TECHNIQUES IN GENETIC ENGINEERING

Işıl KURNAZ

CHAPTER 5

Mutagenesis

“No amount of experimentation can ever prove me right. A single experiment can prove me wrong.”

Albert Einstein (1879 – 1955)

The simplest way to understand the function of a gene is to observe what happens in its absence.

Geneticists rely on patients who display certain symptoms and inquire the loss or deficiency in which gene(s) may be responsible for these symptoms...

Molecular biologists may instead choose a more direct line and mutate or delete all or part(s) of a gene in a model system (could be bacteria, cells, animals or plants) and observe what happens to the physiology or behavior of the organism.

Deletion –based mutagenesis

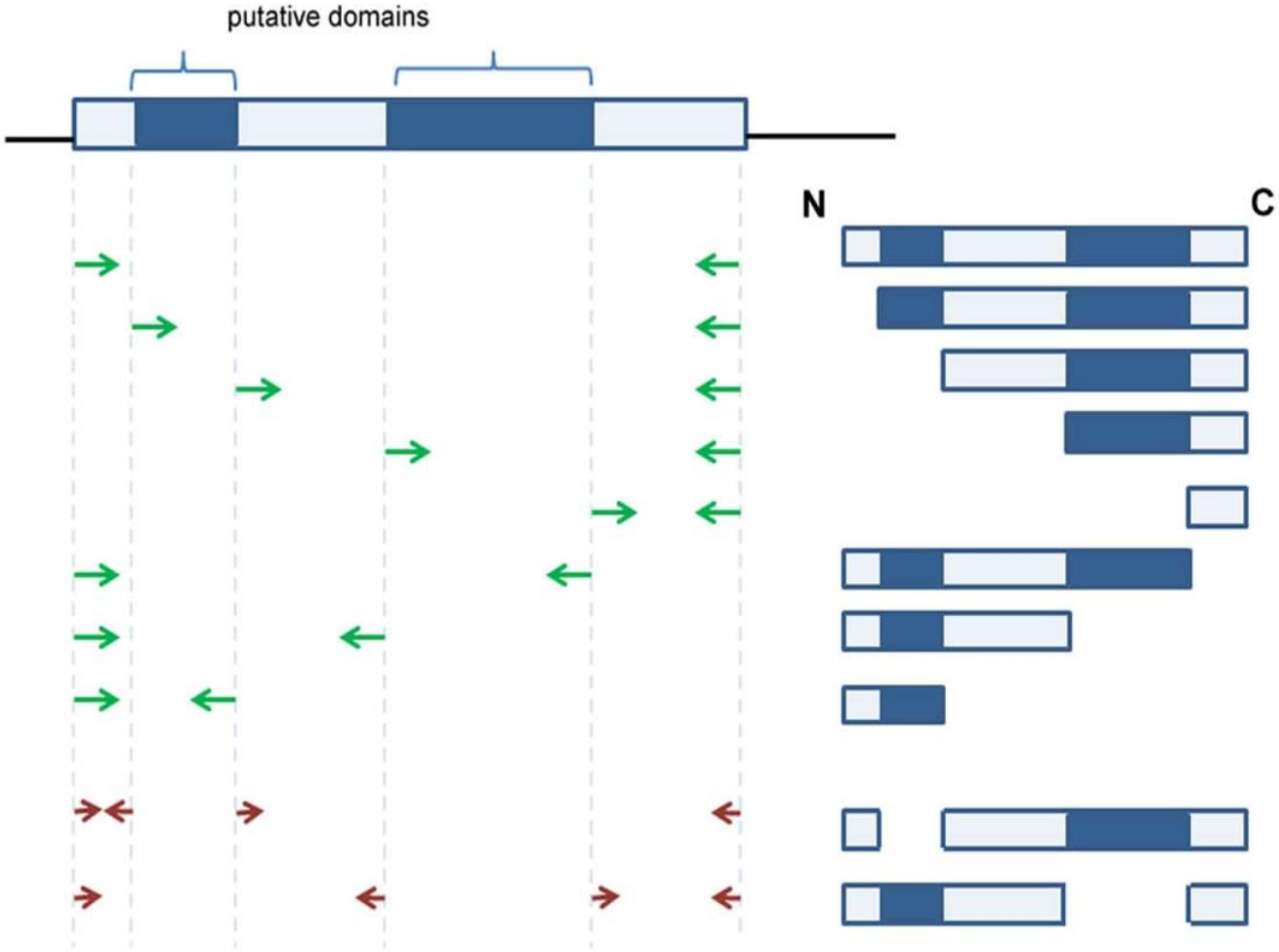
Deletion studies, as the name implies, relies on removal of large “chunks” of the gene sequence through genetic manipulations.

