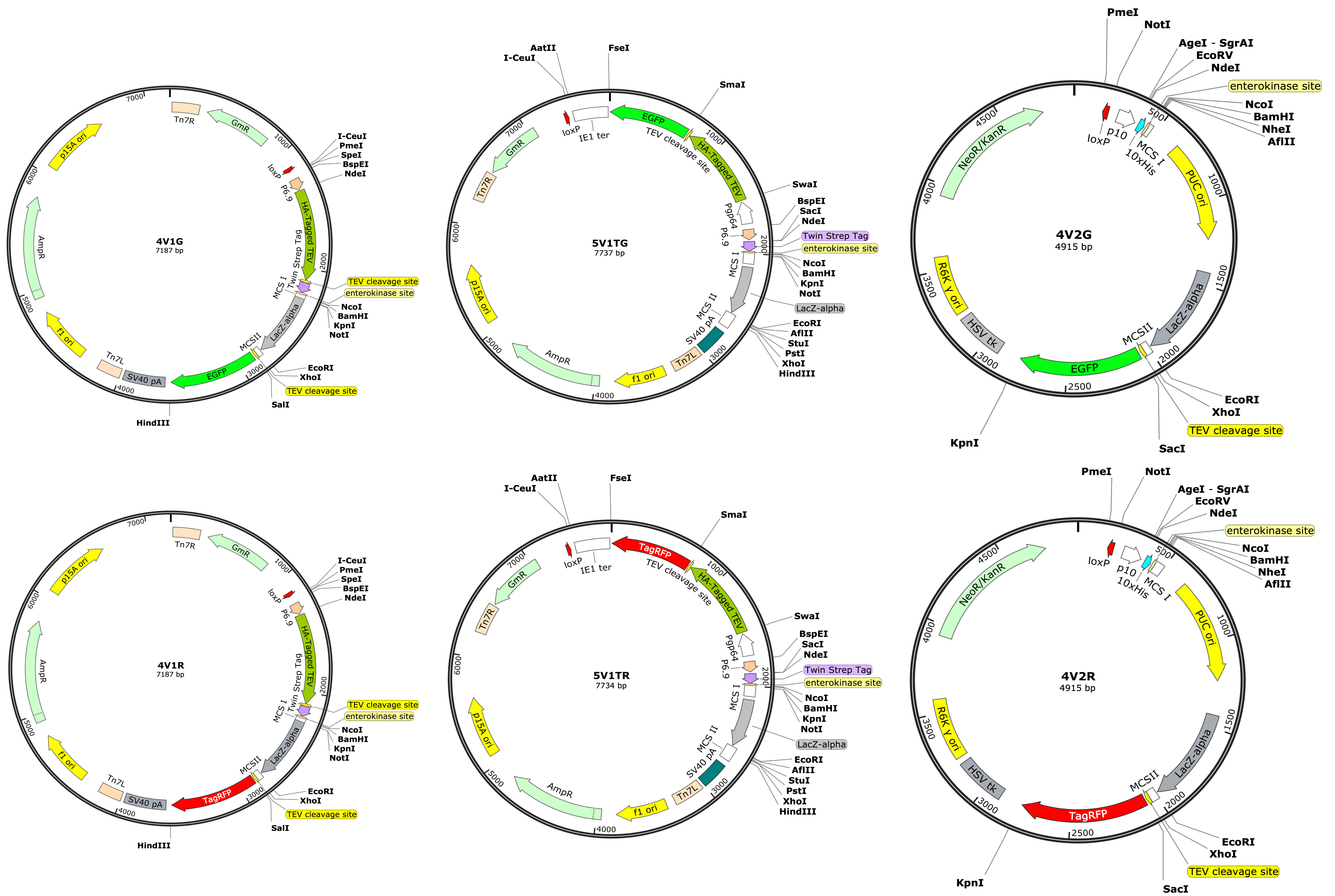


