



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS**

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

MASTER THESIS

**Development of a pipeline for secondary and tertiary
structural analysis of human miRNA targeting**

Panagiotis I. Kakoulidis

Supervisor: **Artemis Hatzigeorgiou**, Professor of Bioinformatics, Department of Electrical & Computer Engineering, Telecommunications and Networks, University of Thessaly

ATHENS

JANUARY 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

**Ανάπτυξη λογισμικού για την δευτεροταγή και τριτοταγή
δομική ανάλυση της στόχευσης των μικρών ριβονουκλεϊκών
οξέων στον άνθρωπο**

Παναγιώτης Ι. Κακουλίδης

Επιβλέπουσα: **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής, Τμήμα
Μηχανικών Υ/Η, Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου
Θεσσαλίας

ΑΘΗΝΑ

ΙΑΝΟΥΑΡΙΟΣ 2017

MASTER THESIS

Development of a pipeline for secondary and tertiary structural analysis of human miRNA targeting

Panagiotis I. Kakoulidis

Student Registration Number: ΠΙΒ0142

SUPERVISOR: **Artemis Hatzigeorgiou**, Professor of Bioinformatics, Department of Electrical & Computer Engineering, Telecommunications and Networks, University of Thessaly

EXAMINING COMMITTEE: **Artemis Hatzigeorgiou**, Professor of Bioinformatics, Department of Electrical & Computer Engineering, Telecommunications and Networks, University of Thessaly
Ioannis Vlachos, Researcher in the Medical School of Harvard University
Maria Paraskevopoulou, Researcher/Scientific Associate

January 2017

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανάπτυξη λογισμικού για την δευτεροταγή και τριτοταγή δομική ανάλυση της στόχευσης των μικρών ριβονουκλεϊκών οξέων στον άνθρωπο

Παναγιώτης Ι. Κακουλίδης

A.M.: ΠΙΒ0142

ΕΠΙΒΛΕΠΟΥΣΑ: **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής, Τμήμα Μηχανικών Υ/Η, Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου Θεσσαλίας

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής, Τμήμα Μηχανικών Η/Υ, Τηλεπικοινωνιών και Δικτύων του Πανεπιστημίου Θεσσαλίας.
Ιωάννης Βλάχος, Ερευνητής Ιατρικής Σχολής του Πανεπιστημίου Χάρβαρντ
Μαρία Παρασκευοπούλου, Ερευνήτρια/Επιστημονικός Συνεργάτης

Ιανουάριος 2017

ABSTRACT

It has been observed that the majority of the current studies concerning miRNA-targeting interactions do not take into account the tertiary structures of RNA and Argonaute-2 (Ago2) protein. However, recently there have been several attempts considering this aspect justifying its significance. Target3D (pronounced "targeted") pipeline was developed following the advancements in the emerging field of the RNA three-dimensional structure prediction along with the growing availability of higher quality experimental crystal structures of RNAs and human Ago2-miRNA-mRNA complexes. It is an attempt that borrows techniques and data from Machine Learning, Molecular Dynamics, Quantum Mechanics, Molecular Modeling, RNA Sequencing and Structural Bioinformatics domains in order to analyze this inherently complex phenomenon. The pipeline yields information about secondary structural motifs that overlap with miRNA Recognition Element (MRE) on target mRNA and intermolecular and interatomic interactions between human Ago2 protein, guide miRNA and target mRNA which are part of RNA-induced Silencing Complex (RISC) that regulates the gene expression. Moreover, it calculates the structural similarity between fragments in bound and unbound states of target's structure that enclose the MRE region. Simulated and modified experimental structures, suitable for further Molecular Dynamics studying and visualization, are also generated. The modified structures are energy minimized with a customized force field.

SUBJECT AREA: Bioinformatics

KEYWORDS: miRNA, miRNA targeting, tertiary structure, secondary structure, RNA Sequencing, Molecular Dynamics, Argonaute 2, RISC, MRE, seed region, guide miRNA, target mRNA, gene expression regulation, RNA motifs, RNA tertiary motifs, structural similarity, molecular interface, energy minimization, force field parameterization, homology modeling, RBP-RNA interaction

ΠΕΡΙΛΗΨΗ

Έχει παρατηρηθεί ότι οι μελέτες έως σήμερα οι οποίες αφορούν την ανίχνευση στόχων των μικρών ριβονουκλεϊκών οξέων, που ρυθμίζουν την έκφραση των γονιδίων, δεν λαμβάνουν υπόψη τις τριτοταγείς δομές των ριβονουκλεϊκών οξέων και της πρωτεΐνης Αργοναύτης 2. Παρολαυτά, προσφάτως έχουν διενεργηθεί κάποιες προσπάθειες που συμπεριλαμβάνουν αυτό το σημείο τονίζοντας την σημαντικότητα του. Το Target3D λογισμικό αναπτύχθηκε υπό το πρίσμα των εξελίξεων στο αναδυόμενο πεδίο της πρόβλεψης των τρισδιάστατων δομών των ριβονουκλεϊκών οξέων σε συνάρτηση με την αυξημένη διαθεσιμότητα υψηλότερα ποιοτικά κρυσταλλικών δομών από ριβονουκλεϊκά οξέα και συμπλέγματα Αργοναύτη 2-μικρού ριβονουκλεϊκού οξέως-ριβονουκλεϊκού οξέως. Είναι ένα εγχείρημα που δανείζεται τεχνικές και δεδομένα από την Μηχανική Μάθηση, Μοριακή Δυναμική, Κβαντομηχανική, Μοριακή Μοντελοποίηση, Αλληλούχιση του ριβονουκλεϊκού οξέως και Δομική Βιοπληροφορική με σκοπό την ανάλυση αυτού του πολύπλευρου φαινομένου. Αυτή η ομοχειρία λογισμικού παράγει πληροφορία για τα μοτίβα δευτεροταγούς δομής που περιλαμβάνονται στην περιοχή του στοιχείου αναγνώρισης του μικρού ριβονουκλεϊκού οξέος, τις διαμοριακές και διατομικές αλληλεπιδράσεις μεταξύ της πρωτεΐνης Αργοναύτης 2, του οδηγού μικρού ριβονουκλεϊκού οξέως και του ριβονουκλεϊκού οξέως στόχου, τα οποία είναι μέρη ενός συμπλέγματος βιομορίων που ρυθμίζει την έκφραση των γονιδίων. Επιπροσθέτως, γίνεται υπολογισμός της δομικής ομοιότητας μεταξύ τμημάτων της δομής του στόχου σε ελεύθερη και δεσμευμένη κατάσταση που αντιστοιχούν σε στοιχεία αναγνώρισης του οδηγού μικρού ριβονουκλεϊκού οξέος. Επίσης, παράγονται ως αποτέλεσμα προσομοιωμένες και τροποποιημένες πειραματικές δομές για περαιτέρω ανάλυση Μοριακής Δυναμικής και οπτικοποίηση. Η ενέργεια των τροποποιημένων δομών έχει ελαχιστοποιηθεί με ειδικά προσαρμοσμένες παραμέτρους προσομοίωσης.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: miRNA, στόχοι miRNA, μικρό ριβονουκλεϊκό οξύ, δευτεροταγής δομή, τριτοταγής δομή, Αλληλούχιση, Μοριακή Δυναμική, Αργοναύτης 2, RISC, MRE, οδηγό miRNA, στόχος mRNA, ρύθμιση έκφρασης γονιδίων, μοτίβα RNA, τριτοταγή μοτίβα RNA, δομική ομοιότητα, μοριακή διασύνδεση, ελαχιστοποίηση ενέργειας δομής