The simplest method of creating such deletions (which can later be used for cloning purposes) is by PCR using primers that define the boundaries of the deletion.

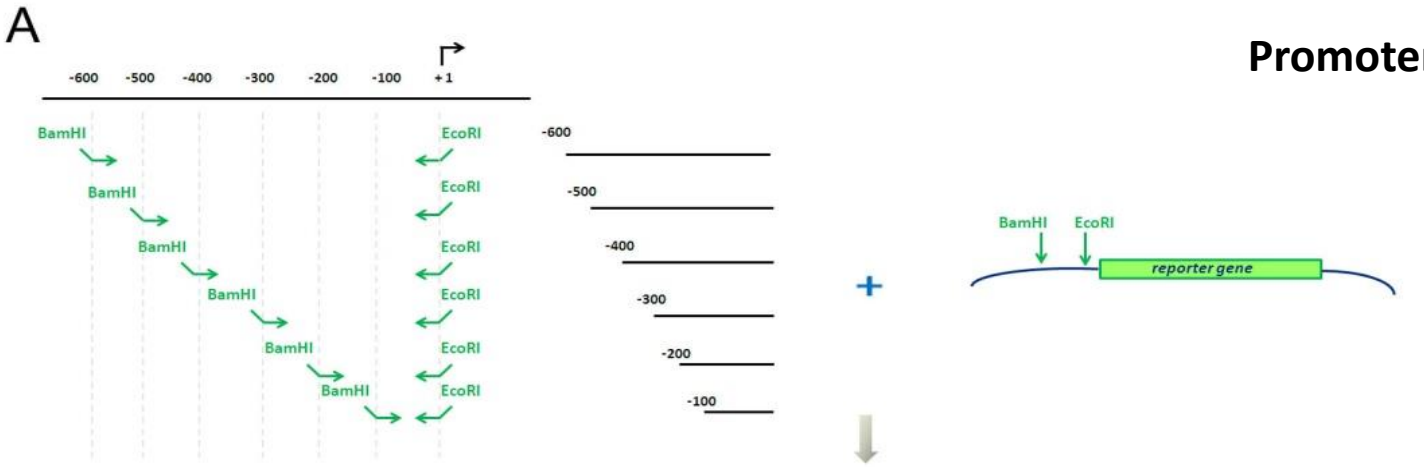
For protein coding sequences, this technique is mostly employed so as to define the function of different protein domains, such as Nuclear Localization Sequences (NLS), DNA binding domains (DBD), activation domains (AD), and so on.

By deleting different regions of the protein and observing the changes in its function, one can then monitor which region is responsible for that particular function.

Figure 5.1. A schematic example of a typical deletion analysis of protein-coding sequences for the analysis of putative domains (active site, binding motif, interaction region, activation domain etc). If very little is known about the predicted domains, then a large series of N- and C-terminal deletions can be used to study the function of these domains (green arrows show the forward and reverse primers that could be used for PCR amplification of these deletions). Additionally or alternatively, depending on what is known and on the design of the experiment, one could also delete an internal segment (such as the putative domain) from the sequence. In that case, two sets of primers may be used, but with well-calculated design of primers before and after the region (considering the region spans several amino acids when translated; red arrows indicate forward and reverse primer sets; note that the primers in the middle must include restriction sites that overlap and that are in-frame with the coding sequence).



