IAE ACADEMY INSTITUTE FOR ENGINEERS

GATE BT

IAE Academy free study material · Updated 27 October 2026

Genetics, Cell and Molecular Biology Formula Sheet

INSIDE

- 45 formulas in 6 topics
- When to use each one
- Taken from the free topic notes on iaeacademy.in

Online, one page per topic: **iaeacademy.in/study/genetics-cell-molecular-biology/formula-sheet**

FREE COUNSELLING

Preparing for GATE BT? Talk to IAE Academy.

Or call **+91 84204 36524**

Every day, 10:30 am to 7:30 pm

Chat on WhatsApp

Free. Not for sale. Pass it on.

Mendelian genetics, gene interaction and linkage mapping

MULTIPLICATION (AND) RULE

$$P(A \text{ and } B) = P(A) \times P(B)$$

For two independent events, e.g. the genotype at one unlinked locus combined with the genotype at another unlinked locus.

ADDITION (OR) RULE

$$P(A \text{ or } B) = P(A) + P(B)$$

For mutually exclusive outcomes, e.g. two different genotypes that give the same phenotype.

NUMBER OF DISTINCT GAMETE TYPES FOR N HETEROZYGOUS, INDEPENDENTLY ASSORTING LOCI

$$2^n$$

n = number of heterozygous gene pairs. When all n loci assort independently, each gamete type has frequency $(1/2)^n$.

BINOMIAL PROBABILITY FOR A SET OF N OFFSPRING

$$P = \frac{n!}{s!\,t!}\,p^s\,q^t$$

The probability that exactly s of n offspring are of one kind (probability p each) and t are of the other kind (probability q each), with s + t = n and p + q = 1. Each offspring is treated as an independent event.

CHI-SQUARE GOODNESS-OF-FIT TEST

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

O = observed count and E = expected count in each class under the hypothesised ratio. Use counts, not percentages. Degrees of freedom = number of classes − 1. If the calculated χ^2 is below the tabulated critical value at the chosen significance level (commonly 0.05), the deviation is attributed to chance and the hypothesised ratio is not rejected.

GENOTYPES POSSIBLE WITH N ALLELES AT ONE LOCUS (DIPLOID)

$$\text{number of genotypes} = \frac{n(n+1)}{2}$$

Made up of n homozygotes and $\frac{n(n-1)}{2}$ heterozygotes. For n = 3 (ABO): 3 + 3 = 6 genotypes.

RECOMBINATION FREQUENCY (MAP DISTANCE)

$$\text{RF} = \frac{\text{number of recombinant offspring}}{\text{total offspring}} \times 100\%$$

Measured in a testcross (heterozygote × fully recessive homozygote), where each offspring's phenotype shows the gamete it received from the heterozygote. The observed RF never exceeds 50%, the value for independent assortment. Over long intervals it underestimates the true map distance, because two (or any even number of) crossovers between the markers restore the parental combination and go undetected. Map distances are therefore built by adding short intervals.

HALDANE MAPPING FUNCTION

$$d = -\frac{1}{2}\ln(1-2r)$$

r = observed recombination frequency as a fraction (r < 0.5); d = map distance in morgans (multiply by 100 for cM). Assumes crossovers occur independently along the chromosome (no interference). For small r, d ≈ r.

COEFFICIENT OF COINCIDENCE

$$\text{c.o.c.} = \frac{\text{observed double crossovers}}{\text{expected double crossovers}}$$

Expected double crossovers = RF(interval 1) × RF(interval 2) × total offspring. Both RFs are taken as fractions and each already includes the double-crossover classes. This expectation assumes that crossovers in the two intervals are independent.

INTERFERENCE

$$I = 1 - \text{c.o.c.}$$

I = 1: complete interference (no double crossovers observed). I = 0: crossovers in the two intervals occur independently (observed = expected). I < 0 (c.o.c. > 1): negative interference, more double crossovers than expected.

Microbial genetics, gene transfer and transposable elements

INTERRUPTED-MATING TIME-OF-ENTRY DISTANCE

$$\text{Map distance (minutes)} = t_2 - t_1$$

t_1 and t_2 are the times of entry (minutes after mixing) of two markers transferred by the same Hfr strain in the same direction. Distances from different Hfr strains are compared only after aligning each strain's order on the circular map.

COTRANSDUCTION FREQUENCY (DEFINITION)

$$C = \frac{\text{transductants selected for marker A that also carry donor marker B}}{\text{all transductants selected for marker A}}$$

C is a fraction (multiply by 100 for a percentage). Select for one donor marker, then score the unselected marker. Cotransformation frequency is defined the same way among transformants.

