

**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗΝ ΒΙΟΙΑΤΡΙΚΗ**

**2η ΕΡΓΑΣΤΗΡΙΑΚΗ ΑΣΚΗΣΗ**

**ΓΙΑ ΤΟ ΕΡΓΑΣΤΗΡΙΟ ΤΟΥ ΜΑΘΗΜΑΤΟΣ:**

**ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΙΙ**

**Θέμα: Πρόβλεψη Διαμεμβρανικών Περιοχών Αμινοξέων με τη χρήση της τεχνικής του Κυλιόμενου Παραθύρου**

**ΕΚΠΟΝΗΘΗΚΕ ΑΠΟ ΤΟΝ ΦΟΙΤΗΤΗ:**

**ΚΩΝΣΤΑΝΤΙΝΙΔΗ ΠΑΥΛΟ**

**ΔΙΔΑΣΚΟΥΣΑ:**

**κ. ΚΟΝΤΟΥ ΠΑΝΑΓΙΩΤΑ**

Λαμία, 2021

**ΠΕΡΙΕΧΟΜΕΝΑ**

* 1. **Πρόλογος.........................................................................................................................σελ. 3**
  2. **Επεξήγηση Κώδικα ...................................................................................................... σελ. 3**
  3. **Παρατηρήσεις ............................................................................................................... σελ. 6**
  4. **Συμπεράσματα .............................................................................................................. σελ. 10**

**1.1 Πρόλογος:**

Γνωρίζουμε από τη θεωρία ότι το εσωτερικό της κυτταρικής μεμβράνης είναι μη πολικό και κατά συνέπεια τα επιφανειακά αμινοξικά κατάλοιπα των διαμεμβρανικών πρωτεϊνών έχουν κυρίως μη πολικές πλευρικές αλυσίδες. Εξαιτίας αυτής της ιδιότητας των αμινοξέων, είναι δυνατή η πρόβλεψη των τμημάτων της πρωτεΐνης που βρίσκονται στο εσωτερικό της μεμβράνης. Για τον λόγο αυτό, έχουν δημιουργηθεί κλίμακες υδροφοβικότητας που δίνουν τιμές στην ιδιότητα της υδροφοβικότητας των αμινοξικών πλευρικών αλυσίδων. Σε ορισμένα αμινοξέα η ιδιότητα της υδροφοβικότητας είναι εμφανής ενώ σε άλλα δεν είναι. Οι κλίμακες υδροφοβικότητας που έχουν κατασκευαστεί στηρίζονται σε μελέτες διαλυτότητας, κρυσταλλογραφικές μελέτες ή σε συνδυασμό και των δύο.

