E.coli expression systems

- Elements of an expression system
  - Vectors
  - Strains
- About your insert
- What affects the expression level and quality
- Information sources

Marko Hyvönen
Crystallography and Bioinformatics group
Sanger Building, lab 3.58/ office 3.61
http://www-cryst.bioc.cam.ac.uk/~marko/methods

E.coli Expression Systems

The aim of this talk is to give you an introduction to protein production in *E. coli* and to give some general advice on designing your expression construct and optimising your protein production.. Some of the most typical problems will be discussed and possible solutions suggested.

I will not give detailed experimental protocols, as most labs will have their own ways and they best learned in the lab anyway. Should you wish to try things our way, many of the protocols used by us are available as PDF files from the above website.

Why express in *E.coli* ?

- Well established system
- Easy to manipulate
- Large variety of vectors, strains, methods
- Low-tech, safe and inexpensive to grow
- Suitable for variety of labellings
  - isotopes for NMR (^{13}C, ^{15}N, ^{2}H)
  - non-natural amino acids (Se-Met for crystallography)
  - radioactive (^{35}S, ^{14}C, ^{3}H)

E.coli Expression Systems

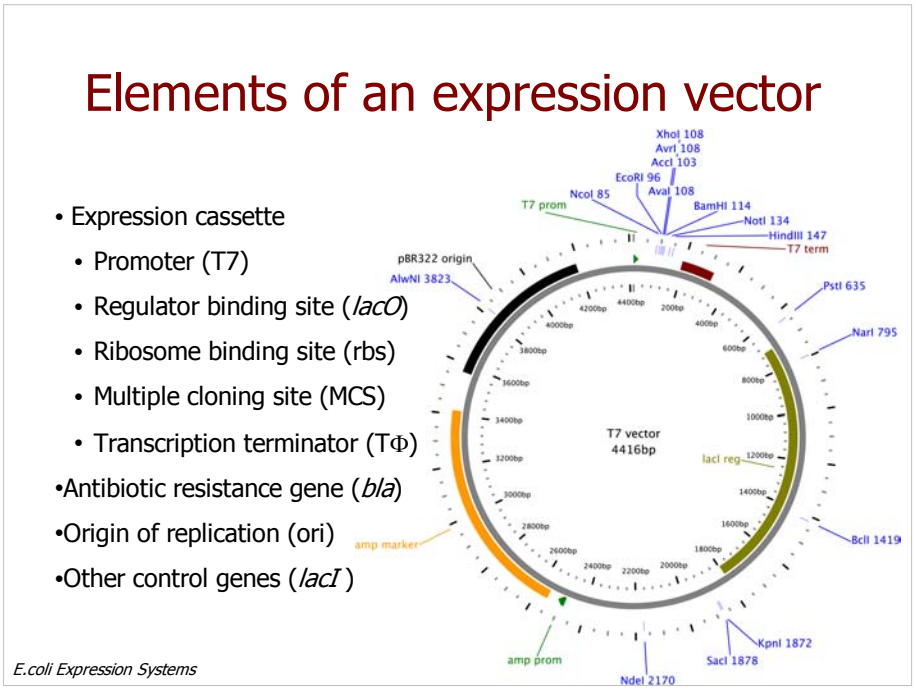
E.coli tends to be the first choice when expressing heteorlogous proteins or their fragments, be it for antigens, ligand hunting, or structural studies. *E.coli* offers several advantages, and is typically the only bacterial expression system people try. More complex systems like yeast, insect cells or mammalian cells are usually tried only after *E.coli* has failed to yield a protein of desired quality.

And why not ?

- Eukaryotic post-translational modifications are missing
  - glycosylation, disulfide bridges, lipidation, proteolytic processing
  - (could be a blessing too...)
- Often problems with solubility of proteins
  - Eukaryotic chaperones missing
  - Too high level and speed of expression compared to the rate of folding of the protein of interest

E.coli Expression Systems

But no system is without its flaws. Bacteria lack most of the post-translational modification eukaryotic cells have, and naturally any protein we express will not be modified.
High level overexpression also causes aggregation problems very often and this can be very difficult to overcome. Reasons for this are multiple, and often most likely a combination of different factors, such as lack of correct chaperones, high speed of expression and lack of obligatory interacting pratners.



Picture above shows diagram of a typical expression vector with an expression cassette containing all the elements needed for regulated high level expression of a protein in E. coli. In addition it has, like in any vector, an antibiotic selection marker, origin of replication and possibly other genes (such as lac repressor) that control the expression system.

The image has been created using PlasMapper plasmid drawing program, available at http://wishart.biology.ualberta.ca/PlasMapper.

Promoters

- *E.coli* natives
  - *lac, trp, tac, trc, ara*
- Viral, but recognised by *E.coli*
  - λ_L, λ_R, T5
- T7, T7*lac*
  - requires its own RNA polymerase

E.coli Expression Systems

Promoters used in E.coli expression vectors can be divided into three categories depending on their origin and mode of function. Examples of all of these can be found in commercial vectors today.

