

## Biology 2250 - Principles of Genetics

---



Dr. Steven M. Carr  
 Department of Biology  
[Memorial University of Newfoundland](#)  
 St. John's NL A1B 3X9, Canada  
 e-mail: [scarr@mun.ca](mailto:scarr@mun.ca)

---

Click on the following for: [Bio2250 Lab Manual](#) v.05c (updated **07 Sept 2009**)  
[Bio2250 Lectures Notes](#) (through first midterm) (Updated **16 Sept 2011**)

Laboratory Investigation #1 (for [weeks of 19 & 26 Sept](#))

Please **print out & review lab instructions** before coming to lab

---

[Bio2250 Lecture & Lab schedule](#) (updated **04 Sept 2011**)

[Orientation to 2250](#) (*Please read !!*)  
[How to use the website](#)

Required text: [Russell \(2010\) iGenetics: A Molecular Approach, 3rd ed.](#) [Igen3]

[Replaces Griffiths *et al.* 2002 [Modern Genetic Analysis, 2nd ed.](#)] [MGA2]

[Assigned readings & suggested homework](#) from iGen3 (updated **20 Sept 2010**)

[Online practice problems](#) from MGA-2

[Sample Quiz & Exam questions](#) (updated **17 Sept 2010**)

Other webpages of interest:

[Genetic Research in my lab](#)

[Bio2900 \(Principles of Systematics & Evolution\)](#) (from **Winter 2001**)

[Bio4241 \(Advanced Genetics\)](#) (next offered **Winter 2011**)

[Bio3950 \(Fundamentals of Genetic Biotechnology\)](#) (offered **Spring 2011, Fall 2012**)

[Genetics in the News](#) (updated **17 Sept 2009**)

Rap version of the [Krebs cycle](#)

*T. rex struttin'* to '[Staying Alive](#)'

Activities of [BIOS](#) (MUN Biology Society)

Click here to [e-mail me](#) questions, comments, or suggestions.

Please include '**2250**' in the subject line

---

Lectures: Tu Th 0900-1015 **Sn-2109** (Science Lecture Theatre)

Labs: M Tu W Th 1400-1700 **Sn-4110** (Genetics & Evolution Laboratory)

[No labs on 9.10 Sept]

Course lecture notes:

These notes are revised before & after lectures; check frequently for revisions.

Topic	Last Revised	Lecture Date
<a href="#">History of the discovery of DNA</a>	<b>16Sept 2011</b>	-----

1	<a href="#">Structure of DNA: the Hereditary Molecule</a>	03 Sept 2011	08 Sept 2011
2	How Genes Work I: <a href="#">DNA Replication &amp; Transcription</a>	13 Sept 2011	13 Sept 2011
3	<a href="#">The Genetic Code</a>	13 Sept 2011	15 Sept 2011
4	How Genes Work II: <a href="#">RNA Translation</a>	15 Sept 2011	20 Sept 2011
5	How Genes Work III: <a href="#">Protein Structure &amp; Function</a>	16 Sept 2011	22 Sept 2011 <a href="#">Practical Genetics, 23 Sept 2010</a>
	<a href="#">Molecular Basis of Heredity</a>	16 Sept 2011	27, 29 Sept 2011
6	Chromosome Genetics <a href="#">I: Cytogenetics</a>	18 Oct 2010 18 Oct 2010	04 Oct 2011
	<a href="#">II: Genome organization</a>		TBA
	<b>Midterm Exam I (topics 1-5)</b> <i>Review TBA</i> [Sn-2109]	<b>Sample</b>	<b>06 Oct 2011</b>  <b>Review, 5-6 P</b> <b>Wednesday, 05 Oct</b>
	Thanksgiving (No lecture)		<i>11 Oct 2011</i>
7	<a href="#">Mendelian Genetics: Dominance, Segregation, &amp; Assortment</a> <a href="#">Extensions to Mendelian Analysis</a>	14 Oct 2009	13 Oct 2011
		25 Oct 2010	18 Oct 2011
			20 Oct 2011
8	<a href="#">Pedigree Analysis</a>	04 Nov 2010	25 Oct 2011
9	<a href="#">Chromosome Linkage Recombination &amp; Mapping</a>	04 Nov 2010	27 Oct 2011 01 Nov 2011
	<b>Midterm Exam II (topics 6-9)</b> <i>Review TBA</i> [Sn-2109]		<b>08 Nov 2011</b>
	Remembrance Day		<i>10 Nov 2011</i> <i>no lecture 11 Nov 2010</i>
10	<a href="#">Molecular Basis of Mutation</a>	17 Nov 2010	03 Nov 2011
11	<a href="#">Genetic Engineering &amp; Biotechnology</a>	19 Nov 2010	15 & 17 Nov 2011
12	<a href="#">Genomics, &amp; Bioinformatics</a>	25 Nov 2010	22 Nov & 24 Dec 2011
13	Genetics & Genomics Research at Memorial U		29 Nov & 01 Dec 2011
	<b>Final Exam (Cumulative)</b>	Reviews: TBA	<b>TBA</b> <b>[Sn-2109]</b>

## History of the hereditary molecule (to 1953)

---

### In principle:

"**Genetics**" was taught for **50 years**

*without* knowledge of the hereditary **substance** or its **structure**

(see [Orientation to Bio2250](#))

The story of the search for the hereditary substance includes superb examples of the **experimental method** in biology.

---

### Two candidates: protein *versus* nucleic acid

Cells contain **H<sub>2</sub>O**, **lipids**, **carbohydrates**, *and* ...

**Mulder** (1838) - Discovery of **protein**

Abundant, water-soluble, **nitrogenous**

*"... complex... regulates cell metabolism...  
most important component of living matter...  
without it, life would not be possible"*

Hydrolysis of protein ⇒ **amino acids** (~20 kinds)

**Miescher** (1868) - Discovery of **nuclein**

Found in cell nucleus, acidic, rich in **PO<sub>4</sub>**,

Lacks **S** (characteristic of protein)

Now know this as **nucleic acid**

**Levene** (1910) - **Tetranucleotide hypothesis**

**nucleic acid** is a repetitive polymer of four **bases**

**A:C:G:T** in the approximate ratio **1:1:1:1**

⇒ Structure seems too simple to carry information

**Griffith** (1928) - **transforming principle**

Killed virulent viruses '**transform**' live avirulent viruses:  
avirulent viruses become virulent, and

[Transformation is inherited](#)

⇒ Hereditary makeup of organisms can be altered

**Avery, MacLeod, & McCarty** (1944) -

**Chemical isolation** of '[transforming principle](#)' from cells

Transformation survives **protease** treatment,

destroyed by **nuclease** treatment (**Homework**):

⇒ It's chemically pure **deoxyribonucleic acid (DNA)** [?!?!](#)

**Hershey & Chase** (1952) - '**blender experiment**'

**Bacteriophages** are grown in [radioactive medium](#)

**Proteins** labeled with <sup>35</sup>**S**

**DNA** labeled with <sup>32</sup>**P**

During infection of *E. coli* by bacteriophages,

[<sup>32</sup>P goes in, <sup>35</sup>S stays out](#)

⇒ **DNA** is the transforming principle

**Watson & Crick** (1953) "**The Double Helix**"

**Schrodinger** (1945) "**What is Life?**":

Are there "other laws of physics?"