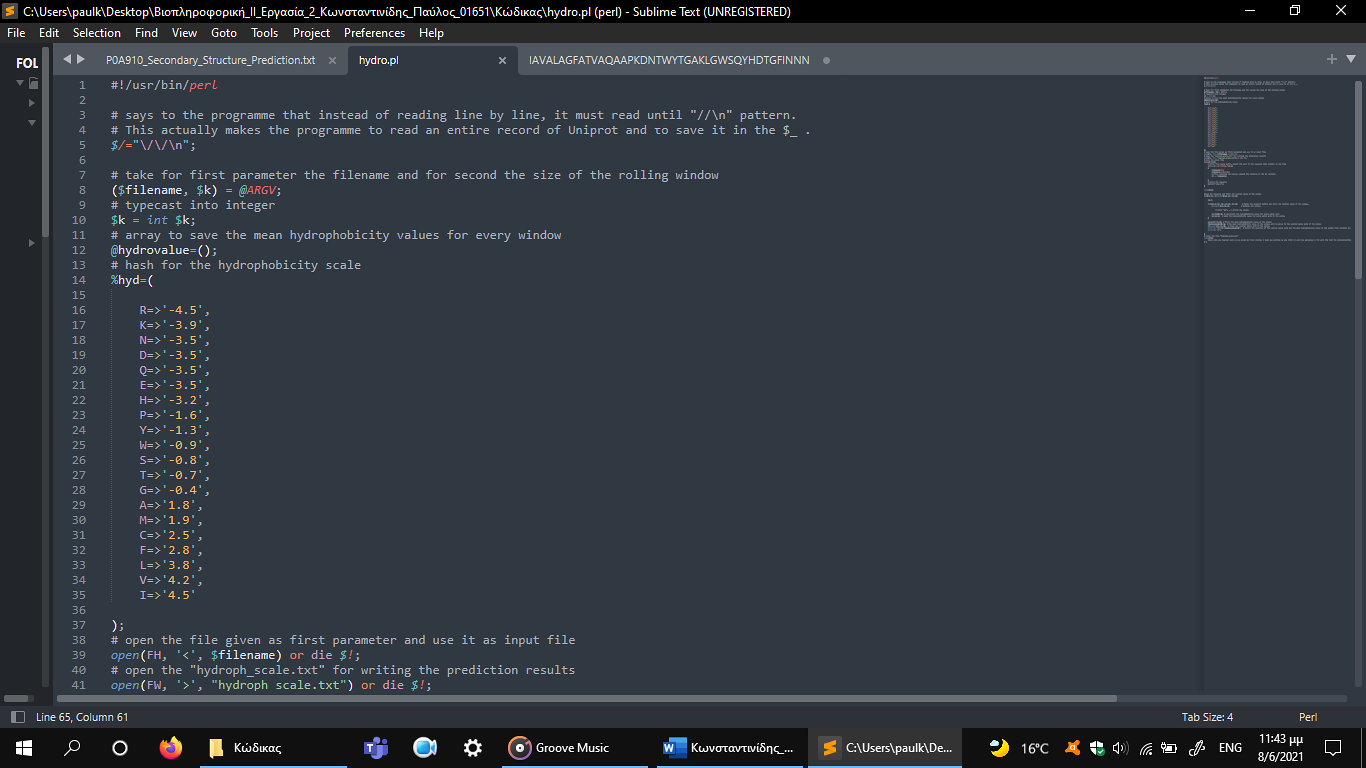
Η τεχνική του κυλιόμενου παραθύρου είναι ένα είδος τοπικής πρόβλεψης που βασίζεται στην στατιστική ομαλοποίηση και σύμφωνα με αυτή, οι ιδιότητες των αμινοξέων που βρίσκονται στο εσωτερικό του παραθύρου, καθορίζουν και την ιδιότητα του κεντρικού αμινοξέος.

Ο παρακάτω κώδικας, χρησιμοποιώντας την κλίμακα υδροφοβικότητας των Kyte-Doolittle, απομονώνει την αμινοξική ακολουθία από ένα αρχείο .swiss και στην συνέχεια, αναλόγως το μέγεθος παραθύρου που θα προσδιορίσει ο χρήστης, θα υπολογίσει τις νέες τιμές υδροφοβικότητας. Μια θετική τιμή υποδηλώνει ότι το αμινοξύ είναι πιθανώς διαμεμβρανικό κατάλοιπο, ενώ μια αρνητική τιμή ότι δεν είναι.

