



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCE  
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATION**

**POSTGRADUATE PROGRAM  
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"**

**MASTER THESIS**

**Current state-of-the-art of the research conducted in  
mapping protein cavities – binding sites of bioactive  
compounds, peptides or other proteins**

**Katerina Evangelos Dalamara**

**Supervisor:** **Evangelia D. Chrysiná**, Senior Researcher at NHRF  
Associate Professor, Örebro University Sweden

**ATHENS**

**SEPTEMBER 2017**



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

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

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Αποτύπωση των ερευνητικών μελετών που αφορούν στον  
χαρακτηρισμό μιας πρωτεϊνικής κοιλότητας – κέντρου  
πρόσδεσης βιοδραστικών ενώσεων, πεπτιδίων ή άλλων  
πρωτεϊνών**

**Κατερίνα Ευάγγελος Δαλαμάρα**

**Επιβλέπουσα:** **Ευαγγελία Δ. Χρυσίνα**, Κύρια Ερευνήτρια, ΕΙΕ  
Αναπληρώτρια Καθηγήτρια, Örebro University, Sweden

**ΑΘΗΝΑ**

**ΣΕΠΤΕΜΒΡΙΟΣ 2017**

## **MASTER THESIS**

Current state-of-the-art of the research conducted in mapping protein cavities - binding sites of bioactive compounds, peptides or other proteins

**Katerina E. Dalamara**

**S.N.:** PIV025

**SUPERVISOR:** **Evangelia D. Chrysina**, Senior Researcher at NHRF  
Associate Professor, Örebro University Sweden

**EXAMINATION  
COMMITTEE:**

**Elias S. Manolakos**, Professor, Dept. of Informatics, Univ. of Athens

**Ioannis Z. Emiris**, Professor, Dept. of Informatics, Univ. of Athens

**Evangelia D. Chrysina**, Senior Researcher NHRF, Assoc. Prof. Örebro University

September 2017

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Αποτύπωση των ερευνητικών μελετών που αφορούν στον χαρακτηρισμό μιας πρωτεϊνικής κοιλότητας - κέντρου πρόσδεσης βιοδραστικών ενώσεων, πεπτιδίων ή άλλων πρωτεϊνών.

**Κατερίνα Ε. Δαλαμάρα**

**A.M.: ΠΙΒ025**

**ΕΠΙΒΛΕΠΟΥΣΑ: Ευαγγελία Δ. Χρυσίνα, Αναπληρώτρια Καθηγήτρια**

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: Ηλίας Σ. Μανωλάκος, Καθηγητής, Τμ. Πληροφορικής & Τηλ/νιών, ΕΚΠΑ**  
**Ιωάννης Ζ. Εμίρης, Καθηγητής, Τμ. Πληροφορικής & Τηλ/νιών, ΕΚΠΑ**  
**Ευαγγελία Δ. Χρυσίνα, Κύρια Ερευνήτρια, ΕΙΕ, Αν. Καθ. Örebro University**

Σεπτέμβριος 2017

## **ABSTRACT**

The aim of this thesis was to report on the current state-of-the-art of the research conducted concerning mapping of protein cavities with a potential function role as binding sites of bioactive compounds, peptides or other proteins.

A literature review was performed with emphasis on the relevant tools developed during the last decade. In addition, the main research findings regarding drug design and druggable targets based on binding sites are presented.

Processes performed in protein cavity detection and analysis, of previous research articles, are compared with the approach described by Anaxagoras Fotopoulos and Athanasios Papathanasiou (2015). The results showed that a competitive advantage of their approach is the multidimensional k-means algorithm for clustering.

For the bibliographic review the scientific knowledgebase has been used, which includes international articles and journals, book chapters, as well as online articles regarding drug design and protein cavity.

Search keywords such as protein cavity dynamics, catalytic sites of enzymes, protein pocket etc. were used to identify bioinformatics tools with text mining. A catalogue of the most recently developed tools is presented followed by a brief description of selected tools. The selection criteria imposed for preparing the catalogue and the detailed description included the publication date, as well as the algorithms and the methods they use. The tools were then classified according to the search keywords.

