

BioGrammatics expression vectors utilize “seamless” cloning schemes; they allow any open reading frame (ORF) to be inserted between the promoter/ATG or leader sequence and the stop codon without any additional bases for restriction enzyme sites. Here, 2 different schemes are outlined: 1) based on Type IIS restriction enzymes that cut outside of their recognition sequence (illustrated below), and 2) by recombination cloning, also known as “Gibson” cloning (see a training video at the NEB website link <https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/nebuilder-hifi-dna-assembly?redi=nebuilderhifi>).

By the first ligation method, the Type IIS restriction enzyme binding sites are removed in the final expression vector and an ORF can be added without unwanted codons. Similarly, by recombination cloning, homology to the exact base in the DNA junction directs the outcome of the final vector.

Circular plasmids, like the BioGrammatics expression vector pJAG or pJAZaMF illustrated below, are linearized with *BsaI* to leave a 4 base, single-strand, “sticky” end on both sides of the cloning junction. An insert with complementary ends, or the proper homology arms, can be rapidly ligated into the vector.

Vectors:

- 1) *vectors for intracellular expression.* All of the promoters in BioGrammatics expression vectors end with the nucleotide G, and an ATG start codon is required for intracellular expression. Therefore a 5'-GATG-3' is the “sticky” end required on an insert – just before the ORF of interest.
- 2) *Vectors for secreted expression encoding the alpha Mating Factor (aMF) secretion signal sequence.* The second to last codon of the aMF ends in a G, and the last codon is GCT, therefore the single strand, “sticky” end created in these vectors is complementary to a 5'-GGCT-ORF “sticky” end required at the 5' end of an insert.

In both cases the other end of the expression vector has a 4 base single-strand, “sticky” end with the TAA stop codon, followed by the nucleotide T. Therefore, the expression vector has a 5'-TAAT-vector “sticky” end. A complementary single strand-end must be created at the 3' end of an ORF.



Illustration of cloning site in a vector with the AOX1 promoter and the alpha Mating Factor (aMF) secretion signal (primers sites for sequence verification of an ORF cloned into the vector are shown).

Vector Preparation.

BioGrammatics expression vectors can be purchased as Cloning Ready expression vectors (CR). CR vectors have been processed, purified and concentrated to facilitate ligation of an insert-of-interest into the expression vector.

Processing of the BioGrammatics expression vectors includes:

- 1) "complete" digestion with BsaI.
- 2) Digestion with NotI. A NotI site is positioned between the inverted BsaI sites to decrease background from undigested plasmid, or plasmids that have been linearized by cleavage at only one of the 2 BsaI sites.
- 3) Gel purification of linearized vector.
- 4) Washing and concentration of the gel purified linear vector.
- 5) The CR expression vector is provided at a concentration of ~30 ng/ul (~10 fmoles/ul, 10 nmolar).

One ul of vector in a 10 ul ligation reaction with ~ 20 ng of a 1 kbp insert should result in 100's of transformants with the correctly sub-cloned insert.

Methods to create ORF-inserts with the proper 5' four base single-strand "sticky" ends.

- 1) *Gene synthesis.* Contract the synthesis of a desired ORF with 11 extra bases on both ends of the ORF to create the proper "sticky" ends. Type IIS restriction enzyme binding sites, inverted relative to one another, are added in the flanking 11 bases to create the 4 base 5' single-strand, sticky ends. Type IIS restriction enzymes that cut outside of their recognition sequence and leave 4 base single-strand ends include *BsaI*, *BsmBI*, and *BbsI*. Following are examples of flanking sequences with *BsaI* binding sites (underlined).

5' end of the ORF:

- a) Intracellular expression, 5'-GGTCTCGGATG-ORF of interest-3'. When cut with *BsaI* a 5'-GATG will remain at the end of the ORF.
- b) For vectors with the aMF secretion signal add: 5'-GGTCTCGGGCT-ORF of interest-3'. Digestion with *BsaI* (recognition site underlined) will result in a 5'-GGCT single stranded end at the 5' end of the ORF (see illustrations below).

3' end of the ORF: in both cases, the addition of the TAA stop codon followed by TCGAGACC-3' (5'-TAATCGAGACC-3') will allow the 3' end of the insert ORF to become complementary to the 5'-TAAT single-strand "sticky" end of the vector, by *BsaI* digestion.