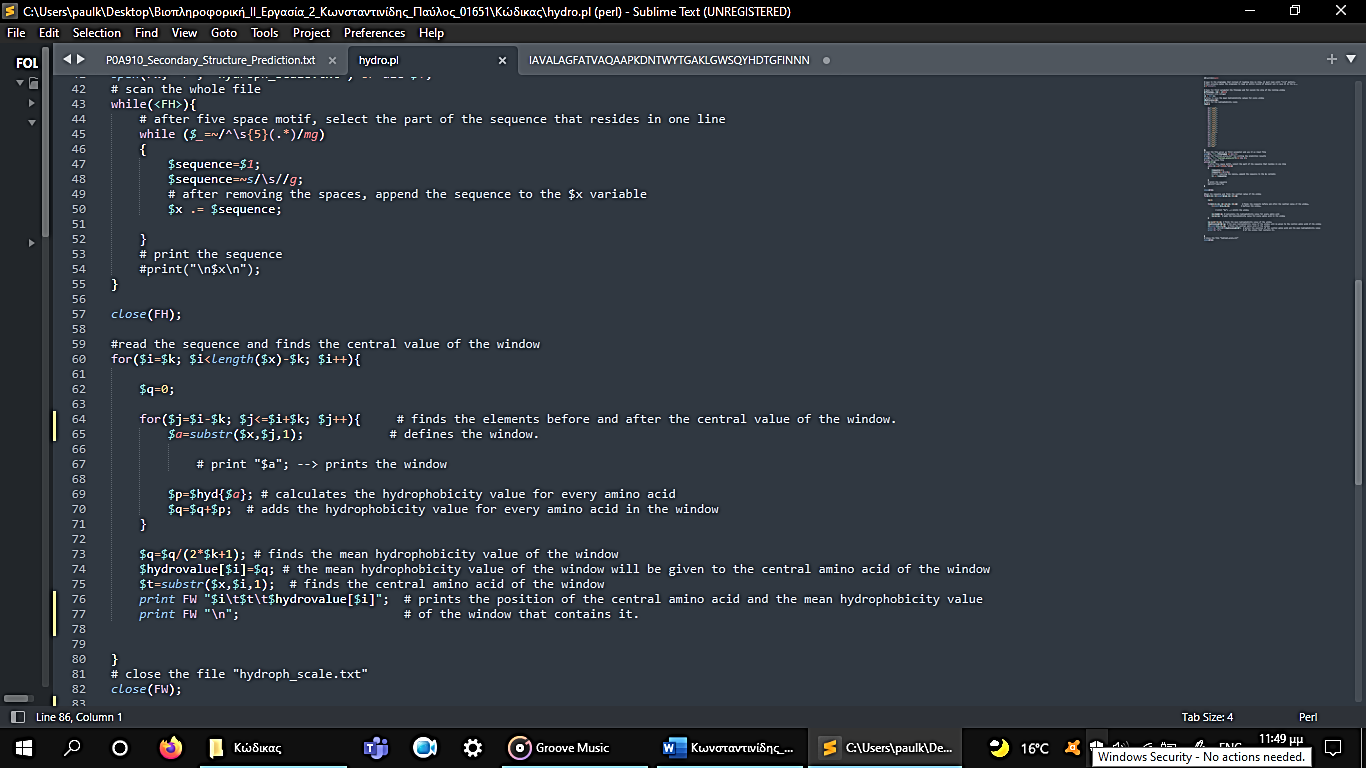
**1.2 Επεξήγηση Κώδικα:**

Αρχικά, ο κώδικας στην γραμμή 5 «λέει» στο πρόγραμμα να διαβάσει μια ολόκληρη εγγραφή από την UniProt και να την αποθηκεύσει στη μεταβλητή $\_. Έπειτα, (γραμμή 8), λαμβάνουμε από τη γραμμή εντολών τα δύο ορίσματα που συμβολίζουν το αρχείο .swiss (διαμεμβρανικής πρωτεΐνης) που θα χρησιμοποιήσουμε για την πρόβλεψη των διαμεμβρανικών τμημάτων και το μέγεθος του παραθύρου αντίστοιχα. Ακολουθούν οι αρχικοποιήσεις ενός πίνακα (γραμμή 12) που θα χρησιμοποιηθεί για την αποθήκευση των μέσων τιμών υδροφοβικότητας για κάθε παράθυρο και ένα λεξικό (γραμμή 14) που περιέχει την κλίμακα υδροφοβικότητας που θα χρησιμοποιηθεί στην πρόβλεψη. Αφού ενεργοποιήσουμε τους περιγραφείς των δύο αρχείων, δηλαδή του .swiss που χρησιμοποιείται σαν είσοδος και του «transmembrane\_proteins.txt» όπου και θα αποθηκευτούν τα αποτελέσματα, μέσω του επαναληπτικού βρόχου while, θα απομονώσουμε την ακολουθία της πρωτεΐνης (γραμμές 43-55). Στη συνέχεια, με δύο εμφωλευμένους επαναληπτικούς βρόχους, (γραμμές 60-79), εντοπίζουμε την κεντρική τιμή του παραθύρου και κατόπιν αφού βρούμε τα αμινοξέα του παραθύρου, αθροίζουμε τις τιμές της υδροφοβικότητας, υπολογίζουμε τη μέση τιμή της, και την ορίζουμε ως τιμή του κεντρικού αμινοξέως του παραθύρου (γραμμές 69,70,73,74). Τέλος, «κλείνουμε» το αρχείο όπου έχουν αποθηκευτεί οι νέες τιμές υδροφοβικότητας.

