

# Guide to expression construct cloning

Marko Hyvönen

October 18, 2004

## Preface

This manual describes how to create an expression construct to over-express a protein or a domain. The example described is using an E.coli expression vector, but the same applies for others systems, such as *Pichia pastoris* or baculovirus expression vectors. The manual is written with the needs of structural biology (or rather, structural biologists) in mind, but the same procedure is fully applicable whenever recombinant protein is being expressed, for example when producing antigens for antibody production.

Links are provided to other manuals which deal either with supporting protocols or later steps of the complete process of producing recombinant proteins.

The reader should remember that there are a number of ways each step can be done and this manual describes the way I work. You should choose the methods and reagent/kits according to availability and the expertise around. Should anything go wrong, it might be difficult to trace the problems with this manual alone. A good source for molecular biology protocols is the Cloning Manual by Sambrook et.al. and Current Protocols for Molecular Biology.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Designing the expression construct</b>	<b>3</b>
<b>3</b>	<b>Cloning</b>	<b>7</b>
3.1	Primer handling . . . . .	7
3.2	PCR . . . . .	7
3.3	Analysing the PCR products . . . . .	9
3.4	Purification of the PCR fragment . . . . .	10
3.5	Restriction digestion . . . . .	11
3.6	Subcloning to a new vector . . . . .	12
3.7	Preparing the plasmid for cloning . . . . .	12
3.8	Ligation . . . . .	13
<b>4</b>	<b>Analysing the transformants</b>	<b>14</b>
4.1	Restriction analysis . . . . .	15
4.2	Final confirmation of the constructs . . . . .	15

## 1 Introduction

Creating an expression vector includes the following steps:

1. Designing the construct and choosing the vector
2. Designing the cloning strategy and ordering oligonucleotides
3. Cloning by PCR
4. Screening of the clones and verifying their correctness
5. Testing for expression in the expression strain

An overview of the whole procedure is given in the Figure 1.

## 2 Designing the expression construct

This is perhaps the most crucial step of the whole experiment and also the one to which you might have to return to many times in the future. If you are expressing a complete proteins from N- to C-terminus, you should consider yourself lucky, as the most difficult decision, where to cut the protein, has been made already. If you are planning to express a part of a bigger protein, definite advice cannot be given and one has to use very subjective judgement in making these decisions. Domain boundaries can sometimes be difficult to define and combination of several techniques should be used to derive the necessary information.

For a structural biologists it should be self-evident, that the expression constructs must make sense structurally: only complete domains (folding units) are expressed and, unless needed for functional reasons, the expressed protein should lack any flexible, non-structured tails.

A multiple sequence alignment is the first step in defining the domain boundaries. Sequence conservation (especially of the hydrophobic residues) is greatest in the structured domains and lowest in the joining linkers. If a structure of a homologous protein is known, this can be used as a guide to define the domain boundaries of the novel protein. Also, secondary structure predictions can be used to identify potential linker regions between domains and secondary structure elements. It might be worth while to express multiple domains in some cases, as the neighbouring domains can stabilise each other and create functional entities. Also, domain boundaries are not definite even between highly related proteins, and one has to keep an open mind.

Once the expressed part has been defined, one needs to decide how the corresponding DNA can be inserted into an expression vector. With the powerful Polymerase Chain Reaction (PCR) procedure one can amplify just the desired part of the DNA and at the same time introduce restriction enzyme recognition sites to both ends to facilitate cloning.

The first thing to do is to check which restriction enzymes cleave the DNA to be cloned, as any enzyme that cleaves the target DNA cannot be used in the subsequent clonings. There are several computer programs which can perform restriction analysis and any of them should give the same result. Just make sure the database of restriction enzymes is relatively recent and contains the