Promoter “bashing” experiments and reporter assays



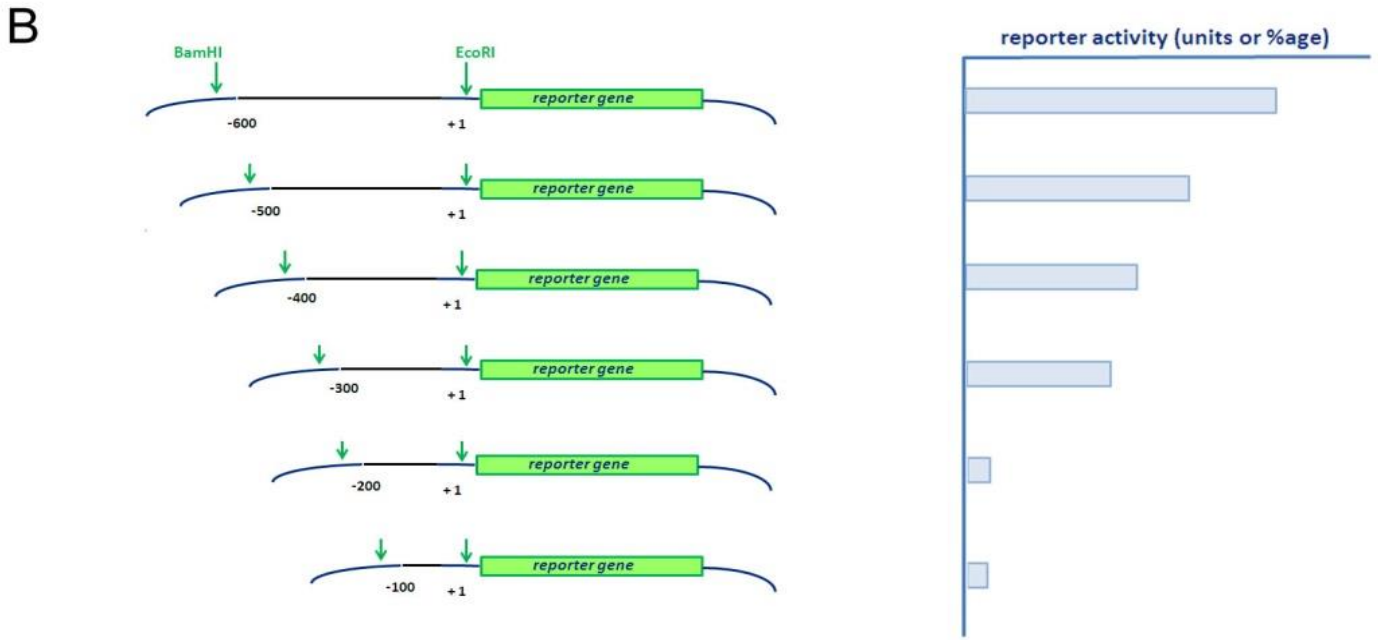


Figure B2.1. A schematic example of the cloning strategy for promoter deletion analysis and reporter assay. **(a)** Deletions of the 5’ end of the promoter can be constructed using different 5’ primers, and cloned into the reporter plasmid as described in Chapter 2; **(b)** the reporter constructs thus obtained can then be used to analyze the activity of these deletions through analysis of reporter gene activity

DNA – protein interaction assays: *Electrophoretic Mobility Shift Assay (EMSA)*

Also known as *Gel Mobility Assay, Gel Shift Assay, Bandshift Assay* or *Gel Retardation Assay*, among others, the basic principle of this assay is that non-denaturing gels are used, thus protein-DNA complexes are not separated or denatured and therefore both size and structure can affect migration through the gel.