**Franklin & Wilkins' X-ray crystallography**

DNA is a helix: two or three strands?

**Chargaff's Rules** : Bases are *not* equimolar, but

[A]=[T] & [C]=[G] ([Table](#))

**Model building:**

Two or three strands, bases inside or outside

Key discovery: A+T pair looks like C+G pair

The **Watson-Crick structure** for DNA

double-stranded helix ([3-D image](#))

Phosphate backbone *outside*

Nitrogenous **bases** *inside*

**H-bonds** hold strands held together

---

**Homework:**

- 1) What is Avery's "**Assay for Transformation**"? How does it relate to the presence of **R** and **S** cells in the cultures on the third line above? How does Avery's "assay" differ from Griffith's?
  - 2) Predict the results of the Hershey-Chase experiment if they had used radioactive <sup>14</sup>C instead of <sup>32</sup>P
- 

**For further reading:**

- J. Cairns, G. Stent, & J. Watson (1966). **Phage and the Origins of Molecular Biology**. Freeman.  
[Biographical essays on the early days by the founders of molecular genetics.]
- F. H. C. Crick (1988). **What Mad Pursuit?** Basic Books.  
[Crick's version of the 'double helix' history, and lots more.]
- L. Gonick & M. Wheelis (1991). **The Cartoon Guide to Genetics**, 2nd ed. Harper Collins.  
[A well-illustrated, entertaining primer of basic Mendelian and molecular genetics for non-biologists.]
- H. F. Judson (1979). **The Eighth Day of Creation**. Simon & Schuster.  
[A general history of molecular biology.]
- A. Sayre (1975). **Rosalind Franklin and DNA**. Norton.  
[A re-appraisal of the role of Franklin, with commentary on the role of women in science.]
- J. D. Watson (1968). **The Double Helix**. Atheneum.  
[An entertaining, irreverent, sexist, account of the discovery of the structure of DNA.  
See the accounts of Crick and Sayre for another view]
- J. D. Watson (2003). [DNA: The Secret of Life](#). Knopf  
[A narrative history of genetics and molecular biology in the 20<sup>th</sup> century,  
written for the 50<sup>th</sup> anniversary of the discovery of the DNA structure.]
-

## Biochemistry of heredity: the structure of Deoxyribonucleic Acid (DNA)

---

### In principle: Genes are made of nucleic acids

The identity of the hereditary substance was unknown until 1940;  
its structure was unknown until 1953

"Genetics" was taught for 50 years without this information (see [Orientation](#))

The [history of the discovery of DNA](#) is a fascinating detective story

The **Watson-Crick structure** for **Deoxyribonucleic acid (DNA)** (1953) (MGA2 Box 2-2, p.31)

a **double-stranded helix**

sugar-phosphate **backbone** outside

nitrogenous **bases** (A,C,G, T) inside

bases held together by **hydrogen bonds** (AKA **H-** or **hydrostatic** bonds)

**Fundamental insight:**

bases on alternative strands pair according to specific rules:

A with T G with C

each pair has similar structure

A second form of nucleic acid is **ribonucleic acid (RNA)**

### [Homework Assignment #1](#)

---

### Building blocks of nucleic acids (DNA & RNA)

#### bases

**pyrimidines** ([single ring](#))

**cytosine(C)** **thymine (T)** [ **uracil** in RNA (U) ] [[iGen3\\_02-08](#)]

"**PYR**amids were **CUT** from stone"

**purines** ([double ring](#))

**adenine (A)** **guanine (G)**

"**AG**s are **PUR**e"

**nucleoside** = [base + sugar](#)

**deoxyribose sugar** in **DNA** (- H on 2'-C) [[iGen3\\_02-07](#)]

**ribose sugar** in **RNA** (- OH on 2'-C)

**deoxyadenosine (dA)** **deoxyguanosine (dG)**

**deoxycytosine (dC)** **deoxythymidine (dT)**

**nucleotide** = nucleoside + phosphate(s) [**PO<sub>4</sub>**] [[iGen3\\_02-09a](#)]

in **DNA**,

one phosphate => **deoxynucleoside monophosphate (dNMP)**

three phosphates => **deoxynucleoside triphosphate (dNTP)**

**deoxyadenosine-5'-phosphate** or **deoxyadenylic acid**

deoxyadenylic acid (**dAMP**) / deoxyguanylic acid (**dGMP**)

deoxycytidylic acid (**dCMP**) / deoxythymidylic acid (**dTMP**)

**polynucleotide** = nucleotide + nucleotide + nucleotide + etc [[iGen3\\_02-09b](#)]

nucleotides are linked by **3' → 5' phosphodiester bonds**

\*\*\*polynucleotides have **directionality**\*\*\*

**hydroxyl (3')** & **phosphoryl (5')** ends

---

## Structure of B-DNA ([3-D model](#): requires [MDL chime](#) plugin) [[iGen3 02-12](#)]

- 1) Two **plectonemic** (twisted) **right-handed** polynucleotide helices ([demo](#))
- 2) Helices **antiparallel** strands wrt **5' → 3'** orientation [[MGA2-02-05](#)]
- 3) Strands held together by **hydrogen (H-) bonds** between bases
- 4) H-bonds form according to specific **base-pairing rules**

[A pairs with T](#): **two** H-bonds



[G pairs with C](#): **three** H-bonds



**A+T & G+C** pairs have [very similar shapes & sizes](#) [[iGen3 02-13](#)]

- 5) Base pairs **co-planar**: interval = **0.34 nM** [= **3.4 Å**ngstroms]
- 6) Period of helix is **10 bp (base pairs) = 3.4 nM**
- 7) **3-D** structure has **major & minor grooves** [[MGA2 02-07](#)]
- 8) Order of bases in each strand **aperiodic**

### [Homework Assignment #2](#)

### [Homework Assignment #3](#)

## [Other structures](#) for nucleic acids [[iGen3 02-14](#)]

[A-DNA](#) : not groovy, [base pairs not co-planar](#)

