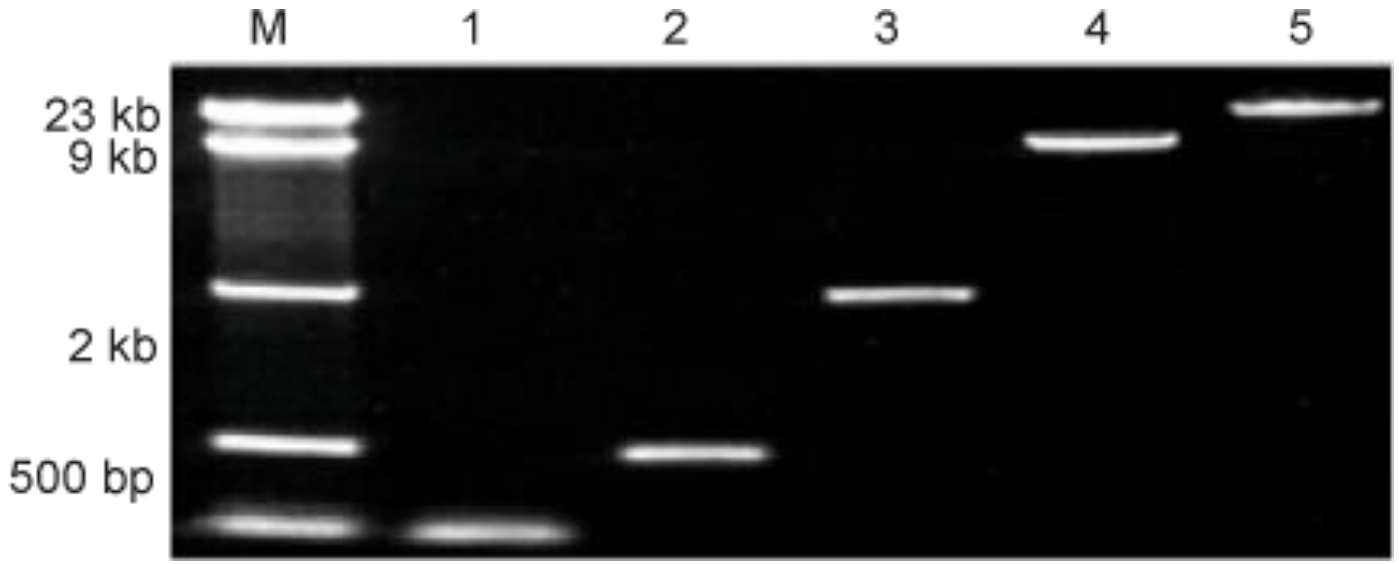
# Quantitative Cellular and Molecular Biology Laboratory
# Computational Biology Department
# Comp Bio 02-261

## PCR Primer Design Lab Introduction

# Gel Electrophoresis

- Experimental method to determine distribution of DNA strand sizes in DNA sample.
- More details in the next lecture…



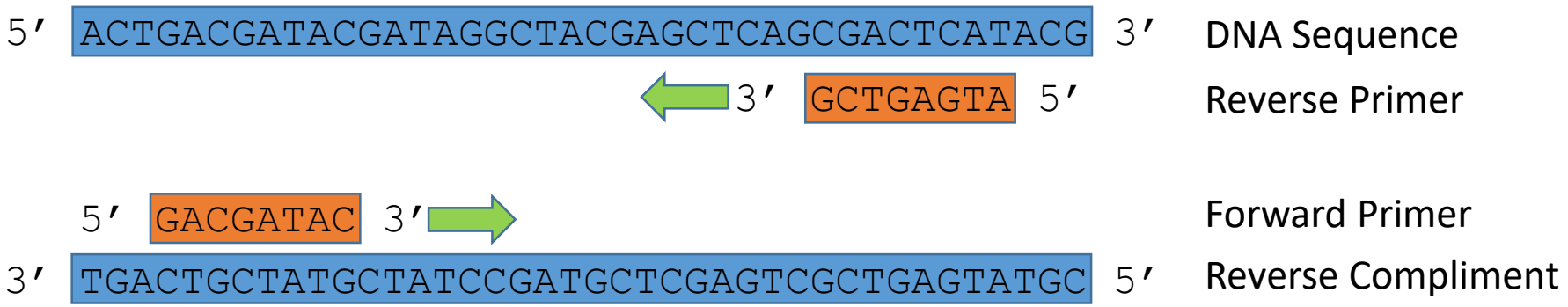
# DNA Notation (primary and secondary structure)

ssDNA = 5’-ACTGCGATAGACGATGTCCGGATGACA-3’ ← Shows sequence

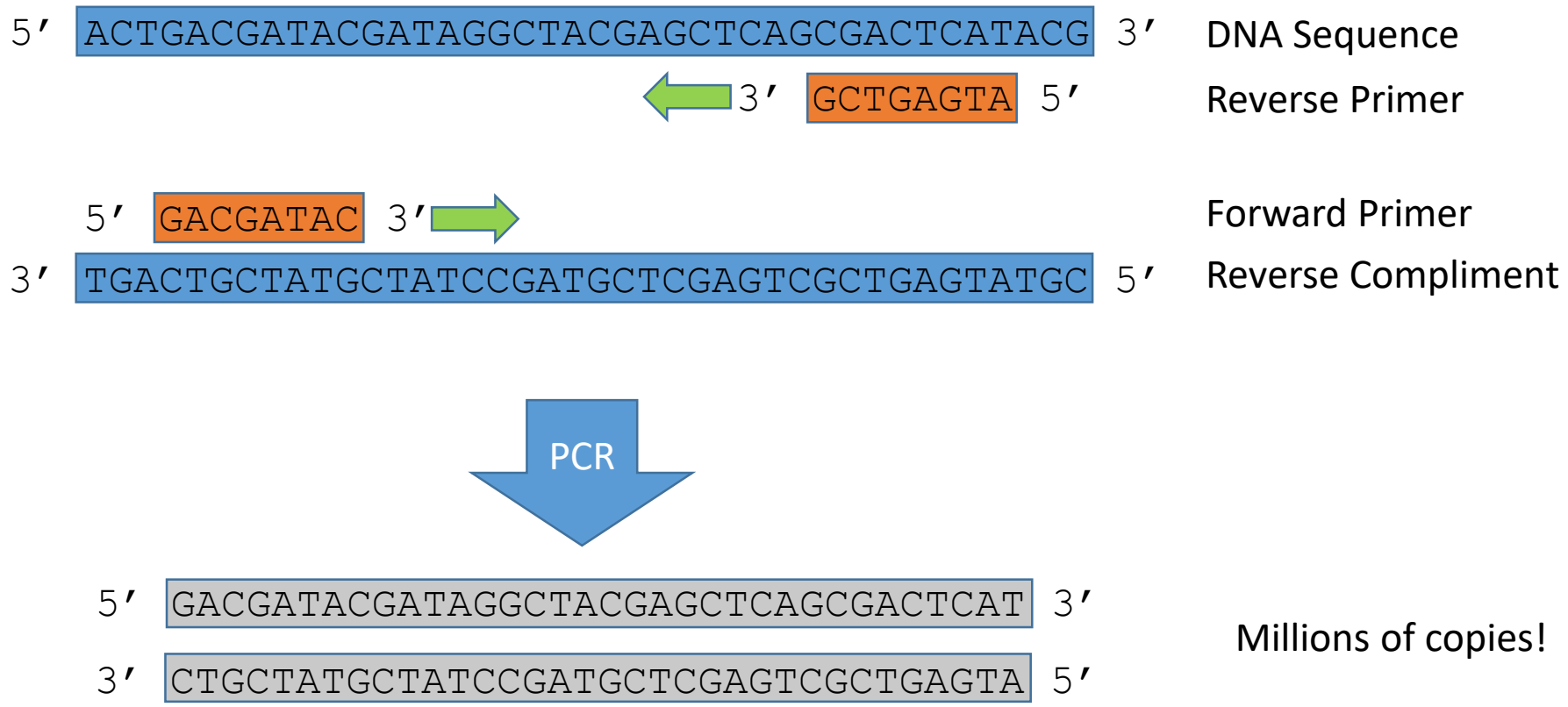
dsDNA = 5’-ACTGCGATAGACGATGTCCGGATGACA-3’
3’-TGACGCTATCTGCTACAGGCCTACTGT-5’ ← Shows sequence and pairing