Figure 2

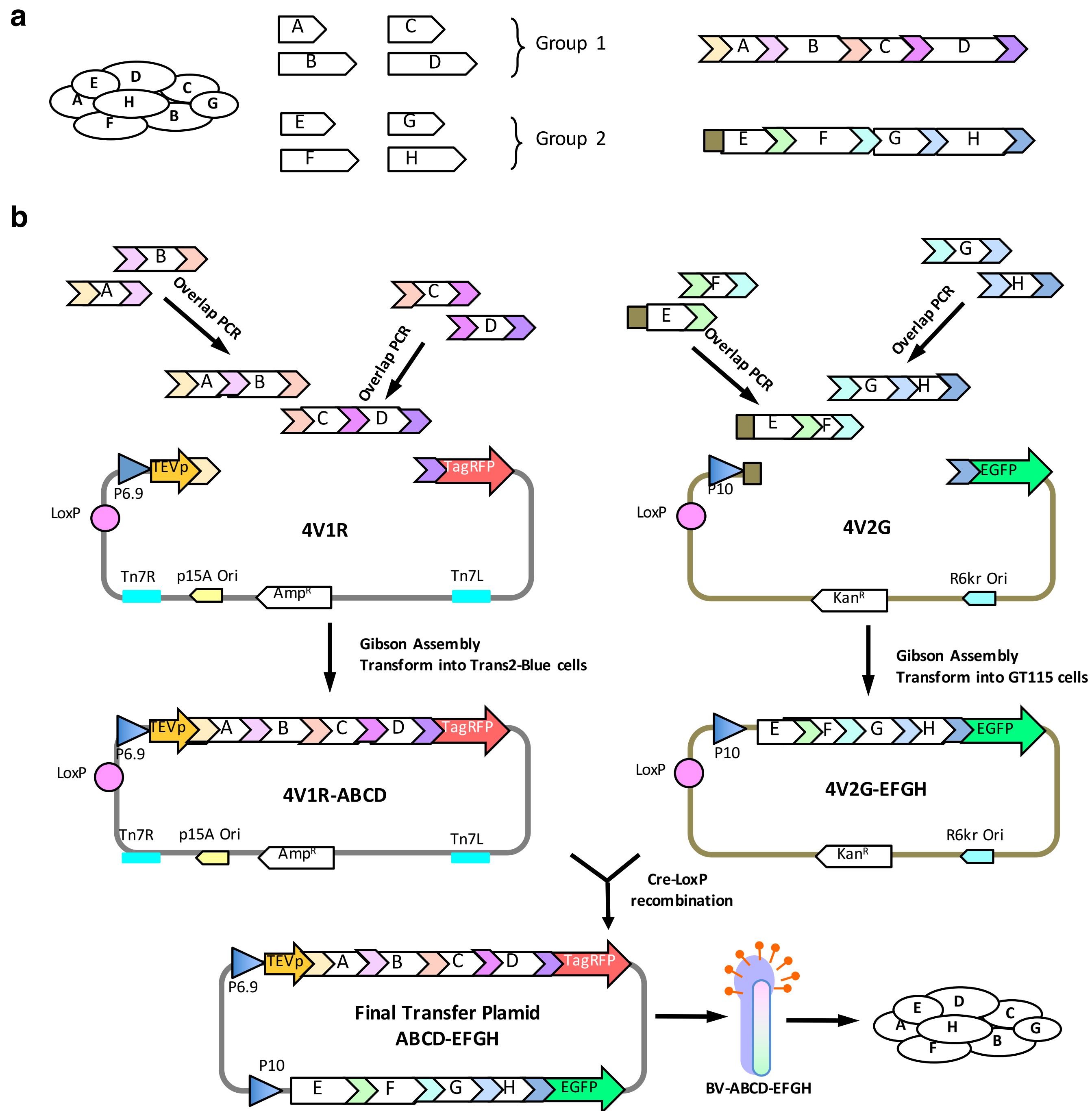