[Z-DNA](#): [left-handed](#) helix ([demo](#))

### Ribonucleic Acid (**RNA**):

substitute [uracil](#) for **thymine** [ **thymine = 5-methyl-uracil** ]

[ribose](#) sugar for **deoxy-ribose**

typically **single-stranded** or with **complex double-stranded** folding:

**mRNA** (**messenger RNA**): long, single-stranded

**rRNA** (**ribosomal RNA**): medium-sized, complex 'stem & loop' folding

**tRNA** (**transfer RNA**): small, 'cloverleaf' structure

[more on **RNA** structures later]

## Implications of DNA structure for its genetic function

*"The sequence of bases on a single chain does not appear to be restricted in any way.... It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."*

([Watson & Crick 1953. Nature 112:753](#))

DNA is an **aperiodic** crystal:

**order of bases** conveys **information**

**Antiparallel** strands are **self-complementary**:

DNA is potentially **autocatalytic**

All text material © 2011 by [Steven M. Carr](#)



## DNA Replication & Transcription

**In principle: DNA replication is semi-conservative**

H - bonds 'unzip', strands unwind,

complementary nucleotides added to [existing strands](#) [iGen3 03-02]

After replication, **each double-helix has one "old" & one "new" strand**

[note alternative **conservative** & **dispersive** models: [Homework #4](#)] [iGen3 03-01]

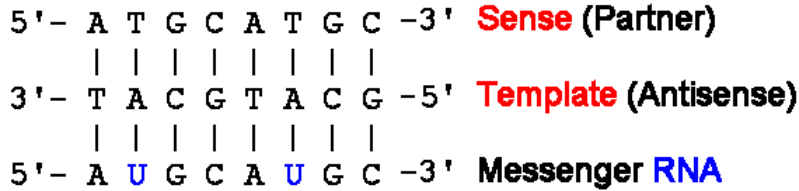
**DNA is *not* the "Genetic Code" for proteins**

information in DNA must first be **transcribed** into RNA

**messenger RNA** transcript is base-complementary to **template strand** of DNA

& therefore **co-linear** with **sense strand** of DNA

DNA & RNA syntheses occur **only** in the **5' → 3'** direction



**DNA synthesis in prokaryotes:**

Nucleotides are added simultaneously to both strands, *but*

**DNA grows in the 5' → 3' direction** **ONLY** [iGen3 03-03]

([online MGA 2 animation](#))

Distinguish:

**Replication:** duplication of a **double-stranded DNA (dsDNA)** molecule  
an exact 'copy' of the existing molecule (*cf.* **xerox** copy)

**Synthesis:** biochemical creation of a new **single-stranded DNA (ssDNA)** molecule  
a *base-complementary* 'copy' of an existing strand (*cf.* **silly putty** copy)  
occurs **only** in the **5' → 3'** direction

### [Homework #5](#)

**DNA Synthesis** in prokaryotes [iGen3 03-04, -05, -06]

- (1) Formation of **replication fork**  
provides two **single-stranded DNA template (ssDNA)**
- (2) Synthesis of **RNA primer**
- (3) Addition of **dNTPs** by **DNAPol III** at **3' end only**  
**continuous synthesis** on **leading strand**
- (4) **discontinuous synthesis** on **lagging strand**  
**Okazaki fragments**  
**proof-reading** by **3'→5' exonuclease** activity
- (5) Excision of **RNA primer** by **DNAPol I**  
**ligation** (connection) of fragment ends at gaps by **DNA ligase**

[A talkie animation of DNA synthesis](#) [onlineMGA2 animation]

DNA synthesis occurs at multiple **replications forks (replicons)** [iGen3 03-09]

DNA synthesis occurs on leading & lagging strands **simultaneously** [iGen3 03-08]

A single, dimeric **DNAPol III** replicates **both strands**

**DNA synthesis in eukaryotes**

Eukaryotic **genomes** are **much larger** [the "**C-value Paradox**"]

⇒ eukaryotic DNA synthesis is more "efficient":

More **DNAPol** molecules, slower rate of synthesis, more replicons,

*E. coli*: 15 **DNAPol** add 100,000 bases/min over 3,500 replicons

⇒ **4.2 x 10<sup>6</sup> bp** genome replicated in **20 ~ 40 min**

*Drosophila*: 50,000 **DNAPol** add 500 ~ 5,000 bases/min over 25,000 replicons

⇒ **330 x 10<sup>6</sup> bp** diploid genome replicated in **< 3 min** : net **600x faster**

**Transcription: synthesis of messenger RNA (mRNA)** ([online MGA2 animation](#))

*What is a "Gene"* [iGen3 05-03]

**RNA** transcribed from **DNA** by **RNA Polymerase (RNAPol I)** [iGen3 05-01]

- (1) **Recognition** of transcriptional unit: ~ '**gene**'  
**Promoters** - short DNA sequences that regulate transcription  
 typically '*upstream*' = '*leftward*' from 5' end of sense strand
- (2) **Initiation & Elongation** [iGen3 05-04ab , -04cd]  
 mRNA synthesized 5'→3' from DNA **template strand**  
 mRNA sequence therefore homologous to DNA **sense strand**  
**Colinear**: mRNA and DNA **sense strand** "*line up*"  
 (in prokaryotes, but *not* eukaryotes: see below)  
 Process similar to DNA replication, **except**  
**No primer** is required  
 Transcription may occur from **either strand**  
 Some (most?) DNA is not transcribed into RNA
- (3) **Termination** [iGen3 05-05]

### Regulation of transcription

In **prokaryotes**, **transcription & translation** may occur simultaneously  
 In **eukaryotes**, **transcription** occurs in **nucleus** [ex.: **Lampbrush chromosomes**]  
**translation** occurs in **cytoplasm** (see next section):  
 ⇒ **RNA must cross nuclear membrane** [iGen3 05-09]  
 transcription & translation are **physically separated**  
 primary RNA transcript is extensively processed  
**heterogeneous nuclear RNA (hnRNA) ⇒ mRNA**

**Post-transcriptional processing** of eukaryotic RNA is **complex**  
**promoters & enhancers** determine initiation & control rate  
**'cap'** (7-methyl guanosine, 7mG) added to 5' end [iGen3 05-10]  
**'tail'** of poly-A (5'~::~~AAAAA~3') added to 3' end [iGen3 05-11]  
**'splicing'** of **hnRNA** : eukaryotic genes are "split" (MGA2 03-12,14,15,16) [iGen3 05-12]  
**intron** DNA sequences removed from hnRNA : "*intervening*" [iGen3 05-14]  
**exon** DNA sequences represented in mRNA : "*expressed*" in protein  
 1 ~ 12's of **exons** / '*gene*'  
 >90% of transcript may be '*spliced out*'  
 [An important note on **terminology**]  
**Eukaryotic genes & mRNA are not colinear!**  
 DNA / RNA hybridization produces **heteroduplexes**  
**DNA introns** '*loop out*'  
**DNA exons** pair with mRNA  
 Eukaryotic exons may be widely separated  
**Alternative splicing** of the *same* transcript produces *different* products [iGen3 18-14]

