3.4: Prokaryotic Expression Vectors

The first vector we will consider is the *pUC* family of vectors

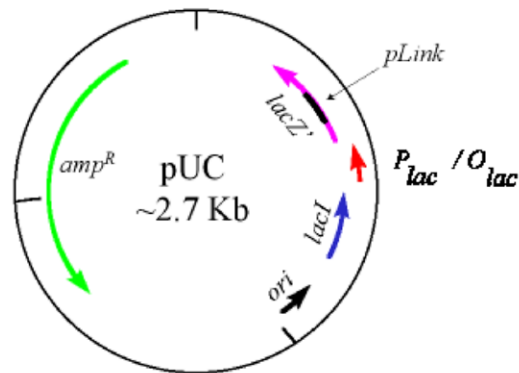

Figure 3.4.1: *pUC vector*

Although it is not typically used for the expression of recombinant proteins, it has all the necessary elements of an expression vector:

1. *An origin of replication*. The pUC family of vectors are *high copy* vectors. They have a ColE1 origin of replication, but a deletion of the *rop* replication regulatory region.
2. *A drug resistance marker*. The pUC vectors contain a gene for *ampicillin resistance* (b-lactamase).
3. *An inducible promoter*. This family of vectors contains the *lac* promoter (P_{lac}) along with the associated *lac* Operator region (O_{lac})

These vectors contain a couple of other elements which add utility to the vector:

1. *The lacI gene*. This gene codes for the *lac* repressor protein. Being a high copy plasmid the host *lac*repressor levels may not be sufficient to efficiently repress the *lac* operator on the plasmids. The plasmid produces *lac* repressor to augment the host levels.
2. *The lacZ' gene*. This gene product is transcribed from the *lac* promoter and produces an amino-terminal fragment of the b-galactosidase protein.
3. *A polylinker (multiple cloning site)*. This short stretch of DNA is located just downstream from the *lac*promoter and after the first few codons of the *lacZ'* gene. It is a short stretch of nucleotides which contains a variety of restriction endonuclease sites.

a-complementation

The *lac* promoter of the pUC plasmid can be induced using the lactose analogue *isopropylthiol-b-D-galactoside*. ("IPTG").

- This will result in the expression of the lacZ' gene.
- This gene codes for an amino-terminal fragment of the b-galactosidase protein
- By itself, this fragment is *non-functional* (it will not hydrolyze lactose or other b-galactosides)

However, b-galactosidase is an interesting protein:

- If we express the carboxy-terminal fragement of the protein it is also non-functional
- But, if we combine both peptide fragments they can combine to yield a *functional* b-galactosidase protein
- This is termed **a-complementation**

The *lac*ZDM15 mutation in a bacteria means that its **genome** has suffered a deletion of the amino-terminal region of the lacZ gene.

- Such a bacteria has a *non-functional* b-galactosidase protein

- If this type of bacteria contains an extra-chromosomal element which expresses the *lacZ'* peptide, it can complement the *lac*ZDM15 protein to produce functional b-galactosidase protein (*a-complementation*)

The galactoside, *5-bromo-4-chloro-3-indoyl-b-D-galactoside* ("X-gal"), can be hydrolyzed by b-galactosidase and produces a **dark blue** color.

- *lac*ZDM15 bacteria in the presence of lactose (or IPTG) and X-gal will appear a normal white-yellow color.
- The same bacteria which harbor the above pUC plasmid will be dark blue in color when IPTG and X-gal are present in the media (due to a-complementation and a functional b-galactosidase protein)

The pUC polylinker region

The pUC polylinker region is a short stretch of DNA which is actually inserted *within* the *lacZ'* gene, just downstream from the start codon:
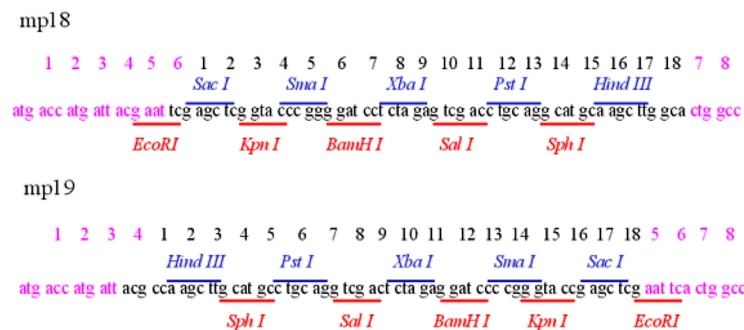

Figure 3.4.2: pUC Polylinker region

> **Note**
>
> In the above diagram the codon numbers for the wild-type b-galactosidase are given in magenta (in this case the initiator methionine is not present on purified protein and the first codon is considered to be the ACC (threonine) residue). The codons from the polyliner (which interrupt the b-galactosidase gene) are given in black.