Figure 2

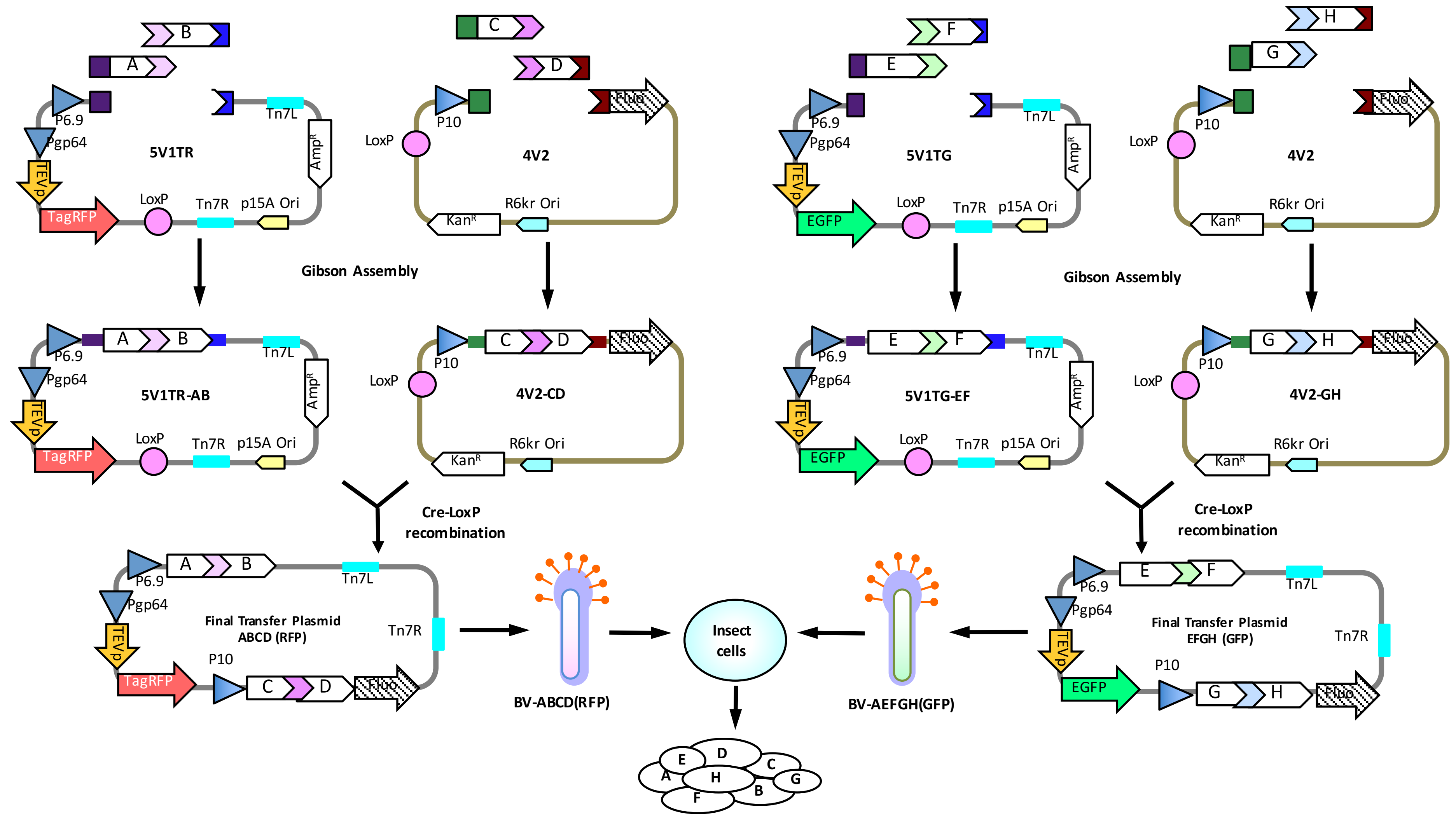
