RECOMBINANT DNA TECHNOLOGY

II M.Sc., MICROBIOLOGY
SEMESTER III

UNIT IV (RNA AND PROTEIN LEVEL EXPRESSION PROFILING)

SURESH S S RAJA
ASSOCIATE PROFESSOR
DEPARTMENT OF MICROBIOLOGY
BHARATHIDASAN UIVERSITY
TRICHY - 24

RNA LEVEL EXPRESSION PROFILING

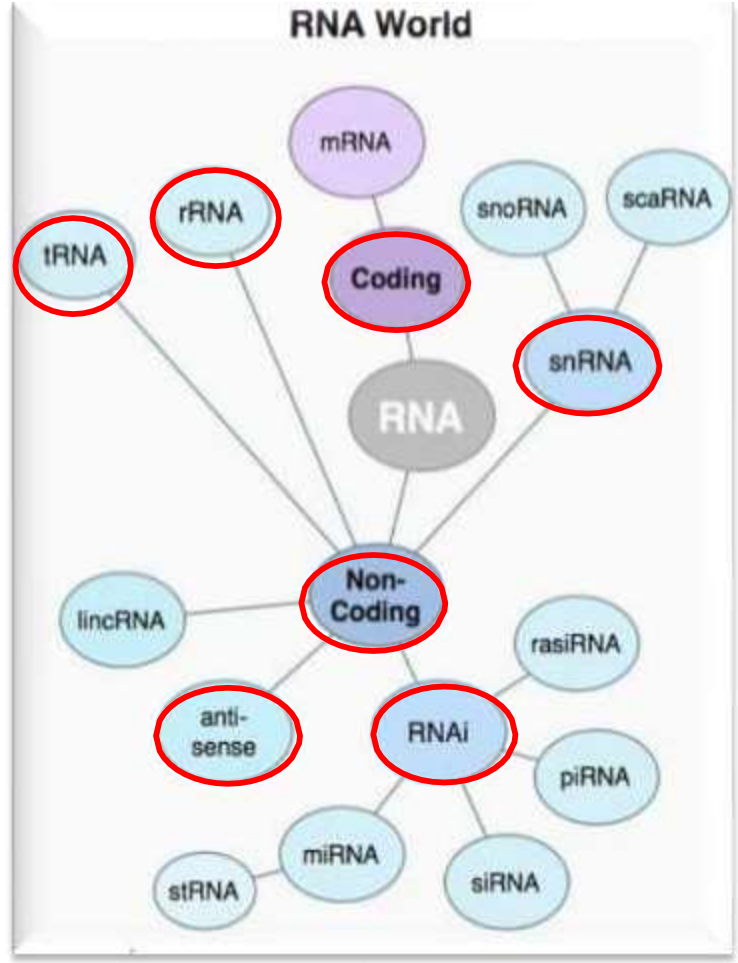
CENTRAL DOGMA OF MOLECULAR BIOLOGY

Complete set of transcripts and relative levels of expression in a particular cell or tissue under defined conditions at a given time

GENOME ⇌ TRANSCRIPTOME → PROTEOME

Complete DNA content of an organism with all its gene and regulatory sequences

Complete collection of proteins and their relative levels in each cell



TRANSCRIPTOMICS

The study of the complete set of RNAs (transcriptome) encoded by the genome of a specific cell or organism at a specific time or under a specific set of conditions is called **Transcriptomics**.

In transcriptomics, the expression of genes by a genome is studied :

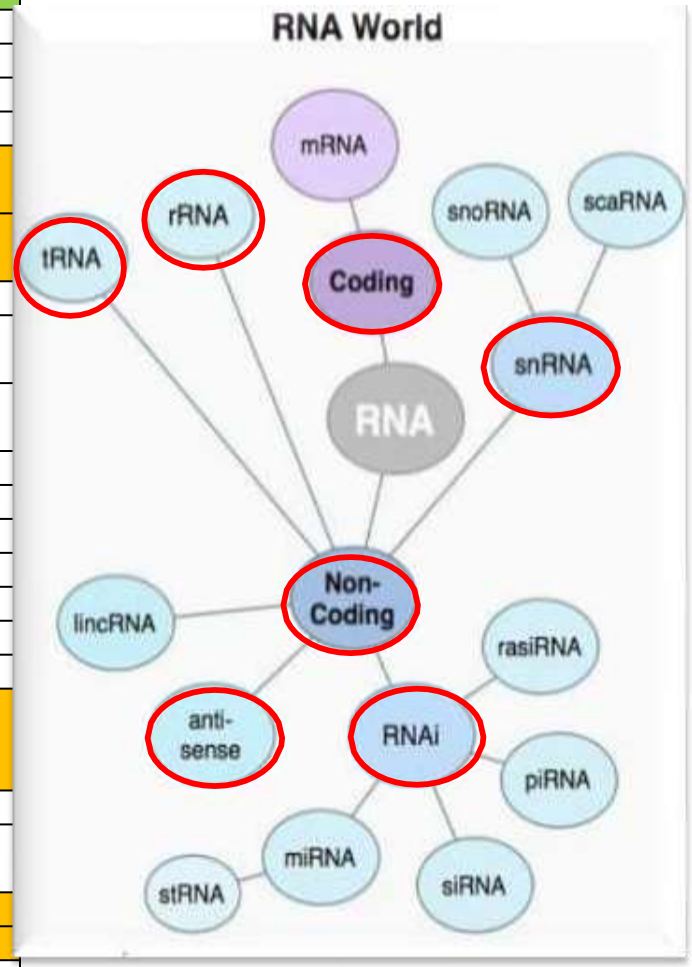
- ✓ Qualitatively (identifying which genes are expressed and which are not)
- ✓ Quantitatively (measuring varying levels of expression for different genes)

The Transcriptome

RNAs involved in protein synthesis			
Messenger RNA	mRNA	Codes for protein	All organisms
Ribosomal RNA	rRNA	Translation	All organisms
Signal recognition particle RNA	7SL RNA or SRP RNA	Membrane integration	All organisms
Transfer RNA	tRNA	Translation	All organisms
Transfer-messenger RNA	tmRNA	Rescuing stalled ribosomes	Bacteria

RNAs involved in post-transcriptional modification or DNA replication			
Type	**Abbr.**	**Function**	**Distribution**
Small nuclear RNA	snRNA	Splicing and other functions	Eukaryotes and archaea
Small nucleolar RNA	snoRNA	Nucleotide modification of RNAs	Eukaryotes and archaea
SmY RNA	SmY	mRNA trans-splicing	Nematodes
Small Cajal body-specific RNA	scaRNA	Type of snoRNA; Nucleotide modification of RNAs	
Guide RNA	gRNA	mRNA nucleotide modification	Kinetoplastid mitochondria
Ribonuclease P	RNase P	tRNA maturation	All organisms
Ribonuclease MRP	RNase MRP	rRNA maturation, DNA replication	Eukaryotes
Y RNA		RNA processing, DNA replication	Animals
Telomerase RNA		Telomere synthesis	Most eukaryotes

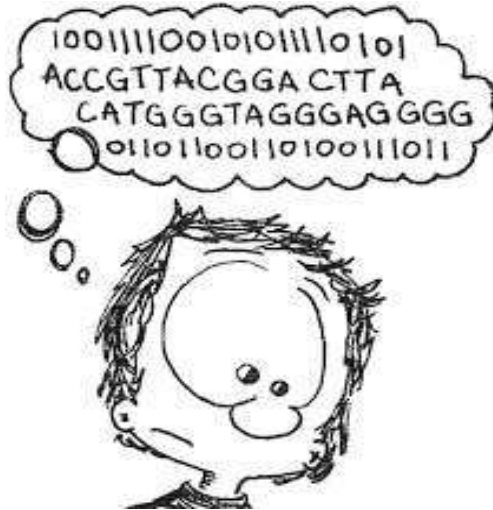
Regulatory RNAs			
Type	**Abbr.**	**Function**	**Distribution**
Antisense RNA	aRNA	Transcriptional attenuation / mRNA degradation / mRNA stabilisation / Translation block	All organisms
Cis-natural antisense transcript		Gene regulation	
CRISPR RNA	crRNA	Resistance to parasites, probably by targeting their DNA	Bacteria and archaea
Long noncoding RNA	Long ncRNA	Various	Eukaryotes
MicroRNA	miRNA	Gene regulation	Most eukaryotes
Piwi-interacting RNA	piRNA	Transposon defense, maybe other functions	Most animals
Small interfering RNA	siRNA	Gene regulation	Most eukaryotes
Trans-acting siRNA	tasiRNA	Gene regulation	Land plants
Repeat associated siRNA	rasiRNA	Type of piRNA; transposon defense	Drosophila



Why transcriptomics is important?

Transcriptome profiling provides clues to:

- ✓Expressed sequences and genes of genome
- ✓ Gene regulation and regulatory sequences
- ✓ Gene function annotation
- ✓ Functional differences between tissue and cell types
- ✓ Identification of candidate genes for any given process or disease



Approaches for transcriptome mining

Candidate gene approaches

- Northern Blotting Analysis
- quantitative real-time RT-PCR

Hybridization based approaches

- Microarray expression profiling

Sequencing based approaches

- EST libraries
- Serial analysis of gene expression (SAGE)
- CAGE
- RNA sequencing
- MPSS

Sequence-based approaches

- Directly determine the cDNA sequence, hence defining the corresponding mRNA

1. Sanger sequencing of cDNA or EST libraries
   - Low-throughput, expensive, generally not quantitative

2. Tag-based methods were developed: SAGE, CAGE, MPSS
   - Still expensive because based on Sanger sequencing, short tags cannot be uniquely mapped to the reference genome, isoforms are generally not distiguishable

3. RNA-seq, based on NGS technologies

- **By analyzing the transcriptome at spectacular and unprecedented depth and accuracy, thousands of new transcripts variants and isoforms have been shown to be expressed in mammalian tissues or organs**

- **it greatly accelerated our understanding of the complexity of gene expression, regulation and networks for mammalian cells**

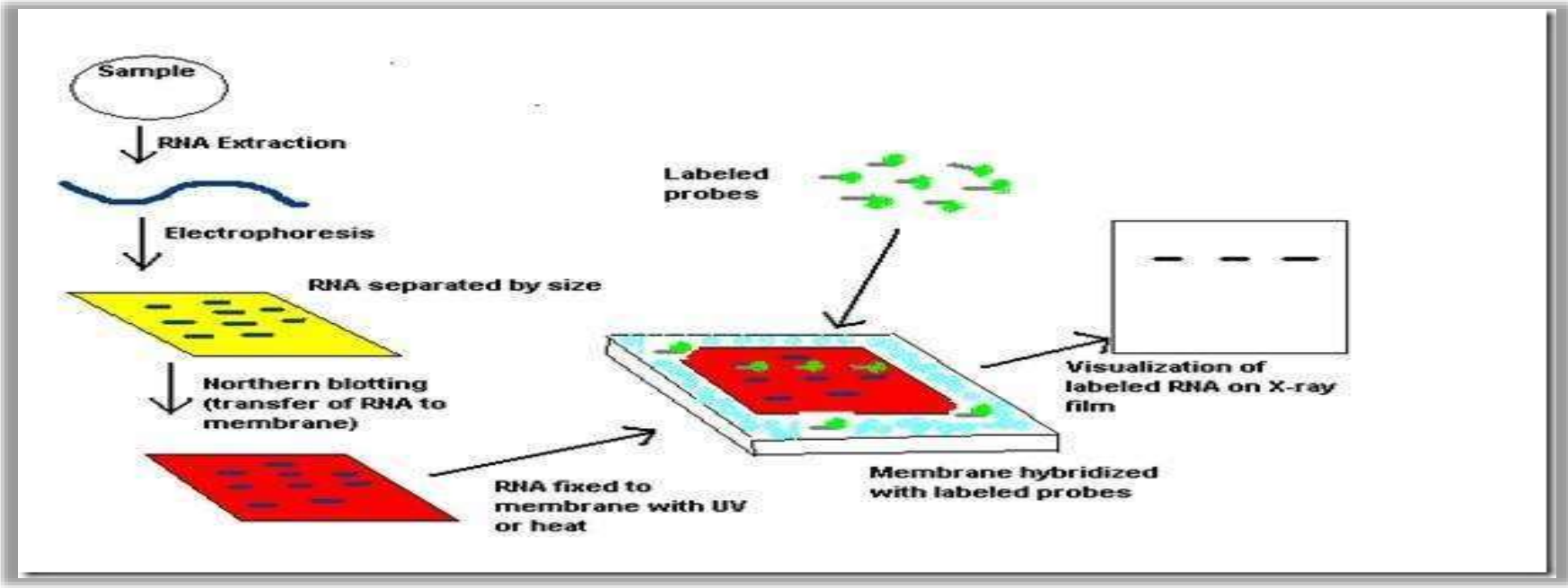
Table 1 | High-throughput technology used for transcriptome analyses

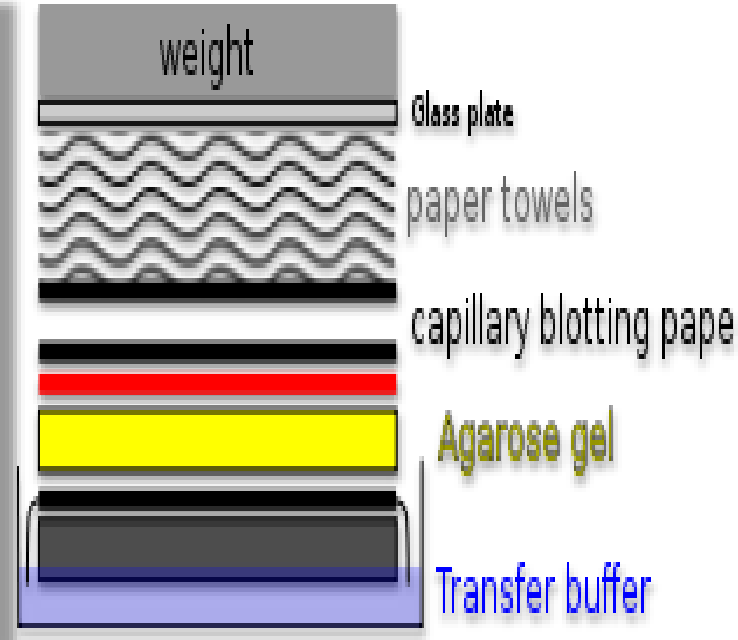
Technique	Principle	Aim	Organism	Main references
SAGE	Reverse transcription to cDNA of the transcriptome Short sequence tag production from unique poly(A)+ transcripts Cloning and sequencing of sequence tag arrays Quantitation of tag recovery	Quantitative gene expression data without prerequisite of sequence knowledge	*S. cerevisiae*	Velculescu *et al*, 1997
3'-long SAGE	Similar to SAGE Introduction of a recognition site for a IIS restriction enzyme at the 3' end of cDNA, recovery of 3' end tags High-throughput sequencing	Identification of 3' ends Distinction between overlapping transcripts	*S. cerevisiae*	Neil *et al*, 2009
CAGE	Similar to SAGE but from CAP-selected RNAs	Identification of transcription start sites	Mouse Human	Carninci *et al*, 2006
ASSAGE	Similar to SAGE RNA treatment with bisulphite (changes cytosine residues to uracil) before reverse transcription	Unambiguous strand specificity	Human	He *et al*, 2008
RNA-Seq	Reverse transcription to cDNA of poly(A)+ RNA High-throughput sequencing	Quantitative transcriptome data High sensitivity, low background, high resolution	*S. cerevisiae* *S. pombe*	Nagalakshimi *et al*, 2008; Wilhelm *et al*, 2008
GRO-seq	Extension of nascent RNA associated with engaged RNAPII (nuclear run-on) Isolation and purification of the RNAs Reverse transcription to cDNA High-throughput sequencing	Genome-wide positional mapping, determination of transcript amount, orientation of engaged RNAPII	Human	Core *et al*, 2008
High-resolution tiling arrays	Reverse transcription of total or selected RNA followed by hybridization on DNA chip Probes overlap and cover the entire genome	Quantitative transcriptome data from sequenced genomes Strand specific	*S. cerevisiae* *S. pombe* Rice Human	David *et al*, 2006; Xu *et al*, 2009; Dutrow *et al*, 2008; Wilhelm *et al*, 2008; L. Li *et al*, 2006; Kapranov *et al*, 2007a; Preker *et al*, 2008
ChIP-chip	Chromatin immunoprecipitation of RNAPII and TFs Hybridization on DNA array	Mapping of chromatin-bound RNAPII and TFs, genome-wide	*S. cerevisiae* Human	Steinmetz *et al*, 2006; Kim *et al*, 2005
ChIP-Seq	Chromatin immunoprecipitation of RNAPII High-throughput sequencing	Mapping of chromatin-bound RNAPII, also at low complexity sequences (such as telomeres and transposons)	Mouse	Seila *et al*, 2008

Northern Blotting Analysis

- The northern blot, or RNA blot, is a technique used in molecular biology research to study gene expression by detection of RNA (or isolated mRNA) in a sample.

- The northern blot technique was developed in 1977 by **James Alwine, David Kemp,** and **George Stark** at Stanford University.





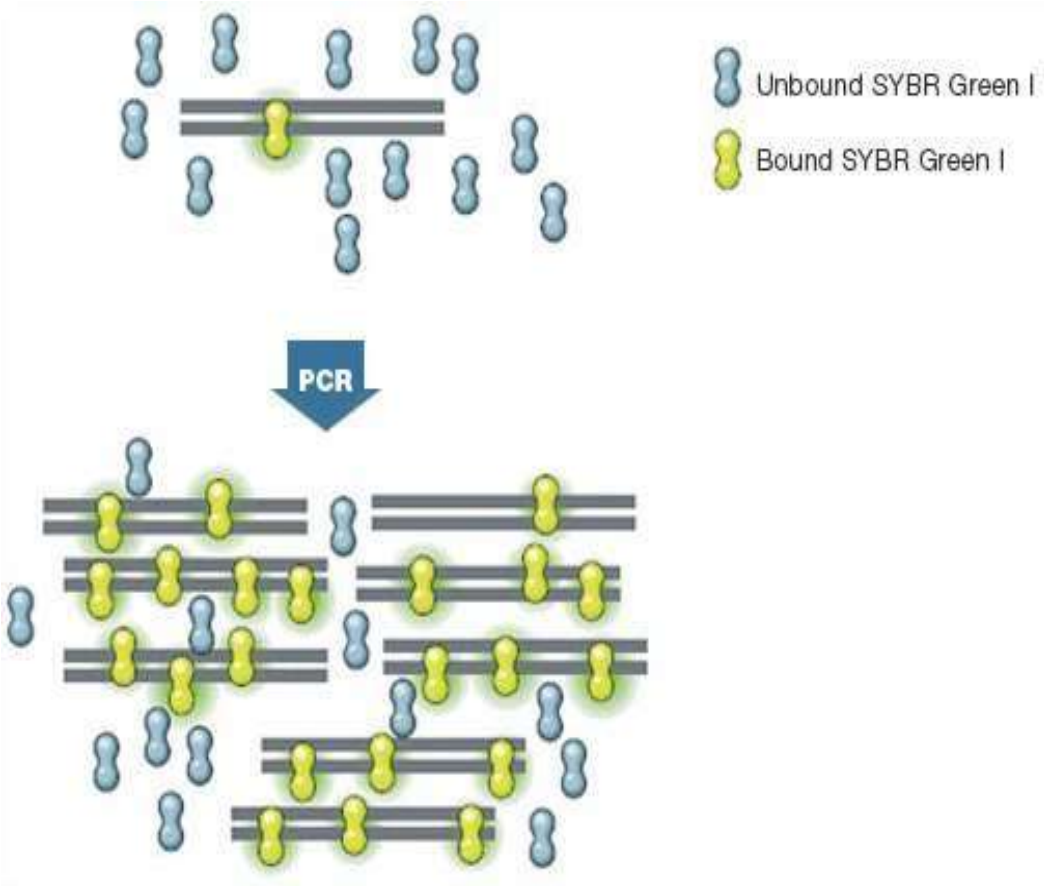
Reverse transcription polymerase chain reaction (RT-PCR)

A variant of polymerase chain reaction (PCR), commonly used in molecular biology to detect RNA expression.

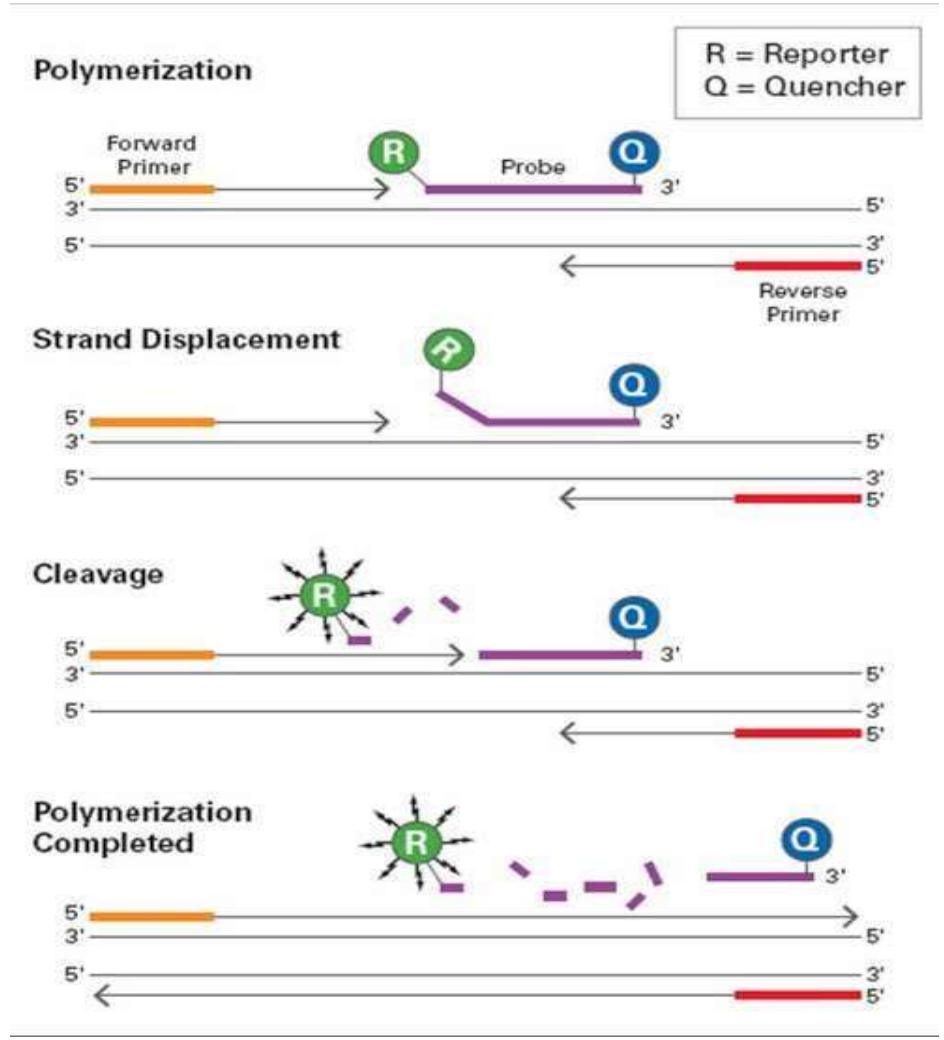
Application

- RT-PCR is used to clone expressed genes by reverse transcribing the RNA of interest into its DNA complement through the use of reverse transcriptase.
- Utilized for quantification of RNA, by incorporating qPCR into the technique. The combined technique, described as **quantitative real-time RT-PCR (RT-qPCR)**

Probes for qRT PCR



SYBR Green



Taqman Probe

PCR amplification plot

Fluorescence emission is measured continuously during the PCR reaction and ▵Rn (increase in fluorescence emission, from which the background fluorescence signal is subtracted) is plotted against cycle number. The threshold cycle (Ct) is the cycle at which the fluorescence exceeds a chosen threshold.

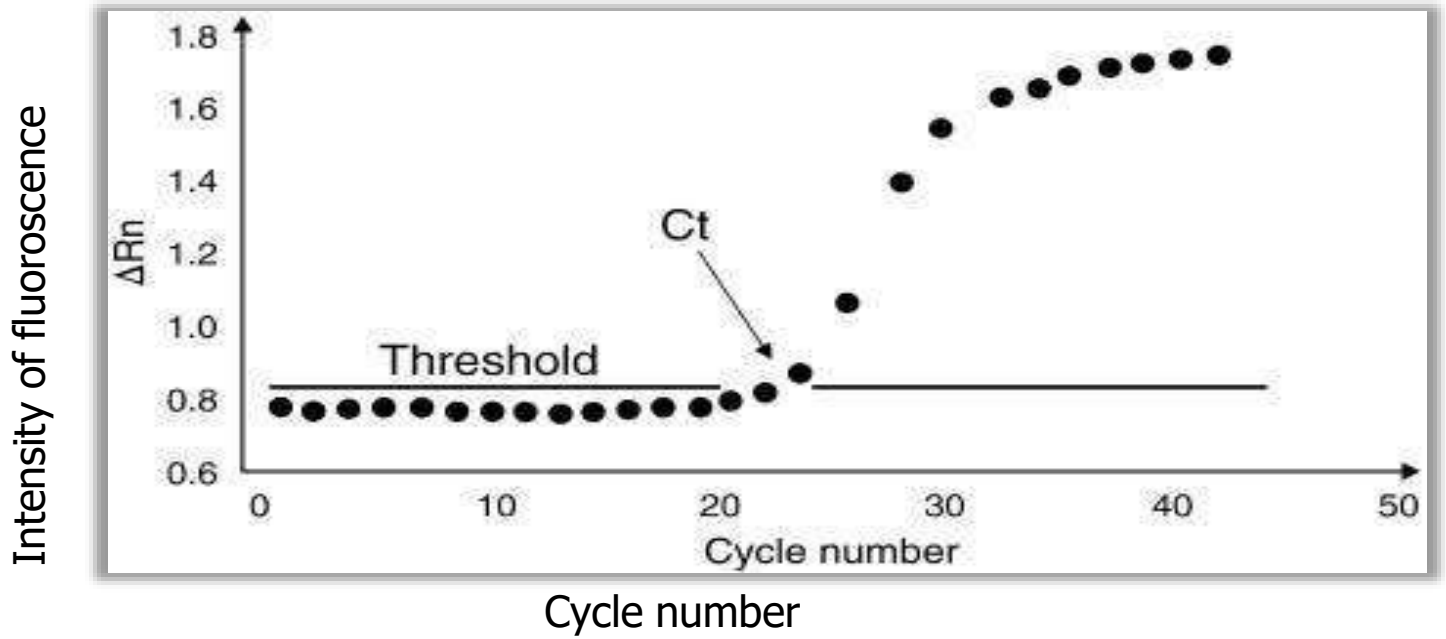
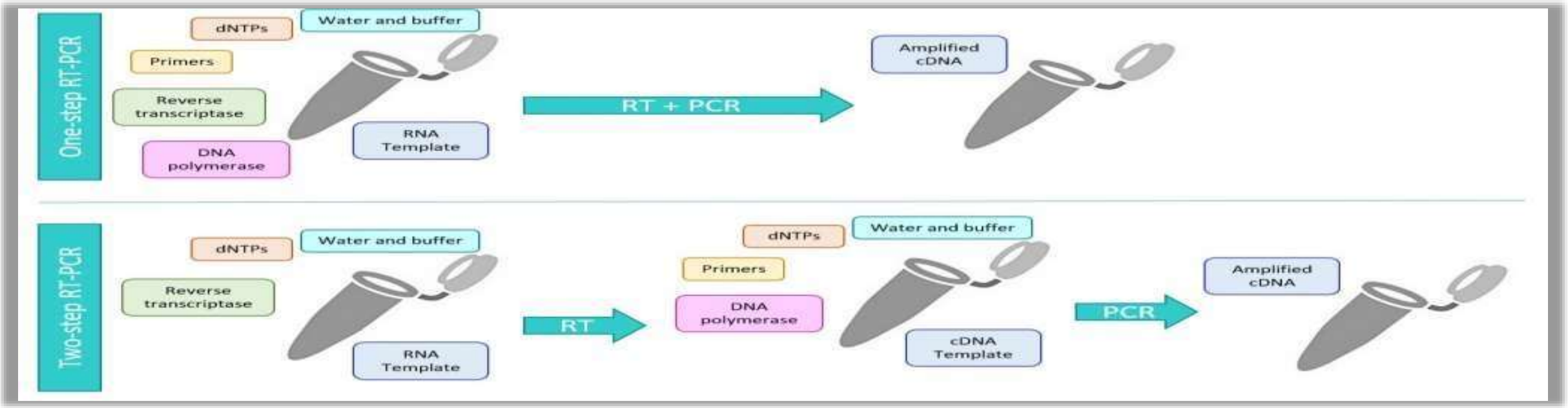


Fig: PCR amplification plot

One-step vs. Two-step RT-qPCR

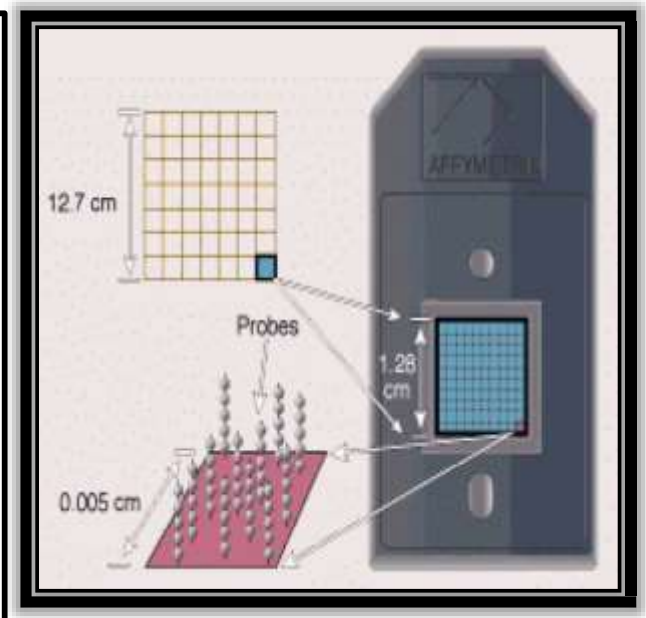
- **One-step assays** combine reverse transcription and PCR in a single tube and buffer, using a reverse transcriptase along with a DNA polymerase. One-step RT-qPCR only utilizes sequence-specific primers.