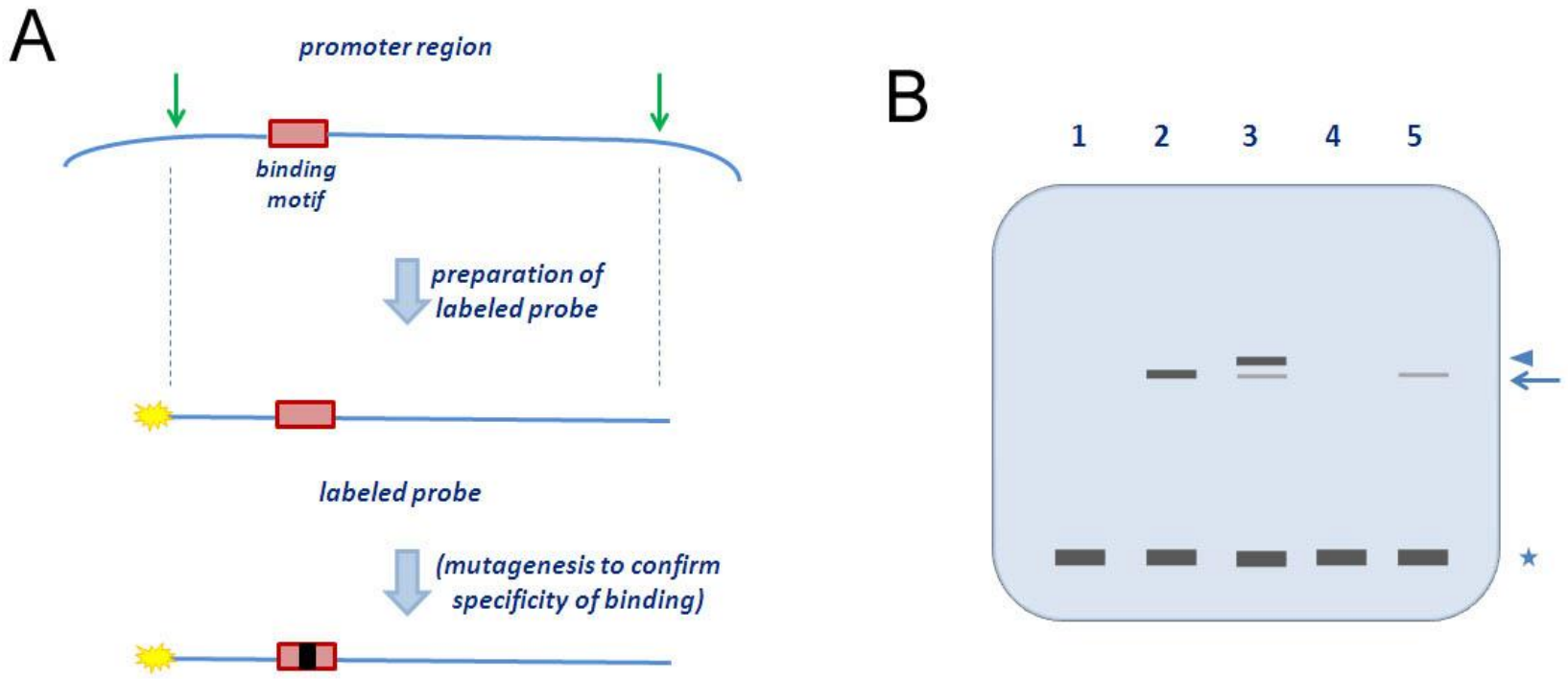


Figure B2.2. A schematic explanation of electrophoretic mobility shift assay (EMSA). **(a)** the DNA fragment of interest (usually a part of the promoter containing the binding motif) is used to prepare a labeled (radioactive or non-radioactive) probe (this DNA fragment could also contain a mutation introduced to disrupt the binding motif, to which the transcription factor in question is no longer expected to bind); **(b)** a cartoon of a hypothetical bandshift assay, where the labeled probe or its mutated counterpart schematized in (a) is incubated together with either purified transcription factor or a nuclear extract in the presence or absence of an antibody and then subjected to nondenaturing gel electrophoresis; 1, labeled probe alone; 2, labeled probe incubated with either purified protein or a nuclear extract; 3, labeled probe incubated with purified protein / nuclear extract in the presence of a specific antibody against the transcription factor of interest; 4; labeled mutagenized probe alone; 5, labeled mutagenized probe with the protein or nuclear extract; the *star* shows the probes without any protein, running fast in the gel; the *arrow* shows the retarded or shifted probe DNA/protein complex; the *arrowhead* indicates the probe DNA/protein/antibody complex, further retarded in the electrophoretic field. (The radioactive EMSA could further be used to calculate binding constants, if the bands on the gel are quantified in experiments designed for binding affinity calculations)

