DNA cloning for protein expression using Gateway© technology

Artem Evdokimov
artem@xtals.org

1. *Introduction.*

Gateway© cloning is a technique based on DNA recombination mechanisms utilized by the phage Lambda for integration into, and subsequent excision from the *Escherichia coli* genome. Gateway system was developed by Life Technologies[1], and further refined by Invitrogen after the former was acquired by the latter in 2000.
In general this technique is no different from any other method that relies on cutting and pasting DNA fragments in order to produce recombinant DNA. Unlike restriction-based cloning, Gateway has extremely low background rates and does not impose any requirements♣ on the sequence to be cloned (such as the need for the absence of particular restriction sites within the region of interest). Gateway also does not place many restrictions on the expression vector environment – in fact nearly any conventional expression vector can be converted into Gateway via a simple procedure (as described later). Finally, Gateway technology is rapid – functional expression vectors can be created within 2-3 days. These features make Gateway technology particularly useful for protein expression experiments in multiple vector backgrounds and expression systems. In addition, large libraries of Gateway-adapted ORFs have been created by both academic and commercial entities, therefore cloning of many genes can often be reduced to very simple procedures following the acquisition of the appropriate vector.

2. *Brief functional overview of the system.*

The process of creating an expression vector using Gateway can be broken down into two major stages:

a) Creation of an **Entry vector** – which is a plasmid that serves as a universal donor of the ORF of interest. **Entry vectors** are expression-silent and therefore convenient for sequencing and storage of the ORFs since they are not likely to be toxic for *E. coli*.

b) Transfer of the ORF from the **Entry vector** into the **Destination vector**(s) – which is the plasmid that supplies the elements necessary for expression, such as selective markers, promoters, replicative or integrative elements, etc. Destination vectors themselves can be shuttle vectors for expression in other hosts (baculovirus/insect cells, yeasts, filamentous fungi, mammalian cells, etc.)

♣ Very rarely, *Att*-like sequences can be found within the DNA of interest. The chances for that are very low since the shortest of all sites (*Att*B) is 25-nt long, with a strictly conserved core comprising 14 nucleotides.

(a) The first stage can be accomplished via at least three different technologies – those being conventional (restriction) cloning, Topoisomerase cloning, or Gateway recombination. The end result of all three approaches is the same – to have the ORF of interest flanked with *Att*L recombination sites, to be used in the second stage.

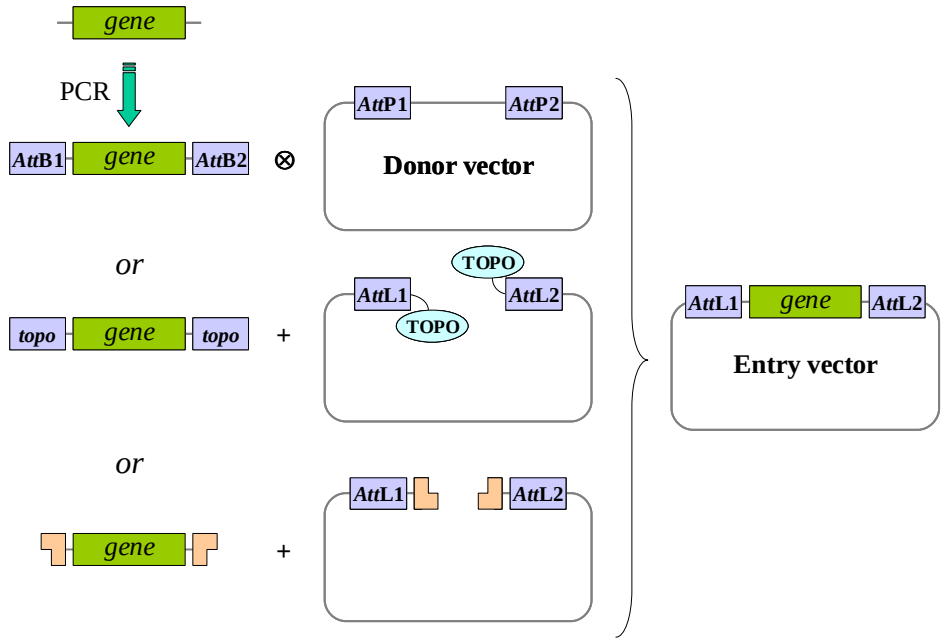


Figure 1. Construction of Gateway Entry vectors. Any of the methods on the left side can be used to generate the vector on the right side.

