# Biochemistry & Molecular Biology

## A Study Guide for the Experienced CS Programmer

Covering DNA Structure, Hybridization, Thermodynamics,
Enzyme Mechanics, and Gel Electrophoresis

With CS analogies, mathematical detail, and a curated reading roadmap

| | |
|---|---|
| **Target audience:** | Programmer with strong chemistry background |
| **Goal:** | Foundation for DNA computing & AI in cancer/drug research |
| **Study time estimate:** | 8–10 weeks at 5–8 hrs/week |

# Contents

## How to Use This Guide

This guide is structured as six sequential modules. Each module builds on the previous one. For each topic you will find: a conceptual overview, the essential mathematics, CS analogies to anchor unfamiliar ideas, key terms to memorize, and specific reading assignments.

> **CS Analogy:** Think of this guide like a compiler pass over an unfamiliar language. You need to parse the syntax (terminology), understand the type system (molecular specificity), then trace execution (biochemical pathways).

**Recommended reading alongside this guide:**

- *Biochemistry* — Berg, Tymoczko & Stryer, 9th ed. (cited as **BTS**)
- *Molecular Biology of the Cell* — Alberts et al., 7th ed. (cited as **MBC**)
- *Molecular Biology: Principles and Practice* — Cox et al. (cited as **Cox**)
- SantaLucia 1998, PNAS (cited as **SL98**) — free via PubMed
- Adleman 1994, Science (cited as **Adl94**) — free via Google Scholar

## Module 1 — DNA Structure and Watson-Crick Base Pairing

*Estimated time: 1 week* **Read: BTS Ch. 4, MBC Ch. 4**

### Overview

DNA (deoxyribonucleic acid) is a polymer. The monomer unit is a **nucleotide**, consisting of three parts: a five-carbon sugar (deoxyribose), a phosphate group, and a nitrogenous base. The sequence of bases encodes biological information; the chemistry of the sugar-phosphate backbone provides structural integrity.

**The four DNA bases and their pairing rules (Watson-Crick):**

| Base | Type | Pairs with |
|---|---|---|
| Adenine (A) | Purine (2 rings) | Thymine (T) — 2 hydrogen bonds |
| Thymine (T) | Pyrimidine (1 ring) | Adenine (A) — 2 hydrogen bonds |
| Guanine (G) | Purine (2 rings) | Cytosine (C) — 3 hydrogen bonds |
| Cytosine (C) | Pyrimidine (1 ring) | Guanine (G) — 3 hydrogen bonds |

G:C pairs are stronger than A:T pairs because of the additional hydrogen bond.

**CS Analogy:** The base-pairing alphabet {A, T, G, C} with complementarity rules is a 4-symbol alphabet with a complement function: `comp(A)=T`, `comp(T)=A`, `comp(G)=C`, `comp(C)=G`. A double-stranded DNA molecule is just a string and its reverse complement. In DNA computing, sequence design is fundamentally a constraint satisfaction problem over this alphabet.

### The Double Helix: Structural Details

Two antiparallel strands wind around a central axis. "Antiparallel" is critical: one strand runs $5' \to 3'$ and the complementary strand runs $3' \to 5'$. This directionality arises from the asymmetric chemistry of the phosphodiester bond.

**Key Terms to Know:**

- **5′ end**: The end of a DNA strand with a free phosphate group on the 5′ carbon of the sugar
- **3′ end**: The end with a free hydroxyl (–OH) on the 3′ carbon
- **Antiparallel**: The two strands of a duplex run in opposite directions
- **Major groove / Minor groove**: Gaps in the double helix accessible to proteins
- **B-form DNA**: The canonical right-handed helix at physiological conditions (10.5 bp/turn)
- **Phosphodiester bond**: The covalent linkage connecting nucleotides in a strand

### Stabilizing Forces in the Double Helix

The double helix is stabilized by two cooperative forces:

1. **Hydrogen bonds** between complementary bases (perpendicular to helix axis)
2. **Base stacking interactions** — hydrophobic and van der Waals forces between adjacent base pairs (parallel to helix axis)

Base stacking is actually the dominant stabilizing force energetically, though hydrogen bonding provides the sequence specificity. This distinction matters for understanding melting thermodynamics.