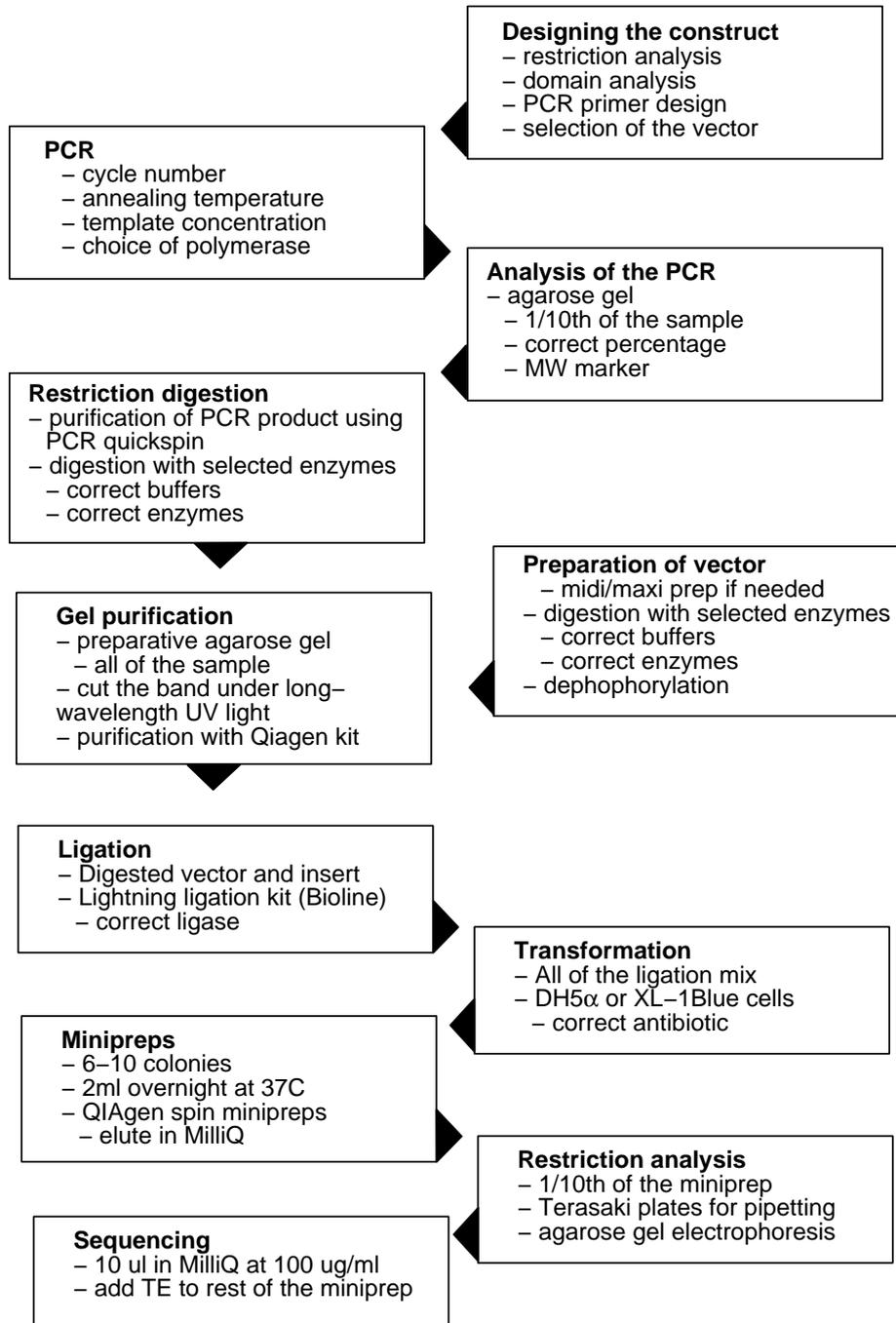


Figure 1: Overview of the cloning protocol

enzymes one might like to use (see below). A good program should list both the enzymes that do and do not cut, as this way you can be sure which ones to choose. One such program is mapplot from the GCG package. It produces a compact graphical output of the cleavage sites in the DNA fragment one has analysed and in addition lists all the enzymes that do not cut, ie. the ones we can use later.

Once the restriction analysis has been done, one can start designing oligonucleotides for PCR. A few simple rules should be kept in mind at this point:

- Include the start and the stop codons in the primers, unless you are using a fusion protein (GST, His-tag etc.) in one or both ends.
- Make sure the amplified DNA will be inserted into the vector in the correct reading frame and in correct orientation.
- Add enough nucleotides in the 5' ends to guarantee efficient cleavage by restriction enzymes. Most enzymes are less efficient the closer the cleavage site is to the 5' end of the DNA. 5-6 nucleotides should be sufficient. See New England Biolabs' (NEB) catalogue for details on the most common enzymes.

The enzymes you can use depends, in addition to the results of the restriction analysis, on the vector's polylinker or multiple cloning site. Any enzyme which does not recognise the cloned DNA and does cleave the polylinker, can be used in the primers.

In an optimal case different vectors would have the same or very similar polylinkers. So when needed, the insert can be subcloned easily to a new vector, carrying a different fusion partner or promoting expression in different host organism, without need to order new oligos and redo the PCR. If you are using some of the new cloning systems such as Echo from Invitrogen or Gateway from Gibco/Life Technologies, you need not to worry about this.<sup>1</sup>

If the start and stop codons are to be inserted by PCR, restriction sites carrying the codon ATG for initiation methionine or TAA/TGA/TAG stop codons become very convenient. It is also wise to plan ahead and anticipate which other vectors you might need to try and choose enzymes which are present in those polylinkers as well.

