Mapping, Plasmid, and Primer Design

Shifra Ben-Dor

WEIZMANN
INSTITUTE
OF SCIENCE

BBCU

Israeli
National
Node

So, you have a DNA sequence.
Now what?

subclone

complement

reverse

clone

translate

Sequence Editing

It is critical to have an accurate copy of the sequence you plan to work with.

Whether you are cloning a known gene, designing a fusion protein, or planning PCR, you should have your ideal sequence in-silico before you start in the lab.

This can save much time, trouble and heartache.

Sequence Editing

There are various programs available for simple sequence editing:

- GCG
- EMBOSS
- DNAstrider
- VectorNTI
- MacVector
- ApE (a plasmid editor)
- Word (Microsoft Office)

-…..

Sequence Editing

The important things to remember when choosing a program:

- Does it let me jump around the sequence based on coordinates? Seqeunce?

- How easy is it to combine two existing files?

- File storage?

Sequence Editing

Don't forget to be very careful with sequence "joins"

- If you are putting sequence into a multiple cloning site, erase what's in-between!

- If you are joining at an enzyme site, be sure you know what each sequence is contributing

Sequence Editing

In gcg, the main sequence editing program is **seqed.**

lishifra34 [~]% seqed

This command opens a text editing window that looks like this:

Sequence Editing

Sequence Editing

To move around in seqed, you use

^d (control d). This allows you to move

from the header to the sequence, and

from the sequence to the command line.

To return to the sequence, press enter.

Sequence Editing

- To move around within a sequence, you use the arrow keys (right and left).

- To add a base or amino acid, just type it in.

- To erase bases, press delete

Sequence Editing

- To 'jump' around a sequence, just type the number of the character you would like to go to and press enter

- To search for a particular string in a sequence, type / and the string, and press enter. For example: /TCTAGA

Sequence Editing

- To add sequence from an existing gcg file, type ^d, then include. You will be prompted for information about the sequence you would like to add.

- To add sequence from a non-gcg or even non-unix file, just copy and paste.

Sequence Editing

- To save the work you have done on your file, type write

- There are two ways to exit seqed: quit or exit

- quit leaves the program without saving changes

- exit performs the write command automatically, and then leaves the program

Restriction Maps
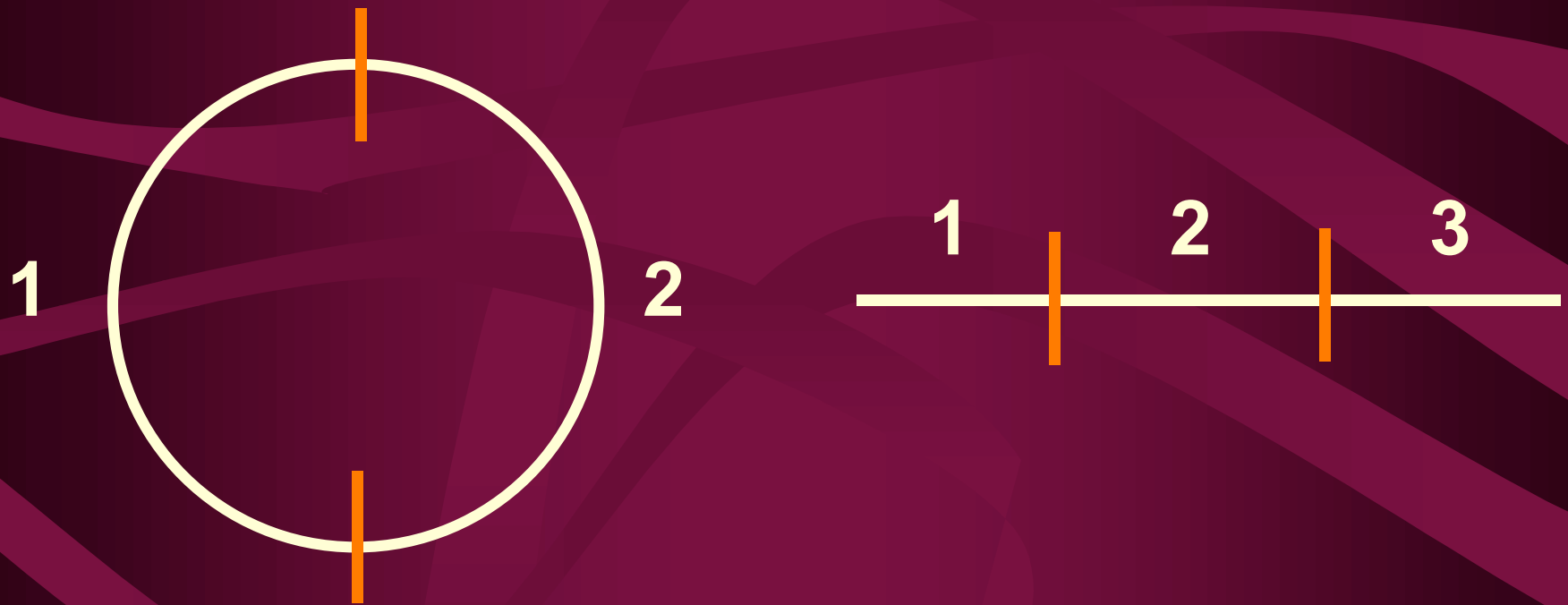
Attributes of mapping programs