- The different available plink regions are all a multiple of 3 nucleotides in length, therefore the lacZ' gene remains *in-frame*.
- The resulting *lacZ'*/plink peptide *can still function in a-complementation.*

What would happen to the lacZ' peptide if we inserted a DNA fragment into the plink region?
- The insertion of a random piece of DNA into the plink region will result in *three possible different reading frames* after the insert (depending on the number of nucleotides in the inserted fragment)
- There is a 33% (one in three) possibility that the downstream *lacZ'* gene will be translated in the correct reading frame
- There is also the possibility that even if the reading frame is correct that the insertion of a large peptide at the start of the *lacZ'* protein will prevent a-complementation.
- Therefore, *most of the time*, a DNA fragment inserted into the plink region will *abolish b-galactosidase function*
- DNA inserts in the plink region can therefore be identified by growing the host bacteria on media containing IPTG and X-gal and looking for white-yellow colonies (with the caveat that some blue colonies may in fact contain an insert)

DNA coding for a gene of interest can be inserted in the plink region in-frame with the lacZ' reading frame.

- The protein coded for by this gene can be expressed by induction with IPTG (i.e. using the P_{lac})
- The expressed protein will have as its amino-terminal sequence the first few amino acids of the b-galactosidase gene (in the absence of any other manipulations/mutagenesis). This is also known as a (short) *fusion protein*.
- Although the *lac* promoter is considered to be relatively weak, the high copy number plasmid may result in a useful level of expression of the protein of interest

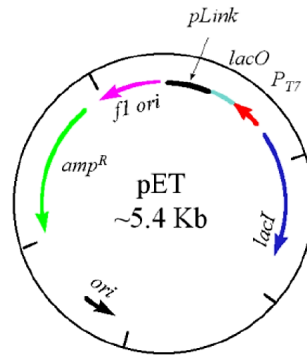## The pET vector system (Novagen, Inc.)

The pET vector looks like this:

Figure 3.4.3: pET Vector

It has the following important elements:

- Ampicillin resistance marker
- ColE1 origin of replication
- f1 origin of replication (allows single stranded vector to be produced when co-infected with M13 helper phage)
- *lacI* gene (*lac* repressor protein)
- T7 transcription promoter (specfic for phage T7 RNA polymerase)
- *lac* operator region 3' to the T7 promoter
- multiple cloning site (polylinker region) downstream of the T7 promoter

The pET vector is a little different from the pUC vector: pUC uses the *lac* promoter and pET uses a promoter from phage T7

- The phage T7 promoter is stronger than the *lac* promoter
- Phage T7 RNA polymerase will specifically recognize the T7 promoter region and will not efficiently transcribe from other promoters
- The T7 promoter will not be efficiently transcribed by *E. coli* RNA polymerase

Where will the phage T7 RNA polymerase come from?

The pET system involves not only an expression vector, but also a *genetically engineered host bacteria*. The host bacteria for the pET vector is typically *E. coli* strain *BL(DE3)*

- This strain has integrated into its chromosome the gene for T7 RNA polymerase
- The T7 RNA polymerase in the host genome is constructed such that it is under the control of a *lac* promoter and operator
- Thus, induction by the lactose analogue, IPTG, causes the host to produce T7 RNA polymerase
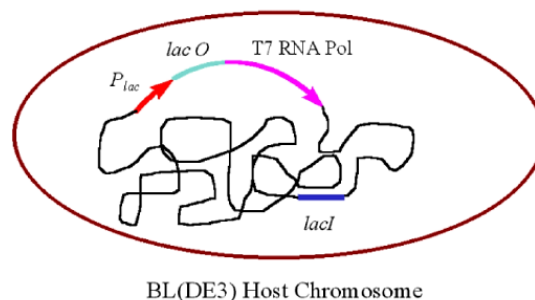- The *E. coli* host genome also carries the *lacI* (repressor) gene

Figure 3.4.4: E.Coli BL(DE3) chromosome

Thus, induction by IPTG results in:

- Derepression of T7 RNA Pol gene on host chromosome with subsequent production of this polymerase
- Derepression of target gene under *lac O* regulation
- Transcription of target gene by T7 RNA Pol

Thus, the system couples a strong promoter with tight regulation (i.e. extremely low level of expression in the repressed state)

