TECHNIQUES IN GENETIC ENGINEERING

Işıl KURNAZ

CHAPTER 4
Protein Production and Purification

“There are no such things as applied sciences, only applications of science.”
Louis Pasteur (1822 – 1895)

Expression vectors and recombinant protein expression

A good expression vector must

i) be suitable for the system or organism where transcription would take place;

ii) ensure transcription of stable mRNA transcripts, with appropriate promoter and enhancer element(s);

iii) have optimal transcription initiation sequences (or transcriptional start sites, TSS);

iv) contain polyadenylation sequences engineered so as to increase the half-life hence stability of mRNA transcripts (important for eukaryotic expression systems);

v) maintain the reading frame if a fusion protein is to be translated;

vi) contain optimal translation initiation sequences suitable for the translation system (such as a Kozak's sequence prior to the START codon for optimal translation in eukaryotes)

If the protein that is thus translated has to be secreted, targeted to an appropriate organelle, or posttranslationally modified, the appropriate sequence motifs (such as localization sequences, transport sequences, phosphorylation motifs, etc) must also be incorporated or engineered to the expression construct.

In short, a number of key design features must be considered well in advance to achieve the maximum product efficiency in your system.

In vitro transcription and translation

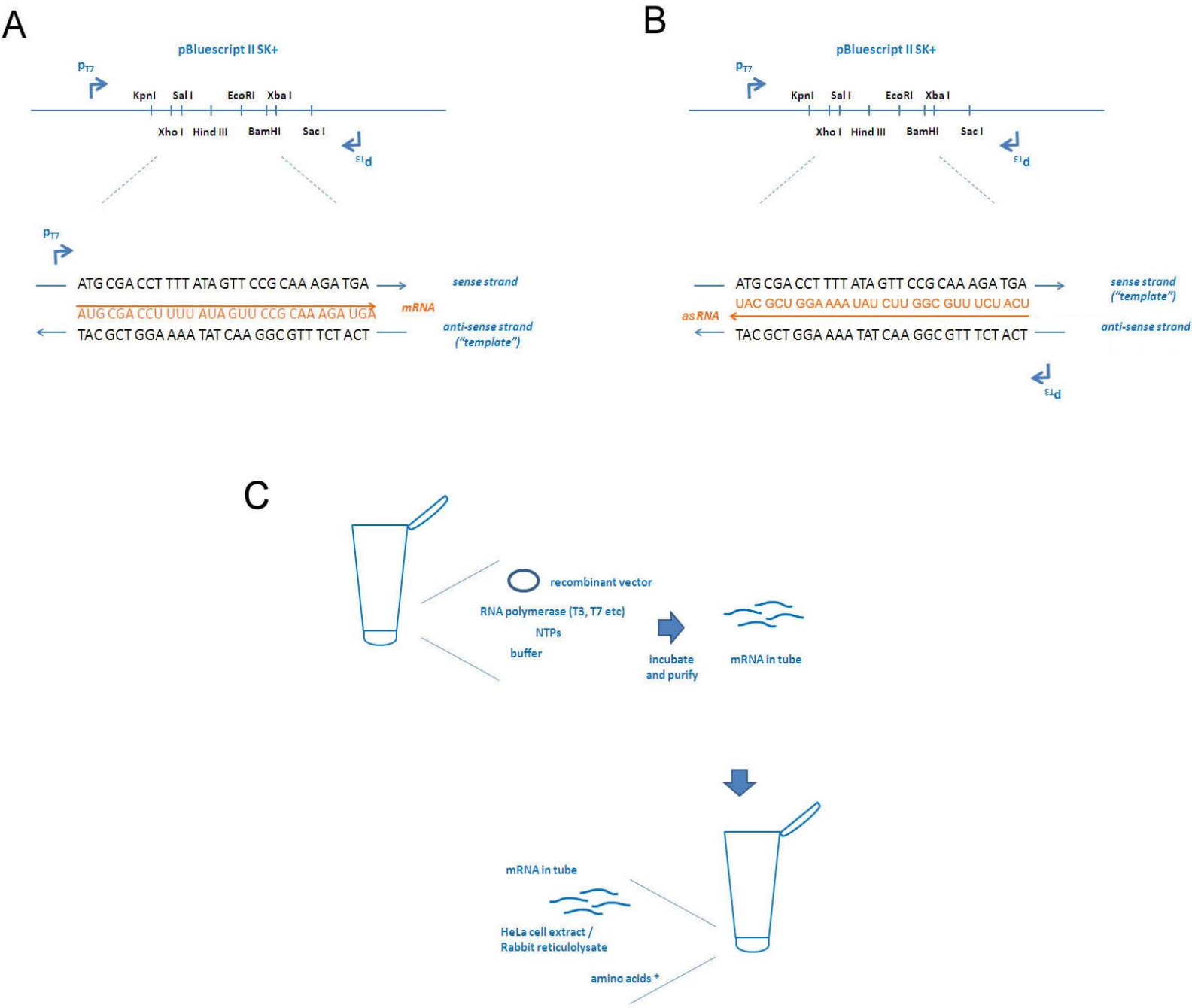


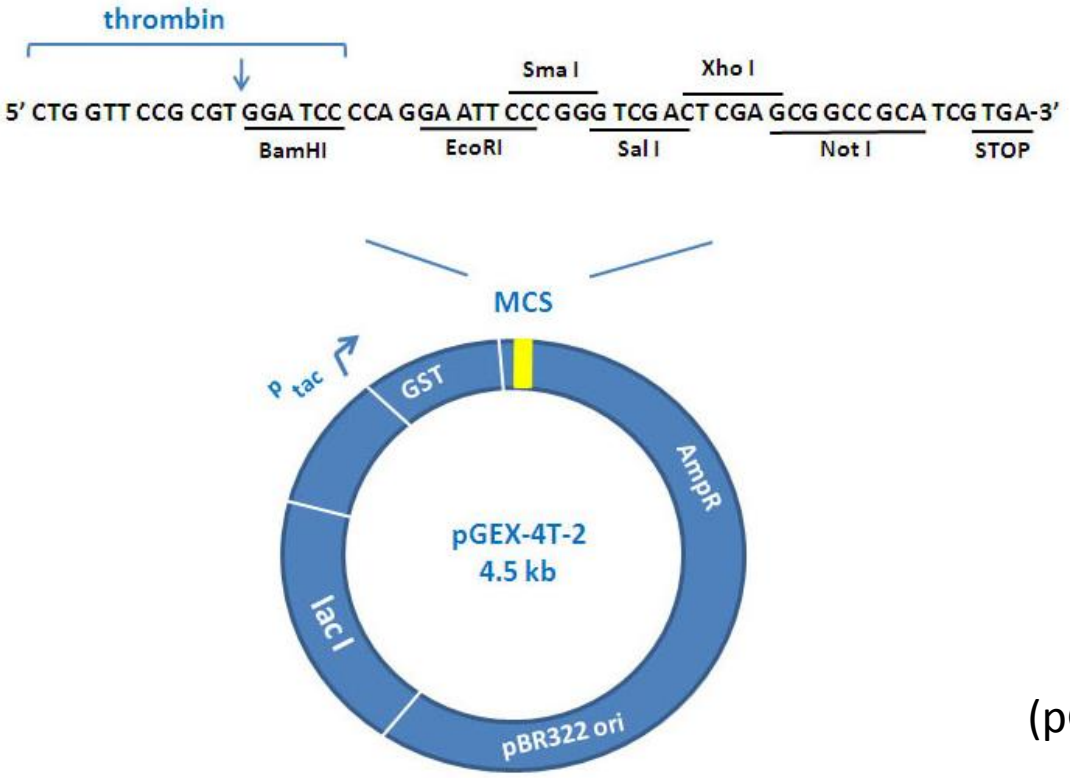
Figure 4.1. Cloning into *in vitro* expression vector and following transcription and translation.
(a) the MCS of the phagemid pBlueScript SK+ is shown (see Fig.2.18 for a plasmid map). Depending on the choice of promoter, either sense or antisense transcript may be transcribed; an example of transcribing sense mRNA from T7 promoter is shown

(b) for anti-sense RNA transcription, the reverse T3 promoter should be used.