- 2) *PCR products as inserts by ligation (see also recombination cloning below).* Add 15 bases to PCR primers (lower case nucleotides can be any nucleotide, and the *BsaI* recognition site is underlined):
 - a) Forward primer for intracellular expression:
5'ggcGGTCTCgGATG-seq of your desired ORF.
 - b) Forward primer for vectors with the aMF secretion signal:
5'ggcGGTCTCgGGCT-seq of your desired ORF.
 - c) In both cases the "Reverse" primer should have: 5'-15 or more bases of the ORF -TAATgGAGACCgcc-3'

Note, the addition of flanking sequences by PCR, and subsequent digestion of the PCR product by a Type IIS enzyme requires that *BsaI* cut the DNA near the end of the PCR amplicon. Therefore, 3 non-specific bases are added to the ends to facilitate, for example, *BsaI* cleavage. These bases are not required when the cloning sites are added during gene synthesis; in this case, the ORF insert is clone into a circular plasmid.

- 3) *PCR product as inserts - ligation (alternative method).* Make two similar PCR products that are different by only 4 bases at the 5' or 3' ends: one with the 5' desired single-strand end (5' GATG for intracellular expression), the other with the TAAT-3' end. Mix the two PCR products, heat to denature, and then cool to create a class of inserts with a 5' and 3' single-strand "sticky" ends.

- 4) *Peptide inserts*. Similar to 3 above, small inserts less than ~200 bp can be made by annealing two single stranded primers so the first 5' four bases of the oligonucleotides are unique and the rest of the oligonucleotide is complementary to the other oligonucleotide. The unique bases should complement the single stranded ends in the expression vector, i.e. "Top strand" = 5'-GGCT-ORF for vectors with the aMF ss, and "bottom strand" = 5' ATTA- ORF (reverse complement).
- 5) ***Recombination cloning*** (also called *Gibson cloning*). More recently popularized as a convenient method to join DNA ends. See information on the New England Biolabs (NEB) website; they have a short video, NEBuilder® HiFi DNA Assembly, or Gibson Cloning, see this link: <https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/nebuilder-hifi-dna-assembly?redi=nebuilderhifi>
Recombination cloning results in the 5' strand of any DNA end, i.e. in the vector and insert(s) being "chewed" back, leaving complementary 3' overhangs of >15 bases to anneal and ligate.

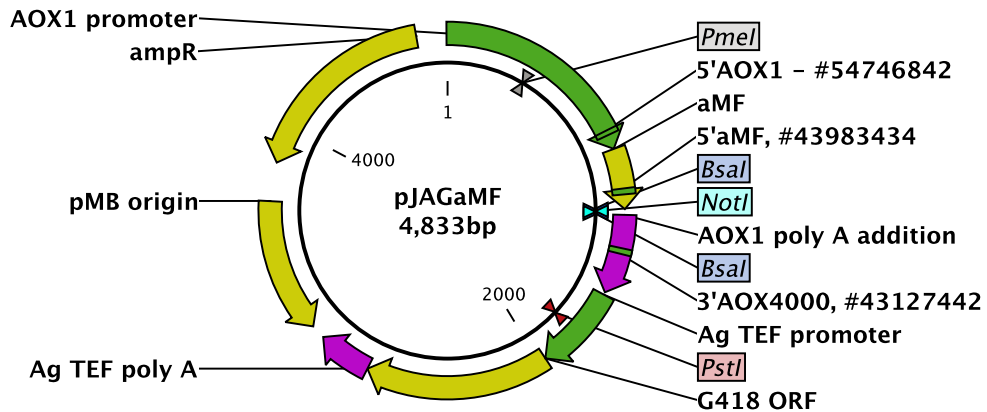
To clone an ORF into the pJAGaMF vector, or any BioGrammatics vector with the alpha Mating Factor secretion signal, the following 5' ends of the primers for amplification of an ORF, if designed to immediately precede the first and last codon, will allow the proper recombination:

- 1) 5' ORF primer: 5'- TGGACAAGAGAGAGGCTGAGGCT- ORF-3'
- 2) 3' ORF primer: 5'- GGCATTCTGACATCCTCTTGATTA-ORF-3' *Note, here this is the "reverse" primer, and, it includes the stop codon.*