Παρακάτω, παρουσιάζεται ο κώδικας του προγράμματος με εξήγηση ανά γραμμή στα αγγλικά:



Εικόνα



Εικόνα 2

**1.3 Παρατηρήσεις**

## Το πρόγραμμα χρησιμοποιήθηκε για την πρόβλεψη των διαμεμβρανικών τμημάτων στην πρωτεΐνη με αριθμό προσπέλασης (AC) στην UniProt, το “P30516”. Η πρωτεΐνη είναι η αλυσίδα α’ ενός αντιγόνου ιστοσυμβατότητας και προέρχεται από τον ασπρόλοφο ταμαρίνο

(επιστημονική ονομασία: Saguinus Oedipus).

Τα αποτελέσματα της πρόβλεψης μετά από γραφική αναπαράστασή τους, για μεγέθη παραθύρων: 7, 11 και 19 φαίνονται παρακάτω :

Εικόνα 3

Εικόνα 4

Εικόνα 5

Εικόνα 6

Στο αρχείο «Bioinf\_ex\_2.xlsx», περιέχονται οι προβλέψεις του προγράμματος για τα διάφορα μεγέθη παραθύρου που χρησιμοποιήθηκαν, καθώς και το πραγματικό διαμεμβρανικό τμήμα της πρωτεΐνης (Συμβολισμοί: Μ -> Διαμεμβρανικό, ΝΜ -> Μη Διαμεμβρανικό, αρνητικές τιμές -> πιθανώς μη διαμεμβρανικό, θετικές τιμές -> διαμεμβρανικό).

* Για τις προβλέψεις του προγράμματος για τα διάφορα μεγέθη παραθύρων, θεωρούμε ότι συνεχόμενα αμινοξέα διαφορετικής τοπολογίας με μήκος μικρότερο του 6, που παρεμβάλλονται μεταξύ μεγαλύτερων ακολουθιών καταλοίπων αμινοξέων, βρίσκονται στην ίδια τοπολογία με τα τελευταία.
* Τα διαμεμβρανικά τμήματα είναι υδρόφοβα, άρα θα έχουν θετικές τιμές υδροφοβικότητας, ενώ το αντίστροφο ισχύει για τα μη διαμεμβρανικά (υδρόφιλα-αρνητικές τιμές).
* Η **ακρίβεια** μια πρόβλεψης δηλώνει κατά πόσο το αποτέλεσμα που προβλέφθηκε είναι σωστό (όσο πιο κοντά στο 1 τόσο καλύτερη είναι η πρόβλεψη).
* Ο **συντελεστής** **[Matthews](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)****[correlation](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)****[coefficient](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)**[, υποδηλώνει άριστη πρόβλεψη για τιμή ίση με 1, τυχαία πρόβλεψη για τιμή 0 και πολύ κακή πρόβλεψη για τιμή που προσσεγγίζει το -1.](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)

Οι τιμές της ακρίβειας και του συντελεστή Matthews για τα διαφορετικά μεγέθη παραθύρων παρουσιάζονται παρακάτω.

**Μέγεθος παραθύρου 7:**

TN=299, FP=42, TP=22, FN=2,

acc=0,879452055,

MCC=0.517197288

**Μέγεθος παραθύρου 11:**

TN=317, FP=24, TP=21, FN=3,

acc=0,926027397,

MCC=0,606585077

**Μέγεθος παραθύρου 19:**

TN=328, FP=13, TP=24, FN=0,

acc=0,964383562,

MCC=0.789886136

* 1. **Συμπεράσματα:**

Από τις παραπάνω στατιστικές τιμές, είναι ευκατανοητό πως το πρόγραμμα με μέγεθος παραθύρου ίσο με 19 προβλέπει καλύτερα τα διαμεμβρανικά αμιβοξέα, διότι και οι δύο τιμές προσεγγίζουν το 1.