



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΣΤΕΡΕΑΣ ΕΛΛΑΔΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**ΠΡΟΓΝΩΣΗ
ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΩΝ
ΤΡΟΠΟΠΟΙΗΣΕΩΝ ΠΡΩΤΕΪΝΩΝ**

Χρυσούλα Μαλέσκου

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνος
Παντελής Μπάγκος
Επίκουρος Καθηγητής**

Λαμία, 2009



ΠΑΝΕΠΙΣΤΗΜΙΟ ΣΤΕΡΕΑΣ ΕΛΛΑΔΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

ΠΡΟΓΝΩΣΗ
ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΩΝ
ΤΡΟΠΟΠΟΙΗΣΕΩΝ ΠΡΩΤΕΪΝΩΝ

Χρυσούλα Μαλέσκου

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Υπεύθυνος

Παντελής Μπάγκος

Επίκουρος Καθηγητής

Λαμία, 2009

*Στους αγαπημένους μου γονείς,
Χρήστο και Μαρία*

*Στο λατρεμένο μου παππού,
Γιώργο*

*Στη λατρεμένη μου γιαγιά,
Ελευθερία*

ΠΡΟΛΟΓΟΣ

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε εξ ολοκλήρου στο τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Στερεάς Ελλάδας κατά τη διάρκεια των ακαδημαϊκών ετών 2007-2009, υπό την επίβλεψη του Επίκουρου Καθηγητή κ. Παντελή Μπάγκου.

Στον πρόλογο αυτής της πτυχιακής θα ήθελα να εκφράσω την εκτίμηση και την ευγνωμοσύνη μου στα άτομα που με βοήθησαν ώστε να ολοκληρώσω την εργασία αυτή.

Αισθάνομαι πως οφείλω να πω ένα μεγάλο “ευχαριστώ” στον κύριο Μπάγκο, πρώτα από όλα γιατί με εμπιστεύτηκε και με επέλεξε να συνεργαστώ μαζί του. Επιπλέον, κατά τη διάρκεια της συνεργασίας μας, ήταν πάντα εκεί για μένα να μου εξηγεί και να επιλύει τις απορίες μου. Υπήρξε το άτομο που με συμβούλευε και με ενθάρρυνε ακόμη και όταν απογοητευόμουν. Με βοήθησε να πιστέψω στον εαυτό μου και συνέβαλε στο να αγαπήσω αυτό με το οποίο ασχολήθηκα και να δώσω τον καλύτερό μου εαυτό.

Ένα μεγάλο ευχαριστώ οφείλω αναμφίβολα και στα υπόλοιπα δύο μέλη της Τριμελούς Συμβουλευτικής Επιτροπής, το Λέκτορα κ. Μάρκου Ευριπίδη και τον Επίκουρο Καθηγητή κ. Πλαγιαννάκο Βασίλειο, οι οποίοι αφιέρωσαν τμήμα του πολύτιμου χρόνου τους στο να παρακολουθήσουν την παρουσίαση της διπλωματικής μου εργασίας.

Θα ήθελα κλείνοντας να ευχαριστήσω τους γονείς μου, τον παππού και τη γιαγιά μου για την ψυχολογική τους στήριξη, την οικονομική τους ενίσχυση και τις θυσίες που κάνουν τόσα χρόνια για μένα. Στερούνται αυτοί για να έχω εγώ, να σπουδάσω, να μορφωθώ και να έχω ένα λαμπρό μέλλον. Ότι είμαι το οφείλω σε αυτά τα τέσσερα σημαντικά άτομα της ζωής μου, τα οποία είναι πάντα δίπλα μου και στα οποία είναι αφιερωμένη αυτή η πτυχιακή εργασία.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ	2
ΠΕΡΙΕΧΟΜΕΝΑ	3
ΠΕΡΙΛΗΨΗ	5
ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ	6
ABSTRACT	7
KEY-WORDS	8
1.ΕΙΣΑΓΩΓΗ	9
1.1 ΠΡΩΤΕΪΝΕΣ.....	9
1.1.1 Πρωτεϊνοσύνθεση	12
1.1.1.1 Από το DNA στις πρωτεΐνες	12
1.1.1.2 Μεταγραφή του DNA σε RNA	13
1.1.1.3 Από το RNA στις πρωτεΐνες	15
1.2 Η ΔΟΜΗ ΤΩΝ ΠΡΩΤΕΪΝΩΝ.....	22
1.2.1 Η δομή μιας πρωτεΐνης καθορίζεται από την αλληλουχία των..... αμινοξέων της.....	23
1.2.2 Αλληλεπιδράσεις και υπάρχοντες δεσμοί στις πρωτεΐνες	24
1.2.3 Επίπεδα οργάνωσης των πρωτεϊνών	25
1.2.4 Οι πρωτεΐνες μπορούν να ταξινομηθούν σε οικογένειες.....	27
1.3 ΛΕΙΤΟΥΡΓΙΕΣ ΤΩΝ ΠΡΩΤΕΪΝΩΝ ΚΑΙ ΤΡΟΠΟΣ ΔΡΑΣΗΣ ΤΟΥΣ.....	28
1.3.1 Λειτουργίες των πρωτεϊνών	28
1.3.2 Τρόπος δράσης των πρωτεϊνών	30
1.3.2.1 Ενζυμική δράση.....	30
1.3.2.2 Αντισώματα.....	31
1.3.2.3 Μεταφορική δράση.....	32
1.3.2.4 Ανάδρομη τροφοδότηση	33
1.3.2.5 Αλλοστερική δράση.....	34
1.3.2.6 Κινητήριες πρωτεΐνες	34
2.ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΕΣ ΤΡΟΠΟΠΟΙΗΣΕΙΣ ΠΡΩΤΕΪΝΩΝ	36
2.1 ΕΙΔΗ ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΩΝ ΤΡΟΠΟΠΟΙΗΣΕΩΝ	37
2.1.1 Φωσφορυλίωση.....	38
2.1.1.1 Η φωσφορυλίωση στα φυτά.....	39
2.1.1.2 Γλυκοζυλίωση	40
2.1.1.2.1 Ο-Γλυκοζυλίωση	41
2.1.1.2.2 N-Γλυκοζυλίωση	42
2.1.1.2.3 Η γλυκοζυλίωση στα αρχαία.....	42
2.1.1.2.4 Η γλυκοζυλίωση στα βακτήρια	43
2.1.1.3 Μυριστοϋλίωση	44
2.1.1.4 Σουλφιδώση	45
2.1.1.5 Παλμιτοϋλίωση	47
2.1.1.6 Σουμοϋλίωση	48
2.1.1.7 Ακετυλίωση	50
2.1.1.8 Γλυκίωση	51
2.2 ΤΕΧΝΙΚΕΣ ΠΡΟΓΝΩΣΗΣ ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΩΝ	
ΤΡΟΠΟΠΟΙΗΣΕΩΝ	52
2.2.1 Τεχνητά Νευρωνικά Δίκτυα(ANNs).....	52
2.2.2 Έλεγχος προτύπου (Pattern Search-PS).....	54
2.2.3 Κρυφά Μαρκοβιανά Μοντέλα(HMMs).....	55

