



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

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

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Ανάλυση των πεπτιδίων μοναδικής αλληλουχίας αμινοξέων στο
ανθρώπινο πρωτέωμα.**

Ευάγγελος Μ. Κοντοπόδης

Επιβλέπων : **Γεώργιος Τσάγκαρης**, Ειδικός Λειτουργικός Επιστήμονας
(βαθμίδα Α')

ΑΘΗΝΑ

ΜΑΙΟΣ 2015

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανάλυση των πεπτιδίων μοναδικής αλληλουχίας αμινοξέων στο ανθρώπινο πρωτέωμα.

Ευάγγελος Μιχαήλ Κοντοπόδης

A.M.: ΠΙΒ0102

Επιβλέπων : **Γεώργιος Τσάγκαρης**, Ειδικός Λειτουργικός Επιστήμονας (βαθμίδα Α')

**ΕΞΕΤΑΣΤΙΚΗ
ΕΠΙΤΡΟΠΗ :**

Γεώργιος Τσάγκαρης, Ειδικός Λειτουργικός Επιστήμονας (βαθμίδα Α')

Γεώργιος Σπύρου, Ειδικός Λειτουργικός Επιστήμονας (βαθμίδα Α')

Ευανθία Αναστασιάδου, Ερευνήτρια Δ'

ΜΑΙΟΣ 2015

ΠΕΡΙΛΗΨΗ

Το ανθρώπινο πρωτέωμα αποτελείται από περισσότερες από 20.000 πρωτεΐνες. Αναλύσαμε το ανθρώπινο πρωτέωμα με σκοπό να εντοπίσουμε τις μικρότερες αμινοξικές ακολουθίες οι οποίες είναι μοναδικές για την κάθε πρωτεΐνη γνωστές ως core unique πεπτίδια (πεπτίδια ελάχιστης μοναδικής ακολουθίας). Στη συνέχεια ορίσαμε και τα composite unique πεπτίδια που προέρχονται από τα επικαλυπτόμενα core unique πεπτίδια.

Αναπτύξαμε αλγόριθμους με σκοπό να κατασκευάσουμε μια βάση δεδομένων που να περιλαμβάνει τα core και τα composite unique πεπτίδια. Στη συνέχεια αναλύσαμε το ανθρώπινο πρωτέωμα διερευνώντας χαρακτηριστικά αυτής της βάσης όπως : το μήκος των πεπτιδίων, τη σύσταση τους από αμινοξέα, τις θέσεις έναρξης των πεπτιδίων μέσα στις πρωτεΐνες. Επίσης ασχοληθήκαμε και με τις ιδιαιτερότητες των πρωτεϊνών που δεν αποτελούνται από τέτοια μοναδικά πεπτίδια.

Περισσότερα από 7×10^6 core unique πεπτίδια σχηματίζουν 6.8×10^4 composite unique πεπτίδια. Η πλειοψηφία των core unique πεπτιδίων (~ 72%) αποτελούνται από 6 αμινοξέα και το 20% από 7 αμινοξέα. Η πλειοψηφία των composite unique πεπτιδίων έχει μήκος ακολουθίας 11-12 αμινοξέα τα οποία σχηματίζονται από την ένωση 5-6 core unique πεπτιδίων τα περισσότερα από τα οποία (~ 30%) εμφανίζονται στην αρχή των πρωτεϊνών. Ειδικές ομάδες πρωτεϊνών έχουν αναλυθεί βάση αυτών των χαρακτηριστικών τα αποτελέσματα των οποίων παρουσιάζονται.

Τα αποτελέσματα αυτής της διπλωματικής είναι πολύ χρήσιμα για την ταυτοποίηση των πρωτεϊνών χρησιμοποιώντας την μέθοδο φασματομετρίας μαζών καθώς και για την μελέτη ιδιαίτερων χαρακτηριστικών σε διάφορες ομάδες πρωτεϊνών.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοτεχνολογία / Πρωτεωμική & Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ : Πρωτεωμική, φασματογραφία μάζας, πεπτίδια μοναδικής αλληλουχίας, πεπτίδια ελάχιστης μοναδικής ακολουθίας, σύνθετα πεπτίδια μοναδικής ακολουθίας.

ABSTRACT

The human proteome consists of more than 20,000 proteins. We analyzed the entire human proteome in order to identify the distribution of the shortest amino acid sequences appearing solely in one protein, defined as core unique peptides. Furthermore, we defined the composite unique peptides that consist of overlapping core unique peptides.

We developed an algorithmic approach to populate a database containing the core and the composite unique peptides. We analyzed the whole human proteome from the perspective of core unique peptides and composite unique peptides, investigating the distribution of peptide length, amino acid composition, starting position within the protein, density and coverage. We also analyzed the proteins that do not contain any unique peptides in order to understand their functionality.

More than 7×10^6 core unique peptides have been identified forming 6.8×10^4 composite unique peptides. The majority of the core unique peptides (~ 72%) consist of 6 amino acids and approximately 20% consists of 7 amino acids. The majority of composite unique peptides has a sequence length of 11-12 amino acids while they comprise by 5-6 core unique peptides while most of them (~ 30%) appear at the protein's first amino acids. Specific protein groups have been analyzed through these perspectives and results will be presented.

The results of the present study will be very useful for the identification of proteins by mass spectrometry and the application of selective reaction monitoring on complex protein mixtures.

SUBJECT AREA: Biotechnology / Proteomics & Bioinformatics

KEYWORDS: proteomics, mass spectrometry, unique peptide, core unique peptide, composite unique peptide.