

Plasmid Library Construction

Small-insert random genomic libraries provide the groundwork for most sequencing projects. Successful genome projects have relied on these libraries to gather the bulk of the sequence data required (shotgun sequencing).

Essentially any source of DNA can be used as starting material for plasmid library construction (genomic, organellar, large-capacity genomic clones such as BACs or YACs, etc). The only requirement is that the DNA prep be free of any contaminating DNA source. Once the target DNA has been purified, it must be fragmented to the desired size before cloning into the chosen plasmid vector.

When generating a plasmid library, care should be taken to eliminate small inserts from the population of fragments to be cloned. Typically, fragments should be no smaller in length than two average sequence reads; this maximizes the amount of sequence information gathered from each clone. It also ensures that sequences derived from unlinked regions in the genome are not colinear and present in the same sequence read, even if the clone *'per se'* contains more than one insert fragment.

The choice of vector depends on the specific needs. Considerations include the ideal fragment size, the desire for screening out "empty" clones (*i.e.*, clones without inserts), compatibility with other cloning systems, downstream uses of the sequencing clones, etc. In general, all-purpose cloning vectors such as pUC18/pUC19 provide a simple, inexpensive vehicle. A variety of commercial cloning systems are available, some of them with clever features to speed up construction, to minimize empties, etc. Also, low and intermediate copy number vectors are available (such as pSMART, Lucigen); these are useful in instances where libraries in high copy number vectors are biased. The TOPO (Invitrogen) family of cloning vectors provides a very fast and useful, albeit more expensive, way of generating libraries with high efficiency.

Fragmentation of target DNA

Enzyme-based methods

Although enzyme-based methods to fragment DNA have the disadvantage of being nonrandom (since digestion is influenced by sequence context), the advantages are that these methods are generally quite robust, the digestions produce DNA ends that are readily clonable without the need for further manipulations, and they are generally quite quick. They are commonly used for small projects or when the objective of the project is to collect sequence data without aiming for completion.

The two more widely used enzymatic methods for fragmenting DNA are: 1) using DNaseI in the presence of Mn^{2+} to generate random double stranded breaks, followed by repair of the ends using T4 DNA polymerase and/or Klenow; and 2) partial digestion with a restriction enzyme, usually one with a 4-basepair recognition site. Of the two, partial digestion with restriction enzymes is the favored method because of its simplicity. The choice of enzyme may be guided by prior knowledge; for example, target DNAs with high GC content can be digested with *Sau3A* and/or *MboI*, whereas for AT-rich sequences one is better served by using *Tsp509I*. Whatever the decision as to which enzyme to use, after running partial digestions, the DNA fragments should be separated on an agarose gel and fragments of the desired size excised and purified before cloning into the plasmid vector.

Mechanical methods

Mechanical methods for fragmenting DNA are preferred for large sequencing projects. They provide the most random collection of DNA fragments since they are not biased by sequence composition. The disadvantage, and hardest problem when generating libraries using these methods, is that the fragmented DNA has to be treated extensively to repair the ends to render them clonable.

The most widely used methods for fragmenting DNA into the 1-10 kbp size range are use of a French press, sonication and shearing by nebulization (using either a clinical nebulizer or the specifically designed Hydro-Shear). The first two require that special equipment be available (either a mini-French pressure cell or a sonicator fitted with a microtip probe). Mechanical shearing using a nebulizer cup provides a fast and inexpensive alternative. The main disadvantage is that it requires sizeable quantities of DNA: whereas the first two methods require as little as 5 µg of DNA, shearing by nebulization requires 4-5 times that amount. The Hydro-Shear is probably the best alternative, except for the fact that it is a quite expensive piece of equipment; however, it is easy to operate and requires the least amount of DNA of all mechanical methods.

End-repair and size selection of DNA fragments

The mechanical methods of DNA fragmentation produce a collection of DNA fragments with heterogeneous ends: flush, 5' and 3' overhangs, with or without phosphate groups at the ends. Some of the jagged ends could be dozens of nucleotides long. Therefore, it is necessary to repair the ends of the resulting fragments prior to cloning. This is probably the most crucial and difficult step in the generation of random sequencing libraries. Although numerous protocols are cited in the literature, no single one is foolproof. In our experience, researchers are well advised to try several repair protocols and then decide what works best in their hands.