2.2.3.1 Θεωρία	55
2.2.3.2 Ορισμός του μοντέλου	55
2.2.3.3 Εφαρμογές των HMMs στη βιολογία	56
2.3 ΜΕΘΟΔΟΙ ΠΡΟΓΝΩΣΗΣ ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΩΝ	
ΤΡΟΠΟΠΟΙΗΣΕΩΝ	56
2.3.1 Πρόγνωση φωσφορυλίωσης	56
2.3.1.1 Πρόγνωση της φωσφορυλίωσης στους ζυμομήκυτες	62
2.3.2 Πρόγνωση γλυκοζυλίωσης	64
2.3.2.1 Πρόγνωση της Ο-γλυκοζυλίωσης στα θηλαστικά	64
2.3.2.2 Πρόγνωση της Ο-γλυκοζυλίωσης στο Dictyostelium	
discoideum	66
2.3.3 Πρόγνωση μυριστοϋλίωσης	67
2.3.3.1 Πρόγνωση μυριστοϋλίωσης στα φυτά	68
2.3.4 Πρόγνωση σουλφίδωσης	70
2.3.5 Πρόγνωση παλμιτοϋλίωσης	72
2.3.6 Πρόγνωση σουμοϋλίωσης	74
2.3.7 Πρόγνωση των N-τελικών θέσεων ακετυλίωσης	76
2.3.8 Πρόγνωση της πρωτεϊνικής γλυκίωσης στα θηλαστικά	79
2.3.9 Πρόγνωση των περιοχών διαχωρισμού στις προ-πρωτεΐνες	80
2.3.9.1 Λογότυπο ακολουθίας furin	81
2.3.9.2 Γενικό λογότυπο ακολουθίας PC	82
3.ΠΡΟΓΝΩΣΗ ΤΗΣ ΦΩΣΦΟΥΡΥΛΙΩΣΗΣ ΤΩΝ ΠΡΩΤΕΪΝΩΝ	84
3.1 ΕΙΣΑΓΩΓΗ	84
3.2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	84
3.2.1 Συλλογή δεδομένων ακολουθιών	84
3.2.2 Χρήση του προγράμματος NetPhosk για μεγάλο αριθμό	
πρωτεϊνών	86
3.2.3 Υλοποίηση προγράμματος σε περιβάλλον perl	88
3.2.4 Ταξινόμηση των οργανισμών από την PubMed	91
3.2.5 Εξαγωγή τελικών αποτελεσμάτων με χρήση του προγράμματος	
Stata	99
3.2.6 Παλινδρόμηση Poisson	99
3.3 ΕΦΑΡΜΟΓΗ ΤΗΣ POISSON REGRESSION	101
3.4 ΑΠΟΤΕΛΕΣΜΑΤΑ	101
4.ΣΥΖΗΤΗΣΗ-ΣΥΜΠΕΡΑΣΜΑΤΑ	137
5.ΒΙΒΛΙΟΓΡΑΦΙΑ	140
6.ΠΑΡΑΡΤΗΜΑ	152

ΠΕΡΙΛΗΨΗ

Σχεδόν όλες οι δραστηριότητες που συντελούνται μέσα σε ένα ζωντανό κύτταρο διαμορφώνονται από τις πρωτεΐνες, οι οποίες διαδραματίζουν σημαντικούς και ποικίλους ρόλους. Οι πρωτεΐνες αποτελούνται από 20 διαφορετικά αμινοξέα, το καθένα από τα οποία έχει διαφορετικές χημικές ιδιότητες. Αποτελούνται από τέσσερα επίπεδα οργάνωσης: την πρωτοταγή δομή που περιλαμβάνει την αλληλουχία των αμινοξέων μιας πρωτεΐνης, τη δευτεροταγή δομή που συνίσταται από τα τμήματα της πολυπεπτιδικής αλυσίδας που σχηματίζουν α -έλικες και β -πτυχωτές επιφάνειες, την τριτοταγή δομή στην οποία αναφέρεται η τρισδιάστατη διαμόρφωση μιας πολυπεπτιδικής αλυσίδας ενώ αν μία πρωτεΐνη σχηματίζεται ως σύμπλοκο δύο ή περισσότερων πολυπεπτιδίων, τότε η πλήρης δομή ορίζεται ως τεταρτοταγής δομή. Οι ρόλοι των πρωτεϊνών ποικίλουν, από τη μεταφορά μηνυμάτων από ένα κύτταρο σε ένα άλλο, την αναγνώριση και σύνδεση κυττάρων μεταξύ τους ή με άλλους σχηματισμούς, την επεξεργασία σήματος, τη μετάδοση σημάτων από την κυτταρική μεμβράνη στον πυρήνα ορισμένων κυττάρων έως και την εξειδικευμένη ενζυμική δραστηριότητα. Η σύνθεση των πρωτεϊνών είναι γνωστή ως πρωτεϊνοσύνθεση και επιτυγχάνεται ακολουθώντας το κεντρικό δόγμα της Μοριακής Βιολογίας.

Οι πρωτεΐνες μετά τη μετάφρασή τους υφίστανται τροποποιήσεις γνωστές ως μετά-μεταφραστικές τροποποιήσεις των πρωτεϊνών. Η μελέτη των τροποποιήσεων αυτών είναι σημαντική για την κατανόηση του μηχανισμού παθογένειας καθώς και για την τελική διαμόρφωση των πρωτεϊνών στο χώρο και για την λειτουργικότητα τους. Οι μετά-μεταφραστικές τροποποιήσεις γίνονται κατά κύριο λόγο στο ενδοπλασματικό δίκτυο και στη συσκευή Golgi ακολουθώντας ορισμένες πολύπλοκες διεργασίες και συσχετίζονται με αλλαγές στερεοδιάταξης, με την πρόσδεση στη μεμβράνη καθώς και με τον τελικό προορισμό της πρωτεΐνης μέσα στο κύτταρο. Είναι δυνατό να περιλαμβάνουν τροποποίηση του αμινοτελικού άκρου, τροποποίηση μεμονωμένων αμινοξέων, πρωτεολυτική επεξεργασία και σχηματισμό δισουλφιδικών γεφυρών. Για την πρόγνωση των μετά-μεταφραστικών τροποποιήσεων των πρωτεϊνών έχουν αναπτυχθεί διάφορες μέθοδοι στηριζόμενες σε εφαρμογές Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks-ANNs), ελέγχου προτύπου (pattern search-PS) και Κρυφών Μαρκοβιανών Μοντέλων (Hidden Markov Models-HMMs).

Μία από τις σημαντικότερες μετά-μεταφραστικές τροποποιήσεις είναι η φωσφορυλίωση, η μέθοδος που χρησιμοποιείται συχνά από τα ευκαρυωτικά κύτταρα

για τη ρύθμιση της λειτουργίας μιας πρωτεΐνης. Το ενδιαφέρον μας επικεντρώθηκε στη μετά-μεταφραστική τροποποίηση της φωσφορυλίωσης και προσπαθήσαμε να διερευνήσουμε αν η τροποποίηση αυτή εμφανίζεται σε μεγαλύτερη συχνότητα για κάποιους οργανισμούς των αρχαίων και των βακτηρίων, σε σύγκριση με τους ευκαρυώτες για τους οποίους γνωρίζουμε πως η φωσφορυλίωση αποτελεί βασική μετά-μεταφραστική τροποποίηση. Για τους σκοπούς αυτούς χρησιμοποιήσαμε την τοπική έκδοση του NetPhosK 1.0 κάνοντας προβλέψεις στο σύνολο των ακολουθιών των γονιδίων σε 234 οργανισμούς (28 για τα αρχαία, 171 για τα βακτήρια και 35 για τους ευκαρυώτες) οι οποίοι αναλύθηκαν σε 1.020.130 πρωτεΐνες. Αναπτύξαμε ένα πρόγραμμα το οποίο βασίζεται στο περιβάλλον προγραμματισμού perl για να αναλύσουμε τα αποτελέσματα και στη συνέχεια εφαρμόσαμε παλινδρόμηση Poisson (λογαριθμικό-γραμμικό πρότυπο). Αναλύσαμε ξεχωριστά κάθε μία από τις 18 κινάσες καθώς και τις τελικές μετρήσεις και βρήκαμε πολλές περιπτώσεις όπου ένας αυξημένος αριθμός (σε σύγκριση με τις πρωτεΐνες του ανθρώπου) από κινάσες στόχους βρέθηκε στα είδη των βακτηρίων και των αρχαίων. Είναι απαραίτητο να υπάρξει μελλοντική έρευνα ώστε να διευκρινιστούν περαιτέρω τα αποτελέσματα αυτά και να καθοριστεί ο ακριβής βιολογικός ρόλος της φωσφορυλίωσης σε αρχαία και βακτήρια.

ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ

Πρωτεΐνες, μετά-μεταφραστικές τροποποιήσεις πρωτεϊνών, πρόγνωση μετά-μεταφραστικών τροποποιήσεων πρωτεϊνών, φωσφορυλίωση, πρόγνωση φωσφορυλίωσης.

ABSTRACT

Almost all the biological functions that are necessary for the survival of the cell, are performed by proteins, which are involved in a variety of very important and different biological phenomena. Proteins are built from 20 different amino acids with different chemical properties. They are organized in four levels: the primary structure that includes the sequence of the amino acids of the protein, the secondary structure where the parts of the polypeptide chain make an α -helix and β -pleated leaves, the tertiary structure in which the 3-D structure of the polypeptide chain is referred to and the quaternary structure if the protein is the result of two or more polypeptides. Proteins transport messages between the cells and are involved in cellular recognition and adhesion, in signal process, in transporting substrates through membranes, till their specialized enzymatic activity. The making of proteins is known as protein synthesis and is succeeded by following the Central Dogma of Biology.