(c) The recombinant vector is put in a reaction tube along with the appropriate phage RNA polymerase, the appropriate buffer, and the NTPs. In the second step of *in vitro* translation, the mRNA product from the first reaction and cell lysates (most commonly HeLa cell extract or rabbit reticulocyte extract based systems) will be used along with amino acids. One of these amino acids is most often radioactively labeled to monitor newly-synthesized proteins

Bacterial expression of proteins

The inducible nature of some operons in bacteria has been exploited in bacterial protein expression: some of the bacterial expression plasmids use, for example, the *lac* promoter of the *lac* operon, or modifications of it, to control expression from the cloned gene sequence. For induction, instead of lactose, a lactose-analog IPTG (isopropyl-b-D-thiogalactoside) is usually used. (However, remember that *lac* induction relies on the presence of a regulatable repressor in the host bacteria: if the bacteria that you have chosen to express the protein in has a mutation in its *lacI* gene, which codes for the *lac* repressor, then you will not get IPTG-inducibility either!)



(pGEX vector map already given in Fig.2.21)

Expression in yeast

Yeast is a very popular eukaryotic model organism; it is unicellular and thus both easier to manipulate and easier to propagate; it has a cell cycle profile that is very similar to higher eukaryotes and thus is used as a model for cell cycle or cellular aging studies; its genome is deciphered (the entire 12 Mb coding for around 6000 proteins) and its metabolism is mostly known, just to name a few of the advantages.

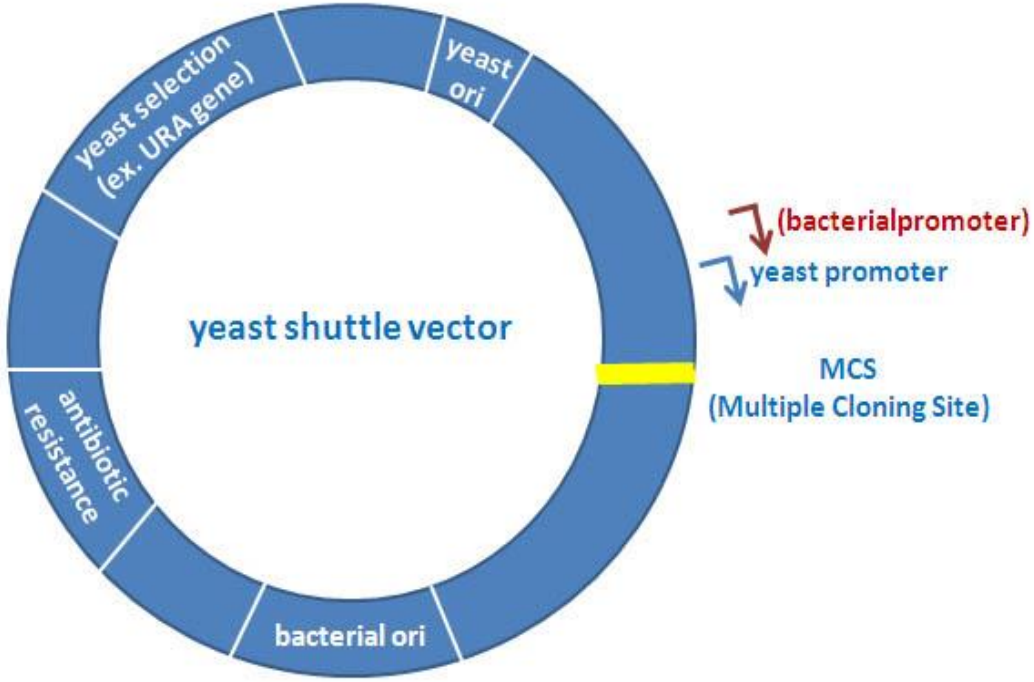


Figure 4.2. An example for a yeast shuttle vector, which could also be transformed into bacteria (for expression, if there is a bacterial promoter, shown here in red). The MCS is used for cloning; yeast promoter is used for expression in yeast. Yeast origin of replication (such as the 2m ori in 2m plasmids) is used to stably maintain the vector without integration into chromosomes) and yeast selection marker (such as the gene of an enzyme necessary for the synthesis of an amino acid, see text for details) are necessary for propagation and selection in yeast. On the other hand antibiotic selection (such as ampicillin resistance) or bacterial origin of replication are used for propagation and selection in bacteria.