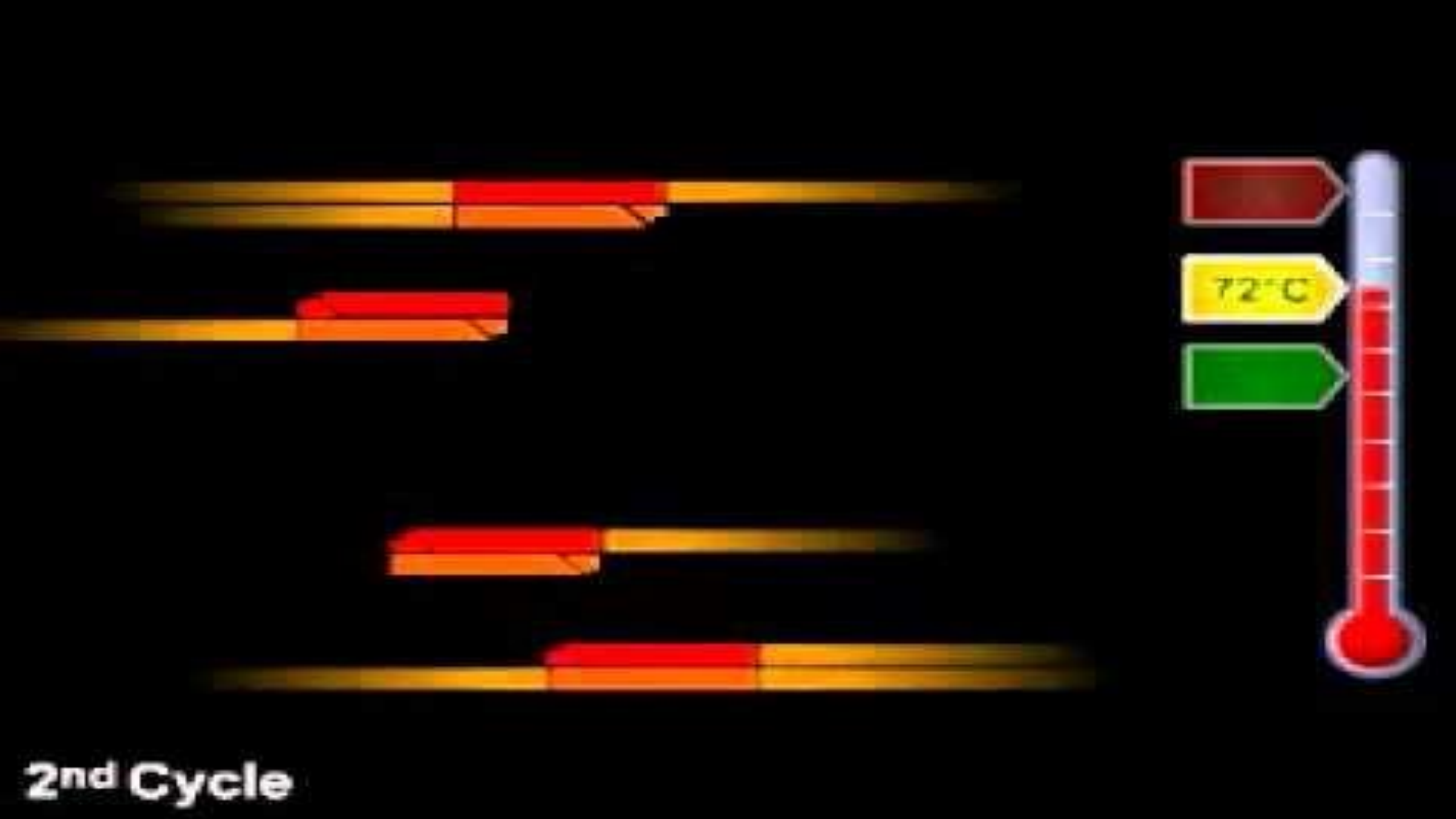
dsDNA = 5’-  -3’
3’--5’ ← Shows pairing

# Polymerase Chain Reaction



# Polymerase Chain Reaction




72°C
2nd Cycle

# Video URL

https://www.youtube.com/watch?v=YJKYSlJREIc

# Tasks for Computational Lab

1. Generate features to allow prediction of primer melting points
2. Implement function for predicting PCR products
3. Design primers for PCR reaction to identify three types of DNA

# Task 1 – Primer Melting Point Prediction

**Features**: Numerical descriptors of an object

Design ***features*** to help predict the melting point for a primer. Implement your feature calculation methods.