We have found it useful to remove most of the jagged ends by treatment with a single strand-specific exonuclease, such as BAL32 nuclease or Mung Bean nuclease, prior to size selection on agarose gels and end repair of the excised fragments. The nuclease treatment is done on the whole collection of randomly sheared fragments (working with larger amounts of DNA at this stage is easier than trying to fine-tune the nuclease treatment to the small amounts of DNA recovered from a specific region of the agarose gel). The nucleases need to be "calibrated" first: typically, one would want the fragments to be reduced in size about 200-300 bp to ensure smaller overhangs that will be easier to repair.

After the DNA ends have been shortened by nuclease treatment, fragments are separated on agarose gels, and the desired size molecules are excised and recovered from the gel. Generally it is best to excise a narrow size range, so that the length of cloned fragments is accurately known (to help identify potential "deleted" clones). Also, the spread of fragments should be such that the smallest potential dimer between two insert fragments will be longer than the largest monomer fragment (this will help identify chimeric clones). The agarose should be one of a variety of "refined" agaroses, such as SeaPlaque (FMC Corporation) or similar. To recover fragments from the gel, a variety of methods can be used: treatment with agarase (especially useful for larger fragments), melting of the agarose plug followed by phenol extraction, or one of a series of commercial kits for gel extraction, such as those sold by Qiagen or Stratagene.

After the fragments have been extracted, the DNA ends are repaired by treatment with T4 DNA polymerase and the Klenow fragment of DNA polymerase. At this stage, an optional additional treatment with T4 polynucleotide kinase to insure that all relevant ends are phosphorylated can be performed. The DNA fragments are now ready for cloning.

In the case of preparing fragments for cloning into TOPO vectors (Invitrogen), they should not be dephosphorylated but rather treated with alkaline phosphatase at this point, since the "TOPO-gation" reaction requires OH⁻ groups at the insert ends.

Preparation of plasmid vector

The chosen plasmid vector has to be digested with an appropriate restriction enzyme to generate the correct ends required for cloning. It is important to ensure that digestion is as complete as possible since undigested vector will contribute to the background of empty clones. Also bear in mind that supercoiled circular DNA will transform more efficiently than nicked,

relaxed molecules (as recombinant molecules will be). After digestion, the plasmid should be dephosphorylated by treatment with a phosphatase to decrease the background of self-ligating empty vector molecules. Calf intestine alkaline phosphatase works well. Also bear in mind that dephosphorylating blunt-ended molecules is less efficient than treating 5' overhangs, and incubations at higher temperatures (above 37°C) may be useful in allowing better access of the enzyme by causing the DNA ends to "breathe". At this point, the digested and dephosphorylated plasmid may be purified on agarose gels as an optional step.

Alternative protocols include preparing vectors with incompatible ends, to prevent self-ligation; inserts should be generated with the appropriate ends (for a very good strategy, refer to Bill Nierman's lecture notes)

Ligations and transformations

Ligations should be set up with an insert:vector ratio close to 1:1 to increase the likelihood of bimolecular reactions. However, one should realize that is true in the ideal situation in which all DNA ends are compatible. One of the biggest uncertainties in these protocols of library construction is the chemical identity of the DNA fragment ends, and there is no guarantee that in mechanically sheared DNA all ends have been efficiently repaired. Therefore, it is advisable to also include, if possible, ligation reactions with increasing amounts of fragment.

After ligation, usually overnight, an aliquot is transformed into a suitable host strain, such as XL1Blue-MRF'. Since the idea is to maximize the number of recovered transformants, electrotransformation is the method of choice. Plating on media containing the appropriate antibiotic will identify transformants clones. If it is desired (and if the cloning system allows it) color selection in the presence of X-gal and IPTG can be used to discriminate recombinant clones from vector-only transformants.

For small sequencing projects, or in cases where one does not mind the fact that the cloned fragments are transcribed (because of the use of color selection), the libraries generated are suitable for sequencing. Also, if the existence of chimeric clones (clones with more than one insert molecule) is of no importance (for example, in the case that the potential proximity of the 2 sequence reads generated for that clone is not going to be utilized in strategies for assembly), the library will be suitable.