- **In two-step assays**, the reverse transcription and PCR steps are performed in separate tubes, with different optimized buffers, reaction conditions, and priming strategies.



MICROARRAY

- Microarray is a nucleic acid hybridization based, high throughput technique developed to quantitate gene expression levels at the whole genome scale.
- A **microarray** is a pattern of ssDNA probes which are immobilized on a surface (called a chip or a slide) in a regular pattern of spots and each spot containing millions of copies of a unique DNA probe.
- **Principle:** base-pairing hybridization
- The technique was first time used for "**Quantitative Monitoring of Gene Expression Patterns with a complementary DNA microarray**" by **Patrick Brown, Mark Schena** and colleague published in Science (1995).



Mark Schena
"Father of Microarray Technology"

Methodology

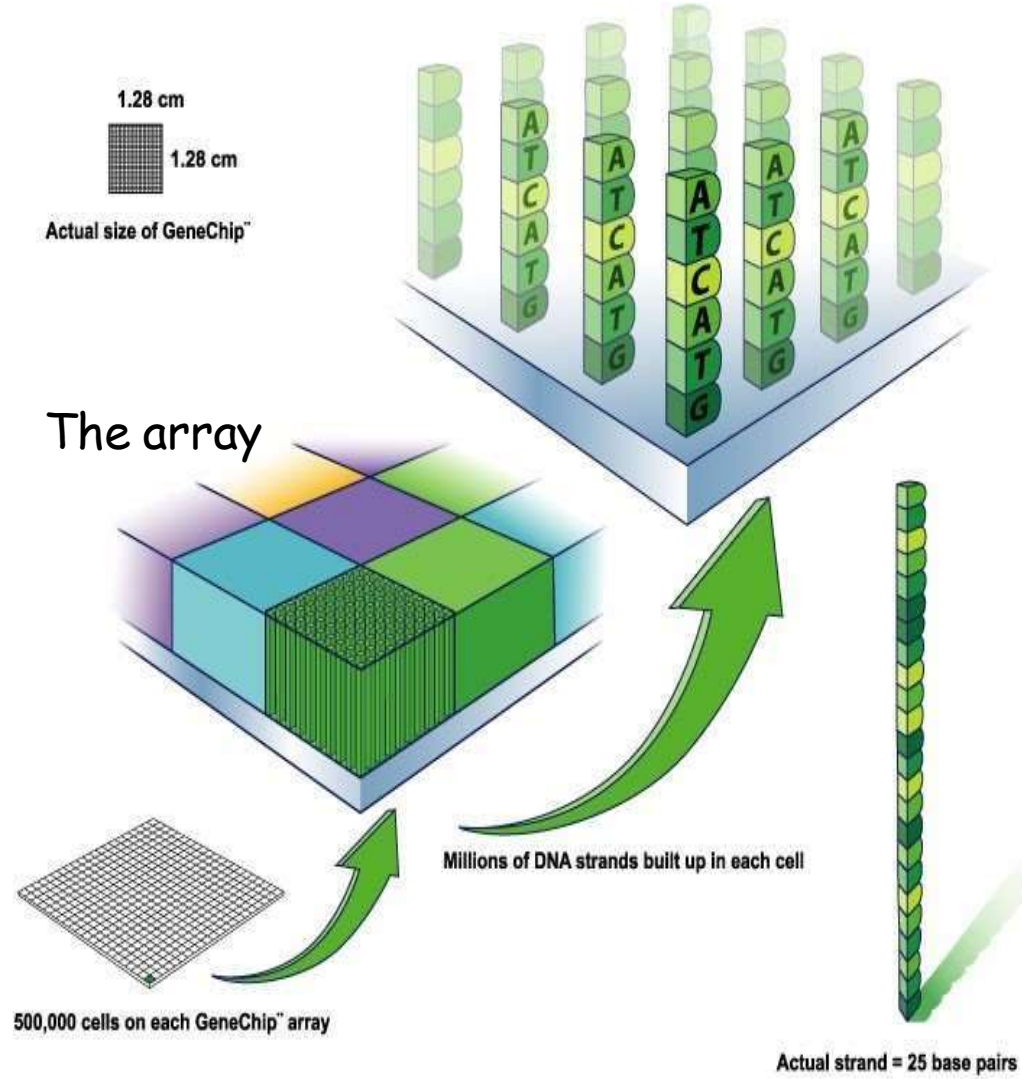
Microarray analysis can be divided into

1. Probe production

Specific sequences are immobilized to a surface and reacted with labelled cDNA targets.

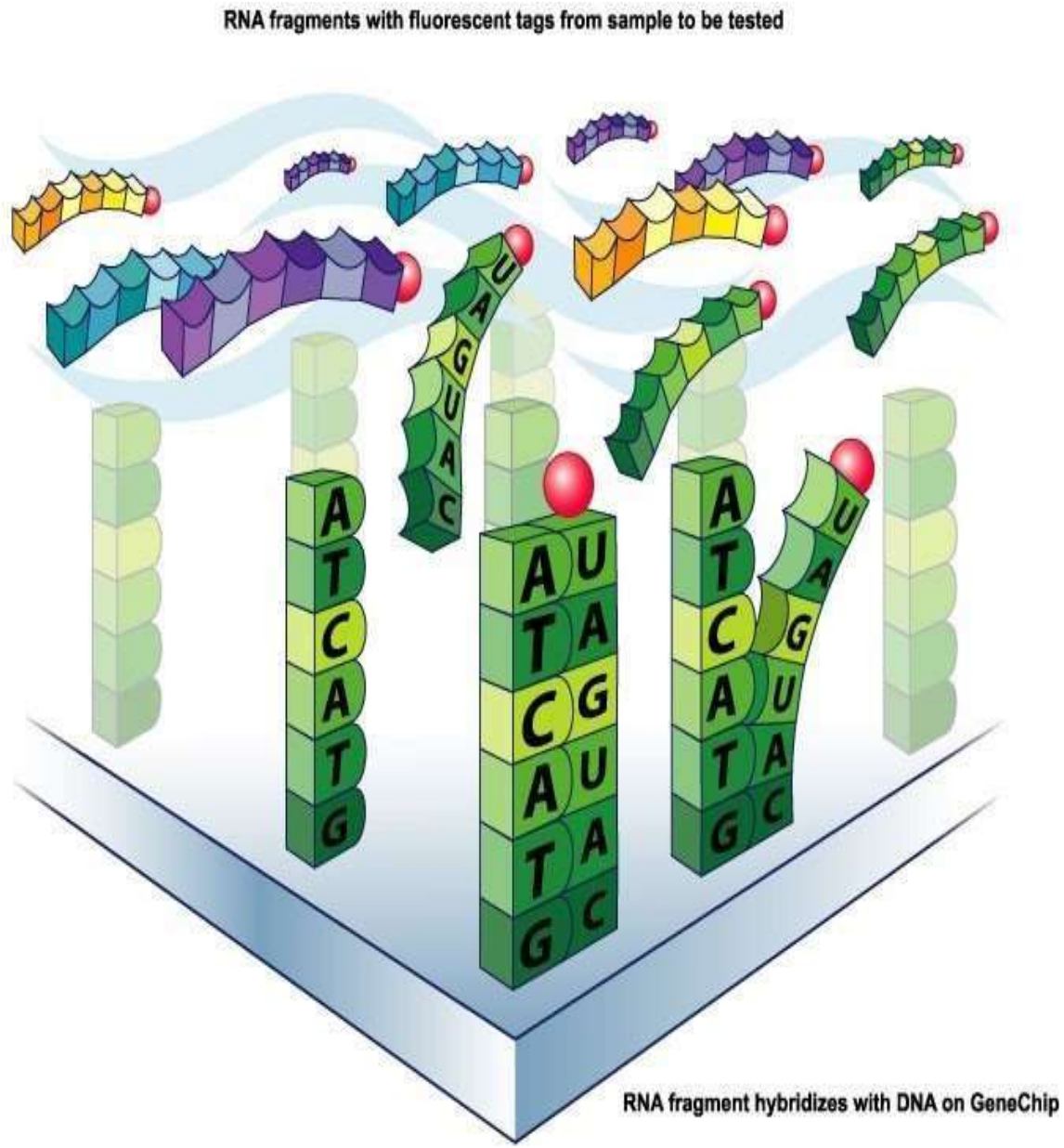
2. Target (cDNA) production

mRNA is extracted from the sample, converted to cDNA and labeled using fluorescent dyes (usually Cy3 or Cy5).



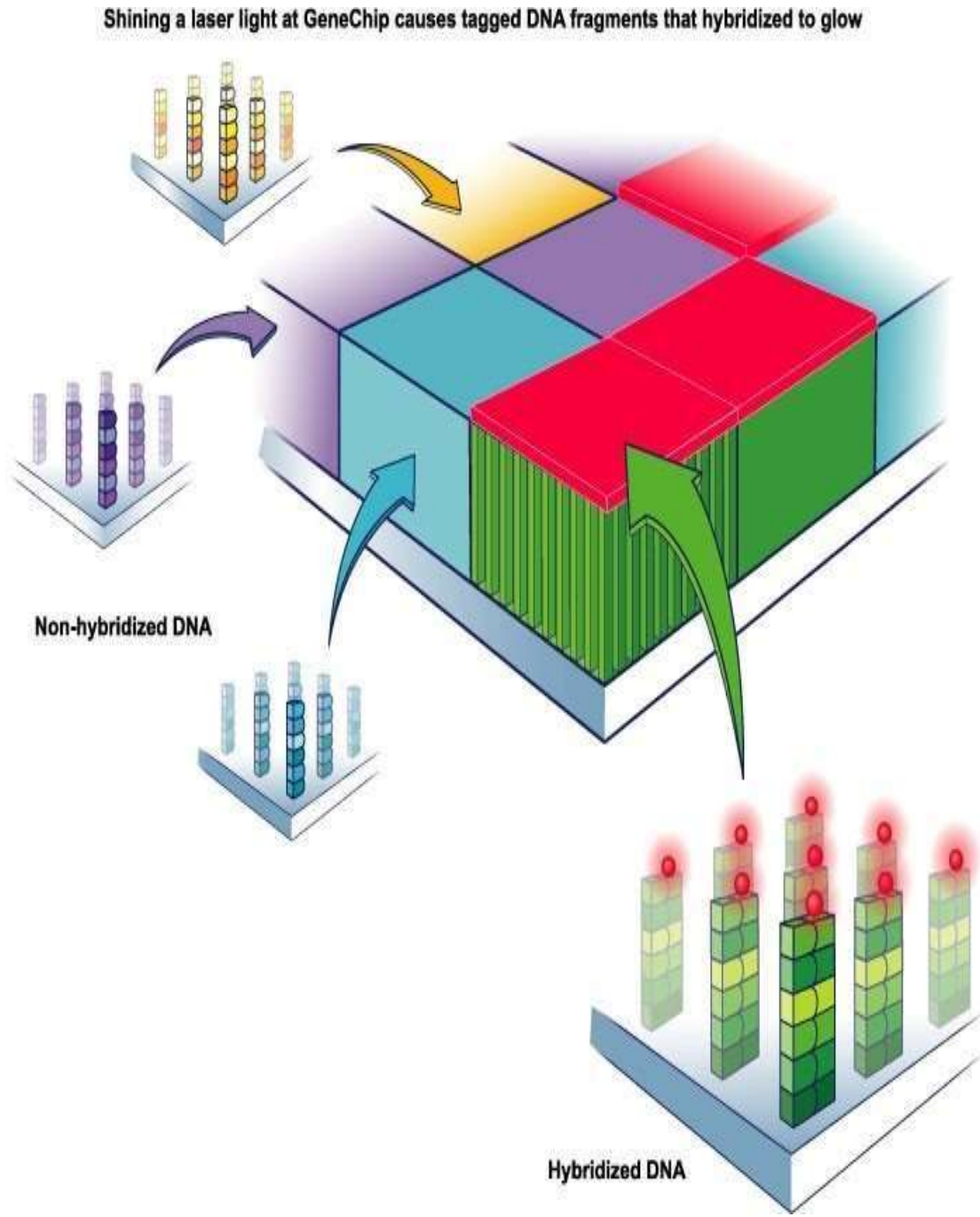
3. Hybridisation

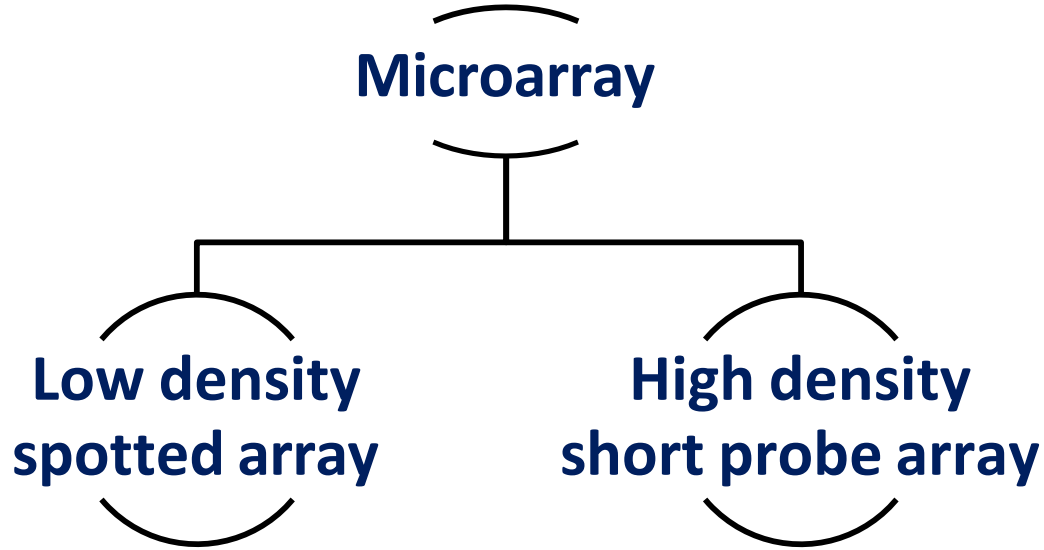
- ✓ The chip is exposed to a solution containing extracted labelled cDNA
- ✓ Complementary nucleic acid sequences get pair via hydrogen bonds.
- ✓ Washing off of non-specific bonding sequences .



4. Scanning

- ✓ The array is scanned to measure fluorescent label.
- ✓ Fluorescently labelled target sequences that bind to a probe sequence generate a signal.
- ✓ The signal depends on.
  - The hybridization conditions,
    ex: temperature
  - washing after hybridization
- ✓ Total strength of the signal, depends upon the amount of target sample.

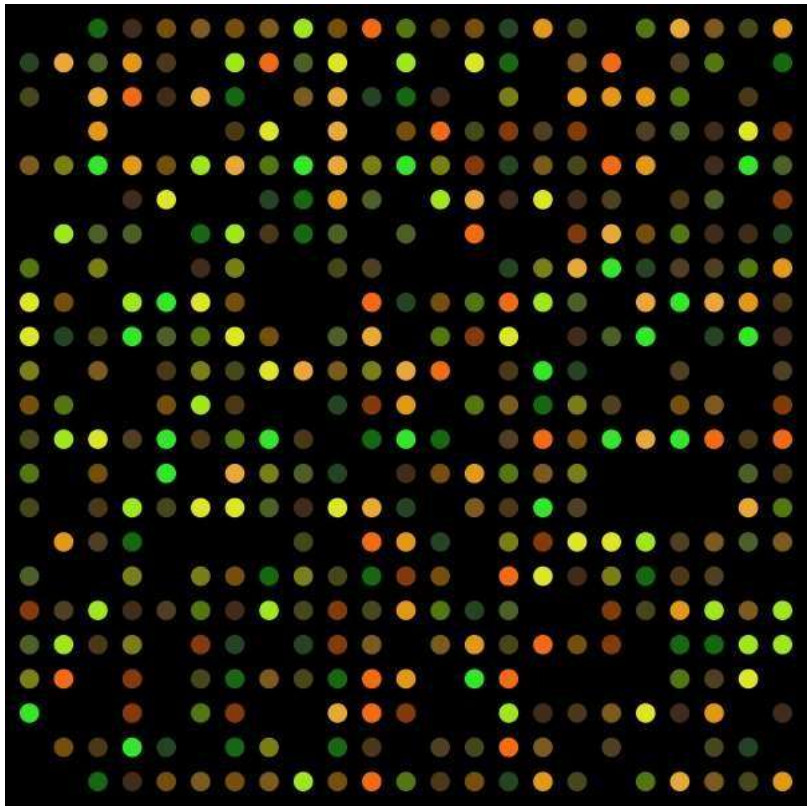




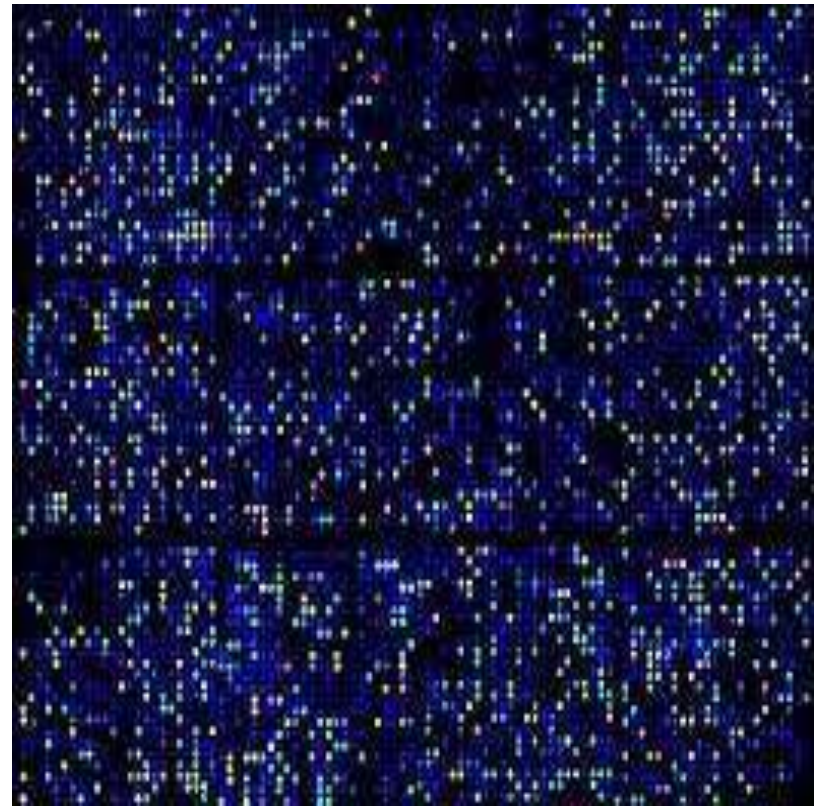
- ✓ prepared by **using cDNA** known as cDNA chips , cDNAs are amplified using PCR than Spotted by **inkjet technology**
- ✓ **Probes are >1000 nucleotides** in length
- ✓ **Two different fluorophores** can be used to label the test and control sample (two-channel detection)
- ✓ Cannot identify alternative splicing events

- ✓ short DNA oligonucleotides are **directly synthesised on solid microarray substrate** using photolithgraphy (Affymatrix) or spotted by ink-jet printing.
- ✓ **Probes are 70-80 nucleotides** in length
- ✓ Uses a **single fluorescent lable** (single channel)
- ✓ Can identify alternative splicing events
- ✓ Commercial oligonucleotide chips are available (e.g. Affymetrix, Inc. GeneChip system)

The Colours of a Microarray



cDNA chip



Affymatrix gene chip

Applications of Microarray

- Gene expression profiling
- Gene fusions detection
- Alternative splicing detection
- SNP detection
- TILING array
- Chip

ADVANTAGES

- High-throughput
- Fast
- Relatively inexpensive

LIMITATIONS

- Reliance upon existing knowledge about genome sequence
- High background levels owing to cross-hybridization
- A limited dynamic range of detection due to both background and saturation of signals
- Comparing expression levels across different experiments is often difficult and can require complicated normalization methods

ESTs

- In 1983, SD Putney for the first time demonstrated the use of cDNA in identification of genome.
- In 1991 Adams and co-workers coined the term EST.

- They are the tiny sequences of cistron randomly selected from genome library and can be used to identify and map the whole genome of any particular species. ESTs are usually 200 to 500 nucleotides long and are generated by sequencing the ends of DNA.

- Use of EST
- ✓ Identify unknown gene and map their position in a genome.
- ✓ Provide simple and inexpensive path for discovering new gene.
- ✓ Genome map construction.
- ✓ Characterization of expressed gene

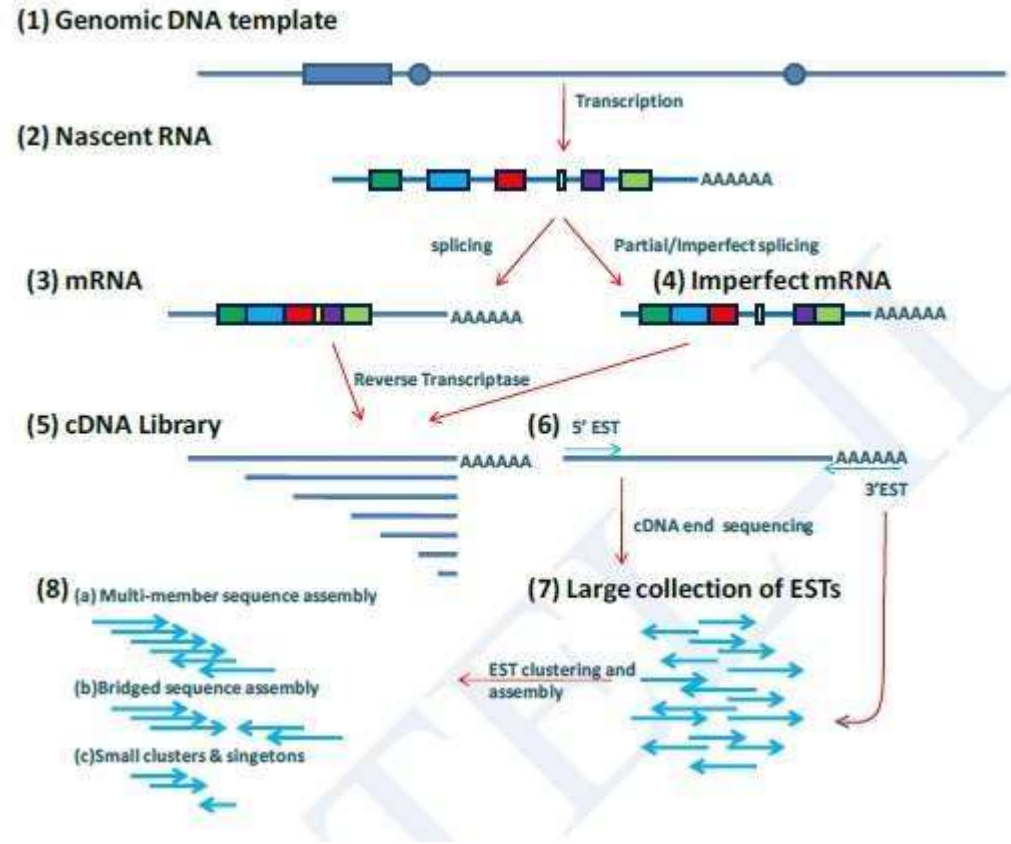


Figure: Method of construction of ESTs from nascent DNA.

RNA-Sequencing

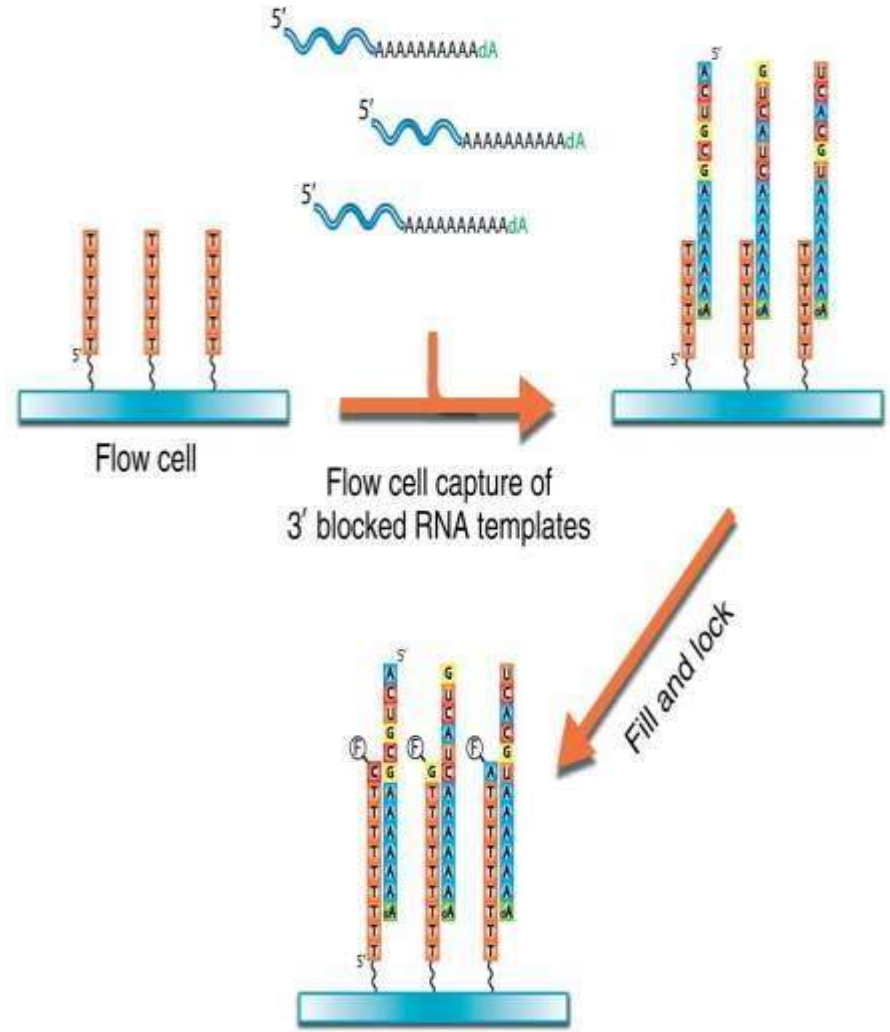
- RNA-seq, also called whole-transcriptome shotgun sequencing, refers to the use of high-throughput sequencing technologies to reveal the presence and quantity of RNA in a biological sample at a given moment.

Types of RNA sequencing

1. Direct sequencing
2. Indirect sequencing (next-generation sequencing)

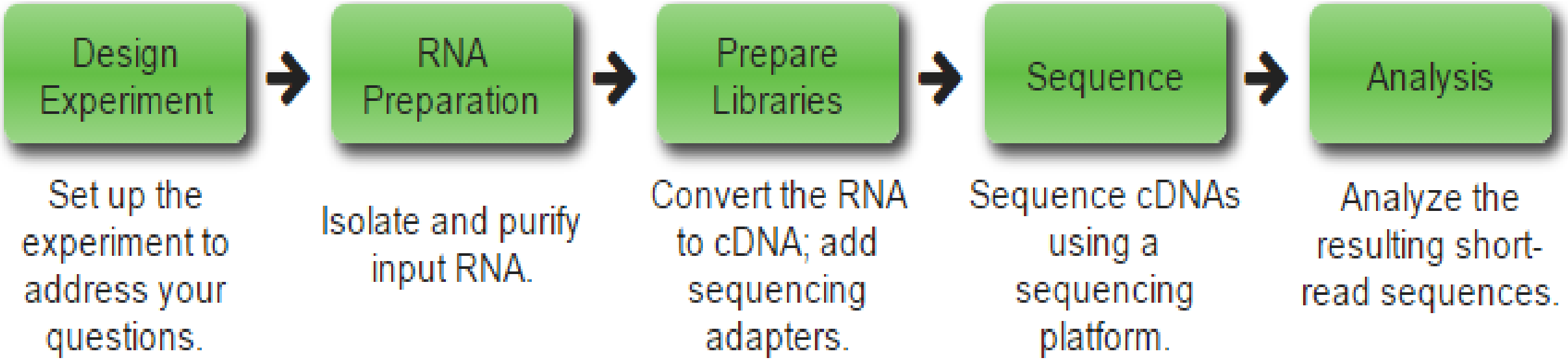
Direct RNA sequencing

- Direct single molecule RNA sequencing without prior cDNA conversion
- Sequencing by synthesis
- Polyadenylation by poly (A) polymerase I (PAPI) from *E. coli*
- Blocking by 3′ deoxyATP
- Poly (A) tail ~ 150 nucleotides
- Each RNA molecule is filled in with dTTP and polymerase and than locked in position VT-A, VT-C, VT-G, stopping the further nucleotide addition.
- Unincorporated dye labelled nucleotides are washed away, images are taken.
- Flourescent dye and inhibitor are cleaved off from the incorporated nucleotide, making it suitable for additional rounds of incorporation.
- Repeating this cycle of rinsing, imaging and cleaving provides a set of images that are aligned and are used to generate the sequence information for each individual RNA molecule with real time image processing.
- Requirement of minor RNA quantities
- No biases due to cDNA synthesis, end repair, ligation and amplification procedures
- Potentially useful to study short RNA species

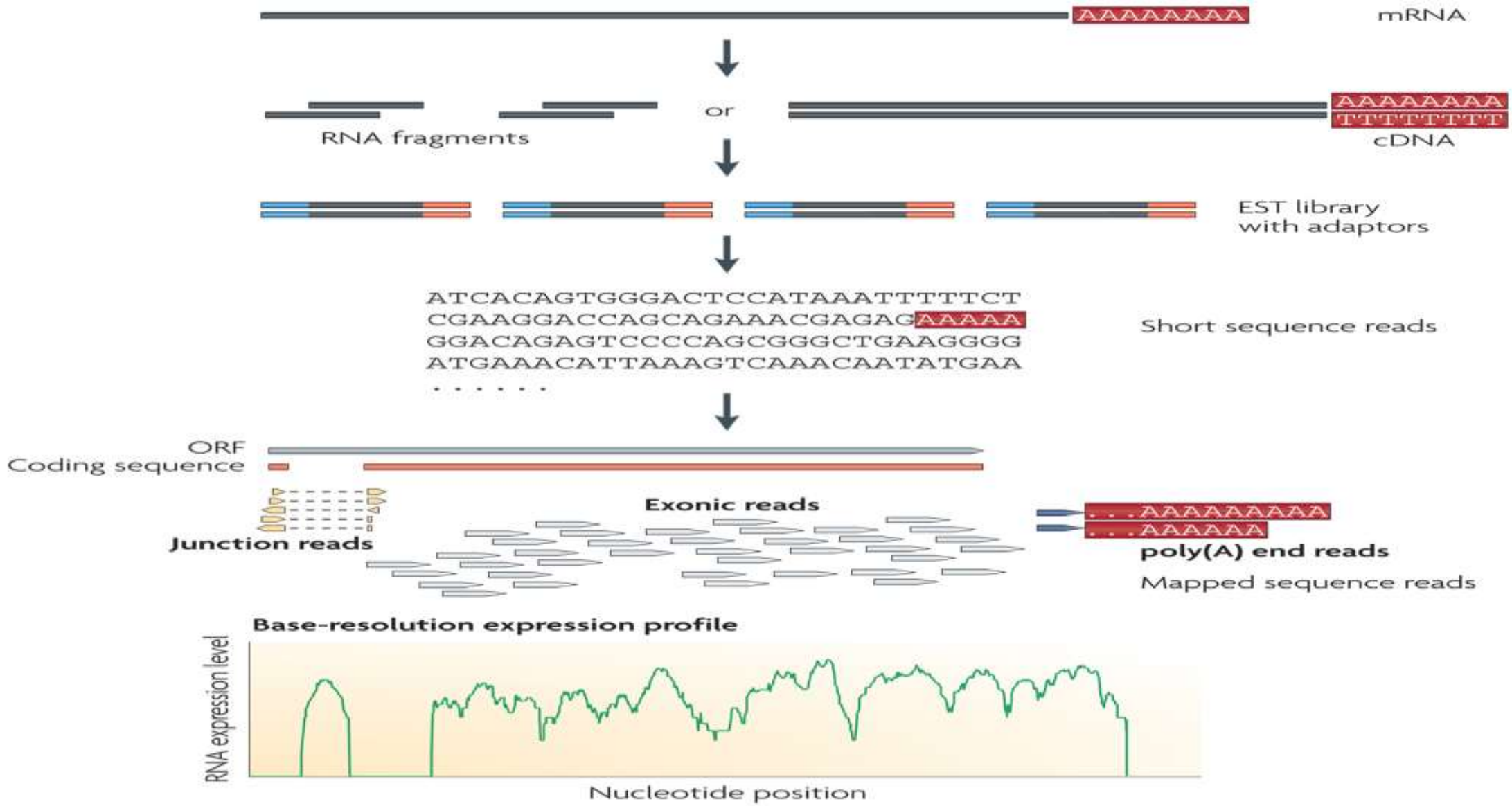


RNA-Sequencing WORKFLOW:

A typical RNA-Seq experiments follows these steps:



A typical RNA-Seq Experiment



Some mRNA-Seq Applications

- Differential gene expression analysis
- Transcriptional profiling
- Identification of splice variants
- Novel gene identification
- Transcriptome assembly
- SNP finding
- RNA editing

Chromatin Immunoprecipitation

Not only do proteins interact with one another, they can also interact with DNA. Chromatin immunoprecipitation (ChIP) is a technique that determines whether a protein of interest interacts with a specific DNA sequence. This technique is often used to study the repertoire of sites on DNA that are bound by particular transcription factors or by histone proteins, and to look at the precise genomic locations of various histone modifications (including acetylation, phosphorylation, or methylation).