Summaries of transcription [& translation] in **prokaryotes** & **eukaryotes**

### Homework #5: Suggested problems from

**MGA2** (2002), Chapter 2, pp. 53-54  
 Solved problems 1 & 2  
 problem ## 7, 8, 9, 11, 14, 15, 18, 19, 21, 26, 27  
 for extra fun: ## 29 & 34

**iGen3** (2010), Chapter 5, pp. 98-101  
 Problems ## 2, 4, 6, 7, 12, 13, 15, 16, 21

### Ongoing Homework problem:

**What is a 'gene'?** How does the discovery of **introns** and **exons** in eukaryotic genomes modify the concept?

All text material © 2011 by **Steven M. Carr**



# The Genetic Code

---

**The Central Dogma:** DNA makes RNA makes protein



**In principle:** The DNA **genotype** does not produce the **phenotype** directly  
 A DNA **gene** contains the **information** necessary for the production of **proteins**, which is expressed **biochemically** through an intermediate molecule, **RNA**, which functions as a **Genetic Code**

The **Genetic Code** ...

specifies **amino acids** that make up **proteins**

Protein expression leads directly (or indirectly) to the phenotype

was "**cracked**" before the details of translation were understood:

⇒ we can talk about the Code *before* describing **RNA** translation

can be used to infer the protein product of a gene **directly** from **DNA**:

see next section, and lab exercise

**Alternative alleles** of genes arise by **mutation**

which alters the **DNA sequence** of **genes**

which *may* cause **amino acid substitutions** in **proteins**

which *may* affect the **function** of those proteins

As a result, most genes are highly **polymorphic**

**The Genetic Code is ...** [iGen3 pp. 106 ff.]

a **messenger RNA (mRNA)** code

*i.e.*, the code is written in **RNA**

**DNA** is a coding molecule,

but *not* 'the genetic code' in the biochemical sense

in **64 triplets (codons)** : 61 for **amino acids** + 3 '**stops**' [iGen3 06-07]

**mRNA codons are read 5'→3'**

20 **amino acids**: note 1- & 3-letter abbreviations

[more on amino acids & proteins in next section]

For example,

```

5' - A U G U U C C C A A G G G U U G A - 3'
    met phe pro lys gly *
    M F P K G *
  
```

**Degenerate:** most amino acids are encoded by more than one codon

first two positions are critical: third position can "**wobble**"

if third can be either **puRine (R)**, or either **pYrimidine (Y)** ⇒

two-fold degeneracy

if third can be *any* base ⇒

four-fold degeneracy

**Leucine (leu)** has six-fold degeneracy with six codons in unusual arrangement

# codons / amino acid	
<i>trp, met</i>	1 @
<i>ser, arg, leu</i>	6 @
<i>ile</i>	3 @
14 others	2 or 4 @

**Unambiguous:** any one triplet codes for only one amino acid  
but not *vice versa*, because of **wobble**

'Always' begins with an **'start' or 'initiator' codon: AUG**

'Always' ends with a **'stop' or 'terminator' codon: UAG, UAA, or UGA**

**Universal** (with some important exceptions)

**Five Kingdoms** (animals, plants, algae, fungi, & monera)

use the same codes for **nuclear DNA (nucDNA)**

**Organelles (chloroplasts & mitochondria)** have separate **genomes:**

**cpDNA & mitochondrial DNA** codes are evolutionarily modified

e.g., **UGA** codes for *trp* in **vertebrate mtDNA code**

termination codons may be formed by addition of "A"s to transcript

**Lab exercises use mtDNA, so this code is important**

## Alterations of the Genetic Code: Mutations

**Mutations** - interchanges of one **base type** for another [MGA2 Table 10-2]

**transitions** - alternative **pyrimidines** [ C↔T ] or **purines** [ A↔G ] [iGen3 07-03a,b]

**transversions** - purine ↔ pyrimidine [ C / T ↔ A / G ]

Recognized in individuals & populations as **SNPs (single nucleotide polymorphisms)**

[**SNPs, Mutations, & Mutants**: a note on terminology & some lessons from history]

**Alternative nucleotide sequences** of a **gene** correspond to alternative **alleles**

or: a single **gene** occurs in **variant forms (alleles)**

### Single-base mutations

Consequences of **exon SNPs** depend on **position in triplet** (MGA2 10-4) [iGen3 07-03cd,fg]

#### 3rd position

*typically* a **silent mutation** - if position "*wobbles*", no change to amino acid

*sometimes* a **missense mutation** - results in different amino acids

**2nd position** - *always* a **missense mutation**

**1st position** - *almost always* a **missense** replacement

[**Leu** codons are **major exception**]

**stop codon** mutations may occur at any position: coding → non-coding triplet

**nonsense (termination) mutation** terminates polypeptide prematurely [iGen3 07-04]

**HOMEWORK: Identify** all codons **one step away** from a termination codon

[*Hint*: there are 18]

mutations in **non-coding DNA** have **variable effects**

Ex.: mutations in **promoter regions**

mutations at **intron / exon splice junctions**

**Missense *mutations* in DNA cause *substitutions* in protein****Proteins do not mutate! Watch your language!**

Consequences depend on position of substitution in polypeptide

**none:** substitution not in active site or binding site

**minor:** substitution of same type (**synonymous** substitution)

**Allozymes** are minor variants (see laboratory exercise)

**major:** substitution affects structure / function (**nonsynonymous** substitution)

Ex.: **Glu** → **Val** in **beta-globin** produces **Sickle-cell hemoglobin (HbS)**

**HOMEWORK:** What is the DNA mutation involved?

**Insertion / Deletion (indel) mutations**

gain or loss of one or more nucleotides

**frameshift mutations** ([examples](#))

**single & double nucleotide indel** ⇒ downstream amino acids change

**nonsense mutation** eventually (quickly) produced

**triplet indel** - insertion / deletion of single amino acid

typically milder consequences

*multiple* triplet insertions produce major effects

Ex.: **CGG** repeats in "[Fragile X](#)"

**length mutations** - larger indels ( $10^{2-6}$  bps)

**Genes are highly polymorphic (w/ multiple alleles) wrt their mutational variation****Phenylalanine Hydroxylase (PAH)** ([OMIM citation 261600](#))

has 14 **exons**, encodes 2.4kb **mRNA** for 452 **amino acid protein**

Of [68 alleles](#) known to affect enzymatic activity of **PAH** [[Current GenBank List](#)]