### Self-Check Questions

1. Write the complement of the sequence `5'-ATCGGCTA-3'` including correct directionality.
2. Why does a G:C-rich sequence have a higher melting temperature than an A:T-rich sequence of the same length?
3. What does "antiparallel" mean structurally? Sketch the $5' \to 3'$ orientation of both strands of a short duplex.
4. Distinguish between a purine and a pyrimidine. Which bases fall in each category?

## Module 2 — DNA Hybridization and Melting Temperature ($T_m$)

*Estimated time: 1.5 weeks* **Read: BTS Ch. 4, SL98**

### What is Hybridization?

**Hybridization** (also called annealing) is the process by which two complementary single-stranded DNA molecules form a double-stranded duplex via hydrogen bonding and base stacking. The reverse process — separation of strands — is called **denaturation** or **melting**.

**CS Analogy:** Hybridization is pattern matching at the molecular level. A single-stranded probe sequence will bind its reverse complement with high specificity. Mismatches destabilize the duplex measurably — this is the physical basis for sequence-specific detection, and it is exactly why DNA computing gates can implement logic.

### Melting Temperature: Definition and Basic Formula

The **melting temperature** $T_m$ is the temperature at which 50% of a given DNA duplex population is single-stranded. It is not a sharp transition but a sigmoidal curve.

### Wallace Rule (quick estimate for short oligos $\leq 20$ bp)

$$T_m \approx 2°\text{C} \times (n_A + n_T) + 4°\text{C} \times (n_G + n_C) \tag{1}$$

where $n_A$, $n_T$, $n_G$, $n_C$ are the counts of each base. This rule is crude but useful for back-of-envelope calculations.

### Nearest-Neighbor Model (the correct model)

The Wallace rule ignores sequence context entirely. In reality, the stability of a base pair depends on its neighbors. The **nearest-neighbor model** (SantaLucia 1998) computes $\Delta H°$ and $\Delta S°$ by summing contributions from each dinucleotide step:

$$\Delta H° = \sum_{i=1}^{n-1} \Delta H°_{i,i+1} + \Delta H°_{\text{init}} \tag{2}$$

$$\Delta S° = \sum_{i=1}^{n-1} \Delta S°_{i,i+1} + \Delta S°_{\text{init}} \tag{3}$$

$$\boxed{T_m = \frac{\Delta H°}{\Delta S° + R \ln(C_T/4)}} \tag{4}$$

where $R = 1.987$ cal mol$^{-1}$K$^{-1}$ is the gas constant and $C_T$ is the total strand concentration. The $\ln(C_T/4)$ term accounts for the concentration dependence of the bimolecular association reaction.

**Nearest-neighbor parameters (selected examples from SL98):**

| Sequence ($5' \to 3'/3' \to 5'$) | $\Delta H^\circ$ (kcal/mol) | $\Delta S^\circ$ (cal/mol·K) |
|---|---|---|
| AA/TT | −7.9 | −22.2 |
| AT/TA | −7.2 | −20.4 |
| TA/AT | −7.2 | −21.3 |
| GC/CG | −9.8 | −24.4 |
| GG/CC | −8.0 | −19.9 |

A full table of all 10 unique dinucleotide pairs is in SL98 Table 2. NUPACK and most primer design tools use this dataset.

### Factors Affecting $T_m$

- **GC content**: Higher GC ⇒ higher $T_m$ (3 H-bonds vs 2)
- **Length**: Longer duplexes are more stable
- **Salt concentration**: Monovalent cations ($Na^+$) shield the negatively charged phosphate backbone, stabilizing the duplex. Standard correction:

$$T_m(\text{corrected}) = T_m + 16.6 \log_{10}\left(\frac{[\text{Na}^+]}{1.0\ \text{M}}\right) \tag{5}$$

- **Strand concentration** $C_T$: Higher concentration shifts equilibrium toward duplex
- **Mismatches**: Each mismatch typically lowers $T_m$ by 5–12°C depending on position and type

### Self-Check Questions

1. Using the Wallace rule, estimate $T_m$ for `5'-GCATGCATGC-3'` (a 10-mer with 5 G/C and 5 A/T).
2. Why does the nearest-neighbor model outperform the Wallace rule for oligos used in PCR?
3. What happens to $T_m$ if you double the strand concentration $C_T$? (Examine the $\ln(C_T/4)$ term and reason qualitatively.)
4. A DNA strand displacement gate relies on a toehold of 6 nt. Why would you want the toehold $T_m$ to be well below the operating temperature?