- Choice of enzymes
  - Single cutters
  - x base cutters (6 base)
  - minimum/maximum of sites
- Linear/circular
- Simulation of double digests

Attributes of mapping programs

- Silent mutations
- Output
  - Annotated seqeunce
  - Table of sites (sorted by enzyme name or position)
  - Table of fragment sizes (sorted by size or position)
  - Restriction site (the actual sequence)
  - Those that do/don't cut

Mapping programs on the web

- Webcutter
- Seqcutter
- TACG4
- NetPlasmid

http://bip.weizmann.ac.il

Under Toolbox; Seq. Analysis by Target; DNA; Mapping and Primers

Mapping programs in GCG

- Map

- Mapsort

- Mapplot

MAP

- Displays the sequence and its complementary strand
- Creates a restriction map of your sequence
- Can display a translation of the sequence in all frames, three forward frames, the frame of your choice, or ORFs (open reading frames)

MAP

Restriction mapping:

all enzymes, type *   or press return

no enzymes, press space

a specific enzyme, type in the enzyme name, using the character i instead of the roman numeral
example: for HindIII type hindiii

MAP

Program options on the command line
can be used to limit the number of enzymes:

-maxcuts=2
-mincuts=2
-onc

For example, I want any enzyme that cuts my
sequence once or twice, so I type:
%map -maxcuts=2 or
%map sample.seq -maxcuts=2

MAP

Program options on the command line
can be used to limit the number of enzymes:

-maxcuts - allows me to choose the maximum
               number of cut sites
-mincuts - allows me to choose the minimum
               number of cut sites
-onc -   gives me all the enzymes that cut
               are single cutters.

MAPSORT

- Finds the coordinates of the restriction sites

- Sorts the fragments by size

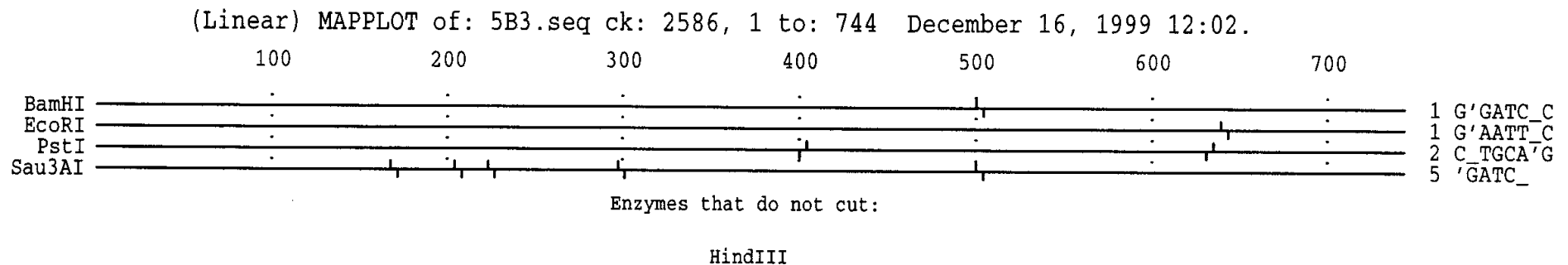- Can do single or multiple enzymes in one run of the program

MAPSORT

An important program option is:

-dig

This performs a digest with multiple enzymes at the same time, and gives an idea of what the gel will look like. It sorts the pieces both by location of sites, and by size.