68% **miss-sense** mutations (many produce **Phenylketonuria (PKU)**)

13% **non-sense** mutations (premature termination)

9% **indel** mutations (single base → 1~5 triplets → whole exon)

10% **splice-site** mutations (including [most common variant allele](#))

Most allelic variants of the **PAH** locus are 3rd position silent

no affect on **PAH** expression

& therefore undetected

**Homework:**

(1) "**What is a Gene?**" Write an essay that that distinguishes **Gene**, **Allele**, and **Locus**

(2) Critique the following statements:

"**PAH is the gene for Phenylketonuria (PKU).**"

"**PKU is a genetic disease caused by absence of the PAH gene.**"

Text material © 2011 by [Steven M. Carr](#)



## RNA Translation: RNA makes Protein

---

### In principle:

Translation of **messenger RNA (mRNA)** takes place on **ribosomes**, which include **ribosomal RNA (rRNA)**, with the help of **transfer RNA (tRNA)**

---

### Structure of rRNA & tRNA

#### ribosomal RNA (rRNA)

rRNA + ribosomal protein  $\Rightarrow$  **ribosomes**

Structure of rRNA: **stems & loops**

**stems**: double-stranded (**dsRNA**)

**loops**: single-stranded (**ssRNA**)

complex 2<sup>o</sup> folding

Structure of **eukaryotic ribosomes** [[iGen3 06-13](#)]

**Large Subunit (LSU)** = 60S = 28S rRNA + 5S rRNA + 50 proteins

**Small Subunit (SSU)** = 40S = 18S rRNA + 33 proteins  
= 80S monosome

**A site** (Aminoacyl), **P site** (Peptidyl), & **E site** (Exit) [[iGen3 06-14](#)]

#### transfer RNA (tRNA)

the **adaptor** molecule: ~30 tRNA types

2-dimensional '**cloverleaf**' model

small: 75 ~ 90 nucs

#### stems & loops

**D-loop** & **T $\Psi$ C-loop** ( $\Psi$  = pseudo-uridylic acid)

tRNA characterized by **2<sup>o</sup>-modified bases**

#### amino-acceptor stem

3' end is **CACCA** - 3'

5' end is **G** - 5'

#### anticodon loop

specificity of tRNA determined by **3-ribonucleotide sequence**

**3-dimensional structure** is an "L":

**D- & T $\Psi$ C-loops** fold back on each other

**Charged tRNA: aminoacyl synthetase<sub>(x)</sub>** forms ester linkage between

**3'-A** of amino-acceptor stem of tRNA<sub>(x)</sub> & **COOH** of amino acid<sub>(x)</sub> [[iGen3 06-10](#) , [-11](#)]

~20 synthetase types '[recognize](#)' [correct anticodon loop](#)

#### isoacceptance:

**one-to-one correspondence** between synthetase & amino acid

A "second genetic code"?

---

## RNA Translation: Protein Synthesis

A three-step process ([Review](#))

**Ribosomes "read" mRNA & assemble polypeptide according to [Genetic Code](#)**  
([online MGA2 animation](#))

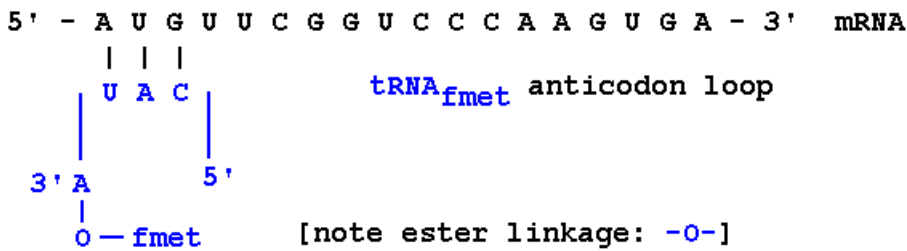
**(1) Initiation** at **start codon (AUG)** [[iGen3 06-15](#)]

**SSU** binds at **Shine-Delgarno sequence** (-6 nucs) [[iGen3 06-16](#)]

**Initiation complex** consists of **mRNA, ribosome, & tRNA**

Multiple complexes form on a single mRNA: **polysome** (**polyribosome**)

tRNA<sub>met</sub> always added first [**N-formyl-methionine** in prokaryotes]



In simplified form,

5' -AUG-3' codon in mRNA

|||

3' -UAC-5' anticodon in tRNA

5' -CAU-3' if anticodon is written 5'→3'

(2) **Elongation**: addition of amino acids [[iGen3 06-17, simplified](#)] according to [Genetic Code](#)

Amino acids are joined via **peptide bonds** (see next section)

Think of mRNA as fixed: **ribosome** moves along it 5'→3'

**peptidyl (P)** site on 5' end,  
**aminoacyl (A)** site on 3' end

first **AUG** codon [for met] is in **P site**

second **UUC** codon enters **A site**

corresponding tRNA<sub>phe</sub> enters **A site**

**peptidyl transferase** forms **peptide bond** between fmet & phe  
**P site amino acid transferred to A site amino acid** [[iGen3 06-18](#)]

tRNA<sub>fmet</sub> ester bond broken, peptide bond to tRNA<sub>phe</sub> formed

uncharged tRNA released from **P site** (passes to **E site**)

⇒ **amino end of fmet remains unchanged**

and so on ... [[online MGA animation](#)] [[iGen3 06-19](#)]

growing polypeptide in **P site** joins single amino acid in **A site**

initial fmet *always* remains unchanged

**"Wobble"**: pairing of codon / anticodon goes 5'→3' on codon

last position can **miss-pair**

Fewer tRNA species needed:

Ex.: *three* tRNA<sub>ser</sub> species for *six* codons

tRNA Anti-codon	Alternative Serine mRNA codons
3'- AG G -5'	5'- UC C / U -3'
3'- AG U -5'	5'- UC A / G -3'
3'- UC G -5'	5'- AG C / U -3'

(3) **Termination**: release of polypeptide [[iGen3 06-20](#)]

mRNA + tRNA<sub>(aa<sub>n</sub>-...-aa<sub>3</sub>-aa<sub>2</sub>-aa<sub>1</sub>)</sub>

here: mRNA + tRNA<sub>(lys-pro-gly-phe-fmet)</sub>

**stop codon** (UAG, UAA, or UGA) enters **A site**

no corresponding tRNA:

**release factor** cleaves polypeptide from terminal tRNA<sub>n</sub>

polypeptide product is: **lys - pro - gly - phe - fmet**

[interactive translation animation](#) [[Genetic Science Learning Center](#), Univ Utah]

[A talkie animation of transcription & protein synthesis](#)

[Griffiths et al. \(1996\) Fig. 13-7](#) is a nice summary (**HOMEWORK**)

## Bioinformatics of DNA, mRNA, & Protein

5' - G T A    A T C    C T C - 3' **DNA sense strand**

5' - G U A    A U C    C U C - 3' **mRNA**

N - val - ile - leu - C **protein**

This is a *logical*, not a *biochemical*, relationship:

Because **mRNA** is transcribed from the template strand,  
it "*looks like*" the *sense* strand (except for 'U').