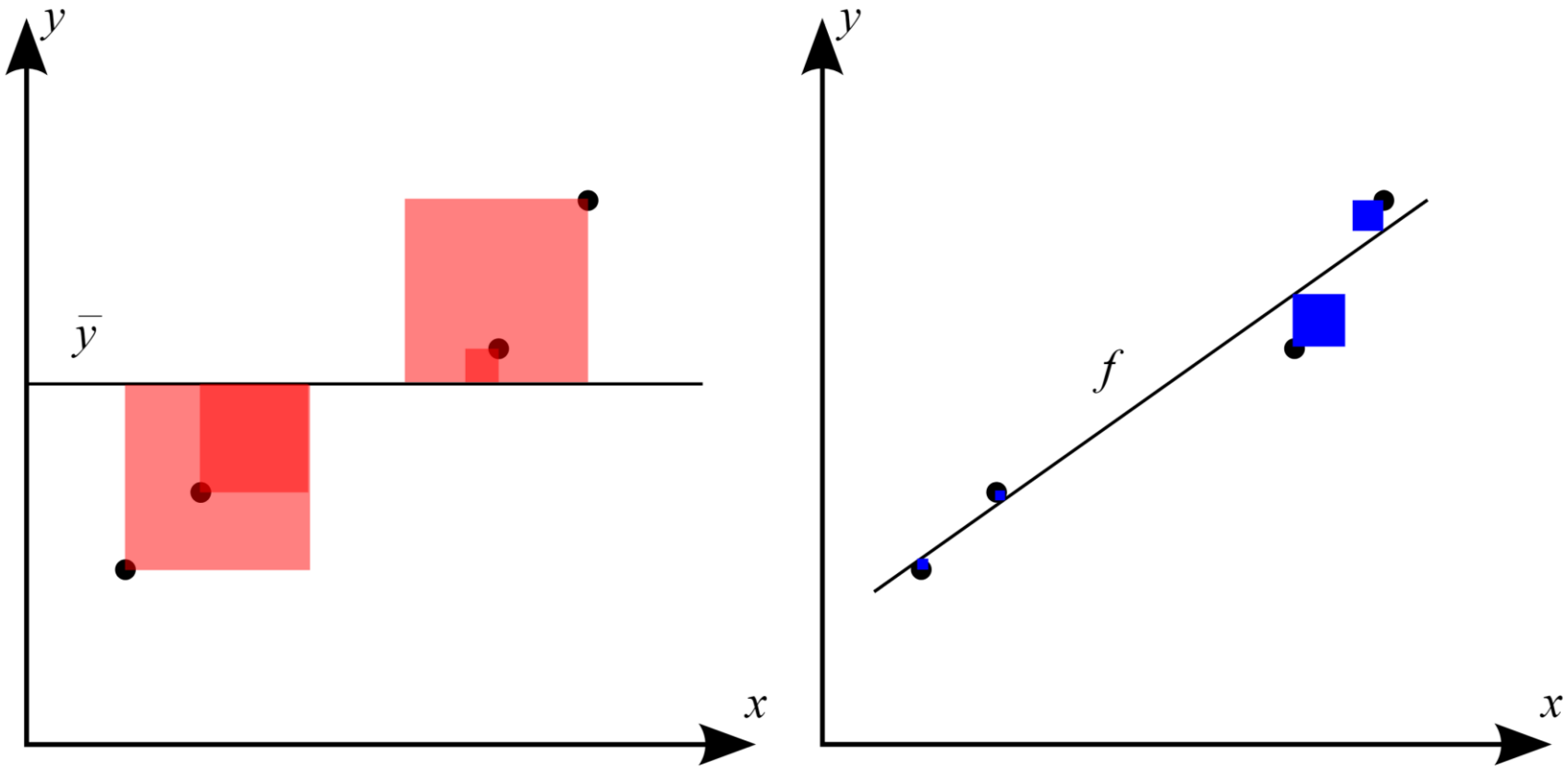
Assess with N-fold cross-validation using a RandomForest regressor model for generating predictions.

- Regressor/regression – predict continuous value
- Classifier – predict discrete class

# How to design features for predicting melting point?

http://www.premierbiosoft.com/tech_notes/PCR_Primer_Design.html

# Assessing Accuracy of Predictions




$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

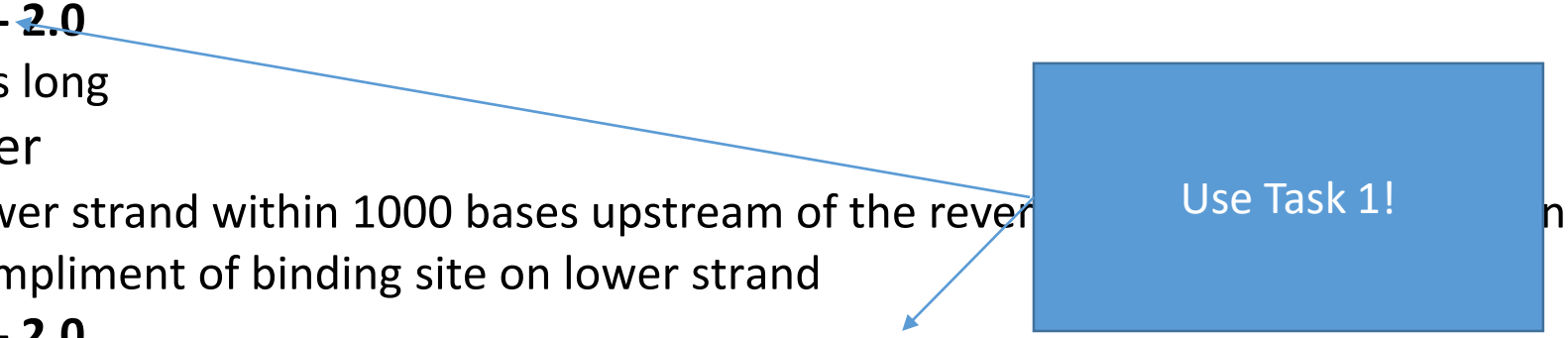
$R^2$ values closer to 1.0 are better.

# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product. If so, predict the resulting product sequence (upper strand).
- Important Primer Pair Characteristics:
  - Reverse Primer
    - Binds to upper strand
    - Reverse compliment of binding site on upper strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-35 bases long
  - Forward Primer
    - Binds to lower strand within 1000 bases upstream of the reverse primer binding location
    - Reverse compliment of binding site on lower strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-35 bases long

# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product.  If so, predict the resulting product.
- Important Primer Pair Characteristics:
  - Reverse Primer
    - Binds to upper strand
    - Reverse compliment of binding site on upper strand
    - **$T_m$ ~ 60$^o$C +- 2.0**
    - 18-25 bases long
  - Forward Primer
    - Binds to lower strand within 1000 bases upstream of the rever
    - Reverse compliment of binding site on lower strand
    - **$T_m$ ~ 60$^o$C +- 2.0**
    - 18-25 bases long



# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product.  If so, predict the resulting product.
- Important Primer Pair Characteristics:
  - Reverse Primer
    - Binds to upper strand
    - **Reverse compliment of binding site on upper strand**
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-25 bases long
  - Forward Primer
    - Binds to lower strand within 1000 bases upstream of the reverse pr
    - **Reverse compliment of binding site on lower strand**
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-25 bases long

Reverse Complement:
ACTG -> CAGT

Complementary Base Pairs:
A <-> T
G <-> C

# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product. If so, predict the resulting product.
- Important Primer Pair Characteristics:
  - Reverse Primer
    - **Binds to upper strand**
    - Reverse compliment of binding site on upper strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-35 bases long
  - Forward Primer
    - **Binds to lower strand within 1000 bases upstream of the reverse primer binding location**
    - Reverse compliment of binding site on lower strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-25 bases long

How do we determine binding?

# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product. If so, predict the resulting product.
- Important Primer Pair Characteristics:
  - Reverse Primer
    - **Binds to upper strand**
    - Reverse compliment of binding site on upper strand
    - $T_m$ ~ 60$^o$C +- 1.5
    - 18-25 bases long
  - Forward Primer
    - **Binds to lower strand within 1000 bases upstream of the reverse primer binding location**
    - Reverse compliment of binding site on lower strand
    - $T_m$ ~ 60$^o$C +- 1.5
    - 18-25 bases long

How do we determine binding?

(Local) Sequence Alignment!

