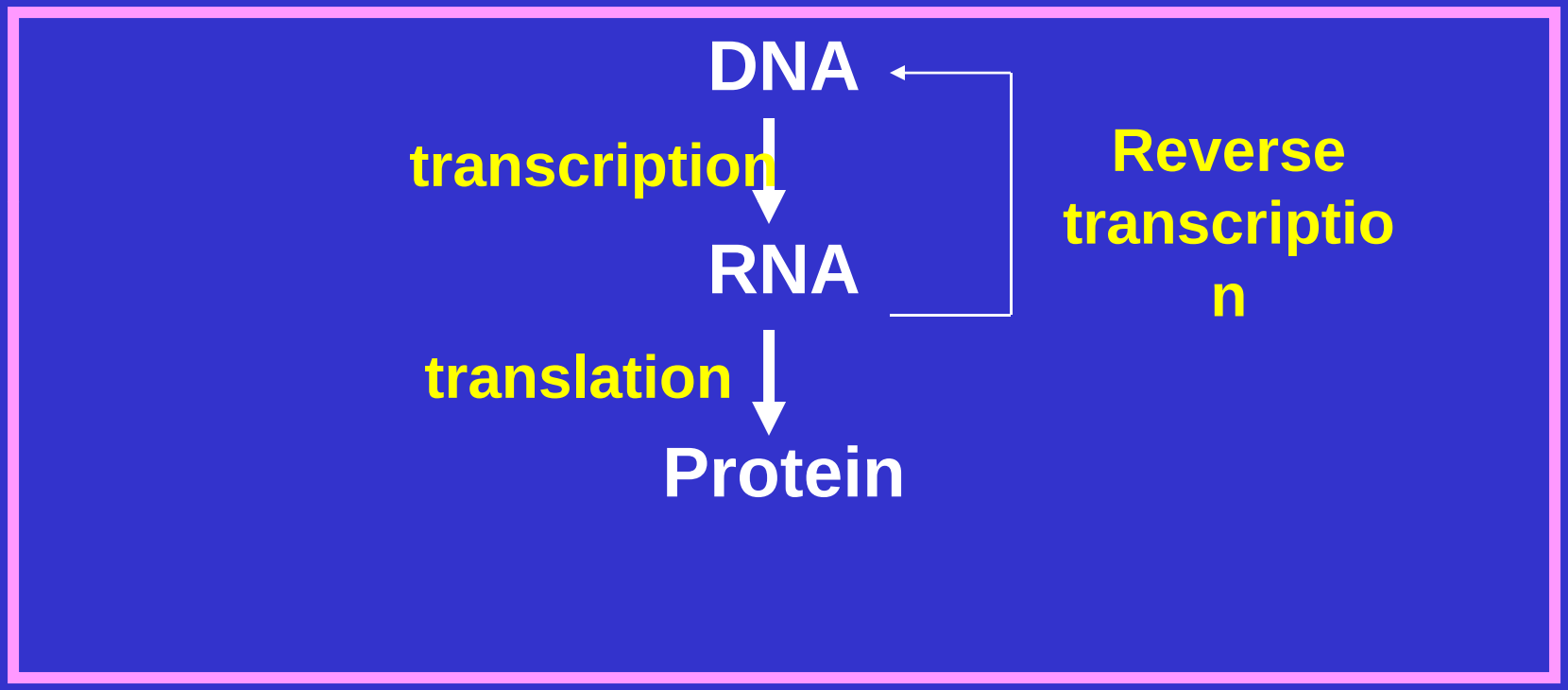
The Genetic material: DNA

The Central Dogma of Genetics



- A, T, G, C in DNA
- A, U, G, C in RNA
- DNA is double stranded
- DNA has polarity 5’ to 3’
- A T base pair, G C base pair
- RNA is single stranded, also has polarity, generally referred as upstream and downstream. In RNA: A U base pair G C base pair. This type of base pairing in RNA causes secondary structure.

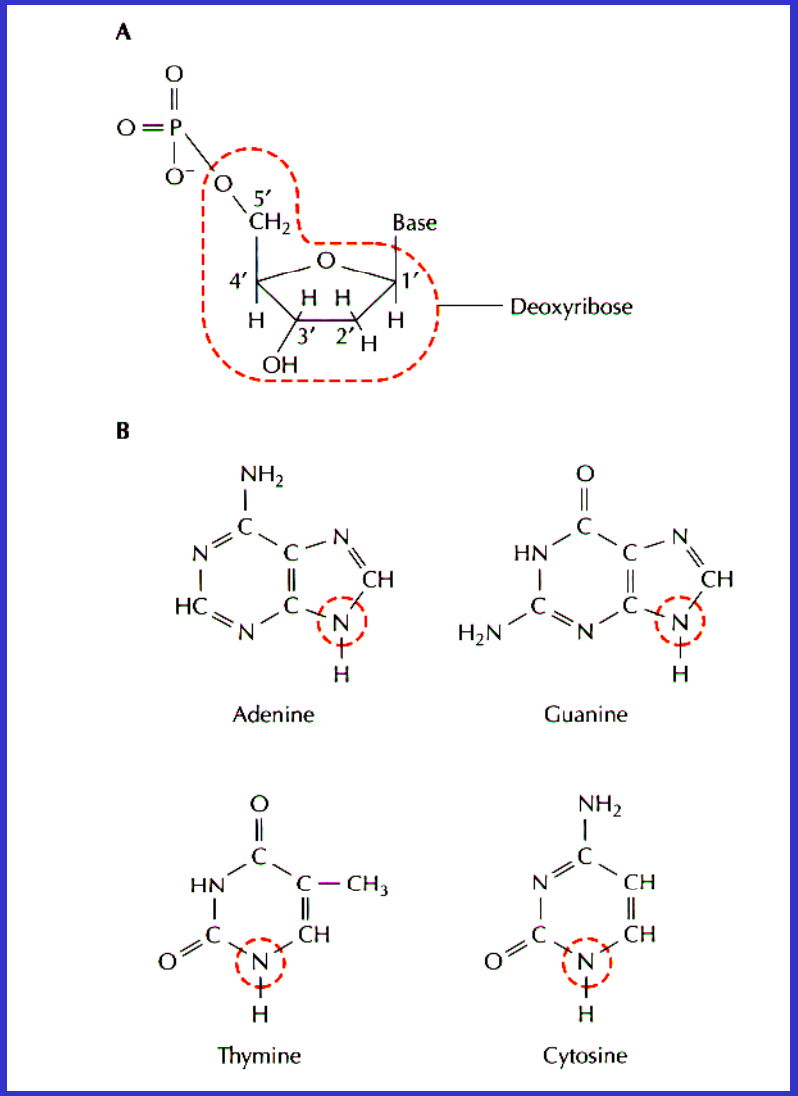
1. 5’ vs 3’