Proteins, after their translation exist modifications known as post-translational modifications of proteins. The study of these modifications is very important so as to understand the pathogenic mechanisms and the final three-dimensional structure of proteins. The post-translational modifications take part, in general, in the endoplasmic reticulum and the Golgi apparatus, following some complicated activities and they have to do with the integration into membranes and the final destination of the protein into the cell. It is also possible that they include modification of the amino acid part, modification of separated amino acids, proteolysis process and the construction of disulfide bridges. As far as it has to do with the prediction of post-translational modifications of proteins there have been many methods developed, using Artificial Neural Networks-ANNs, pattern search-PS and Hidden Markov Models-HMMs.

Phosphorylation, the method that is usually used by eukaryotic cells to regulate the function of a protein, is one of the most important post-translational modifications. We dealt extensively with phosphorylation and we investigated if this post-translational modification appears more often in some organisms of archaea and bacteria than in eukarya, where it is known that phosphorylation is a basic modification. In order to achieve this goal we used the local version of NetPhosK 1.0 for performing predictions in the complete genome sequences of 234 organisms (28 archaea, 171 bacteria and 35 eukarya) which code for a total of 1.020.130 proteins. We developed a program, which is based on perl to parse the results and subsequently

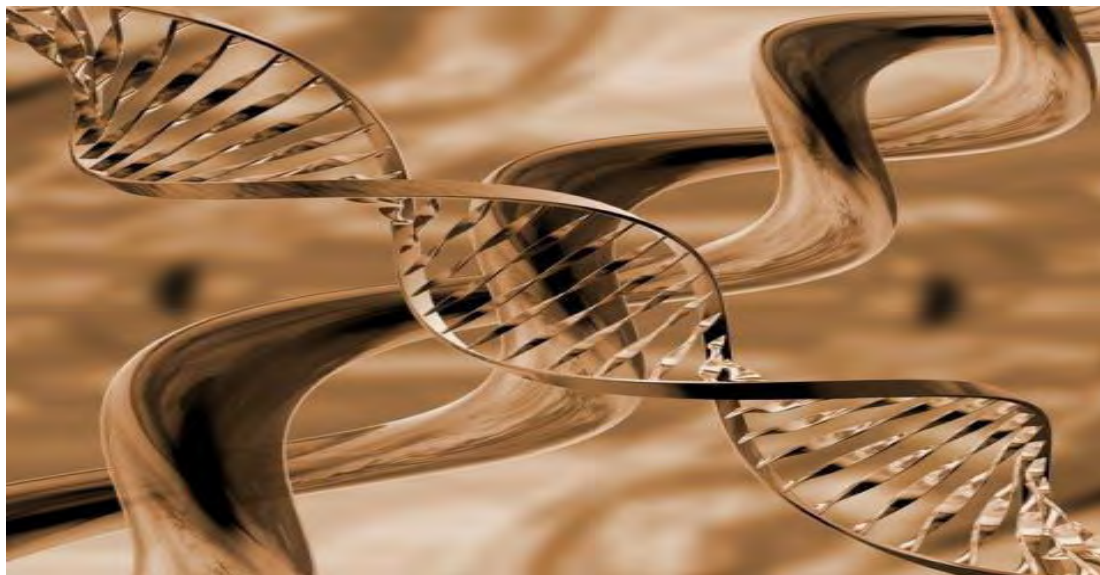
these were analyzed using Poisson regression (log-linear modeling). We analyzed separately each one of the 18 kinases as well as the total counts and we found numerous cases where an increased (compared to human proteins) number of kinase targets were found in bacterial and archaeal species. Future work needs to be done in order to elucidate these findings and unravel the exact biological role of phosphorylation in archaea and bacteria.

KEY WORDS

Proteins, post-translational modifications, protein prediction, phosphorylation, phosphorylation prediction.

1.ΕΙΣΑΓΩΓΗ

1.1 Πρωτεΐνες



Οι πρωτεΐνες είναι τα πιο πολυδύναμα μακρομόρια στους ζώντες οργανισμούς και αποτελούν το μεγαλύτερο μέρος της ξηρής μάζας ενός κυττάρου. Όταν εξετάζουμε ένα κύτταρο με το μικροσκόπιο ή όταν αναλύουμε την ηλεκτρική ή βιοχημική δραστηριότητά του, στην ουσία παρατηρούμε τις πρωτεΐνες οι οποίες εξυπηρετούν βασικές λειτουργίες σε όλες σχεδόν τις βιολογικές διεργασίες.

Για παράδειγμα, τα ένζυμα διαθέτουν τις περίπλοκες μοριακές επιφάνειές τους που διεκπεραιώνουν τις πολλές χημικές αντιδράσεις του κυττάρου. Ορισμένες πρωτεΐνες μεταφέρουν μηνύματα από ένα κύτταρο σε ένα άλλο, ενώ άλλες δρουν σαν επεξεργαστές σήματος και μεταδίδουν σήματα από την κυτταρική μεμβράνη στον πυρήνα ορισμένων κυττάρων. Οι πρωτεΐνες οι οποίες είναι ενσωματωμένες στην κυτταρική μεμβράνη σχηματίζουν διαύλους και αντλίες που ελέγχουν τη δίοδο μικρών μορίων μέσα και έξω από το κύτταρο. Επίσης, υπάρχουν πρωτεΐνες που λειτουργούν σαν μικροσκοπικές μοριακές μηχανές με κινητά τμήματα: για παράδειγμα, η κινεσίνη (kinesin) προωθεί τα οργανίδια μέσα στο κυτταρόπλασμα ενώ η τοποϊσομεράση (topoisomerase) μπορεί να ξεμπλέξει κόμπους στο DNA. Άλλες εξειδικευμένες πρωτεΐνες δρουν ως αντισώματα, τοξίνες, ορμόνες, αντιψυκτικά μόρια, ελαστικές ίνες ή πηγές φωταύγειας. Προτού επιχειρήσουμε να κατανοήσουμε πώς λειτουργούν τα γονίδια, πώς συστέλλονται οι μύες, πώς άγουν τα νεύρα

ηλεκτρισμό, πώς αναπτύσσονται τα έμβρυα ή πώς λειτουργεί το σώμα μας πρέπει να κατανοήσουμε τις πρωτεΐνες. Στον πίνακα 1-1 παρουσιάζονται μελέτες που αποτελούν ιστορικά ορόσημα για την κατανόηση των πρωτεϊνών.

Η δυνατότητα των πρωτεϊνών να επιτελούν τόσο πολλές και διαφορετικές λειτουργίες προκύπτει από τον τεράστιο αριθμό διαφορετικών τρισδιάστατων στερεοδομών που μπορεί να προσλάβουν εξηγώντας τη φράση “η λειτουργία ακολουθεί τη δομή”. Στο πρώτο μέρος αυτού του κεφαλαίου παρουσιάζονται η τρισδιάστατη δομή των πρωτεϊνών και των ιδιοτήτων που προσδίδει η δομή αυτή καθώς και τα επίπεδα οργάνωσης αυτών, ενώ στο δεύτερο μέρος του κεφαλαίου εξετάζουμε τον τρόπο με τον οποίο λειτουργούν οι πρωτεΐνες.