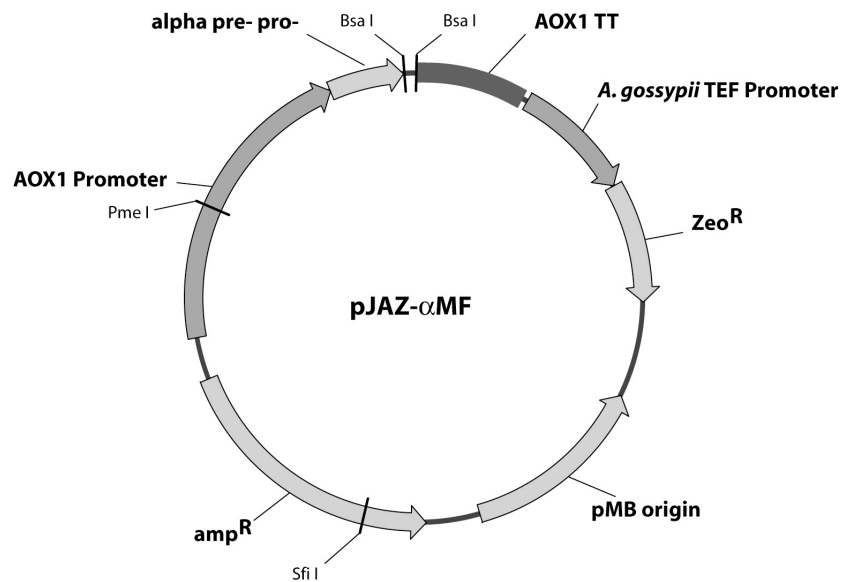
Tips.

- An ORF that is synthesized and provided in a circular plasmid should be put into a plasmid with, for example, a Kan resistant marker, not AmpR. The alternative selection marker allows sub-cloning without gel purification.
- Use a proof-reading polymerase for higher fidelity, particularly for larger inserts.
- Add at least 3 bases of non-specific nucleotides to the 5' end of PCR primers to facilitate cleave of a Type IIS enzyme near the end of a PCR fragment.
- If all available Type IIS restriction enzyme sites are in your ORF, use the "alternative PCR method".
- Sequence verify the complete ORF and cloning junctions of all expression vectors prior to transformation into *Pichia*.
- If possible, remove restriction enzyme sites from an ORF of interest if the sites are required to linearize the expression vector for integration into the *Pichia* genome. For example, the *PmeI* site in the AOX1 promoter is good to linearize an expression vector to target integration to the AOX1 locus.
- *BsaI* and *BsmBI* create 4 base single-strand ends, one base away, from their recognition site. The type IIS restriction enzyme *BbsI* leaves a 4 base single strand end 2 bases from the recognition site. This must be considered in the design of the cloning ends.

Plasmid map of pJAGaMF (also referred to as pJAG-s1).

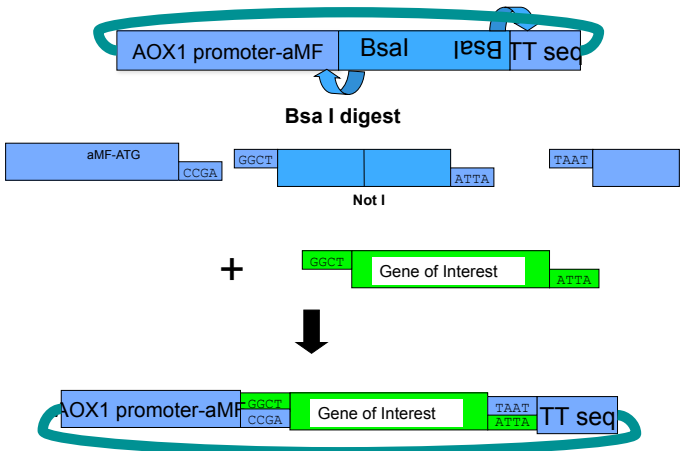


Vector map, pJAZ-aMF



Modified Biogrammmatics vectors include: vectors without a signal peptide (i.e. pJAZint), alternative signal peptides (new and modified aMF; aMF-KR, i.e. no EAEEA), different promoters (i.e. GAP, Biogrammmatics inducible and constitutive), and selectable Markers (ie. Resistance to G418, Nat, Hygro, Bsd).

Seamless Cloning Scheme (vector with aMF, i.e. pJAZ-aMF)



“Seamless” Cloning

Vector cloning junction	G-A
end of aMF	C-T-C-C-G-A