2. Purines vs. Pyrimidines

3. A vs. G

4. C vs. T

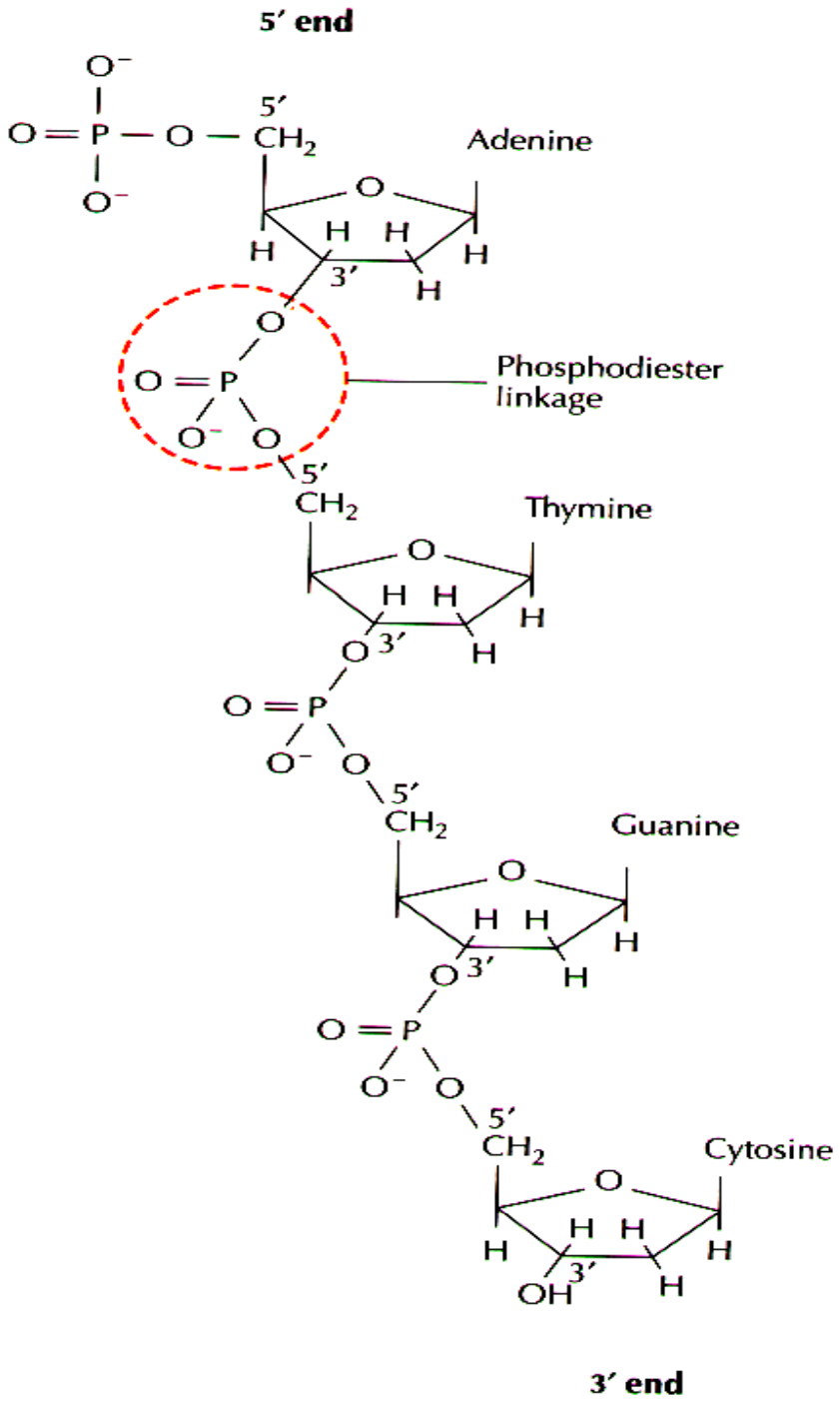
5. Transitions vs. transversions

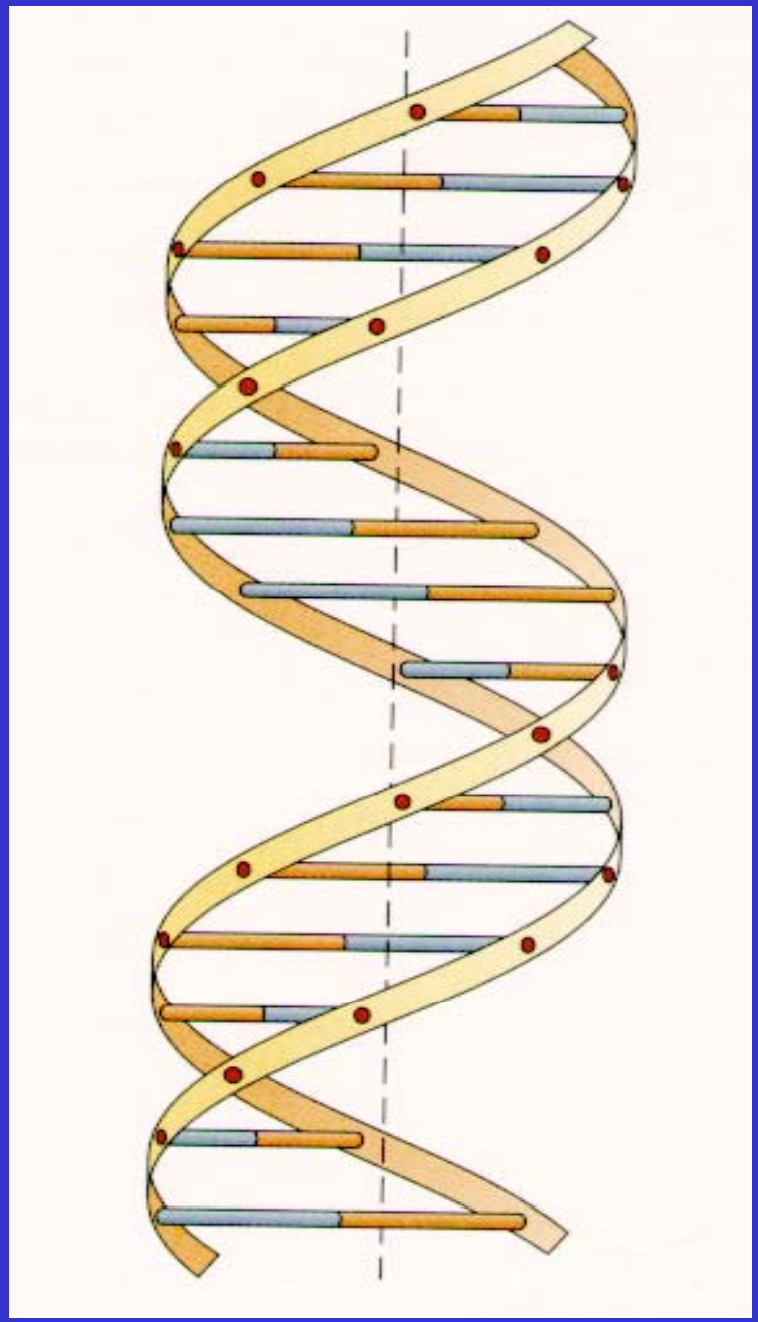


AT/GC ratios and their applications

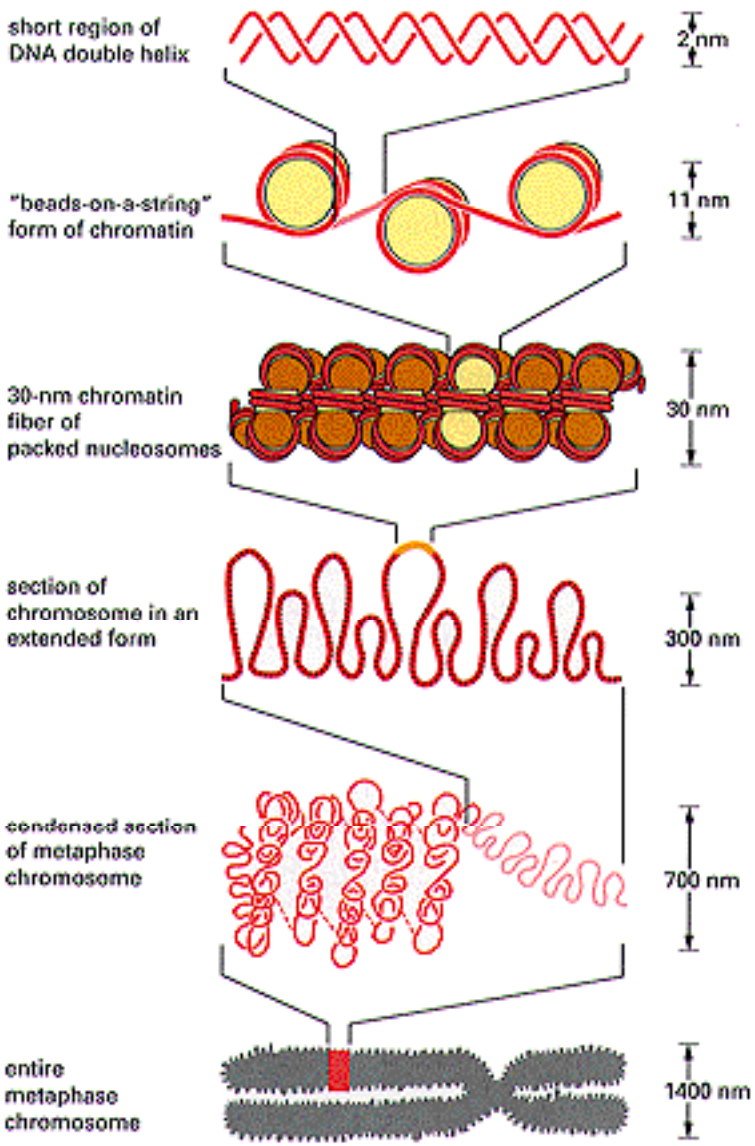
- Genome composition and characterization
- Implications in sequencing
- Primer design
- PCR yields
- …

The Polarity of DNA





Higher order organization of genomes



Most chromosomal DNA does not code for proteins or RNAs:

e.g.,

Human genome 3 billion base pairs

25,000 genes x 2,000 bp per gene

= 5×10^7 bp

$5 \times 10^7 / 3 \times 10^9 = 1.67\%$

Molecular Probes:

The tools of molecular genetics

Concept of probes

- **“For diagnostic tests, the agent that is used to detect the presence of a molecule in the sample”.**
- **“A DNA sequence that is used to detect the presence of a complementary sequence by hybridization with a nucleic acid sample”.**

Need for probes

- Screen for the gene of interest
- Southern blot to understand genomic structure and gene copy numbers
- Northern blot for analysis of RNA expression
- Verification of allelic amplification in PCR
- ...