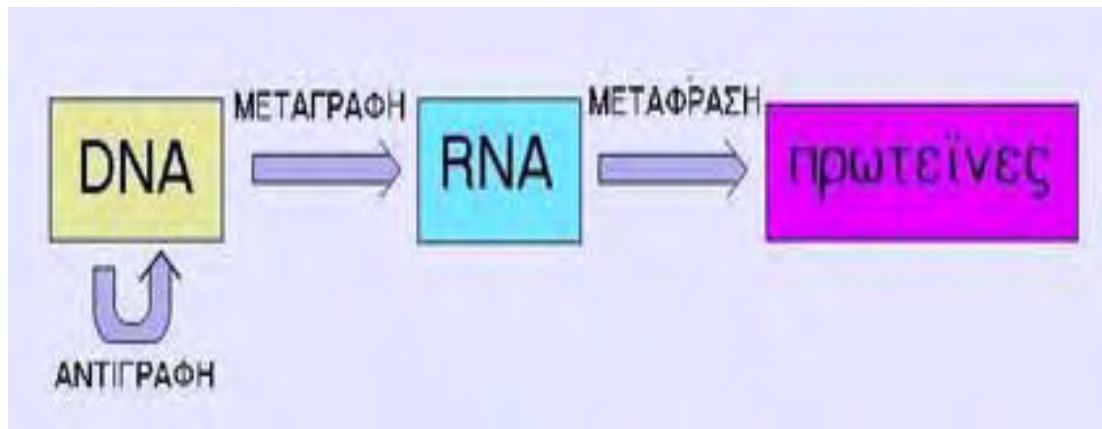
Πίνακας 1-1. Ιστορικά ορόσημα στις μελέτες για την κατανόηση των πρωτεϊνών	
1838	Ο Berzelius πρότεινε τον όρο “πρωτεΐνη” (από την Ελληνική λέξη πρώτος) για την πολύπλοκη, οργανική, αζωτούχο ουσία που απαντά στα κύτταρα όλων των ζώων και των φυτών.
1819-1904	Ανακαλύφθηκαν τα περισσότερα από τα 20 αμινοξέα που υπάρχουν στις πρωτεΐνες.
1864	Ο Hoppe-Seyler κρυστάλλωσε και ονόμασε την πρωτεΐνη αιμοσφαιρίνη.
1894	Ο Fisher πρότεινε ένα ανάλογο κλειδαριάς και κλειδιού για την αλληλεπίδραση ενζύμου-υποστρώματος.
1897	Οι Buchner και Buchner έδειξαν ότι τα εκχυλίσματα ζύμης ελεύθερων κυττάρων έχουν τη δυνατότητα να επιτύχουν τη ζύμωση της σουκρόζης και να σχηματίζουν διοξείδιο του άνθρακα και αιθανόλη, θέτοντας έτσι τα θεμέλια της ενζυμολογίας.
1926	Ο Sumner κρυστάλλωσε την ουρεάση σε καθαρή μορφή και έδειξε ότι οι πρωτεΐνες μπορεί να διαθέτουν την καταλυτική δράση των ενζύμων. Ο Svedberg ανέπτυξε την πρώτη αναλυτική υπερφυγόκεντρο και τη χρησιμοποίησε για να υπολογίσει το μοριακό βάρος της αιμοσφαιρίνης.

- 1933 Ο Tiselius εισήγαγε την τεχνική της ηλεκτροφόρησης για το διαχωρισμό πρωτεϊνών σε διάλυμα.
- 1934 Οι Bernal και Crowfoot παρουσίασαν τα πρώτα λεπτομερή διαγράμματα περίθλασης ακτίνων X μιας πρωτεΐνης που πήραν από κρυστάλλους του ενζύμου πεψίνη.
- 1942 Οι Martin και Synge ανέπτυξαν τη χρωματογραφία, μια τεχνική η οποία χρησιμοποιείται σήμερα για το διαχωρισμό πρωτεϊνών.
- 1951 Οι Pauling και Corey πρότειναν τη δομή της ελικοειδούς διαμόρφωσης μιας αλυσίδας L-αμινοξέων, την α-έλικα και τη δομή της β-πτυχωτής επιφάνειας. Αργότερα, τόσο η α-έλικα όσο και η β-πτυχωτή δομή βρέθηκαν σε πολλές πρωτεΐνες.
- 1955 Ο Sanger ολοκλήρωσε την ανάλυση της αλληλουχίας των αμινοξέων της Ινσουλίνης. Η ινσουλίνη ήταν η πρώτη πρωτεΐνη της οποίας η αλληλουχία των αμινοξέων καθορίστηκε πλήρως.
- 1956 Ο Ingram παρήγαγε τα πρώτα πρωτεϊνικά “αποτυπώματα” (protein fingerprints) και έδειξε ότι η διαφορά ανάμεσα στη δρεπανοκυτταρική και τη φυσιολογική αιμοσφαιρίνη οφείλεται σε μεταβολή ενός μόνο αμινοξέος.
- 1960 Ο Kendrew περιέγραψε την πρώτη λεπτομερή δομή μιας πρωτεΐνης, της μυοσφαιρίνης του σπέρματος της φάλαινας, με διακριτική ικανότητα 0.2 nm και ο Perutz περιέγραψε με μικρότερη διακριτική ικανότητα τη δομή της αιμοσφαιρίνης.

1.1.1 Πρωτεϊνοσύνθεση

1.1.1.1 Από το DNA στις Πρωτεΐνες

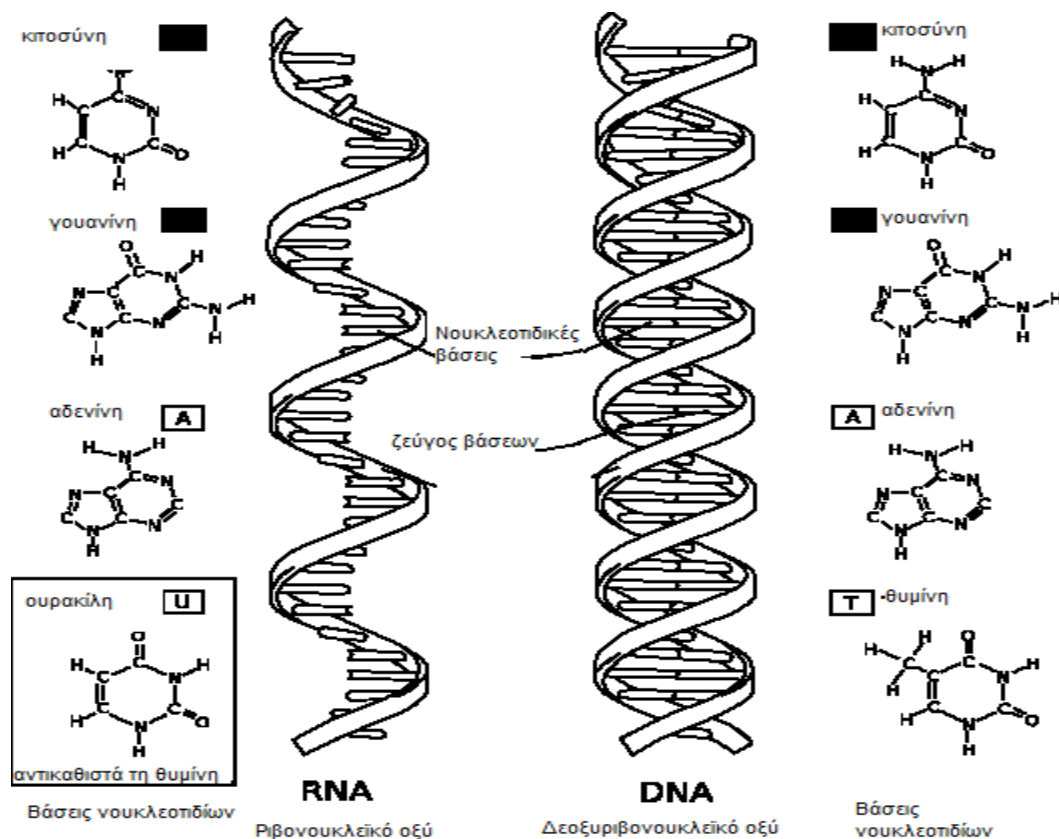
Πριν ακόμα αποκρυπτογραφηθεί ο γενετικός κώδικας, ήταν γνωστό ότι οι γενετικές πληροφορίες κατευθύνουν με κάποιο τρόπο τη σύνθεση των πρωτεϊνών. Καθώς κάθε πρωτεΐνη έχει τη δική της μοναδική αλληλουχία αμινοξέων, η οποία υπαγορεύει πώς θα διπλωθεί η αλυσίδα ώστε να σχηματισθεί ένα μόριο με χαρακτηριστική στερεοδομή και χημεία επομένως, οι γενετικές οδηγίες που μεταφέρονται από το DNA πρέπει με κάποιο τρόπο να καθορίζουν την αλληλουχία των αμινοξέων των πρωτεϊνών. Η σύνθεση των πρωτεϊνών είναι γνωστή ως πρωτεϊνοσύνθεση. Για να επιτευχθεί η πρωτεϊνοσύνθεση αρχικά τα κύτταρα αντιγράφουν το DNA σε RNA, μια διεργασία γνωστή ως μεταγραφή και κατόπιν χρησιμοποιούν τις πληροφορίες του RNA για την παραγωγή πρωτεϊνών, μια διεργασία γνωστή ως μετάφραση. Πρόκειται για μία αρχή η οποία είναι τόσο θεμελιώδης, ώστε αποκαλείται το κεντρικό δόγμα της Μοριακής Βιολογίας [1, 2, 99].



Εικόνα 1-1. Το κεντρικό δόγμα της Μοριακής Βιολογίας. Το DNA αντιγράφεται, μεταγράφεται σε RNA και στη συνέχεια μεταφράζεται σε πρωτεΐνες.