NcoI recognizes hexa-nucleotide CCATGG and is my favourite for the 5' end of the gene. The reasons are many:

- It cleaves a sequence with the ATG codon, it leaves a four-base overhang (unlike the much used and almost as often cursed NdeI which leaves only two-base overhang)
- There are several enzymes creating compatible overhangs (AflIII, BspHI/RcaI)
- It can be filled-in with Klenow fragment or T4 DNA polymerase to create blunt-end ATG codon
- It is found in most of my favourite expression vectors.

---

<sup>1</sup>Thermostable polymerases are error-prone enzymes and any piece of DNA created by PCR *must* be sequenced to verify its correctness. Subcloning a previously sequenced DNA fragment on the other hand is safe and resulting constructs need not be sequenced.

Some of the other restriction sites commonly found in the 5' end of the polylinkers in expression vectors include BamHI (GGATCC, and again there are several enzymes that create compatible overhangs), EcoRI (GAATTC) and NdeI (CATATG), but you really do need to check your vector map for exact details. Also, pay attention to the open reading frame, both in the N- and C-termini of the expressed protein, to avoid introducing frameshifts and hence ruining your construct. Many vectors are also available in three versions which differ in the reading frame of (part of) the polylinker and one should choose carefully the correct version.

*HindIII* is a good choice (my favourite, that is) for the 3' end as its recognition sequence AAGCTT can be fused with the stop codon TAA. But other enzymes can naturally be used, depending on your preferences and the vectors limitations. Again some of the more often used sites include NotI (GCGGCCG-CGC), XhoI (CTCGAG) and SpeI (ACTAGT).

It is highly advisable to introduce a different restriction site to the 5' and 3' oligos as this forces the orientation of the fragment in the vector and reduces the possibility of the vector to self-ligate and thus lower the cloning efficiency. It might seem appealing to use the same site in both ends and save some money by using a single enzyme, but the few pennies you save and the little (if any) time you will save are not worth the extra work and potential troubles you will have later. You can always use any other enzyme compatible with your expression system, but check carefully details of the use of these enzymes, as there is a growing number of thermophilic enzymes requiring higher incubation temperatures etc. Also, if you choose a new, fancy, and perhaps expensive, enzyme for each cloning, the cost of a single construct increases significantly.

Include always at least five to six nucleotides (I use Ts and As as they have lower annealing temperature as Cs and Gs) before the restriction enzyme recognition site to ensure the enzymes can cleave the DNA. Some enzymes are very reluctant to act on sequences at the very end of a DNA fragment; some examples are listed in the Appendices of the New England Biolabs' (NEB) catalogue (which is a good source of information on restriction enzymes in general).

The annealing part of the oligo should be ca. 21-24 nt and carry similar and fairly even composition of nucleotides in both primers for efficient annealing in the PCR tube. *Do remember* to reverse-complement the 3' primer, so that it promotes DNA synthesis towards the 5' primer. You can again use computer programs to aid the primer design, but it is relatively straightforward without them as well. Some people recommend to end the primer in a C or G to make sure the end where polymerase acts upon is tightly bound to the template. I don't think it is absolutely necessary, but there is certainly no harm in doing that either.

Additional factor to keep in mind while designing the 5' primer is the codon usage of the gene. Different organisms use the 64 triplet codons with different preferences, and there are several codons which are used frequently in mammalian genes, but are hardly ever encountered in *E.coli*. It might be that translation of these codons is unusually slow or that the corresponding tRNA is normally found in very small quantities and during a high overexpression the intracellular pool of this tRNA is depleted. A high number of such codons, especially close to the translational start, can lower the expression level of the

target protein tremendously. To correct for this, you can mutate the codons in the amplified DNA fragment to highly used codons of *E.coli* by mutating the 5' primer sequence accordingly. This will of course affect the annealing properties of the primer and should be kept in mind while designing the constructs.<sup>2</sup>

Now that you have designed your primers, you need to order them from the DNA synthesis service and wait a few days. For PCR you need very small amounts and I would suggest you to order the smallest scale the service offers. From a 40 nmol synthesis you can easily get enough oligo to run hundreds of PCRs. And if everything works OK, a single run will be enough.