The **information content** of the **DNA sense strand** and **mRNA** are *identical*

**Protein sequences can be read directly from DNA:**

Read the *sense* strand in the 5'→3' direction,  
Substitute 'T' for 'U' in the code table.

Computer programs (**Chromas**, **Sequencher**, etc.) do this automatically

There are [six possible ways](#) of reading a piece of **dsDNA**

**two** 5'→3' strands X **three** reading frames in each strand

**Open Reading Frames** suggest protein sequences

**Deducing protein sequences from random DNA sequences is a major research activity**

**Bioinformatics:** extraction of information from large macromolecular datasets

The following clues are useful:

**Remember** that all **prokaryotic coding** sequences:

are read *only* in the 5'→3' direction

begin with a "**start**" (**AUG**) codon

end with a "**stop**" (**UAG**, **UAA**, or **UGA**) codon.

Ex.: a typical problem is to identify a **polypeptide** of six amino acids encoded in a **dsDNA** molecule

**But.** in real life, your **eukaryotic** (cloned) **DNA** fragment

may *not* have **start** or the **stop** codon for a complete protein,

[and not all **AUG** codons are '**start**' codons]

and may include part of an **intron** with one or more '**stop**' sequences.

**Do not assume** that a **dsDNA** molecule is read from left to right, on the top strand

### Suggested problems for review

MGA2, pp. 86-87

Solved problem 1

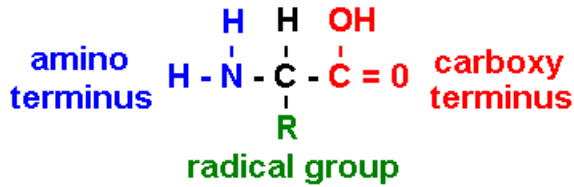
Problem ## 1, 2, 3

[Practice DNA "Translation" problems](#) [[PDF download version](#)]



## Protein Structure & Function

In principle: Proteins are polymers of **amino acids**



[sometimes  $\text{NH}_3^+$  &  $\text{COO}^-$  : depends on **pH**] [[iGen3-06-01](#)]

**R = radical group:**

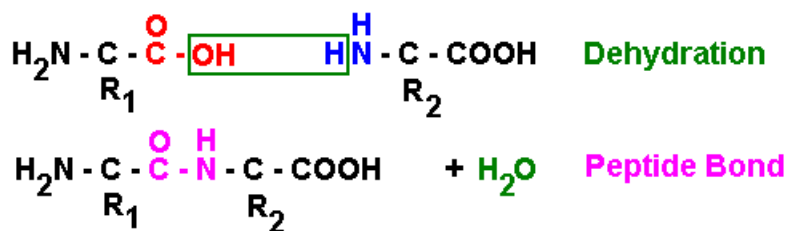
asymmetric (**Homework**); levo (L)-rotatory [cf. dextro (D)-rotatory]

determines **biological properties: 20 types** (note 1- & 3-letter codes) [[iGen3-06-02](#)]

Group properties	Three- & Single-letter codes
<u>neutral, non-polar (hydrophobic)</u>	gly, ala, val, leu, ile, pro, met, phe, trp
	<b>G A V L I P M F W</b>
<u>neutral, polar (hydrophilic)</u>	gly, ser, thr, cys, tyr, asn, gln
	<b>S T C Y N Q</b>
<u>polar basic (positive charge)</u>	lys, arg, his
	<b>K R H</b>
<u>polar acidic (negative charge)</u>	asp, glu
	<b>D E</b>

[Memorization of the abbreviations is *not* required for exams, but *will* make your lives as biologists easier!]

**Dehydration** of carboxy & amino termini forms **peptide bond** [[iGen3-06-03](#)]



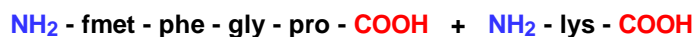
Peptidyl Transferase catalyzes analogous reaction

**carboxyl (C)** terminus of growing **polypeptide** in **P** site

cleaved from the **tRNA** &

joined to **amino (N)** terminus of new **amino acid** in **A** site [[iGen3-06-18](#)]

=> **carboxyl end "grows"**: in the last example,



Recall that amino acid in **A** site is linked to **tRNA** through **COOH** terminus

Repeating, remnant backbone subunit [**N - C(R) - C**] is an **amino acid residue**

**Proteins have four levels of structure** [[iGen3-06-04](#)]

**Primary Structure** - order of amino acid residues in polypeptide

$20^N$  possible, where **N** is number of residues

Potential for enormous variety:

e.g.,  $20^5 = 3.2 \times 10^6$  possible pentapeptides

**Secondary Structure** - configuration of [-N-C(R)-C-] backbone

**alpha helix**: a right-handed helix

**beta-pleated-sheet**: parallel / antiparallel chains

both stabilized by **H-bonds**

**Tertiary Structure** - **3-Dimensional folding** of backbone

**cys + cys** pairs form **disulfide bridges** (- S - S -)

**pro** residues form **hydrophobic "corners"**

**hydrophilic** residues occur on *exterior*,  
participate in reactions in aqueous environments

**hydrophobic** residues occur in *interior*,  
interact with membrane lipid bi-layer

**gly** fits in both **hydrophobic & hydrophilic** environments

**Quaternary Structure** - assembly of **multiple subunits**

monomers / dimers / oligomers

e.g., **hemoglobin** is a **tetramer**: two **alpha** + two **beta** chains

**charged** residues (**asp, glu, lys, arg, his**) form ionic bonds bx subunits

## Post-translational processing

**Chemical modification** of amino acids

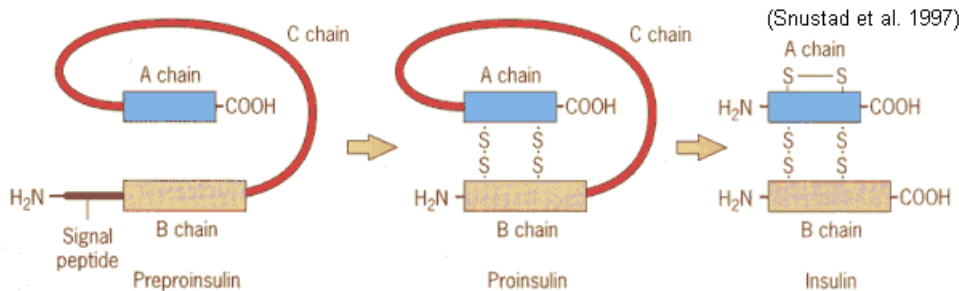
addition of **formyl** group to **met** → **fmet**