1.1.1.2 Μεταγραφή του DNA σε RNA

Το πρώτο βήμα στην έκφραση ενός τμήματος των γενετικών οδηγιών ενός κυττάρου είναι η διεργασία της μεταγραφής (transcription). Είναι δηλαδή, η αντιγραφή της κατάλληλης αλληλουχίας των νουκλεοτιδίων του DNA σε μια αλληλουχία νουκλεοτιδίων του RNA. Η μεταγραφή αρχίζει με τη διάνοιξη και το ξεδίπλωμα ενός μικρού τμήματος της διπλής έλικας του DNA, ώστε να αποκαλυφθούν οι βάσεις του κάθε κλώνου. Στη συνέχεια ο ένας από τους δύο κλώνους της διπλής έλικας του DNA δρα ως εκμαγείο για τη σύνθεση του RNA. Η αλληλουχία των ριβονουκλεοτιδίων της αλυσίδας του RNA καθορίζεται από το εκμαγείο-DNA με βάση τους κανόνες συμπληρωματικότητας των βάσεων. Όταν ένα εισερχόμενο ριβονουκλεοτίδιο ταιριάζει καλά με ένα δεοξυριβονουκλεοτίδιο του εκμαγείου-DNA, τότε θα συνδεθεί ομοιοπολικά με την αυξανόμενη αλυσίδα του RNA με μία ενζυμικά καταλυόμενη αντίδραση.



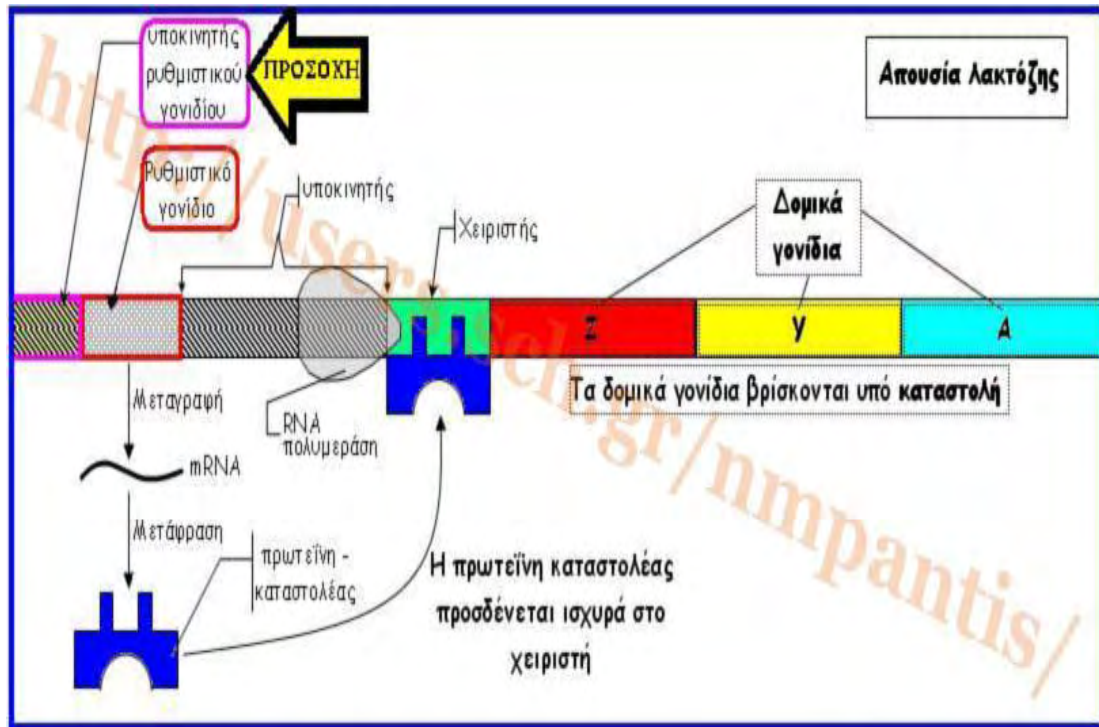
Εικόνα 1-2. Μεταγραφή του DNA σε RNA.

Επομένως, η αλυσίδα του RNA που παράγεται από τη μεταγραφή, γνωστή και ως μετάγραφο, επιμηκύνεται με ρυθμό ένα νουκλεοτίδιο τη φορά και έχει αλληλουχία νουκλεοτιδίων ακριβώς συμπληρωματική με την αλληλουχία του κλώνου του DNA ο οποίος χρησιμοποιήθηκε ως εκμαγείο. Η μεταγραφή μπορεί να γίνει μόνο προς την κατεύθυνση 5' → 3'. Τα ένζυμα που επιτελούν τη μεταγραφή αποκαλούνται RNA πολυμεράσες και στους ευκαρυωτικούς οργανισμούς είναι τρεις στο σύνολο. Ρόλος τους είναι η κατάλυση του σχηματισμού των φωσφοδιεστερικών δεσμών που συνδέουν τα νουκλεοτίδια μεταξύ τους και η δημιουργία του σακχαροφωσφορικού σκελετού της αλυσίδας.

Μετά τη μεταγραφή, στα κύτταρα παράγονται αρκετά είδη RNA. Αυτά τα είδη RNA χαρακτηρίζονται ως : **ριβοσωματικό RNA (rRNA)** το οποίο σχηματίζει το κέντρο των ριβοσωματίων, πάνω στο οποίο πραγματοποιείται η μετάφραση του mRNA σε πρωτεΐνη, το **μεταφορικό RNA (tRNA)** το οποίο σχηματίζει τους συναρμολογητές που επιλέγουν τα αμινοξέα και τα συγκρατούν στην κατάλληλη θέση πάνω στο ριβοσωμάτιο για να ενσωματωθούν σε μία πρωτεΐνη και τέλος το **αγγελιοφόρο RNA (mRNA)** που περιλαμβάνει τα μόρια του RNA που μεταγράφονται από τα γονίδια που κατευθύνουν τη σύνθεση των πρωτεϊνών.

Όσον αφορά τη διαδικασία της μεταγραφής στα βακτήρια [1, 2], όταν ένα μόριο RNA πολυμεράσης συναντηθεί τυχαία με το DNA, τείνει να προσκολληθεί ασθενώς πάνω του και στη συνέχεια, ολισθαίνει γρήγορα κατά μήκος του DNA έως ότου συναντήσει μία περιοχή γνωστή ως υποκινητής, οπότε και προσδέεται ισχυρά στο DNA. Ο υποκινητής περιέχει μία αλληλουχία νουκλεοτιδίων που σηματοδοτεί τη θέση έναρξης για τη σύνθεση του RNA. Η πολυμεράση αναγνωρίζει αυτή την αλληλουχία του DNA δημιουργώντας ειδικές επαφές με τα τμήματα των βάσεων που είναι εκτεθειμένα στο εξωτερικό της έλικας. Αφού έρθει σε επαφή με το DNA του υποκινητή και συνδεθεί ισχυρά, η RNA πολυμεράση διανοίγει την περιοχή της διπλής έλικας που βρίσκεται ακριβώς μπροστά της και αποκαλύπτει τα νουκλεοτίδια σε ένα βραχύ τμήμα του DNA κάθε κλώνου. Στη συνέχεια, ο ένας από τους δύο εκτεθειμένους κλώνους του DNA δρα ως εκμαγείο για να σχηματισθούν συμπληρωματικά ζεύγη βάσεων με τα εισερχόμενα ριβονουκλεοτίδια. Δύο από αυτά συνδέονται μεταξύ τους από την πολυμεράση και έτσι αρχίζει η αλυσίδα του RNA. Έπειτα η αλυσίδα επιμηκύνεται, έως ότου η RNA πολυμεράση συναντήσει ένα δεύτερο σήμα πάνω στο DNA, την αλληλουχία τερματισμού (stop site), οπότε

σταματά και απελευθερώνει τόσο το DNA εκμαγείο όσο και τη νεοσυνθετημένη αλυσίδα του RNA.



Εικόνα 1-3. Η διαδικασία της μεταγραφής στα βακτήρια.

1.1.1.3 Από το RNA στις Πρωτεΐνες

Όπως υποδηλώνει ο όρος μεταγραφή [1, 2], η όλη διεργασία αντιστοιχεί στη μετατροπή ενός χειρογράφου μηνύματος σε ένα δακτυλογραφημένο κείμενο. Η ίδια η γλώσσα και η μορφή του μηνύματος δεν αλλάζουν, ενώ τα σύμβολα που χρησιμοποιούνται είναι παρεμφερή.