lac & trp

lac

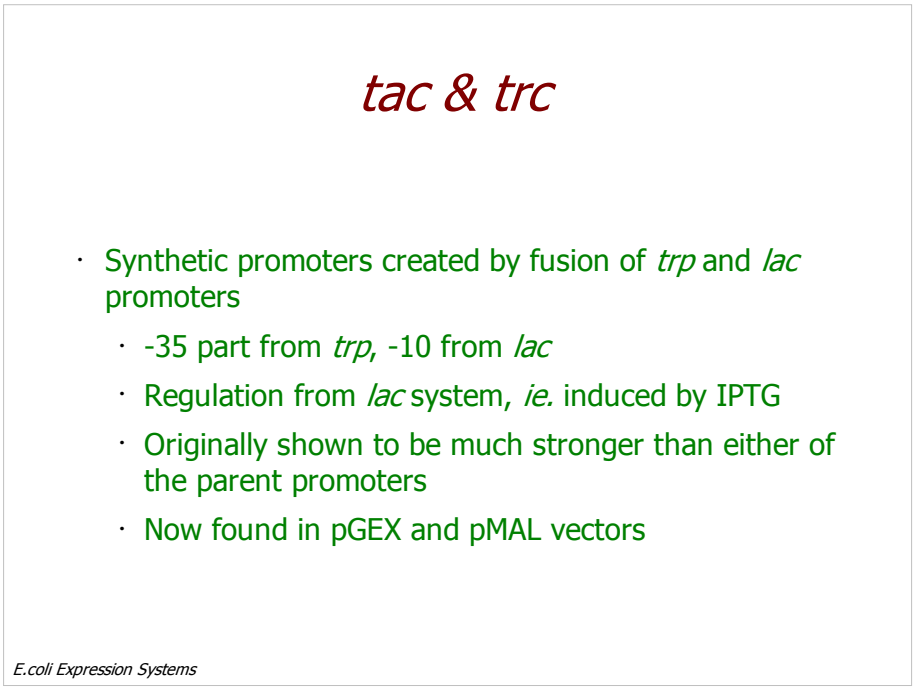
- Promoter of the *lac* operon
- Repressed by *lacI* gene, which binds downstream of the promoter
- Regulated by galatose or its analogues, in expression work non-hydrolysable IPTG used.

trp

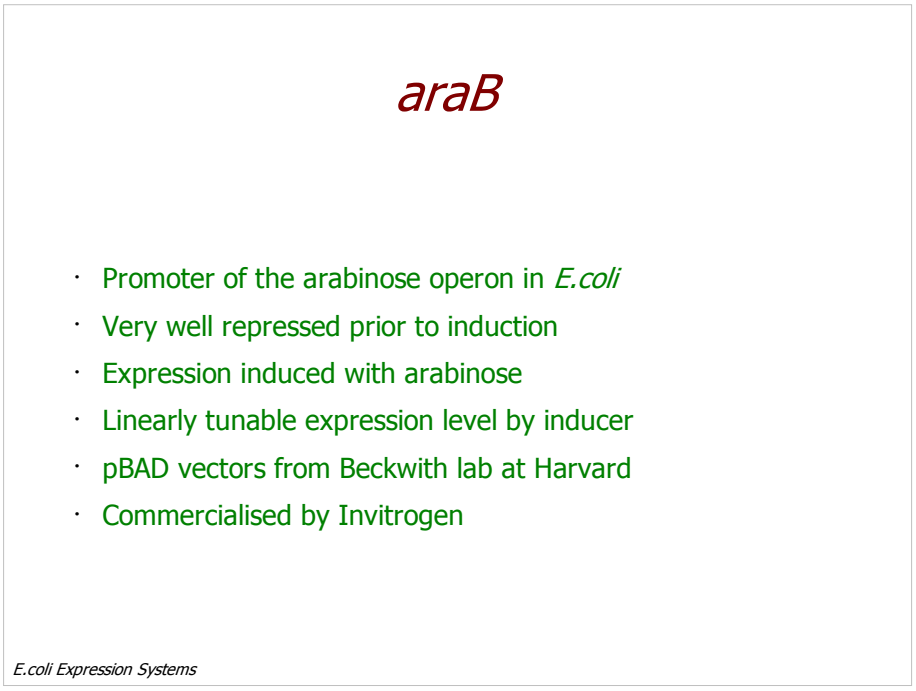
- Promoter of tryptophane biosynthetic enzymes
- Repressed by Trp, so induction done by causing a Trp deficiency with indole-2-acrylic acid

E.coli Expression Systems

E.coli's own promoters are the first ones ever used to drive overexpression of proteins in bacteria. These are strong promoters, and can be induced with relatively inexpensive chemicals, but like we will se in a minute, they are usually superceeded by other promoters.



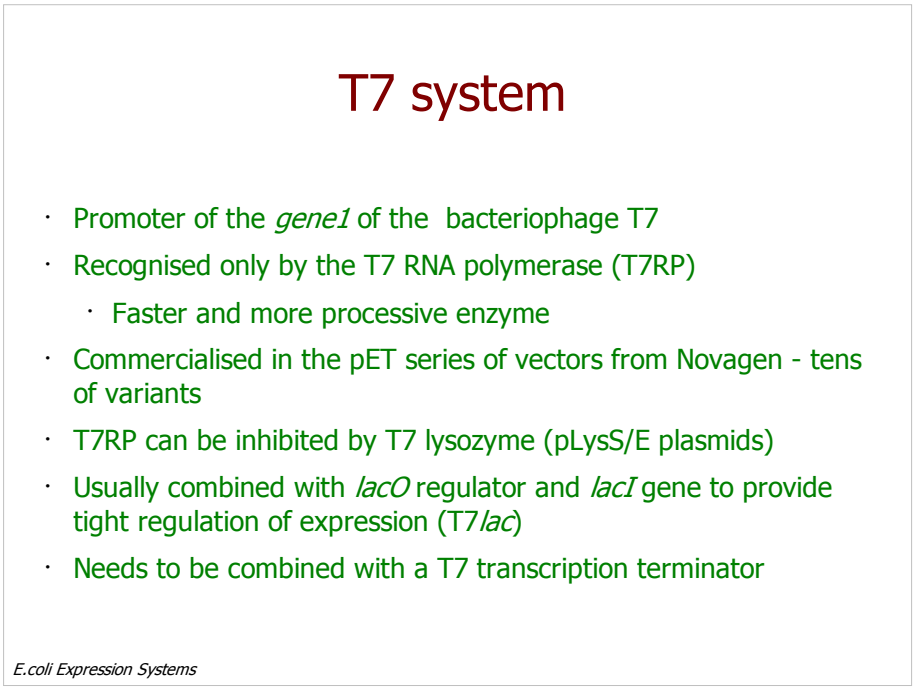
Although not naturally found in *E.coli* the synthetic tac and trc promoters can be classified as *E.coli* promoters, as they are created by fusing different elements of the *lac* and *trp* promoters making them more powerful than either of the parental promoters alone. Several commercial vector systems still use these, inclusing pMAL and pGEx series, and pTRC series from Invitrogen,



Arabinose promoter is perhaps the latest entry to the E.coli promoter family, and offers very tight control of the expression. Several vectors with ara pormoter are available from Invitrogen and in particular the thioredoxin fusions are worth having a look at. One of the advantages of the pBAD vectors is the broad range of inducer (L-arabinose) concentrations where expression occurs and the ability to fine tune the expression level to maximise solubility.