Addition of **carbohydrate** side chains (**glycoproteins**)

**ABO blood group** proteins

Amino acids may be cleaved out of primary structure

e.g., biologically active **insulin** is less than half the primary sequence



**preproinsulin** → **proinsulin** → **insulin**  
(110 aa's)      (86 aa's)      (51 aa's)

**signal peptide** (24 aa's) clipped from amino terminus

**C peptide** (31 aa's) excised from center

Tertiary (active) structure of insulin is

**A chain** (30 aa's) & **B chain** (21 aa's) held together by 3 disulfide bridges

## Detection of genetically-based variation in Proteins

**allelic variation** translates to **proteins** with minor changes in **electric charge**

**allozymes** arise from different **alleles** of the *same* protein gene locus

substitution of alternatively charged amino acids changes overall charge

[cf: **Isozymes** are homologous proteins encoded by *different* gene loci]

Ex.: **Sickle-cell hemoglobin (HbS)** is a variant of **standard HbA beta globin** protein

**glu** → **val** in sixth residue of beta chain

because **GAG** → **GUG** in sixth codon of beta-chain mRNA

**HOMEWORK: What is the corresponding mutation in the DNA?**

"minus" → "neutral" charge : net negative **electrophoretic charge** decreased [MGA2 14-7]

**val** stabilizes **crystalline** form of hemoglobin => '**sickling**' of rbc's [MGA2 14-6]

**Homework: Critique the following statement:**

"Electrophoresis of hemoglobin shows two alleles, **F** and **S**, for the **Hb** gene."

---

## Overview of protein function

**Enzymatic catalysis** of biological reactions

**Substrates** are bound in **active sites**: the **Induced-Fit Model**

Lowered **energy of activation**

biological reactions occur at body temperature  
with lower energy input

**Anabolic** - *synthesis* of complex molecules from simpler components

Ex., **transferases** synthesize peptide bonds

**polymerases** assemble nucleotides

**Catabolic** - *break-down* of complex molecules into simpler

Ex., **dehydrogenases** remove protons (**H<sup>+</sup>**)

**Amphibolic** - *reversible* reactions

direction depends on relative concentration of precursor and product

**Structural motifs** recur in proteins with similar functions

Identification of motifs allows inferences about function

**Helix - loop - helix** motifs binds **Ca<sup>++</sup>**

**Zinc - finger** motifs binds **major & minor DNA grooves**

**Other protein functions**

**Structural**

**Collagen** constitutes 25% of human protein

**Histones** are the major components of **chromosomes**

[**online MGA animation** of **DNA** packing into chromosomes]

**Nucleic Acid binding proteins**

**Polymerases, nucleases, helicases, ligases**, etc.

**Transport**

**Hemoglobin** in blood & **myoglobin** in muscle bind **O<sub>2</sub>**

**Miscellaneous**

immunoglobulins, hormones , etc.

**Major Histocompatibility Complex (MHC)** determines transplant success

**Drosophila Genome Project** has cataloged >11,000 genes with protein products

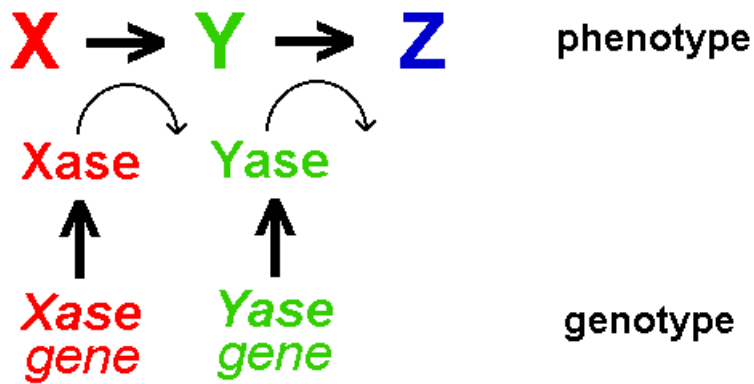
**2 / 3** have **unknown functions**

---

All text material ©2011 by [Steven M. Carr](#)

---



**"One Gene, One Enzyme"**

In Principle: **Proteins** are the products of **genes**.

Proteins catalyze **biochemical reactions**.

Such reactions produce **phenotypes**, either directly or indirectly.

Different **alleles** produce different phenotypes

Interaction between gene alleles in *diploid* organisms is the classic subject of **Genetics**

**How do genotypes produce phenotypes?**

**Beadle & Tatum experiment** (1940s) on **haploid *Neurospora*** bread mold [iGen3 04-02]

**haploid** organisms have **one allele at each gene locus**

**prototroph** ("self feeding") **wild-type** grows on **simple medium**

**auxotroph** ("other feeding") **mutants cannot** grow on simple medium,  
require supplementation with **specific amino acids** [iGen3 04-03]

[These are also known as **autotrophs** and **heterotrophs**, respectively]

**Hypothesis:** "No-growth" *phenotype* results from a change in the *genotype*:  
inability to synthesize amino acid is the result of loss of enzyme activity  
each *mutant* corresponds to a *defect* in a *particular enzyme*:

**"One gene, one enzyme"** (Homework)

[ **Remember:** Beadle & Tatum did not know about **DNA** in 1940]

Ex.: ***arg<sup>-</sup>*** mutants *cannot* grow without **arginine**, *always* grow with added **arginine**

particular mutant classes *sometimes* grow with **other amino acids**

(such as **citrulline** and/or **ornithine**)

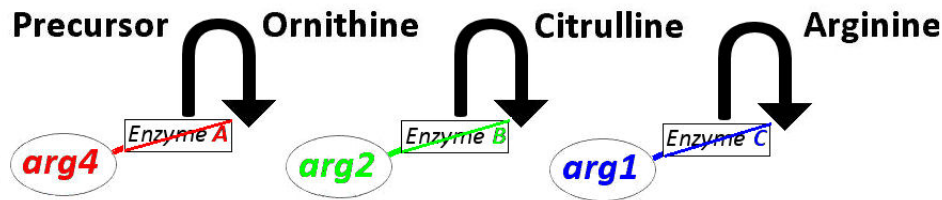
[cf. **methionine** metabolism: [iGen3 04-04]