Second-ligation libraries

There are instances where it is necessary to screen out chimeric clones, or where it is desirable to avoid inducing the cloned fragments (for example, to allow potentially deleterious sequences to be cloned). In these instances, one may opt to proceed with additional steps to ensure a minimum of chimeric and empty clones in the library.

After ligation (as above) the products of the reaction are resolved in an agarose gel. If ligation was efficient, there should be a linear vector band and a smear of insert fragments; and above them a ladder of discreet fragments, each rung representing (vector + one insert), (vector + two inserts), etc. The (vector + one insert) band is excised and the DNA recovered. After a round of re-polishing with T4 DNA polymerase, the fragments are allowed to self-ligate and an aliquot of the reaction is transformed into a suitable host (as above). These second-ligation libraries, when successful, show very low (<2%) background of empty clones with essentially no multiple-insert recombinants.

In the case of TOPO vectors, the generation of recombinant clones does not proceed through a typical ligation step. Rather, the enzyme topoisomerase (covalently bound to the 3' ends of the linearized vector) mediates a series of cleavages and chemical bond formations resulting in "ligated" recombinant molecules. Hence, in this case there are no clones with multiple inserts. However, the "TOPO-gation" reaction is very sensitive to the concentration of insert, since a vector molecule reacting with two different insert fragments at each end will be unproductive. Therefore, in this case it is advisable to set up reactions with different vector:insert ratios.

Plasmid Library Construction Protocol

Preparation of plasmid vector for cloning of blunt end inserts

In a microcentrifuge tube, mix:

10 µg pUC18 DNA (or another plasmid of your choice)
H₂O to 100 µL final volume
10 µL 10x *Sma*I reaction buffer
20 units *Sma*I

10x *Sma*I reaction buffer:
200 mM Tris-HCl, pH 7.4
50 mM MgCl₂
500 mM KCl

Incubate at 30°C for 2 hours.

In a small agarose gel, run an aliquot of the reaction mix to check that digestion has been complete.

Extract the reaction with an equal volume of Phenol:Chloroform:Isoamyl alcohol (25:24:1) (PCI)

Recover the aqueous (upper) phase and extract with an equal volume of Chloroform:Isoamyl alcohol (24:1) (CI)

Recover the aqueous phase and precipitate the DNA with 1/10th vol 3 M NaOAc and 2.5 vol cold ethanol. Leave at -20°C (from 1 h to overnight)

Centrifuge at 15,000 x g for 15 min at 4 °C.
Rinse the DNA pellet with 1 mL 80% cold ethanol.
Spin 5 min at 15,000 x g at 4°C.
Drain and dry the pellet (either air dry for about 10 min on the benchtop, or in a SpeedVac for 3-5 min).

Resuspend the DNA pellet in 85 µL TE (10 mM Tris-HCl, pH 8.0/1 mM EDTA)
Add: 10 µL 10x CIAP buffer
10 units calf intestinal alkaline phosphatase (CIAP)

10x CIAP buffer:
500 mM Tris-HCl, pH 9.3
10 mM MgCl₂
1 mM ZnCl₂
10 mM spermidine

Incubate 15 min at 37°C, followed by 15 min at 55°C
Add another 10 units CIAP
Incubate 15 min at 37°C, followed by 15 min at 55°C

Extract with 100 µL PCI, followed by 100 µL CI, as above. Recover DNA by ethanol precipitation, as above.

Resuspend DNA pellet in 50 µL TE and quantitate DNA.
Vector is now ready for ligation.

Notes:

Before using the vector in ligations, it is advisable to test the batch. Transform an aliquot of supercoiled plasmid and an equal amount (usually 0.1 ng) of *SmaI*/CIAP vector. There should be at least 1,000 times fewer transformants for the treated vector.

Preparation of insert DNA fragments*Nebulization*

In the cup of a clinical nebulizer, mix:

30-50 μ g DNA

500 μ L 50% sterile glycerol

100 μ L 3 M NaOAc

H₂O to 1 mL final volume

Remove a small aliquot for gel analysis later (10 μ L is enough)

Nebulize with a stream of nitrogen, keeping the nebulizer on ice. The most important parameter determining the extent of shearing is the pressure of the gas stream.

Start with 10 seconds at ca. 12 psi. Remove an aliquot as above and check the extent of shearing on an agarose minigel. If fragments are still too big for your needs, increase the pressure, nebulize again for 10 sec, and check the results again. Repeat if necessary until you are satisfied. Remember that the BAL31 nuclease digestion that follows (see below) will reduce the size of the fragments an additional 100-300 bp.