Αντιθέτως, η μετατροπή των πληροφοριών του RNA σε πρωτεΐνη αντιπροσωπεύει μία μετάφραση των πληροφοριών σε μία άλλη γλώσσα που χρησιμοποιεί πολύ διαφορετικά σύμβολα. Γνωρίζουμε πως στις πρωτεΐνες υπάρχουν είκοσι διαφορετικά αμινοξέα και στο mRNA υπάρχουν μόνο τέσσερα διαφορετικά νουκλεοτίδια. Το γεγονός αυτό καθιστά αδύνατη την αντιστοιχία ένα προς ένα ανάμεσα σε ένα νουκλεοτίδιο του RNA και σε ένα αμινοξύ της αντίστοιχης πρωτεΐνης. Είναι λοιπόν απαραίτητη η ύπαρξη κάποιων κανόνων με τους οποίους η

αλληλουχία των νουκλεοτιδίων ενός γονιδίου, μέσω mRNA, μεταφράζεται στην αλληλουχία των αμινοξέων μιας πρωτεΐνης. Οι κανόνες αυτοί είναι γνωστοί ως **γενετικός κώδικας** (genetic code) και αποκρυπτογραφήθηκαν στις αρχές της δεκαετίας του 1960. Στον γενετικό κώδικα [1, 2] κάθε ομάδα τριών διαδοχικών νουκλεοτιδίων του RNA αποκαλείται **κωδικόνιο** (codon) και κάθε κωδικόνιο καθορίζει ένα αμινοξύ. Επειδή το RNA είναι ένα γραμμικό πολυμερές από τέσσερα διαφορετικά νουκλεοτίδια, υπάρχουν $4 \times 4 \times 4 = 64$ δυνατοί συνδυασμοί από τρία νουκλεοτίδια: AAA, AUA, AUG κ.ο.κ. Ωστόσο, στις πρωτεΐνες υπάρχουν συνήθως μόνο 20 διαφορετικά αμινοξέα. Επομένως, είτε ορισμένες τριπλέτες νουκλεοτιδίων δεν χρησιμοποιούνται ποτέ είτε ο κώδικας είναι πλεονάζων και ορισμένα αμινοξέα καθορίζονται από περισσότερες από μια τριπλέτες. Υπάρχουν 64 κωδικόνια από τα οποία τα 3 είναι τα κωδικόνια λήξης (UAA, UAG, UGA). Για τη σύνθεση των πρωτεϊνών είναι απαραίτητη η ύπαρξη μορίων του tRNA τα οποία συνταιριάζουν τα αμινοξέα στα κωδικόνια του mRNA.

Τα βραχέα τμήματα του διπλωμένου tRNA εμφανίζουν μορφή που αποδίδεται ως ένα τριφύλλι. Το τριφύλλι διπλώνεται ακόμα περισσότερο για να σχηματίσει μία συμπαγή δομή σχήματος L, η οποία συγκρατείται από επιπρόσθετους δεσμούς υδρογόνου ανάμεσα σε διάφορες περιοχές του μορίου. Υπάρχουν δύο περιοχές αζευγάρωτων νουκλεοτιδίων που εντοπίζονται στα δύο άκρα της δομής του σχήματος L και έχουν αποφασιστική σημασία για τη λειτουργία του tRNA στην πρωτεϊνοσύνθεση. Η μία δομή αποτελεί το αντικωδικόνιο (anticodon), μία ομάδα από τρία διαδοχικά νουκλεοτίδια του συμπληρωματικού κωδικονίου ενός μορίου mRNA. Η άλλη δομή είναι η θέση στην οποία προσδένεται το αμινοξύ που αντιστοιχεί στο συγκεκριμένο κωδικόνιο. Υπάρχουν ειδικά ένζυμα τα οποία συνδέουν τα μόρια tRNA με τα κατάλληλα αμινοξέα. Η αναγνώριση και σύνδεση με το κατάλληλο αμινοξύ βασίζεται στη λειτουργία ενζύμων γνωστών ως συνθετάσες των αμινοακυλο-tRNA (aminoacyl-tRNA synthetases), τα οποία συνδέουν ομοιοπολικά κάθε αμινοξύ με τα ιδιαίτερα και κατάλληλα μόρια tRNA.

Πρώτη βάση	Δεύτερη βάση				Τρίτη βάση
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Ser	Tyr	C
	Leu	Ser	Λήξη	Λήξη	A
	Leu	Ser	Λήξη	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Όπου: A=αδεΐνη, C=κυτοσίνη, G=γουανίνη, U=ουρακίλη

Εικόνα 1-4. Ο γενετικός κώδικας.

Για κάθε αμινοξύ υπάρχει μια διαφορετική συνθετάση (συνολικά, οι διάφορες συνθετάσες είναι 20): μια από αυτές συνδέει τη γλυκίνη σε όλα τα μόρια tRNA, που αναγνωρίζουν κωδικόνια για τη γλυκίνη, μια άλλη συνδέει την αλανίνη σε όλα τα μόρια tRNA που αναγνωρίζουν κωδικόνια για την αλανίνη κ.ο.κ. Ειδικά νουκλεοτίδια τα οποία εντοπίζονται τόσο στο αντικωδικόνιο όσο και στον βραχίονα-αποδέκτη τον αμινοξέος (amino acid accepting arm) βοηθούν κάθε συνθετάση να αναγνωρίζει το κατάλληλο tRNA. Οι συνθετάσες έχουν εξίσου μεγάλη σημασία με το tRNA για τη διεργασία της αποκωδικοποίησης καθώς χάρη στη συνδυασμένη δράση των συνθετασών και των μορίων tRNA, κάθε κωδικόνιο ενός μορίου mRNA σχετίζεται με ένα συγκεκριμένο, ξεχωριστό αμινοξύ.

Το μήνυμα του RNA αποκωδικοποιείται στα ριβοσωμάτια. Η αναγνώριση ενός κωδικονίου από το αντικωδικόνιο ενός μορίου tRNA βασίζεται στο ίδιο είδος ζευγαρώματος των συμπληρωματικών βάσεων που χρησιμοποιείται κατά την αντιγραφή και τη μεταγραφή του DNA. Η ακριβής και ταχεία μετάφραση του mRNA σε πρωτεΐνη προϋποθέτει τη λειτουργία μιας μεγάλης μοριακής μηχανής, η οποία μετακινείται κατά μήκος της αλυσίδας του mRNA, συλλαμβάνει συμπληρωματικά

μόρια tRNA, τα συγκρατεί σε κατάλληλες θέσεις και διασυνδέει τα αμινοξέα που μεταφέρουν ώστε να σχηματισθεί μια πρωτεϊνική αλυσίδα. Αυτή η μηχανή παρασκευής πρωτεϊνών καλείται **ριβосоμάτιο**. Το ριβосоμάτιο είναι ένα μεγάλο σύμπλοκο που αποτελείται από 50 και πλέον διαφορετικές πρωτεΐνες (γνωστές ως ριβосоματικές πρωτεΐνες (ribosomal proteins) και αρκετά μόρια RNA τα οποία αποκαλούνται ριβосоματικά RNA (ribosomal RNA, rRNA). Ένα τυπικό ζωντανό κύτταρο περιέχει στο κυτταρόπλασμα του εκατομμύρια ριβосоμάτια (Εικόνα 1-5).



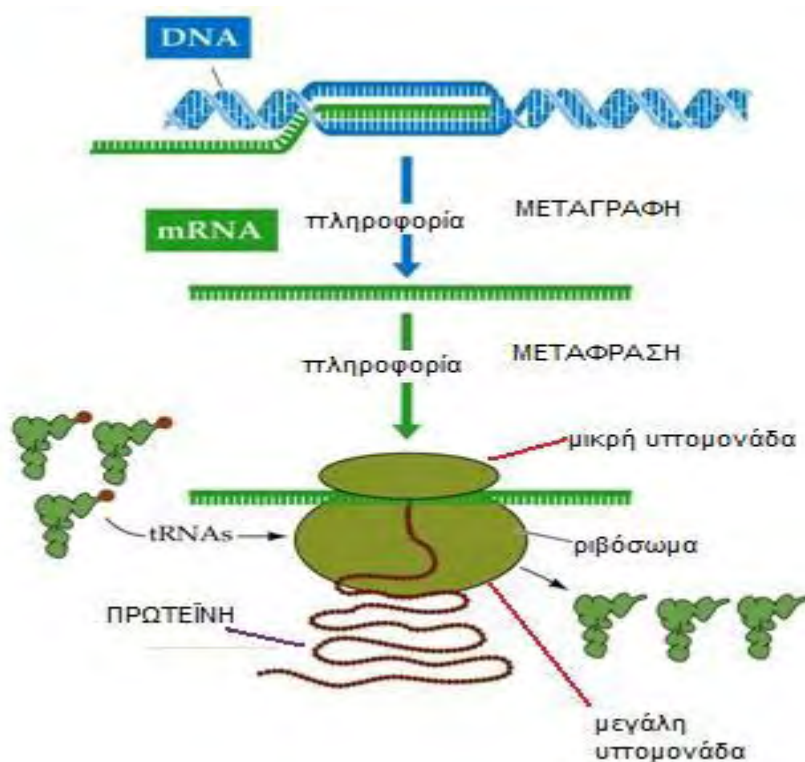
Εικόνα 1-5. Ριβосоμάτια στο κυτταρόπλασμα ενός ευκαρυωτικού κυττάρου. Στην ηλεκτρονιογραφία αυτή φαίνεται μία λεπτή τομή από μια μικρή περιοχή του κυτταροπλάσματος. Τα ριβосоμάτια φαίνονται σαν μαύρα στίγματα (κόκκινα βέλη).