Selection in yeast

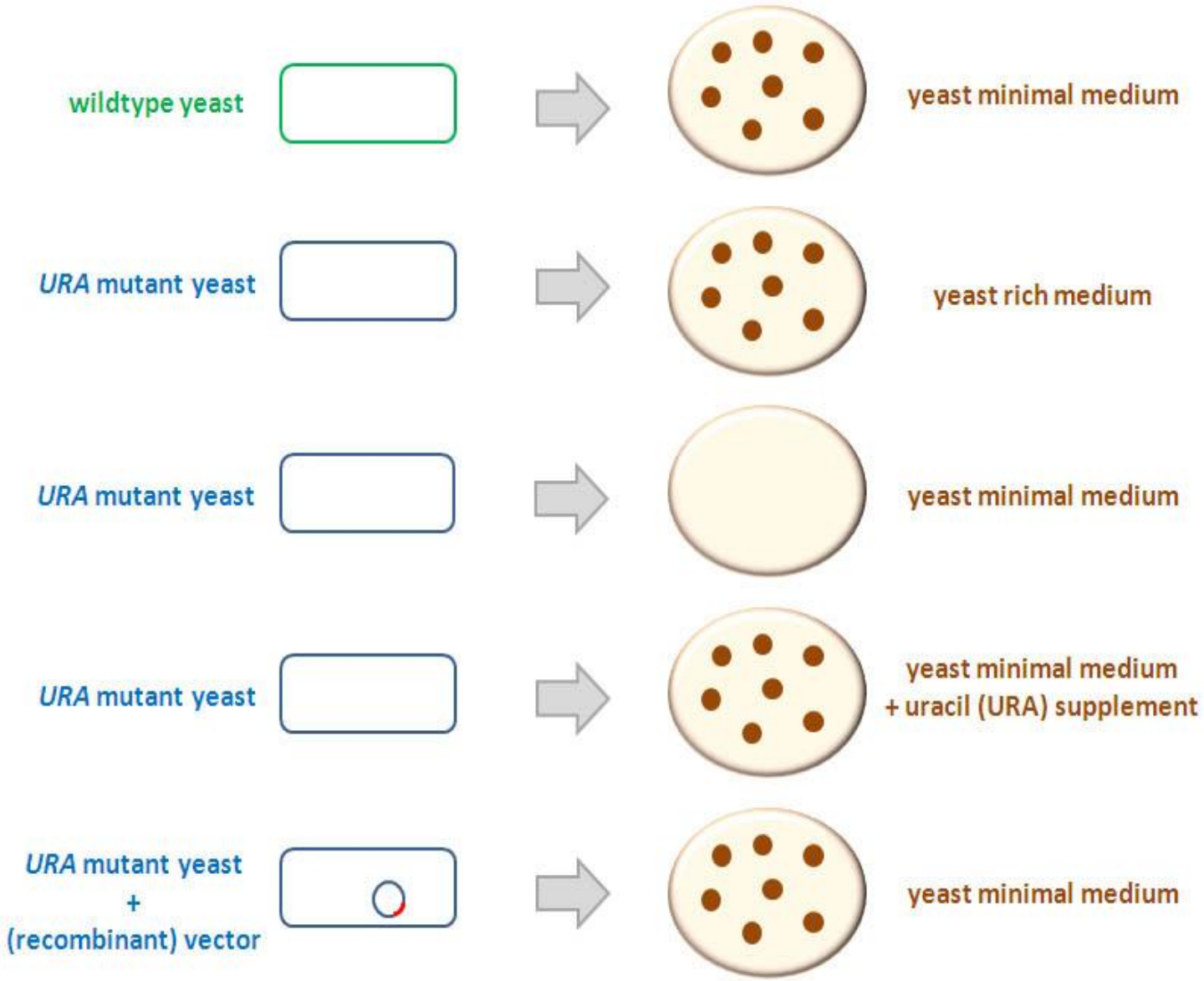


Figure 4.3. The basic principle behind selection in yeast. Wildtype yeast have all the metabolic enzymes necessary for synthesizing their entire set of amino acids, starting from basic precursors present in minimal medium (top panel). Yeast mutant for *URA*, an enzyme necessary for the synthesis of uracil, can grow on rich medium, which contains all the nutrients, not just the basic essentials (second panel from top), but not on minimal medium that lacks uracil (middle panel). The same mutant can, however, be grown on minimal medium that is supplemented with uracil (second panel from bottom). The only other way this mutant can grow on minimal medium is if it has a vector that carries the synthetic enzyme for uracil, ie wildtype *URA* (bottom panel).

Expression in insect cells

Insect cells are frequently used to express high amounts of proteins from baculoviral expression vectors. Insect cells, similar to yeast and other eukaryotics, are preferred for expression of proteins towards the analysis of post-translational modifications or the intracellular localizations, which cannot be addressed in a simple prokaryote lacking such modifications or membrane-bound intracellular compartments.

The most common insect expression system involves the Baculovirus-based expression systems, available from a number of different companies. In this system, genes are usually cloned after a strong polyhedrin promoter of *Autographa californica* (the moth alfalfa looper) nuclear polyhedrosis virus (AcNPV), thus high levels of expression takes place during late stages of infection.

