

НАЦИОНАЛЕН ЕСЕНЕН ТУРНИР ПО ИНФОРМАТИКА И ИНФОРМАЦИОННИ ТЕХНОЛОГИИ
“Джон Атанасов”

Шумен, 27 ноември 2004г

ТЕМА ПО ИНФОРМАТИКА ЗА 12 КЛАС (ГРУПА А)

Задача А2. ПРОЧИТ НА ГЕН

От биологическата наука е известно, че всеки ген се състои от наредени кодони, а всеки кодон се състои от три наредени азотни бази. Азотните бази биват четири вида, които се бележат с главните латински букви **A**, **C**, **G** и **T**. Всеки кодон кодира аминокиселина. Еднаквите кодони кодират еднакви аминокиселини, а за целите на тази задача ще приемем, че различните кодони кодират различни аминокиселини. Аминокиселините, наредени в редица, образуват белтък. Когато клетъчният апарат прочита даден ген, той прочита **някои** от кодоните и конструира съответните им аминокиселини. След това нарежда аминокиселините в същия ред в който са съответните им кодони върху гена и така конструира белтък. Броя на прочетените кодони може да бъде от 1 до броя на кодоните в гена.

Напишете програма **GENE**, която по зададен ген намира броя на **различните** белтъци, които могат да бъдат конструирани от него.

Входните данни се четат от стандартния вход и се състоят от единствен ред, съдържащ последователност от главни латински букви измежду **A**, **C**, **G** и **T** — азотните бази на гена. Броят на азотните бази се дели на три и е не по-голям от 12 000 000 (т.е. в гена има не повече от 4 000 000 кодона). В 50% от тестовите броят на азотните бази ще е не по-голям от 1 500 000 (т.е. не повече от 500 000 кодона).

Резултатът трябва да бъде изведен на стандартния изход. За ваше улеснение, не трябва да извеждате целия резултат, а само остатъка му по модул 1 048 576 (това е 2 на степен 20). Изведеното на стандартния изход трябва да съдържа единствен ред с единствено число на него — броя на **различните** белтъци, които могат да се конструират от гена, по модул 1 048 576.

ПРИМЕР 1**Вход**

ACGTTTACG

Изход

6

Възможни прочетени кодони

ACG

ACGACG

ACGTTT

ACGTTTACG

TTT

TTTACG

ПРИМЕР 2**Вход**

AGGCTACTACTA

Изход

7

Възможни прочетени кодони

AGG

AGGCTA

AGGCTACTA

AGGCTACTACTA

CTA

CTACTA

CTACTACTA