WU FORMULA (GENERALIZED TRANSDUCTION)

$$C = \left(1 - \frac{d}{L}\right)^3$$

C is the cotransduction frequency (a fraction), d the distance between the two markers, and L the length of the packaged fragment. d and L must be in the same unit, either kb or map minutes; for P1, L is about 2 min of the E. coli map. The formula assumes the fragment ends fall at random and holds for d < L. C falls to 0 as d approaches L and is taken as 0 for d ≥ L.

DISTANCE FROM COTRANSDUCTION FREQUENCY (WU FORMULA REARRANGED)

$$d = L\left(1 - C^{1/3}\right)$$

Use this to find d from a measured cotransduction frequency C (as a fraction, not a percentage). The same conditions as the Wu formula apply.

RECOMBINATION FREQUENCY AND MAP UNITS

$$\text{RF} = \frac{\text{number of recombinant progeny}}{\text{total progeny scored}} \times 100\%$$

RF is expressed as a percentage, and 1% RF = 1 map unit (cM). It applies to crosses scored for recombinant and parental classes (including phage crosses). This scale cannot be converted directly into the minutes of an interrupted-mating map.

Chromosomal variation, genetic disorders, population genetics and evolution

ALLELE FREQUENCY SUM (TWO ALLELES)

$$p + q = 1$$

p = frequency of allele A, q = frequency of allele a in the gene pool.

ALLELE FREQUENCIES FROM GENOTYPE FREQUENCIES

$$p = f_{AA} + \frac{1}{2} f_{Aa}, \qquad q = f_{aa} + \frac{1}{2} f_{Aa}$$

Valid whether or not the population is in Hardy-Weinberg equilibrium: each heterozygote carries one copy of each allele.

HARDY-WEINBERG GENOTYPE FREQUENCIES

$$p^2 + 2pq + q^2 = 1$$

At equilibrium: p^2 = frequency of AA, $2pq$ = frequency of Aa, q^2 = frequency of aa.

THREE ALLELES (E.G., ABO)

$$(p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

Homozygote frequencies are the squares; each heterozygote frequency is twice the product of its two allele frequencies.

X-LINKED LOCUS AT EQUILIBRIUM

$$\text{males: } f(X^A Y) = p,\ f(X^a Y) = q; \qquad \text{females: } p^2 + 2pq + q^2$$

The fraction of males showing an X-linked recessive trait estimates q directly. The fraction of affected females is q^2, which is why such traits are far more frequent in males.

CHI-SQUARE TEST FOR HARDY-WEINBERG EQUILIBRIUM

$$\chi^2 = \sum \frac{(O - E)^2}{E}, \qquad \text{df} = (\text{genotype classes}) - (\text{alleles}) = 3 - 2 = 1$$

Expected counts E come from $p^2 N$, $2pqN$, $q^2 N$, using p and q estimated from the same data, which costs one degree of freedom. With two alleles, df = 1 (critical value 3.84 at the 5% level).

GENOTYPE FREQUENCIES WITH INBREEDING

$$f_{AA} = p^2 + Fpq, \qquad f_{Aa} = 2pq(1 - F), \qquad f_{aa} = q^2 + Fpq$$

F = inbreeding coefficient (the probability that the two alleles in an individual are identical by descent). Inbreeding raises homozygosity without changing allele frequencies. $F = 1 - \frac{H_{\text{obs}}}{2pq}$.

SELECTION AGAINST A RECESSIVE HOMOZYGOTE

$$q' = \frac{q(1 - sq)}{1 - sq^2}, \qquad \Delta q = -\frac{s\, q^2 (1 - q)}{1 - s\, q^2}$$

Fitnesses $AA = 1$, $Aa = 1$, $aa = 1 - s$. Selection is slow when q is small, because almost all copies of a are hidden in heterozygotes.

COMPLETE SELECTION AGAINST A RECESSIVE (RECESSIVE LETHAL, S = 1)

$$q_n = \frac{q_0}{1 + n\, q_0} \quad \Leftrightarrow \quad n = \frac{1}{q_n} - \frac{1}{q_0}$$

From $q' = q/(1 + q)$, each generation adds 1 to $1/q$. n is the number of generations to go from q_0 to q_n.

EQUILIBRIUM UNDER HETEROZYGOTE ADVANTAGE

$$\hat{q} = \frac{s}{s+t}, \qquad \hat{p} = \frac{t}{s+t}$$

Fitnesses $AA = 1-s$, $Aa = 1$, $aa = 1-t$; q = frequency of a. The equilibrium is stable. The allele whose homozygote pays the larger cost stays rarer.