This is rather tedious, and if you have plenty of DNA available, you may want to "calibrate" the fragmentation first by running a "pressure course": nebulize for 10 sec at 12 psi, remove an aliquot; nebulize another 10 sec increasing the pressure, etc. Analyze all samples together and determine the optimal pressure to be used thereafter.

When the DNA has been fragmented to an appropriate size range, remove the solution from the nebulizer and precipitate the DNA by adding 2 vol of cold ethanol. Recover the DNA by centrifugation, rinse with 80% cold ethanol, and drain and dry the pellet.

BAL31 Nuclease treatment

Resuspend the DNA in 49 μ L TE.

Add 50 μ L 2x BAL31 buffer

1 μ L BAL31 Nuclease (1 unit, from NEB, a mix of "fast" and "slow")

2x BAL31 buffer:

40 mM Tris-HCl, pH 8.0

1.2 M NaCl

2 mM EDTA

24 mM MgCl₂

24 mM CaCl₂

Incubate 5 min at 30°C.

Clean up the reaction either by organic extraction (with PCI and CI) followed by ethanol precipitation, or using one of several commercial kits, such as Qiaquick PCR purification kit from Qiagen.

Resuspend (or elute, if using the kit) the DNA in 40 μ L EB (elution buffer: 10 mM Tris-HCl, pH 8.5)

Size fractionation of insert fragments

Pour a 1% agarose gel in 1x TAE (for recipe, see Molecular Cloning: A laboratory manual) at least 10-12 cm long, with lanes about 2 cm wide. Use a good low-melting agarose, such as SeaPlaque (FMC Corporation).

To the resuspended DNA pellet, add an appropriate gel loading buffer (containing bromophenol blue). Run out the fragments at about 8 V/cm (preferably in a cold room) until the bromophenol blue has migrated $3/4^{\text{th}}$ down the length of the gel. Make sure that you have good size markers (such as 1-kbp DNA ladder from Invitrogen) on both sides of the experimental lane.

Visualize the DNA by staining with ethidium bromide and ultraviolet light. Excise the regions of the gel containing the desired size fragments. Work quickly to avoid excessive nicking of the DNA.

Recover the DNA fragments from the gel: Melt the gel chunk by incubation for a few minutes at 65°C. Extract once with an equal volume of TE-saturated phenol, followed by extractions with PCI and CI, as above. Recover the DNA by precipitation, as above.

A faster alternative is to use one of a variety of commercial kits, such as Qiaquick gel extraction kit (Qiagen).

End repair of recovered fragments

Resuspend (or elute if using a kit) the DNA in enough EB to bring the final volume of the repair reaction to 100 μ L.

Add:

20 μ L 5x T4 DNA polymerase buffer

1 μ L 20 mM each dNTP's

5 units T4 DNA polymerase

5x T4 DNA polymerase buffer:

165 mM Tris-acetate, pH 7.9

330 mM NaOAc

50 mM MgOAc

500 μ g/mL BSA

2.5 mM DTT

Incubate 45 min at 16°C.

Add 5 units Klenow fragment of DNA polymerase.

Incubate 30 min at 37°C.

Alternatively, you can perform both reactions simultaneously, incubating the mix at room temperature.

Extract the reaction with an equal volume of PCI, followed by CI, as above.

Precipitate the DNA with cold ethanol, as above.

Make sure you rinse the DNA pellet extensively with 80% ethanol, since Na⁺ ions inhibit the ligation reaction that will follow.

At this point some researchers find it useful to add a phosphorylation step. Treat the recovered DNA fragments with T4 polynucleotide kinase (as suggested by the manufacturers). After treatment, make sure you extract the reaction and precipitate the DNA as above.

Resuspend the recovered DNA fragments in 20 μL TE.

Estimate the amount of DNA by running an aliquot (1-2 μL) of the preparation in a small agarose minigel; use known amounts of DNA (plasmid, for example) as standards for comparisons.

The DNA fragments are now ready for ligation.