References:

Guzman LM, Belin D, Carson MJ, Beckwith J.(1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.
J Bacteriol.177:4121-30.

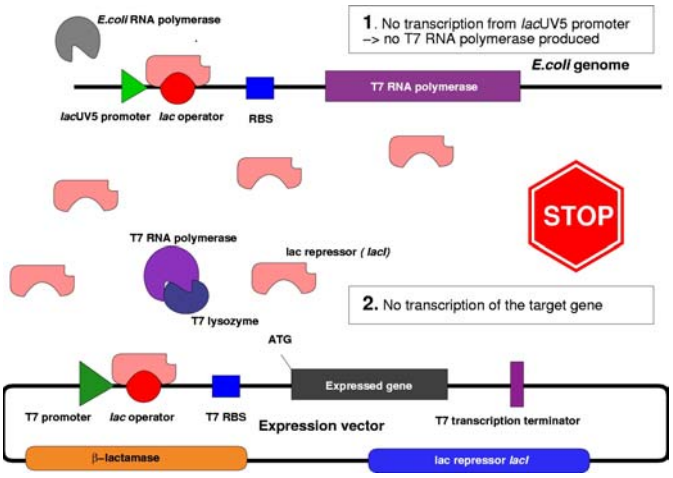


But the system of choice to me is the T7 system which is based on the powerful promoter of gene 1 of T7 phage and the fast and processive RNApolymerase of the same phage.
Originally developed by William Studier in the late 80s it has become the most popular expression system today. Novagen sells the pET system commercially, and they have tens of different vectors with different fusions etc.

References:

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 185:60-89.

Rosenberg AH, Lade BN, Chui DS, Lin SW, Dunn JJ, Studier FW. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene. 56:125-35.



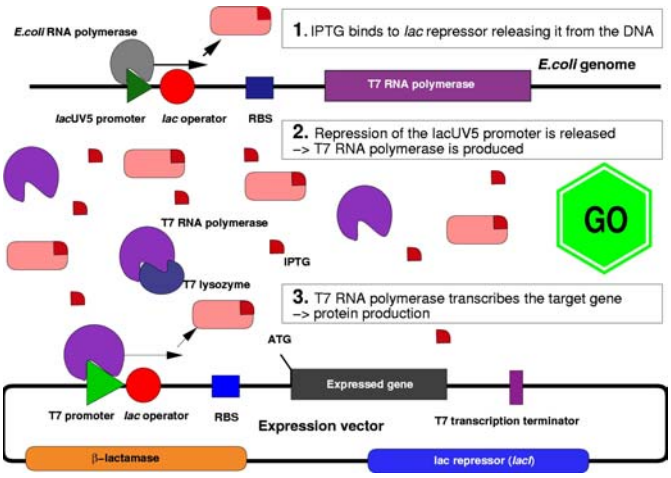
E.coli Expression Systems

The T7lac system offers several levels of control. The lacI repressor, the gene of which is typically in the expression vector, prevents transcription from the lacUV5 promoter and production of T7RP. The same repressor binds also downstream of the T7 promoter preventing transcription in case some T7RP is made. And should this still not be enough, T7 lysozyme can be expressed from a separate plasmid (pLysS or pLysE) to inhibit the polymerase. And as a truly last resort, expression can also be done in a strain with no T7RP gene in the genome, but instread introducing it by λ or M13 phage infection. I am still to see a case whem this would be needed though.

References:

Studier FW.(1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J Mol Biol. 219:37-44.

Dubendorff JW, Studier FW. (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol. 219:45-59.



E.coli Expression Systems

Upon addition of IPTG to the cells the repression by lacI is released and the cells start first making the T7 RNA polymerase. This is turn starts transcribing from the T7 promoter and protein production starts. If pLysS or pLysE plasmids are present, the T7 lysozyme present in the cells will be titered out with the increasing level of T7RP.

Some commercial vector families

- pET from Novagen
  - T7 promoter, lots of variants, pBR322 based
- pGEX from Amersham (Pharmacia)
  - original GST fusions, *trc* promoter
- pQE from QIAGEN
  - T5 promoter, the original His-tag vectors
- pBAD from Invitrogen
  - ara promoter, a few useful fusions
- pMAL from NEB
  - maltose binding protein fusions

E.coli Expression Systems

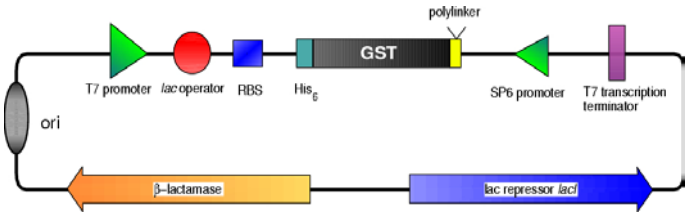
This is by no means a comprehensive list of vectors available, but those that I think are used most frequently and that are commercially available. Check the websites and recent catalogues by these companies for further details. Justa few notes here:

•Some of the pGEX vectors have very long linkers after the thrombin site and beginning of the target gene. These linkers can have recognition sequences for kinases, other targeting or detection sequences, but unless really needed, choose a version without them.
• original pQE vectors did not contain a copy of the repressor (lacI) gene in the expression plasmid, and hense a separate plasmids cvarrying this needed to be used in order to keep the basal expression under control. Later versins (pQE80 series) fixed this and should be used if at all possible.
•pMAL vectors come in two genral variations. the pMAL-p2 series contains the natural signal sequence for MBP and should be used only is secretion to extracellular space is desired. The pMAL-c2 series is for cytoplasmic expression and is the more commonly used variant.