Growth response to added amino acids				
mutants	none	ornithine	citrulline	arginine
<b><i>arg<sup>-</sup> 4</i></b>	-	+	+	+
<b><i>arg<sup>-</sup> 2</i></b>	-	-	+	+
<b><i>arg<sup>-</sup> 1</i></b>	-	-	-	+

=> These other amino acids are involved in the **arginine** pathway:

**Enzyme defect blocks interconversion of precursors in the pathway.**

Inference of a *haploid* **biosynthetic pathway**



Each mutant class (*arg4*, *arg2*, & *arg1*) affects a different enzyme in **arginine** biosynthesis

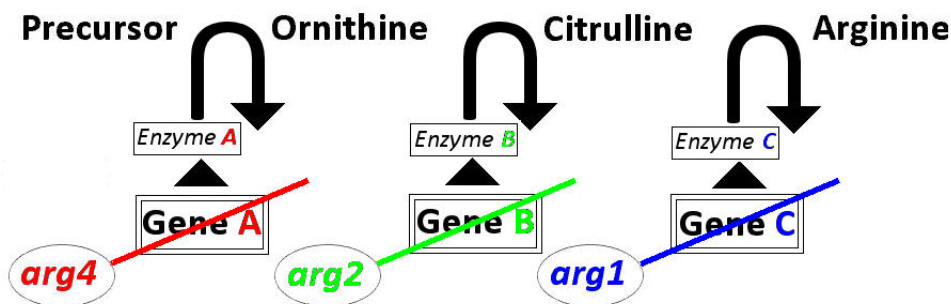
### Homework:

- In the above discussion, I have used "*mutant*" but carefully avoided "*mutation*": **WHY?**
- Critique the following statements:
  - "*arg<sup>-</sup>* mutants result in defective arginine."
  - "*arg<sup>-</sup>* mutants are defects of arginine."
  - "*arg<sup>-</sup>* mutants are due to absence of the gene for arginine."
  - "*arg<sup>-</sup>* mutants are enzymes that block synthesis of arginine."
- Would you expect to observe an ***arg*** mutant class with the following phenotype? Explain.

Growth response to added amino acids				
mutant	none	ornithine	citrulline	arginine
<i>arg<sup>-</sup> X</i>	-	+	-	+

## Biochemical Basis of Human Genetic Diseases

In a **haploid** organism, discovery of **DNA** shows nature of metabolic "*mutants*":



**Mutations** affect the **genes** responsible for different **enzymes**:  
mutate the gene  $\Rightarrow$  eliminate (or modify) the enzyme

What about **diploid** organisms?

**Diploid** organisms have **two alleles** at each **gene locus**: one from each parent  
Interactions between alleles at a locus are the subject matter of **genetics**

"**Online Mendelian Inheritance in Man**" (**OMIM**) database

examples from human biochemical genetics: "**Inborn errors of metabolism**" [[iGen3 04-Table02](#)]

First three involve disruptions of **phenylalanine metabolism** [[iGen3 04-01](#)]

**Phenylketonuria (PKU)** (Folling 1934) ([OMIM citation 261600](#)) [[iGen3 p. 66 ff.](#)]  
**phenylalanine** accumulates in **Central Nervous System**  $\Rightarrow$  mental retardation

A defect of [phenylalanine hydroxylase](#)  
**phenylalanine** metabolized to **phenylpyruvic acid** in alternative pathway

Detection & treatment

biochemical testing of new-borns: [Guthrie Test](#)

phenylalanine-restricted diet corrects inborn condition (**Euphenics**)

**Maternal PKU** results from high fetal [**phe**] in treated, asymptomatic mothers

[ Further information on [PKU & related Inborn Errors of Metabolism](#) ]

**PKU** arises from variation at the [Phenylalanine Hydroxylase \(PAH\) gene locus](#)

**Important:** This gene is not a gene "for" **PKU**: it is a gene "for" **PAH**

**Diploid** humans each have **two alleles** at this locus

**Allelic variants** produce different levels of **PAH activity**

Consider three (hypothetical) alleles: **A**, **B**, & **C** :

Phenotypic consequences of interactions between alleles at the PAH locus

Genotype	PAH Activity	[phe] uM	PKU Phenotype
<b>AA</b>	<b>100%</b>	60	<b>Standard</b>
<b>AB</b>	<b>30%</b>	120	<b>Standard</b>
<b>CC</b>	<b>5%</b>	200 ~ 300	<b>Hyperphenylalanemia:</b> no special diet required
<b>BB</b>	<b>0.3%</b>	600 ~ 2400	<b>Classic PKU:</b> special diet required

[Alleles **B** & **C**] arise from [DNA mutations](#) in the **PAH gene**]

>> **PKU** is a classic "**recessive**" genetic disease <<

What does this mean?

**AB** genotype shows same **PKU phenotype** as **AA** genotype

that is, **A** allele shows **haplosufficiency**: one 'dose' is sufficient to produce standard phenotype  
 or, expression of **A** allele "**masks**" expression of **B** allele

**A** is therefore "**dominant**" to **B** in influencing **PKU phenotype**

**B** is therefore "**recessive**" to **A**

but **PAH activity phenotype** of **AB** is *intermediate* between **AA** & **BB**

**AB phenotype** is closer to **BB** than **AA** ( $0\% < 30\% \ll 100\%$ )

**B** is an "**incomplete dominant**" to **A**

and **B** produces a *higher* [**phe**] phenotype than **A**:

Isn't **B** therefore "**dominant**" to **A** ?

Be careful to [distinguish molecular & phenotypic expression](#) (**Homework**)

#### Homework

Predict the **PAH activity**, [**phe**], & **PKU phenotypes** of the **AC** and **CB** genotypes.  
 Explain your reasoning.

Would you expect to find a **dominant** mutation in this pathway?  
 Why or why not? What might be the nature of such a mutation?

Patterns of molecular phenotypic expression in genetic diseases

"**Recessive**" diseases: [Alkaptonuria](#) & [Albinism](#)

Tay-Sachs, p. 68 ff. [[iGen3 04-05, -06](#)]

Cystic Fibrosis, p. 71 ff. [[iGen3 04-12, -13](#)]

"**Co-dominant**" diseases: [Sickle-Cell Anemia](#) [[iGen3 04-07](#) ,[-08,-10](#)]

"**Dominant**" diseases: [Huntington Disease](#)

---

**Homework:**

**MGA2**, pp. 86-87

Problems ## 8, 9, 10, 12, 18, 21, 23, 24

---

All text material © 2011 by [Steven M. Carr](#)

---

[Home](#)