Σ' έναν ευκαρυώτη, οι υπομονάδες των ριβосоματίων παρασκευάζονται στον πυρήνα. Εκεί μόρια νεοσυντεθειμένου rRNA συνδυάζονται με ριβосоματικές πρωτεΐνες, οι οποίες έχουν εισαχθεί στον πυρήνα από το κυτταρόπλασμα, όπου είχαν συντεθεί. Κατόπιν, οι ξεχωριστές υπομονάδες εξάγονται στο κυτταρόπλασμα για να συμμετάσχουν στην πρωτεϊνοσύνθεση. Τόσο τα ευκαρυωτικά όσο και τα προκαρυωτικά ριβосоμάτια αποτελούνται από μία μεγάλη και μία μικρή υπομονάδα, οι οποίες προσαρμόζονται κατάλληλα η μία στην άλλη για να σχηματίσουν ένα πλήρες ριβосоμάτιο με μάζα αρκετών εκατομμυρίων daltons. Η μικρή υπομονάδα συνταιριάζει τα μόρια tRNA στα κωδικόνια του mRNA, ενώ η μεγάλη καταλύει τον σχηματισμό των πεπτιδικών δεσμών οι οποίοι συνδέουν τα αμινοξέα σε μια πολυπεπτιδική αλυσίδα. Οι δύο υπομονάδες φθάνουν μαζί πάνω σ' ένα μόριο mRNA,

συνήθως κοντά στην αρχή του (5' άκρο), για να αρχίσουν τη σύνθεση μιας πρωτεΐνης. Στη συνέχεια, το ριβοσωμάτιο μετακινείται κατά μήκος του mRNA και μεταφράζει την αλληλουχία των νουκλεοτιδίων σε μια αλληλουχία αμινοξέων με ρυθμό ένα κωδικόνιο τη φορά. Το ριβοσωμάτιο χρησιμοποιεί τα μόρια tRNA για την προσθήκη του κάθε αμινοξέος στο άκρο της αυξανόμενης πολυπεπτιδικής αλυσίδας με τη σωστή σειρά. Τελικά, μόλις ολοκληρωθεί η σύνθεση της πρωτεΐνης οι δύο υπομονάδες του ριβοσωματίου δίστανται. Τα ριβοσωμάτια λειτουργούν με αξιοσημείωτη αποτελεσματικότητα: μέσα σ' ένα δευτερόλεπτο, ένα ριβοσωμάτιο ενός ευκαρυωτικού κυττάρου προσθέτει σε μια πολυπεπτιδική αλυσίδα περίπου δύο αμινοξέα. Τα ριβοσωμάτια των βακτηριακών κυττάρων λειτουργούν ακόμα πιο γρήγορα, με ταχύτητα περίπου 20 αμινοξέων ανά δευτερόλεπτο.

Ένα ριβοσωμάτιο περιέχει τέσσερις θέσεις σύνδεσης για μόρια RNA: η μια προορίζεται για το mRNA ενώ οι άλλες τρεις (γνωστές ως θέση A, θέση P και θέση E) για τα tRNA. Ένα μόριο tRNA θα συγκρατείται σφιχτά στις θέσεις A και P μόνο υπό την προϋπόθεση ότι το αντικωδικόνιό του θα ζευγαρώνει (έστω και με ταλάντευση) με ένα συμπληρωματικό κωδικόνιο του μορίου mRNA, το οποίο είναι συνδεδεμένο στο ριβοσωμάτιο. Οι θέσεις A και P βρίσκονται αρκετά κοντά η μια στην άλλη με συνέπεια τα μόρια tRNA που προσδέονται σε αυτές να εξαναγκάζονται να σχηματίσουν ζεύγη βάσεων με γειτονικά κωδικόνια του μορίου mRNA. Αφότου αρχίσει η πρωτεϊνοσύνθεση, κάθε νέο αμινοξύ προστίθεται στην αυξανόμενη αλυσίδα με έναν κύκλο αντιδράσεων. Ένα μόριο tRNA που μεταφέρει το επόμενο αμινοξύ στην αλυσίδα έχει προσδεθεί στην κενή θέση A, σχηματίζοντας ζεύγη βάσεων με το κωδικόνιο του mRNA που βρίσκεται εκτεθειμένο στη θέση A. Στο βήμα 2, το καρβοξυτελικό άκρο της πολυπεπτιδικής αλυσίδας απελευθερώνεται από το tRNA που βρίσκεται στη θέση P (με διάσπαση του δεσμού υψηλής ενέργειας ανάμεσα στο tRNA και το αμινοξύ του) και ενώνεται με πεπτιδικό δεσμό με την ελεύθερη αμινομάδα του αμινοξέος που μεταφέρεται από το tRNA το οποίο βρίσκεται στη θέση A. Αυτή η κεντρική αντίδραση της πρωτεϊνοσύνθεσης καταλύεται από το ένζυμο πεπτιδυλοτρανσφεράση (peptidyl transferase), η οποία αποτελεί τμήμα του ριβοσωματίου. Στην περίπτωση αυτή, το καταλυτικό τμήμα του ριβοσωματίου δεν είναι μια από τις πρωτεΐνες του αλλά ένα από τα μόρια του rRNA της μεγάλης υπομονάδας. Στο βήμα 3, η μικρή υπομονάδα μετακινείται ακριβώς 3 νουκλεοτίδια κατά μήκος του μορίου mRNA και επανέρχεται στην αρχική της θέση σε σχέση με τη μεγάλη υπομονάδα, ενώ το tRNA που καταλάμβανε τη θέση E δίσταται. Ο πλήρης

κύκλος των τριών βημάτων επαναλαμβάνεται κάθε φορά που ένα αμινοξύ προστίθεται στην πολυπεπτιδική αλυσίδα, η οποία αυξάνει από το αμινοτελικό προς το καρβοξυτελικό της άκρο έως ότου το ριβοσωμάτιο συναντήσει ένα κωδικόνιο τερματισμού.



Εικόνα 1-6. Πρωτεϊνοσύνθεση. Κάθε ριβοσωμάτιο (Ribosome) αποτελείται από μία μεγάλη και μία μικρή υπομονάδα (large και small subunit αντίστοιχα) που συνδέεται στο mRNA και η μετάφραση γίνεται με κατεύθυνση 5' → 3'.