PRINCIPLE BEHIND CHIP

ChIP can be used to examine the presence of protein-DNA interaction at steady state, or to quantify changes in interaction at specific phases of the cell cycle, or following a treatment of interest. Protein and associated chromatin are temporarily cross-linked in live cells or tissues (using formaldehyde or UV) and sheared using enzymatic digestion or sonication to yield ~300-1000 bp fragments of DNA. The protein of interest, along with any associated DNA fragments, is immunoprecipitated from the cell debris using a specific antibody. The cross-link is then reversed and DNA fragments are purified. The amount of eluted DNA can be assessed through quantitative real-time PCR (qRT-PCR) using primers flanking the genomic locus of interest. DNA amplification is an indication of enrichment in binding of the protein of interest.

Modified ChIP Techniques

DNA fragments purified by ChIP can be utilized for a number of downstream analysis techniques. Furthermore, the basic ChIP protocol described above can be modified to answer additional biological questions.

ChIP-on-chip: Genome-wide analysis of protein binding sites using microarray analysis of purified DNA fragments.

ChIP-Seq: Genome-wide analysis of protein binding sites using deep sequencing of purified DNA fragments.

Native ChIP: Omits the cross-linking step and uses micrococcal nuclease digestion to cut DNA at histone linkers to examine the DNA target of histone modifying proteins.

ChIP-exo: Addition of an exonuclease digestion step to obtain increased resolution of protein binding sites, up to a single base pair.

ChIA-PET (chromatin interaction analysis by paired-end tag sequencing): A technique that combines the principles of ChIP with chromosome conformation capture (3C) to detect long-range chromatin interactions mediated via a protein of interest.

iChIP (indexing-first chromatin immunoprecipitation): A high-sensitivity technique that reduces the number of cells required for a ChIP experiment by initially barcoding total cellular chromatin.

enChIP (engineered DNA-binding molecule-mediated chromatin immunoprecipitation): A technique which employs the CRISPR/Cas9 system to target specific genomic regions. A guide RNA complementary to the desired genomic region is expressed in combination with a tagged, enzymatically inactive Cas9 protein. ChIP is then performed using an antibody against the modified Cas9. This technique can help evaluate cis- and trans-interacting chromosomal looping events.

RIP-Chip/RIP-Seq: Similar techniques used to analyze protein-RNA interactions.

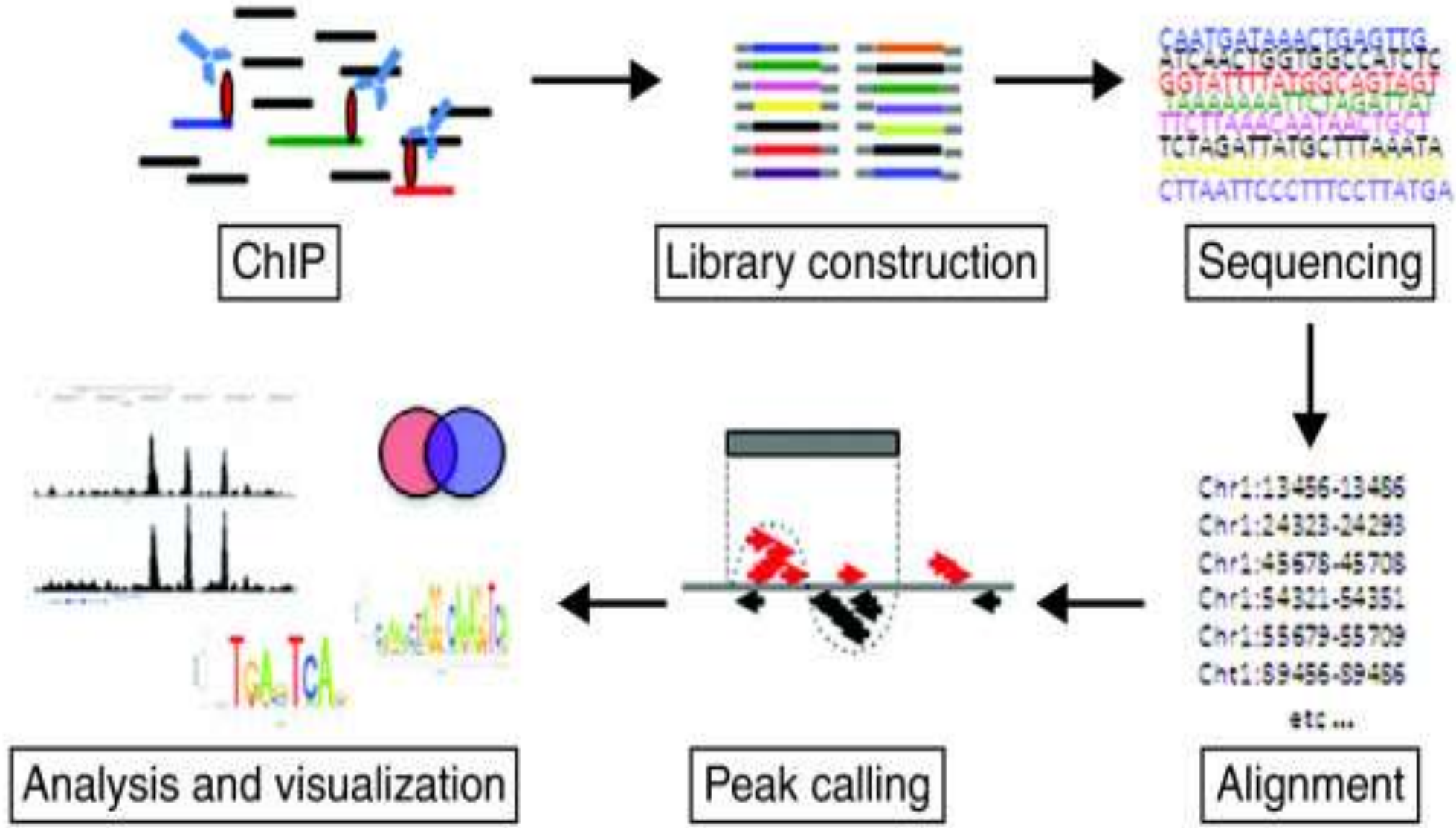
ChIP Sequencing

ChIP sequencing, also known as **ChIP-Seq,** is a method used to analyze protein interactions with DNA.

ChIP stands for Chromatin Immuno-Precipitation and **seq** refers to the high throughput sequencing to detect bound genomic locations.

ChIP-Seq combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins.

It can be used to map global binding sites precisely for any protein of interest


ChIP
Library construction
Sequencing
Chr1:13486-13486
Chr1:24323-24293
Chr1:45678-45708
Chr1:54321-54351
Chr1:55679-55709
Chr1:69486-69486
etc...
Alignment
Peak calling
Analysis and visualization

USES

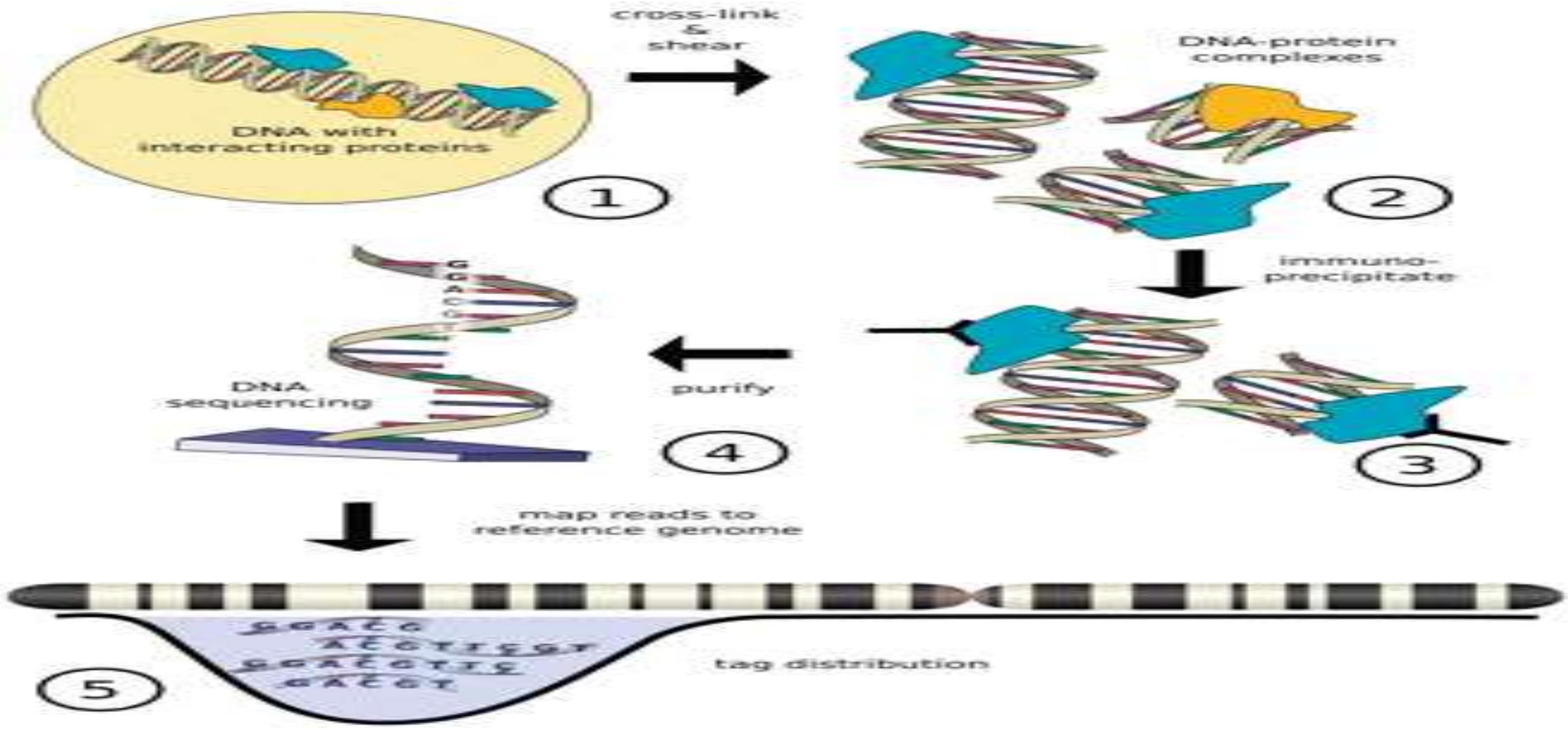
ChIP-seq is used primarily to determine how transcription factors and other chromatin-associated proteins influence phenotype affecting mechanisms.

Determining how proteins interact with DNA to regulate gene expression is essential for fully understanding many biological processes and disease states.

Specific DNA sites in direct physical interaction with transcription factors and other proteins can be isolated by chromatin immunoprecipitation.

ChIP produces a library of target DNA sites bound to a protein of interest *in vivo*

Overview of ChIP sequencing



GENE EXPRESSIONS ANALYSIS BY MASSIVELY PARALLEL SIGNATURE SEQUENCING (MPSS)

DNA microarray, serial analyses of gene expression (SAGE), cDNA sequencing and a variety of other technologies are available for analysing the expression of hundreds to thousands of genes simultaneously.

Each of these existing technologies has limitations when it comes to generating complete data sets for building relational databases.

Massively Parallel Signature Sequencing (MPSS) is a an open-ended platform that analyses the level of gene expression in a sample by counting the number of individual mRNA molecules produced by each gene.

Massively Parallel Signature Sequencing (MPSS) a *sequencing* technique is developed by Sydney Brenner which is bacteria-free bead-based library preparation, "Megaclone" technology.

In MPSS, mRNA transcripts did not need to be known and could be discovered *de novo.*

Genes with low level expression could be quantified by MPSS.

All clones in a microbead library can be sequenced simultaneously (so, called "massively parallel").

MPSS produces data in a digital format. MPSS Captures data by counting virtually all mRNA molecules in a tissue or cell sample.

All genes are analysed simultaneously, and bioinformatics tools are used to sort out the number of mRNAs from each gene relative to the total number of molecules in the sample.

Even genes that are expressed at low levels can be quantified with high accuracy.

Counting mRNAs with MPSS is based on the ability to identify uniquely every mRNA in a sample by generating a 17-base sequence for each mRNA at a specific site upstream from its poly (A) tail.

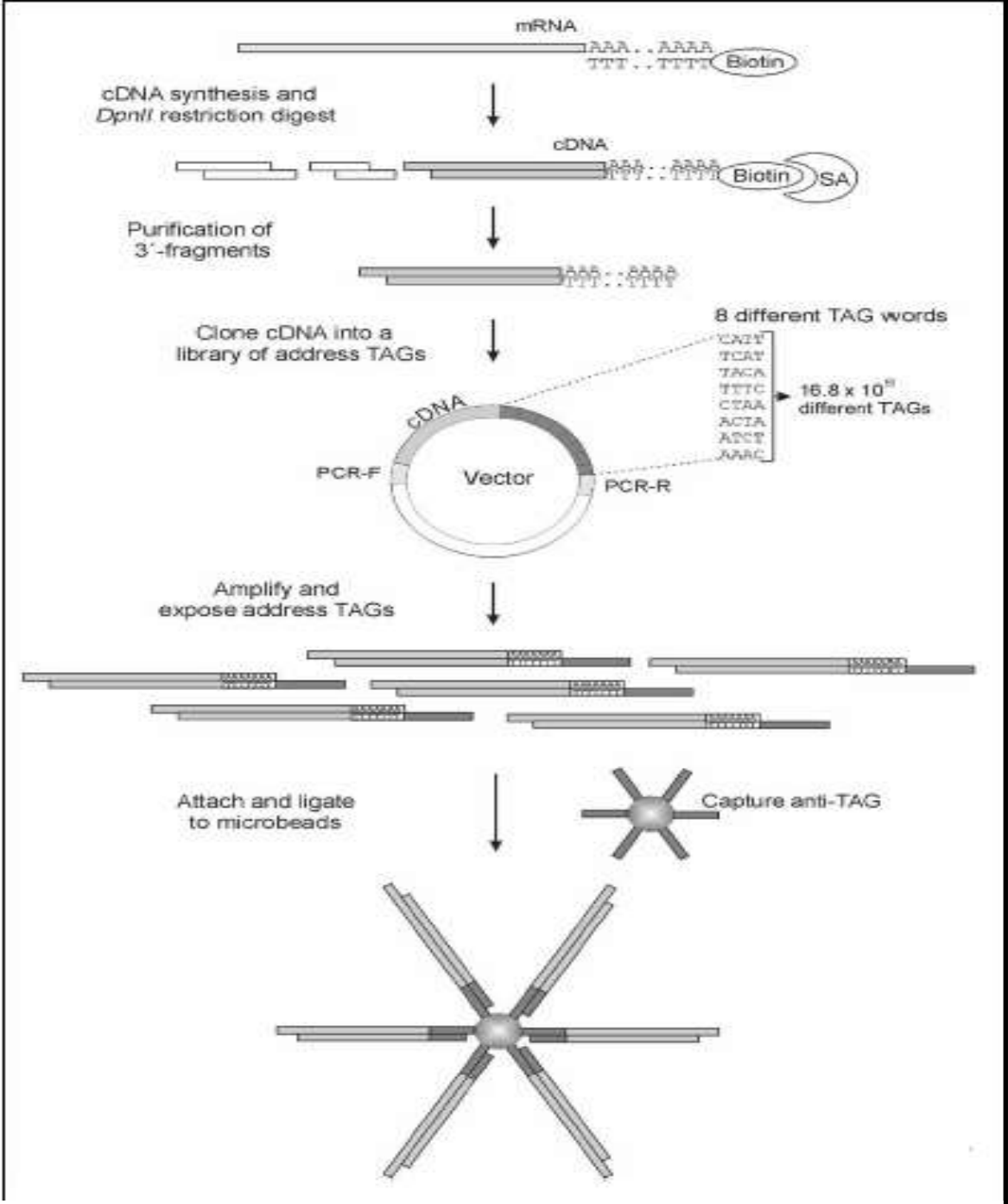
This 17 base sequence is used as mRNA identification signature.

To measure the level of expression of any given gene, the total number of signatures for that gene’s mRNA is counted.

Principle of MPSS

A sample's mRNA are first converted to cDNA using reverse transcriptase, which are fused to a small oligonucleotide "tag" which allows the cDNA to be PCR amplified and then coupled to microbeads. After several rounds of sequence determination, using hybridization of fluorescent labeled probes, a sequence signature of ~16-20 bp is determined from each bead.

Fluorescent imaging captures the signal from all of the beads, so DNA sequences are determined from all the beads in parallel, approximately 1,000,000 sequence reads are obtained per experiment.

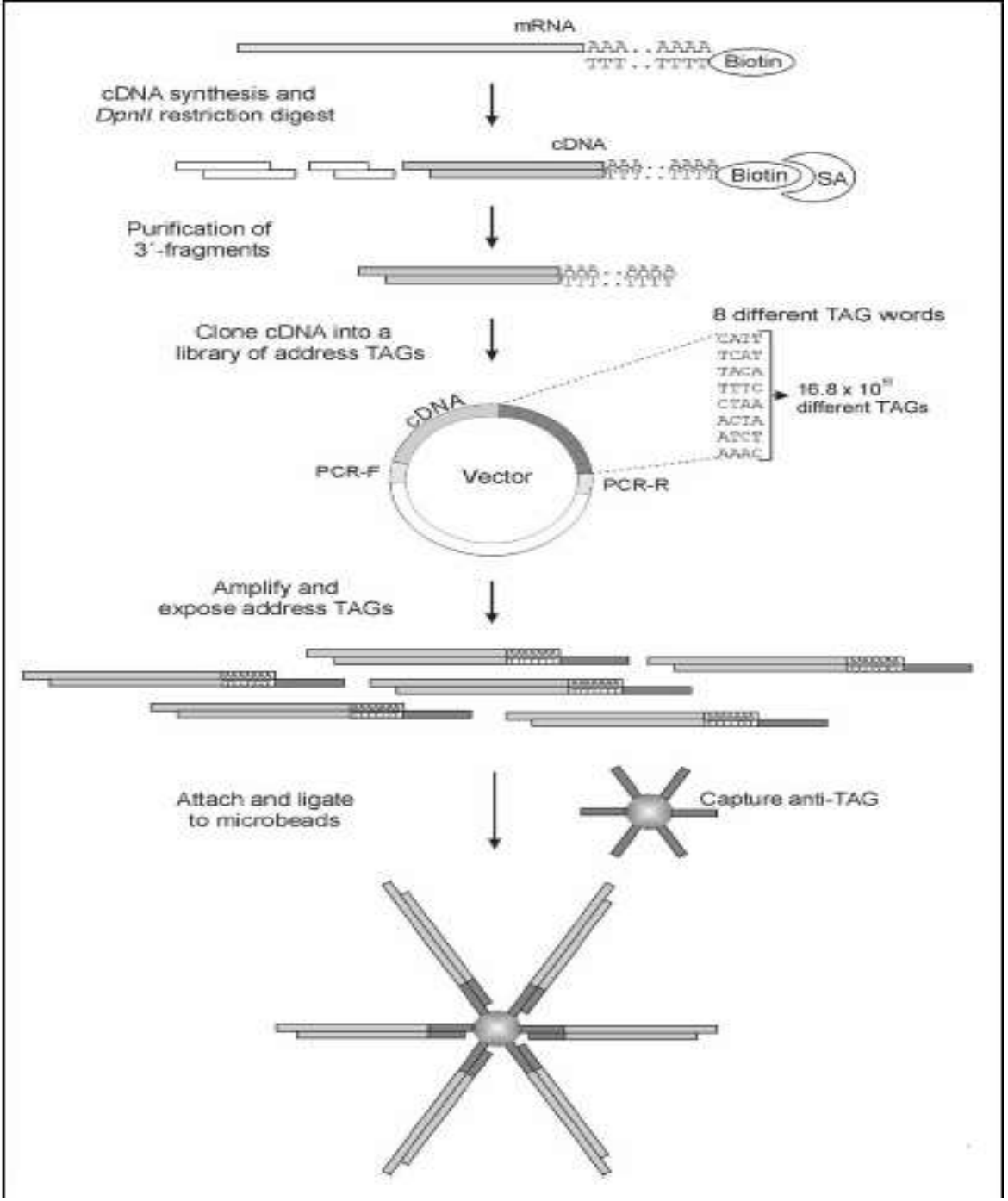


Procedure of Cloning and Sequencing cDNA Fragments on Beads

MPSS signatures for mRNAs in a sample are generated by sequencing ds cDNA fragments cloned onto microbeads using the Lynx Megaclone technology

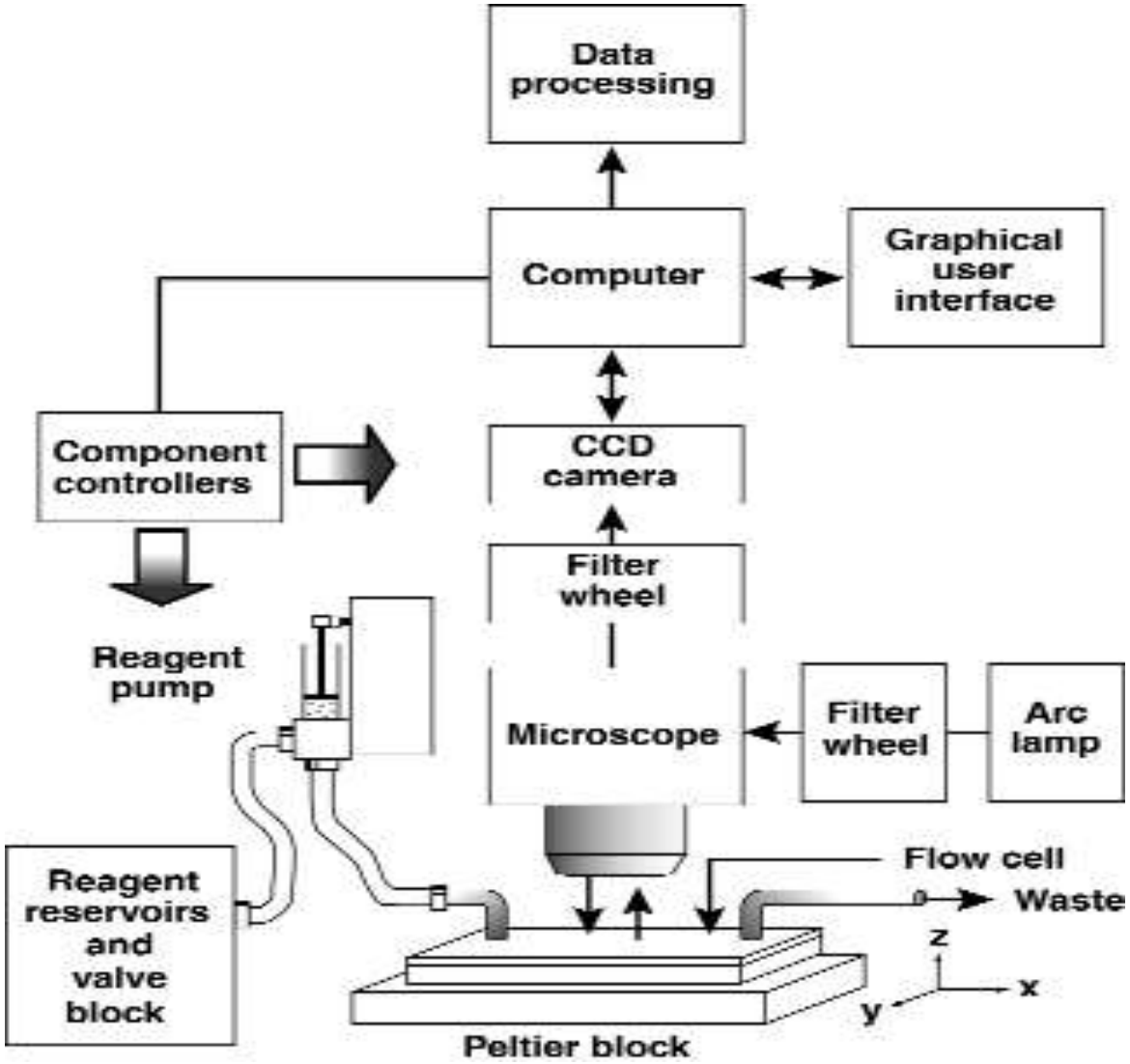
- Poly (A) mRNA molecules are converted into double-stranded cDNA molecules using biotynalated oligo dT primer. Streptavidin is use to purify biotynalated cDNA.
- cDNA digested with DpnII
- cDNA fragments cloned into a specially designed plasmid vector containing a unique barcode tag.
- Total 16.8 x 10^6 million different 32-base sequences available in the reference tag library, and each cDNA clone contains a different sequence.
- The library of cDNA inserts with oligonucleotide tags are PCR- amplified.
- The resulting linear molecules are partially treated with an exonuclease to

make the 32-base tag single stranded.



The 32-base tags at the end of each of the cDNA molecules are hybridised to 32-base complementary tags of microbeads.

The end-product is a microbead with approximately 100,000 identical cDNA molecules covalently attached to the surface.

