

Virtual Restriction Mapping

Sandra Porter, Ph.D.
Geospiza, Inc.

Learning objectives:

To understand and be able to name factors that affect the frequency and occurrence of restriction sites within a DNA sequence. These same factors affect the design of oligonucleotides used for PCR primers and for probes.

To learn:

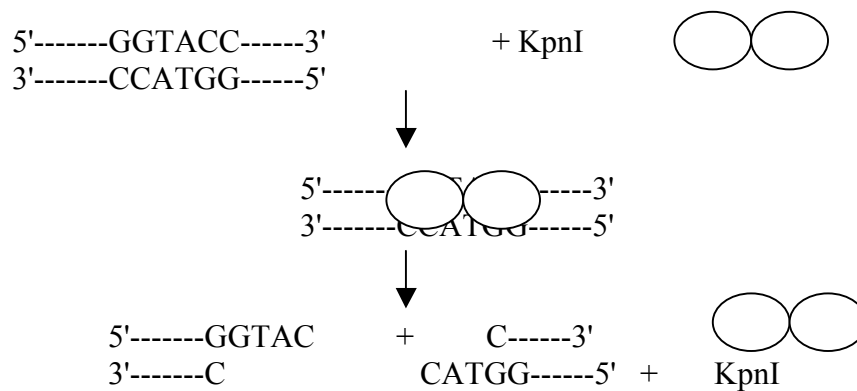
- How to use a computer to locate restriction sites and create a map.
- About factors that influence the probability of finding a random DNA sequence.

To see an example of how pattern recognition is used in bioinformatics.

Background

Restriction enzymes are proteins that recognize specific DNA sequences and cut phosphodiester bonds in the DNA backbone. Type II restriction enzymes are dimers that cut both strands of DNA at once. The two pieces of DNA created by this reaction break apart after digestion because the hydrogen bonds that might remain aren't strong enough to hold the molecule together. The DNA sequences recognized by restriction enzymes are called "restriction sites". In this exercise, you will simulate a variety of restriction digests and make a map showing the location of restriction sites in a specific sequence of DNA. You will also make and test predictions regarding the frequency of different restriction sites.

Many restriction enzymes are dimers (contain two identical subunits). They bind to DNA and move along the DNA molecule until they recognize a specific sequence. Then they cut the phosphodiester bonds within that sequence. Because the recognition sequence is found on both of the two DNA strands, they cut both strands. Hydrogen bonds between complementary base pairs are too few in number and too weak to hold the DNA fragments together, so the two molecules of DNA separate.



The recognition sites for a variety of restriction enzymes are shown in the table below. Each DNA sequence also contains an arrow showing where that sequence is cut by the restriction enzyme. For example, Kpn I binds to DNA, moves back and forth, searching for the sequence, 5' GGTACC 3'. Once the sequence has been found, Kpn I cuts the DNA backbone between the two cytosines. Since the two strands of DNA are complementary, only the sequence of one strand is shown in the table.

Enzyme name	Recognition site
SacI	GAGCT↓C
SmaI	CCC↓GGG
KpnI	GGTAC↓C
HindIII	A↓AGCTT
SphI	GCATG↓C
BamHI	G↓GATCC
EcoRI	G↓AATTC
PstI	CTGCA↓G
DraI	TTT↓AAA
MseI	T↓TAA
HhaI	GCG↓C

Procedure

Analysis of example sequences. Work with a partner to complete the steps below. Assign one person to take notes and one person to perform the computer activities.

1. Go to the NCBI home page (<http://www.ncbi.nlm.nih.gov>)

Choose the nucleotide database from the finder bar.

2. Type in the following accession number: Z96795

Change the first pull-down menu to read "FASTA"

Click on the button marked "Display"

3. Copy the sequence and paste in a WORD document (or some other kind of text document). Name the file "T. brucei" and save it.
4. Repeat the same steps for the second sequence. The accession number is: L39157. Name this file "S. laurentii" and save it.
5. Take five minutes and write down everything that you can think of that would be different between a member of the genus, *Streptomyces* and *Trypanosma bruceii*. You may want to do a quick search with Google and find out two or three things about each organism, before the discussion.
6. Examine the table showing recognition sites for restriction enzymes. Write down anything that you notice about the restriction sites. Note: there are many different restriction enzymes, this is *not* a comprehensive list.
7. Next you'll use WORD to locate restriction sites, separate the DNA sequences apart and determine the sizes of each "fragment".

Start with KpnI. Cut the DNA sequence into fragments in the following way.

- A. Under the Edit menu is a command called "Find". Select this command.
- B. A window will appear with a box requesting the text that you wish to find. Type a recognition sequence, for the enzyme you're working with, in that box. For Kpn I, that would be "GGTACC".
- C. Then click the button that says "Find Next".
- D. Each KpnI site in the DNA sequence will be highlighted, in order.

Ex:
GGTCTACATACGGTACCCCTTATACTATGATAGTAGGTACCGGATCTACTA

At each KpnI site, move the cursor to the position where the DNA is cut, push "return or enter" to mark where the sequence would be broken by cutting with KpnI.

Ex: GGTCTACATACGGTAC
CCCTTATACTATGATAGTAGGTAC
CGGATCTACTA

Continue this process until you have broken the sequence at the correct location for each KpnI site.

E. Determine the size of each fragment and the location of all KpnI sites within each sequence.

Start with the first fragment. Select the fragment with the mouse, and choose the "Word count" command in the "Tools" menu.

The number of characters in a "fragment" is equal to the number of letters, or bases, in this DNA sequence. There are no KpnI sites within the DNA sequence from the garden pea. The *Streptomyces* DNA sequence contains one KpnI site. Separating the DNA sequence at the cut site for KpnI creates two DNA sequences, a long fragment of 2916 characters or 2916 bases and a smaller fragment that's only 25 bases long.

4. Determine the location and size of the restriction fragments that would be created by digesting both DNA sequences each enzyme in the table *except* for the last two, MseI and HhaI.

Record your data in the table shown on the next page.

5. Draw a map for each sequence showing the location of all restriction sites.

Enzyme	Recognition site	number of fragments		Location of restriction site	
		T. bruceii	Streptomycete	T. bruceii	Streptomycete
SacI	GAGCT↓C				
SmaI	CCC↓GGG				
KpnI	GGTAC↓C				
HindIII	A↓AGCTT				
SphI	GCATG↓C				
BamHI	G↓GATCC				
EcoRI	G↓AATTC				
PstI	CTGCA↓G				
DraI	TTT↓AAA				
MseI	T↓TAA				
HhaI	GCG↓C				

6. Review the information in the table. Write down any conclusions or observations that you can make about restriction sites.

Approximately, how often did you see a "six base" sequence appear?

Predict how the frequency would change if you were to look for a DNA sequence of four bases.

7. Look for the recognition sequences for MseI and HhaI. Record your data in the table on the previous page.

Was the frequency different? If so, how?

Do you notice anything else about the frequency of cutting?

Compare the DNA sequences from the two different organisms. Do you notice anything about the sequences that might affect the probability of finding a restriction site?

8. The product rule can be used to estimate the probability of finding a restriction site of a specific size. The product rule states that the probability of two events, A and B, both occur, equals the probability of event A occurring multiplied by the probability of event B, occurring.

This can be expressed in the following way: $P(A\&B) = P(A)P(B)$

Use the product rule to determine the probability of finding the following:

Six base sequence:

Four base sequence:

a 20 base sequence:

9. Given the probability of finding a random 20 base sequence, how often would you expect to find a random 20 base sequence in the human genome? (The human genome is approximately 3 billion bases in size)