DNA – protein interaction assays: *Chromatin immunoprecipitation (ChIP)*

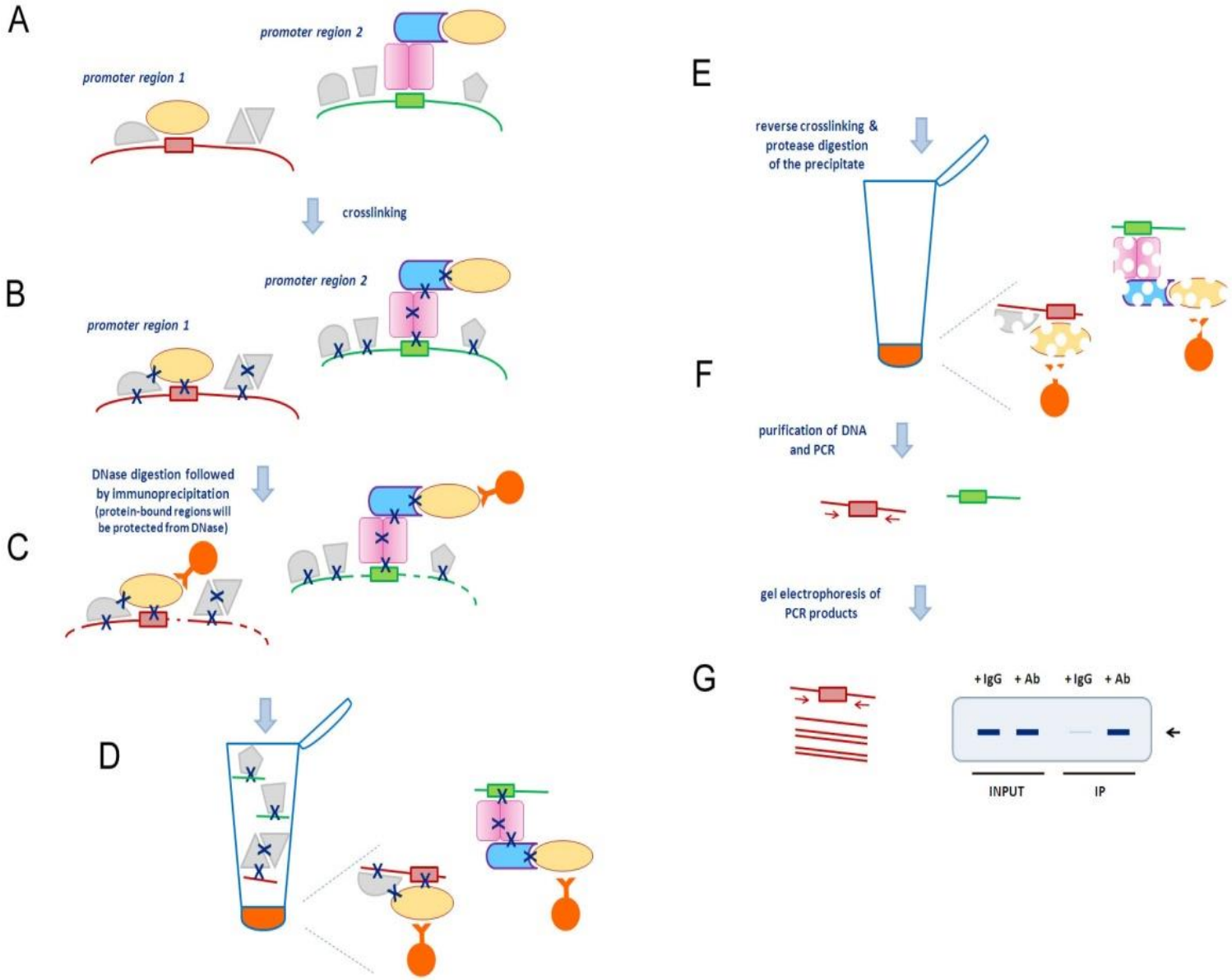


Figure B2.3. A schematic explanation of chromatin immunoprecipitation.

The question in this case is whether a protein is actually bound to a putative binding site *in vivo*.

To that end, cells are treated with a crosslinking reagent. This mix is then digested with micrococcal DNase (or else sonicated) to shear away any unprotected DNA.

You would then immunoprecipitate the protein of interest with a specific antibody, along with any interaction partner, be it protein or DNA. After crosslink is reversed and any protein is digested using proteases; the DNA that has been precipitated along with protein-antibody complexes are purified, and subjected to PCR amplification using primers specific to the region of interest.

After gel electrophoresis of both the “input” (ie, sample before immunoprecipitation, in other words the cell lysate) and the “IP” amplification products are analyzed.