The first stage invariably utilizes PCR to generate the DNA fragment of interest. Therefore, clones obtained in the first stage can contain PCR errors, and should be uniquefied (e.g. by single colony picking) and sequenced. During the following steps there is virtually no chance that the DNA of interest will change, therefore sequencing has to be performed only once – on the Entry vector.

(b) The second stage is Gateway-specific, and involves recombination between *Att*L (Entry) and *Att*R (Destination) sites resulting in the transfer of the ORF of interest into the background of the expression vector.

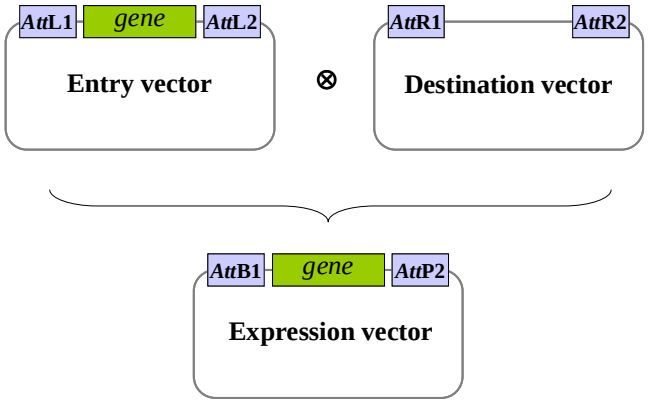


Figure 2. Construction of expression vectors by LR recombination

As a result, an expression vector (or expression clone) is created (unless the Destination vector is in itself a shuttle plasmid, bacmid, etc.). It is noteworthy that *Att*L x *Att*R recombination reaction produces *Att*B and *Att*P sites (this is due to the cyclic nature of the phage propagation in *E. coli*).

3. *How does Gateway achieve nearly 100% background-free cloning?*

Gateway relies on two types of selective pressure to achieve >99% efficiency of cloning (based on numbers of colonies containing correct recombination product versus all other colonies). The following facts are important:

1. The initial *Att*B-containing PCR amplicon cannot propagate in E. coli

2. The AttP-containing Donor vector is Km^R. Also, between the *Att*P sites lays a cassette containing chloramphenicol acetyl transferase (CAT) and a small protein ccdB, both under a low-level constitutive promoter. ccdB is lethal to normal *E. coli* cells because it is a component of a plasmid addiction system. Toxic action of ccdB can be neutralized by either an overexpression of its partner ccdA or by site-directed mutagenesis of *E. coli* DNA gyrase♦. As a result, the Donor vector is Km^R, Clm^R, and is toxic to normal *E. coli*.

3. The AttL-containing Destination vector is Amp^R.♥ The same CAT-ccdB cassette is also nested between the *Att*L sites.
   As a result, the Destination vector is Amp^R, Clm^R, and toxic to normal *E. coli*.

During BP recombination, the *Att*B-containing amplicon reacts with the *Att*P-containing donor vector. As the result of this, *Att*L and *Att*R sites are formed and the DNA between

♦ Currently at least two cell lines are commercially available that can propagate plasmids expressing ccdB – specifically the "**DB3.1"** (DNA gyrase mutation) and "**ccdB-survival"** (overexpressing ccdA).
♥ Other antibiotic resistance genes are allowed as long as they are orthogonal to Km^R and Clm^R

the initial pairs of sites is swapped. The resulting reaction mixture contains four types of DNA molecules (Fig. 3a):

1. The initial PCR amplicon (does not count as it cannot propagate in *E. coli*)
2. The unreacted Donor vector (Km^R, Clm^R, and toxic to normal *E. coli*)
3. *Att*R product – a linear DNA molecule containing the CAT-ccdB cassette nested between the *Att*R sites (also cannot propagate)
4. The Entry vector (*Att*L product) – a circular DNA molecule containing the ORF of interest nested between *Att*L sites (Km^R only).