# Task 2 – Predict PCR Products (Alignment)

> alignment.local_align("ACTG", "ACTG", print_output = True)
Scoring: match = 10; mismatch = -5; gap_start = 0; gap_extend = -7
A matrix =

| | * | A | C | T | G |
|---|---|---|---|---|---|
| * | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 10 | 3 | 0 | 0 |
| C | 0 | 3 | 20 | 13 | 6 |
| T | 0 | 0 | 13 | 30 | 23 |
| G | 0 | 0 | 6 | 23 | **40** |

Best Alignment:
ACTG
ACTG

Optimal Score = 40
Max location in matrix = (4, 4)

# Task 2 – Predict PCR Products (Alignment)

> alignment.local_align("ACTGACTGACTG", "ACTG", print_output = True)

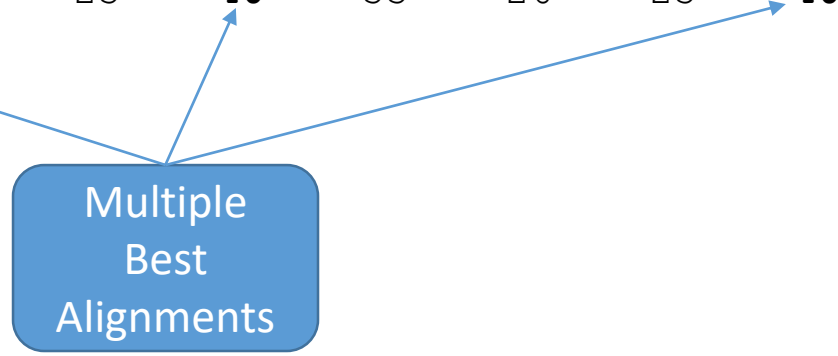
Scoring: match = 10; mismatch = -5; gap_start = 0; gap_extend = -7

A matrix =

| | * | A | C | T | G | A | C | T | G | A | C | T | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 10 | 3 | 0 | 0 | 10 | 3 | 0 | 0 | 10 | 3 | 0 | 0 |
| C | 0 | 3 | 20 | 13 | 6 | 3 | 20 | 13 | 6 | 3 | 20 | 13 | 6 |
| T | 0 | 0 | 13 | 30 | 23 | 16 | 13 | 30 | 23 | 16 | 13 | 30 | 23 |
| G | 0 | 0 | 6 | 23 | **40** | 33 | 26 | 23 | **40** | 33 | 26 | 23 | **40** |

Optimal Score = 40

Max location in matrix = (12, 4)



# Task 2 – Predict PCR Products (Alignment)

> alignment.local_align("AGTCACTGGCTT", "ACTG", print_output = True)

Scoring: match = 10; mismatch = -5; gap_start = 0; gap_extend = -7

A matrix =

| | * | A | G | T | C | A | C | T | G | G | C | T | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 10 | 3 | 0 | 0 | 10 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 3 | 5 | 0 | 10 | 3 | 20 | 13 | 6 | 0 | 10 | 3 | 0 |
| T | 0 | 0 | 0 | 15 | 8 | 5 | 13 | 30 | 23 | 16 | 9 | 20 | 13 |
| G | 0 | 0 | 10 | 8 | 10 | 3 | 6 | 23 | **40** | 33 | 26 | 19 | 15 |

Optimal Score = 40

Max location in matrix = (8, 4)

Position in String 1 of the last character in optimal alignment

Position in String 2 of the last character in optimal alignment

Best Alignment:
----ACTG----
ACTG

Best Score:
40/40
Best score possible for alignment of 4 characters.

Binding defined by 80% + alignment.

# Local Alignment Function

def local_align(x, y, score=ScoreParam(10, -5, -7), print_output = False):

x = sequence 1

y = sequence 2

score = Score Parameter (match = +10, mismatch = -5, gap = -7)

(optional)

print_output = binary indicating whether or not you want pretty output printed from alignment

(optional)

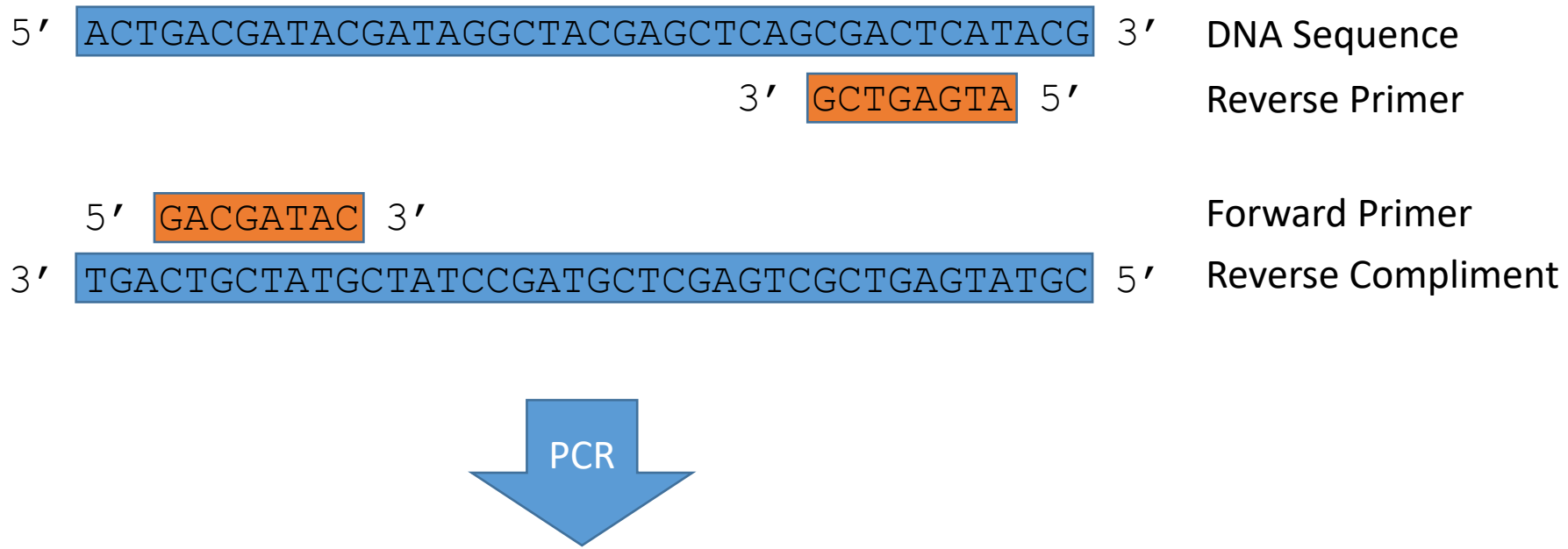
# Task 2 – Predict PCR Products

- Given a sequence and a pair of primers, predict whether or not there will be a product. If so, predict the resulting product.
- Important Primer Pair Characteristics:
  - Reverse Primer
    - **Binds to upper strand (80%+ alignment)**
    - Reverse compliment of binding site on upper strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-30 bases long
  - Forward Primer
    - **Binds to lower strand within 1000 bases upstream of the reverse primer binding location (80%+ alignment)**
    - Reverse compliment of binding site on lower strand
    - $T_m$ ~ 60$^o$C +- 2.0
    - 18-30 bases long