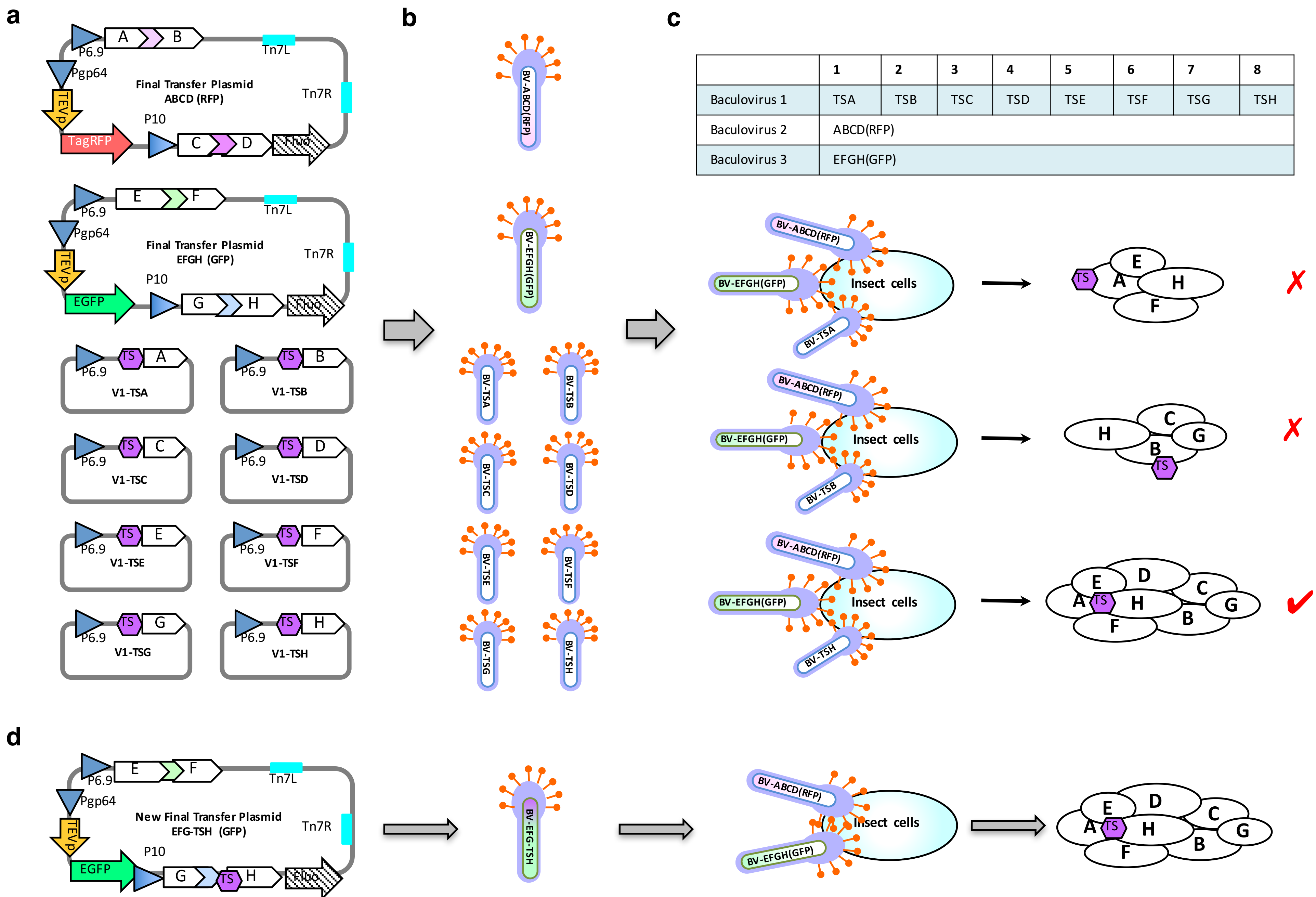