If you haven't done cloning before, this is the time to make sure you have all the necessary equipment and reagents in the laboratory ready. Make also sure you have done appropriate risk assessments for the techniques involved.

## 3 Cloning

### 3.1 Primer handling

Primers are typically delivered de-salted and lyophilised. As DNA is prone to self-degradation in acidic pH, they should be resolubilized either in MilliQ quality water and pH adjusted to 7-8 or to TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). Depending on the service you use, they either determine the yield of synthesis or not. In case they do, you should resuspend the DNA to final concentration of 1  $\mu\text{g}/\mu\text{l}$ . If you do not know how much oligo you have, resuspend it first to 100  $\mu\text{l}$  (40 nmol scale synthesis) and measure absorbance of a 1:50-1:100 diluted sample at 260 nm. Multiply the absorbance by dilution factor and by a constant 33  $\mu\text{g AU}^{-1}$  to obtain the total amount of oligo in the sample. Dilute the original sample to final concentration of 1  $\mu\text{g}/\mu\text{l}$ . Store the oligos at -20 °C and when in use, keep on ice.

### 3.2 PCR

The PCR is performed in small 200  $\mu\text{l}$  tubes with very thin walls that allow quick heat transfer. In these tubes pipette:

10 $\mu\text{l}$	10x polymerase reaction buffer (comes with the enzyme)
10 $\mu\text{l}$	8 mM dNTP mix (2 mM of each ATP, CTP, GTP and TTP)
1 $\mu\text{l}$	5' primer
1 $\mu\text{l}$	3' primer
1 $\mu\text{l}$	Template DNA (<10ng/ $\mu\text{l}$ )
76 $\mu\text{l}$	MilliQ
<hr/>	
99 $\mu\text{l}$	Total volume

You need just a few nanograms of the template, so dilute your original DNA sample sufficiently in order not to quench the PCR reaction.

As a last, add 1 ul of Taq (or other thermophilic) DNA polymerase and heat to 95 °C for 5 minutes. Many thermophilic DNA polymerase differ for example from *E.coli* DNA polymerase in lacking the proofreading activity. This

<sup>2</sup>A codon usage table compiled from the complete genome of *E.coli*, can be found at the WWW pages of Kazusa DNA Research institute (<http://www.kazusa.or.jp/>).

makes the enzyme less accurate and it will introduce errors more often to the synthesised DNA product. Although in most cases the products are correct, at least when PCRing larger fragment (>500-1000 nt) it is advisable to use other polymerases, such as Vent (NEB) or Pfu, which are more accurate. If possible, it is good idea to use such enzymes routinely. In the few PCRs a structural biologist does in a year, the price of the polymerase will hardly be an issue either. Go for the best. The non-proofreading enzymes introduce also an additional nucleotide (A) to the 3' end of the DNA creating an single base overhang. Sometimes this can be used to aid cloning (T/A cloning), but it needs to be removed if the PCR product is to be cloned as a blunt-end fragment without restriction enzyme digestion. Proofreading enzymes do not introduce additional nucleotides (or more accurately, they remove it), creating blunt-end products. Other factors that affect the quality of the PCR product include the quality of the template, PCR conditions (temperatures) and number of cycles you run. As each cycle introduces more errors, minimising the number of cycle is very advisable—25 cycles should be enough in most cases.

If the PCR machine does not have a thermal lid which prevents condensation of water inside the cap of the tube and subsequent concentration of the reaction mix, add a layer of mineral oil (light paraffin oil) on top of the sample to prevent evaporation and spin the tube in a microfuge for 5 seconds.

If you are preparing several PCR samples at the same time you can make a “master mix” of the buffer, nucleotides, water and enzyme, and add this to the tubes containing templates and primers. In addition to speeding up your work, you will also minimise the handling of the stock solutions. Each time you pipette from a tube, you have a risk of contaminating its content with proteases, nucleases, phosphatases etc.

Times and temperatures for PCR depend on the fragment length and primers length and composition. As rule of thumb use 1 minute extension for each kilobase and 58°C annealing temperature for primers with 24 bp annealing part.

If you have cDNA library as a template, run 30-35 cycles and if you have isolated cDNA for template, run 25 cycles. In the end you can add a single 5 minute elongation step to ensure all the fragments are full-length. A typical program could look like this:

```

                                1 cycle:
5' at 94 °C  Initial denaturation

                                25 cycles:
1' at 94 °C  Denaturation
1' at 58 °C  Annealing
2' at 72 °C  Elongation

                                1 cycle:
5' at 72 °C  Filling-in partial products

                                1 cycle:
  4 °C  Keep the sample cooled until you continue with the next step

```

Place the PCR sample(s) into the PCR machine and start the run.

