

DNA

สายดีเอ็นเอประกอบด้วยลำดับของนิวคลีโอไทด์ 4 ชนิด แสดงด้วยสตริงของตัวอักษร 'A', 'T', 'G' และ 'C' เช่น "AAAACCCGGT" โดยนิวคลีโอไทด์ 'A' จะมี 'T' เป็นคู่สับ และนิวคลีโอไทด์ 'G' จะมี 'C' เป็นคู่สับ ถ้าสายดีเอ็นเอมีตัวอักษรอื่นแทรกอยู่ จะถือว่าสายดีเอ็นเอนั้นไม่ถูกต้อง (Invalid DNA) อักษรตัวพิมพ์เล็กและตัวพิมพ์ใหญ่ไม่ต่างกัน ในโจทย์ข้อนี้ ให้นิสิตเขียนโปรแกรมเครื่องประมวลผลสายดีเอ็นเอ

ข้อมูลนำเข้า

บรรทัดแรก สายดีเอ็นเอในรูปแบบของสตริง

บรรทัดที่สอง คือชื่อโอเปอเรเตอร์ โดยมีทั้งหมด 3 โอเปอเรเตอร์ที่รองรับคือ

- 1) R คือหา reverse complement ของสายดีเอ็นเอ โดยโอเปอเรเตอร์นี้จะทำการเปลี่ยนนิวคลีโอไทด์ในแต่ละลำดับให้เป็นคู่สับของมัน และเมื่อเปลี่ยนเสร็จแล้วต้องกลับด้านสายดีเอ็นเอใหม่นี้ด้วย เช่น
AAAACCCGGT จะถูกเปลี่ยนเป็น TTTTGGGCCA และจากนั้นแสดงผลกลับด้านจากขวามาซ้ายเป็น
ACCGGTTTT
- 2) F คือหาความถี่ของแต่ละนิวคลีโอไทด์ แสดงตามลำดับ A, T, G, C เช่น AAAACCCGGT จะได้ผลลัพธ์เป็น
A=4, T=1, G=2, C=3 เป็นต้น (สังเกตว่า มีช่องว่าง 1 ช่องหลังเครื่องหมายคอมมาด้วย !!)
- 3) D คือการหาจำนวนคู่ของสองนิวคลีโอไทด์ที่อยู่ติดกัน ดังนั้น ถ้าเป็นโอเปอเรเตอร์นี้ ต้องมีการรับข้อมูลเพิ่มอีก 1 บรรทัด คือคู่ของสองนิวคลีโอไทด์ที่สนใจ เช่น GC, AA เป็นต้น โดย GC ไม่เท่ากับ CG และในการนับจำนวนคู่ จะเทียบการหาคู่ไปทีละ 1 นิวคลีโอไทด์ เช่น AAAA แล้วต้องการหา AA จะได้ 3 คู่ เป็นต้น

Hint ให้ตรวจสอบ input เฉพาะกรณีที่เกิด Invalid DNA เท่านั้น รับประกันว่าโอเปอเรเตอร์จะถูกต้อง

*** อย่าลืม ต้อง strip() ข้อมูลจาก input() ก่อนนำไปประมวลผล

ข้อมูลส่งออก

ข้อมูลส่งออกมีลักษณะแตกต่างกันไปตามโอเปอเรเตอร์ที่ถูกเรียกใช้งาน

หมายเหตุ สีของแต่ละนิวคลีโอไทด์มีไว้เพื่อให้สังเกตแต่ละตัวได้ง่ายขึ้นเท่านั้น

ตัวอย่าง	
input (จากแป้นพิมพ์)	output (ทางจอภาพ)
ATTTGCGGCATATCC R	GGATATGCCGCAAA T
ATTTGCGGCATATCC F	A=3, T=5, G=3, C=4
ATTTGCGGCATATCC D GC	2
ATTTGCGGCATATCC D TT	2
ATTTGCGGCATATCC F	Invalid DNA
atTTTgcggcAtaTCC R	GGATATGCCGCAAA T