

Analysis of Nucleic Acids

A. Description of Size / Length

Usually measured as length in bases (nucleotides)

1 Kb = 1,000 bases (pairs); 1 Mb = 1 million bp

Early studies – Sedimentation Rate During Centrifugation

S = Svedberg Coefficient

Determined by mass, density, shape

Used for nucleic acids (e.g. rRNAs)

And protein complexes (e.g. proteasome)

Higher S value = Larger size/mass

B. Sizes of RNA Molecules

rRNAs: Several types in bacterial ribosome:

5S (120 b); 16S (1540 b); 23S (2900 b)

tRNAs: ~30 Different Types; Quite small (75 - 90 b)

mRNAs: Many Different Types in a Single Cell

Variable in Size (100 b - 10 kb)

Depends on size of protein encoded

C. Sizes of DNA Molecules

Usually measure genome instead of chromosome size

Viruses: 1 Kb - 1 Mb

Bacteria: 1 - 10 Mb

Plants: 100 Mb - 100 Gb

Animals: 100 Mb - 100 Gb

Humans: 3 Gb (3 billion base pairs)

Eukaryotes: More DNA than needed to code for protein

Much of this DNA is found between genes

Humans: Large introns within genes

DNA is highly condensed in chromosomes

Humans: >1 m DNA uncoiled; Nucleus: 10^{-6} m diameter

D. Estimating [DNA] in Cells and Solutions

Cells: *Stain with fluorescent dyes that bind to DNA*
 Intensity of fluorescence proportional to [DNA]

Solutions: *Measure absorbance at 260 nm*

E. Separation of Nucleic Acids

Sedimentation Equilibrium Centrifugation

DNA moves to position equal to buoyant density
Final position in tube determined by:
 Relative proportion of G:C and A:T pairs
 Physical nature of DNA (ss/ds; linear/circular)

Sedimentation Velocity Centrifugation

Measure rate of movement (analytical centrifuge)
Used to calculate S coefficient of macromolecules

Gel Electrophoresis

Separate nucleic acids in gel matrix (electric field)
Negative charge of phosphates directs movement
Small molecules move faster than large molecules
Fragments of different sizes separate into bands
Cut out fragments or transfer to filters for analysis

F. Denaturation and Renaturation

ds Nucleic acid <----> ss Nucleic acid

Denaturation / melting:

Double stranded molecule becomes single stranded
H bonds between strands broken

Renaturation / hybridization / reassociation:

Single stranded molecules pair; form double strands
H bonds form between complementary strands
Specificity determined by Temperature, [salt]

Hybridization with Labeled ss Nucleic Acid

*Identify complementary sequences in cells / solutions
Use 32-P or fluorescent label attached to nucleic acid*

Examples of Hybridization Methods

*in situ: Locate sequences in intact chromosomes
FISH Fluorescent in situ hybridization
Southern: Locate DNA fragments separated in gel
Northern: Locate RNA molecules separated in gel*

G. Reassociation Kinetics

*Goal: Measure RATE of spontaneous reassociation of
denatured DNA fragments in solution*

Result: Information on sequence abundance in genome

*Method: Cleave DNA ---> Many ds DNA fragments
Random shearing or enzyme digestion
Denature DNA in solution (ds --> ss)
Allow spontaneous hybridization
Measure rate of reassociation with time*

Unexpected result with eukaryotic genomes:

*Significant portion of genome hybridizes rapidly
Some DNA sequences are present in many copies*

How Often Should Random Sequence Appear in Genome?

Calculations: Assume: [A + T] = [G + C]

*Example: ---- C A T ---- Answer: $(1/4)^3 = 1/64$
 ---- G T A ----*

*Example: --- C A T C A T C A T C A T C A T ---
 --- G T A G T A G T A G T A G T A ---*

Answer: $(1/4)^{15} = 1/1,073,741,824 = 1/10^9$

Perhaps once or twice in human genome

Conclude: Short Sequences can Provide Unique Information

Sequence Classification Based on Abundance in Genome:

Highly Repetitive

Usually short sequences (< 100 bp)
Often rich in A-T base pairs
Up to 1 million copies / genome
Common in centromere and telomere regions
Often tandemly duplicated (adjacent copies)
Not transcribed (not genes)
Not essential yet maintained

Moderately Repetitive

Often long sequences (100 – 1,000 bp)
Typically 10 to 1,000 copies / genome
Some tandemly duplicated; others dispersed
Variety of functions; some transcribed
Some genes (e.g. rRNA genes)
Transposable elements
Regulatory sequences

“Unique” Sequences

Few copies / genome
Includes most genes
Focus of modern genetics