

Actividades a entregar¹

1. Imaginemos que el acceso a un banco de datos genómico se realiza introduciendo un **identificador** (número decimal) y un código alfanumérico de tres caracteres. El sistema tiene que comprobar si son correctos, accediendo a un *string* que almacena de forma consecutiva por número de id dichos códigos. Realizar un programa que devuelva el código a partir del identificador (que siempre será correcto, con valores entre 1 y 12). No se pueden usar otros tipos de datos como listas, tuplas o diccionarios. Por ejemplo: para 4 usuarios, si el identificador es 3, el *string* devuelto sería HAL:

1	2	3	4
U	D	C	B
B	8	H	A
L	H	2	0

2. Modifica el ejercicio previo para usar un diccionario en lugar del *string*.
3. Escribir en un programa sentencias para que dada una cadena de texto:
 - a. Eliminar el guion bajo y poner en mayúscula la primera letra a continuación.
 - b. Devolver TRUE si es reversible o FALSE en caso contrario. La cadena NO contendrá acentos. P.ej.: Anita lava la tina
 - c. Devolver TRUE si contiene todas las vocales y FALSE en caso contrario. Pista: usa un conjunto.
 - d. Devolver otra cadena codificada sustituyendo cada vocal por el número de orden en 'aeiou'. La cadena NO contendrá acentos. P.ej.: 'Esto es una prueba' produce la cadena de salida '3st4 3s 5n1 pr52b1'
 - e. Dada una frase correctamente codificada según el paso previo devolverla a su formato original. Recuerda poner las mayúsculas iniciales o después del punto.
4. Escribir un programa que a partir de una secuencia ADN correcta (puede ser siempre la misma) determine e imprima por pantalla:
 - a. El número de ocurrencias de una subsecuencia (que debe existir) en la secuencia.

¹ Para la resolución sólo es posible usar conceptos del tema actual y previos. No usar ni estructuras repetitivas (bucles) ni selectivas (ifs)

- b. La posición de la primera ocurrencia de dicha subsecuencia
- c. La longitud de la secuencia y el número de bases ADN distintas
- d. El contenido de GC (guanina y citosina) de la secuencia que equivale al porcentaje de bases que son G o C.
- e. La cadena resultado de rotar a la izquierda $n > 0$ caracteres (o eliminarlos del final e insertarlos de manera inversa por el principio). P.ej.: AAAATGC con $n = 3$ produce **CGTAAAA**

La salida será similar a ésta²:

```
¿Introducir secuencia? AAAATGC
¿Introducir subsecuencia? AT
Número de ocurrencias AT = 1
Posición de la primera ocurrencia 'AT' = 3
Longitud de la secuencia = 7
Número de bases diferentes = 4
Contenido GC = 28.5%
Nº bases a rotar? 3
Cadena ADN rotada = CGTAAAA
```

5. Construir un programa que pida al usuario una secuencia ADN, realice un recuento del número de bases que contiene y de qué tipo es cada una e imprima por pantalla el resultado. La salida podría ser:

```
Introducir una secuencia: ATtACñ
Contiene 5 bases, con la siguiente distribución:
adenine: 2
thymine: 2
cytosine: 1
guanine: 0
y 1 caracteres extraños
```

La cadena puede contener mayúsculas y minúsculas. Incluir en el recuento el número de caracteres extraños (que no son A, T, G o C)

Pista: Usar **obligatoriamente** un diccionario para almacenar el recuento

² Para salida por pantalla con "formato" hay que usar el método *format* con especificadores de formato