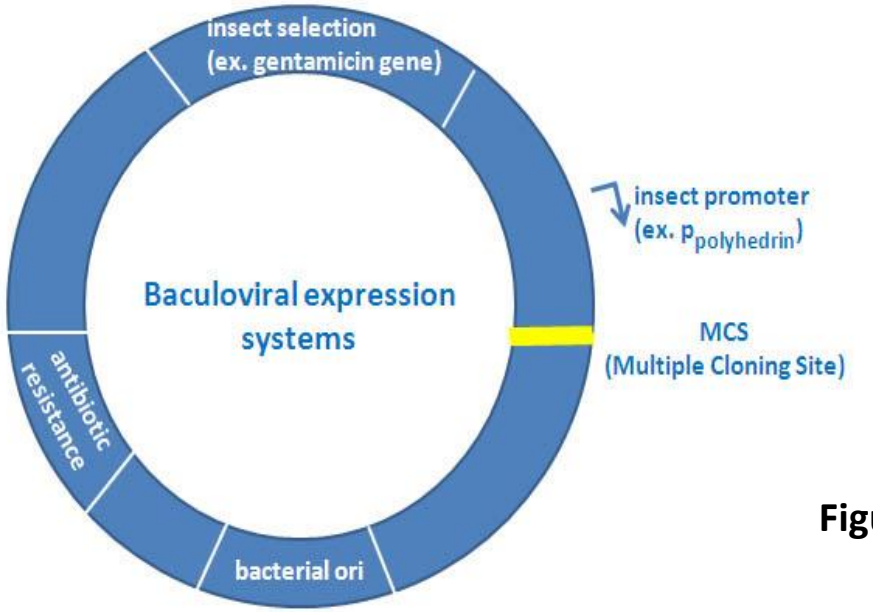


Figure 4.4. A generic baculoviral expression vector

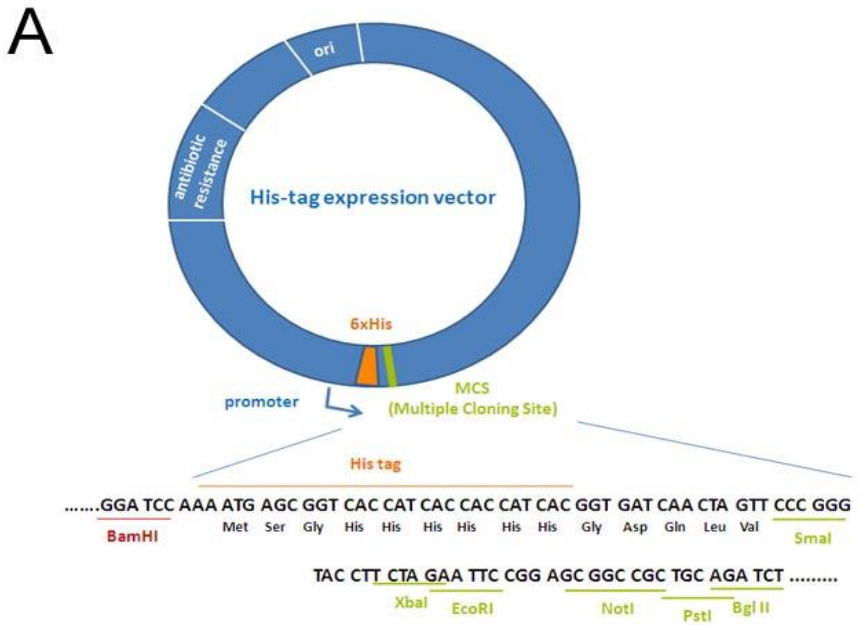
Purification of proteins

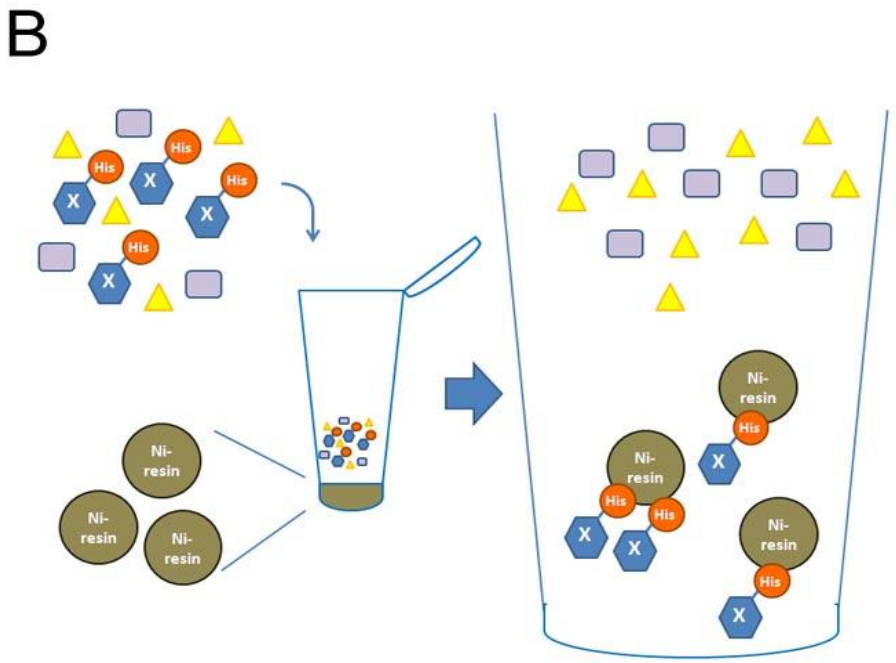
Purification of proteins

Biochemical and analytical-grade purification through various chromatography methods are largely covered in a variety of biochemistry textbooks.

Being a genetic engineering textbook, here we only concentrate on crude purifications (or “enrichments”), based on affinity purification methods.... such as immunoprecipitation or GST pulldown assays.

Affinity purification by nickel columns





Affinity purification relies on the presence of **tags**, or short peptide labels, on the protein analyzed.

A typical polyhistidine tag consists of at least 5 or 6 Histidines at either the N- or C-terminus of a protein. His-tagged protein thus can be purified using the affinity of Histidines to a metal such as Nickel, Cobalt or similar resins.

Histidines present in the tag will show affinity for the Nickel in the resin or agarose “beads” and thus will be retained in the “precipitate”, whereas all other proteins will be in the supernatant and can thus be discarded.

Figure 4.5. A schematic of His-tag fusion protein purification strategy. (a) A generic His-tag expression vector; (b) Purification of His-tagged proteins using Nickel affinity beads.

After purification, the column or beads could either be directly used for further identification of interaction partners in co-immuneprecipitations, or the protein of interest may be eluted and separated (cleaved) from its tag to obtain pure protein. If this last option is desired, then a cleavage site for a proteolytic enzyme such as thrombin should be present in the vector used for cloning, between the tag and the multiple cloning site where the protein is to be cloned .

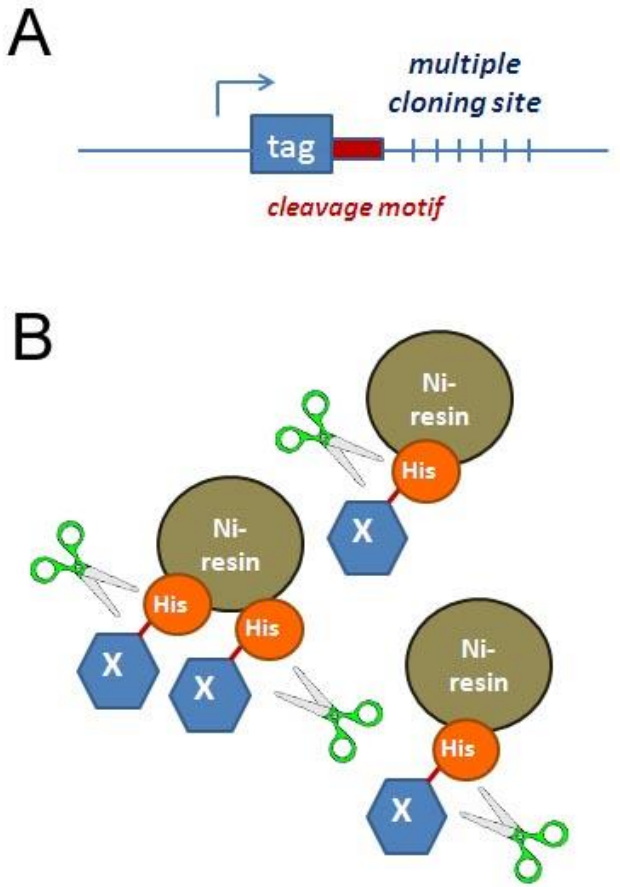


Figure 4.6. A schematic of enzymatic cleavage of partner protein after tag purification. (a) A general scheme for incorporation of protease cleavage motif between a tag and the multiple cloning site; (b) Removal of the protein X from its Histidine tag that is affinity-bound to a nickel resin

Affinity Purification using Monoclonal and Polyclonal antibodies

In many of the cases, we will use indirect detection methods for analysis or purification of proteins, and these methods will in some cases rely on the affinity of antibody to their respective antigens.

