H. Drossman **CH 345: Bioanalytical Chemistry** **Block 6, 2008**

MS DNA Sequencing Approaches

MODEL 1 TOF Mass Spectrometry

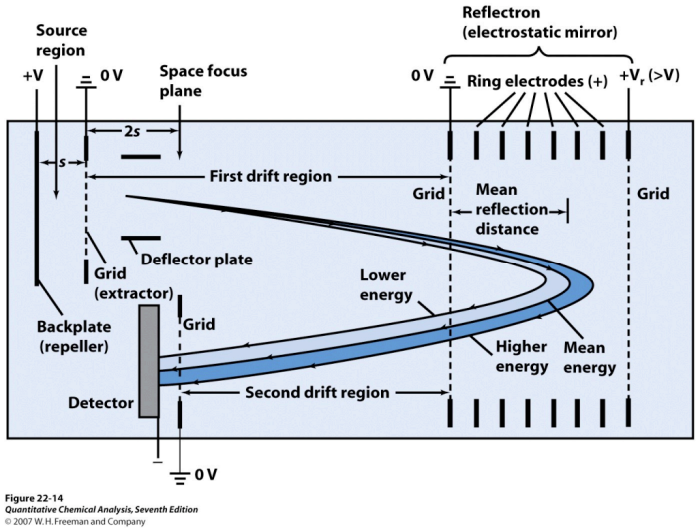


Figure 22-14
Quantitative Chemical Analysis, Seventh Edition

1. Edwards et. al. state that "in its simplest form, conservation of energy can be used to relate the electrical potential energy of the DNA fragment to its kinetic energy." Write an equation to solve for the velocity as a function of m/z and V where m is the mass, z is the charge and V is the MS voltage.

2. Explain the role of the source region and the ring electrodes in the TOF diagram above for determining the different m/z.

3. How does MALDI allow large molecules like oligonucleotides to enter an instrument that was traditionally designed for gas phase ion analysis?

4. A typical MS experiment can be done quite rapidly so as many as 5,000 samples could be run on an instrument in one day. As we will see the maximum resolution is ~100nt. How do CE and MS compare for genomic sequencing?

MODEL 2 **Interpreting mass data derived from mass spectrometry**

Base	Monomer	$m_{tetramer}/4$	$m_{10\text{-}mer}/10$	$m_{100\text{-}mer}/100$
A	331.2	318.5	315	313.4
T	322.2	308.7	306	304.4
C	307.2	293.6	291	289.4
G	347.2	333.7	331	329.4
Ave	327.0	313.6	310.7	309.1

1. Why does the mass (in Da) change as a function of fragment length? Which set of mass data is most similar to the masses noted in the paper?

2. The following data was available from a MS sequencing reaction of Sanger fragments. Determine the sequence.

Signal	m/z	Δm/z	Base
Primer	6417.6	-------	-------
1	6731.9		
2	7035.9		
3	7340.9		
4	7669.9		
5	7960.6		
6	8274.0		
7	8587.2		
8	8892.2		
9	9205.3		
10	9495.1		

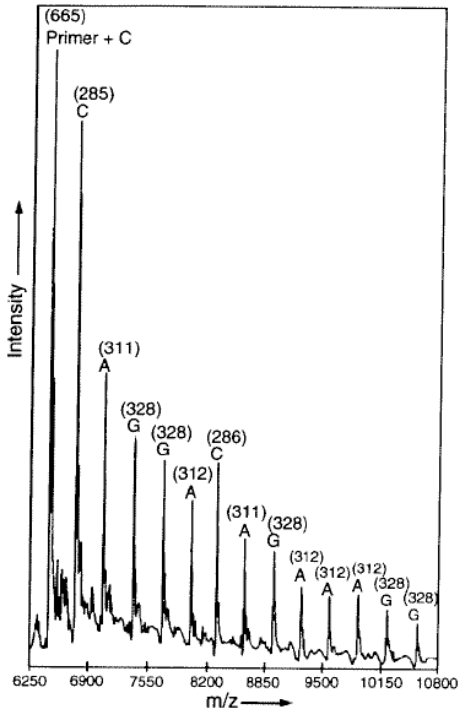
3. What two adjoining bases would lead to the minimum resolution? As such, which base might you choose to terminate the primer if you had a choice?

4. Approximately how many DNA bases are in a 6417.6 Da oligonucleotide? Does this seem appropriate for a primer sequence?

5. The resolution of a mass spectrometer is given by the equation: $Rs = \frac{m}{\Delta m}$. What is the resolution required between peaks 1 - 2 and between peaks 9 - 10? Which is more difficult to resolve?

MODEL 3 Mass Spectral Data

Using the peak widths and m/z values for the last G-G pair in the mass spectrum below, provide a rough estimate of the mass spectrometer's resolution.