Web links:

pET vectors: http://ww.novagen.com (Look for pET system manual)
pGEX vectors: http://www.gehealthcare.com
pQE vectors http://www.qiagen.com
pBAD vectors http://www.invitrogen.com
pMAL vectors http://www.neb.com

pBAT series of vectors



T7*lac* promoter based vectors and similar to pET, but:
- smaller in size and higher copy number
- with more convenient fusions
  - His6, His6-GST, His6-MBP, His6-Trx

E.coli Expression Systems

The pBAT series of T7 vectors we use in our lab were developed originally by Dr Johan Peranen at the Intitute of Biotechnology, University of Helsinki. They are similar to pET vectors, with T7lac promoter, but with few differences. Although far fewer variants exists compared to pET series, the pBAT vectors come with several useful fusions and with compatible polylinkers between the variants. They are also higher copy number compared to pETs and thus cloning is a bit more convenient. For a full list of these and more details, check my website.

References:
Peranen J, Rikkonen M, Hyvonen M, Kaariainen L. (1996) T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. Anal Biochem. 236:371-3.

BL21(DE3)

- Workhorse of the T7 system
- Carries lysogenic λ phage (DE3) which contains a copy of the T7 RNA polymerase under the control of *lac*UV5 promoter
- Relatively wild strain, and grows fast
- Recent studies have shown that it is safe strain to work with, *ie.* unlikely to be pathogenic
- Numerous derivatives for special applications
- defective in *OmpT* (outer membrane protease) and, as a B strain, of *lon* protease

E.coli Expression Systems

BL21(DE3) is the original expression strain developed by William Studier et al. for the T7 system. It remains as the strain of choice in many cases, and many of the variants listed in the next slides are are based on this strain. It is relatively wild strain, and grows fast and as such is well suited for expression work. Some doubt existed over its safety and ability to colonise human (and other animals) gut, but this seems to have been settled after a specially comissioned study found it to be similar in its pathogenesity to commonly used, safe cloning strains like DH5α.

More expression strains

- BL21(DE3) derivatives
  - Rosetta2(DE3): more Arg, Pro, Gly, Leu and Ile tRNAs
  - BL21(DE3)CodonPlus RP/RIL: similar to one above
  - Tuner(DE3): *lac* permease deletion, linear IPTG titration to control the rate of transcription
  - BL21(DE3)Star: reduced RNaseE activity
  - BL21-AI: arabinose inducible T7RP
  - BL21-SI: Salt inducible variant
  - C41(DE3), C43(DE3): high expressing variants

E.coli Expression Systems

Please see my webpages for more detailed description of these strains at http://www-cryst.bioc.cam.ac.uk/~marko/methods.

And more

- Redox modified strains (allow disulfide formation in the cytoplasm):
  - AD494(DE3), BL21(DE3)*trxB*: *trxB* deletions
  - Origami(DE3), OrigamiB(DE3): *trxB,gor* deletions
- The rec- options
  - BLR(DE3): rec- BL21
  - JM109(DE3): Good old cloning strain
  - HMS174(DE3): Original Studier strain

E.coli Expression Systems

Many extracellular eukaryotic proteins contain disulfide bonds that stabilise their structure, but production of such proteins in *E.coli* can be problematic as the cytoplasm of the bacteria is reducing and formation of disulfide bonds is unlikely and if formed, their stability is very low.
To overcome this problem, strains with more oxidising intracellular environment have bene developed, aminly by the group of Jon Beckwith of Harvard University, and commercialised by Novagen. These strains have deletion of either thioredoxin B (*trxB*) alone or *trxB* and glutathione oxidoreductase (*gor*), and as a result of this allow (some) disulfide bond formation in the cytoplasm. While these strains have been used succesfully in many cases, they are a far cry from the conditions of the native secretory pathway of mammlian cells, and their usefulness is still limited.
The rec- strains are useful if you wish to do your cloning in the expresion strain directly, or are experiencing insert instability in other strains. They are more slowely growing however.

Your insert

- An insert for expression needs to:
  - make sense structurally
  - be in frame with ATG codon and/or any fusion
  - be in frame with the stop codon
  - contain suitable ends for cloning
  - correct in sequence (!)
  - be compatible with the restriction sites you plan to use

E.coli Expression Systems

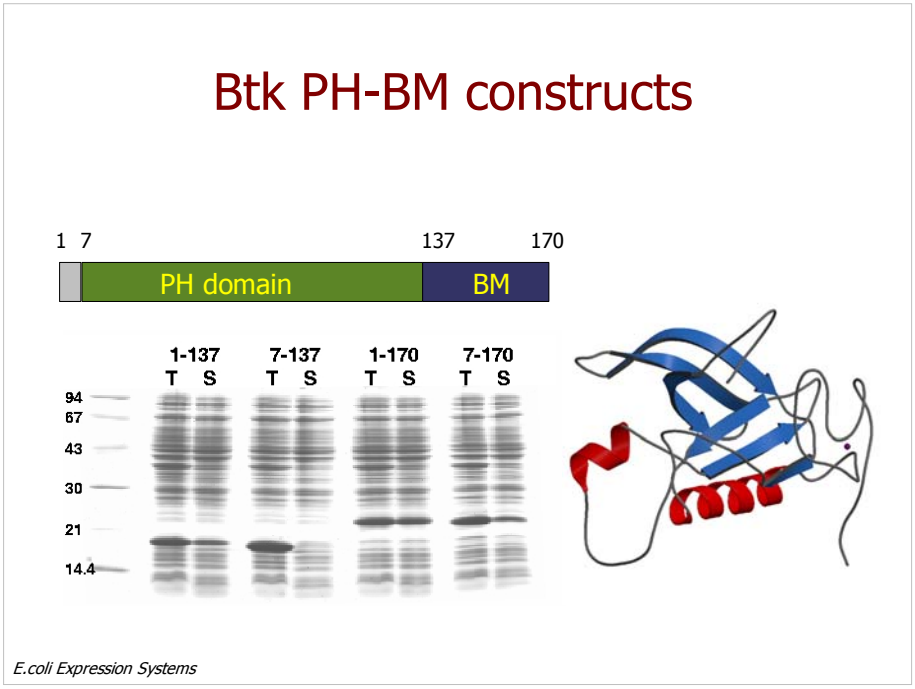
As the vectors used for expression work carry all the elements needed for high level expression, all that is left for you to do is to create an insert that takes advantage of that design.
Perhaps the most important thing is make sure your insert is cloned in the correct orientation and in the right translation frame to produce the protein you have decided to make. If you have no N-terminal fusion, you need to make sure an intiation codon ATG for methionine is present and in frame with the rest of the protein. You also need to make sure a sensible stop codon exists to avoid producing exessively long tails.Most commercial vectors will have stop codons in all three frames, but rather than relying on this, introduce one yourself in the PCR primer in the ideal position. Of course if you are using a C-terminal fusion, you ahve to ensure the reading frame continues in the right frame and the stop codon will be the one provided by the vector. Again, C-terminal fusion vectors tend to come in three version with different reading frame for the fusion, and you have to make sure to clone into the same vector you used for designing the construct in the first place. You will also need to make sure you can clone the insert to the vector(s) of choice by computing a restriction map with your insert's sequence.
Once you have done the cloning, made your minipreps and verified that correct insert is found in the plasmid, you will still need to confirm the correctness of the sequence by sending a sample to sequencing service. PCR (even with the fanciest
error-free polymerases) will create some errors and it would not be sensible to skip sequencing of the construct at this stage.