It is easy to see that transformation of this reaction mixture into normal cloning-grade E. coli followed by selection on **LB-Kan** agar results in only *one* kind of DNA molecule propagating in the cells – namely the Entry vector (4).

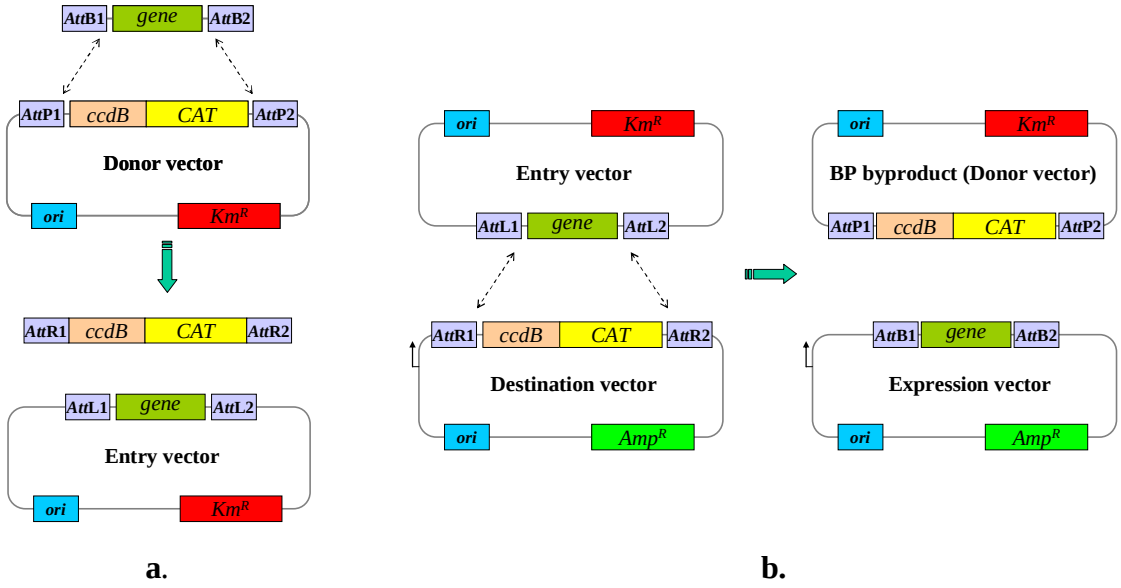


Figure 3. (a) The BP reaction. (b) The LR reaction

After the Entry vector colonies are grown and sequenced, the result is a unique Entry vector which then enters the next step of Gateway cloning - the LR recombination. In this step the *Att*L-containing Entry vector reacts with the *Att*R-containing destination vector(s). As the result of this, *Att*B and *Att*P sites are formed and the DNA between the initial pairs of sites is swapped. The resulting reaction mixture contains four types of DNA molecules (Figure 3b):

1. The unreacted Entry vector (Km^R only)

2. The unreacted Destination vector (Amp^R, Clm^R, and toxic to normal *E. coli*)

3. *Att*P product – what used to be the Entry vector, but now containing the CAT-ccdB cassette nested between the *Att*P sites (Kan^R, Clm^R, and toxic to normal *E. coli*)

4. The Expression vector (*Att*B product) – a plasmid derived primarily from the Destination vector, and containing the ORF of interest nested between *Att*B sites (Amp^R only).

It is easy to see that transformation of this reaction mixture into normal cloning-grade *E. coli* followed by selection on **LB-Amp** agar results in only *one* kind of DNA molecule propagating in the cells – namely the Expression vector (4).