Positive signal with the antibody-IP (along with a negative signal in IgG-IP) confirms specific binding of the protein in question to the target DNA region

Site-directed mutagenesis

Site-directed mutagenesis, as the name implies, refers to the generation of mutations at specific regions of the gene sequence, with a specific purpose (such as changing a phosphorylation site, catalytic activity, ligand binding property etc).

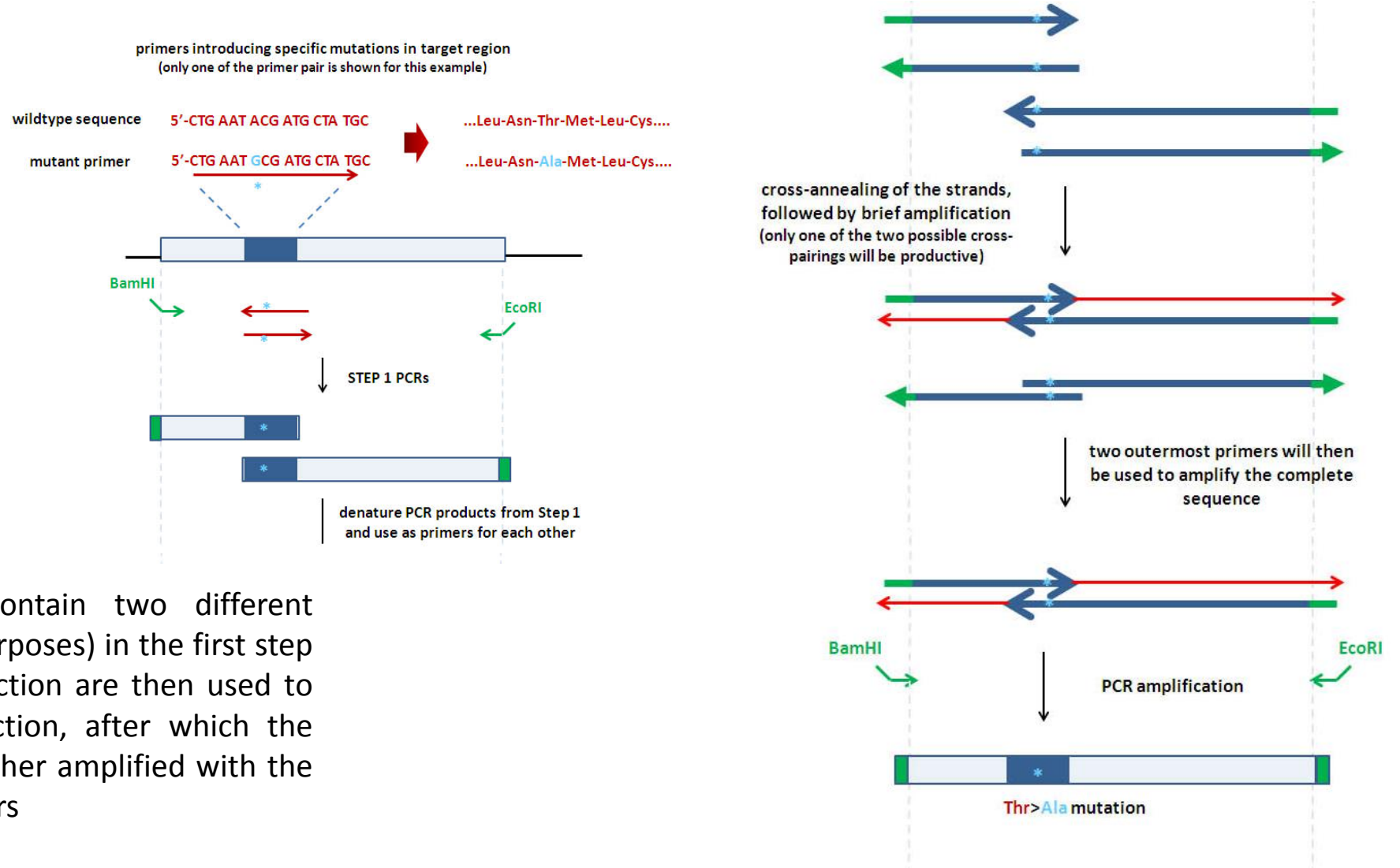


Figure 5.3. A schematic diagram of PCR-based site-directed mutagenesis strategy. In this particular example, a potential phosphorylation motif, Threonine, is to be mutated into an Alanine (T>A mutation). A mutant primer pair is designed to contain a mismatched nucleotide that creates this conversion, and teamed up with forward and reverse primers (shown in this example to contain two different restriction enzymes for cloning purposes) in the first step PCR. The products from this reaction are then used to cross-anneal in a brief PCR reaction, after which the complete mutant sequence is further amplified with the help of forward and reverse primers

Random mutagenesis

Mutations have originally been introduced to genes (or rather, organisms) by either irradiation or chemical mutagenesis (for example, in Morgan's *Drosophila* experiments).

