



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

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

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Υλοποίηση εφαρμογής για την αυτοματοποίηση της
ανάλυσης γεγονότων συγχώνευσης πρωτεϊνικών δομικών
ενοτήτων**

Δημοσθένης Α. Τσαγκρασούλης

Επιβλέπουσα: Σοφία Κοσσίδα
Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών – ΙΙΒΕΑΑ

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2010

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Υλοποίηση εφαρμογής για την αυτοματοποίηση της ανάλυσης γεγονότων συγχώνευσης πρωτεϊνικών δομικών ενοτήτων

Δημοσθένης Α. Τσαγκρασούλης

A.M.: ΠΙΒ10

ΕΠΙΒΛΕΠΟΥΣΑ:

Σοφία Κοσσίδα

Ερευνήτρια Γ' (Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών – ΙΙΒΕΑΑ)

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Σοφία Κοσσίδα,

Ερευνήτρια Γ' (Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών – ΙΙΒΕΑΑ)

Δέσποινα Σανούδου,

Ερευνήτρια Γ' (Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών – ΙΙΒΕΑΑ)

Ηλίας Μανωλάκος, Αναπληρωτής Καθηγητής ΕΚΠΑ

Φεβρουάριος 2010

ΠΕΡΙΛΗΨΗ

Οι υπολογιστικές μέθοδοι που χρησιμοποιούνται παραδοσιακά για τη διερεύνηση των λειτουργιών και των αλληλεπιδράσεων των πρωτεϊνών στηρίζονται στη σύγκριση των αλληλουχιών των προς εξέταση πρωτεϊνών με άλλες γνωστών και χαρακτηρισμένων πολυπεπτιδίων. Η προσέγγιση αυτή παρουσιάζει κάποια εγγενή προβλήματα, όπως ότι μπορεί να εφαρμοστεί μόνο για πρωτεΐνες που έχουν διακριτές ομόλογες, οι οποίες είναι γνωστής λειτουργίας. Τα τελευταία χρόνια αναπτύσσονται μεθοδολογίες που ξεπερνάνε τα παραπάνω εμπόδια. Μία τέτοια μέθοδος είναι η ανάλυση γεγονότων συγχώνευσης πρωτεϊνικών δομικών ενοτήτων. Βασίζεται στο γεγονός ότι όταν δύο ξεχωριστές πρωτεΐνες ενός οργανισμού εντοπιστούν συγχωνευμένες σε μία πρωτεΐνη ενός άλλου, είναι πολύ πιθανό να αλληλεπιδρούν και να παρουσιάζουν συγγενείς λειτουργίες. Αντικείμενο της παρούσας εργασίας είναι η υλοποίηση πλατφόρμας για την πλήρη αυτοματοποίηση μεθοδολογίας που εντοπίζει γεγονότα συγχώνευσης πρωτεϊνικών δομικών ενοτήτων. Η πλατφόρμα υλοποιήθηκε με τη χρήση του προγραμματιστικού πακέτου Matlab R2009b. Για κάθε ανάλυση δημιουργούνται αρχεία κειμένου με τα αποτελέσματα της διαδικασίας και παρέχονται δυνατότητες για αξιολόγηση των γεγονότων συγχώνευσης με τη χρήση πινάκων και γραφικών αναπαραστάσεων. Τέλος, στα αποτελέσματα παρέχονται σύνδεσμοι προς κατάλληλη ιστοσελίδα πρωτεϊνικής βάσης δεδομένων (NCBI). Στην εργασία περιγράφεται επιπλέον ένα παράδειγμα ανάλυσης γεγονότων συγχώνευσης μεταξύ τριών οργανισμών (*Aquifex aeolicus*, *Thermofilum pendens*, *Entamoeba histolytica*), που πραγματοποιήθηκε με τη χρήση της πλατφόρμας.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: βιοπληροφορική, πρόβλεψη λειτουργίας πρωτεϊνών, πρόβλεψη πρωτεϊνικών αλληλεπιδράσεων, συγχώνευση πρωτεϊνών, συγχώνευση πρωτεϊνικών δομικών ενοτήτων.

ABSTRACT

Traditional computational methods for predicting protein function and protein-protein interactions are based on amino acid sequence comparisons between proteins of unknown function and others that are characterized. These methods have some innate disadvantages, e.g. they cannot produce any result if a protein does not have a homolog of known function. In the recent years, methodologies are being developed in order to overcome the aforementioned problems. One such method is the protein domain fusion analysis. It is based on the fact that if two separate proteins from one organism are fused in a protein from another organism, then they are likely to interact and have relevant functions. The main purpose of this thesis is to create an application for the automation of protein domain fusion analysis. The application was created in Matlab R2009b. Text files are created for each analysis' results. Comprehensive tables and graphical representations of the fused events are also provided for better evaluation of the results. Finally, the application creates links to the appropriate web pages from the NCBI protein database. The thesis also includes an example of domain fusion analysis between three organisms (*Aquifex aeolicus*, *Thermofilum pendens*, *Entamoeba histolytica*) that was performed by the developed software.

SUBJECT AREA: Bioinformatics

KEYWORDS: bioinformatics, protein function prediction, protein-protein interaction prediction, protein fusion, protein domain fusion.