MAPPLOT

- Displays restriction map graphically

- Requires plotter or defined printer

```
(Linear) MAPPLOT of: 5B3.seq ck: 2586, 1 to: 744  December 16, 1999 12:02.
          100           200           300           400           500           600           700
          .             .             .             .             .             .             .
BamHI     .             .             .             .             .             .             .      1 G'GATC_C
EcoRI     .             .             .             .             .             .             .      1 G'AATT_C
PstI      .             .             .             .             .             .             .      2 C_TGCA'G
Sau3AI    .             .             .             .             .             .             .      5 'GATC_

                        Enzymes that do not cut:

                              HindIII
```

Translation Programs

- Translates nucleotide sequences into peptide sequences
- Some do all reading frames, some have choice of frames
- Some do full translations, others only ORFs
- Usually have option to reverse sequence
- Can sometimes add multiple exons from one parent sequence

Translation Programs

- The definition of an ORF
  - Start to Stop
  - Stop to Stop
- Minimum sequence to be considered an ORF
- Alternate start codons (mainly microbial)
- Multiple ATGs

Translation Programs on the web

- ORF finder (NCBI)
- Translate (Expasy)
- Transeq (EBI - EMBOSS)

TRANSLATE

- Translates nucleotide sequences into peptide sequences

- Has option to reverse sequence

- Can add multiple exons from one parent sequence

REVERSE

- Can reverse, complement or reverse and complement a nucleotide sequence

- File remains nucleotide sequence, does not translate

Primer Design

When do we need primers?

- Sequencing (one primer)
- PCR (two, one for each strand)
  - Exact (cloning, add tags, add enzyme sites, site directed mutagenesis, …)
  - Degenerate
- Real time quantitative PCR (qPCR)
- RNAi
  - One primer (synthetic)
  - Hairpin (plasmid)

Primer Design

- Things to keep in mind:

    - Primer length
    - AT/GC ratio should be around 50%
    - 3' end should be G/C
    - melting/annealing temperature
    - secondary structure
    - primer dimers

Primer Length

- Primers have to be long enough to be specific, but short enough to detach efficiently from the template

- Ideal lengths are from 18-24 bp long

- For some applications, we use longer ones (adding enzyme sites, tags, changing the end of a sequence…)

- We rarely use shorter ones

GC ratio

- If there are too many Gs and Cs, it will be hard to separate the primer from the template (G and C have 3 hydrogen bonds)
- We generally try to keep the G/C percentage as close to 50% as possible, with a range of 45% - 55%
- If nothing is found, expand the range

3' clamp

- There is a running argument in the literature as to what base is prefereable at the 3' end. Some maintain that an S clamp (G or C) makes for better priming, others say it makes it worse. We generally recommend using an S clamp (unless you're doing qPCR, in which case an A is recommended)

Melting temperature

- The melting temperature of the primers directly effects the temperature of the annealing step of PCR.

- Currently accepted norms: primer melting temperatures in the 58°C - 60°C range

- The difference in melting temperatures of primers should be as little as possible, but can be up to 5°C

Annealing temperature

- The "rule of thumb" for annealing temperature: it should be 5^oC less than the melting temperature

- Optimally, it should be determined for each set of primers on a gradient cycler

- Currently accepted: a minimum of 50^oC

- It works down to 37^oC, but specificity may become an issue

- If you're working with degenerate primers, you need lower temperatures, though you can use them for a few initial cycles

Secondary structure

- Internal complementarity:

  There should be no self matching stretches of 3 bases or more, or the primer will bind to itself in a hairpin, and not be able to prime

Other Primer Issues

- Primer Dimers

  When the 3' end of the one primer is complementary to the other primer, the primers can anneal to each other and create a new template