While such chemical or physical mutagenesis methods are quite robust in introducing mutations to genes, they are highly hazardous in nature and thus not thoroughly desirable.

A number of other, non-hazardous, methods can be used to generate random mutations in genes, including error-prone PCR, degenerate oligonucleotide primers, or mutant bacterial strains (such as the *E. coli* strain XL1red that is defective in its DNA repair proteins so that mutation rate is around 5000-fold higher than wildtype strain) among many others.

Error-prone PCR simply relies on the fact that *Taq* DNA polymerase or similar polymerases can incorporate wrong nucleotides in the presence of high levels of Mg^{2+} along with Mn^{2+}.

Degenerate primers, on the other hand, would generate site-directed random mutants

A

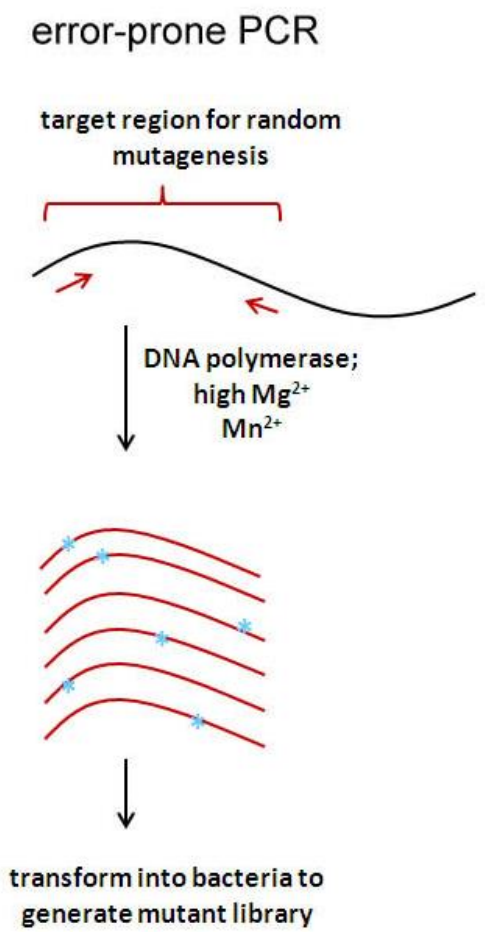


Figure 5.4. **(a)** error-prone PCR simply relies on the loss of specificity of certain DNA polymerases in the presence of high amounts of magnesium and manganese ions.