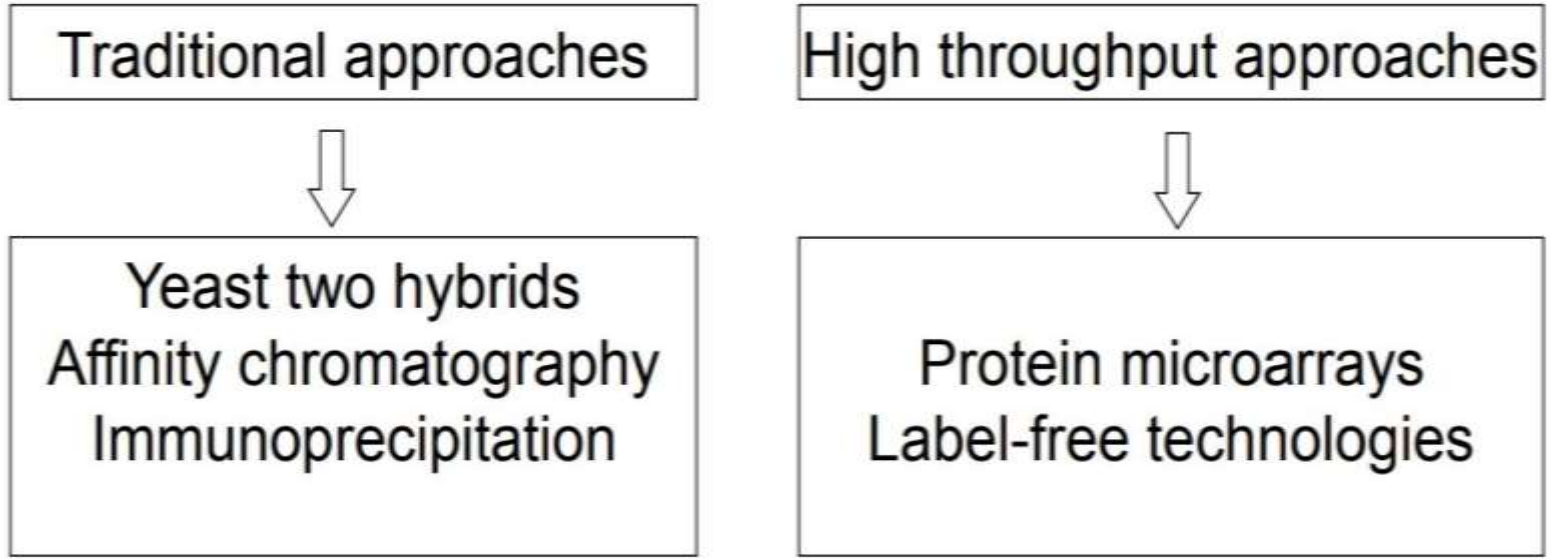


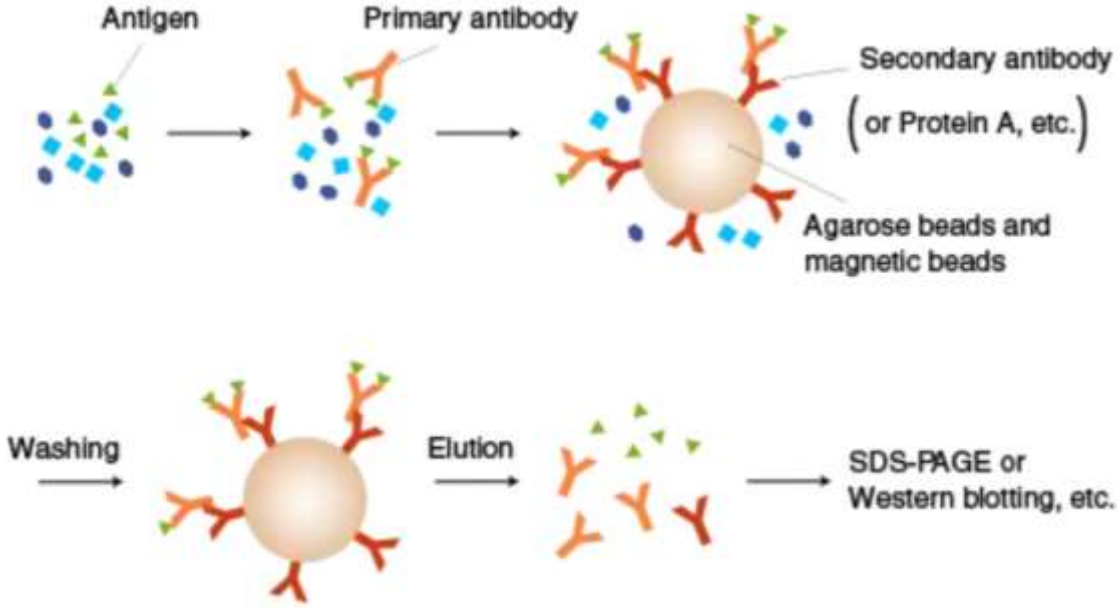
MPSS system

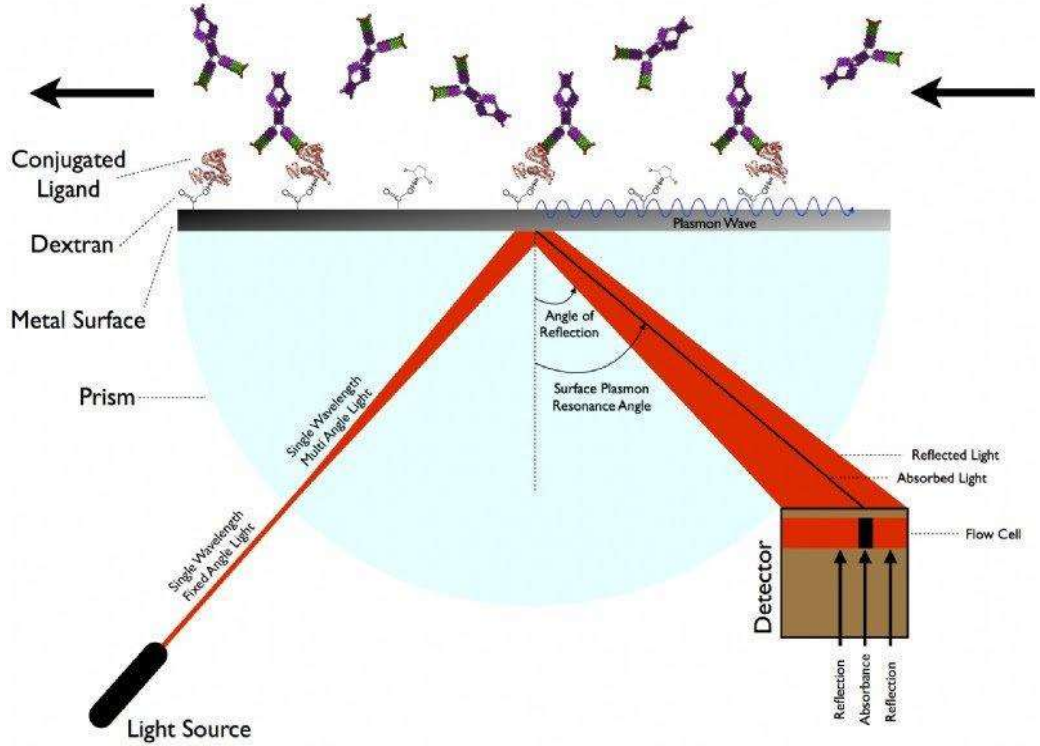
A high-resolution CCD camera is positioned directly over the flow-cell in order to capture fluorescent images from the microbeads at specific stages of the sequencing reactions.

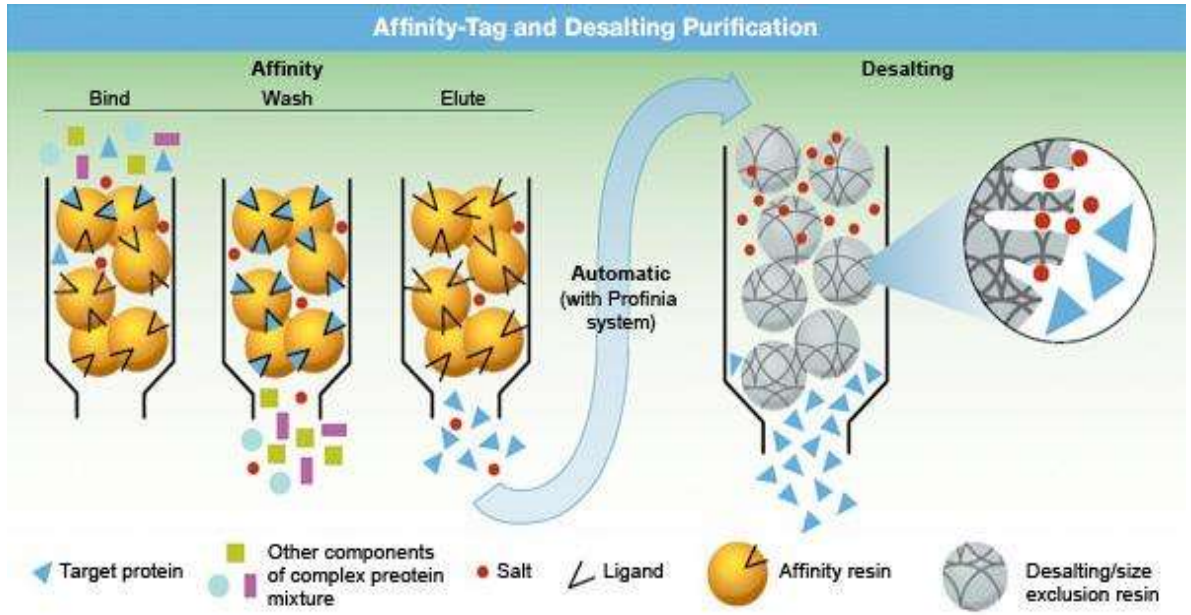
Protein Level Expression Profiling

Experimental methods to map Protein Protein Interaction (interactome):



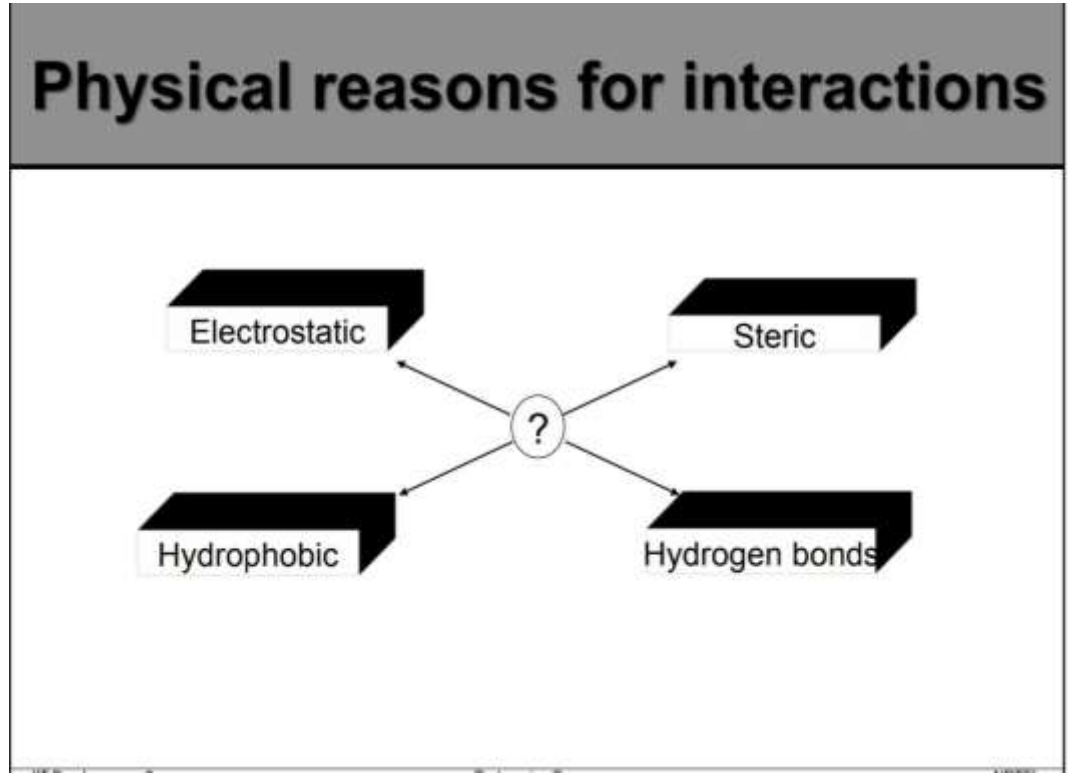

Antigen
Primary antibody
Secondary antibody
(or Protein A, etc.)
Agarose beads and
magnetic beads
Washing
Elution
SDS-PAGE or
Western blotting, etc.


Conjugated
Ligand
Dextran
Metal Surface
Prism
Plasmon Wave
Angle of
Reflection
Surface Plasmon
Resonance Angle
Single Wavelength
Multi Angle Light
Single Wavelength
Fixed Angle Light
Reflected Light
Absorbed Light
Flow Cell
Detector
Reflection
Absorbance
Reflection
Light Source

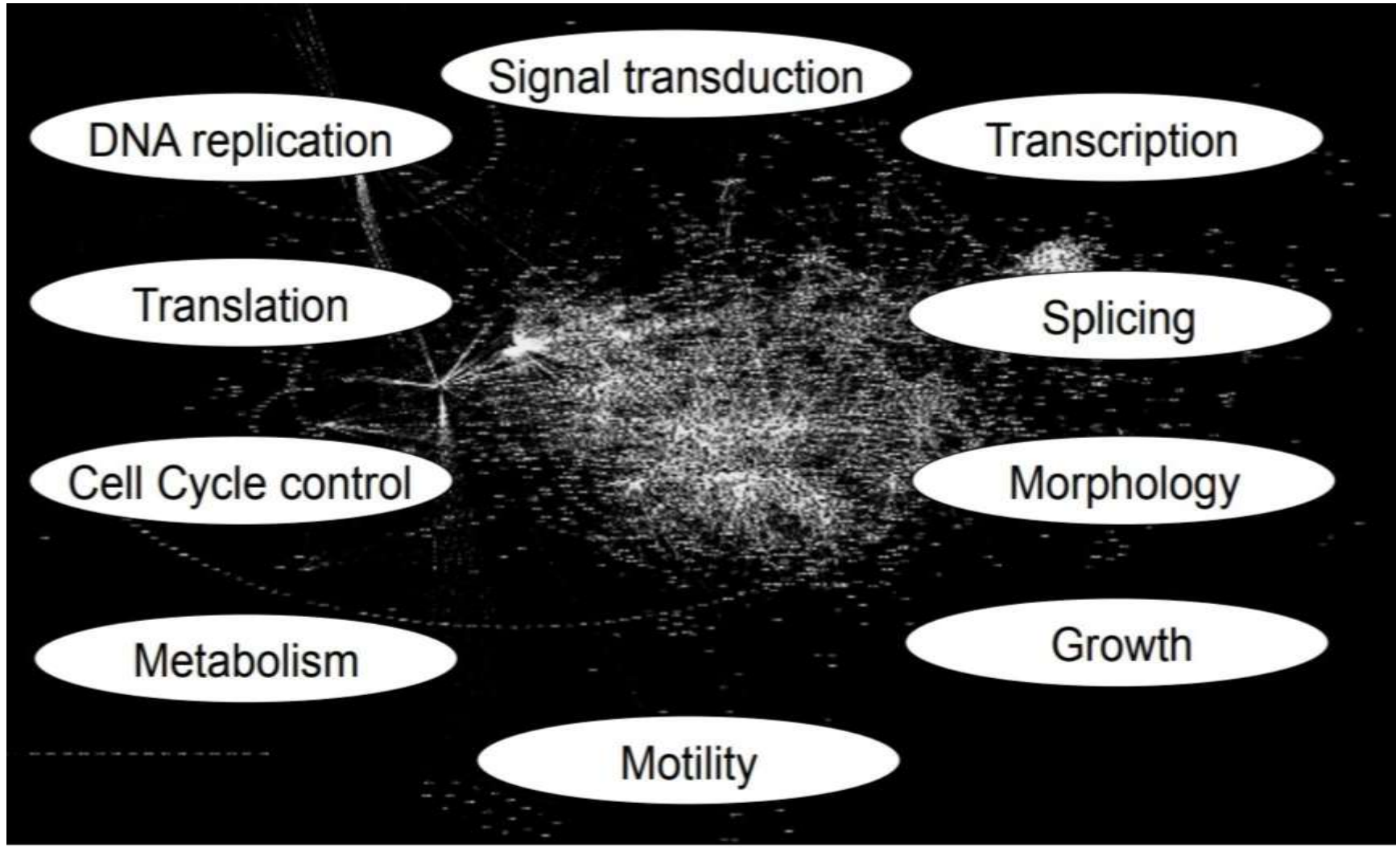

Affinity-Tag and Desalting Purification
Affinity
Bind
Wash
Elute
Automatic
(with Profinia
system)
Desalting
Target protein
Other components
of complex preotein
mixture
Salt
Ligand
Affinity resin
Desalting/size
exclusion resin

Protein protein interactions

Protein–protein interactions (**PPIs**) are the physical contacts of high specificity established between two or more protein molecules as a result of biochemical events steered by electrostatic forces including the hydrophobic effect.



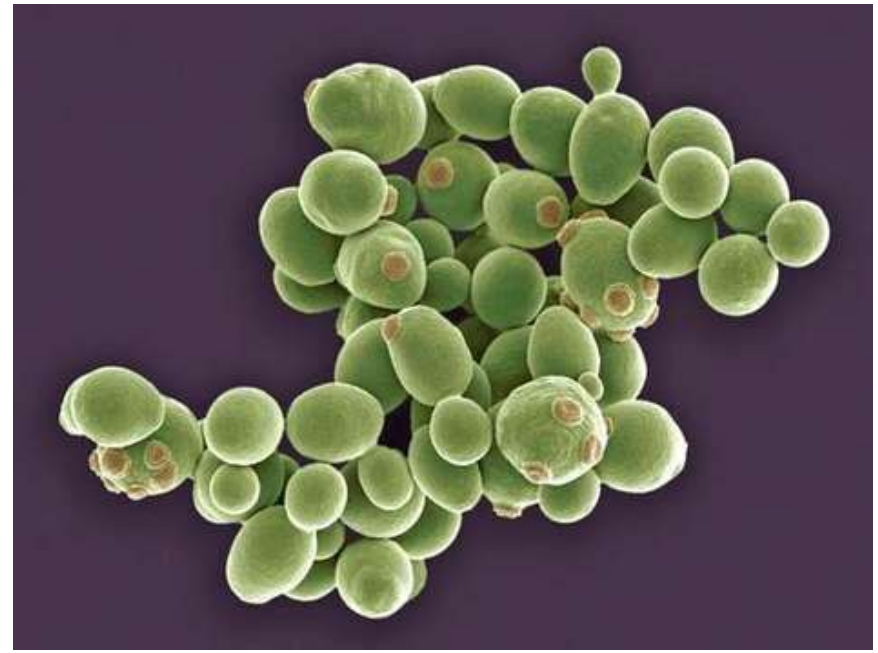
Essentiality of protein interactions:



YEAST TWO-HYBRID SYSTEM

Two hybrid Screening:

Two-hybrid screening (originally known as **yeast two-hybrid system** or **Y2H**) is a molecular biology technique used to discover protein–protein interactions (PPIs) and protein–DNA interactions by testing for physical interactions (such as binding) between two proteins or a single protein and a DNA molecule, respectively.



Need of yeast two hybrid system:

Identifies novel protein-protein interactions
Can identify protein cascades
Identifies mutations that affect protein-protein binding
Can identify interfering proteins in known interactions (Reverse Two-Hybrid System)

Yeast two- hybrid system:

Yeast two-hybrid is based on the reconstitution of a functional transcription factor (TF) when two proteins or polypeptides of interest interact.

In 1989, Fields and Song revolutionized protein interaction analysis by describing a genetic system to detect direct protein-protein interactions in the yeast *Saccharomyces cerevisiae*.

This takes place in genetically modified yeast strains, in which the transcription of a reporter gene leads to a specific phenotype, usually growth on a selective medium or change in the color of the yeast colonies.

The most popular reporter genes are HIS3 to select yeast on a medium lacking histidine, and LacZ to screen yeast in a colorimetric assay.

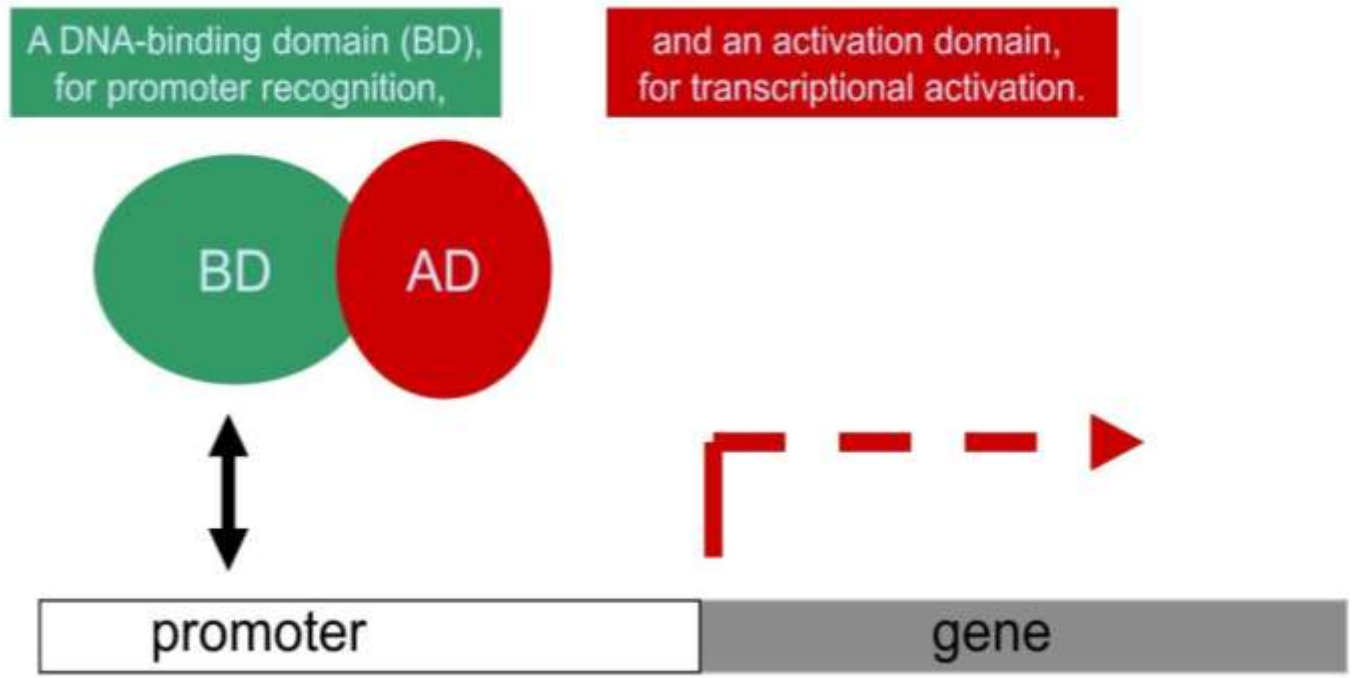
Two fusions (‘hybrids’) are constructed between each protein of interest and either the DNA Binding Domain (DBD) or the Activation Domain (AD) of the TF.

The protein fused to the DBD is referred to as the ‘bait’, and the protein fused to the AD as the ‘prey’.

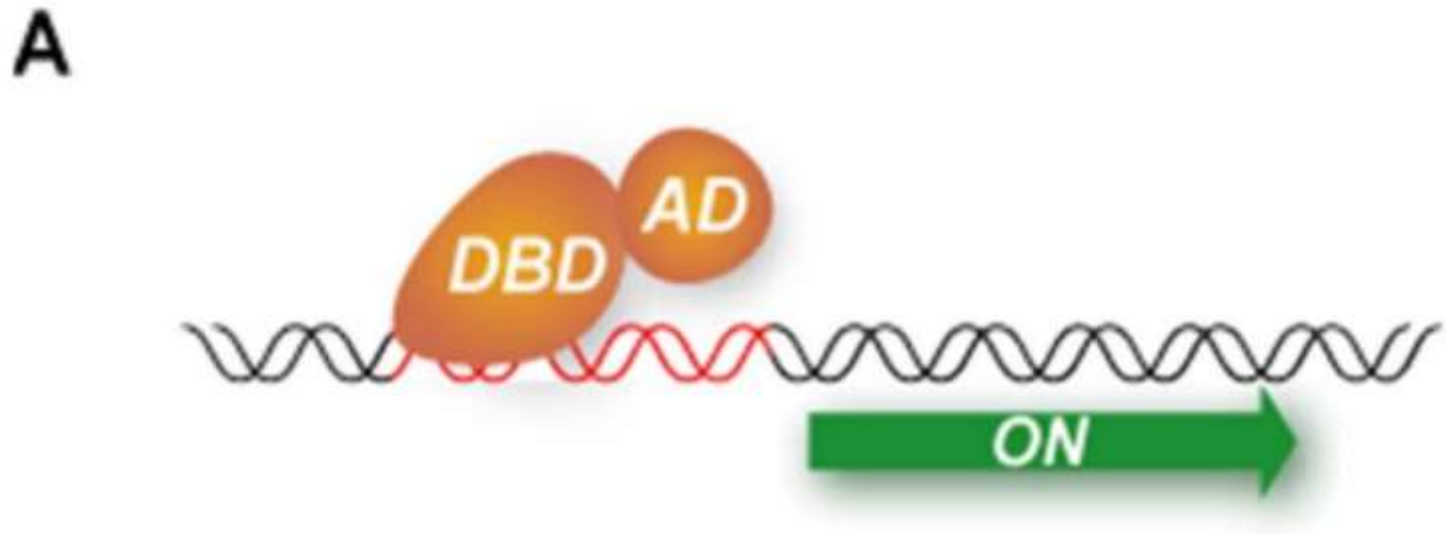
Upon interaction between the bait and the prey, the DBD and AD are brought in close proximity and a functional TF is reconstituted upstream of the reporter gene. The most popular fusions use the DBD and AD of the yeast TF Gal4. The bacterial protein LexA is also frequently used as a DBD in combination with Gal4 AD.

How this system works:

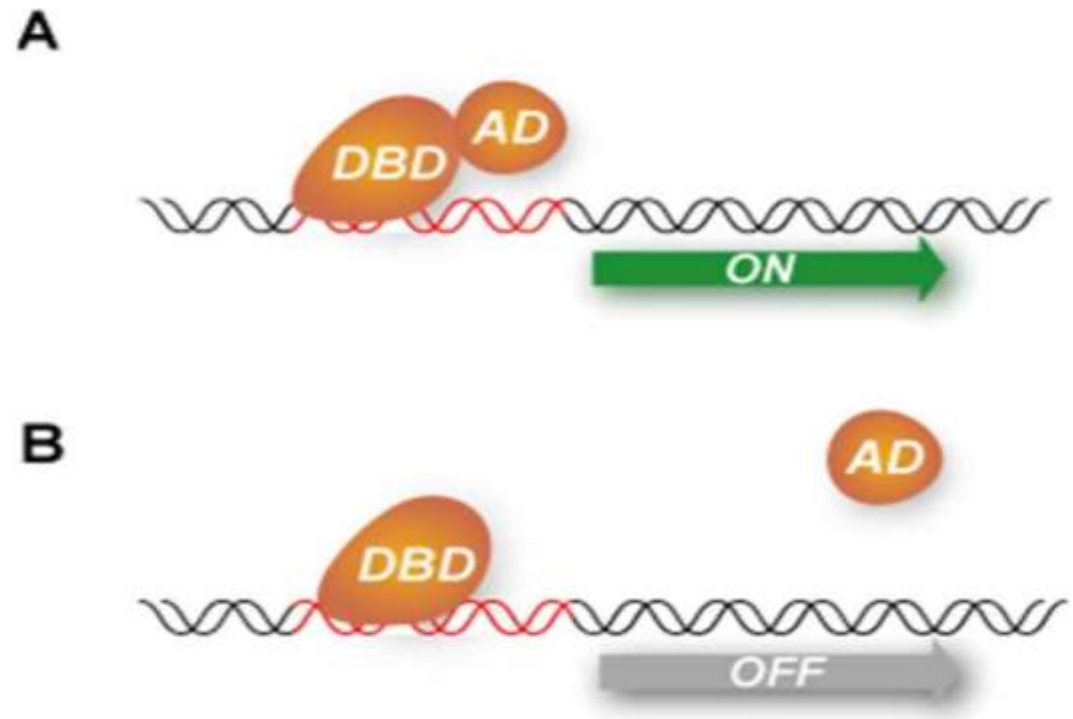
To look at the principle of yeast two hybrid Screening, let's forget about protein protein interactions for a minute, and look at a transcription factor. Many transcription factor have these 2 domains.



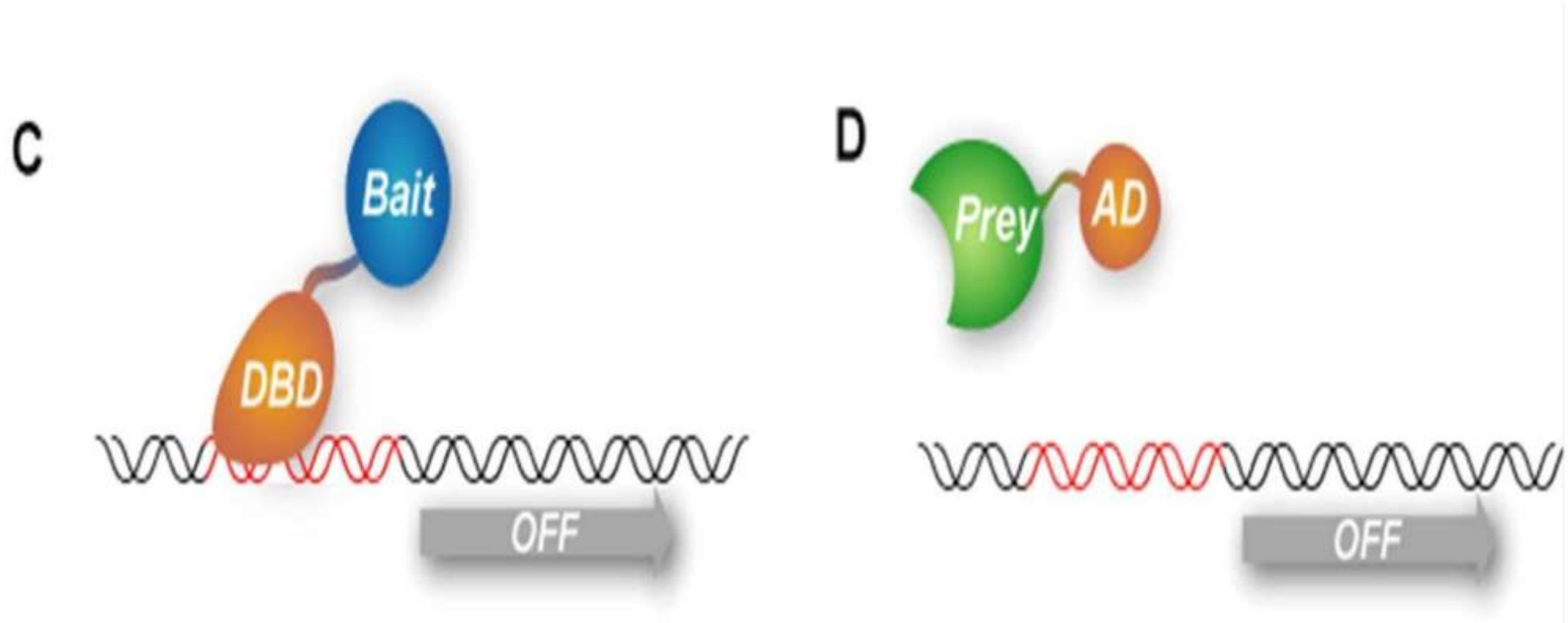
Transcription factor have a DNA binding domain and an activation domain.



