

Βιοπληροφορική I

Εργαστηριακές Ασκήσεις

Άσκηση 1: Να γίνει αναζήτηση στη βάση δεδομένων Uniprot με σκοπό την ανεύρεση ανθρώπινων υποδοχέων συζευγμένων με G-πρωτεΐνες οι οποίοι να έχουν γνωστή (προσδιορισμένη) τρισδιάστατη δομή:

keyword:"G-protein coupled receptor [KW-0297]" AND organism:"Human [9606]" AND existence:"evidence at protein level" AND database:(type:pdb)

Ποιες από τις πρωτεΐνες έχουν δομή στη βάση δεδομένων PDB; Για τις παραπάνω πρωτεΐνες, να σημειωθεί ποιες αντιστοιχούν στην περιοχή της πρωτεΐνης στην οποία βρίσκονται το σύνολο των διαμεμβρανικών τμημάτων, ποιες αντιστοιχούν σε μέρος της περιοχής των διαμεμβρανικών τμημάτων και ποιες δεν περιλαμβάνουν κανένα τμήμα της ακολουθίας το οποίο να αντιστοιχεί σε ακολουθία διαμεμβρανικών τμημάτων.

Άσκηση 2: Να γίνει αναζήτηση στις βάσεις δεδομένων Uniprot, PDB, CATH, SCOP, Pfam για την πρωτεΐνη DHFR του ανθρώπου και του βακτηρίου *Lactobacillus casei* και να καταγραφούν δομικές διαφορές μεταξύ των δύο πρωτεϊνών.

Άσκηση 3: Δίνονται οι ακολουθίες: GAWGHEE και WHEAE. Να γίνει ολική στοίχιση και τοπική στοίχιση των δύο ακολουθιών με τη χρήση του πίνακα BLOSUM 50. Να γίνει επαλήθευση των αποτελεσμάτων με τη χρήση του αλγορίθμου στοίχισης που βρίσκεται στη διεύθυνση <http://www.itu.dk/~sestoft/bsa/dinaws/bsapplet.html>.

Άσκηση 4: Για την πρωτεΐνη OMP40_PORGI πραγματοποιήστε αναζήτηση ομόλογων ακολουθιών με τη χρήση του αλγορίθμου BLASTp.

A) Υπολογίστε το E-value της 30^{ης} ομόλογης πρωτεΐνης από το αποτέλεσμα του BLASTp.

B) Στη συνέχεια να γίνει κατά ζεύγη στοίχιση των πρωτεϊνών OMP40_PORGI και OMP41_PORGI και να βρεθεί αν υπάρχει διαφορά στο E-value της στοίχισης των δύο πρωτεϊνών με το BLASTp σε σχέση με αυτό της κατά ζεύγη στοίχισης.

Γ) Για τη 10^η έως και την 20^η ομόλογη πρωτεΐνη να γίνει πολλαπλή στοίχιση ακολουθιών με τον αλγόριθμο ClustalW.

Άσκηση 5: Να γίνει πρόβλεψη της δευτεροταγούς δομής της πρωτεΐνης bovine pancreatic trypsin inhibitor με τη μέθοδο των Dufton & Hider. <http://www.compgen.org/material/courses/bioinformatics1/uoa/secondary-structure>.

Άσκηση 6: Να γίνει πρόβλεψη της δευτεροταγούς δομής των πρωτεϊνών ompA Ecoli (P0A910) και NAD-dependent malic enzyme (P27443) με τη χρήση των προγραμμάτων **Secstr**, **Porter** και **Psipred**. Στη συνέχεια να υπολογιστεί ο δείκτης Q₃ για την πρόβλεψη της δευτεροταγούς δομής της πρωτεΐνης από τα προγράμματα **Porter** και **Psipred**.

<http://www.compgen.org/material/courses/bioinformatics1/uoa/membrane-proteins>

(Πρόγνωση Δευτεροταγούς Δομής)

Άσκηση 7: Να γίνει πρόβλεψη των διαμεμβρανικών τμημάτων της πρωτεΐνης Rhodopsin (P02699) με τη χρήση των προγραμμάτων TMHMM, HMMTOP, MEMSAT, TOPCONS και PHOBIUS. Στη συνέχεια, σε αρχείο excel να καταχωρήσετε μόνο τις ακολουθίες των διαμεμβρανικών τμημάτων και στις επόμενες γραμμές την αντίστοιχη πρόβλεψη του κάθε προγράμματος για κάθε αμινοξύ του διαμεμβρανικού τμήματος. Αν για κάθε αμινοξύ του διαμεμβρανικού τμήματος η πρόβλεψη συμφωνεί για τέσσερα προγράμματα πρόβλεψης και πάνω τότε να καταχωρείτε ότι το αμινοξύ αυτό ανήκει σε διαμεμβρανικό τμήμα (π.χ. συμβολίστε το με 1) ενώ αν λιγότερες από τέσσερις προβλέψεις συμφωνούν τότε να καταχωρείτε ότι δεν ανήκει σε διαμεμβρανικό τμήμα (π.χ. συμβολίστε το με 0). Τέλος, να απεικονιστεί γραφικά η τελική πρόβλεψη για τα διαμεμβρανικά τμήματα της πρωτεΐνης η οποία θα έχει προκύψει από το συνδυασμό των παραπάνω προγραμμάτων πρόβλεψης τοπολογίας διαμεμβρανικών πρωτεϊνών.