## Module 3 — Thermodynamics: $\Delta G$, $\Delta H$, $\Delta S$ in Biology

*Estimated time: 1 week* **Read: BTS Ch. 1, Ch. 4**

### Why Thermodynamics Matters Here

Every molecular event in biochemistry — hybridization, enzyme catalysis, membrane transport — is governed by thermodynamics. The key insight your chemistry background gives you: *the same laws apply, but biological systems operate far from equilibrium and at nearly constant temperature and pressure*, so Gibbs free energy $\Delta G$ is the dominant framework.

### Core Thermodynamic Relations

$$\Delta G = \Delta H - T\Delta S \tag{6}$$

$$\Delta G^\circ = -RT \ln K_{eq} \tag{7}$$

$$\Delta G = \Delta G^\circ + RT \ln Q \tag{8}$$

where $Q$ is the reaction quotient. At equilibrium, $\Delta G = 0$ and $Q = K_{eq}$.

**Sign conventions — what they mean biologically:**

| $\Delta H$ | $\Delta S$ | $\Delta G$ | Interpretation |
|---|---|---|---|
| $-$ | $+$ | Always $-$ | Spontaneous at all temperatures (e.g., many hydrophobic folding events) |
| $-$ | $-$ | $-$ at low $T$ | Spontaneous only below $T = \Delta H/\Delta S$ (e.g., DNA hybridization!) |
| $+$ | $+$ | $-$ at high $T$ | Spontaneous only above $T = \Delta H/\Delta S$ (e.g., denaturation) |
| $+$ | $-$ | Always $+$ | Never spontaneous under standard conditions |

### DNA Hybridization is Enthalpy-Driven

DNA duplex formation is **exothermic** ($\Delta H < 0$, from H-bonds and stacking) and **entropically unfavorable** ($\Delta S < 0$, two strands become one ordered structure). Therefore:

$$\Delta G_{\text{hyb}} = \underbrace{\Delta H}_{\text{negative}} - T \underbrace{\Delta S}_{\text{negative}} = \Delta H - T\Delta S \tag{9}$$

Hybridization is favorable ($\Delta G < 0$) when $|\Delta H| > T|\Delta S|$, i.e., at temperatures below $T_m$. At $T = T_m$ by definition $\Delta G = 0$, so:

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln(C_T/4)} \tag{10}$$

This recovers the $T_m$ equation from Module 2 — it is purely thermodynamic.

### Standard vs. Physiological Conditions

Biochemistry uses $\Delta G^{\circ\prime}$ (standard free energy at pH 7.0, 37°C, 1 M solutes except water and $H^+$) rather than the chemist's $\Delta G^\circ$ (pH 0, 25°C). This matters when comparing published values.

### Coupled Reactions and Energy Currency

Cells drive thermodynamically unfavorable reactions by coupling them to ATP hydrolysis:

$$\mathrm{ATP} + \mathrm{H_2O} \rightarrow \mathrm{ADP} + \mathrm{P}_i \quad \Delta G^{\circ\prime} \approx -30.5 \text{ kJ/mol} \tag{11}$$

Many enzymes you will encounter (helicases, ligases, kinases) use ATP as energy currency.

**CS Analogy:** Think of $\Delta G$ as the biological equivalent of potential energy in a system. Spontaneous reactions flow "downhill" in free energy. Enzymes do not change $\Delta G$ — they only change the *rate* by lowering activation energy $\Delta G^\ddagger$. This is analogous to caching: the result is the same, but faster.