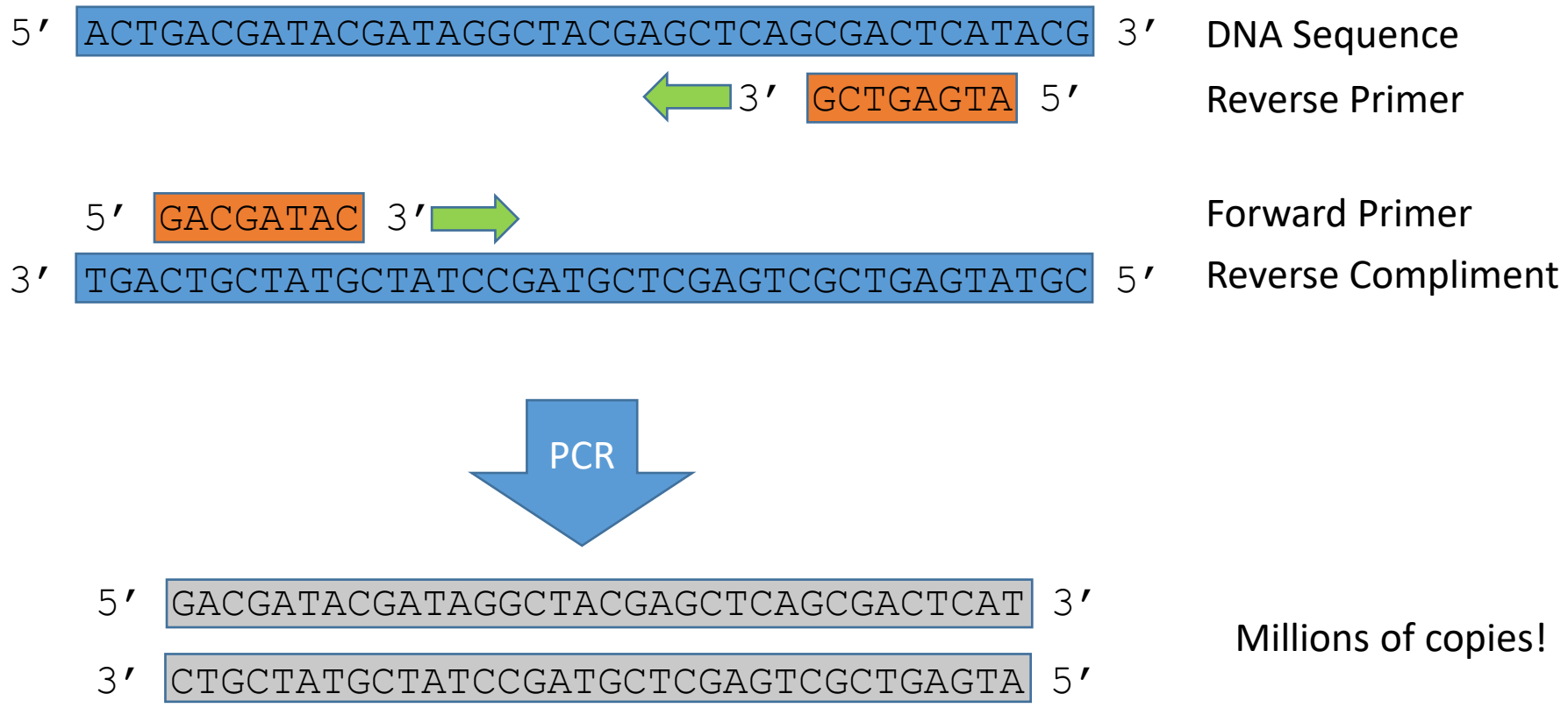
How do we determine binding?

(Local) Sequence Alignment!

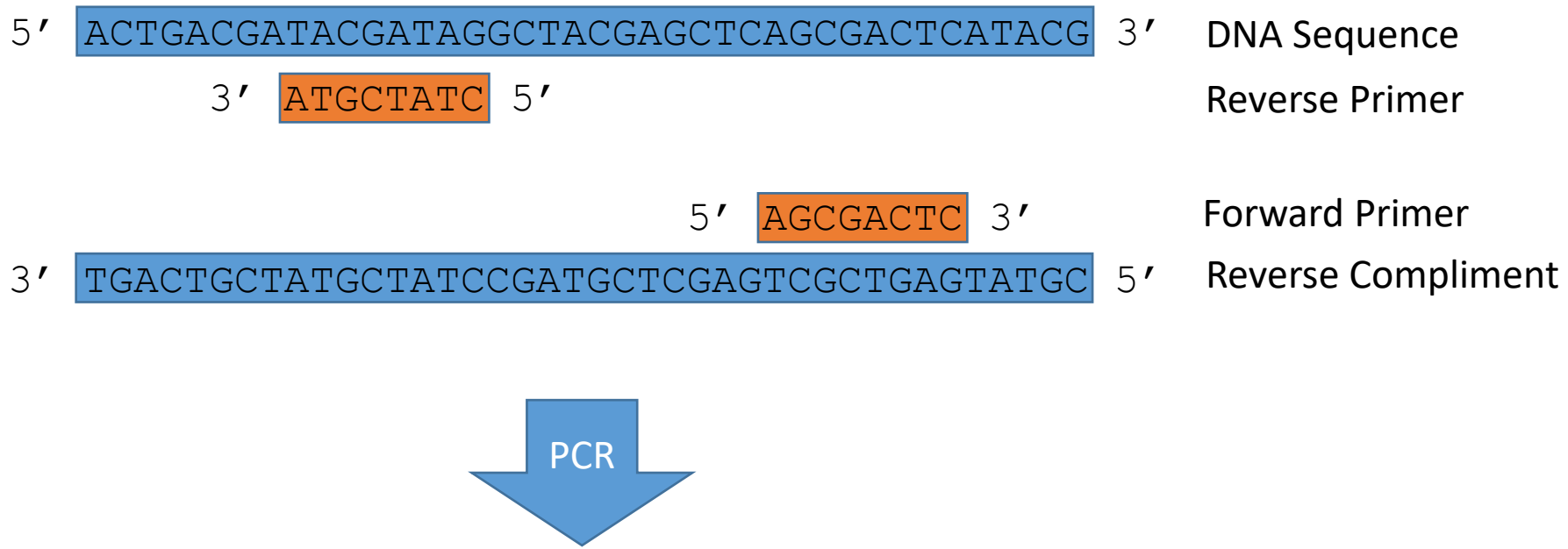
# Polymerase Chain Reaction Quiz!