(First) Ligation

In a microfuge tube, mix:

200 ng *Sma*I/CIAP vector

equimolar amounts of repaired fragments (*e.g.*, *ca.* 200 ng of 3-kbp fragments)

H₂O to adjust the total volume of Vector+Fragment to 15 μL

4 μL 5x T4 DNA ligase buffer

1 μL T4 DNA ligase (400 units as defined by NEB, or 6 Weiss units)

5x T4 DNA ligase buffer:

250 mM Tris-HCl, pH 7.5

50 mM MgCl₂

5 mM ATP

5 mM DTT

25% w/v polyethylene glycol-8000

If enough repaired fragments are available, set up a second ligation mix with 3 times the amount of fragment DNA.

Incubate overnight at 16°C.

Dilute the ligation reaction 2-fold with H₂O and heat to 65°C for 15 min.

Use 0.5-1 μL of the ligation reaction to electrotransform a suitable host strain, such as *E. coli* XL1Blue-MRF'. Follow electroporator manufacturer's instructions. After delivering the pulse, resuspend the cells in 1 mL of LB broth. Incubate for 30 min at 37°C with very mild agitation. Plate out 40- and 200- μL aliquots on LB plates supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin (or other suitable antibiotic according to the vector used). Supplementing the plates with X-gal and IPTG will provide an estimate of the ratio of recombinant (white) to empty (blue) transformants.

For some projects, if enough transformants for the specific needs are recovered, this will be the end of library construction. However, when wanting to eliminate chimeric clones and reduce the background of empty transformants, the following steps will be useful.

Second Ligation

After first ligation, add an appropriate gel loading buffer, then run out the reaction on a 0.75-1% SeaPlaque agarose gel in TAE, in a lane about 1 cm wide. After the bromophenol blue has reached 3/4th down the length of the gel, visualize the DNA by staining with ethidium bromide.

If the repair of fragments and ligation have been successful, you should see a tight band of cut vector, a smear of repaired fragments and a ladder of bands above representing vector+1 insert; vector+2 inserts, etc, and growing fainter as the size increases.

Excise the vector+1 insert band from the gel and extract the DNA, as above. Resuspend (or elute if using a kit) the DNA in 38.5 μL TE.

Add:

10 μ L 5x T4 DNA polymerase buffer
 1 μ L 20 mM each dNTP's
 2 units T4 DNA polymerase

Incubate 30 min at 37°C.

Extract the reaction with PCI and CI and precipitate the DNA, as above.
 Resuspend the recovered DNA in 15 μ L H₂O.

Add:

4 μ L 5x T4 DNA ligase buffer
 1 μ L T4 DNA ligase (as above)

Incubate 4 hours at 16°C.

Dilute the reaction 2-fold with H₂O and heat to 65°C for 15 min.
 Proceed to electrotransformation, as above.

Comments:

1) Transforming after first ligation gives you lots of transformants, but in my hands at least 10% are empties. After second ligation, only 2% are empties. This is important when clones are not screened for presence of insert prior to sequencing.

2) If libraries are constructed with the 2-ligation steps protocols, there is no need to add IPTG when plating out transformants since so few of them will be empty. Also, minimizing transcription through the insert DNA increases the likelihood that deleterious sequences be recovered in clones; the libraries will be less biased.

3) Good sources of enzymes (my favorite):

BAL31 nuclease: New England Biolabs. This is important; their enzyme is a mixture of "fast" and "slow" and is very consistent batch-to-batch. My recommendation is: don't even try any other source.

*Sma*I: New England Biolabs, although most other labs' are adequate.

CIAP: Promega (high concentration preparation)

T4 DNA polymerase, Klenow, T4 DNA ligase: Invitrogen. Actually, I used to like the Gibco enzymes, but Invitrogen has bought them out; and the quality of the enzymes has suffered (oh well...)

4) Without meaning to sneak in a plug for a commercial lab, my experience is that Qiagen kits work really well for clean-up and gel extraction.

5) For small- to medium-sized projects, I like the TOPO vectors; they are fast, very easy to use and essentially every clone you recover is a transformant.

6) When constructing libraries for sequencing, be very clean with the materials you use. If possible, dedicate equipment to that purpose solely. For example, thoroughly clean gel boxes and, if possible, do not use them for DNA's other than the one from which you want to generate the library. Also, do not reuse nebulizers.

Claudia Reich, PhD
 Department of Microbiology
 University of Illinois at Urbana-Champaign
 claudia@life.uiuc.edu