- Primer Complementarity

  If the primers are complementary anywhere else, it can interfere with hybridization

- Primer/Template:

  Avoid stretches of 3 bases or more in a row of the same base - it can lead to mispriming (G, C) or breathing (A, T)

Primer Design

- If you are changing the beginning of a coding region:

    – ATG start codon

    – Kozak sequence
    (GCC) GCC (A/G)CC ATG G

    – signal sequence (secreted, membrane bound)

Reverse (**not complement**) 3' primer

5 '   3 '

GATAAGCTTGATATCGAATTGCCATGTTGAAGCCATCATTACCATT
CTATTCGAACTATAGCTTAACGGTACAACTTCGGTAGTAATGGTAA

5 '   3 '
GATAAGC

CTATTCGAACTATAGCTTAACGGTACAACTTCGGTAGTAATGGTAA

Primer = **GATAAGC**

5 '   3 '
GATAAGCTTGATATCGAATTGCCATGTTGAAGCCATCATTACCATT
                                                                          ATGGTAA

Primer = **AATGGTA**   3 '   5 '

Primer Design

- Always make sure that you are in frame!

- Double check the orientation of the sequence before you submit it for synthesis!

Primer Design

Always sequence PCR products!!!!

(preferably after subcloning, unless you are just checking for presence of product)

```
1                                                              60
233  GATAAGCTTG ATATCGAATT GCCA.GTTGA AGCCATCATT ACCATTCACA TCCCTCTTGT
240  GATAAGCTTG ATATCGAATT GCCATGTTGA AGCCATCATT ACCATTCACA TCCCTCTTAT
239  GATAAGCTTG ATATCGAATT GCCATGTTGA AGCCATCATT ACCATTCACA TCCCTCTTAT
gamma         G AAGAGCAAGC GCCATGTTGA AGCCATCATT ACCATTCACA TCCCTCTTAT

61                                                             120
233  TCCTGCAGCT GCCCCTGCTG GGAGTGGGGC TGAACACGAC AATTCTGACG CCCAATGGGA
240  TCCTGCAGCT GCCCCTGCTG GGAGTGGGGC TGAACACGAC AATTCTGACG CCCAATGGGA
239  TCCTGCAGCT GCCCCTGCTG GGAGTGGGGC TGAACACGAC AATTCTGACG CCCAATGGGA
gamma TCCTGCAGCT GCCCCTGCTG GGAGTGGGGC TGAACACGAC AATTCTGACG CCCAATGGGA

121                                                            180
233  ATGAAGACAC CACAGCTG.. .......... .......... .......ATTT CTTCCTGACC
240  ATGAAGACAC CACAGCTGGT GGGAAATCTG GGACTGGAGG GGGCTGATTT CTTCCTGACC
239  ATGAAGACAC CACAGCTG.. .......... .......... .......ATTT CTTCCTGACC
gamma ATGAAGACAC CACAGCTG.. .......... .......... .......ATTT CTTCCTGACC

181                           210
233  ACTATGCCCA CTGACTCCCT CAGTGTTTCC
240  ACTATGCCCA CTGACTCCCT CAGTGTTTCC
239  ACTATGCCCA CTGACTCCCT CAGTGTTTCC
gamma ACTATGCCCA CTGACTCCCT CAGCGTTTCC
```

Primer Programs in GCG

- prime

- primepair

- melttemp

Prime

- Based on Primer3
- Looks for primers in a given sequence
- Compares primers to input sequence
- Has many parameters that can be changed and optimized

PrimePair

- Compares a set of primers

- Sequence **<u>independent</u>** so ideal for checking existing pairs of primers, or for checking primers that don't match the parent sequence (for example, after adding a linker, or enzyme restriction sites)

- Has many parameters that can be adjusted

The main reason that the GCG programs fail to find primers (if they fail) is the default difference in melting temperature between two primers - which is set at $2^{o}C$. This can be raised up to $5^{o}C$, and can help many times when no primers are chosen otherwise.