# Polymerase Chain Reaction Quiz!

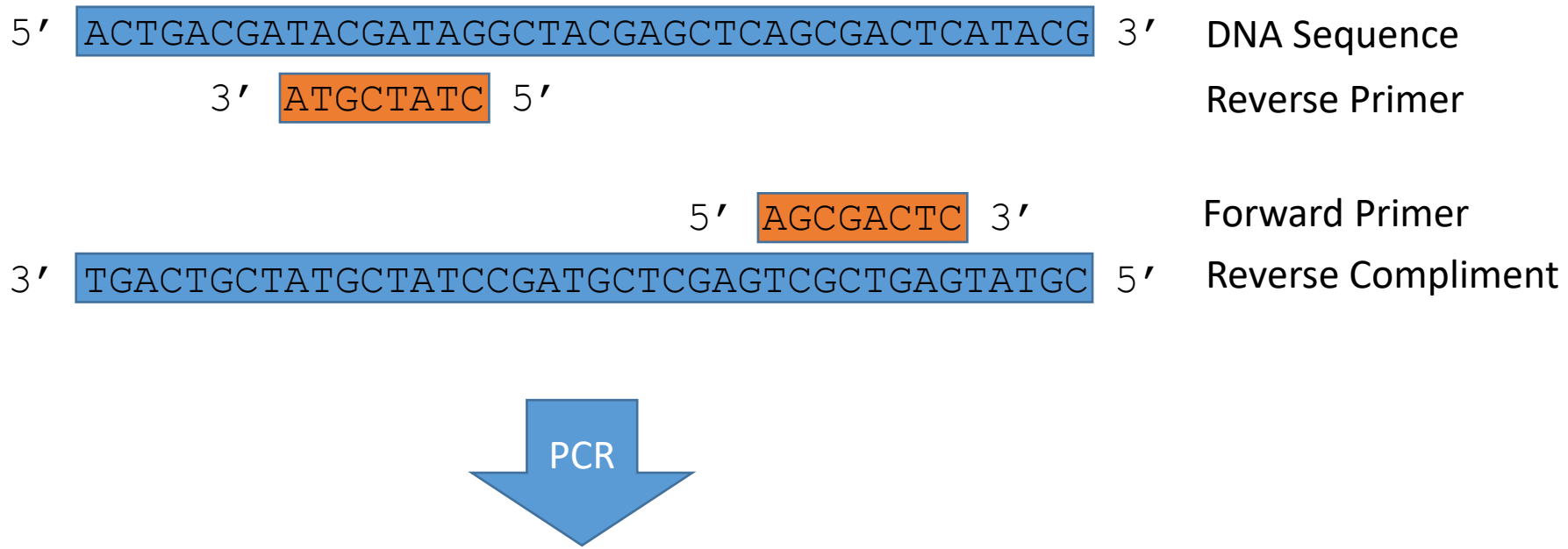


# Polymerase Chain Reaction



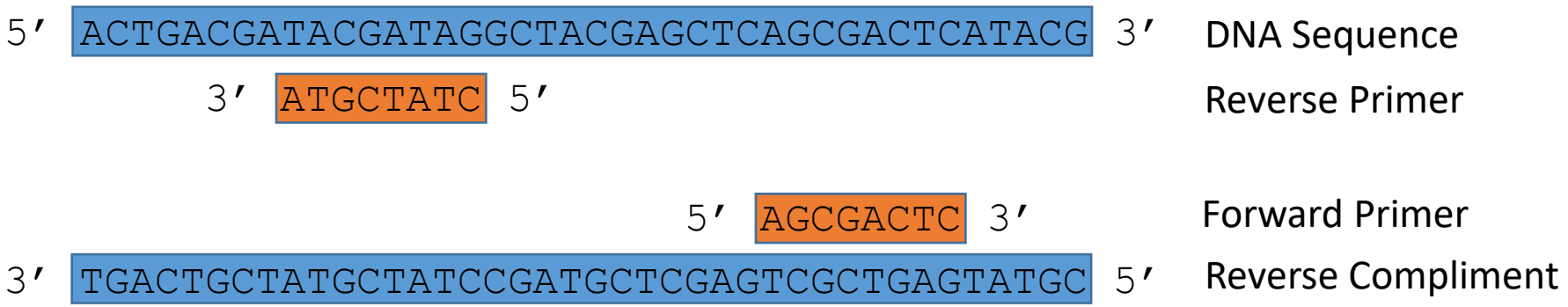

What product would we see?

# Polymerase Chain Reaction




What product would we see?

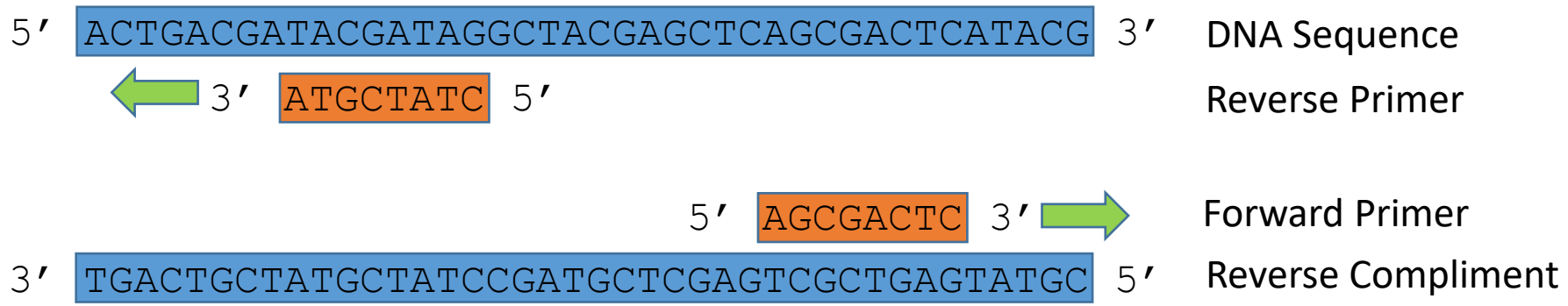
# Polymerase Chain Reaction




PCR

What product would we see?
NO PRODUCT. Why?