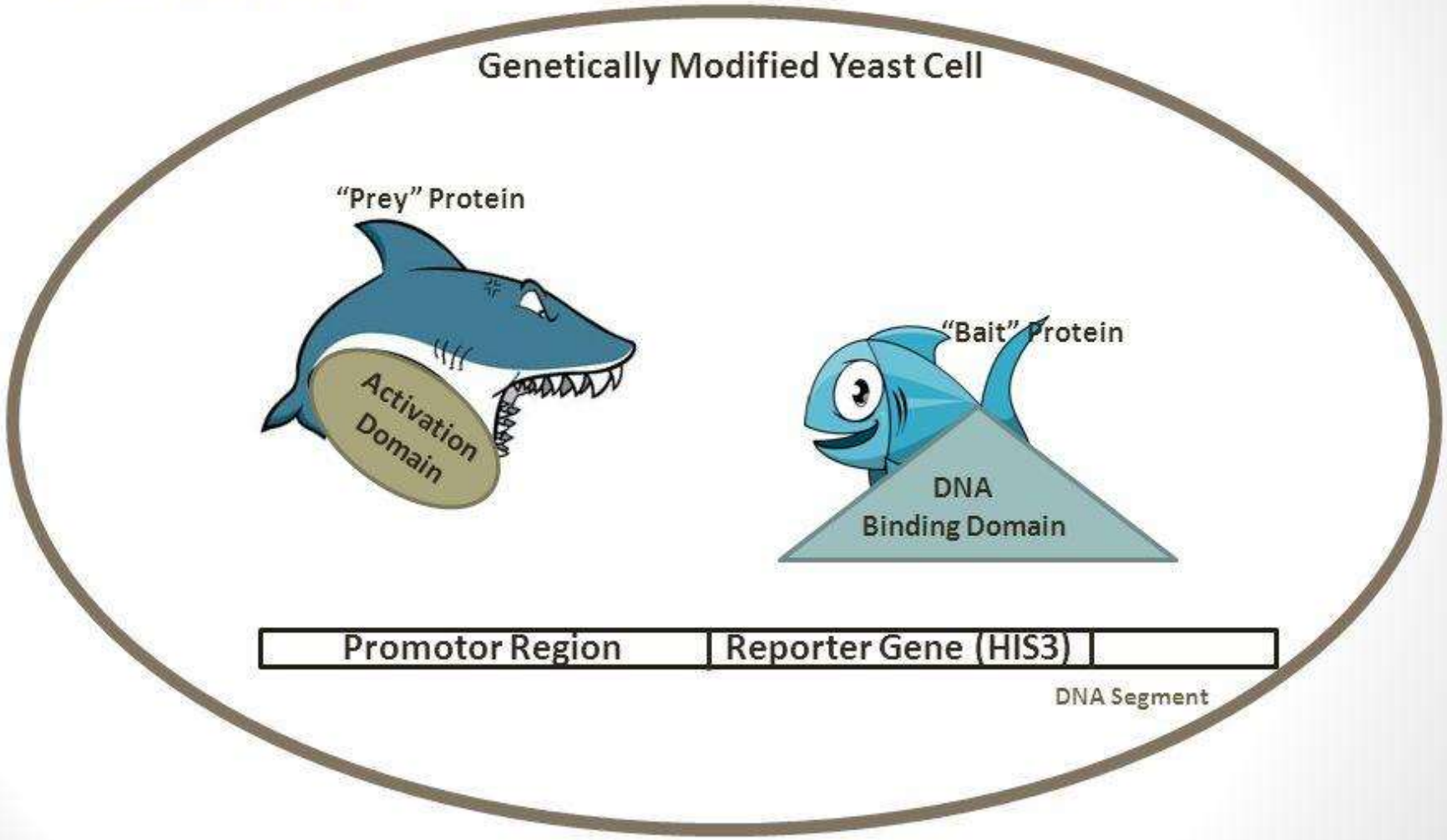
If DBD and AD are bounded then transcription is turned on but if due to any reason if they are unable to be in bound form then it will halt the process of transcription.



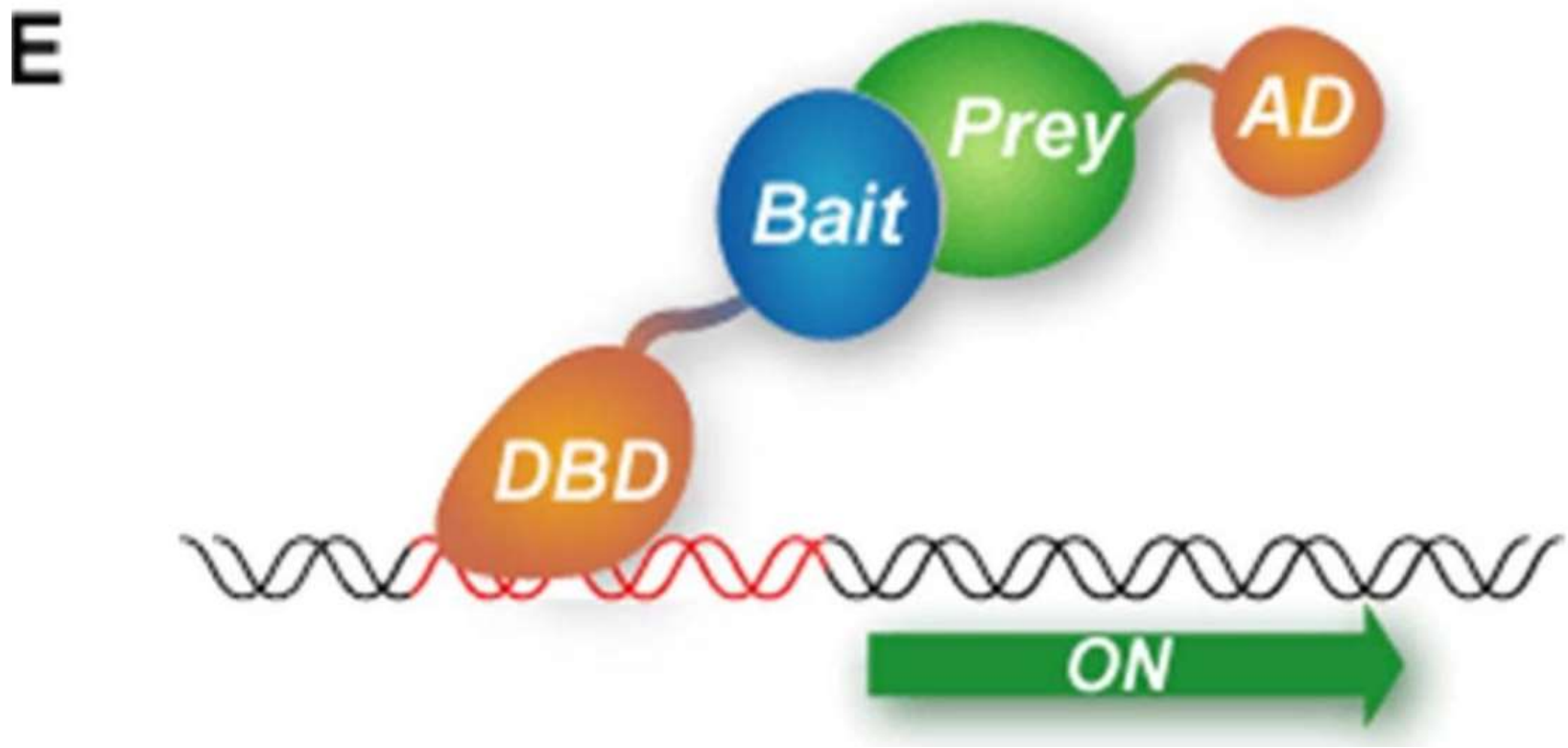
Bait (protein of interest); prey (other protein of interest or protein library:



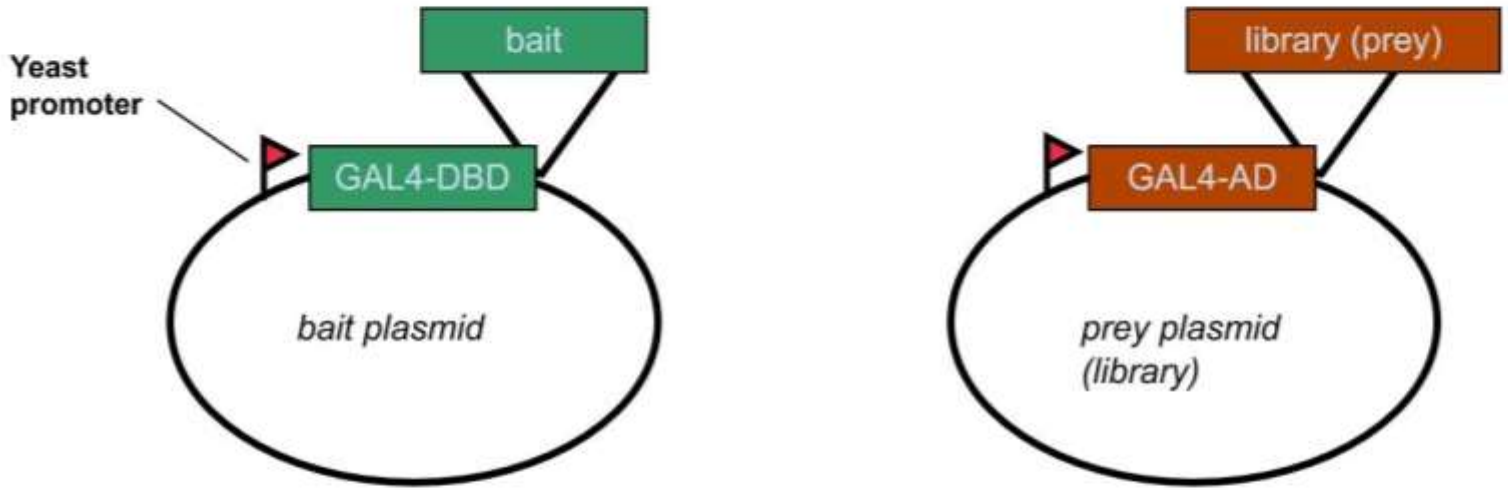
Yeast Two Hybrid (Y2H) Method



If they have any kind of interactions between them or complementarity they will bind and the process of transcription will turn on.



Bait and prey plasmids

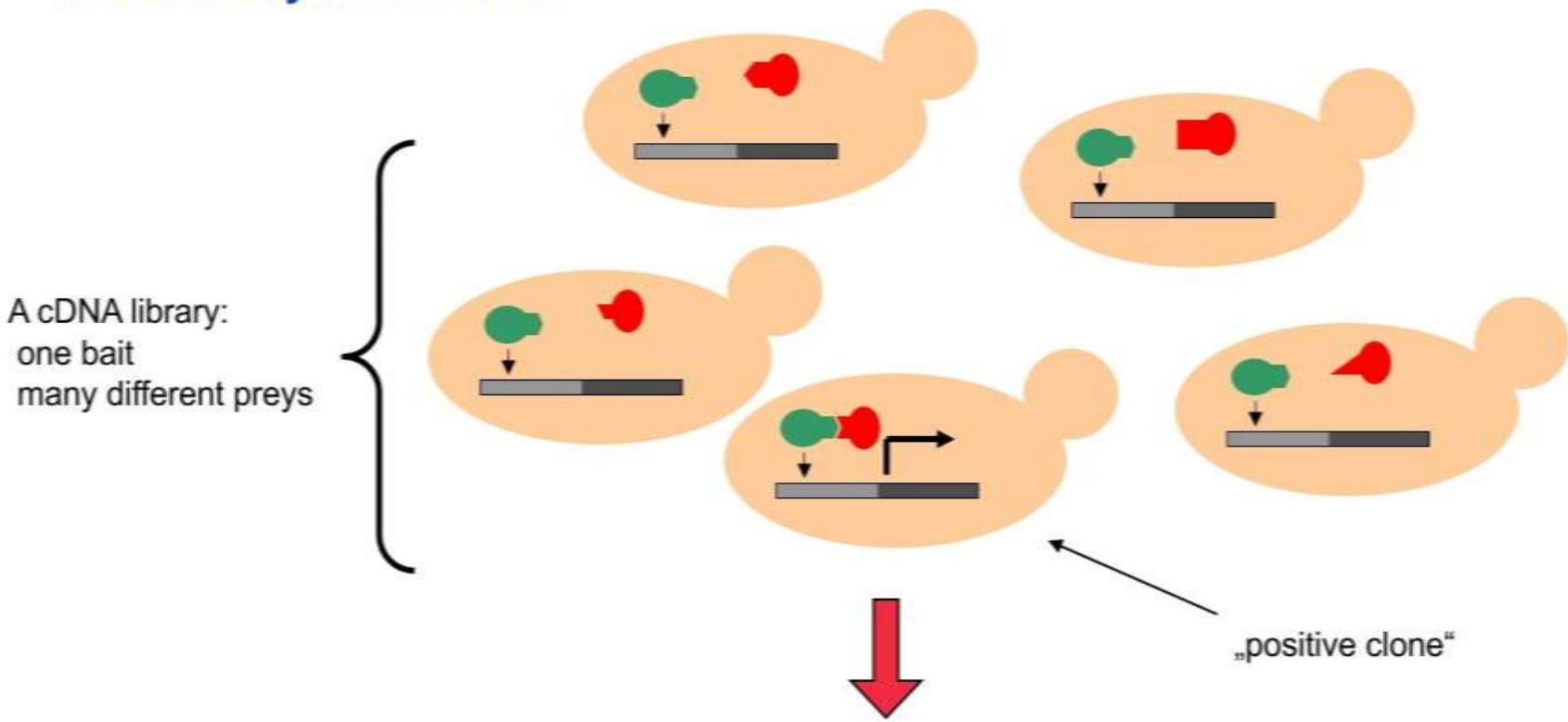


These plasmids express fusion proteins of ...

...your favorite gene with the DNA-binding domain of a transcription factor

....a cDNA library with the activation domain of a transcription factor

Yeast two-hybrid screens



Selection based on reporter gene expression
→ isolation of yeast cells harbouring a prey cDNA fragment coding for a protein that binds to the bait.

Yeast auxotrophies and their exploitation for transformation:

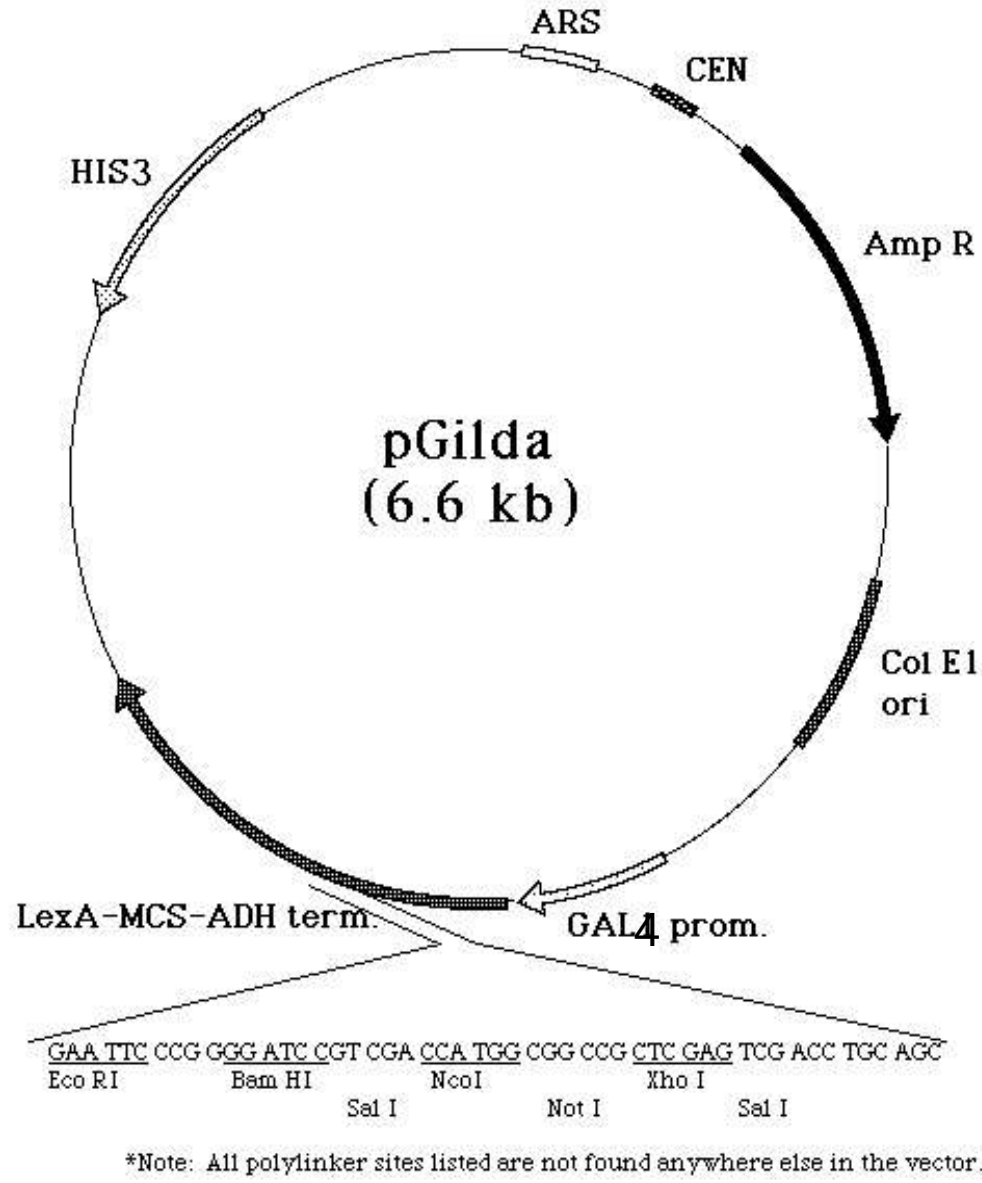
Wild type yeast does not need amino acids and nucleotides for growth.
Lab wild type yeast are auxotrophs for specific amino Acids and nucleotides.
For eg. Leu2Δ trp1Δ ura3Δ these cells do not grows if leucine, tryptophan and uracyl are not added to the medium.

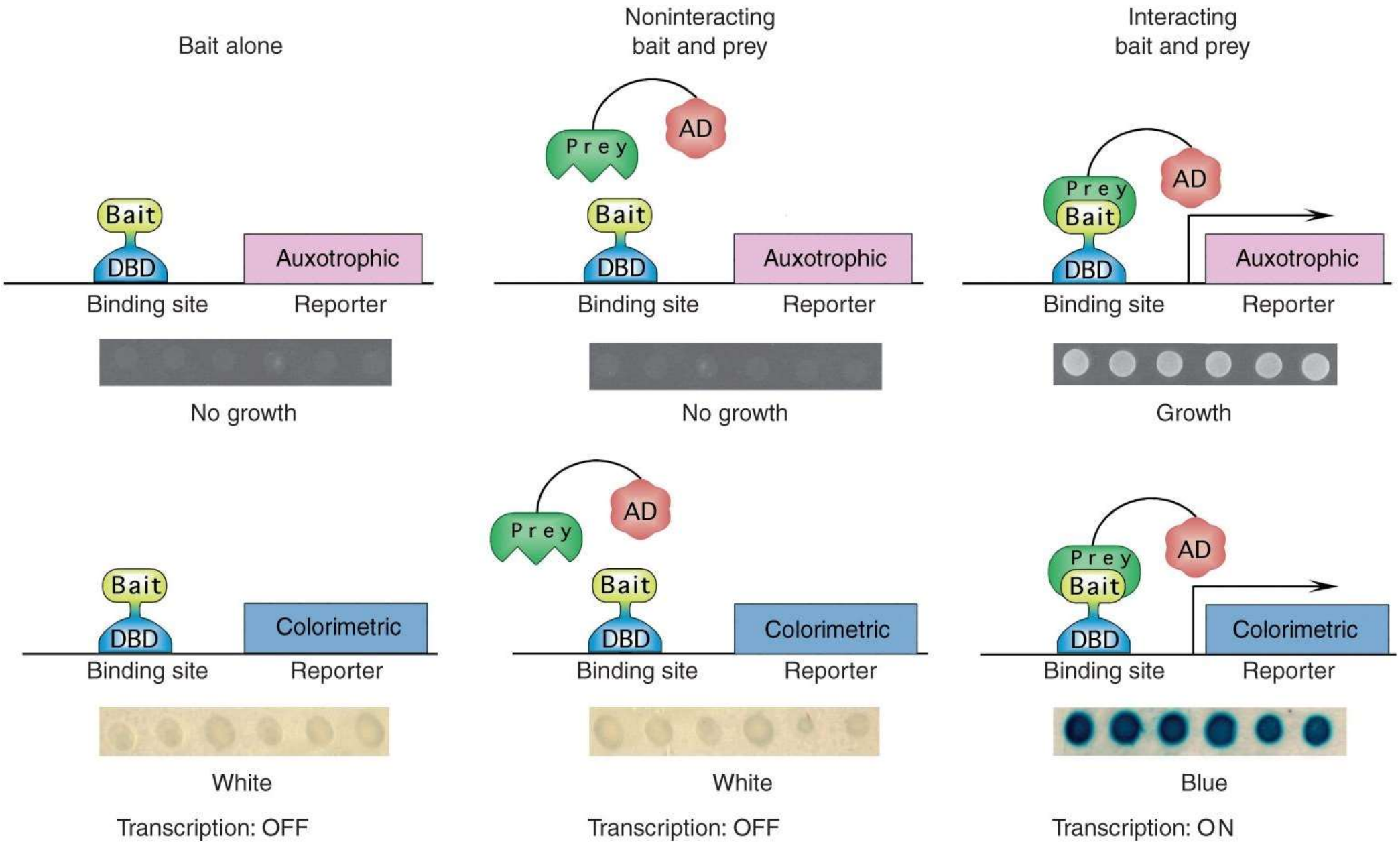
Reporter gene:

LacZ reporter - Blue/White Screening
HIS3 reporter - Screen on His+ media (usually need to add 3AT to increase selectivity)
LEU2 reporter - Screen on Leu+ media
ADE2 reporter - Screen on Ade+ media
URA3 reporter - Screen on Ura+ media (can do negative selection by adding FOA)

Yeast plasmid

ARS: autonomously replicating site
Amp(r): ampicillin resistant
Yeast HIS3 gene
CEN: centromere chromosome
MCS: multiple cloning site
ColE1 ori: ori for replication in bacterial cells.
GAL4 promoter: DNA binding domain

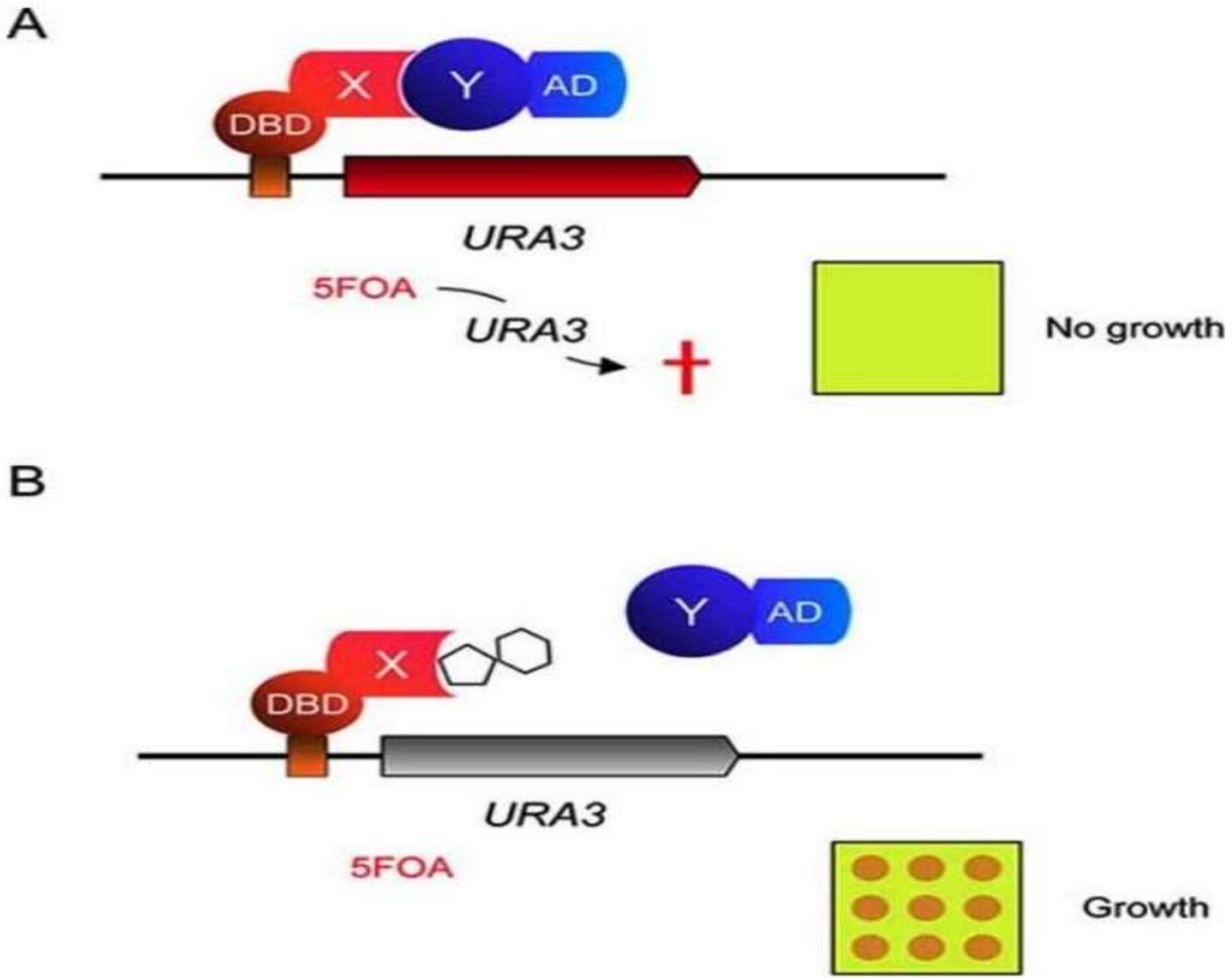



Bait alone
Noninteracting bait and prey
Interacting bait and prey
Prey
AD
Bait
DBD
Auxotrophic
Binding site
Reporter
No growth
No growth
Growth
Colorimetric
White
White
Blue
Transcription: OFF
Transcription: OFF
Transcription: ON

Reverse yeast two hybrid assay.

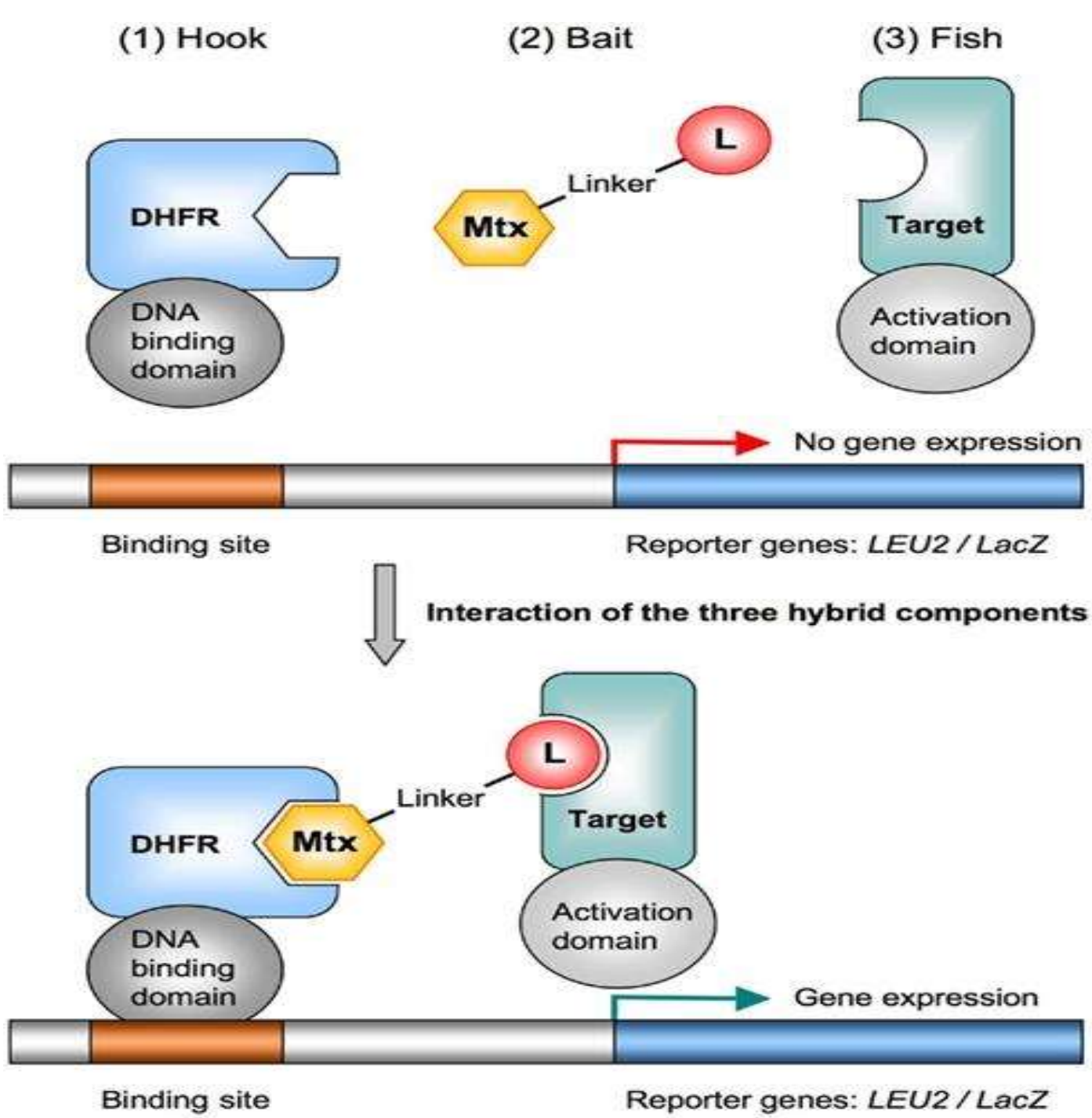
(A) Co-expression of an interacting DBD-X bait and an AD-Y prey activates expression of the URA3 gene, which converts 5FOA into a metabolite that is toxic to the cell. Consequently, yeast co-expressing the interaction partners will be unable to grow on 5FOA containing medium.
(B) Addition of a small molecule that disrupts the interaction between DBD-X and AD- Y abolishes the hybrid transcription factor and silences expression of URA3. Yeast which has been treated with this small molecule is therefore able to grow on 5FOA containing plates.