Mind your ends*

- If you express a part of a protein, do avoid
  - cutting in the middle of a domain
  - cutting in the middle of a well predicted secondary structure element
  - long unstructured tails
- Small difference in the length can make a huge difference in the expression outcome
  - try several slightly different constructs
  - several combinations of adjacent domains

E.coli Expression Systems *© Dr Luca Pellegrini

One of the most crucial decisions you need to make is to choose where to cut the N- and the C-termini (unless you are expressing the full-length protein). Analyse the sequence for known domains (SMART server at smart.embl-heidelberg.de, pfam at www.sanger.ac.uk, etc), and make a secondary structure prediction (JPRED at www.dundee.ac.uk).
Do not start or end your construct in the middle of a well defined domain, as this will most certainly result in inclusion bodies. Also, try not to cut the sequence in the middle of a well predicted secondary structure element. It is not unheard of that an individual domain has longer secondary structure elements at either end of the domain, and seconadary structure analysis can provide hints to this effect. You can also analyse your sequece for the presence of characteristically unstructured regions that might be linkers between the domains.
It is very difficult even for an experience person to get this right, and to maximise the chance of success, you might want to design several constructs with different N- and C-termini. Combinatorial design will allow you to maximise the chances by making all the possible constructs using the limited set of primers. At the extreme, one could envisage a setup where 12+8 primers will be combined to create 96 different constructs, all of which could be tested for solubility, activity etc. using 96 format technologies, robotics etc.



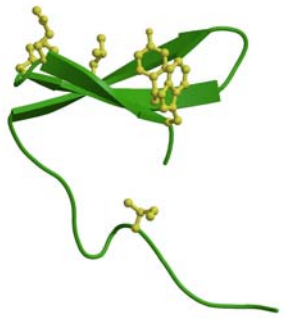
This is an example of an expression construct design where expressing an individual domain (the PH domain), as determined by multiple sequence alignment analysis, resulted in completely insoluble protein.Only inclusion of the adjacent small domain and few extra residues in the N-termini, a fully soluble protein was expressed. Subsequent structure determination revealed a close association between the two domains, and provided us with the explanation to the expression behaviour, although it was not quite clear why also the 6 amino-terminal residues were required for expression of fully soluble protein. And even then the expression was done at 15°C to promote the solubility.
Figure on the right is the final structure solved using construct 1-170 and shows the typical PH domain fold with seven stranded β-barrel and C-terminal α-helix, followed by Btk motif which adheres to the side of the domain and coordiates a zinc atom (shown in purple) with its conserved cyteine and histidine residues.

References:

Hyvonen M, Saraste M (1997) Structure of the PH domain and Btk motif from Bruton's tyrosine kinase: molecular explanations for X-linked agammaglobulinaemia.
EMBO J. 16:3396-404.

YAP WW domain

1	GSM DDVPLPAGW...WQDPRKAMLSQ	degraded
2	GSM SFE**I**PDDVPLPAGW...WQDPRKAMLSQMNVTAPTS	**folded**
3	GS PAGW...WQDPRKAMLSQMNVT	degraded
4	GS VPLPAGW...WQDPRKAMLSQMNVT	degraded
5	GS VPLPAGW...WQDPRKAMLSQMNVTAPTS	unfolded



E.coli Expression Systems

This is another expression construct design problem, where the canonical WW domain sequence was not stable, and got degraded during purification. Only a construct with a long N-terminal sequence was working; Again a structure determination of revealed that an isoleucine in the N-terminal stretch interacted with the domain stabilising it.

References:

Macias MJ, Hyvonen M, Baraldi E, Schultz J, Sudol M, Saraste M, Oschkinat H.(1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature. 382:646-9.

Construct design

- Codon usage
  - rare codons: AGA/AGG, CCC, CTA, ATA, GGA
  - N-terminal codons most important
    - optimisation of 6-10 first codons
    - see www.proteoexpert.com, a Roche facility
- Simplify cloning
  - choose vectors with compatible restriction sites
    - you might need to try different vectors/fusions

E.coli Expression Systems

Other thing to remember when designing you expression construct is to look for the codon usage. A few codons are particularly seldom used in the E.coli and these can become limiting when expressing heterologous proteins. If such codons are found in the v ery N-teminus of the expressed protein, they can be silently mutated to more favourable ones during the initial PCR cloning.
It is also a good idea to design the primers so that you can easil;y use the same insert in different vectors with for example different fusion partners.