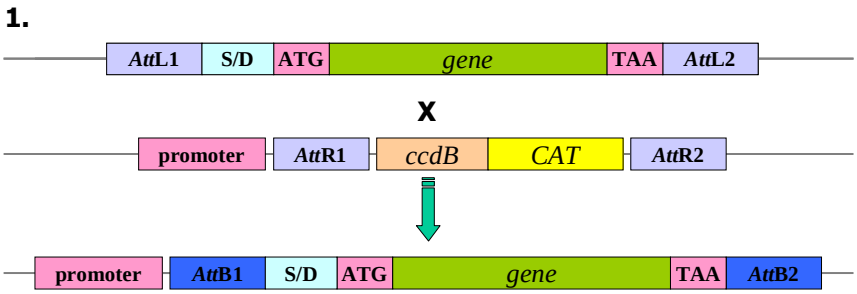
4. *Detailed look at cloning via Gateway recombination for the purposes of protein expression.*

In order to produce recombinant protein in various hosts, the recombinant DNA used to express the protein has to contain several key elements:

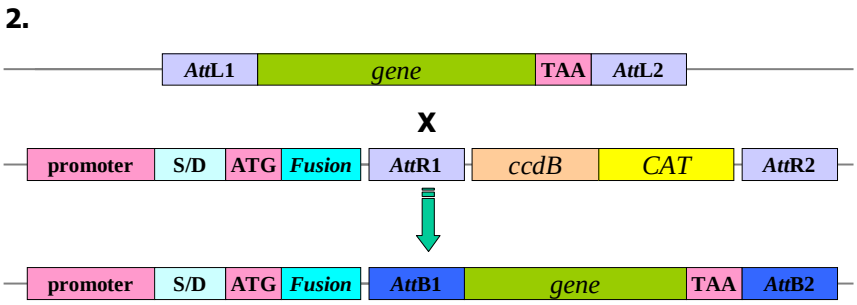
1. The ORF encoding the protein of interest.
2. A promoter, properly positioned upstream of the ORF.
3. A Kozak or Shine-Dalgarno sequence properly positioned at the 5' of the ORF
4. Regulatory elements necessary for inducible expression (if desired).
5. Elements necessary for selection (not always needed) such as resistance genes, complements of host auxotrophy, etc.
6. Elements responsible for maintenance of recombinant DNA in the host of choice (not needed for transient expression) such as replication origins or chromosomal integration loci.

Typically, all these elements are brought together on a plasmid (usually propagated in *E. coli* for convenience). For the efficient use of Gateway, it is customary to prepare a set of Destination vectors – plasmids that carry all the above elements minus the ORF of interest. Destination vectors also carry the *Att*R sites for recombination, positioned in such a way that recombination product contains the ORF of interest properly positioned with respect to the promoter and the control elements. N- and C-terminal fusion proteins can be constructed in the same fashion by incorporating *Att*R sites within the linker between the host and the guest proteins. Because the DNA positioned between the *Att*L sites in the Entry vector is transferred *entirely* and in a *specific frame*, it is necessary to make a choice whether to design the Entry vector for untagged, un-fused expression, or to design the Entry vector for expression of tagged (fusion) proteins. For untagged or C-terminally tagged expression the Entry vector should contain a 5'- ATG codon in the context of either Kozak or Shine-Dalgarno sequences (assuming that the promoter is supplied by the Destination vector) – schemes 1, 3. For N-terminally tagged expression,

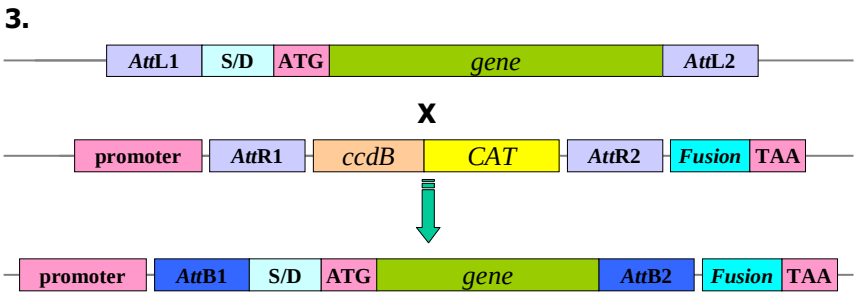
the Entry vector need not contain the 5' elements, and obviously for C-terminally fused constructs the STOP codon at the 3' end should be removed. It is often convenient to incorporate proteolytic cleavage site(s) into the 5' region of the Entry vector, to allow for cleavage of the expressed fusion proteins with site-specific proteases. Cleavage sites can alternatively be incorporated into the Destination vector, however this results in non-native residues present on the N-terminus of the cleaved protein of interest.