### Self-Check Questions

1. DNA hybridization has $\Delta H = -200$ kJ/mol and $\Delta S = -0.6$ kJ/mol $\cdot$ K. At what temperature is $\Delta G = 0$? Is this the $T_m$?
2. A reaction has $\Delta G^{\circ\prime} = +15$ kJ/mol. Is it possible for this reaction to proceed spontaneously in a cell? Explain using the $\Delta G = \Delta G^\circ + RT \ln Q$ equation.
3. Why does an enzyme not appear in $\Delta G$ calculations for a reaction it catalyzes?
4. Explain why increasing temperature eventually denatures DNA even though $\Delta H < 0$ for duplex formation.

## Module 4 — Enzyme Behavior: Polymerases, Ligases, and Restriction Enzymes

*Estimated time: 1.5 weeks* **Read: MBC Ch. 5, BTS Chs. 8–9**

### Enzymes: The Molecular Machines

An enzyme is a biological catalyst — almost always a protein, occasionally RNA (ribozymes). Key properties:

- Highly specific (structural/chemical complementarity between enzyme and substrate)
- Not consumed in the reaction
- Lower activation energy $\Delta G^{\ddagger}$ but do not change $\Delta G$ of the overall reaction
- Regulated (allosterically, by post-translational modification, by concentration)

### DNA Polymerases

DNA polymerases synthesize new DNA strands by extending a **primer** along a **template** strand. The reaction:

$$(\text{dNMP})_n + \text{dNTP} \rightarrow (\text{dNMP})_{n+1} + \text{PP}_i \tag{12}$$

where dNTP = deoxynucleoside triphosphate and $\text{PP}_i$ = pyrophosphate (immediately hydrolyzed, driving the reaction forward).

**Critical rules for DNA polymerases:**

1. Synthesize only in the $5' \rightarrow 3'$ direction
2. Cannot initiate a new strand — require a pre-existing primer with a free 3'-OH
3. Read the template $3' \rightarrow 5'$
4. Most have proofreading (3'→5' exonuclease) activity to correct errors

### Polymerases You Need to Know

| Enzyme | Source | Use |
|---|---|---|
| *Taq* polymerase | *T. aquaticus* (thermophile) | PCR (heat-stable, no proofreading) |
| *Pfu* polymerase | *P. furiosus* | High-fidelity PCR (has proofreading) |
| Phi29 polymerase | Bacteriophage $\phi$29 | Isothermal amplification, strong strand displacement |
| Klenow fragment | *E. coli* Pol I (truncated) | Fill-in reactions, second-strand synthesis |
| Reverse transcriptase | Retroviruses | RNA → cDNA (crosses central dogma!) |

**CS Analogy:** DNA polymerase is a read-and-copy machine with a strict input specification: you must provide a template and a primer. The primer is the instruction pointer — it tells the polymerase where to start. No primer, no output.

### DNA Ligases

Ligases seal **nicks** in the sugar-phosphate backbone — a break where the phosphodiester bond is missing between adjacent nucleotides on the same strand. The reaction requires energy (ATP in eukaryotes, $NAD^+$ in bacteria):

$$5'\text{-phosphate} + 3'\text{-OH} + \text{ATP} \rightarrow \text{phosphodiester bond} + \text{AMP} + \text{PP}_i \tag{13}$$

**Important constraint:** ligase joins nicks only, not gaps. The two ends must be in direct contact on a complementary template strand.

### Restriction Enzymes (Restriction Endonucleases)

Restriction enzymes are bacterial defense proteins that cut double-stranded DNA at specific recognition sequences. They are the molecular scissors of biotechnology.