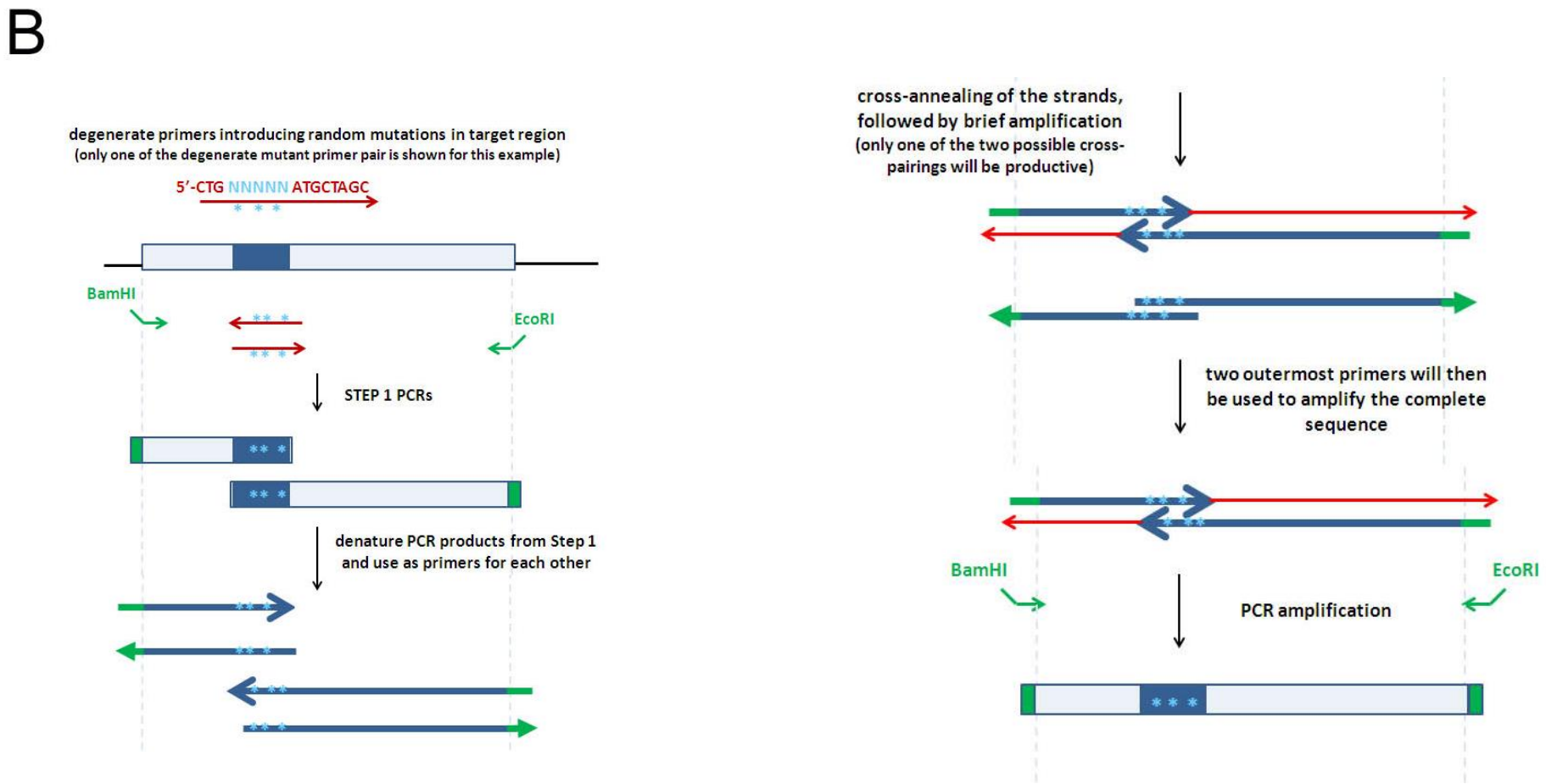


Figure 5.4. **(b)** site-directed random mutagenesis is essentially based on the same principle as site-directed mutagenesis, with the single difference of using degenerate primers across the target region.

This strategy is essentially based on the same principle as site-directed mutagenesis, with the simple difference that rather than mutating a specific nucleotide (or a set of nucleotides) into other known nucleotides (such as A>G in Fig. 5.3, introducing a specific Thr>Ala mutation), a degenerate sequence is created in certain residues (N, for any nucleotide, in Fig.5.4b) to introduce relatively random mutations at specific residues.

Such random mutagenesis products can be used to generate a mutant library so as to study structure-function relations of proteins or enzymes, improve catalytic properties of enzymes, or design proteins with desired properties.

However, since a rather large collection of mutants are generated, high-throughput screening methods need to be designed for selection of these desired properties

Directed evolution, protein engineering and enzyme engineering

Directed evolution experiment is briefly intended to artificially “evolve” a protein or an enzyme, ie “optimize” it for the desired properties (or “direct” it towards a certain goal).

There are different means to achieve this end, albeit the most common ones include random mutagenesis, and mostly error-prone PCR, or the use of a highly mutagenic strain.

The gene of interest is thus mutated randomly, creating a large mutant library from which to select the desired trait under a selection pressure, such as heavy metal tolerance, salt tolerance, higher affinity to substrate etc. Once the positives are selected, the mutants are then amplified and their DNA isolated so that the mutation resulting in the desired trait can be sequenced and identified.

This represents the first “round” or “generation” of directed evolution: in many cases, multiple generations have to be obtained for the “survival of the fittest”.

These molecules represent the engineered protein or the enzyme.