Making ds DNA probes

- 5’ end labeling
  - Polynucleotide Kinase
- 3’ end labeling
  - Fill in a restriction enzyme site
  - Terminal transferase labeling
- Continuous labeling
  - Nick translation
  - Random primer labeling

Criteria for a good probe

- Size of the probe
- Specific activity of the probe
- Yield of the probe

Enzymatic reaction

- Substrate
- Environment (pH etc.)
- Enzyme

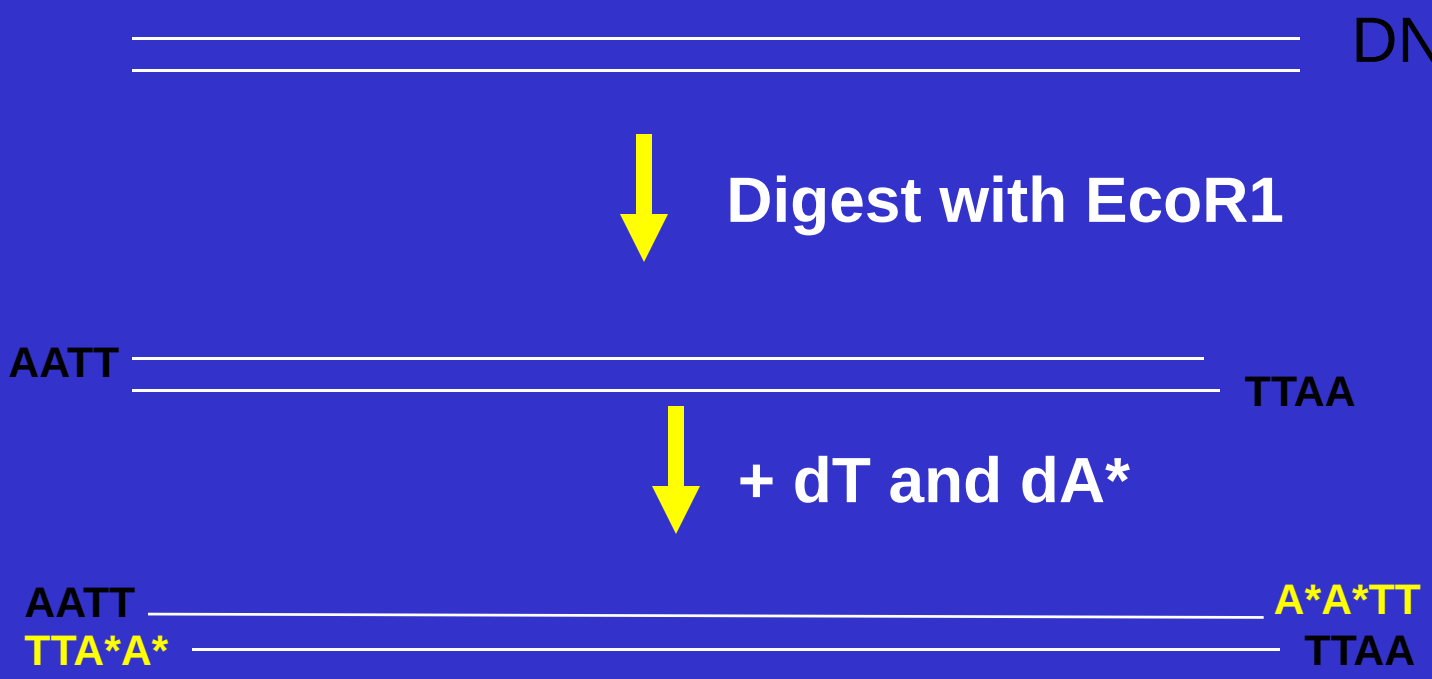
5’ end labeling

- Substrate: DNA, γ-P*-ATP
- environment: buffer
- Enzyme: polynucleotide kinase
- Can label both double-stranded DNA and single-stranded DNA.
- Good for oligonucleotide labeling
- Can label RNA as well
- Do not need template

3’ end labeling by DNA polymerase filling in restriction site

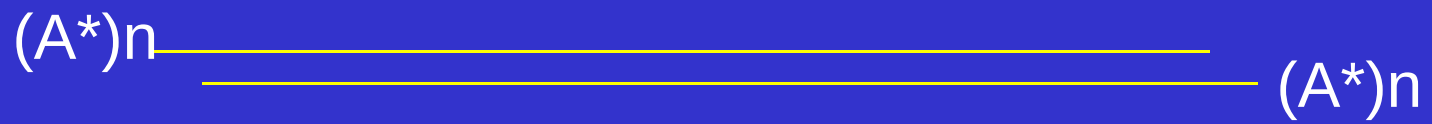
- Substrate: DNA, dNTPs (with one of them labeled)
- Buffer
- Enzyme: polymerase (T4 DNA polymerase, DNA polymerase I Klenow fragment; Taq polymerase
- Template required

3’ end labeling by DNA polymerase filling in restriction site



3’ end labeling by terminal transferase

- **Substrate: DNA, dNTPs (labeled)**
- **Buffer**
- **Enzyme: terminal transferase**
- **Template NOT required**



Continuous labeling

- Substrate: DNA + dNTPs
- Buffer
- Making the new strand DNA with one (or more) labeled dNTPs:

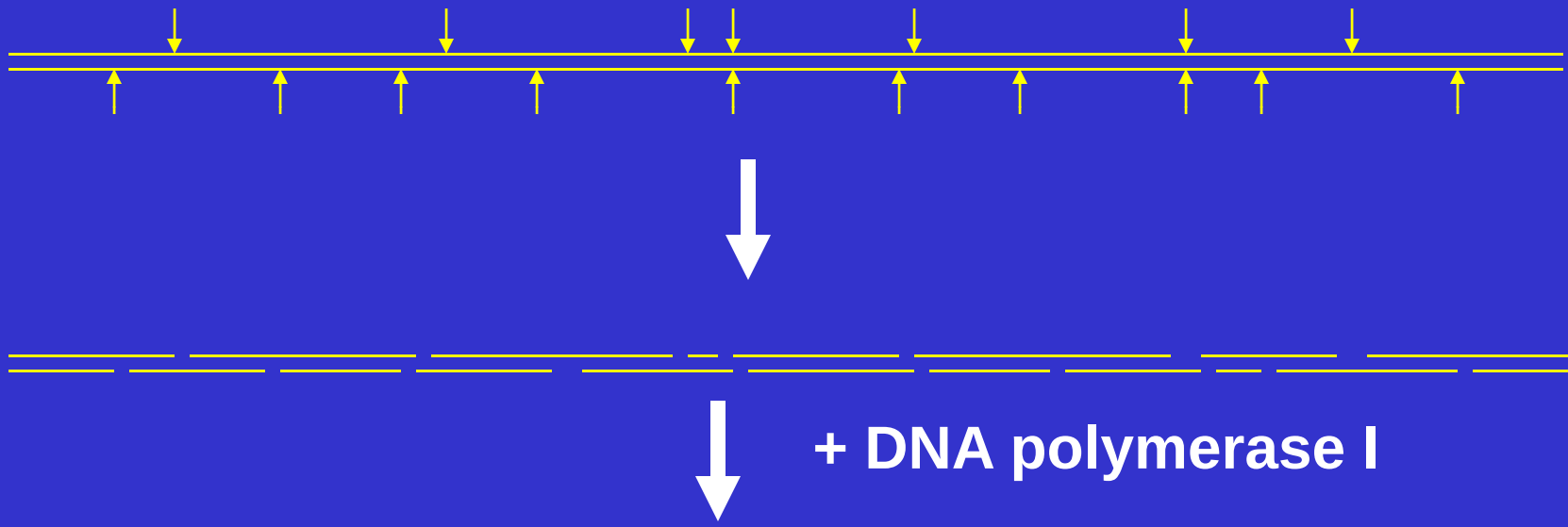
1. Nick translation;
2. Random primer labeling

Nick translation:

- **Making nicks on double stranded DNA with DNase I**
- **DNA polymerase I has two major activities: 5’ to 3’ exonuclease activity and 5’ to 3’ polymerase activity**
- **DNA polymerase I making the new strand DNA with labeled dNTP while degrading the old strand of the DNA.**

Nick translation

ds DNA + DNase I



+ DNA polymerase I

Synthesizing 5’ to 3’ while degrading the original strand

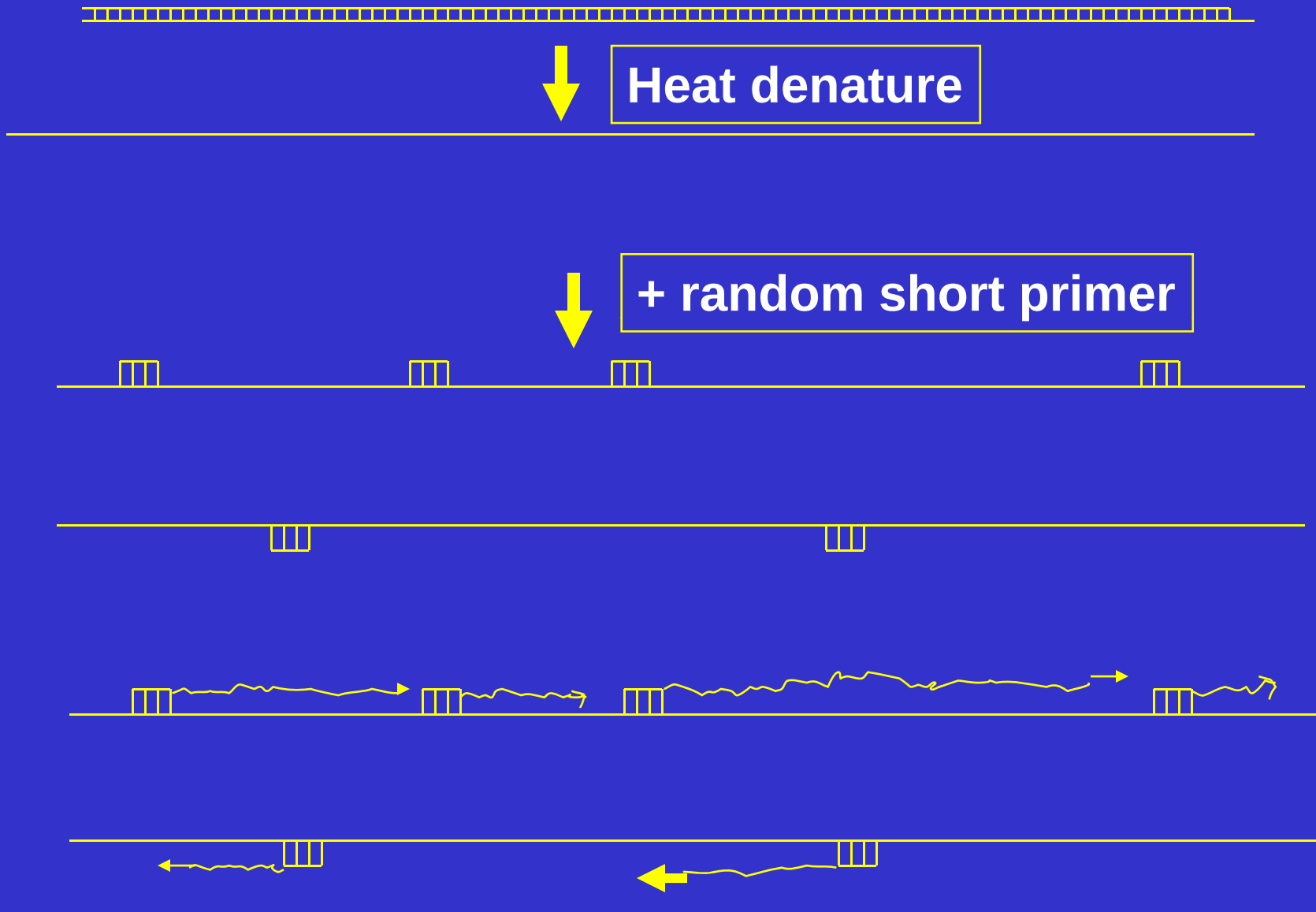
Factors to consider for nick translation