**Classification:**

- **Type I**: Cut at non-specific sites far from recognition sequence; rarely used in labs
- **Type II**: Cut at or near the recognition sequence — the workhorses of molecular biology
- **Type III**: Cut downstream of recognition site

### Type II Restriction Enzymes in Detail

Most Type II recognition sequences are **palindromic** (read the same on both strands $5' \rightarrow 3'$):

$$\underbrace{5'\text{–G|AATTC–}3'}_{\text{top strand}} \quad \underbrace{3'\text{–CTTAA|}G\text{–}5'}_{\text{bottom strand}}$$

*EcoRI* cuts as shown, producing 4-nucleotide **sticky ends** (overhangs). *SmaI* cuts between GGG and CCC producing **blunt ends**.

| Enzyme | Recognition site | Cut type | Overhang |
|---|---|---|---|
| *EcoRI* | 5'-G↓AATTC-3' | Staggered | 4 nt sticky (5' overhang) |
| *BamHI* | 5'-G↓GATCC-3' | Staggered | 4 nt sticky (5' overhang) |
| *HindIII* | 5'-A↓AGCTT-3' | Staggered | 4 nt sticky (5' overhang) |
| *SmaI* | 5'-GGG↓CCC-3' | Blunt | None |
| *PstI* | 5'-CTGCA↓G-3' | Staggered | 4 nt sticky (3' overhang) |

**Key application:** Sticky ends produced by the same enzyme are complementary and can be ligated together — this is the physical basis of classic DNA cloning.

**CS Analogy:** A restriction enzyme is a pattern-matching cutter: it scans the string of DNA and executes a cut at every occurrence of its recognition substring. The 6-base-pair recognition sites (e.g., *EcoRI*: GAATTC) occur on average every $4^6 = 4096$ base pairs in random sequence — a useful fact for estimating fragment sizes.

### Enzymes in DNA Computing Context

- **Strand displacement polymerases** (Phi29, Bst) extend primers while physically displacing the non-template strand — essential for isothermal amplification (LAMP, RPA) used in point-of-care diagnostics
- **Nicking enzymes** (a restriction enzyme variant) cut only one strand — used as clocks and amplification triggers in DNA circuit design
- **T7 RNA polymerase** transcribes DNA into RNA in vitro — used in SELEX and aptamer selection

### Self-Check Questions

1. A DNA polymerase is given a circular single-stranded template with no primer. Will it synthesize a complementary strand? Why or why not?
2. *EcoRI* and *BamHI* both produce 4 nt 5' overhangs. Can a fragment cut with *EcoRI* be ligated into a vector cut with *BamHI*? Why?
3. What is the difference between a **gap** and a **nick** in DNA, and which one can ligase seal?
4. How many *EcoRI* sites would you statistically expect in a 40,000 bp plasmid? (Recognition sequence = 6 bp, 4 bases)

## Module 5 — Gel Electrophoresis

*Estimated time: 0.5 weeks* **Read: OpenWetWare, Cox Ch. 6**

### Principle

Gel electrophoresis separates DNA fragments by size. DNA is negatively charged (phosphate backbone) and migrates toward the positive electrode in an electric field. A porous gel matrix (agarose or polyacrylamide) acts as a molecular sieve: smaller fragments migrate faster.

$$\mu = \frac{q}{f} \tag{14}$$

where $\mu$ is electrophoretic mobility, $q$ is charge, and $f$ is the frictional coefficient (which depends on fragment size and gel porosity).

**CS Analogy:** Gel electrophoresis is sorting by size. Think of it as bubble sort implemented in physical chemistry — smaller items (fragments) move to the front of the gel (positive electrode end) while larger items lag behind. The gel matrix is the comparison operator.

### Agarose vs. Polyacrylamide

| | Agarose gel | Polyacrylamide gel (PAGE) |
|---|---|---|
| Size range | 100 bp – 50 kb | 5 – 500 bp |
| Resolution | Lower | Very high (single-nucleotide) |
| Setup | Simple, room temperature | More complex |
| Use | Routine DNA analysis | Sequencing, short oligo work |

### Running a Gel: Step by Step

1. Cast agarose (typically 0.8–2%) in TAE or TBE buffer with ethidium bromide (EtBr) or SYBR Safe intercalating dye
2. Load samples mixed with loading dye (adds density, tracking dyes)
3. Include a **DNA ladder** — a mixture of fragments of known size as a reference
4. Apply voltage (80–120 V); run 30–60 min
5. Visualize under UV light (EtBr fluoresces when intercalated in DNA)
6. Estimate fragment sizes by comparing to ladder bands

### Reading a Gel

- Bands at the **top** = large fragments (slow)
- Bands at the **bottom** = small fragments (fast)
- **Bright band** = high concentration of that fragment
- **Smear** = degraded DNA or many overlapping sizes
- **No band** = reaction failed or sample too dilute

### Quantitative Analysis: Log-Linear Relationship

Migration distance $d$ is approximately log-linear with fragment size $N$ (in bp):

$$\log_{10}(N) \approx a - b \cdot d \tag{15}$$

where $a$ and $b$ are empirical constants for given gel/run conditions. This is why gel analysis software plots fragment size on a log scale.