Reverse yeast two hybrid assay.



Yeast three hybrid system:

The yeast three-hybrid system for target identification. The principle components of the Y3H system are
(1) the hook, a fusion protein consisting of a DNA-binding domain and the Mtx-binding enzyme DHFR,
(2) the bait, a hybrid ligand of Mtx chemically linked to a small molecule ligand (L) of interest, and
(3) the fish, a transcriptional activation domain fused to a protein from a cDNA library (a potential target). The transcriptional activation of the reporter gene (LEU2 or lacZ) only occurs upon formation of the trimeric complex. To validate the Y3H system, the Mtx–Dex hybrid ligand can be used in combination with the high-affinity Dex-binding glucocorticoid receptor (GR) fused to the activation domain. DHFR, dihydrofolate reductase; Mtx, methotrexate.


(1) Hook
(2) Bait
(3) Fish
DHFR
DNA binding domain
L
Linker
Mtx
Target
Activation domain
No gene expression
Binding site
Reporter genes: LEU2 / LacZ
Interaction of the three hybrid components
DHFR
Mtx
Linker
L
Target
Activation domain
DNA binding domain
Gene expression
Binding site
Reporter genes: LEU2 / LacZ

Advantages of yeast two hybrid system:

An invivo technique.
Higher eukaryotic system.
Only the cDNA of the gene
Sensitive : weak and transient interactions.
Semi quantitatively : discrimination between high, intermediate, and low affinety binding.
Functional screens

PHAGE DISPLAY TECHNOLOGY

INTRODUCTION

➢Phage display technology has emerged in 1985 by George Smith.

Phage display is a laboratory technique for the study of protein–protein, protein–peptide, and protein–DNA interactions that uses bacteriophages to connect proteins with the genetic information that encodes them.

In this technique, a gene encoding a protein of interest is inserted into a phage coat protein gene, causing the phage to "display" the protein on its outside while containing the gene for the protein on its inside, resulting in a connection between genotype and phenotype.

In this way, large libraries of proteins can be screened and amplified in a process called *in vitro* selection, which is analogous to natural selection.

These displaying phages can then be screened against other proteins, peptides or DNA sequences, in order to detect interaction between the displayed protein and those other molecules.

The most common bacteriophages used in phage display are filamentous phages *viz.*, fd, f1, M13

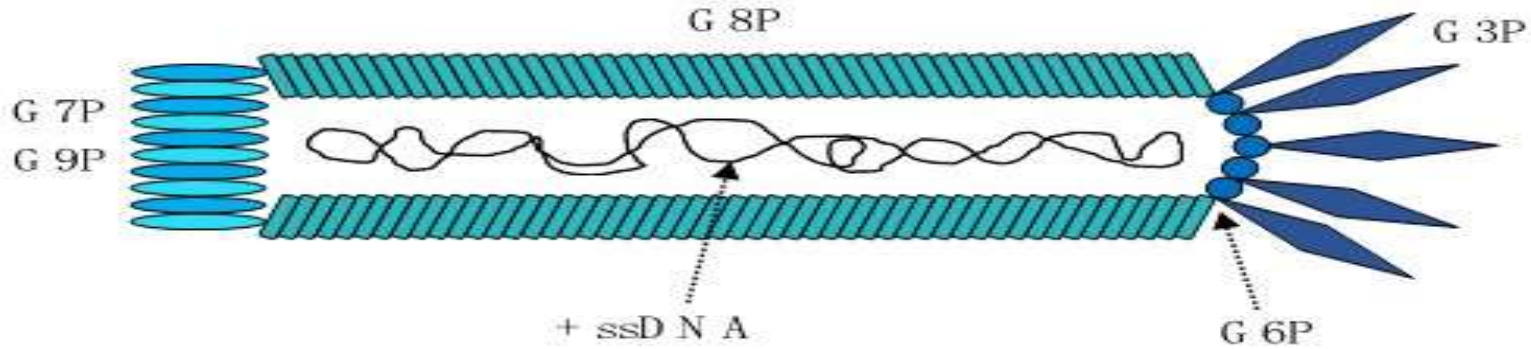


Figure 1. Schematic representation of the M13 bacteriophage, which is a filamentous phage carrying a single-stranded DNA (ssDNA) chromosome. The genome contains nine genes, which encode 11 proteins. Five of these proteins are coat proteins (G3P, G6P, G7P, G8P and G9P), while the remaining six proteins are used for replication of the genome, assembly of the phage, and phage extrusion.

Table 1. Gene name, protein name, protein size, and the function of the genes carried by the M13 phage [16].

Gene Name	Protein Name (Abbreviation)	Size (kDa)	Function
I	Gene 1 protein (G1P) Gene 11 protein (G11P)	39.6 12.4	Assembly Assembly
II	Replication-associated protein (G2P) Gene 10 protein (G10P)	46.2 12.7	Replication Replication
III	Attachment protein (G3P)	44.7	Coat protein Adsorption and extrusion
IV	Virion export protein (G4P)	45.9	Assembly and extrusion
V	DNA-binding protein (G5P)	9.7	Replication
VI	Head virion protein (G6P)	12.4	Coat protein Infection and budding
VII	Tail virion protein (G7P)	3.6	Coat protein Assembly and budding
VIII	Capsid protein (G8P)	7.6	Coat protein
IX	Tail virion protein (G9P)	3.7	Coat protein Assembly and budding

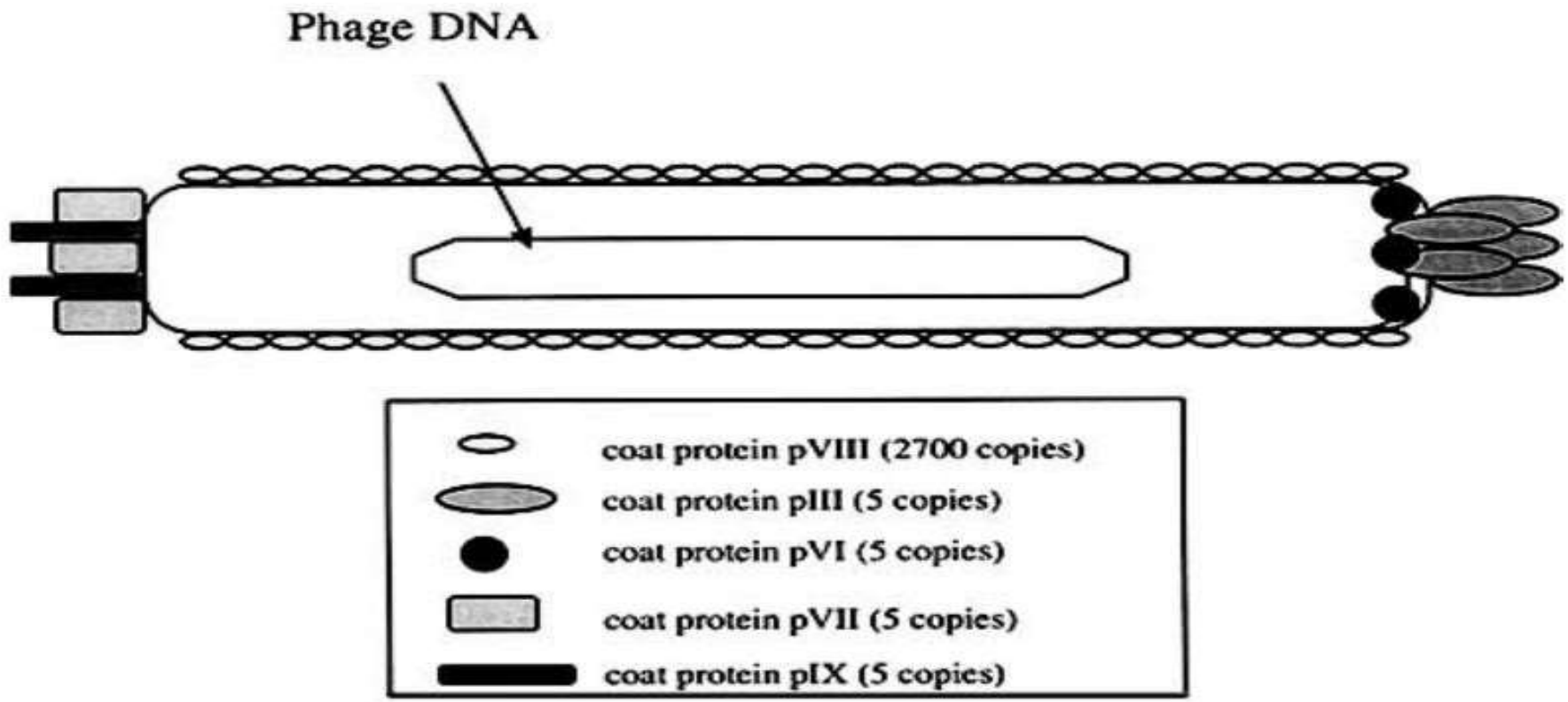
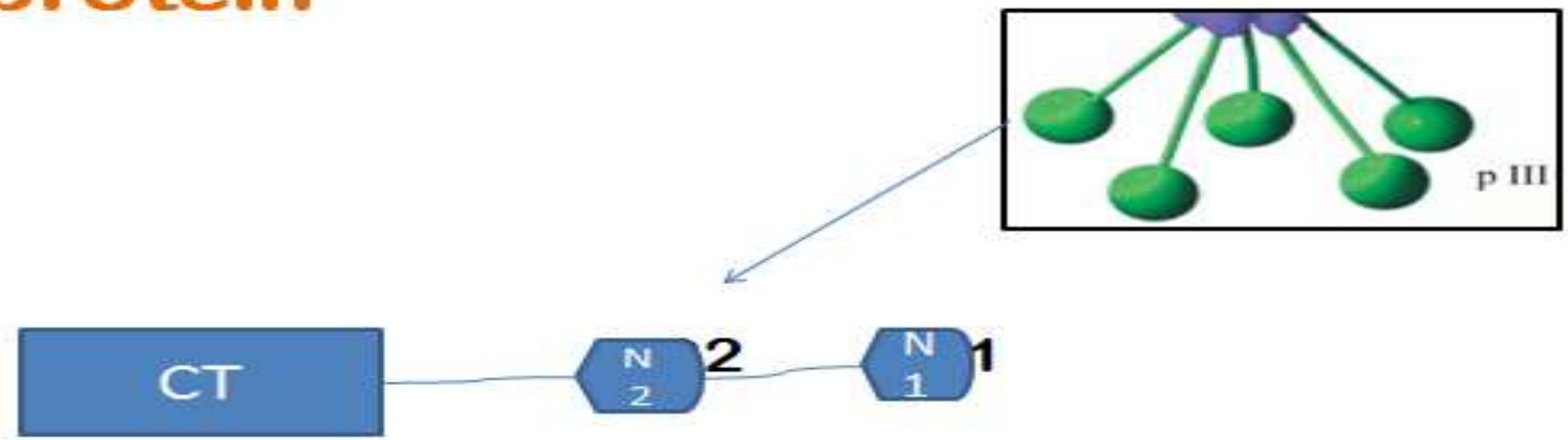


Fig. (1). Schematic representation of a filamentous phage.

Gene 3 protein



N1- interacts with TolA
N2- binds to F pilus

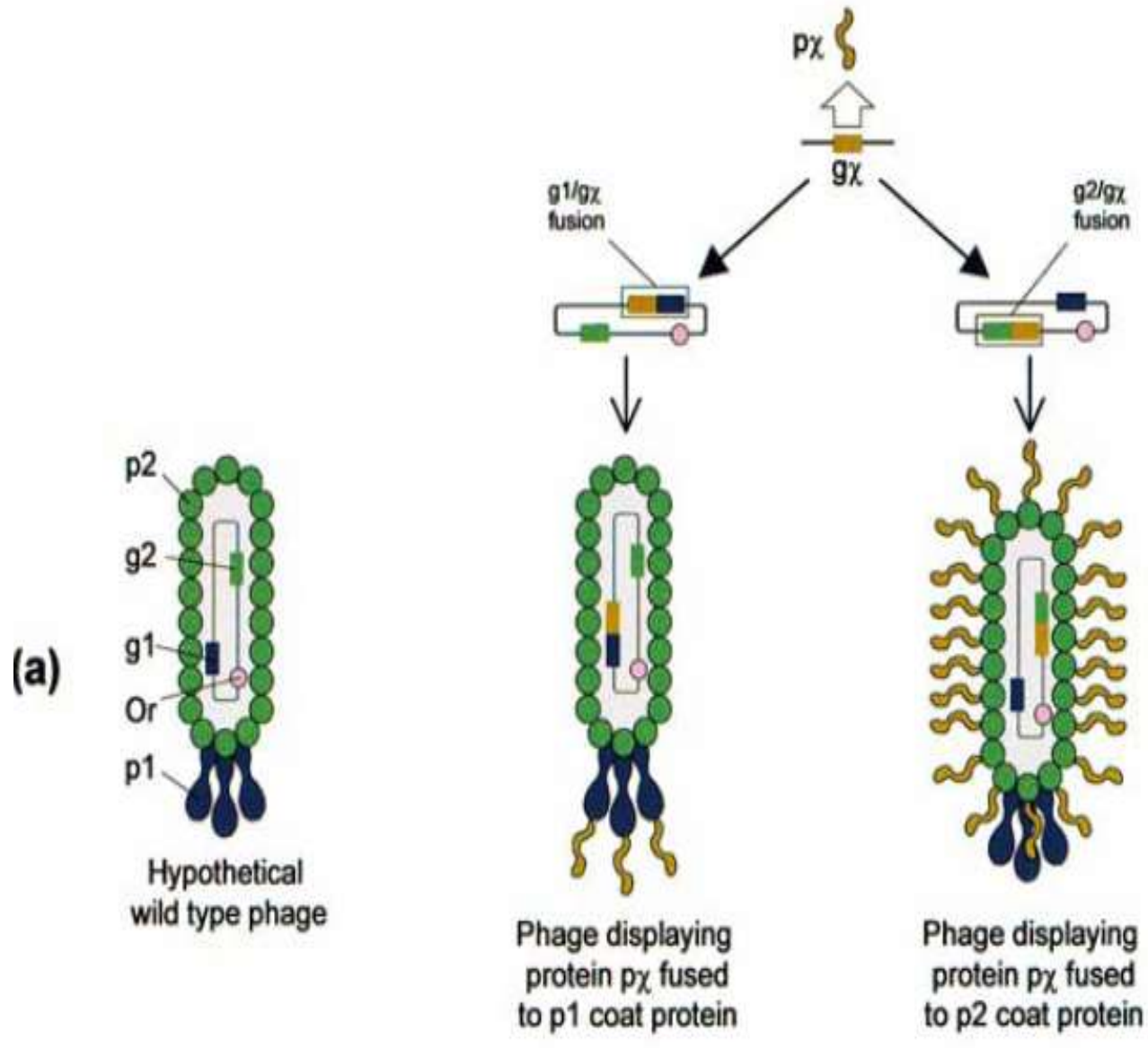
Fig. the modular structure of g3p

The principle of phage display

The DNA encoding the protein or peptide of interest is ligated into the pIII or pVIII gene, encoding either the minor or major coat protein, respectively.
The phage gene and insert DNA hybrid is then inserted (a process known as "transduction") into *Escherichia coli* (E. coli) bacterial cells.
By immobilizing a relevant DNA or protein target(s) to the surface of a micro titer plate well, a phage that displays a protein that binds to one of those targets on its surface will remain while others are removed by washing.
Those that remain can be eluted, used to produce more phage (by bacterial infection with helper phage) and so produce a phage mixture that is enriched with relevant (i.e. binding) phage.

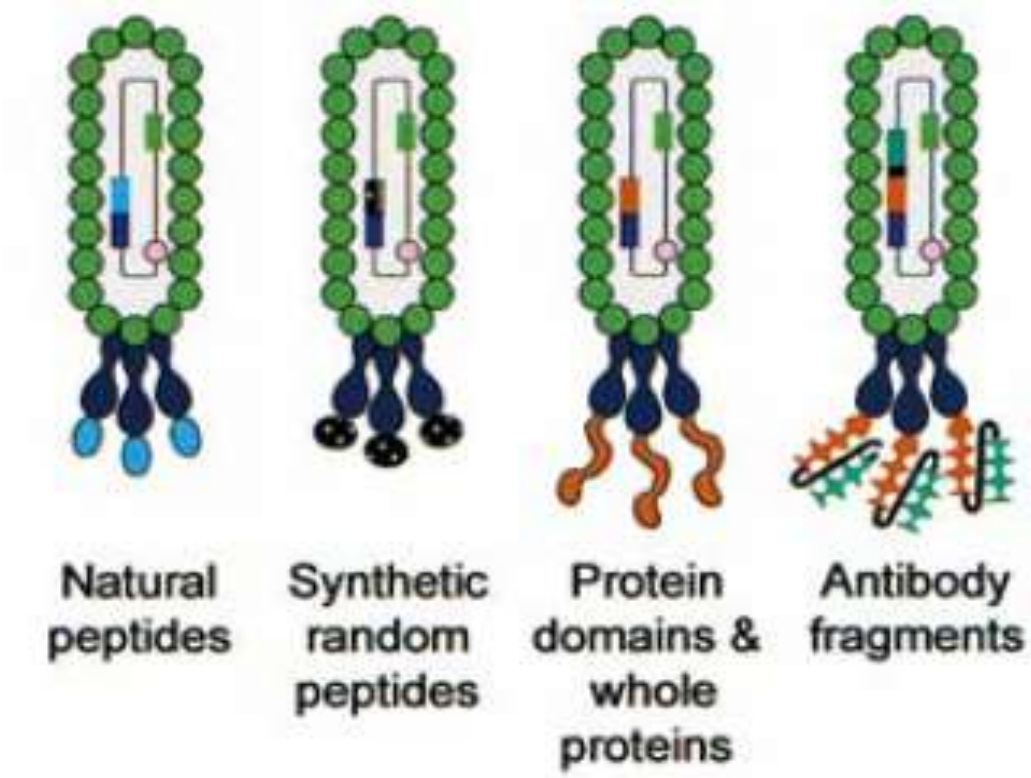
The repeated cycling of these steps is referred to as 'panning', in reference to the enrichment of a sample of gold by removing undesirable materials.

Phage eluted in the final step can be used to infect a suitable bacterial host, from which the phagemids can be collected and the relevant DNA sequence excised and sequenced to identify the relevant, interacting proteins or protein fragments.

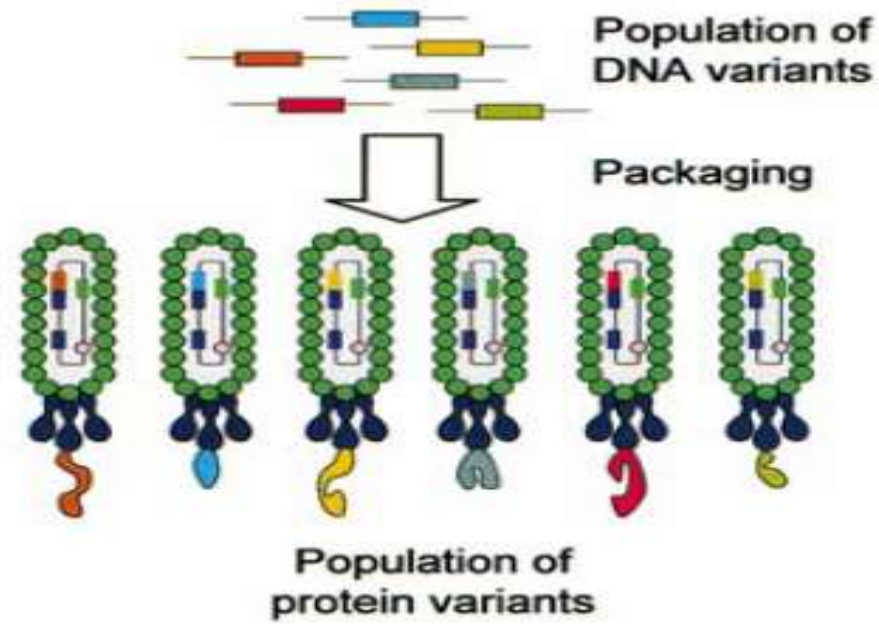


Continued..

(b)



(c)

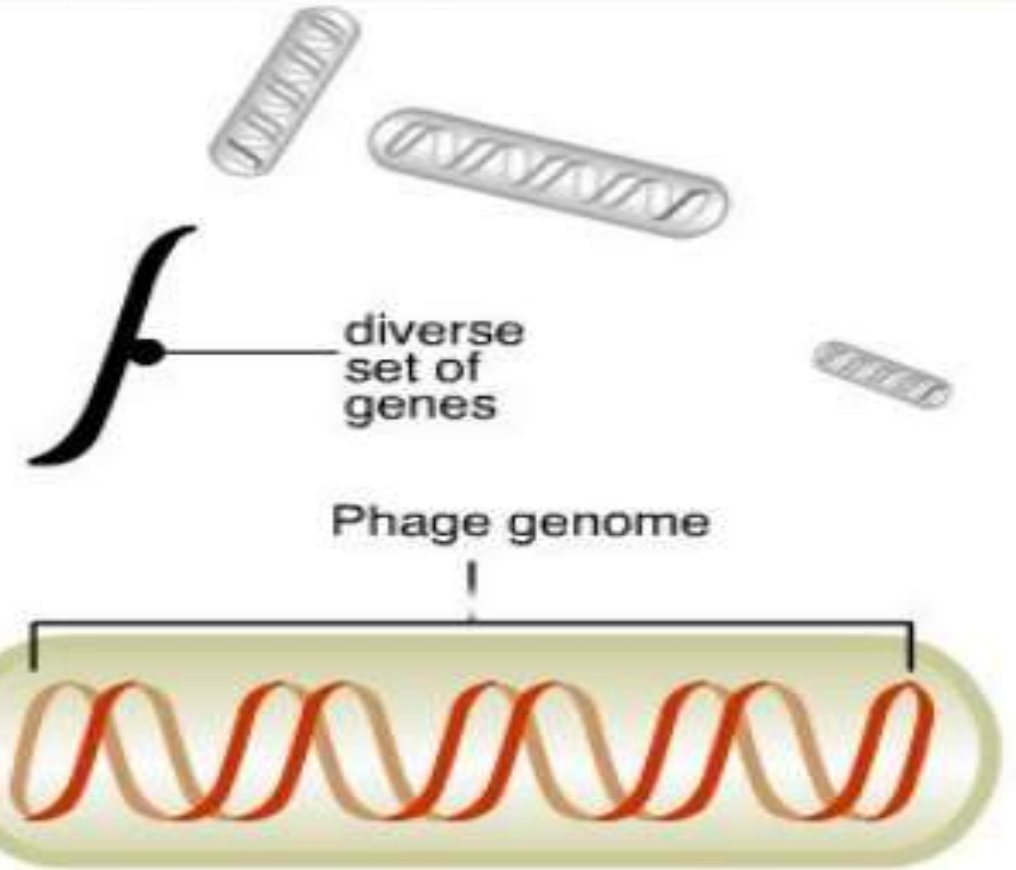


PHAGE DISPLAY PROCESS

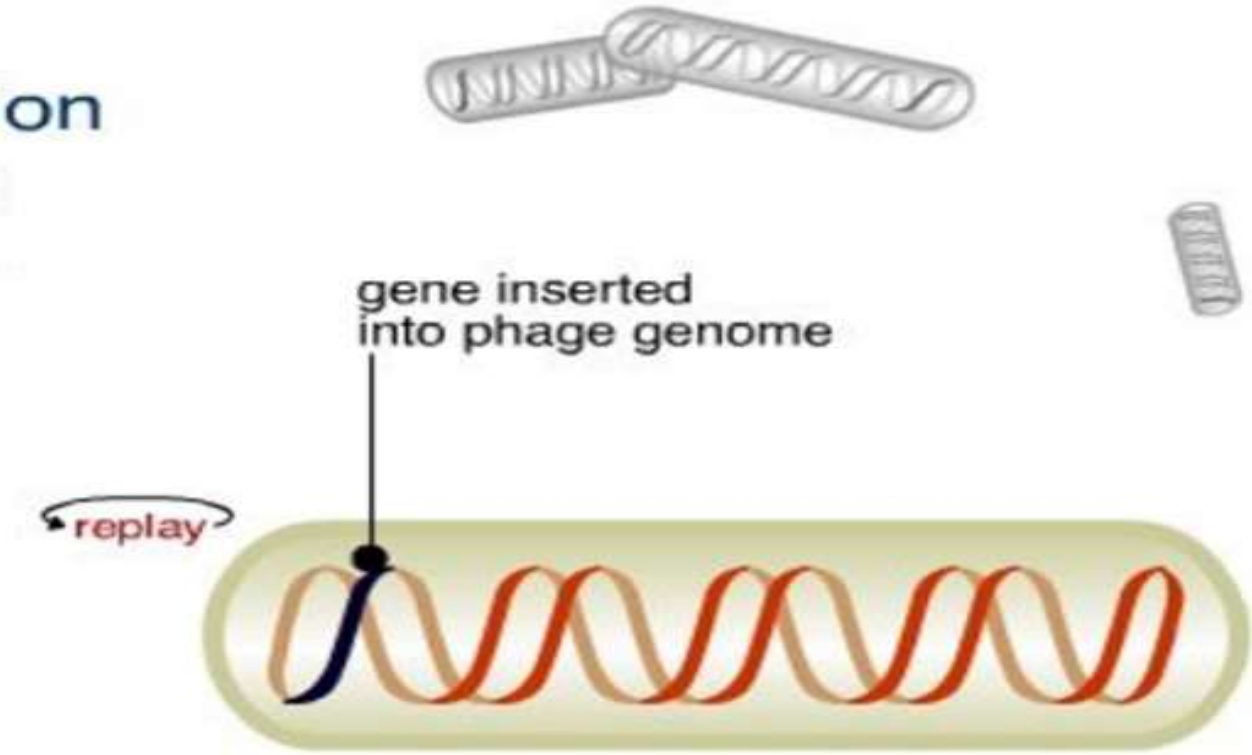
Step 1: Gene Insertion

Phage display begins by inserting a diverse set of genes into the phage genome.

Each phage receives a different gene.



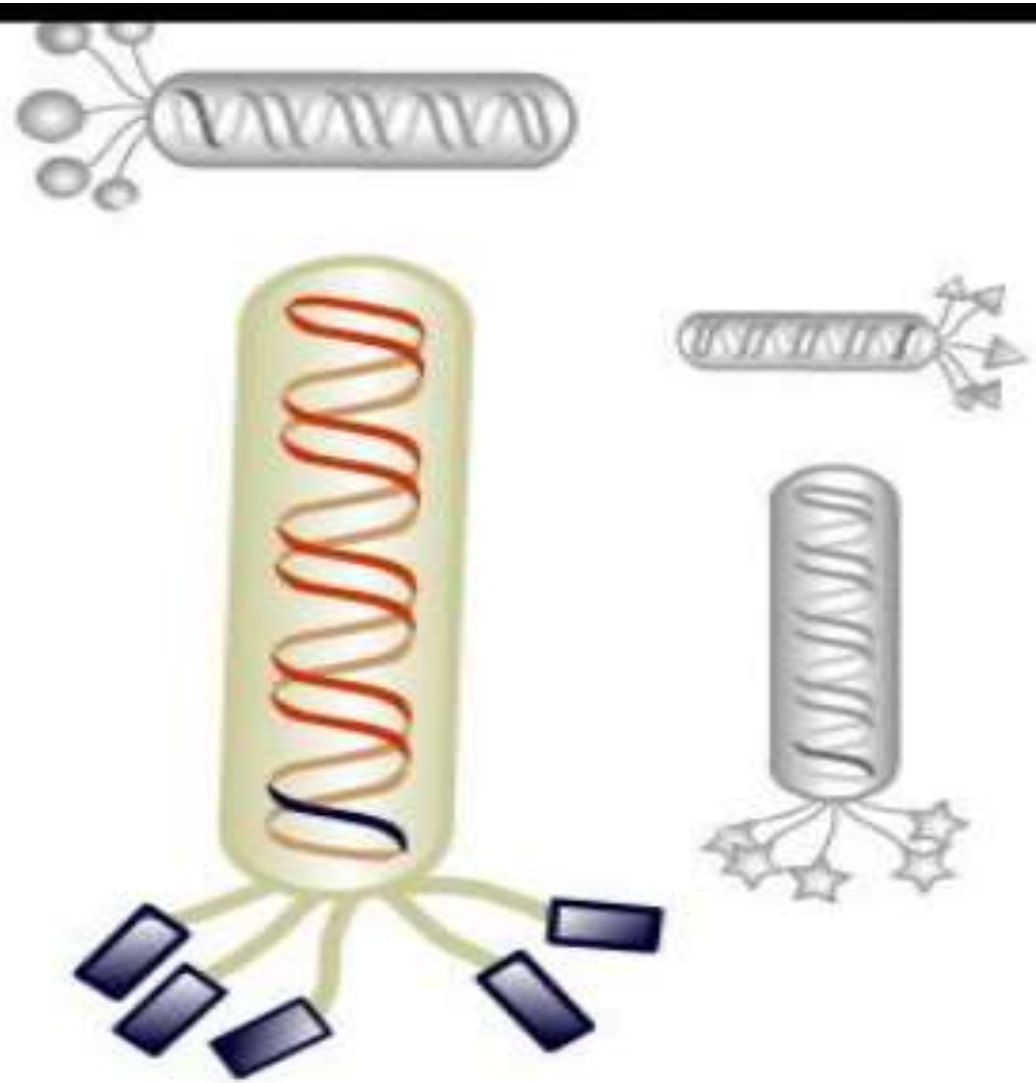
Step 1: Gene Insertion



Step 2: Protein Display

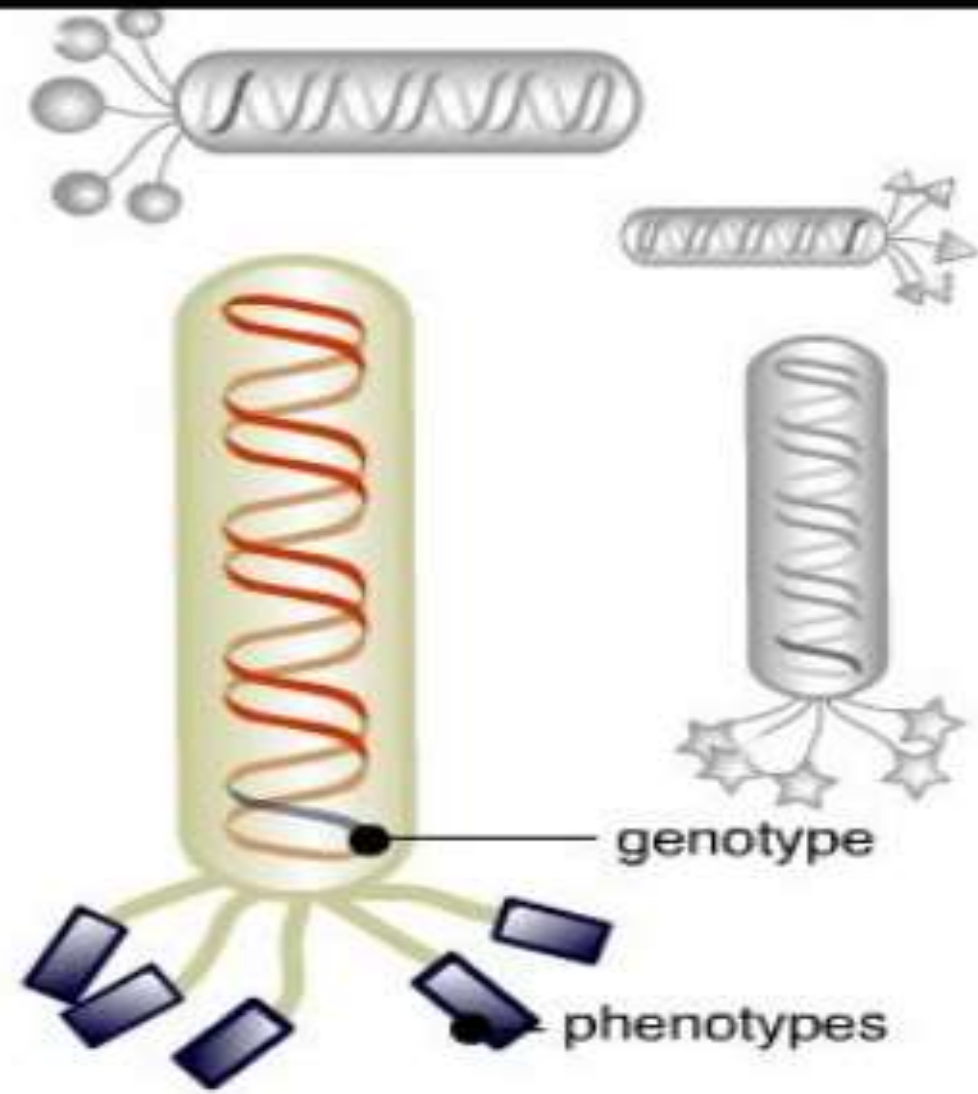
The modified gene 3 contains an added segment — an antibody, small protein, or peptide — which is expressed on the surface of the phage.