### 3.3 Analysing the PCR products

While the PCR reaction is running, pour a 0.8–2 % agarose gel. The percentage depends on the size of the fragments you wish to analyse. Use 2 % gels for fragments smaller than 500 nt, 1.5 % gels for fragment up to 1000 nt and 1 % gels for bigger ones. 0.8 % gels are suitable for analysing digested plasmids which can be several kilobases in size.

An agarose gel is cast as follows:

- Weight required amount of agarose into a 250 ml bottle or flask.
- Add 50 ml of Tris-borate-EDTA (TBE) buffer and boil in a microwave until agarose is completely dissolved. Be careful not to burn your hands: agarose solutions have a tendency to overheat and the sample can “boil” vigorously out of the flask when stirred. Do not boil too long as the water evaporates quickly and you end up with higher agarose and buffer concentration in the gel.
- Cool the solution to ca. 50 °C under running water. Do not cool too much or the agarose solidifies. You can melt the agarose again in the microwave should this happen.
- Add 5  $\mu$ l of 1 % ethidium bromide (this is a carcinogen and should *only* be handled with gloves on). Dispose of any material that has been in touch with EtBr correctly into the EtBr waste.
- Cast the gel into the gel chamber which has been fitted with correct combs—for analytical gels such as this, use combs with small teeth (typically 3–4 mm wide) For preparative gels use large well size, often 8–10 mm wide. Again, the gel apparatus and its components are most likely contaminated with EtBr, so handle with gloves only.
- Let the solution cool and agarose to solidify.
- Remove the stoppers and the combs from the gel and pour enough of TBE buffer to cover the wells.

Once the PCR has finished, remove the tube(s) from the machine and place on ice. Take 10  $\mu$ l of the reaction mixture (lower phase if you added mineral oil) and add into it 1  $\mu$ l of 10x DNA loading buffer (98 % glycerol, 0.1 % SDS and bromophenol blue). Remember also to take out molecular weight (MW) markers for the gel. 100 bp markers are available from several suppliers and they are very nice for fragments up to 1000 bp in length.

Load sample(s) and MW marker next to each other in the gel and run the gel at 80 volts until the dye is almost at the edge of the gel. DNA, as you should know, is negatively charged and migrates towards the positive pole of the electrophoresis chamber—connect the leads correctly or you will lose your samples.

After the gel is finished, pour the buffer into appropriate waste bottle (wear gloves as the gel and the buffer contain ethidium bromide). Take the gel to the UV transilluminator and take a picture of the gel for your laboratory book (UV light can do serious damage to your eyes, so very UV rated protective eye wear). If you see a DNA fragment in your PCR sample of correct molecular

weight, you can continue to the next step. Otherwise better think what went wrong and repeat the PCR. If there are no other bands on the sample or no smear, it might be that one or both of the primers failed to anneal and the annealing temperature should be lowered. If there are bands of wrong size, the annealing has been unspecific and you should probably raise the annealing temperature.

Discard the gel properly in a special container. Ethidium bromide containing gels should never be put into normal waste.

### 3.4 Purification of the PCR fragment

It is important to purify the PCR product at least once from a gel, but you can do it after the restriction digest, purify the DNA after the PCR using simple spin columns. Protocol for Qiagen's PCR Quickspin purification kit is as follows:

- Add 5 times the PCR reaction volume of PB buffer from Qiagen's Qiaquick PCR purification kit.
- Pipette the sample into a spin-column.
- Spin the sample for 30"-60" max. rpm on a minifuge.
- Discard the flow-through.
- Add 750  $\mu$ l of PE buffer (make sure ethanol has been added to it !). Spin as above and discard the flow-through.
- Spin once more to dry the column.
- Place the column into a new eppendorf and add 45  $\mu$ l of EB (elution) buffer on the filter of the column. Let stand for one minute, this will increase the yield of elution.
- Spin for 60" and keep the flow-through. There is your DNA.

Now you have your DNA fragment purified from the oligos, polymerase, buffer, free nucleotides etc. and it is ready for restriction digestion.

For gel purification of the PCR fragment, should you wish to do it at this stage, you will need to run the the whole PCR reaction plus 10  $\mu$ l of DNA loading buffer in a wide slot. The gel is run as above and after it is finished you should do as follows:

- After electrophoresis cut the band under UV light and put the agarose piece into an eppendorf tube. Minimise the exposure of DNA to UV light as this can modify the DNA. The longer the wavelength, lower the intensity and shorter the exposure, the better.
- Weight out the agarose in the tube.
- Add 300  $\mu$ l of buffer QG from Qiagen's Qiaquick gel extraction kit per 100 mg of agarose.