Antibodies are the body's defense systems, produced by the immune cells. Particularly, B lymphocytes of the adaptive immunity produce antibodies specific for the antigens. **Polyclonal antibody** refers to a mixture of antibodies (more commonly, immunoglobulin IgG) from different B cells circulating in the blood of the immunized animal; these antibodies are specific for a different **epitope** of the antigen, in our case this antigen is mostly the protein that we want to study . **Monoclonal antibodies**, however, are derived from a single B cell clone against a single epitope.

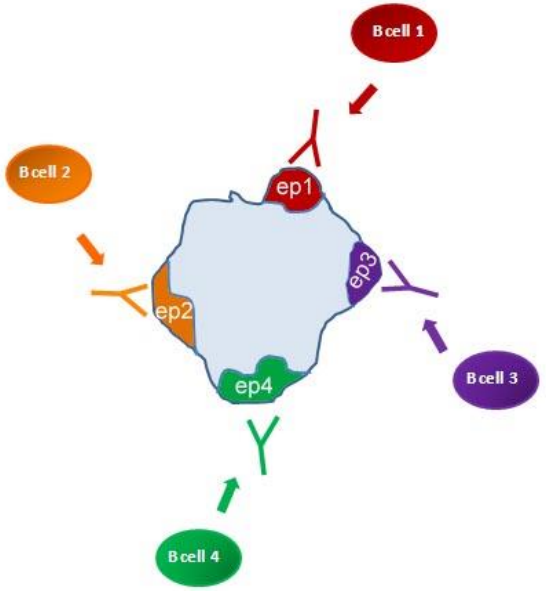
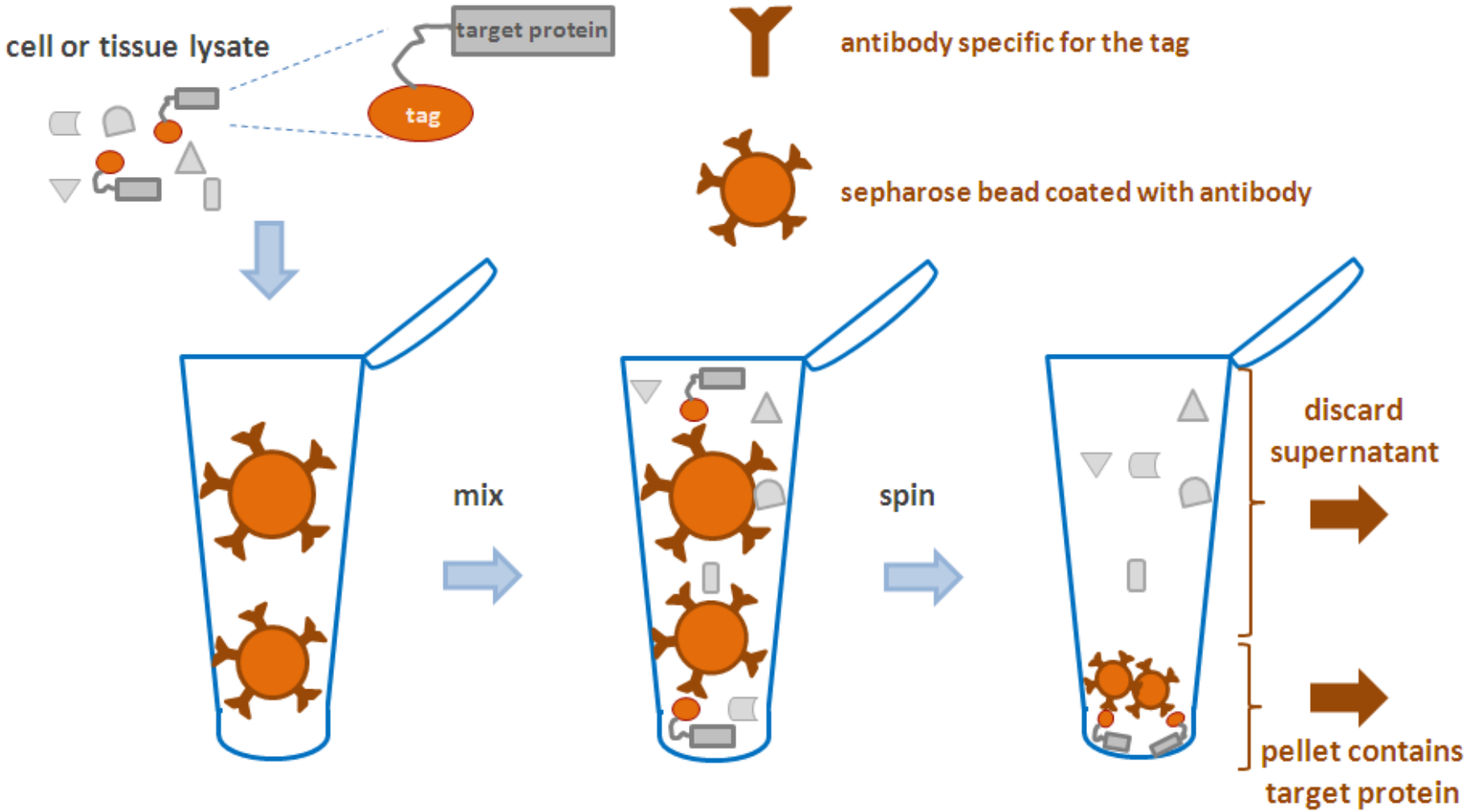


Figure B1.1. A schematic of four hypothetical epitopes of a protein antigen, and four different antibodies produced by different B cell clones against each epitope

Similar to Nickel columns or beads, antibody affinity purification relies on sepharose or agarose beads coated with antibody specific for the tag (such as Flag, HA or Myc tag), to which the target protein is fused.