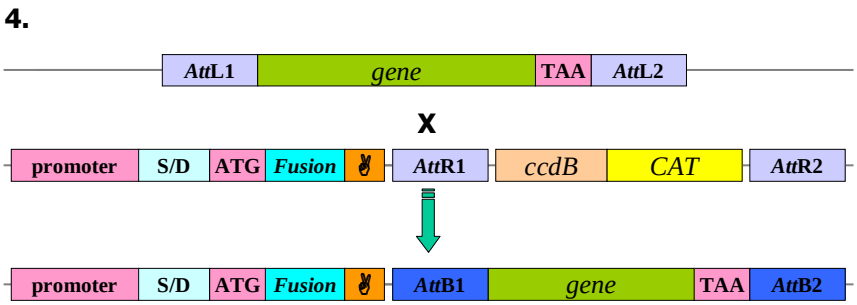
Expression of an untagged protein. S/D represents Shine-Dalgarno or Kozak consensus.



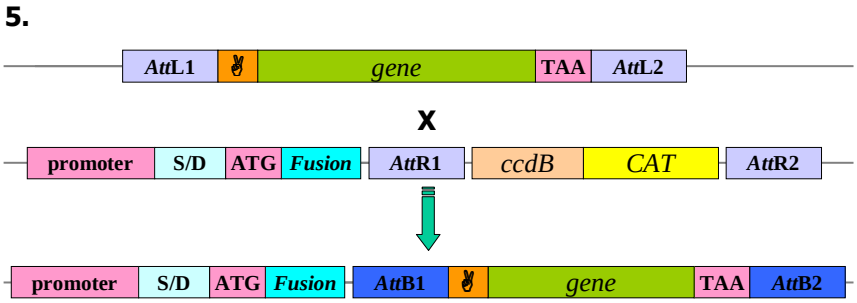
Expression of an N-terminally tagged protein



Expression of a C-terminally tagged protein



Expression of an N-terminally tagged protein with a protease cleavage site (✌).



Expression of an N-terminally tagged protein with a 'clean cut' protease cleavage site.

Many cloning strategies can be designed depending on the particular experiment requirements; however scheme (5) is very convenient for simultaneous multi-host, multi-fusion expression of target proteins typically required for X-ray studies. The reason for this is that in (5) all the expression-regulating elements come from the Destination vector (which means that Entry vector is universal to any expression system) whereas the protease cleavage site is directly attached to the expressed protein, meaning that all the unauthentic sequences can be cleaved off the final purified product. We routinely use TEV[2] and TVMV[3] protease cleavage sequences♠ for this purpose – both proteases are extremely specific (no non-specific cleavage observed to-date) and can be expressed in-house in large quantities.

References:

1. Hartley JL, Temple GF, Brasch MA. (2000) *"DNA cloning using in vitro site-specific recombination."* Genome Res. Nov;10(11):1788-95.
2. Kapust, R. B., Tözsér, J., Copeland, T. D. & Waugh, D. S. (2002). *"The P1' specificity of tobacco etch virus protease."* Biochem. Biophys. Res. Commun. 294, 949–955.
3. J. Tözsér, J.E. Tropea, S. Cherry, P. Bagossi, T.D. Copeland, A.Wlodawer, D.S. Waugh *"Comparison of the substrate specificity of two potyvirus proteases."* (2005) FEBS Journal, 272(2), 514.

♠ TEV = -ENLYFQ/X-, TVMV = -ETVRFQ/X- where the best X is G or S.