- Place the into 50°C heat-block until the agarose has dissolved completely. Mix frequently to speed the melting.
- Add 100  $\mu\text{l}$  of isopropanol per 100 mg of gel and pipette the sample into a spin-column.
- Spin the sample for 30''-60'' max. rpm on a minifuge.
- Discard the flow-through.
- Add 750  $\mu\text{l}$  of PE buffer (again, make sure it has ethanol !). Spin as above and discard the flow-through.
- Spin once more to dry the column.
- Place the column into a new eppendorf and add 45  $\mu\text{l}$  of EB (elution) buffer on the filter of the column.
- Spin for 60'' and keep the flow-through. There is your DNA.

### 3.5 Restriction digestion

Which restriction enzymes to use in digestion will depend on the sites introduced in primers. You can do digestion either at the same time with both enzymes or one after another. Normally a double-digestions work just fine and you can put both enzymes in at the same time. Consult the catalogue of the enzyme manufacturer to see which buffer gives optimum activity for both enzymes. Some enzymes require BSA for stability and if either of the two enzymes need it, add it in—no enzyme is inhibited by it.

44 $\mu\text{l}$	PCR product from previous step
5 $\mu\text{l}$	10 x restriction enzyme buffer
(0.5 $\mu\text{l}$	acetylated BSA, 10 mg/ml)
1 $\mu\text{l}$	Enzyme I
1 $\mu\text{l}$	Enzyme II
<hr/>	
50 $\mu\text{l}$	Total volume

Incubate at least 2 hours, preferably over night, at 37 °C (or higher if the enzymes are thermophilic). If the other enzyme needs for example higher salt and a double-digestion is not possible, you can add some sterile NaCl to give desired concentration after 2 hours, add the second enzyme and continue incubation for further 2 hours. If one of the enzymes requires higher incubation temperature, add that first, incubate minimum 2 hours at its optimal temperature, place the sample on ice to cool, add the second enzyme and continue incubation at 37 °C.

After incubation, purify the digested fragment by running a preparative agarose gel as described above for the PCR product. By doing this you will isolate your desired DNA fragment from other sized DNA (unspecific PCR products, leftovers from template DNA) and from the small fragments released from both ends by the restriction enzymes.

### 3.6 Subcloning to a new vector

Sometimes your gene fragment has already been cloned to an (expression) vector and you only wish to transfer it to a new vector with new fusion, promoter, or whatever. If the restriction sites in the old and new vector are compatible with each other and in the same reading frame, you can simply digest the old construct to release the insert and insert it straight in to the new vector avoiding the PCR step. But make sure you really do have compatible restriction sites in same orientation and that the reading frame remains correct.

The cloning is done very much the same way as with PCR cloning, the only difference is that you digest the old expression construct instead of the PCR fragment. You might need to transform the old construct to DH5 $\alpha$  or other good cloning strain and make a miniprep to get enough DNA for cloning. Assuming the old construct has been sequenced and verified to be correct before, the resulting new construct should also have correct sequence—introduction of mutations in subcloning is very unlikely. But to be on the safe side, do sequence this construct too.

So to do a subcloning, use the protocol above, but start from the restriction digest.

### 3.7 Preparing the plasmid for cloning

While you are digesting the PCR fragment you can also prepare the expression vector for cloning. What you need to do is to digest the vector with restriction enzymes and dephosphorylate to prevent self-ligation of a partially digested vector.

You will of course need to prepare larger quantities of the vector beforehand, unless someone is kind enough to give you some. Again, several ways to do this, I prefer to use QIAGEN's midi- and maxi-prep kits with columns to purify the plasmid. Which ever method you choose to use, always propagate the plasmid in a good, safe cloning *E.coli* strain, such as DH5 $\alpha$  or XL-1Blue. This will yield more and better quality DNA. Never use BL21(DE3) or similar "wild" strain for plasmid preps.

Use same restriction enzymes as you used for your PCR product digestion, unless you are using restriction sites which create compatible cohesive ends (RcaI and NcoI for example) or you are doing blunt-end cloning.