Primer Prediction on the Web

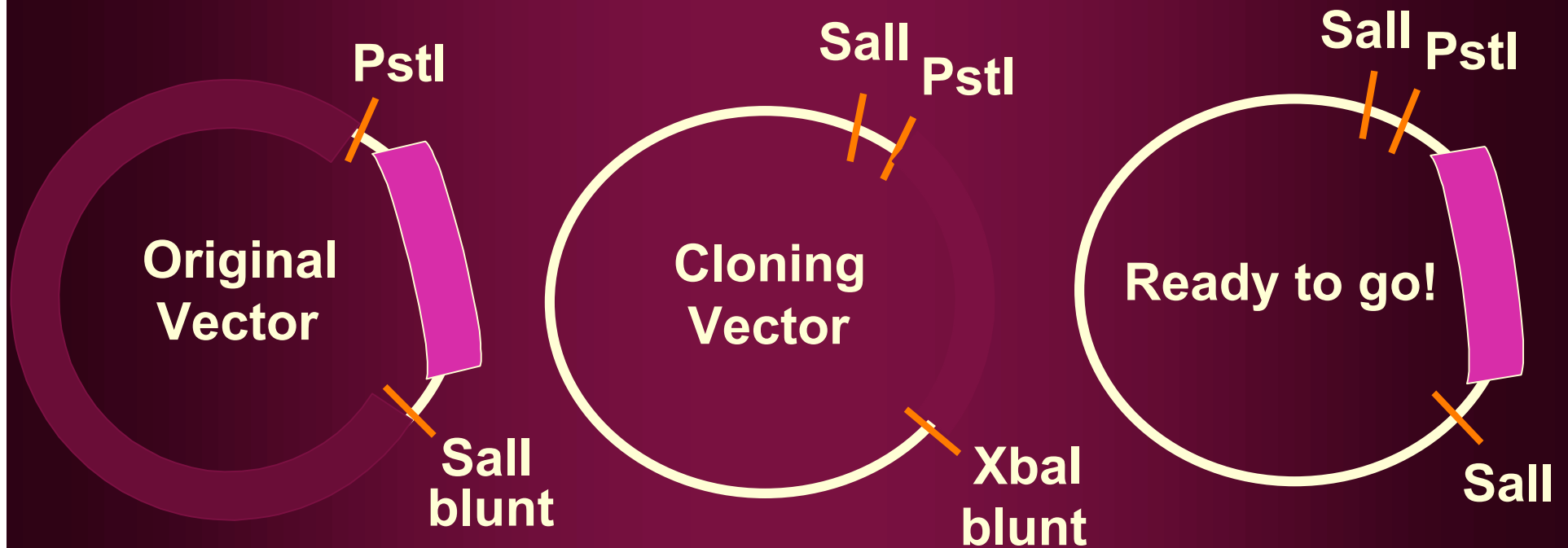http://bip.weizmann.ac.il/toolbox/target/dna/dna_primers.html