The cell or tissue lysate containing the target protein-tag fusion is then mixed with the beads, and after a brief spin the beads that retain the tag as well as the fusion partner settles (or precipitates) at the bottom. The rest of the proteins are mostly in the supernatant and can be discarded.



Monitoring expression in cells

In order to monitor the expression of proteins, once again the specificity of antigen-antibody interaction is exploited.

In Western blots, only a semi-quantitation is possible, and a housekeeping gene such as actin or β-tubulin is used for normalization of samples; the antibody reactivity for our protein of interest is then interpreted relative to this housekeeping gene.

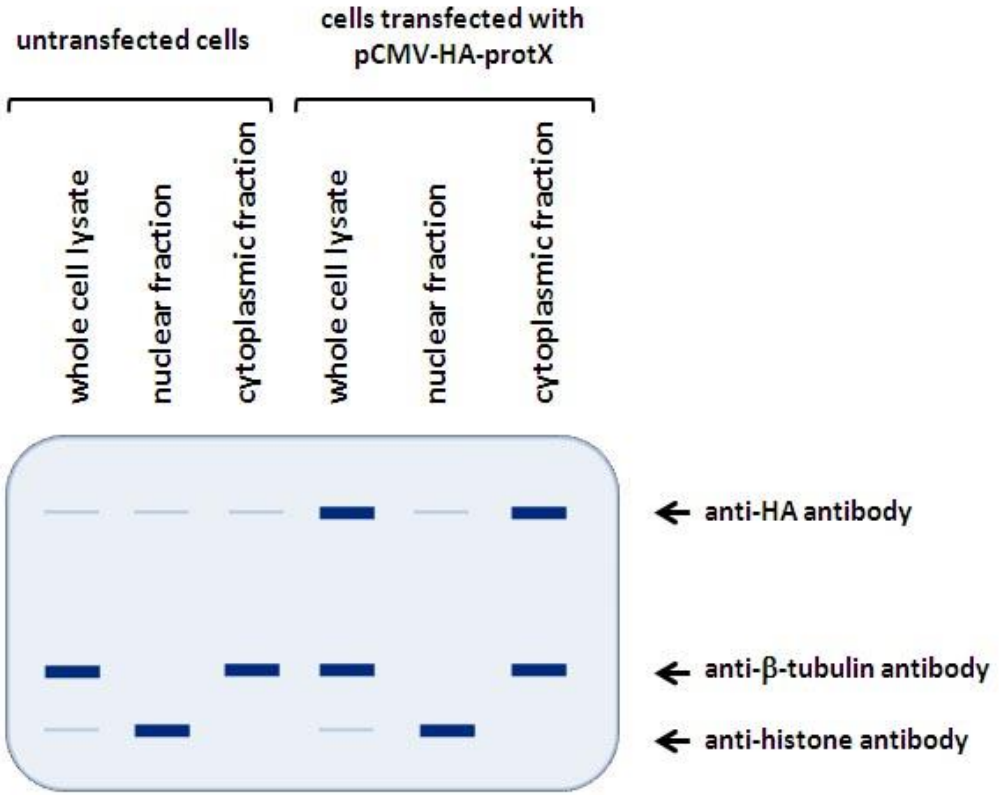
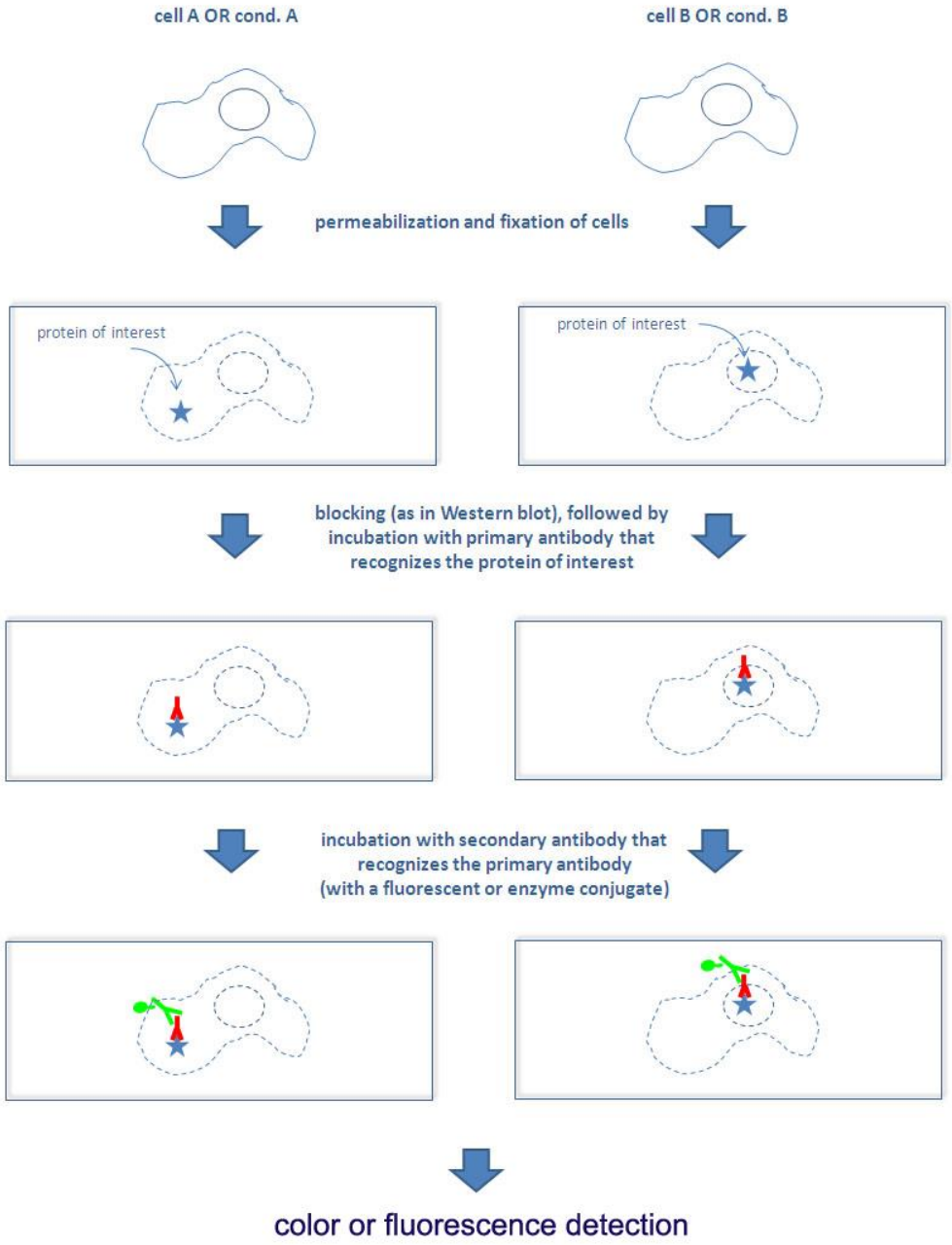