### Self-Check Questions

1. You run an agarose gel with three samples. Sample A shows a band at 1.5 kb, Sample B at 500 bp, and Sample C has no band. Your ladder shows bands at 100, 250, 500, 1000, 2000, and 5000 bp. Rank A, B, C by distance migrated from the well.
2. Why is the DNA ladder essential? Could you determine fragment size without it?
3. A restriction digest of a 5000 bp plasmid with *EcoRI* should yield two fragments of 3500 bp and 1500 bp. Instead your gel shows one band at 5000 bp and no smaller bands. What likely went wrong?
4. Why is polyacrylamide preferred over agarose when you need to resolve fragments differing by only one base pair?

## Module 6 — Synthesis: How These Concepts Connect to DNA Computing and Cancer Research

*Estimated time: 0.5 weeks — review and integration*

### The Complete Picture

You now have the molecular vocabulary to understand how DNA computing and diagnostics actually work at the chemical level. Here is how the modules connect:

**Diagnostic circuit example (SHERLOCK / similar):**

**Input:** Cancer biomarker miRNA-21 present in patient sample
**Step 1** [Module 1, 2]: Probe DNA hybridizes to target miRNA via Watson-Crick base pairing. Hybridization occurs because $\Delta G < 0$ at 37°C and the sequences are complementary.
**Step 2** [Module 3]: Binding is thermodynamically favorable; $T_m$ of the probe-target duplex is designed to be ∼15°C above operating temperature for stability.
**Step 3** [Module 4]: A nicking enzyme recognizes a site exposed only in the bound state and cuts one strand. A polymerase then extends, displacing and releasing a signal strand.
**Step 4** [Module 5]: Product is run on a gel or detected by fluorescence — band presence/absence = positive/negative result.

### Key Connections to DNA Strand Displacement (DSD) Circuits

- **Toehold**: A short single-stranded overhang ($T_m$ set by nearest-neighbor model, Module 2). Acts as the binding site for an invading strand.
- **Strand displacement**: An incoming strand hybridizes to the toehold and then displaces the incumbent strand via branch migration. Thermodynamics (Module 3) drives the reaction: the product duplex is more stable.
- **Cascade amplification**: Phi29 or Bst polymerase (Module 4) can amplify signal strands isothermally.
- **Output detection**: Gel electrophoresis (Module 5) or fluorescent quencher/reporter pairs confirm circuit execution.

### Key Connections to AI in Cancer Research

| AI Application | Biochemistry foundation needed |
|---|---|
| AlphaFold / protein design | Thermodynamic stability, H-bonding, hydrophobic interactions (Modules 3, 4) |
| cfDNA liquid biopsy ML | Hybridization specificity, methylation patterns, sequencing artifacts (Modules 1, 2) |
| Drug-target binding prediction | Enzyme kinetics, $\Delta G$ of binding, induced fit (Modules 3, 4) |
| DNA circuit design automation | All modules — $T_m$, $\Delta G$, enzyme constraints, gel verification |
| Biomarker discovery from omics | Gel/sequencing data interpretation, expression units (Module 5) |