Plasmid Design

Things to remember when designing plasmids

- What is your target cell line?
  - Eukaryotic / Prokaryotic
  - Promoter, Origin of replication….
- How are you going to replicate this plasmid?
  - Bacterial origin of replication
  - Copy number control
- What is your target cell "space"
  - Intracellular, extracellular, vesicular
  - Leader sequence
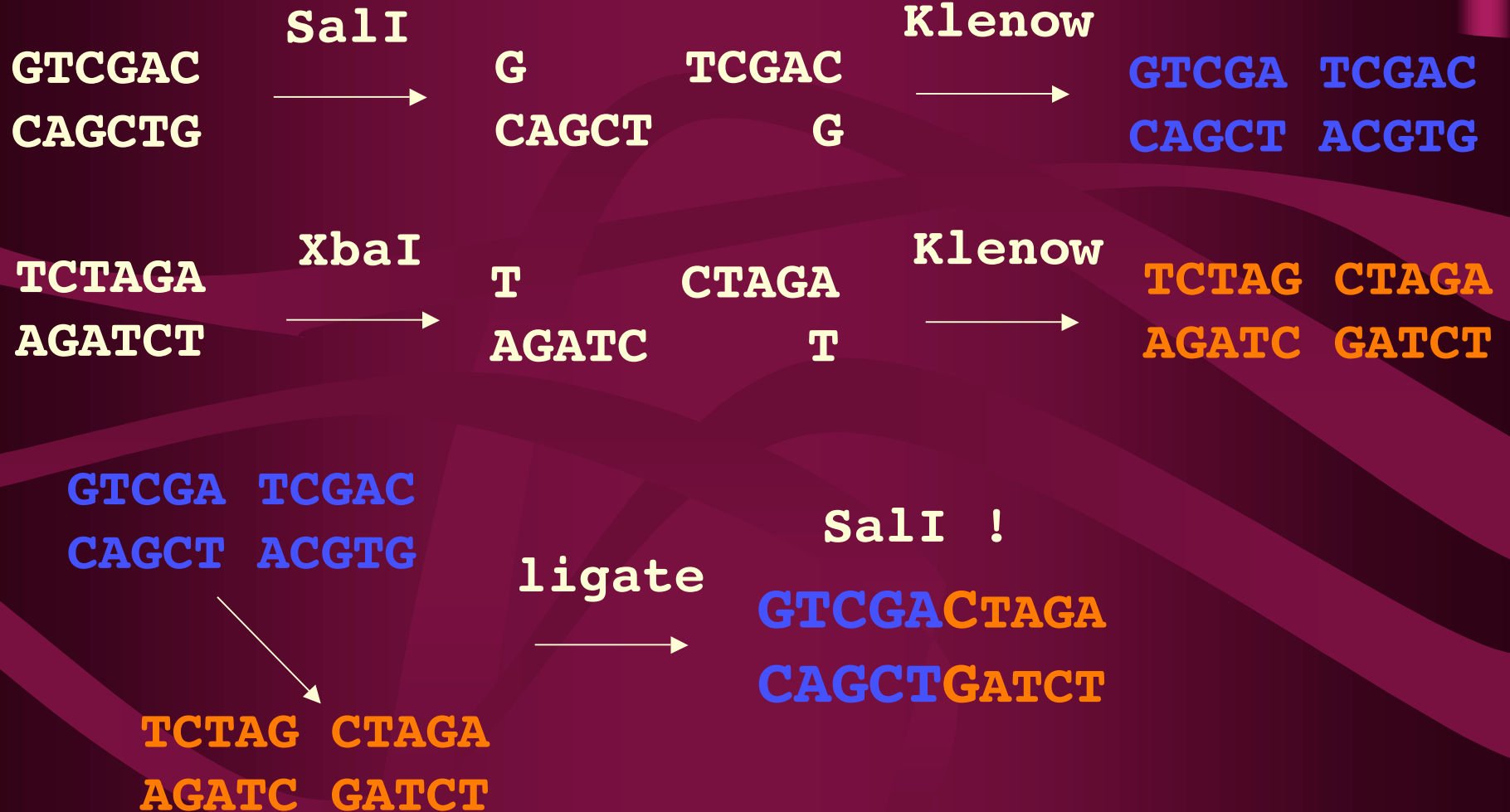
What to do when you're stuck

- 3-way, 4-way, 5-way…ligations
- "Plasmid Shuffle"
- Linkers
- Add via PCR

- **ALWAYS REMEMBER TO CHECK YOUR READING FRAME!!!!**

Problem: cloning site is SalI, but only have Sal on one side of the gene

Cloning Tricks: Sal +Xba = Sal

```
GTCGAC      SalI         G          TCGAC     Klenow     GTCGA TCGAC
CAGCTG     ───────►      CAGCT          G    ───────►    CAGCT ACGTG


TCTAGA      XbaI         T          CTAGA     Klenow     TCTAG CTAGA
AGATCT     ───────►      AGATC          T    ───────►    AGATC GATCT


GTCGA TCGAC                                            SalI !
CAGCT ACGTG
               ligate                                 GTCGACTAGA
             ───────►                                 CAGCTGatct
TCTAG CTAGA
AGATC GATCT
```

Cloning Tricks: Sal +Xho

SalI   XhoI

GTCGAC   CTCGAG
CAGCTG   GAGCTC

↓   ↓

G   TCGAC   C   TCGAG
CAGCT   G   GAGCT   C

GTCGAG   CTCGAC
CAGCTC   GAGCTG

Can also be done with BamHI and BgIII