Figure 4.7. A hypothetical Western blot for monitoring expression of protein X fused to HA tag. The expression is studied in whole cell lysates, as well as in nuclear and cytoplasmic fractions separately. For whole cell and cytoplasmic fraction, b-tubulin antibody is used for normalization; for nuclear fractions histone antibody is used. Relative expression of protein-X-HA fusion is then monitored by anti-HA primary antibody



APPENDIX II – PROTEIN TECNHIQUES

iii) Immunofluorescence and Immunohistochemistry

Fig. A2.4. Brief overview of the steps of immunofluorescence & immunohistochemistry. Two samples are shown here; either a protein's amount and localization can be studied in cell A vs cell B, or the level of expression and localization of a protein can be studied under two (or more) different conditions (cond. A and cond. B).

In essence, the basic logic is the same as in Western blots – we are trying to detect expression of a protein. However, since cells are used, information about the subcellular localization of proteins can also be obtained in addition to level of expression .

The basis of both immunofluorescence and immunohistochemistry is in principle the same, with primary antibodies used as probes to recognize and identify target proteins, and secondary antibodies (in indirect fluorescence) against the primary antibody to amplify the signal

Creating fusion proteins: Green Fluorescent Protein

Immunofluorescence is a useful method for monitoring expression of proteins in cells, as well as their subcellular localizations, however as with immunohistochemistry, immunofluorescence also relies on the use of antibodies, and as such, fixation and permeabilization of cells so that antibodies can penetrate through the lipid bilayer. However, the cells are no longer live after this treatment, and therefore the assay can only show the presence and/or location of the protein in a snapshot in time. Given that proteins exhibit dynamism within cells, this snapshot would not present the big picture.

The discovery of fluorescent proteins has, therefore, been invaluable for the study of proteins in live cells, and for that reason has been awarded with the Nobel Prize in Chemistry in 2008 to its discoverers, Osamu Shimomura, Martin Chalfie and Roger Y. Tsien, *"for the discovery and development of the green fluorescent protein, GFP"*

(http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/press.html).

GFP, first isolated from the jellyfish *Aequorea victoria,* has shortly become a popular tool in molecular biology (Tsien, 1998). Later, red fluorescent proteins have also been isolated from other species, as well as new enhanced versions of GFP such as EGFP (enhanced GFP; Fig.4.8), BFP (blue fluorescent protein) and YFP (yellow fluorescent protein). Since these proteins all have a fluorophore domain that emits fluorescence upon excitation at a certain wavelength, fixation of cells is not necessary and hence live assays can be conducted.

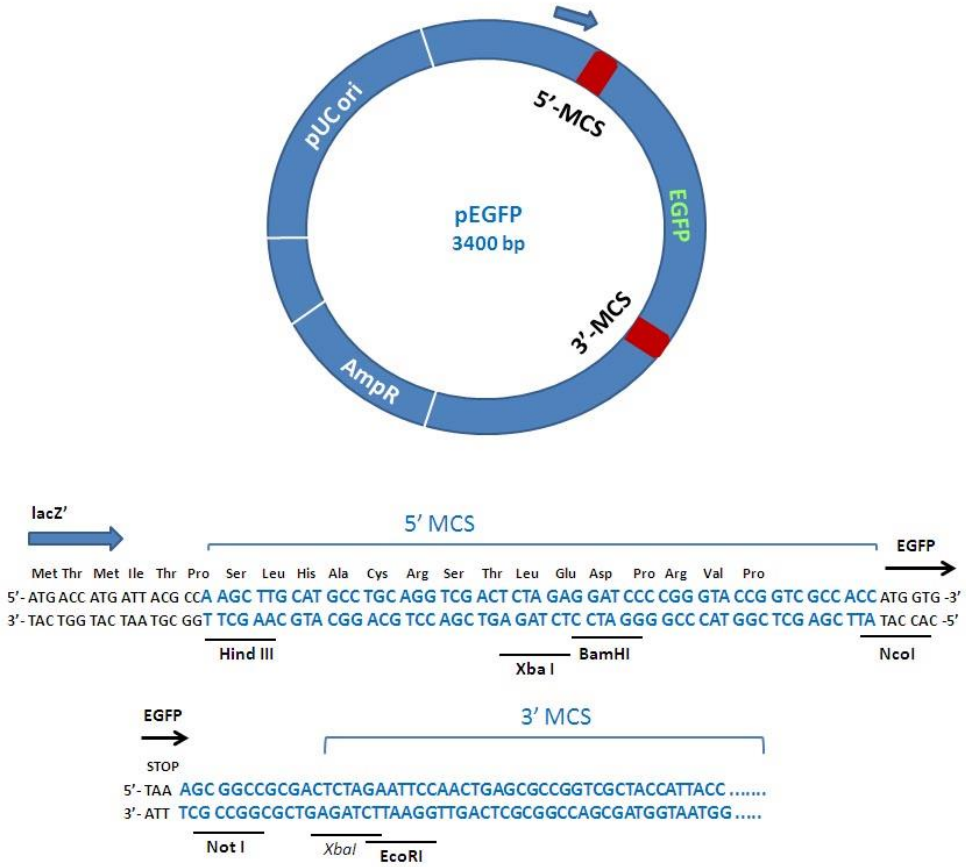


Figure 4.8. A generic map of pEGFP vector, including the 5'- and 3'- MCS sequences. Different pEGFP variants, such as pEGFP-N1 or pEGFP-C3 etc, exist that contain either 5'- or 3'- MCS.