MUTATION–SELECTION BALANCE

$$\hat{q} \approx \sqrt{\frac{\mu}{s}} \ \text{(deleterious recessive)}, \qquad \hat{q} \approx \frac{\mu}{s} \ \text{(deleterious dominant)}$$

μ = mutation rate per generation towards the deleterious allele; s = selection coefficient against the affected genotype. These approximations hold when $\mu \ll s$.

MIGRATION (ONE-WAY GENE FLOW)

$$q_1 = (1-m)\,q_0 + m\,q_m, \qquad \Delta q = m\,(q_m - q_0)$$

m = fraction of the recipient population's genes that come from migrants each generation; q_m = allele frequency among migrants.

LOSS OF HETEROZYGOSITY BY DRIFT

$$H_t = H_0 \left(1 - \frac{1}{2N_e}\right)^t$$

N_e = effective population size; t = generations. The smaller N_e, the faster variation is lost.

EFFECTIVE POPULATION SIZE WITH UNEQUAL SEX RATIO

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

N_m, N_f = numbers of breeding males and females. N_e is always ≤ the census number of breeders, and is dominated by the rarer sex.

HERITABILITY AND RESPONSE TO SELECTION (BREEDER'S EQUATION)

$$R = h^2 S, \qquad h^2 = \frac{V_A}{V_P}, \qquad H^2 = \frac{V_G}{V_P}$$

S = selection differential (mean of selected parents − population mean); R = response (mean of offspring − population mean). h^2 = narrow-sense heritability (additive variance / phenotypic variance); H^2 = broad-sense heritability (total genetic variance / phenotypic variance). Only h^2 predicts the response to selection.

NEUTRAL MUTATIONS: FIXATION PROBABILITY AND SUBSTITUTION RATE

$$P_{\text{fix}} = \frac{1}{2N}, \qquad k = 2N\mu \times \frac{1}{2N} = \mu$$

In a diploid population of N individuals, $2N\mu$ new neutral mutations arise per generation, and each is eventually fixed with probability $1/(2N)$. The neutral substitution rate k therefore equals the neutral mutation rate μ, independent of population size. This is the basis of the **molecular clock**.

DETECTING SELECTION IN PROTEIN-CODING GENES

$$\omega = \frac{d_N}{d_S}$$

d_N = non-synonymous substitutions per non-synonymous site; d_S = synonymous substitutions per synonymous site. $\omega < 1$: purifying (negative) selection; $\omega \approx 1$: neutral evolution; $\omega > 1$: positive (adaptive) selection.

Cell structure, cell cycle, protein trafficking and cell death

EXPONENTIAL GROWTH OF A CELL POPULATION

$$N = N_0 \cdot 2^{t/t_d}$$

Here N_0 is the initial cell count, N the count after time t, and t_d the doubling time. It holds only while every cell is cycling with the same doubling time, with no cell death and outside any lag or stationary phase.

NUMBER OF GENERATIONS (DOUBLINGS) IN TIME T

$$n = \frac{t}{t_d} = \frac{\log_{10} N - \log_{10} N_0}{\log_{10} 2} \approx 3.32(\log_{10} N - \log_{10} N_0)$$

This follows directly from the exponential-growth formula under the same conditions.

SPECIFIC GROWTH RATE

$$\mu = \frac{\ln 2}{t_d}, \qquad N = N_0 e^{\mu t}$$

This is the same exponential law written with base e. μ has units of reciprocal time.

PHASE DURATION FROM THE FRACTION OF CELLS IN THAT PHASE (E.G., MITOTIC INDEX)

$$t_{\text{phase}} \approx \frac{n_{\text{phase}}}{n_{\text{total}}} \times T$$

T is the total cell-cycle time and $n_{\text{phase}}/n_{\text{total}}$ is the fraction of cells observed in the phase; for M phase this fraction is the mitotic index. This is an approximation for an asynchronous culture that treats cells as evenly spread across cycle ages. In an exponentially growing culture young cells slightly outnumber old ones, so the estimate is not exact.

INDEPENDENT ASSORTMENT

$$\text{number of possible chromosome combinations} = 2^n$$

Here n is the haploid chromosome number. The count ignores crossing over, which adds further variation. For humans ($n = 23$), $2^{23} = 8{,}388{,}608$.