- Concentration of DNase I
- Temperature of incubation (16C)
- Concentration of template DNA

Random primer labeling method:

- Heat denature DNA
- Annealing of short random primers to ss DNA
- Klenow polymerase makes the new strand with labeled dNTP
- DNA synthesis continue until it reaches the next primer.

Random primer labeling



Factors to consider for random primer labeling

- Concentration of random primers
- Concentration of template DNA
- Primer template radio

Labeling PCR products

- Label PCR primer
- Label during PCR with one of the dNTPs labeled.
- Label 3’ end by dATP at the end of PCR
- Label 5’ end afetr PCR by kinase
- Nick translation
- Random primer

Single-strand probes

- DNA probes using phagemids
- RNA probes using in vitro transcription:
  - T3 RNA polymerase
  - T7 RNA polymerase
  - Sp6 RNA polymerase
- DNA probe using reverse transcriptase

Oligonucleotide probes

1. You know amino acid sequences, need to obtain DNA sequences

2. Only a segment is conserved among genes in different species known to date, and you are interested in isolating the gene from your species of interest.

How long my probe should be?

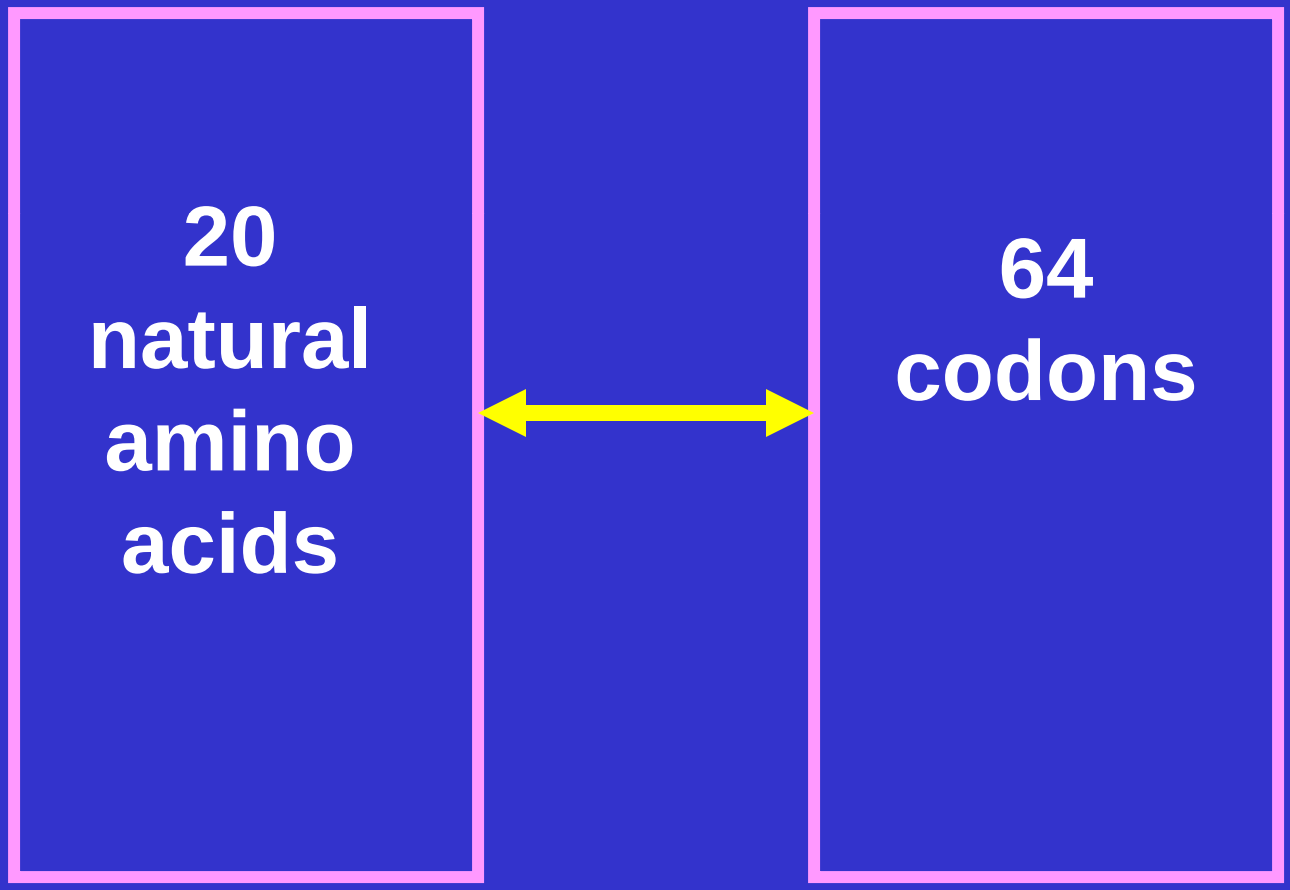
Uniqueness of sequences:

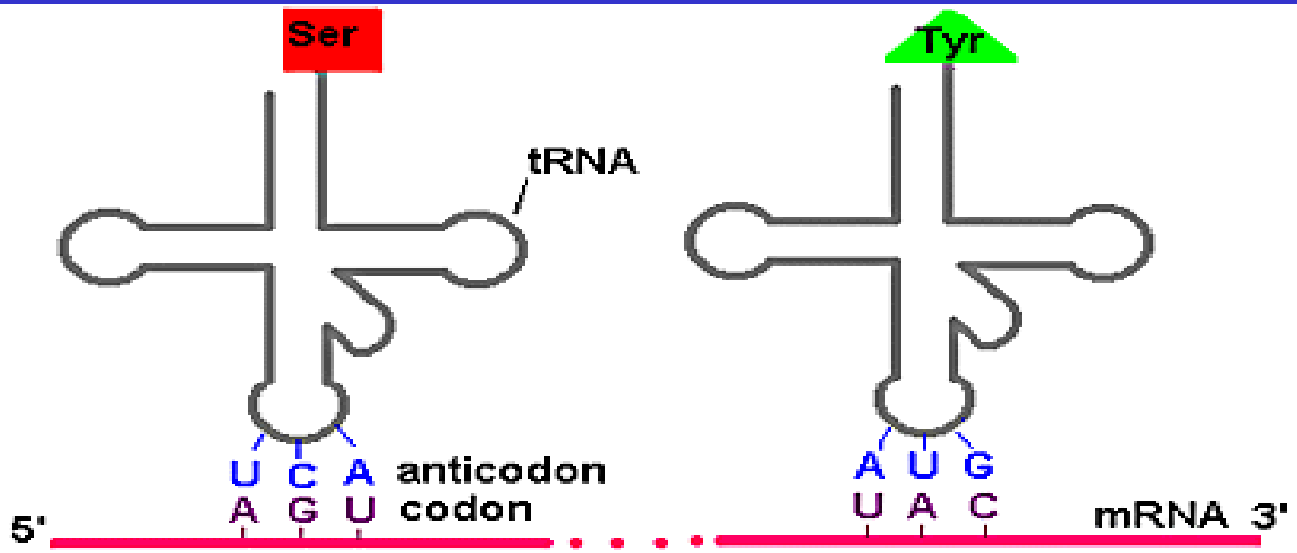
At each base position, there are four possibilities (A, C, G, T)