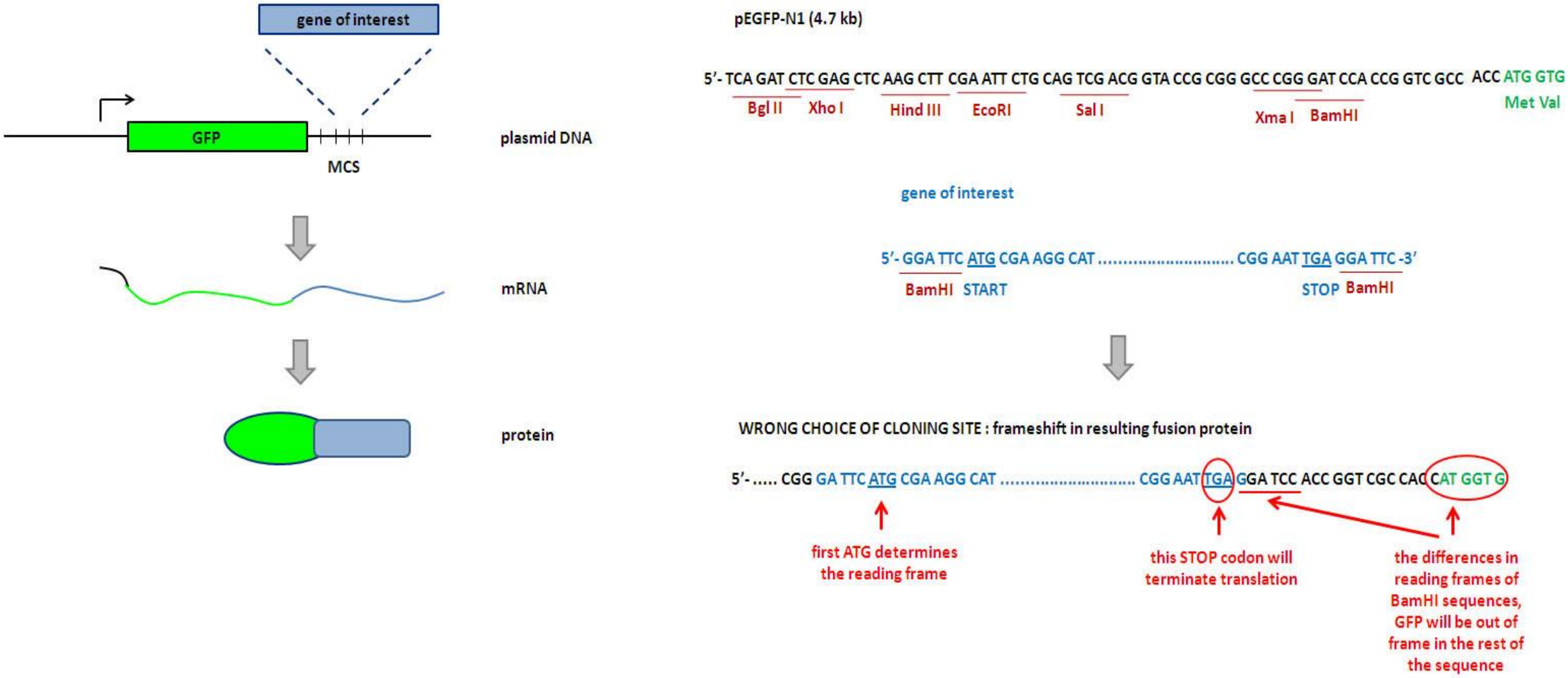


Figure 4.9. A schematic diagram showing that care must be given to START and STOP codons, as well as to reading frames, while cloning fusion proteins. Left panel shows a plasmid where upon transcription and translation, GFP is at the N terminus and the protein of interest is at the C terminus of the translation product. Right panel shows a section of the pEGFP-N1 plasmid, wich GFP at the C-terminus, and a hypothetical gene of interest (in blue). A wrong choice of restriction site for cloning (BamHI in this case) results in a STOP codon in the middle of the fusion protein, and frameshift in the GFP coding sequence

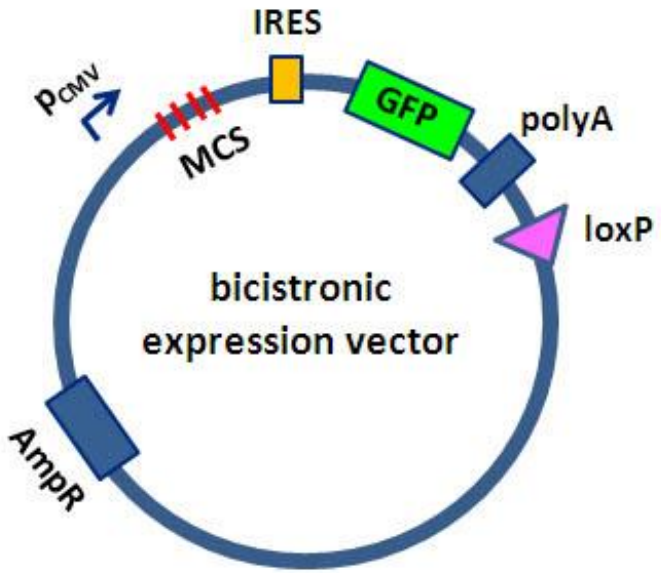


Figure 4.10. A schematic diagram of a generic bicistronic expression vector, where the gene cloned into the multiple cloning site (MCS) will be transcribed into a single mRNA along with GFP under the CMV promoter. This transcript will be used to translate two different proteins, the protein of interest, and GFP. LoxP sites are recombination sites and will be discussed in Chapter 7

Post-translational modifications of proteins

In eukaryotes, the translation product (ie polypeptide) is not necessarily functional upon synthesis and folding: a number of post-translational modifications such as glycosylation may be necessary to render the protein functional or deliver it to the correct subcellular destination via the Golgi network.

Alternatively, a protein's activity may be modified by a number of modifications such as phosphorylation of an enzyme, SUMOylation of a transcription factor, acetylation or methylation of histone proteins. Or else, the protein may be misfolded or no longer needed, and thus it gets labeled for proteolytic degradation by ubiquitination (in fact, there are many different formats of ubiquitination, each with a different "meaning"). Or a protein may need to be cleaved for activation (such as cleavage of a pro-caspase into an active caspase), or disulfide bridges may be formed for functional secreted proteins such as insulin.

Such post-translational modifications are in fact chemical modifications of a polypeptide after its synthesis and folding, and are highly regulated events. 2D gel electrophoresis and other proteomic approaches, however, would also provide information about not only the level of protein expression, but also of any post-translational modifications. Modification-specific antibodies may also be used to study post-translational modifications.

“The vitality of thought is in adventure. Ideas won’t keep. Something must be done about them.”

Alfred North Whitehead (1861 - 1947)