## 8-Week Reading Roadmap

| Week | Topic | Assignments |
|---|---|---|
| 1 | DNA structure, base pairing | BTS Ch. 4 (pp. 1–30); Watson & Crick 1953 (2 pages); MBC Ch. 4 sections 1–3 |
| 2 | Hybridization, $T_m$ basics | BTS Ch. 4 (remainder); Marmur & Doty 1962 (skim); begin SL98 |
| 3 | Nearest-neighbor thermo, NUPACK | Finish SL98; install NUPACK; compute $T_m$ and $\Delta G$ for 5 self-designed sequences |
| 4 | Thermodynamics ($\Delta G$, $\Delta H$, $\Delta S$) | BTS Ch. 1; BTS Ch. 4 thermodynamics sections; re-read Module 3 of this guide |
| 5 | Polymerases & replication | MBC Ch. 5 (DNA replication); BTS Ch. 25 |
| 6 | Restriction enzymes & ligases | MBC Ch. 5 (continued); Sambrook CSHL protocols (restriction digest section, free online); practice gel interpretation |
| 7 | Gel electrophoresis | OpenWetWare gel protocol; iBiology video on gel electrophoresis; Cox Ch. 6 |
| 8 | Integration & DNA computing | Adl94; Winfree 1998 (DNA tile assembly); Qian & Winfree 2011 (strand displacement circuits) |

**Ongoing throughout:** iBiology lecture series; NUPACK design exercises; maintain a personal glossary

## Master Glossary

| Term | Definition |
|---|---|
| Adenine (A) | Purine base; pairs with Thymine via 2 H-bonds |
| Annealing | Hybridization; formation of a duplex from two complementary single strands |
| Antiparallel | Two DNA strands in a duplex run in opposite $5' \to 3'$ directions |
| Base stacking | Hydrophobic + van der Waals interactions between adjacent base pairs; major stabilizing force |
| Branch migration | Process by which a strand displacement reaction propagates along a duplex |
| $C_T$ | Total strand concentration; appears in $T_m$ equation |
| Cytosine (C) | Pyrimidine base; pairs with Guanine via 3 H-bonds |
| Denaturation | Strand separation; melting of a duplex |
| dNTP | Deoxynucleoside triphosphate; the monomer substrate for DNA polymerase |
| $\Delta G$ | Gibbs free energy change; negative value = spontaneous reaction |
| $\Delta G^\circ$ | Standard free energy change (1 M, 25°C, pH 0) |
| $\Delta G^{\circ\prime}$ | Biochemical standard free energy (1 M, 37°C, pH 7) |
| $\Delta H$ | Enthalpy change; negative = exothermic |
| $\Delta S$ | Entropy change; negative = more ordered product |
| Exonuclease | Enzyme that degrades DNA from the end (5' or 3') |
| Gel electrophoresis | Separation of DNA by size in electric field through porous matrix |
| Guanine (G) | Purine base; pairs with Cytosine via 3 H-bonds |
| Hybridization | See annealing |
| Intercalating dye | Molecule (e.g., EtBr) that inserts between base pairs and fluoresces under UV |
| Ladder (DNA) | Reference mixture of fragments of known size for gel size estimation |
| Ligase | Enzyme that seals nicks (missing phosphodiester bonds) in DNA; requires ATP |
| Melting temperature ($T_m$) | Temperature at which 50% of a duplex population is single-stranded |
| Nearest-neighbor model | Thermodynamic model computing $T_m$ from dinucleotide parameters |
| Nick | Single-strand break in DNA (missing phosphodiester bond, no missing nucleotides) |
| Nucleotide | Monomer of DNA: sugar + phosphate + base |
| Palindrome | Recognition sequence reading the same on both strands $5' \to 3'$ |
| Phosphodiester bond | Covalent bond linking nucleotides in a strand (backbone bond) |
| Polymerase | Enzyme synthesizing DNA (or RNA) from a template |
| Primer | Short oligonucleotide providing the 3'-OH needed to initiate synthesis |
| Purine | Double-ring base: Adenine, Guanine |
| Pyrimidine | Single-ring base: Cytosine, Thymine |
| Restriction enzyme | Endonuclease cutting dsDNA at a specific recognition sequence |
| Sticky ends | Single-stranded overhangs produced by staggered restriction |

## Self-Assessment Answer Key

### Module 1

1. Complement of `5'-ATCGGCTA-3'` is `3'-TAGCCGAT-5'`, or written conventionally $5' \to 3'$: `5'-TAGCCGAT-3'`. Note the reversal: complement then reverse.
2. G:C pairs have 3 hydrogen bonds versus 2 for A:T. More bonds to break means more energy (higher $T_m$) required.
3. In a double helix, if the top strand runs left→right as $5' \to 3'$, the bottom (complementary) strand runs right→left as $5' \to 3'$ — opposite direction. Each strand's 3' end is adjacent to the other strand's 5' end.
4. Purines: Adenine, Guanine (double-ring structures). Pyrimidines: Cytosine, Thymine (single-ring structures).