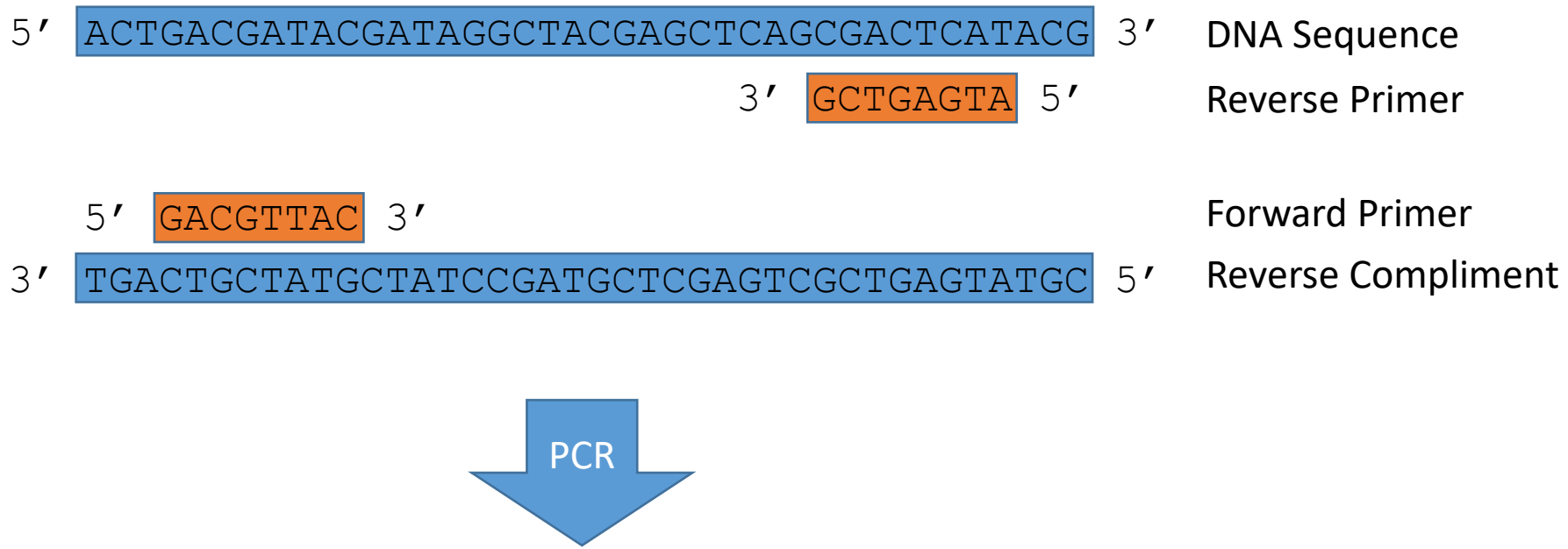
# Polymerase Chain Reaction



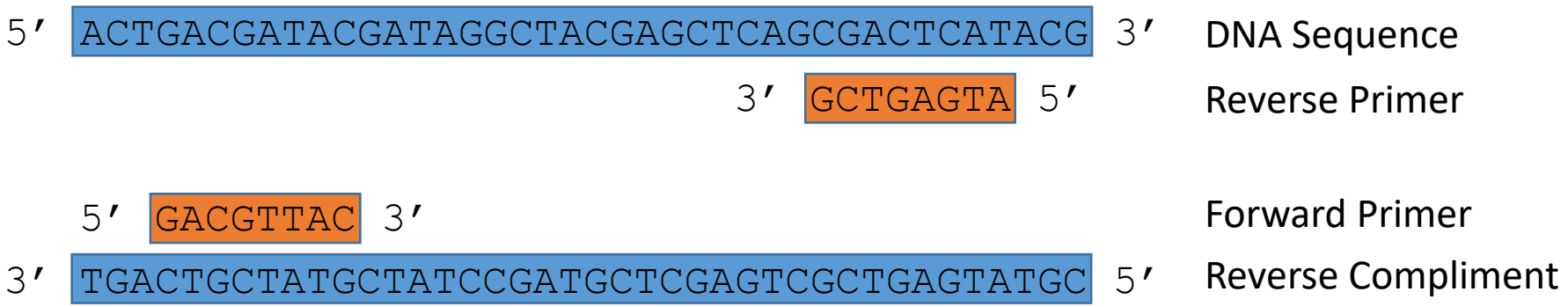

PCR

What product would we see?
NO PRODUCT. Why?
Polymerase extension (5'-> 3') would not yield geometric amplification.

# Polymerase Chain Reaction Quiz!



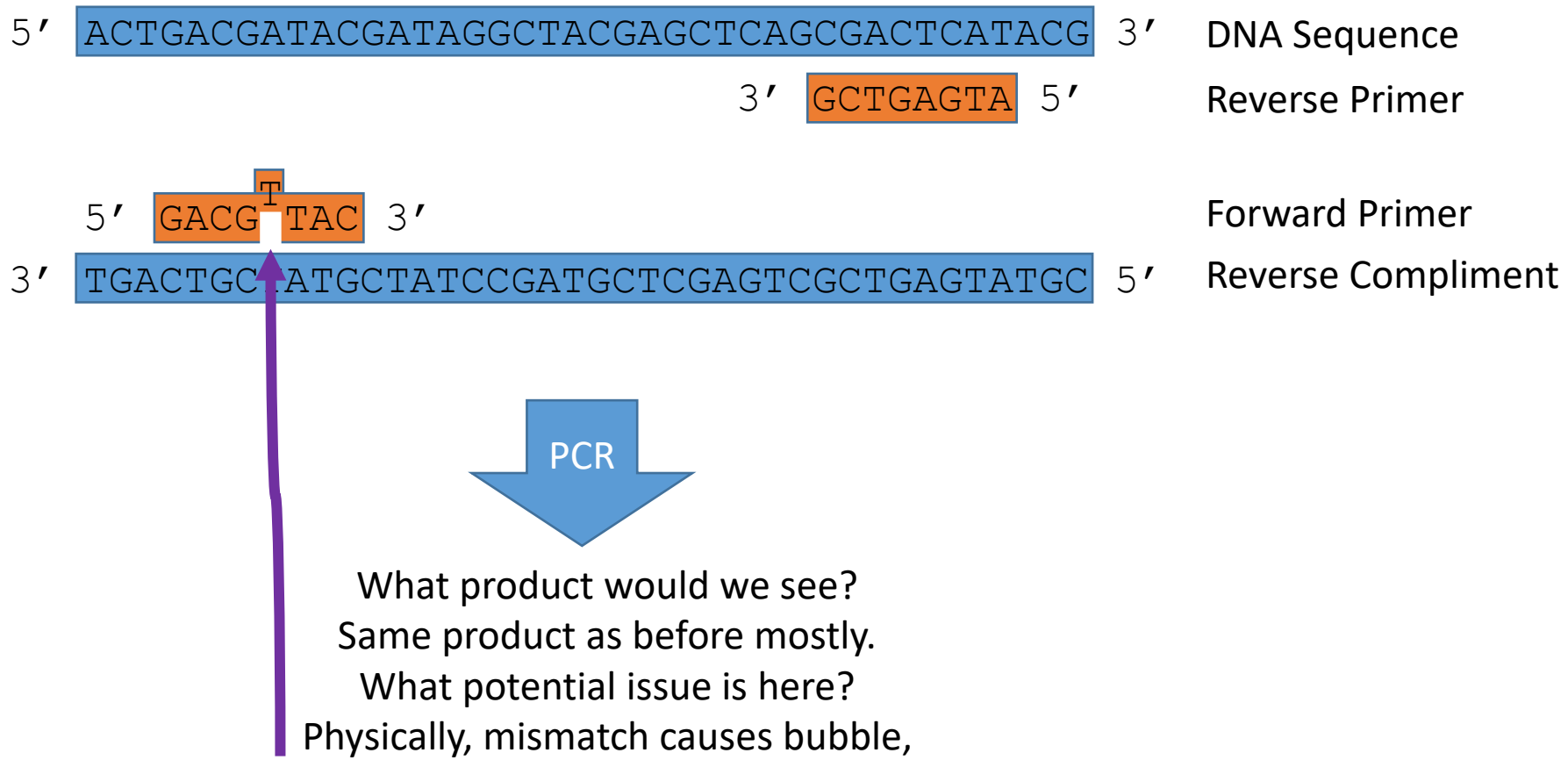
# Polymerase Chain Reaction Quiz!




PCR

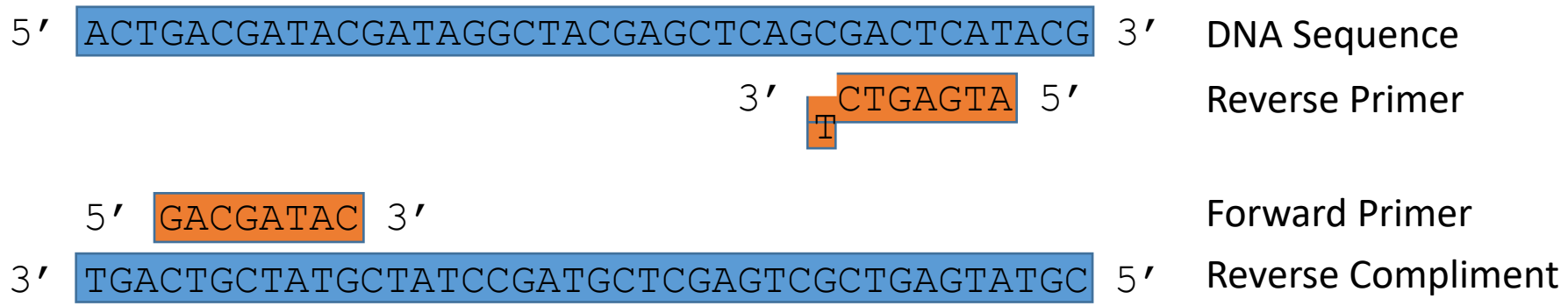
What product would we see?
Same product as before mostly.
What potential issue is here?