Figure 3



	1	2	3	4	5	6	7	8
Baculovirus 1	TSA	TSB	TSC	TSD	TSE	TSF	TSG	TSH
Baculovirus 2	ABCD(RFP)							
Baculovirus 3	EFGH(GFP)							

Figure 4

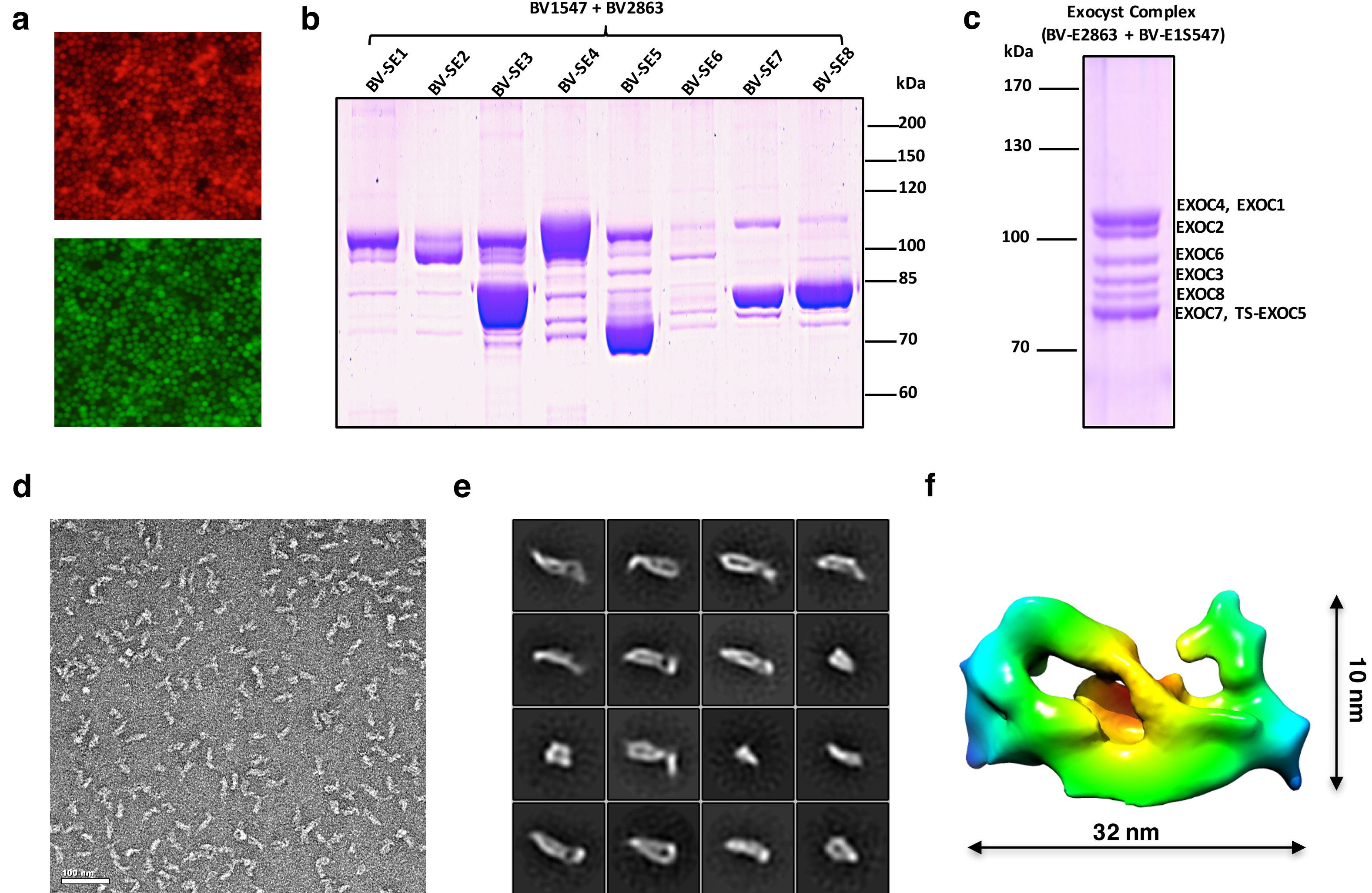


Figure 4