### Module 2

1. `GCATGCATGC`: 5 G/C and 5 A/T. Wallace rule: $T_m \approx 4(5) + 2(5) = 20 + 10 = 30°\text{C}$.
2. Wallace rule sums base counts without sequence context. The nearest-neighbor model accounts for stacking interactions between adjacent pairs, which vary by dinucleotide sequence context. This matters significantly for PCR primer annealing efficiency.
3. $T_m = \Delta H^\circ / (\Delta S^\circ + R \ln(C_T/4))$. Doubling $C_T$ increases $\ln(C_T/4)$, which increases the denominator, which slightly *increases* $T_m$. Higher concentration favors the bimolecular association (Le Chatelier's principle).
4. If the toehold $T_m$ is well below operating temperature, the toehold is single-stranded and available for invader binding, but does not spontaneously form spurious duplexes. This gives kinetic control: the gate responds to the correct trigger strand but not to background.

### Module 3

1. $\Delta G = 0$ when $\Delta H = T\Delta S$. $T = \Delta H / \Delta S = -200 / -0.6 = 333$ K $= 60°\text{C}$. Yes, this is the $T_m$ by definition.
2. Yes. $\Delta G = \Delta G^\circ + RT \ln Q$. If $Q \ll K_{eq}$, then $\ln Q$ is a large negative number, making $\Delta G < 0$ even when $\Delta G^\circ > 0$. Cells maintain $Q \ll K_{eq}$ by continuously consuming products.
3. An enzyme does not appear in the overall $\Delta G$ calculation because it is regenerated unchanged — it is not a reactant or product. It only lowers $\Delta G^\ddagger$ (the transition state energy), not $\Delta G$ of the overall reaction.
4. At high temperature, the $T\Delta S$ term dominates. Since $\Delta S < 0$ for hybridization, $-T\Delta S$ becomes large and positive, eventually making $\Delta G > 0$ and favoring denaturation.

### Module 4

1. No. DNA polymerase requires a primer with a free 3'-OH. With no primer and a circular template, there is no 3'-OH to extend from. The polymerase cannot initiate de novo.
2. No. *EcoRI* produces `5'-AATT` overhangs; *BamHI* produces `5'-GATC` overhangs. These are not complementary to each other and will not hybridize. Ligation would fail.

3. A **nick** is a break in the phosphodiester bond with no missing nucleotides (the two ends are in contact on the template). A **gap** has one or more missing nucleotides. Ligase can seal nicks only; gaps require polymerase fill-in first.
4. Expected frequency = $1/4^6 = 1/4096$ sites per bp. In 40,000 bp: $40000/4096 \approx 9.8$ sites. Approximately 10 *EcoRI* sites expected.

## Module 5

1. Smaller fragments migrate farther. Sample B (500 bp) migrated farthest, then Sample A (1.5 kb). Sample C shows no migration (no band). Distance from well: B > A > C (none).
2. The ladder is essential. Without knowing the exact run conditions, you cannot assign absolute sizes to bands. The ladder calibrates the distance-to-size relationship for that specific gel and run.
3. Most likely the restriction digest failed. Possible causes: enzyme was inactive (expired, stored incorrectly), buffer was wrong, incubation time was insufficient, or the plasmid lacked the expected *EcoRI* sites (wrong plasmid prep). The single 5000 bp band is the intact uncut plasmid.
4. Polyacrylamide gels have smaller pore sizes and much higher resolving power for small fragments. The tighter matrix creates greater differential migration for fragments of nearly identical size, allowing single-base resolution.