Pipette into an eppendorf:

10 $\mu$ l	Plasmid DNA (10 ug)
5 $\mu$ l	10x restriction enzyme buffer
33 $\mu$ l	MilliQ water <sup>3</sup>
(0.5 $\mu$ l	acetylated BSA, 10 mg/ml)
1 $\mu$ l	Enzyme 1
1 $\mu$ l	Enzyme 2
<hr/>	
50 $\mu$ l	Total

Incubate at 37 °C at least 2 hours (preferably o/n). To reduce unwanted self-ligation of the plasmid, use alkaline phosphatase (AP) to dephosphorylate the vector. It is not strictly necessary, especially when using two different enzymes leaving non-compatible hanging ends, but dephosphorylation will reduce the

background of clones lacking the insert, and as a result you will typically need to analyse only few clones in the end.

You can do the dephosphorylation during the digestion by adding the phosphates in the sample for the last 1-2 hours. After dephosphorylation the enzyme has to be completely inactivated. In case of Shrimp Alkaline Phosphatase (SAP) it is done by incubating 15 minutes at 65 °C. If you have used calf intestine AP or *E.coli* AP, the sample has to be purified by phenol extraction and ethanol precipitation or using the Qiaex kit. Any residual phosphatase activity will inhibit the subsequent ligation by dephosphorylating the insert too.

1 hour before the end of the digestion, add  
1  $\mu$ l Shrimp Alkaline Phosphatase (SAP)

Incubate 1-2 hours at 37 °C and inactivate for 15 minutes at 65 °C.

Run the whole sample (+ DNA loading buffer) in a preparative 0.8 % agarose gel. Cut the band out of the gel and purify as instructed above. The vector is ready for ligation.

If you did not dephosphorylate the vector during restriction digestion, you can also do it after the gel purification. Pipette into an eppendorf:

45 $\mu$ l	Purified, digested plasmid
5 $\mu$ l	SAP buffer
1 $\mu$ l	Shrimp Alkaline Phosphatase (SAP)
<hr/>	
50 $\mu$ l	total volume

Incubate 1-2 hours at 37 °C and inactivate for 15 minutes at 65 °C. The vector is ready for ligation.

If you are planning to do more clonings with the same vector using the same restriction sites, it is convenient to prepare a larger preparation of digested, dephosphorylated plasmid in one go (you will need only few hundred nanograms of it for each ligation), and store it at -20 °C.

### 3.8 Ligation

In ligation the vector and insert should be in equimolar concentration. You can measure the DNA concentrations accurately using spectrophotometer or semi-quantitatively using ethidium bromide agarose plate and concentration standards, but in most cases this is not necessary. You can more or less guess the concentration of the vector assuming that very little was lost in the digestion and dephosphorylation and as long as you have enough of insert (in my experience: slight excess of it), it should be OK. I use Quick Ligase from NEB, which thanks to its "magic" buffer, allows ligation to be done in only five minutes at room temperature. I follow the manufacturers protocol, but half the volumes they recommend—this way the final volume is 10.5  $\mu$ l.

The protocol is as follows:

3 $\mu$ l	plasmid
2 $\mu$ l	insert
5.0 $\mu$ l	2X Quick Ligase Buffer
<hr/>	
0.5 $\mu$ l	T4 ligase

Incubate 5 minutes at room temperature. Ligation is ready for transformation. Dilute 1  $\mu\text{l}$  with 9  $\mu\text{l}$  of TE buffer, and use this to transform DH5 $\alpha$  or other good cloning strain. Transformation is inhibited by high DNA concentrations, and hence the dilution is strongly recommended. You can transform the rest of the ligation mix undiluted, and compare the results in the end.

Do also a control ligation with vector alone (no insert) and transform to *E.coli* to check how well your dephosphorylation has worked. The fewer colonies you get, the better. Hopefully none.

If you use more traditional ligases, follow this protocol:

1.2 $\mu\text{l}$	ligase buffer
6 $\mu\text{l}$	plasmid
4 $\mu\text{l}$	insert
0 $\mu\text{l}$	MilliQ water
0.8 $\mu\text{l}$	T4 ligase
<hr/>	
12 $\mu\text{l}$	total volume

Incubate at room temperature for at least 2 hours, preferably over-night. Optimum temperature for ligation is 16 °C, so if you have a suitable incubator for this, do use it. In practice room temperature tends to work well enough. If you are doing blunt-end ligations, you should be more careful with the ligation conditions.

The last, but not the least, you should transform your ligation into *E.coli* as described earlier. Again, use 1:10 diluted sample for transformation. If everything is OK, you should get plenty of colonies using 2-5  $\mu\text{l}$  of the ligation mix for transformation. Follow a separate protocol for transformation.

## 4 Analysing the transformants

Now that you have successfully obtained hundreds of antibiotic resistant colonies of *E.coli*, you have to make sure that at least one of them carries your gene. There are two common ways to analyse the transformants: colony PCR or minipreps.

