



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Δημιουργία Διαδικτυακού Εργαλείου για Ομαδοποίηση και Ανάλυση
Δεδομένων Πρωτεϊνικών Αλληλεπιδράσεων**

Νικόλαος Α. Σάκκος

Επιβλέπουσα : Κοσσίδα Σοφία

Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών -ΙΙΒΕΑΑ

Αθήνα

Ιανουάριος 2010

ΠΕΡΙΛΗΨΗ

Οι πρωτεΐνες δε λειτουργούν αυτόνομα μέσα στο κύτταρο, αλλά σχηματίζουν πρωτεϊνικά σύμπλοκα, τα οποία με τη σειρά τους λειτουργούν ως πολύπλοκες μοριακές μηχανές. Καθώς η λειτουργικότητα του κυττάρου βασίζεται στις επαφές και την αλληλεπίδραση των πρωτεϊνών, η μελέτη του συνόλου των αλληλεπιδράσεων των πρωτεϊνών ενός οργανισμού καθίσταται ιδιαίτερα σημαντική.

Η μελέτη αυτών των αλληλεπιδράσεων γίνεται με τη βοήθεια της θεωρίας γράφων. Στην αναπαράσταση των πρωτεϊνικών αλληλεπιδράσεων με τη μορφή γράφων, κάθε κόμβος αναπαριστά μια πρωτεΐνη, ενώ μια ακμή μεταξύ δυο κόμβων αναπαριστά την αλληλεπίδρασή τους.

Οι πρωτεϊνικοί γράφοι όλων των οργανισμών μοιράζονται κοινά χαρακτηριστικά, όπως την αντιστοιχία πυκνών περιοχών στους γράφους αλληλεπίδρασης πρωτεϊνών, με πρωτεϊνικά σύμπλοκα που επιτελούν μια συγκεκριμένη λειτουργία. Οι πυκνές περιοχές των γράφων μπορούν να εντοπιστούν με τη χρήση αλγορίθμων clustering.

Σκοπός αυτής της διπλωματικής εργασίας είναι η κατασκευή ενός εύχρηστου διαδικτυακού εργαλείου, που να επιτρέπει την επεξεργασία, ανάλυση και οπτικοποίηση δεδομένων που προκύπτουν από high-throughput πειράματα εντοπισμού πρωτεϊνικών αλληλεπιδράσεων (Yeast Two Hybrid System, φασματοσκοπία μάζας και Microarrays πρωτεϊνών), στα πλαίσια μιας συνεχούς ροής εργασίας. Ιδιαίτερο βάρος δίνεται στην αξιοποίηση αλγορίθμων ομαδοποίησης (clustering) δεδομένων για τον εντοπισμό πιθανών πρωτεϊνικών συμπλόκων.

Μέσω του εργαλείου, παρέχονται στο χρήστη λειτουργίες μετατροπής format αρχείων που περιέχουν δεδομένα πρωτεϊνικών αλληλεπιδράσεων (Adjacency Matrix, Sif, Tab-delimited, Dot), οπτικοποίησης πρωτεϊνικών γράφων, χρήσης αλγορίθμων clustering (MCL, RNSC, SIDES, HCS) για την επεξεργασία των πρωτεϊνικών

δεδομένων καθώς και πρόβλεψης λειτουργίας πρωτεϊνών και πρωτεϊνικών συμπλόκων αξιοποιώντας δεδομένα της βάσης δεδομένων λειτουργικής κατηγοριοποίησης πρωτεϊνών FUNCAT (FUNctional CATalogue) του MIPS (Munich Information Center for Protein Sequences) και των αποτελεσμάτων που προκύπτουν από τη χρήση των clustering αλγορίθμων.

Οι λειτουργίες παρέχονται τόσο ξεχωριστά, όσο και σαν μέρη μιας συνεχούς ροής εργασίας, που προσφέρει στο χρήστη τη δυνατότητα μιας ολοκληρωμένης ανάλυσης των παρεχόμενων δεδομένων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: βιοπληροφορική, πρωτεωμική, ομαδοποίηση πρωτεϊνών, διαδικτυακό εργαλείο, δίκτυα αλληλεπίδρασης πρωτεϊνών