Links:

Codon usage tables: http://www.kazusa.or.jp/codon/

N-terminus

- The 5' PCR primer needs to contain:
  - a restriction enzyme recognition site (plus few extra in the 5' end for effective digestion)
  - an ATG codon in frame for initiation methionine
    - followed by small aa to allow for its removal
  - sequence that anneals with the target gene (in frame)
  - optimised codons for *E.coli*
- For example:

```
            M  A   K   T   D   S   A   P
TATATA CCATGGCT AAG ACT GAT TCT GCT CCG
       NcoI
```

E.coli Expression Systems

C-terminus

- The 3' PCR primer needs to have
  - Restriction enzyme recognition site (plus few extra in the 5' end for effective digestion)
  - STOP codon (TAA/TGA/TGA) in frame with your gene
  - Sequence that anneals with the target gene
    - remember to reverse complement !
- For example:

```
        HindIII
TATAT  AAGCTTA GCT AAG GAT TCT GCT CCC
          TAA AGC CTT ATC AGA AGC GGG
         stop  S   L   I   R   S   G
```

E.coli Expression Systems

Typical problems

- Low or no expression
  - codon usage
  - instability of RNA and/or protein
- Degradation
  - ... or premature termination
- Insolubility
  - inclusion body formation

E.coli Expression Systems

Low or no expression

- Are you sure the insert is correct and in frame ?
- Have you checked for rare codons ?
  - Use a strain such as Rosetta(DE3) or Rosetta2(DE3)
- Is the protein being degraded ?
  - Western blot if antibodies available
- Do you have correct expression controls ?
  - same strain with the same plasmid but without an insert

E.coli Expression Systems

Many of the checks above should have been done already before you are as far as expression tests. Careful sequence analysis, and verification of the correct sequence of the insert will allow you to choose the approriate strain for expression and to draw correct conclusions from the test quickly. But not all factors influencing expression behaviour of a protein, and it might be worthwhile to try as many different things as possible in parallel.

Codon usage problem

- Different organisms use codons at different frequencies
- A highly used codon in human might be rare in *E.coli* and vise versa.
  - Arg AGA/AGG (translated by same tRNA in *E.coli*):
    - 2.34% of all codons in human, 0.33% in E. coli
- High number of rare codons in the expressed gene can cause severe reduction in expression level
- It can also lead to mutations at translational level
- The best documented and most common problem is with arginine codons AGA/AGG

E.coli Expression Systems

To check for codon usage, you can either use some of the several sequence analysis packages to calculat the codon usage in your gene, or like I tend to do, go through the sequence by eye and highlight the rare codons, arginine codons AGA/AGG in particular. Highlighting the codons in the sequence will also allow you to see if there are clusters of rare codons which are likely to cause more problems. See the last slide for link to a website with codon usage tables of various species.
If you protein contains fewer than 2 in 100 of the rare Arg codons, they are most likely not lowering the expression yield significantly. If there are tandem pairs of these codons, your are more likely to see an effect even if the total number is under 2% of all codons.
In addition to lowering the expression yield of the target protein, some of the these tRNAs can get substituted with low (but detectable) frequency by other tRNAs once their availablility becomes a lmiting factor in translation process. This will create microheterogeneity in your final protein preparation, and this in turn can have a negative effect on for example crystallisation experiments.

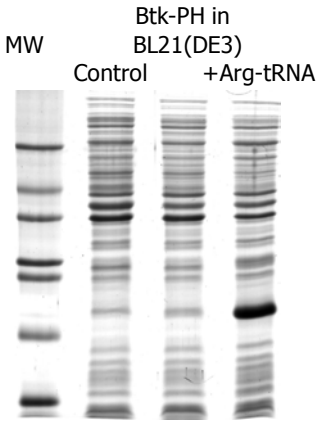
References:

Kane JF. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli.
Curr Opin Biotechnol. 1995 Oct;6(5):494-500. Review.

Del Tito BJ Jr, Ward JM, Hodgson J, Gershater CJ, Edwards H, Wysocki LA, Watson FA, Sathe G, Kane JF (1995) Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli.
J Bacteriol. 177::7086-91.

Effect of too many rare Args

- 170 N-terminal residues of Bruton's tyrosine kinase (Btk)
- 7 AGA/AGGs = 4% of total codons
- 2 tandem pairs
- Expression in BL21(DE3) with or without a plasmid carrying extra copies of the rare tRNA gene
  - pUBS520 plasmid
  - Rosetta, BL21(DE3)RIL strains



E.coli Expression Systems

In this example, a protein with a high content of rare arginine codons (AGA/AGG) is being expressed in BL21(DE3) either with or without an additional plasmid encoding for the corresponding tRNA (*argY* gene). As the concentration of this tRNA is the limiting factor when overexpressing proteins such as this, extra copies of the gene will generate more tRNA and overcome the problem.
Similar systems to the pUBS520 vector used here are nowadays available commercially from Invitrogen and Novagen. The pRARE2 plasmid used in Rosetta2(DE3) strain supplements the cells with extra copies of rare tRNAs for Arg, Pro, Leu, Ile and Gly.
In an extreme case people have synthesised a fully synthtic version of a gene from overlapping oligonucleotides and optimised the codons of the full length coding sequence to match the codon usage of the expression host. While this is more labour intensive and expensive to do, it could be justified if nothing else had helped.
Given the ease by which this still often overlooked problem of differential codon usage can be solved, there is no reason why not to take the necessary precautions and not let your research be limited by it.

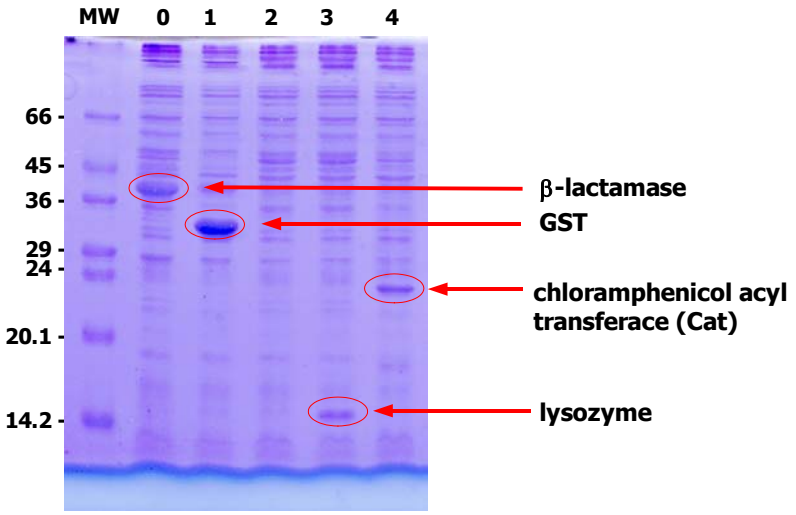
More references:

Goldman E, Rosenberg AH, Zubay G, Studier FW. (1995) Consecutive low-usage leucine codons block translation only when near the 5' end of a message in *Escherichia coli*. J Mol Biol. 245:467-73.

