

## Perl Project 2018

Κατασκευάστε ένα πρόγραμμα που θα πραγματοποιεί πρόγνωση της θέσης μεμβρανικών τμημάτων σε διαμεμβρανικές πρωτεΐνες και ελέγξει την αποτελεσματικότητα του σε μια ομάδα πρωτεϊνών από την Uniprot. Οι γνωστής δομής μεμβρανικές πρωτεΐνες, έχουν στο πεδίο FT το χαρακτηριστικό TRANSMEM το οποίο δείχνει την θέση των διαμεμβρανικών περιοχών. Στο πρώτο στάδιο, το πρόγραμμα θα πρέπει να διαβάσει την εγγραφή και θα πρέπει να τυπώνει μια ιδιαίτερη μορφή fasta oneline, στην οποία θα υπάρχει και σήμανση. Για παράδειγμα το αρχείο θα είναι κάπως έτσι:

|    |  |           |                                |
|----|--|-----------|--------------------------------|
| ID | 140U_DROME   | Reviewed; | 261 AA.                        |
| AC | P81928; Q9VFM8;  |           |                                |
| FT | CHAIN 1  | 261       | RPII140-upstream gene protein. |
| FT |  |           | /FTid=PRO_0000064352.          |
| FT | TRANSMEM 67  | 87        | Potential.                     |
| FT | TRANSMEM 131   | 151       | Potential.                     |
| FT | TRANSMEM 183   | 203       | Potential.                     |
| FT | CONFLICT 64  | 64        | S -> F (in Ref. 1).            |
| SQ | SEQUENCE 261 AA; 29182 MW; 5DB78CF6CFC4435A CRC64;                 |           |                                |
|    | MNFLWKGRRF LIAGILPTFE GADEIVDKE NKTYKAFLAS KPPEETGLER LKQMFTIDEF   |           |                                |
|    | GSISSELNSV YQAGFLGFLI GAIYGGVTQS RVAYMNFEMEN NQATAFKSHF DAKKKLQDQF |           |                                |
|    | TVNFAKGGFK WGWVGLFTT SYFGIITCMS VYRGKSSIYE YLAAGSITGS LYKVSLLGLRG  |           |                                |
|    | MAAGGIIGGF LGGVAGVTSL LLMKASGTSM EEVRYWQYKW RLDRDENIQQ AFKKLTEDEN  |           |                                |
|    | PELFKAHDEK TSEHVSLDTI K  |           |                                |

Το πρόγραμμα θα πρέπει να μετατρέπει τις αλληλουχίες στη μορφή:

```
>P81928
MNFLWKGRRF LIAGILPTFE GADEIVDKENKTYKAFLASKPPEETGLER LKQMFTIDEF
GSISSELNSV YQAGFLGFLI GAIYGGVTQS RVAYMNFEMEN NQATAFKSHF DAKKKLQDQF
-----MMMMMMMMMMMM
```

Στο `eclass` έχω ανεβάσει τον κώδικα που φτιάξαμε μαζί στο εργαστήριο καθώς επίσης και το αρχείο με την πρωτεΐνη που θα χρησιμοποιηθεί για την εκτέλεση του κώδικα.

Το πρόγραμμα που φτιάξαμε μαζί τυπώνει τις 2 από τις 3 γραμμές της τελικής μορφής ενώ έχει γίνει η προεργασία για τη δημιουργία της τρίτης.

Υπενθυμίζω ότι έχουν δημιουργηθεί στον κώδικα 2 πίνακες οι οποίοι αποθηκεύουν την αρχή και το τέλος των διαμεμβρανικών τμημάτων αντίστοιχα.

Πιο συγκεκριμένα ο πίνακας `tmstart` περιέχει τις τιμές 67, 131 και 183 ενώ ο πίνακας `tmend` περιέχει τις τιμές 87, 151 και 203 (σύμφωνα με το παράδειγμα της παραπάνω εικόνας).

Αυτό που πρέπει να γίνει είναι η εκτύπωση από το 1<sup>ο</sup> αμινοξύ (1<sup>ο</sup> γράμμα-θέση της πρωτεΐνης) να τυπωθεί “ – “ έως τη θέση 66, από την 67 έως και την 87 να τυπωθεί **M**, από την 88 έως και την 130 “ – “ από την 131 έως 151 **M** κοκ.

Τέλος, στο φάκελο `eclass` έχω ανεβάσει και ένα αρχείο με πολλές πρωτεΐνες έτσι ώστε η εκτέλεση του προγράμματος να πραγματοποιείται για πολλές πρωτεΐνες.

Όποιος τροποποιήσει τον κώδικα κατάλληλα έτσι ώστε να διαβάσει το αρχείο με τις πολλαπλές εγγραφές και να μετατρέπει τις αλληλουχίες στην παραπάνω μορφή θα έχει bonus στο βαθμό της εργασίας.