# Polymerase Chain Reaction Quiz!




What product would we see?
Same product as before mostly.
What potential issue is here?
Physically, mismatch causes bubble,
but probably no issues.

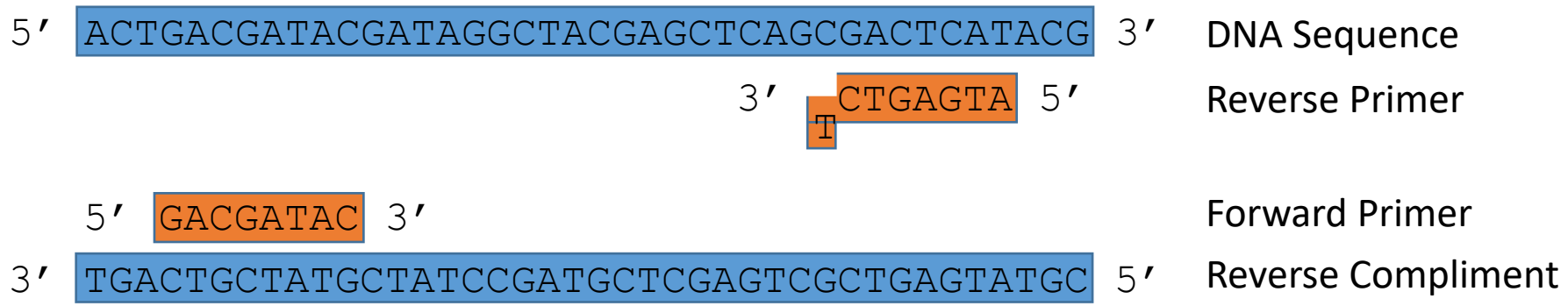
# Polymerase Chain Reaction Quiz!




PCR

What product would we see?

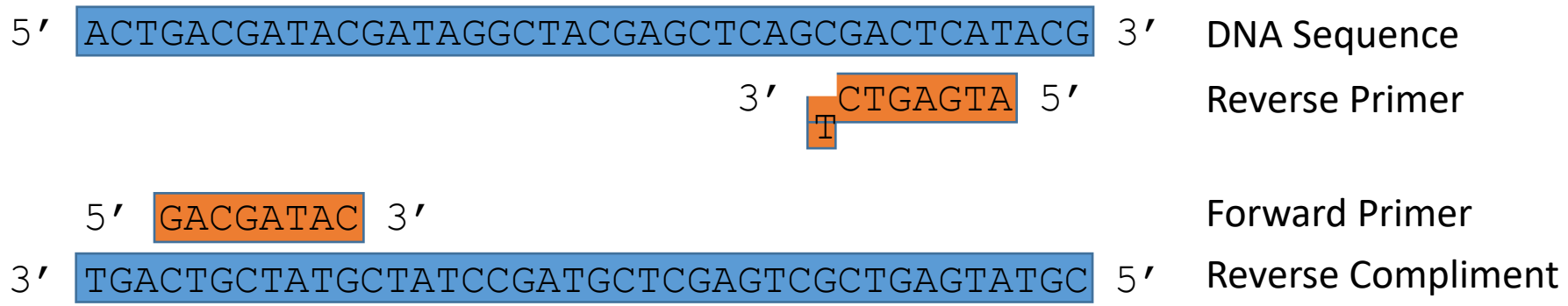
# Polymerase Chain Reaction Quiz!




PCR

What product would we see?
No product. Polymerase can't bind correctly to induce extension.

# Polymerase Chain Reaction Quiz!




PCR

What product would we see?
No product. Polymerase can't bind correctly to induce extension.

For the purposes of this assignment, ignore issues with bumps and first base mismatch.

# PCR for Bacteria Identification

# Do I want to identify all species of bacteria present in a sample (next gen sequencing) or determine whether a single species is present?

Both require PCR…

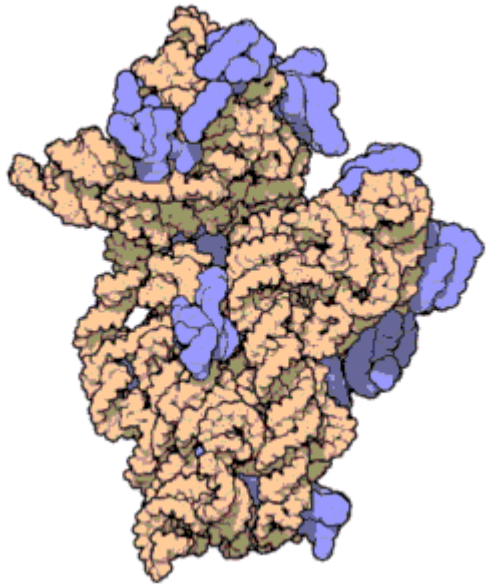
# Identification of all bacteria present…

For sequencing, we want to make lots of copies of DNA for all bacteria present.  Then we can sequence these.

The best way to do this is to look for a common gene across all bacteria and use PCR to make copies of that gene.

# 16s ribosomal RNA subunit

- Portion of ribosome (purple).
- Ribosomal function is essential for life so the sequence is mostly conserved across bacterial species.
- But we still see mutations in this gene. Why?



# Task 3

# How would we design our PCR primers for sequencing?

- We want to make copies of DNA in the 16s rRNA gene regardless of the species.

# How would we design our PCR primers for sequencing?

- We want to make copies of DNA in the 16s rRNA gene regardless of the species.


What parts of our gene allow us to identify species?

# How would we design our PCR primers for sequencing?

- We want to make copies of DNA in the 16s rRNA gene regardless of the species.


What parts of our gene allow us to identify species?

Where should we design our primers to bind such that when we sequence the PCR product, we can identify any species?

- Generate a function which takes as input a list of sequences (strings) and returns a forward and reverse primer which will generate a PCR product for sequences in the list.

# Task 4

# How would we design our PCR primers to detect presence of each of three bacterial species?


Strategy 1: Design primers which yield products of unique length for each species.

Strategy 2: Design primers to yield products specifically for each species.

- Generate a function which takes as input a list of sequences (strings) and returns a set of primer pairs which will generate products for each sequence specifically (or *None* if the task is impossible).

# Strategy Hints:

*n* = number of sequences in list

- 2 primers:

- *n* + 1 primers:

- 2*n* primers:

# Task 5

- For your group’s bacterial sequences, run Task 3 and Task 4. Generate a text file where each line contains an identifier and the sequence separated by a space.

- Example:

Group_10_JK_Task1_FWD ACTGCTACGGACGACT

Group_10_JK_Task1_REV TCAGCGACGAACGCTCT

# Task 6

- Describe the design of an experiment to determine whether or not any of your *n* bacteria in Task 5 are present in a random sample of DNA.  Be sure to include the list of reactions including the template DNA and primers used for each.