Cell communication and signal transduction

EQUILIBRIUM DISSOCIATION CONSTANT

$$K_d = \frac{[R][L]}{[RL]} = \frac{k_{off}}{k_{on}}$$

For a simple 1:1 interaction. K_d has units of concentration (k_{off} in s^{-1}, k_{on} in $\mathrm{M}^{-1}\mathrm{s}^{-1}$). A smaller K_d means tighter (higher-affinity) binding. The association constant is $K_a = 1/K_d$.

FRACTIONAL RECEPTOR OCCUPANCY (SINGLE SITE, NON-COOPERATIVE)

$$\theta = \frac{[RL]}{[R]_{total}} = \frac{[L]}{K_d + [L]}$$

At $[L] = K_d$, $\theta = 0.5$. Rearranged, $[L] = K_d \dfrac{\theta}{1-\theta}$, so moving from 10% to 90% occupancy needs an 81-fold rise in $[L]$ (from $K_d/9$ to $9K_d$). Valid only for one independent binding site per receptor.

SCATCHARD EQUATION

$$\frac{B}{F} = \frac{B_{max} - B}{K_d}$$

B = bound ligand ($[RL]$), F = free ligand, B_{max} = total receptor concentration. For a single class of independent sites, plotting B/F against B gives a straight line with slope $-1/K_d$ and x-intercept B_{max}. A curved Scatchard plot indicates cooperativity or more than one class of binding site.

HILL EQUATION (COOPERATIVE BINDING)

$$\theta = \frac{[L]^n}{K_d + [L]^n}$$

Here K_d has units of concentration^n and equals $(K_{0.5})^n$, where $K_{0.5}$ is the ligand concentration that gives half saturation. n is the Hill coefficient: $n > 1$ means positive cooperativity, $n < 1$ means negative cooperativity (or several classes of site), and $n = 1$ reduces to the single-site equation. n is an empirical measure: it cannot exceed the number of binding sites and is usually smaller.

HILL PLOT

$$\log_{10}\left(\frac{\theta}{1-\theta}\right) = n\log_{10}[L] - \log_{10} K_d$$

A straight line whose slope is the Hill coefficient n.

Replication, transcription, translation and regulation of gene expression

OKAZAKI FRAGMENTS PER COMPLETE ROUND OF REPLICATION

$$N_{\text{Okazaki}} \approx \frac{G}{\ell}$$

G = genome (or region) length in bp and ℓ = average Okazaki fragment length in nt. Every base pair is copied by exactly one fork, and each fork makes one strand continuously and the other discontinuously. So the total lagging-strand DNA equals G nt, whatever the number of origins. Primer lengths and end effects at origins and termini are ignored.

CODON COMBINATORICS

$$4^3 = 64$$

4 bases in 3 positions give 64 codons: 61 sense codons (including AUG) and 3 stop codons (UAA, UAG, UGA).

ORF LENGTH FROM POLYPEPTIDE LENGTH

$$L_{\text{ORF}} = 3\,(N_{\text{aa}} + 1)$$

L_{ORF} = nucleotides from the first base of the start codon to the last base of the stop codon, inclusive. N_{aa} = number of amino acids in the primary translation product, including the initiator Met. The +1 is the stop codon, which is not translated. If the initiator Met is removed after translation, add it back: a mature chain of n residues gives $N_{\text{aa}} = n + 1$. Any signal peptide or propeptide that is cleaved off must be added back in the same way.

FROM IAE ACADEMY

Preparing for GATE BT?

IAE Academy teaches GATE BT in Ballygunge, Kolkata, and online. Counselling is free: ask us about batches, fees or your preparation.

FREE COUNSELLING

Chat on WhatsApp

Call +91 84204 36524

Open every day, 10:30 am to 7:30 pm.

Scan to chat on WhatsApp

THE GATE BT COURSE

Classroom batches in Ballygunge and online batches. Fees and details are on the course page.

Scan: course page

See the course

FREE ON IAEACADEMY.IN

Notes and solved previous-year questions for Genetics, Cell and Molecular Biology, topic by topic.

Scan: the notes

Open the notes

Found this useful? Send it to a friend preparing for GATE.

Send to a friend on WhatsApp

IAE Academy free study material. The formulas come from the free Genetics, Cell and Molecular Biology notes on iaeacademy.in. Prepared and published by IAE Academy, Kolkata. Free to read and free to share, unaltered and with attribution.

IAE ACADEMY INSTITUTE FOR ENGINEERS

GATE coaching in Kolkata since 2012 · EC, EE, IN, CS and IT, DA, BT and CE · in the classroom and online

227A Rash Behari Ave, Ballygunge, Kolkata 700019

+91 84204 36524 · iaeacademy.in · Updated 27 October 2026

Get directions