Université Louis Pasteur

MAGISTERE DE CHIMIE-BIOLOGIE

Première année

Examen d'informatique
17 mai 1990

Durée deux heures

1. Que signifient les termes *bit*, *octet*, *byte*? Qu'est ce qu'un programme `FORTRAN`, qu'est ce qu'un fichier?

2. A chaque variable d'un programme `FORTRAN` est associé un type. Que signifient les types `integer`, `real`, `character`, `logical`?

3. A quoi sert la déclaration `implicit none`?

4. Que représentent les variables définies par:

```
integer          i, m(3,2,5)
character        a, z(10), f(5,8)
character*800    seq, zzz(10)
```

5.
   - Comment déclarer une matrice `R` de rotation dans R^3?
   - Comment déclarer un tableau `TC` devant contenir au maximum 1000 coordonnées tridimensionnelles d'atomes?
   - Comment déclarer le tableau `TN` devant contenir les noms de ces atomes (20 caractères au maximum par atome).

6. Ecrire le sous-programme `lecture_sequence(seq,long)` qui doit:

   - demander à l'utilisateur quel est le nom du fichier contenant la séquence à lire,

- ouvrir ce fichier,
- lire la séquence, sachant qu'il y a un caractère par ligne, en la stockant dans la chaîne de caractères `seq`; dans l'entier `long` figurera le nombre de caractères lus.
- fermer le fichier.

Ecrire *précisément* toutes les déclarations et instructions.

7. Le programme `recherche_ega` indique si la séquence `seq` lue dans un fichier contient exactement une petite séquence `pep` entrée par l'utilisateur. Ecrire le programme `recherche_hom` (et ses sous-programmes) pour qu'il indique s'il existe dans `seq` une sous-séquence qui soit homologue à plus de 80% à la petite séquence `pep`. On supposera qu'il existe une fonction `score`, définie par `real function score(s1,s2,l)`, qui donne le pourcentage d'homologie des deux séquences `s1` et `s2` de longueur `l`.

8. En supposant qu'une ligne de `FORTRAN` est executée en une microseconde, que la banque de données UWGCG contient environ 20000 séquences de protéines d'une longueur moyenne de 300 acides aminés, donner une durée approximative d'un programme qui dresserait la liste de toutes les séquences de la banque de données qui contiennent une sous-séquence homologue à plus de 80% à une sequence `pep` de 20 acides aminés. On supposera également que la fonction `score(s1,s2,20)` s'exécute en 40 μs.

9. On peut améliorer ces programmes pour que leur temps d'exécution soit significativement plus court. Indiquer, sans programmer, ce qu'il serait possible de faire.

10. Ecrire le programme `recherche_dup` qui teste si une séquence `seq` contient une suite de 10 acides aminés qui y figure deux fois, comme par exemple

    CTAPREMAUFOOTLESJOYEUX**MAGISTERES**CROIENT
    POUVOIRBATTRELACRISTALLOMAISCENESERAPASSISI
    MPLEPARCEQUEENFAITCESPAUVRES**MAGISTERES**VONT
    BELETBIENPERDRE