2. What is the minimum required resolution between the last two G peaks.

3. Could this MS resolve a 100-mer oligonucleotide at the largest mass difference in bases? At the smallest? Hint: Remember that when you calculate resolution with Δm and m (as opposed to using peak separation and peak width observed), it is the resolution required, NOT the resolution observed.

4. Why does each peak have a width of several Da?

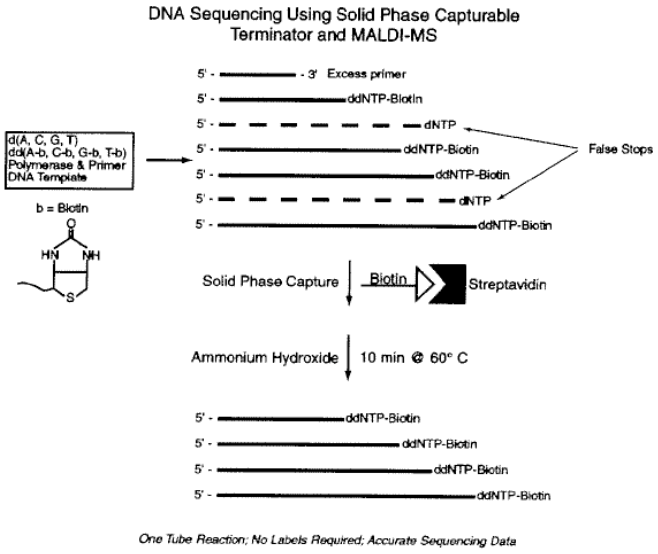
5. Edwards et. al. claim that fragmentation of DNA limits the maximum detectable range of nucleotides by MALDI-TOF. Is this evident in the MS above? Explain what they mean.

6. The authors claim that false stops lead to extension differences with a mass difference of 16 Da less that the correctly terminated fragment. Explain why.

MODEL 4 SNP and mutation analysis by MS

1 Edwards et. al. claim that excess primer and false stops make MS sequencing more difficult. Explain why.

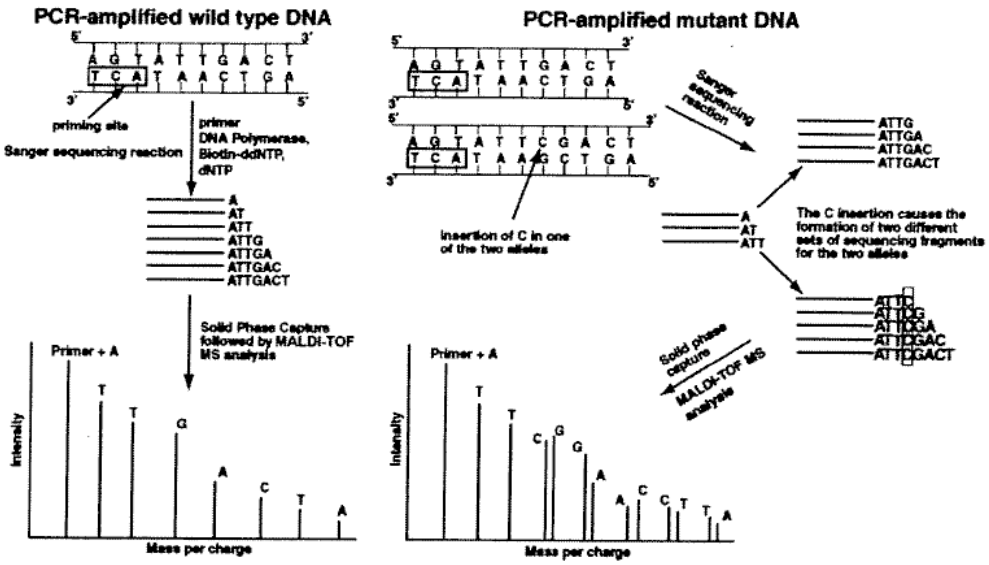
2. In the figure below, explain how biotin and strepatvidin remove the excess primer and false stops.



3. In the MS sequencing experiment outlined in the diagram below, explain why the initial peaks are single peaks and after a point, all peaks become double peaks.

4. How might such information be useful as a diagnostic tool for geneticists?

5. The authors claim that RNA is subject to less fragmentation and lower interaction with salts. Explain why each of these features might make MS sequencing of RNA more attractive than MS DNA sequencing. One problem is that the mass difference between C and U is only 1 Da. How can this problem be circumvented?



5. The figures below compare capillary electrophoresis sequence data with MS sequence data. The authors claim that the MS data is superior. Explain why. Could CE be used differently to remove this problem? What are the advantages and disadvantages of CE and MS in this application?