The probability for any given sequence to exist in a genome is

$$1/4^n$$

Degeneracy of the genetic code

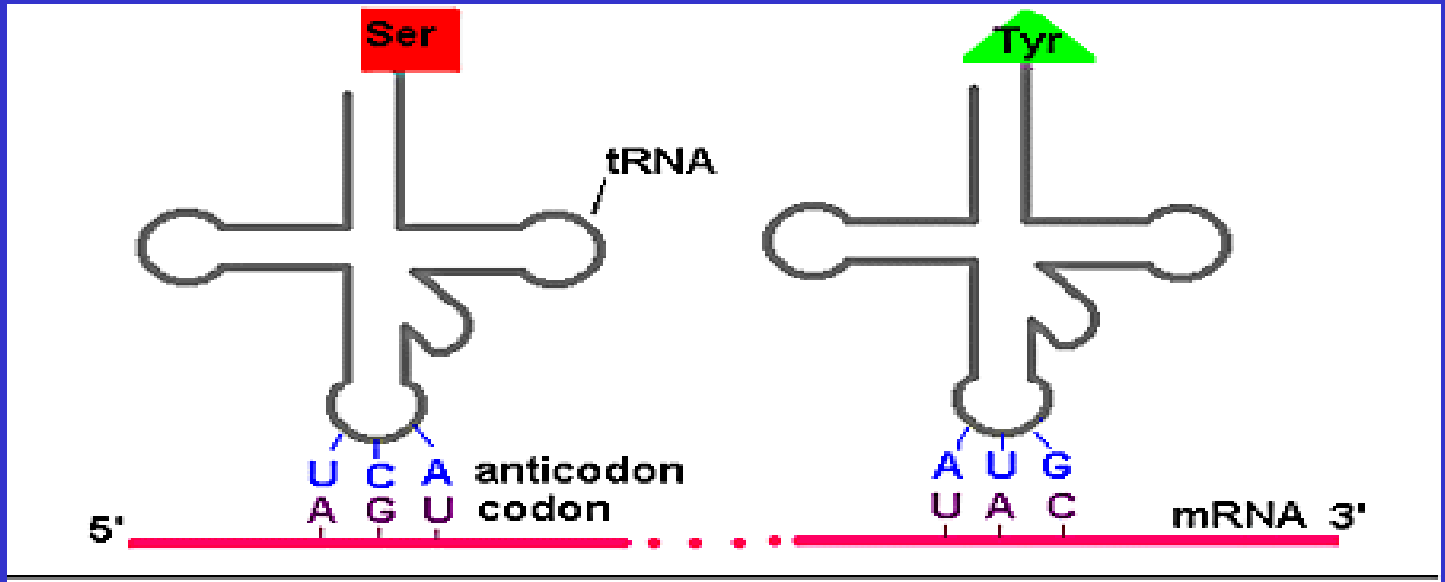




2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

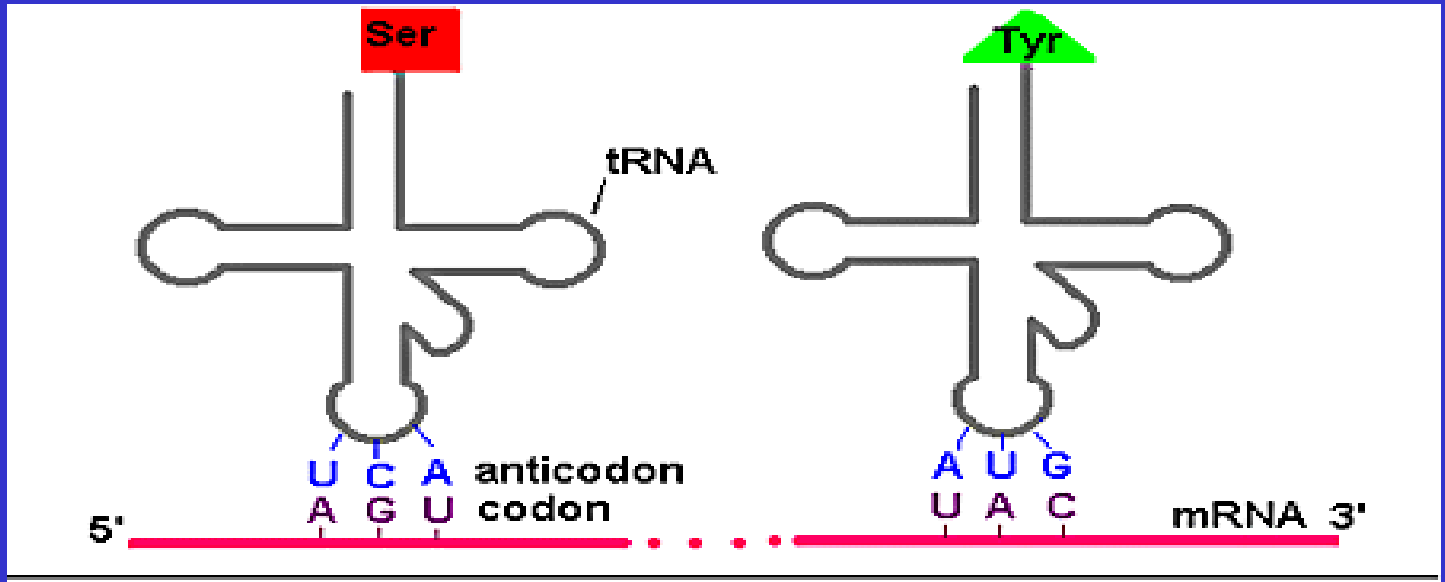
The Genetic Code



2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The Genetic Code



2nd base in codon

1st base in codon	U	C	A	G	3rd base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

The Genetic Code

Exceptions to the genetic code table

- http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Codons.html#Exceptions

Exceptions to the Code

The genetic code is almost universal. The same codons are assigned to the same amino acids and to the same START and STOP signals in the vast majority of genes in animals, plants, and microorganisms. However, some exceptions have been found. Most of these involve assigning one or two of the three STOP codons to an amino acid instead.

Mitochondrial genes

When mitochondrial mRNA from animals or microorganisms (but not from plants) is placed in a test tube with the cytosolic protein-synthesizing machinery (amino acids, enzymes, tRNAs, ribosomes) it fails to be translated into a protein.

The reason: these mitochondria use UGA to encode tryptophan (Trp) rather than as a chain terminator. When translated by cytosolic machinery, synthesis stops where Trp should have been inserted.

In addition, most animal mitochondria use AUA for methionine not isoleucine and all vertebrate mitochondria use AGA and AGG as chain terminators. Yeast mitochondria assign all codons beginning with CU to threonine instead of leucine (which is still encoded by UUA and UUG as it is in cytosolic mRNA).

Plant mitochondria use the universal code, and this has permitted angiosperms to transfer mitochondrial genes to their nucleus with great ease.

Nuclear genes

Violations of the universal code are far rarer for nuclear genes. A few unicellular eukaryotes have been found that use one or two (of their three) STOP codons for amino acids instead.

Nonstandard Amino Acids

The vast majority of proteins are assembled from the 20 amino acids listed above even though some of these may be chemically altered, e.g. by phosphorylation, at a later time. However, two cases have been found where an amino acid that is not one of the standard 20 is inserted by a tRNA into the growing polypeptide.

Selenocysteine. This amino acid is encoded by UGA. UGA is still used as a chain terminator, but the translation machinery is able to discriminate when a UGA codon should be used for selenocysteine rather than STOP. This codon usage has been found in certain Archaea, eubacteria, and animals (humans synthesize 25 different proteins containing selenium).

pyrrolysine. In one gene found in a member of the Archaea, this amino acid is encoded by UAG. How the translation machinery knows when it encounters UAG whether to insert a tRNA with pyrrolysine or to stop translation is not yet known.

Considerations for mixed oligo probes

- # of total oligo in the probe = # of degeneracy. E.g.

Met-Gln-Ile-Gly-Leu = 15 mer

1 x 2 x 3 x 4 x 6 = 144 oligos

Antibody probes

- Antibodies recognize its protein antigen
- You can express the antigen in expression libraries
- Screening a DNA expression library with antibodies

How to select probes

- You have the gene cloned: ds DNA
- Someone cloned the gene
- Someone cloned the gene from cattle, but I am working with pig
- There are only a short region that is evolutionarily conserved: Oligonucleotide or Guessmers
- I only know the protein sequence: Mixed oligonucleotide or antibodies

Reference books

- Sambrook et al., 1989, Molecular cloning, a laboratory manual:

  Chapter 10, Preparation of radiolabelled DNA and RNA probes

  Chapter 11, Synthetic oligonucleotide probes

  Chapter 12, Screening expression libraries with antibodies and oligonucleotides
- Cureent protocols in molecular biology, ed. Ausubel et al., 1987, John Wiley and Sons;
- Nucleic acid hybridisation, ed. Hames and Higgins, IRL press, 1986.