In colony PCR you take a bit of *E.coli* cells with a toothpick and put them into a PCR tube together with primers you used to amplify your insert. Assuming the primers only amplify the fragment you cloned and nothing in the *E.coli*'s genome, the result of the PCR should tell which of the colonies carries the insert.

The other, and my opinion preferred method, is to grow small (2 ml for high copy number plasmids) cultures of cells over-night and purify plasmid DNA from them. By digesting the purified plasmids with appropriate restriction enzymes you can see if the fragment of correct molecular weight can be released. The latter method offers a few advantages, although you must wait until the next to let the bacteria grow. Firstly the miniprep DNA can be used to transform expression strain (in the T7 system this would normally be BL21(DE3) or its derivative). A restriction analysis can be used to verify the presence of an insert and later the same DNA can be used for sequencing. And, of course, you will have pure DNA which is a lot safer way of storing your vector than plates or glycerol stocks of transformed bacteria. To prepare miniprep DNA,

follow a separate protocol again. And make sure to elute/resuspend the DNA in the end to MilliQ, as salt and buffers will interfere with automated DNA sequencing.

#### **4.1 Restriction analysis**

Once you have plasmid DNA prepared, you can cut it with selected enzymes to identify the insert. Normally the enzymes to use would be the same as the ones you used in the cloning. If, however, the ligation did not regenerate the cleavage site (as could be the case using compatible cohesive ends created by different enzymes), other enzymes has to be selected. All T7 expression vectors carry an XbaI site just upstream of the ribosome binding site and this can be used for restriction analysis. Released fragment will have only 15-20 extra nucleotides in front of the coding region. Alternatively you can check from the restriction map of the insert, if any enzyme would cleave close to the 5' end. If this enzyme does not cleave the expression vector, it can be used in combination with a 3' enzyme to verify both the size and orientation of the insert, and the confirm the presence of the recognition sequence.

In practice, you will need only few hundred nanograms of DNA for this analysis. Pipetting a large number of samples can be time consuming, since you will need to pipette separately the buffer, DNA and both of the enzymes. Small microtiter plates with conical wells, called Terasaki plates, are very handy in this step. The wells take up to 20  $\mu$ l of sample, so you can easily do restriction analysis in them. Pipetting is fast, since the sample wells are next to each other and there are no tube caps to open and close. Pipette first (without changing the pipette tip) the restriction buffer to each well. Add them the DNA sample to each well with own tips (you do not want to cross-contaminate your minipreps). In the end, add mixture of the two enzymes to the samples. Put drops of water to the corner of the plate to limit sample evaporation, close the lid and place to 37 °C incubator. After two hours add DNA loading buffer to all the samples and load them on a gel with a molecular weight markers. Run the gel, take a picture under UV light and analyse the results.

If you used alkaline phosphatase to dephosphorylate the vector and worked carefully throughout, the efficiency of cloning (correct clones vs. incorrect) should be close to 100 %. But even a single correct clone will be sufficient.

#### **4.2 Final confirmation of the constructs**

Restriction analysis (or colony PCR) indicated you which clones to use for expression tests. Pick a couple of them, transfer the DNA to an expression host and check for expression (see a separate protocol). Remember to include a proper control for the expression test - empty vector in the same expression host as a negative control and for example GST fusion construct for a good positive control. Hopefully all the clones you chose express large amounts of soluble, functional protein.

As a final check on the expression construct you should sequence the insert. So far you have verified that the correct insert is in the plasmid and it expresses a protein with correct molecular weight. But these assays are only based on the molecular weight of the DNA or protein, but you cannot be sure that the DNA sequence is 100 % correct. This you can only verify by sequencing. For this

you need sufficient amount of DNA (check from your local DNA sequencing service for their requirements) and a pair of primers that anneal outside the coding region of your insert. Often expression vectors have common elements which are used in other vectors (like T7, SP6 or T3 promoters) and primers for these sequences are provided by the DNA sequencing service. In other cases you might need to order a new set of primers for this particular purpose. You shouldn't use the PCR primers for sequencing as you will not be able to get sequence immediately downstream of the priming site.

When you get the results back, compare the sequence with the expected one. As it is the protein that you are interested in, you should not worry of silent mutations which do not change the translated protein sequence. So do the sequence comparison on the translated protein sequences in addition to comparing the DNA sequences. If in doubt, always return to the original data sent by the sequencing service and make sure there are no mistakes. Once the you are confident that the protein sequence is correct, you should put a full effort on the expression and purification work.