The pET vector itself is available with several different polylinker sequences. They contain the same restriction sites, but differ in the *reading frame* leading into the pLink region:

pET 21a

```
  1    2    3    4    5    6    7    8    9   10   11   12   13   14
atg  gct  agc  atg  act  ggt  gga  cag  caa  atg  ggt  cgc  gga  tcc
met  ala  ser  met  thr  gly  gly  gln  gln  met  gly  arg  gly  arg
```

pET 21b

```
  1    2    3    4    5    6    7    8    9   10   11   12   13   14
atg  gct  agc  atg  act  ggt  gga  cag  caa  atg  ggt  cgg  gat  ccg
met  ala  ser  met  thr  gly  gly  gln  gln  met  gly  arg  asp  pro
```

pET 21c

```
  1    2    3    4    5    6    7    8    9   10   11   12   13   14
atg  gct  agc  atg  act  ggt  gga  cag  caa  atg  ggt  cgg  atc  cga
met  ala  ser  met  thr  gly  gly  gln  gln  met  gly  arg  ile  arg
```

plink

Figure 3.4.5: Polylinker sequences

- The *ggatcc* site (BamH I restriction endonuclease site) is the first restriction site in the polylinker.
- The polylinker region is thus available in the three different possible reading frames
- A gene of interest can thus be cloned to be *in-frame* with the transcribed region downstream from the T7 promoter. Again, like the pUC vector the expressed protein will be a fusion protein (with 12-13 amino acids at the amino-terminal end of the polypeptide)

The pTrcHis vector system (Invitrogen, Inc.)
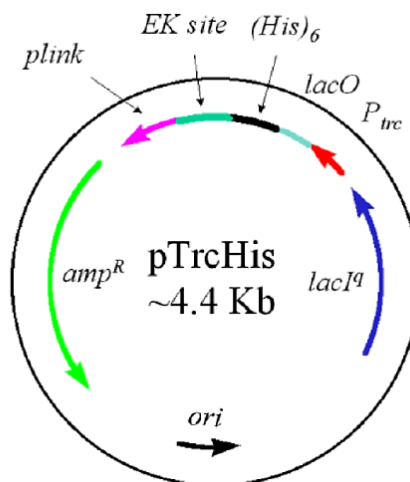
The pTrcHis vector looks like this:

Figure 3.4.6: pTrcHis Vector

The important elements are:

- Ampicillin resistance marker

- ColE1 origin of replication
- *lacI^q* gene. This is a *lac* repressor mutation which is *upregulated* (produces more *lac* repressor than normal)
- *Trc* promoter. This is a hybrid of the *lac* and *trp* promoters which is a stronger promoter in comparison to the *lac* promoter
- The *lac* operator region downstream of P_{trc}. This allows regulation by IPTG.
- An initial stretch of 6 histidine residues in the amino terminal region of the translated protein, also known as a "*His tag*" region.
- An enterokinase (*EK*) cleavage recognition sequence (*asp asp asp asp lys* , with cleavage after the *lys* residue)
- A polylinker region downstream from the EK site

Purpose of the His tag
- Histidine residues can coordinate to form a metal (Ni^{2+}) binding site in a protein.
- The stretch of six His residues forms such a binding site.
- Proteins with a His tag have affinity for *metal chelating resins*, and this characteristic can be used to *selectively purify* such proteins.

Purpose of the EK cleavage site
- Although the His tag can allow rapid and selective purifiction of a cloned protein, the presence of these his residues may prevent normal function of the protein
- The His tag can be cleaved away from the protein by introducing a specific recognition sequence for an endopeptidase
- The sequence "asp asp asp asp lys" is recognized and cleaved by enterokinase. This sequence is not common and it is doubtful that the protein of interest contains another such sequence

Figure 3.4.7: *Enterokinase cleavage*

- His tags/EK sites are typically introduced at the amino terminus of proteins, but can be introduced at the carboxyl terminus as well

Expression hosts

One of the important characteristics of expression hosts (bacteria) is that they allow the expressed protein to accumulate.

- Some expressed proteins may not be folded correctly. Although they may be refolded later to yield functional protein, misfolded proteins may be rapidly proteolyzed in the host bacterium
- Host proteases may selectively cleave precursor forms of expressed proteins to produce mature active forms, which may not be desirable.
- *OmpT* and *lon* genes code for proteases which can degrade expressed proteins.
- Two common mutations in *E. coli* to eliminate the action of host proteases includes the *OmpT^-* and *lon^-* strains.
- *E. coli* strain BL(DE3) is both *OmpT^-* and *lon^-*