Each phage receives only one gene, so each expresses a single protein or peptide.



Step 2: Protein Display

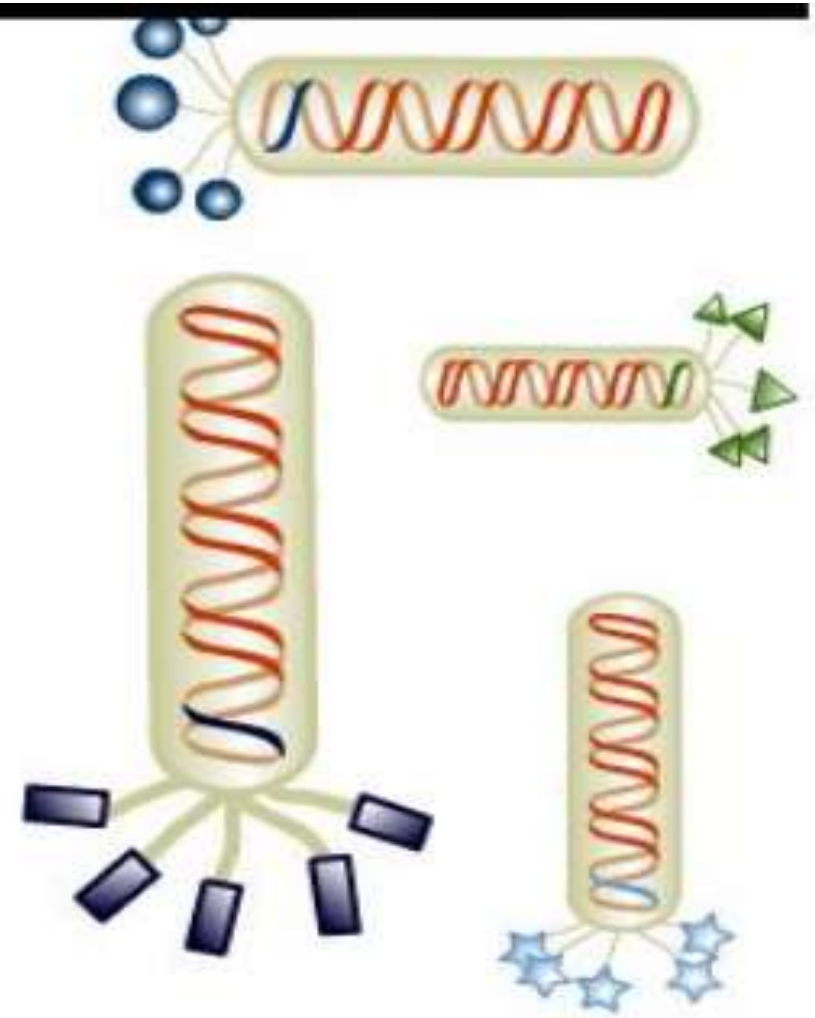
Collectively, the population of phage can display a billion or more proteins or peptides, each tied to its own gene.



Step 3: Library Creation

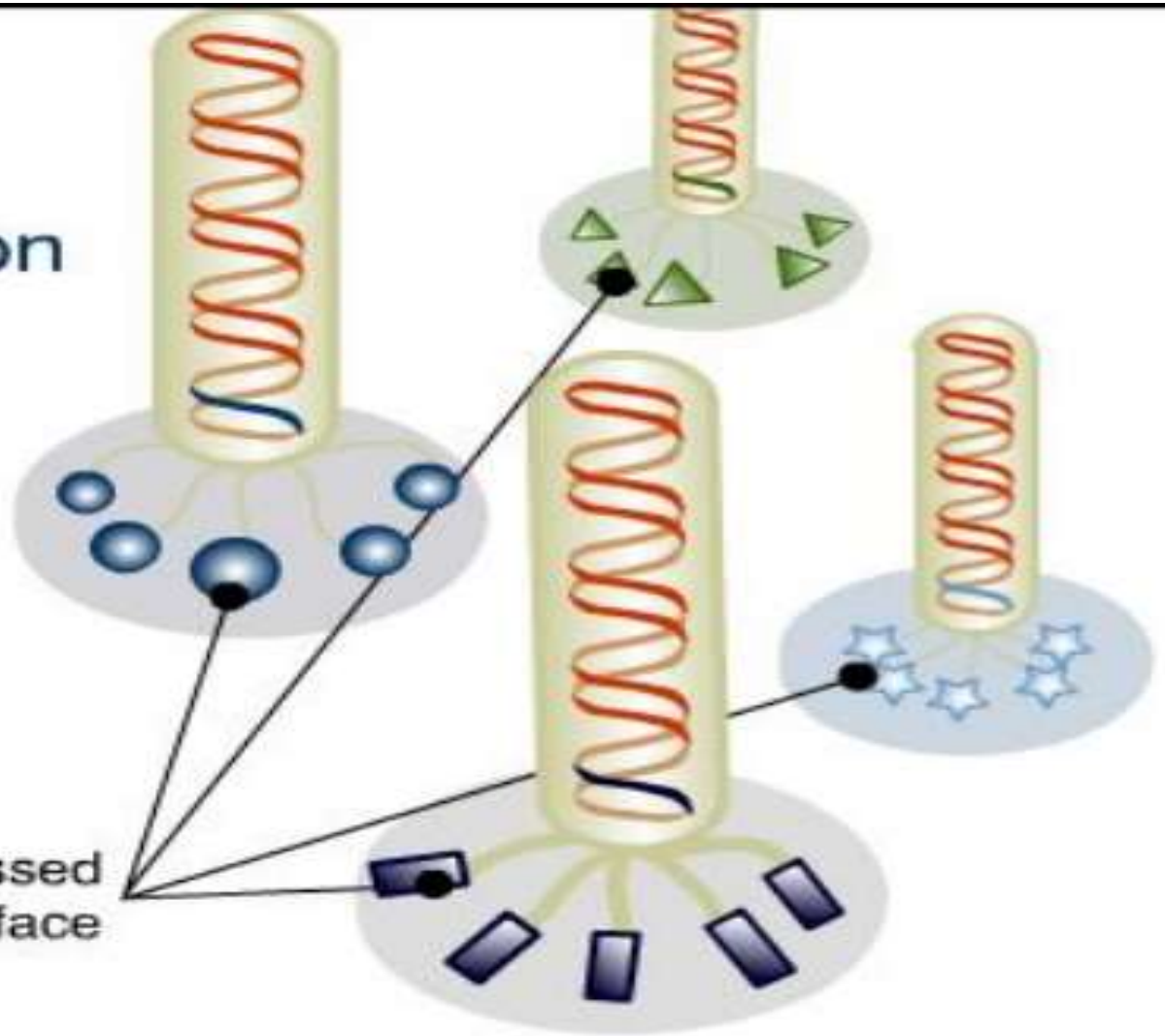
A collection of phage displaying a population of related but diverse proteins or peptides is called a library.

The related proteins keep most of the physical and chemical properties of their parent protein. These proteins are the initial leads to providing medical therapies that may cure certain diseases.



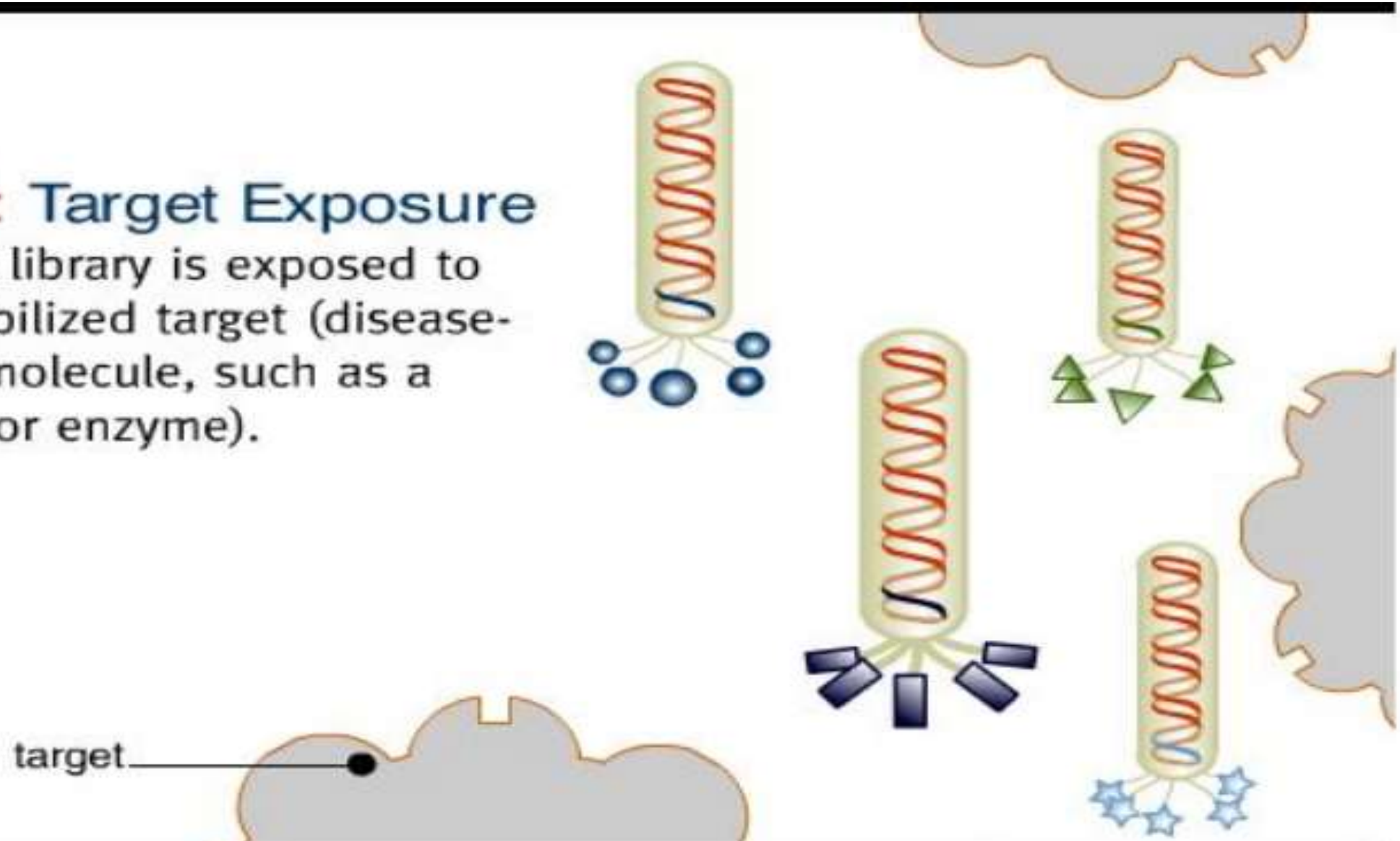
Step 3: Library Creation

Look at that diversity!



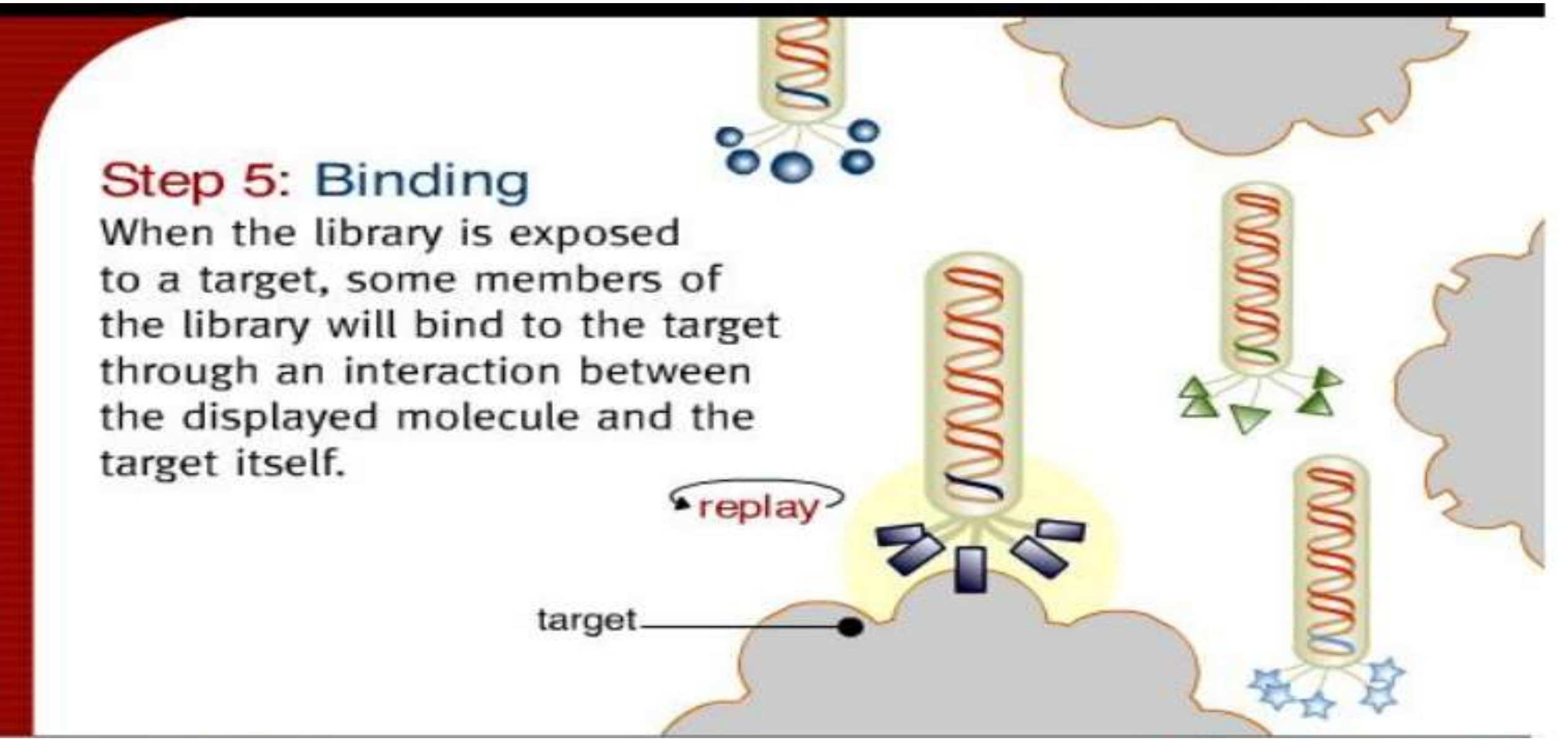
Step 4: Target Exposure

Next, the library is exposed to an immobilized target (disease-causing molecule, such as a receptor or enzyme).



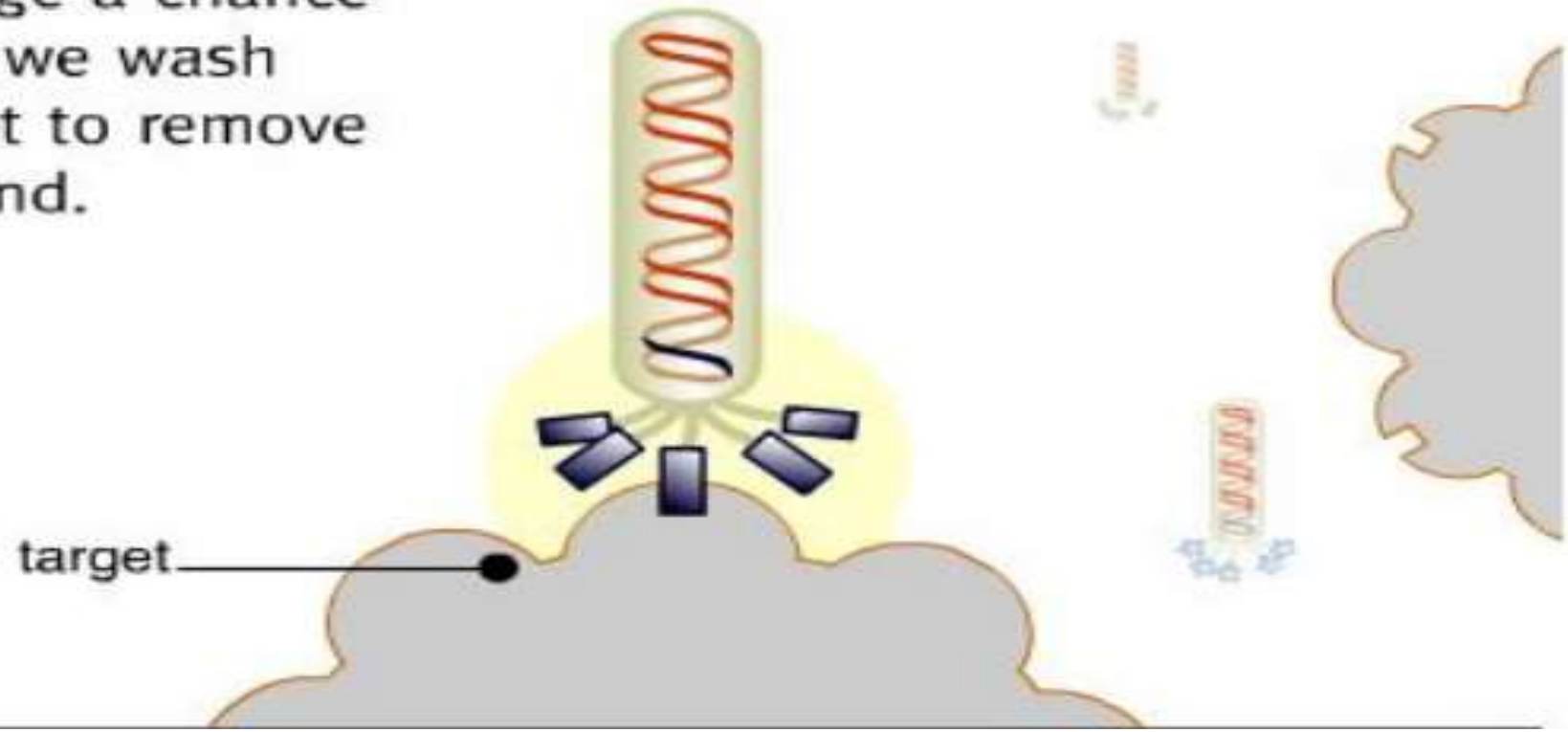
Step 5: Binding

When the library is exposed to a target, some members of the library will bind to the target through an interaction between the displayed molecule and the target itself.



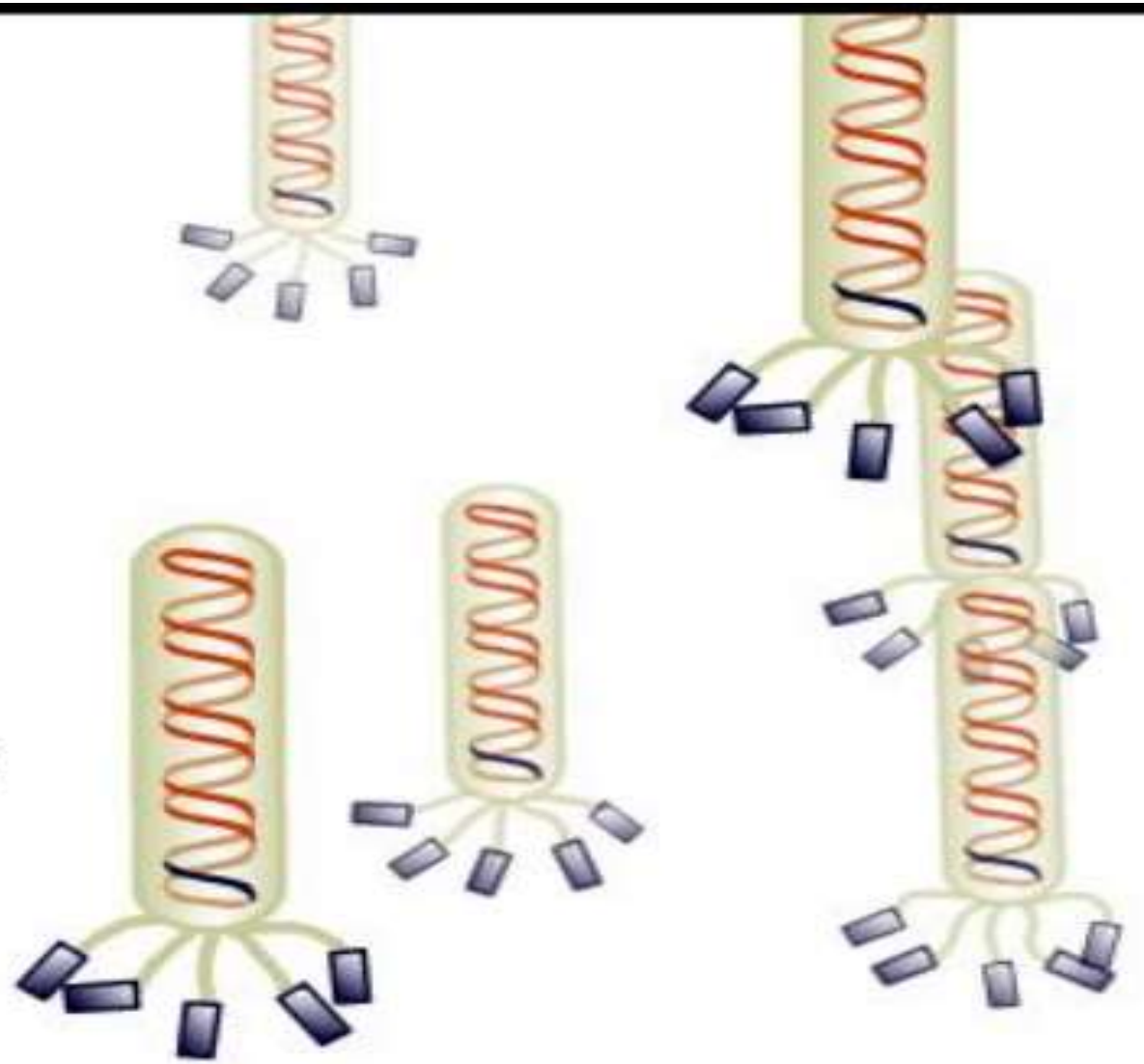
Step 6: Isolation

After we give the phage a chance to bind to the target, we wash the immobilized target to remove phage that did not bind.



Step 7: Repopulation

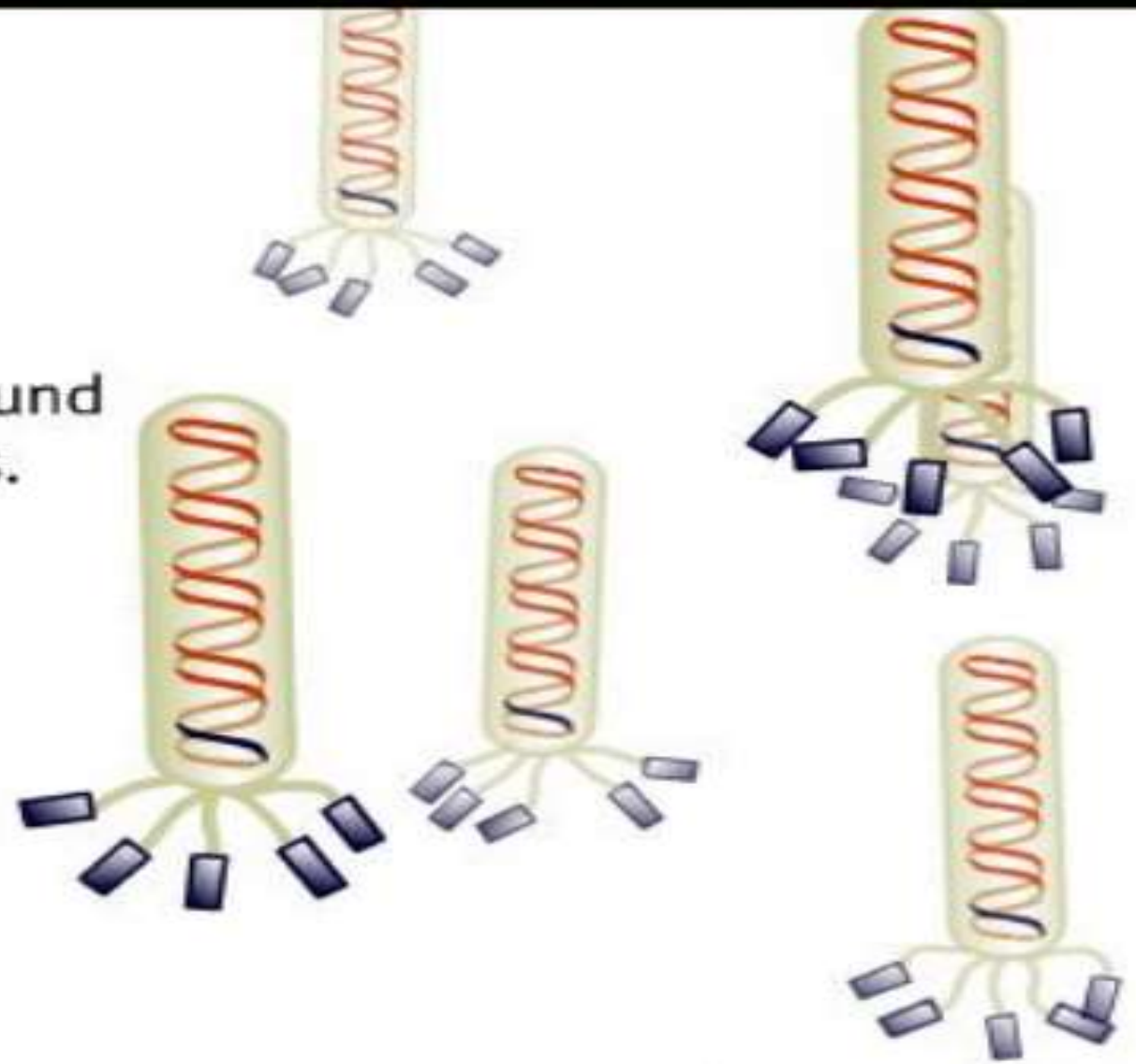
Phage display allows rapid discovery of drug candidates because we need to capture only one phage from the library. Replicating this phage in bacteria increases the amount of this phage several million-fold overnight, providing enough material for sequencing.