The findings of this research are discussed, and the algorithms and methods they use are compared, highlighting the advantages of protein cavity detection.

**SUBJECT AREA:** Protein Structure

**KEYWORDS:** protein cavity dynamics, catalytic sites of enzymes, ligand binding, in silico docking, molecular dynamic simulations, active sites, protein pocket, protein conformer, rotamers

## ΠΕΡΙΛΗΨΗ

Ο σκοπός της διπλωματικής εργασίας είναι η διερεύνηση και αποτύπωση των ερευνητικών μελετών που αφορούν στον χαρακτηρισμό μιας πρωτεϊνικής κοιλότητας – κέντρου πρόσδεσης βιοδραστικών ενώσεων, πεπτιδίων ή άλλων πρωτεϊνών.

Στην παρούσα εργασία χρησιμοποιήθηκε η μέθοδος της βιβλιογραφικής επισκόπησης.

Παρουσιάζονται τα κυριότερα ευρήματα προηγούμενων ερευνών που σχετίζονται με τη διαδικασία σχεδιασμού φαρμάκων και τον εντοπισμό φαρμακοφόρων με βάση ένα σύνολο προσδετών.

Στη συνέχεια συγκρίνονται διαδικασίες επεξεργασίας και ανάλυσης της πρωτεϊνικής κοιλότητας προγενέστερων ερευνών με τη προσέγγιση που προτάθηκε από τους Παπαθανασίου και Φωτόπουλου το 2015. Αναδεικνύονται βασικά πλεονεκτήματα της προσέγγισης αυτής, όπως η εφαρμογή του αλγορίθμου πολυδιάστατη k-means ομαδοποίηση (multidimensional k-means clustering).

Η εύρεση βιβλιογραφίας βασίστηκε σε αναζήτηση επιστημονικών άρθρων σε ξενόγλωσσα επιστημονικά περιοδικά, σε κεφάλαια βιβλίων και σε διάφορα άρθρα σε ηλεκτρονικούς ιστότοπους σχετικά με τον σχεδιασμό φαρμάκων και τις κοιλότητες που απαντώνται στις πρωτεΐνες.

Στην παρούσα εργασία παρουσιάζονται εν συντομία εργαλεία που εντοπίστηκαν χρησιμοποιώντας λέξεις κλειδιά όπως για παράδειγμα δυναμική πρωτεϊνικής κοιλότητας, καταλυτικό κέντρο ενός ενζύμου, πρόσδεση, πρωτεϊνική θήκη κλπ. Στη συνέχεια συγκροτήθηκε κατάλογος με τα εργαλεία βιοπληροφορικής ανάλυσης που βρέθηκαν και ακολούθησε εκτενής αναφορά επιλεκτικά σε κάποια από αυτά. Κριτήριο επιλογής αυτών των εργαλείων αποτέλεσε η ημερομηνία δημοσίευσής τους, οι αλγόριθμοι και η μεθοδολογία που χρησιμοποιούν. Τα εργαλεία αυτά κατηγοριοποιήθηκαν με βάση της λέξεις κλειδιά που χρησιμοποιήθηκαν για την εξόρυξη των δεδομένων από την βιβλιογραφία. Τέλος πραγματοποιήθηκε συγκριτική μελέτη αυτών αναδεικνύοντας τα πλεονεκτήματα και εστιάζοντας στην περαιτέρω αξιοποίησή τους.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Πρωτεϊνική Δομή

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** δυναμική πρωτεϊνικής κοιλότητας, καταλυτικό κέντρο ενός ενζύμου, πρόσδεση προσδέτη, προσομοίωση πρόσδεσης, προσομοίωση μοριακής δυναμικής, κατηγοριοποίηση ενεργών περιοχών ενζυμικής κοιλότητας, πρωτεϊνική θήκη, πρωτεϊνικές διαμορφώσεις, διαμορφωμερή πλευρικής αλυσίδας