Brinkmann U, Mattes RE, Buckel P.(1989) High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product.Gene. 85:109-14.

Calderone TL, Stevens RD, Oas TG. (196) High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*.
J Mol Biol. 262:407-12.

Where is my protein ?



E.coli Expression Systems

When making an expression test, just like in any other experiment, it is vital to have correct controls.
The above gel is highlighting a few of the proteins people commonly think as their own overexpressed protein. Apart from the lane with GST (lets call it the positive control), induced by IPTG, the other highlighted proteins are either antibiotic resistance proteins (*bla* and *cat*) or lysozyme that is sometimes used for cell lysis.
These are also some of the more common examples proteins people purify by mistake (well, thinking it is their epverexpressed protein) and submit for N-terminal sequencing or mass spec analysis.
An urban legend even tells of a biochemist who provided a crystallographer with a purified prep of their pet protein which crystallised very readily. The work progressed quickly with qood quality crystals, but in the end the structure turned out to be hen egg white lysozyme that had been used to lyse the cells. No wonder it crystallised so easily :-)
So when expressing a protein for the first (or nth) time, make sure to run correct negative controls in the gel as well.

Degradation problems

- Does it happen during expression ?
- Or after lysis during purification ?
- Minimise expression time
  - do a time course of expression
    - up to 3 h at 37°C, overnight at 15°C
- Keep samples chilled as soon as expression is finished
- Try a protease deficient strain
- Use protease inhibitors at all stages
- Make another construct

E.coli Expression Systems

Sorry folks, nothing more to say.

Insolubility

- Very common problem with *E.coli*
- Formation of inclusion bodies (IBs) can occur even with natural *E. coli* proteins (such as β-galactosidase)
- Their formation is a combination of many effects, most noticably the high level expression during which partly folded proteins aggregate before gaining their native structure
- Best to try lowered expression temperature all the way down to 15°C possibly combined with partial induction
- IBs can be beneficial too, as protein in IBs is protected from proteases and relatively pure (but can you refold it ?)

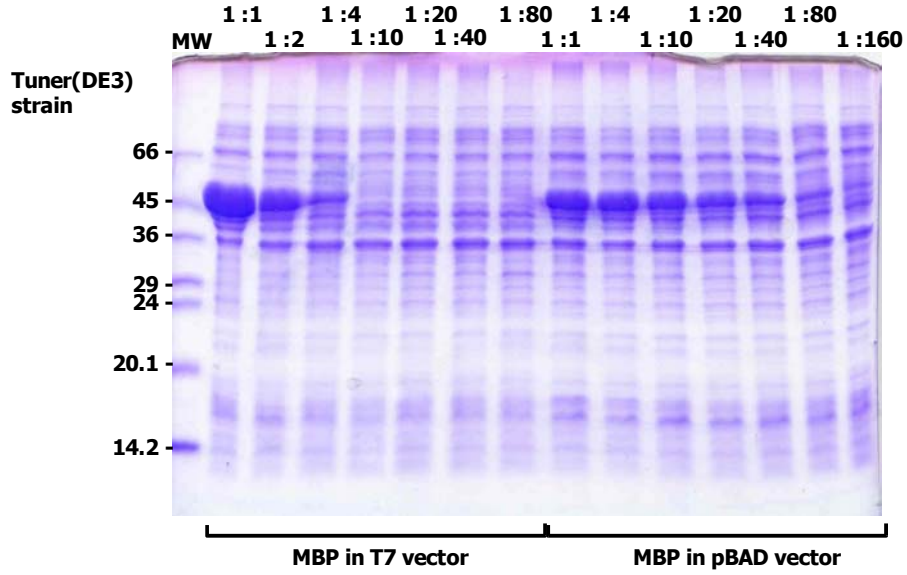
E.coli Expression Systems

References:

Schein, C.H and Noteborn, M. H. M. (1988) Formation of soluble recombinant proteins in Escherichia coli if favoured by lower growth temperature. Bio/technology 6:291-294

Mitraki, A. and King, J. (1989) Protein folding intermediates and inclusion body formation Bio/technology 7:690-697