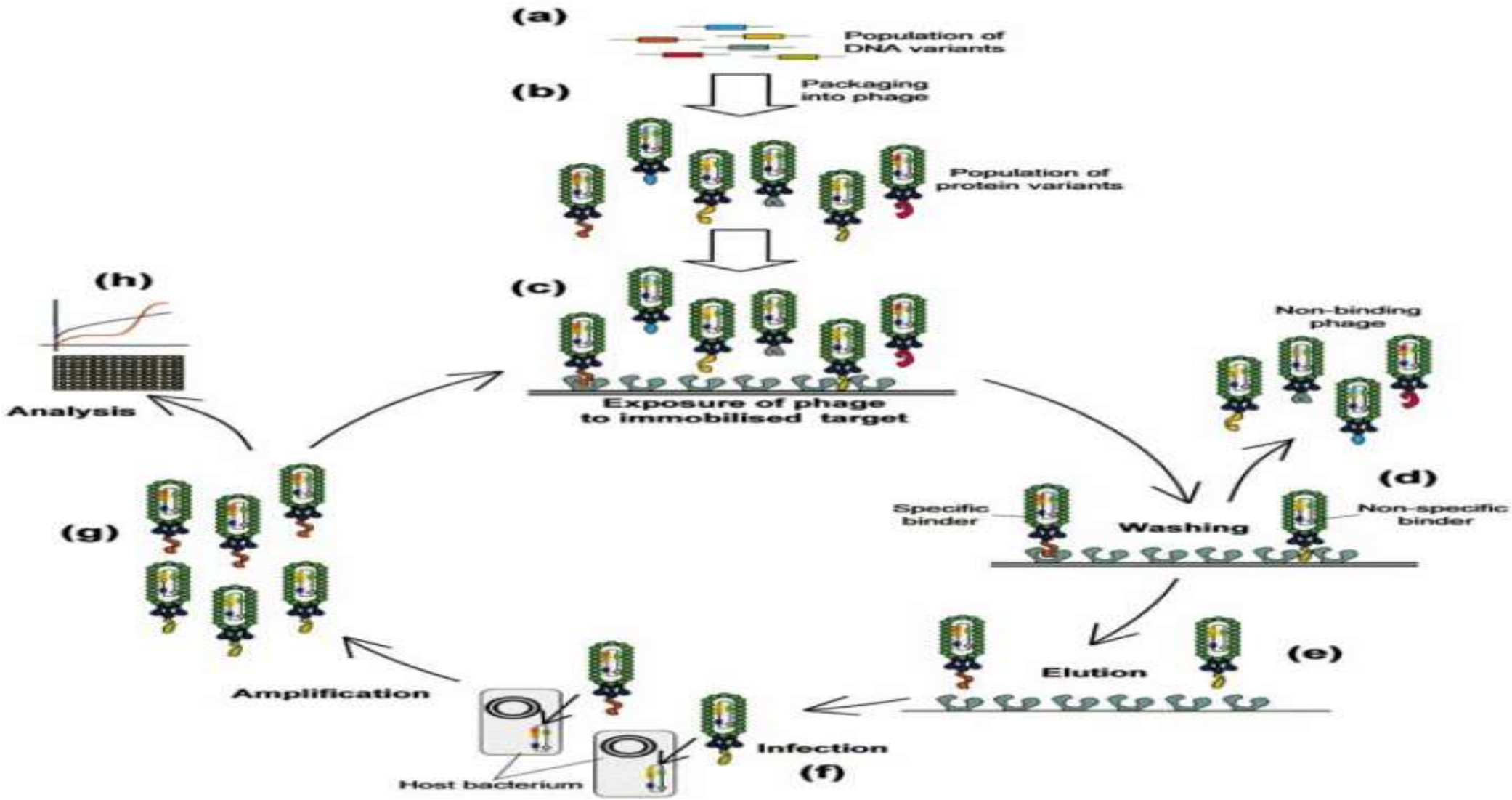
Step 7: Repopulation

Sequencing of the phage DNA tells us what the active compound is before we make more of this.

Good candidates will undergo additional testing and development.

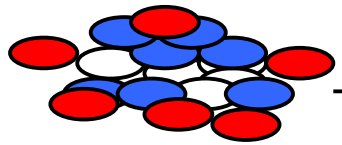


Phage display cycle

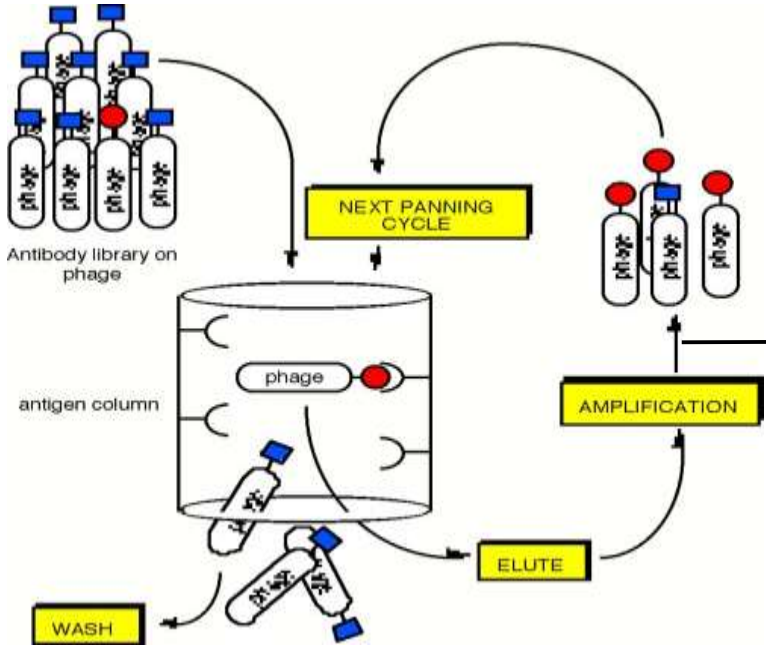


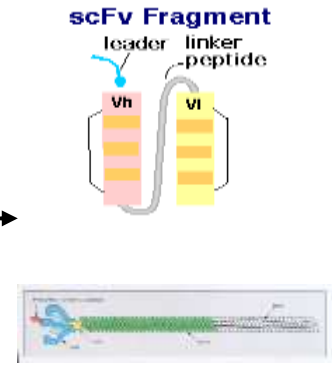
Biopanning

Displaying Targets on Filamentous Phage	Panning (Selection) of The Phage Library	Expression the Protein Fragment or Isolation of Affinity-Purified Phage Clones



Library clones





selection and screening

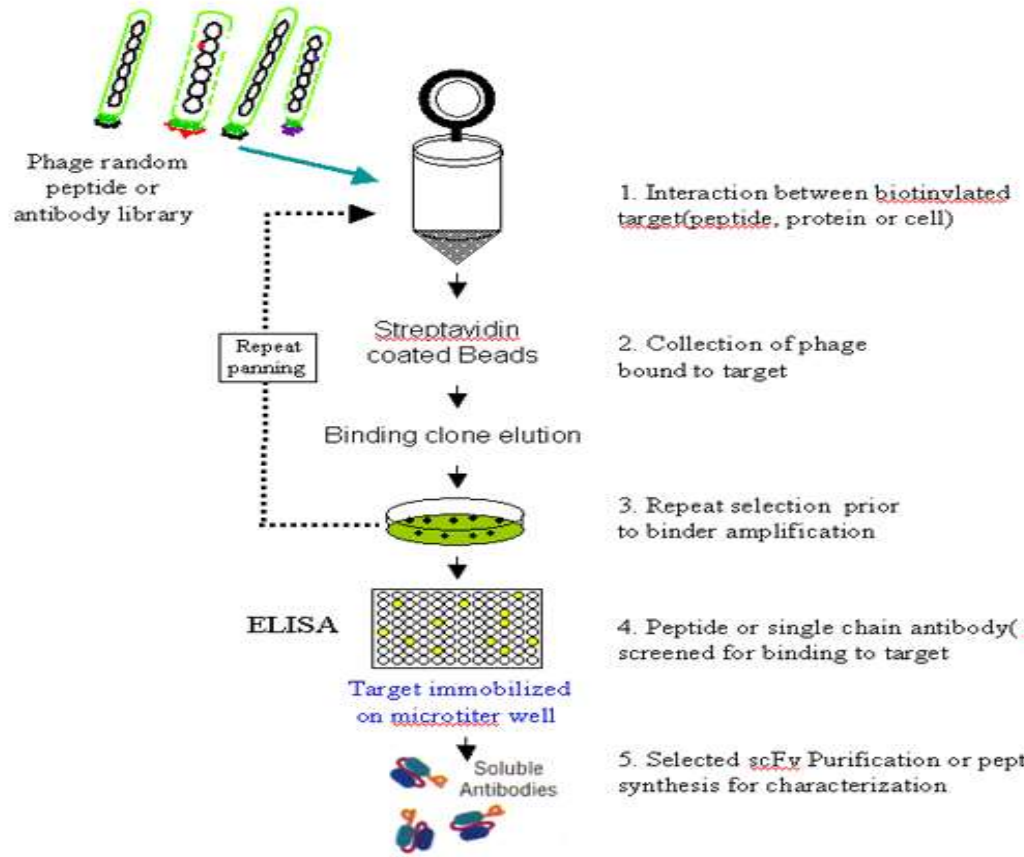
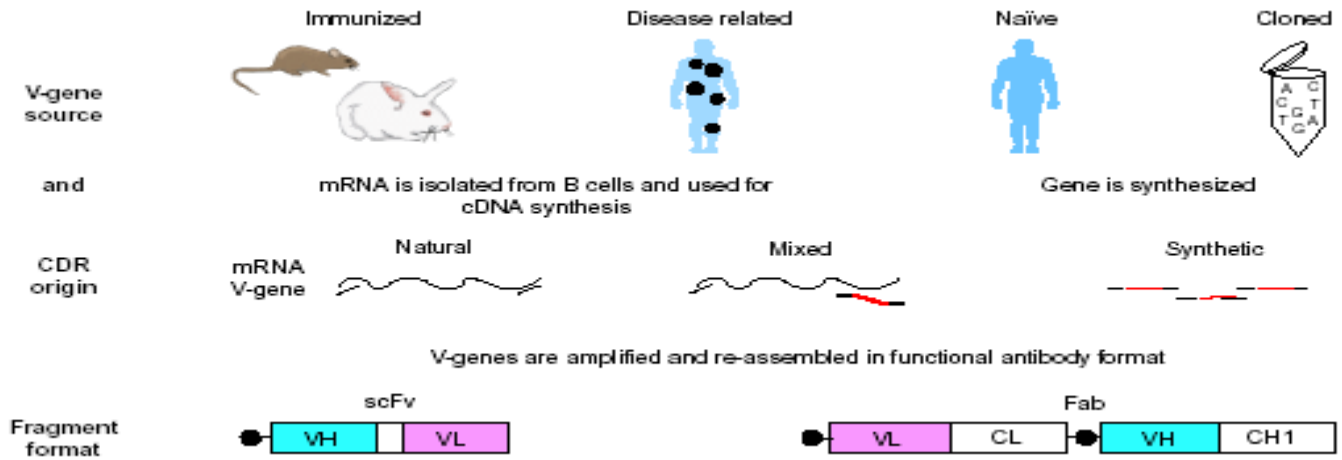


Figure. Diagram showing selection and screening of phage-displayed peptide and antibody

Phage Display Library

- The sources of genetic material



The proteins that are created and isolated by phage display process have a specific interaction with a known disease target,making this a rapid, effective and focused drug discovery method. The proteins are candidates for effective drug therapies.

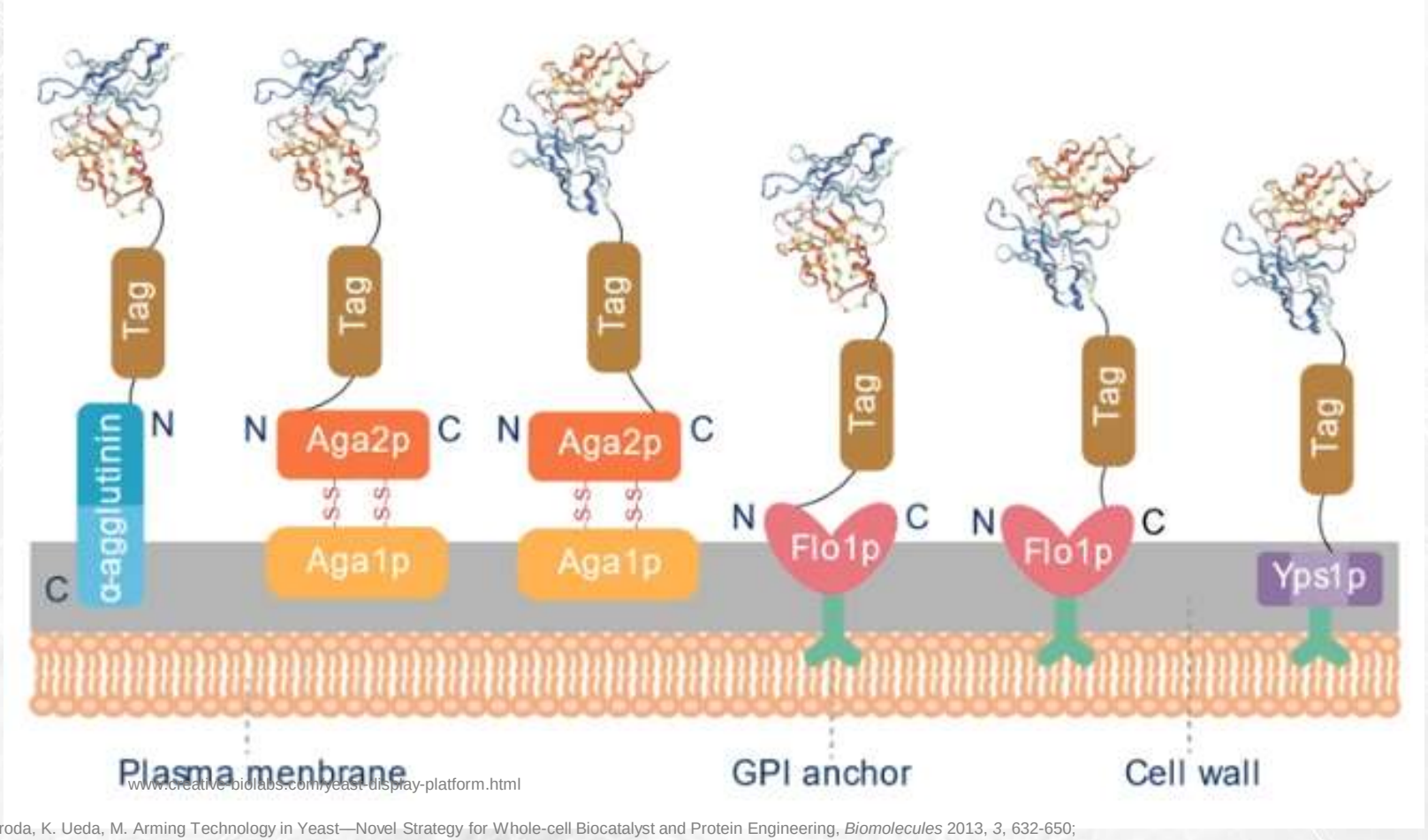
Yeast Display

Yeast display or yeast surface display is a novel technique widely used to express the proteins at the yeast surface after translation and maturation in a eukaryotic system. Thus, the yeast display is superior to phage display for display of prote-in or antibody because the eukaryotic expression environment of yeast cells ensures fine tuning such as folding, modification, and translocation prior to display on the surface.
Yeasts are suitable hosts for proteins that have a high molecular mass or require glycosylation modification.
Simultaneous display of multiple kinds of proteins/peptides on the same cell surface can be performed in yeasts

Imunoglobulin	Non-Imunoglobulin
single-chain variable fragment (scFv), Fab portion, Fc IgG, whole IgG, camelid single-domain antibody (VHH) Single chain T cell receptor (scTCR)	cystein knot peptides (knottins), human tenth domain of fibronectin type III (Fn3), lamprey variable lymphocyte re-ceptor (VLR), hyperthermophilic DNA binding protein Sso7d, human serum albumin (HSA), green fluorescent protein (GFP), major histocompatibility complex (MHC), cytokines (e.g., IL-2, IL-4), growth factors (e.g. EGF, VEGFHGF), the kringle domain, and hormones (e.g., leptins)

Yeast Display System

The yeast display indicates exogenous proteins/peptides are expressed and immobilized on cell surfa-ce by linked or anchored to yeast cell wall composition. To achieve this, exogenous genes are fused with cell wall proteins genes. A protein of interest fused to the C- or N-terminus of an anchor protein typically results in the display of the fusion protein on the cell surface of *Saccharomyces cerevisiae* or *Pichia pastoris*.



www.creative-biolabs.com/yeast-display-platform.html

Kuroda, K. Ueda, M. Arming Technology in Yeast—Novel Strategy for Whole-cell Biocatalyst and Protein Engineering, *Biomolecules* 2013, *3*, 632-650; doi:10.3390/biom3030632

A-aglutinin-based yeast display system

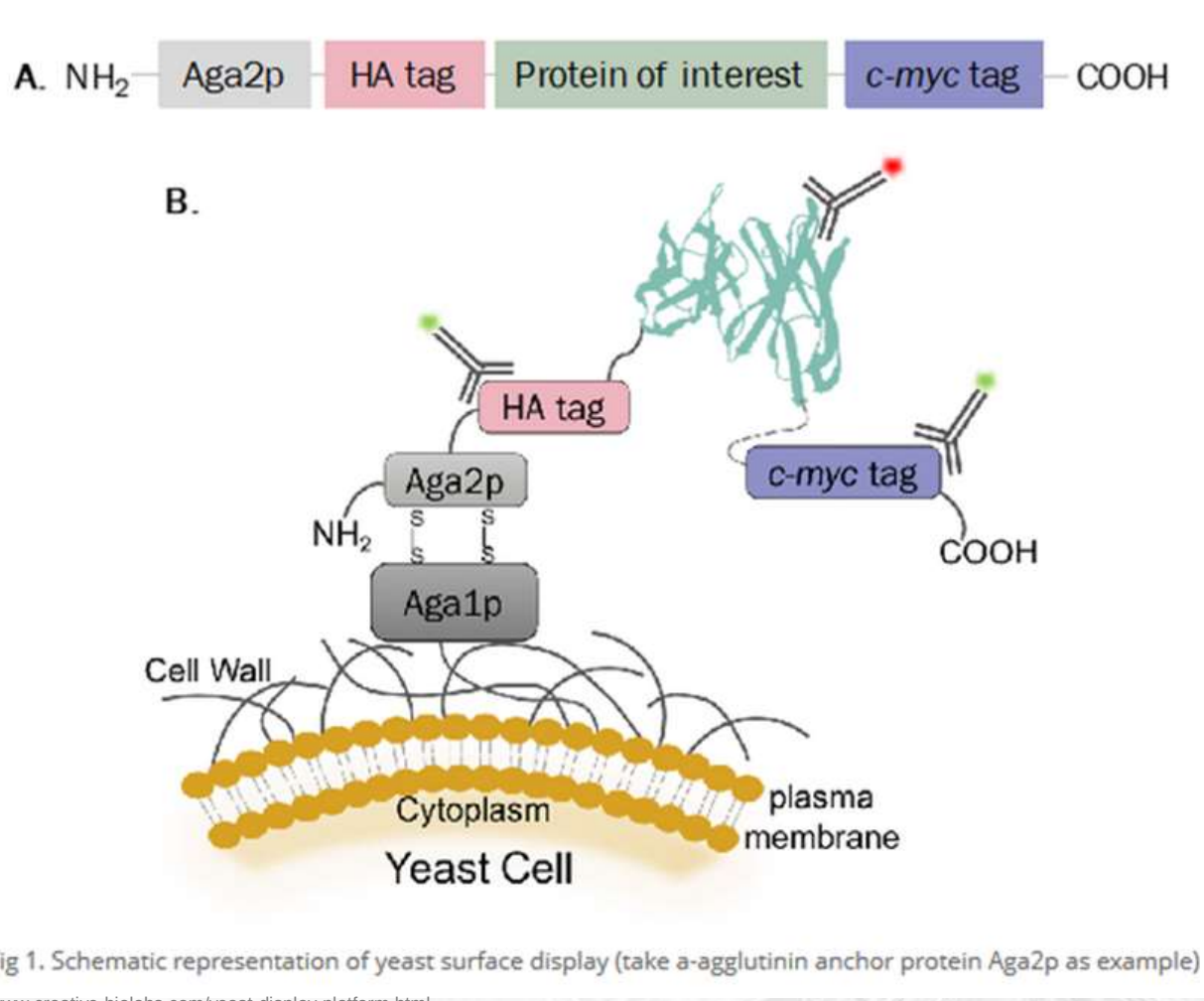
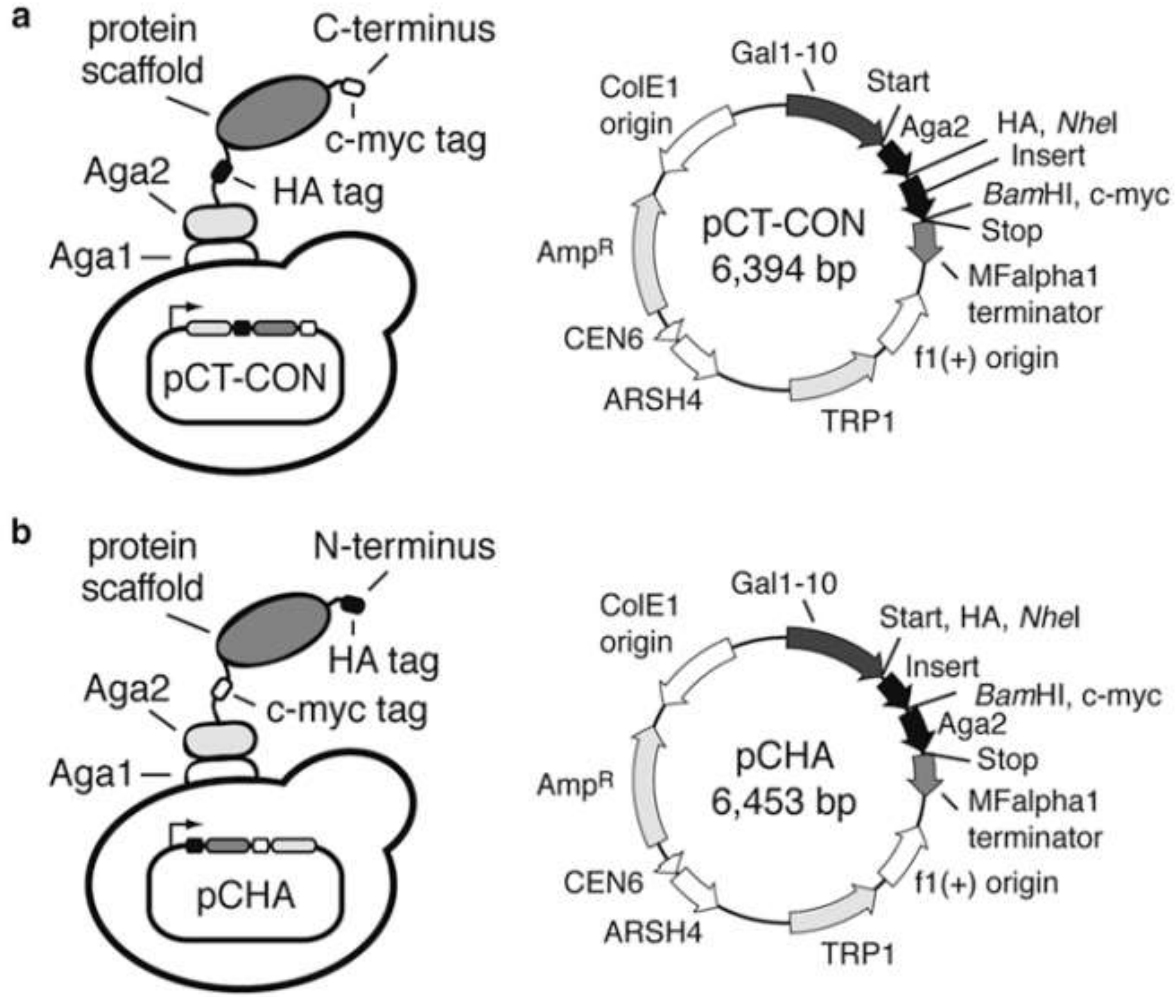


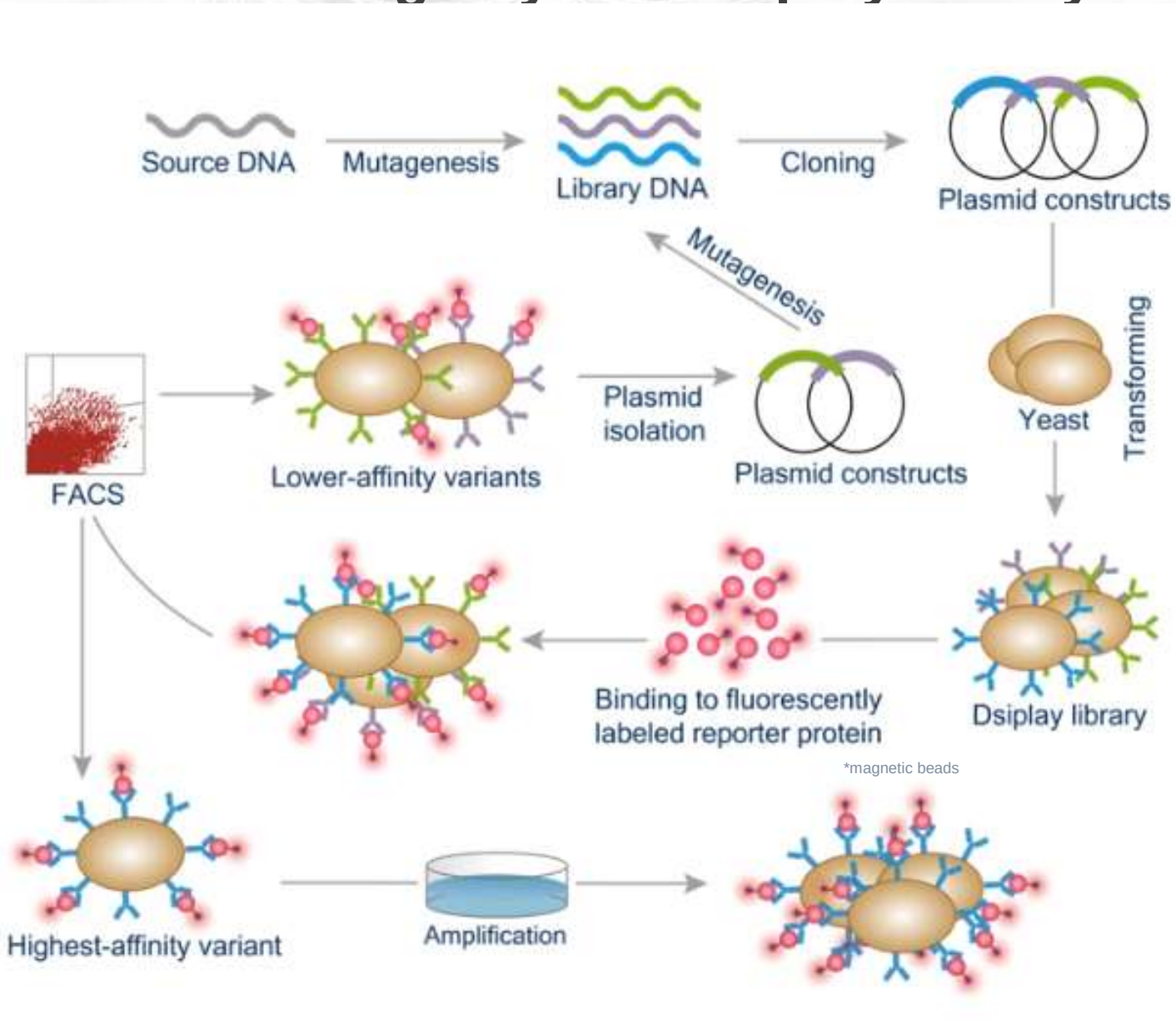
Fig 1. Schematic representation of yeast surface display (take a-agglutinin anchor protein Aga2p as example)

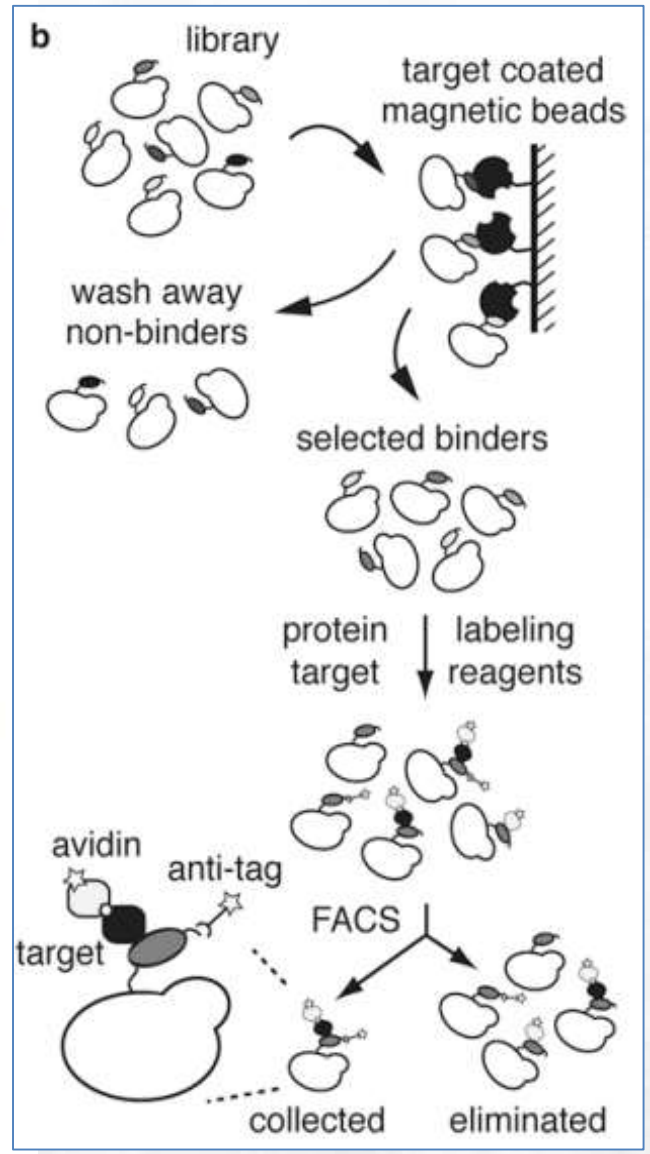
www.creative-biolabs.com/yeast-display-platform.html



Bin Liu (ed.), *Yeast Surface Display: Methods, Protocols, and Applications*, Methods in Molecular Biology, vol. 1319, DOI 10.1007/978-1-4939-2748-7_1, © Springer Science+Business Media New York 2015

Construction and screening of yeast display library





FACS :fluorescence activated cell sorting

www.creative-biolabs.com/yeast-display-platform.html