Regulating the expression level by partial induction



E.coli Expression Systems

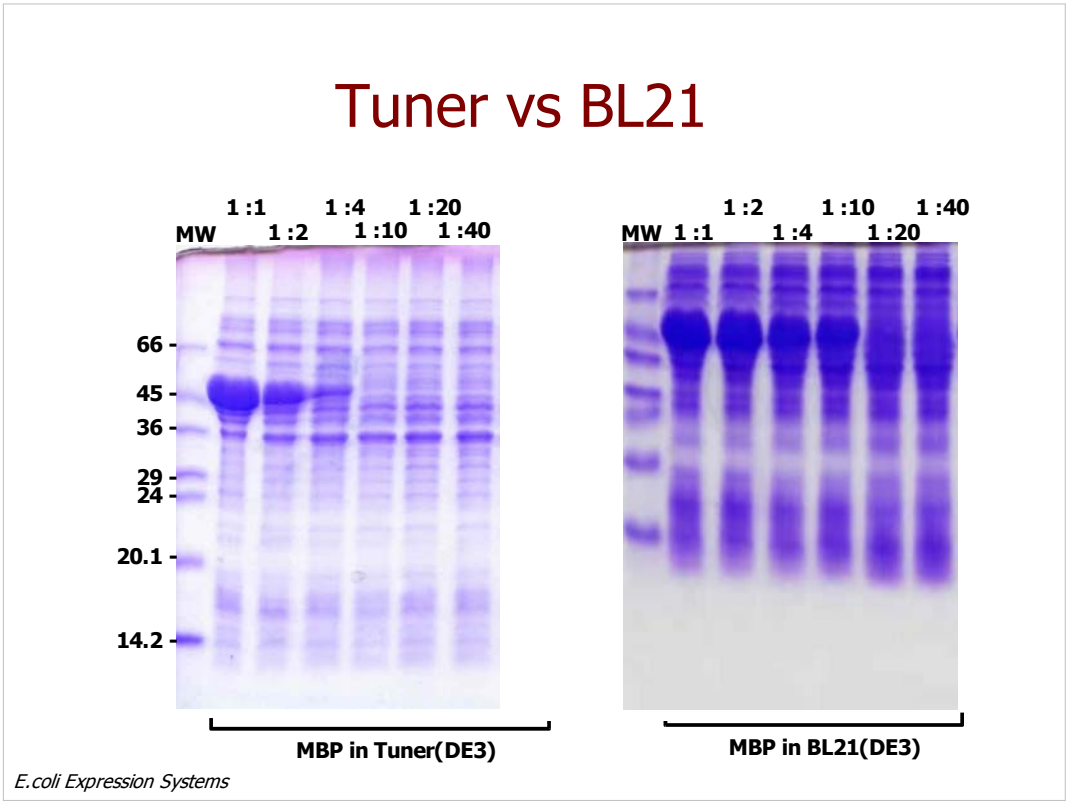
One of most important reasons for inclusion body formation is the very high expression level achieved with modern expression vectors.By slowing down the rate of expression try to five the proteins more time fold and avoid aggregation with other unfolded or partially folded proteins.

There are two principal ways of achieving this. Perhaps the first to try is reduced temperature, whichnot only slows the rate of transcription and translation, but the rate of all processes in the cell too. Typical reduced temperatures to use apart from “standard” 37°C are 30°C, 24°C and 15°C. As things slow down with reduced temperature, expression times need to be adjusted too. From a typical 3 hourr induction at highest temperature, I would increase the induction time to 4 hours at 24°C and overnight at 15° C.

Other option, and something that should be tried in parallel and in combination with reduced temperature, is reducing the amount of inducing agent. The gel above shows an example of inducer titration in Tuner(DE3) cells (more on this in the next slide) both for IPTG induced T7 vector and L-arabinose induced pBAD promoter. The expressed gene in both cases is exactly the same maltose binding protein (MBP).

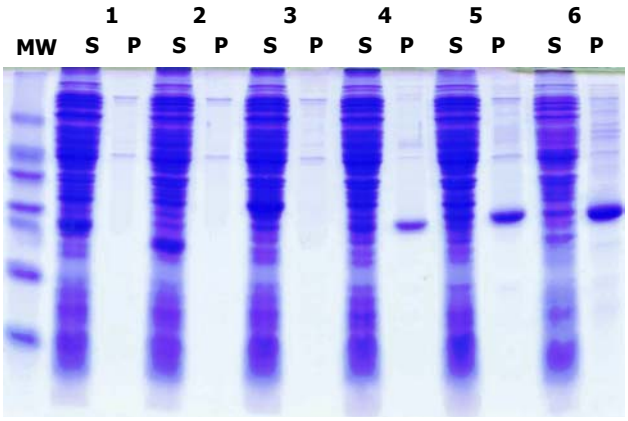
The T7 promoter can be regulated over a relatively narrow window, ranging from maximal induction at 400 μM IPTG to between 40 and 100 μM IPTG.

The arabinose induced pBAD vector expresses clearly less protein at highest inducer concentration of 0.2 % (ca. 13 mM), but its expression can be regulated very finely over nearly 200-fold range.



Traditionally the T7 promoters have been very difficult to control. This is due to lactose permease, product of the *lacY* gene, which actively transports lactose (and its analogues) from the periplasmic space into the cytolasm. As a result, variation of the in the inducing agent (IPTG) concentration inside the cells does not follow the concentration of inducer in the growth medium, and hense the T7 system is virtually an on/off system with very little room for fine-tuning the expression level.
Solution to overcome this has been to create a strain devoid of *lacY* gene so that IPTG is not anymore actively transported into the cells, but rather diffuses in passively. Now we can achieve liner control of the induction.
The two gels above show indution titration of the same MBP expression construct in either Tuner(DE3) or BL21(DE3) cells.
While expression of MBP in the Tuner(DE3) strain is clearly reduced already at half the concentration of IPTG, and is further reduced at ¼ concetration, expression in BL21 (DE3) continues at relatively high level all the way to ten-fold reduction of IPTG concentration, and then stops completely. Although it is possible to achieve intermediate level of expression by more finely sampled titration in BL21(DE3) cells, in practice this can be become a hit-and-miss exercise. With the Tuner(DE3) strain we have wider window of useful inducer concentration and will achieve better control of the experiment.

Testing for solubility



To check for solubility

- cells need to be properly lysed
  - in small scale we use BugBuster *with* lysozyme and DNase to ensure complete lysis.
- you need to have appropriate controls

E.coli Expression Systems

But is there a solution to all these problems?

FUSIONS can

- boost the expression level
- facilitate purification
- increase solubility

More next time....

E.coli Expression Systems

In summary

For successful protein production in *E.coli* one should:

- design the construct(s) *very* carefully
  - domain boundaries, predicted sec. structure, codon usage
  - introduce start and stop codons, restrictions sites etc.
- choose the most approriate vector and most likely try several
  - different promoters, different tags
- optimise expression conditions
  - temperature, induction level, rare codon compensation
  - try different expression strains
- treat an expression test as an experiment
  - appropriate controls, well lysed cells

E.coli Expression Systems

Sources of information

- This talk and more at
  - **http://www-cryst.bioc.cam.ac.uk/~marko/methods**
- Novagen's T7 expression system guides
  - **http://www.novagen.com**
- Invitrogen's manuals and discussion groups
  - **http://www.invitrogen.com**
- Codon usage tables
  - **http://www.kazusa.or.jp/codon/**

E.coli Expression Systems