Ειδικά κωδικόνια του mRNA σηματοδοτούν τη θέση έναρξης και το σημείο λήξης της πρωτεϊνοσύνθεσης. Η θέση από όπου αρχίζει η πρωτεϊνοσύνθεση πάνω στο mRNA έχει μεγάλη σημασία επειδή καθορίζει το πλαίσιο ανάγνωσης για ολόκληρο το μήνυμα. Στο στάδιο αυτό, ένα σφάλμα έστω και κατά ένα νουκλεοτίδιο θα προκαλέσει λανθασμένη ανάγνωση όλων των επόμενων κωδικονίων του μηνύματος, οδηγώντας σε παραγωγή μιας ελαττωματικής πρωτεΐνης με λανθασμένη αλληλουχία

αμινοξέων. Το εναρκτήριο βήμα έχει επίσης μεγάλη σημασία και από μια άλλη άποψη, επειδή είναι το τελευταίο σημείο στο οποίο το κύτταρο μπορεί να αποφασίσει αν θα μεταφρασθεί ένα mRNA για να παραχθεί η αντίστοιχη πρωτεΐνη. Συνεπώς, ο ρυθμός έναρξης της μετάφρασης καθορίζει την ταχύτητα σύνθεσης μιας πρωτεΐνης. Η μετάφραση ενός mRNA μεταγράφου αρχίζει με το κωδικόνιο AUG. Για την έναρξη της μετάφρασης είναι απαραίτητο ένα ειδικό tRNA, γνωστό ως εναρκτήριο tRNA (initiator tRNA), το οποίο πάντοτε μεταφέρει το αμινοξύ μεθειονίνη (στα βακτήρια χρησιμοποιείται φορμυλομεθειονίνη, μια τροποποιημένη μορφή της μεθειονίνης). Επομένως, όλες οι νεοσυντεθειμένες πρωτεΐνες θα έχουν μεθειονίνη ως πρώτο αμινοξύ στο αμινοτελικό άκρο τους, το άκρο μιας πρωτεΐνης που συντίθεται πρώτο. Αργότερα, αυτή η μεθειονίνη συνήθως αφαιρείται από μια ειδική πρωτεάση. Το εναρκτήριο tRNA διαφέρει από το tRNA το οποίο συνήθως μεταφέρει τη μεθειονίνη. Από όλα τα "φορτωμένα" μόρια tRNA (charged tRNAs) του κυττάρου, μόνο το εναρκτήριο tRNA διαθέτει την ικανότητα να προσδένεται ισχυρά στη μικρή ριβοσωματική υπομονάδα. Στη συνέχεια, η φορτωμένη υπομονάδα συνδέεται με το 5' άκρο ενός μορίου mRNA, το οποίο εν μέρει αναγνωρίζεται από την καλύπτρα που υπάρχει στο ευκαρυωτικό mRNA. Κατόπιν, η μικρή ριβοσωματική υπομονάδα μετακινείται προς τα εμπρός (5' → 3') κατά μήκος του mRNA σε αναζήτηση του πρώτου AUG. Μόλις συναντήσει αυτό το κωδικόνιο, αρκετοί παράγοντες έναρξης δίστανται από τη μικρή υπομονάδα, αφήνοντας χώρο για να συναρμολογηθεί η μεγάλη υπομονάδα και έτσι να ολοκληρωθεί το ριβοσωμάτιο. Επειδή το εναρκτήριο tRNA είναι συνδεδεμένο με τη θέση P, η πρωτεϊνοσύνθεση είναι έτοιμη να αρχίσει με την προσθήκη του επόμενου "φορτισμένου" με το αμινοξύ του μορίου tRNA.

Διαφορετικός είναι ο μηχανισμός επιλογής ενός κωδικονίου έναρξης στα βακτήρια. Τα βακτηριακά mRNA μετάγραφα δεν διαθέτουν στο 5' άκρο τους καλύπτρα που να υποδεικνύει στο ριβοσωμάτιο από πού πρέπει να αρχίσει να αναζητά την αφετηρία της μετάφρασης. Αντίθετα, περιέχουν ειδικές αλληλουχίες σύνδεσης των ριβοσωματίων (ribosome binding sequences), μήκους έως και έξι νουκλεοτιδίων, οι οποίες εντοπίζονται λίγα νουκλεοτίδια πριν από τα κωδικόνια AUG όπου πρόκειται να αρχίσει η μετάφραση. Αντίθετα από τους ευκαρυώτες, τα ριβοσωμάτια των βακτηρίων δεν δυσκολεύονται να συνδεθούν απευθείας με ένα κωδικόνιο έναρξης που βρίσκεται στο εσωτερικό ενός mRNA μεταγράφου, αρκεί λίγα νουκλεοτίδια πριν από αυτό να υπάρχει μια ειδική θέση σύνδεσης. Για το λόγο αυτό, τα προκαρυωτικά mRNA μετάγραφα συχνά είναι πολυσιστρονικά

(polycistronic): ο όρος αυτός σημαίνει ότι κωδικοποιούν αρκετές διαφορετικές πρωτεΐνες, οι οποίες μεταφράζονται όλες από το ίδιο mRNA. Αντίθετα, ένα ευκαρυωτικό mRNA μετάγραφο συνήθως φέρει πληροφορίες μόνο για μια πρωτεΐνη.

Το τέλος του μηνύματος που κωδικοποιεί την πρωτεΐνη σηματοδοτείται από την παρουσία ενός από τρία ειδικά κωδικόνια (UAA, UAG, ή UGA), γνωστά ως κωδικόνια τερματισμού (stop codons). Τα κωδικόνια τερματισμού δεν αναγνωρίζονται από κάποιο tRNA και δεν εξειδικεύουν κάποιο αμινοξύ, αλλά αντίθετα ειδοποιούν το ριβοσωμάτιο να σταματήσει τη μετάφραση. Πρωτεΐνες γνωστές ως παράγοντες απελευθέρωσης (release factors) συνδέονται με όποιο κωδικόνιο τερματισμού φτάσει στη θέση A του ριβοσωματίου. Η σύνδεση αυτή μεταβάλλει την ενεργότητα της πεπτιδυλοτρανσφεράσης του ριβοσωματίου, κάνοντας την να καταλύει την προσθήκη ενός μορίου νερού αντί ενός αμινοξέος στο πεπτιδυλο-tRNA. Η συγκεκριμένη αντίδραση απελευθερώνει το καρβοξυτελικό άκρο της αυξανόμενης πολυπεπτιδικής αλυσίδας από ένα μόριο tRNA, που αποτελεί τον μοναδικό σύνδεσμο της με το ριβοσωμάτιο και με τον τρόπο αυτό η ολοκληρωμένη πρωτεϊνική αλυσίδα απελευθερώνεται αμέσως στο κυτταρόπλασμα.

Το ριβοσωμάτιο απελευθερώνει το mRNA μετάγραφο και δίσταται στις δύο επιμέρους υπομονάδες του, οι οποίες μπορεί να συναρμολογηθούν πάνω σ' ένα άλλο μόριο mRNA για να αρχίσουν ένα νέο γύρο πρωτεϊνοσύνθεσης. Η σύνθεση των περισσότερων πρωτεϊνών διαρκεί μεταξύ 20 δευτερολέπτων και αρκετών λεπτών.

1.2 Η δομή των πρωτεϊνών

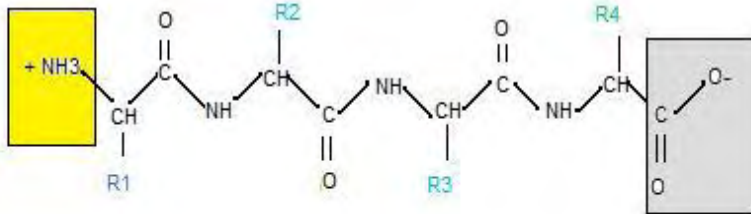
Οι πρωτεΐνες εμφανίζουν τη μεγαλύτερη δομική και λειτουργική πολυπλοκότητα από όλα τα γνωστά μόρια, γεγονός που δεν προξενεί έκπληξη, από χημική σκοπιά, έπειτα από τα δισεκατομμύρια χρόνια εξελικτικής ιστορίας που χρειάστηκαν για να αναπτυχθεί και να συντονιστεί η δομή και η χημεία κάθε πρωτεΐνης. Η κατανόηση της δομής μιας πρωτεΐνης σε ατομικό επίπεδο θα μας επιτρέψει να περιγράψουμε πώς η ακριβής δομή της πρωτεΐνης καθορίζει τη λειτουργία της.

1.2.1 Η δομή μιας πρωτεΐνης καθορίζεται από την αλληλουχία των αμινοξέων της

Όπως είναι γνωστό κάθε πρωτεΐνη αποτελείται από 20 διαφορετικά είδη αμινοξέων, το καθένα από τα οποία έχει διαφορετικές χημικές ιδιότητες. Ένα α-αμινοξύ αποτελείται από ένα κεντρικό άτομο άνθρακα, που λέγεται α-άνθρακας, συνδεδεμένο με μία αμινική ομάδα, μια καρβοξυλική ομάδα, ένα άτομο υδρογόνου και μια χαρακτηριστική ομάδα R. Η ομάδα R ονομάζεται και πλευρική αλυσίδα. Οι πλευρικές αλυσίδες αποτελούν ομάδες του μορίου των αμινοξέων που δε συμμετέχουν στο σχηματισμό του πεπτιδικού δεσμού και που προσδίδουν στο αμινοξύ τις μοναδικές του ιδιότητες. Ορισμένες είναι μη πολικές και υδρόφοβες, άλλες φέρουν αρνητικό και άλλες θετικό φορτίο, άλλες είναι δραστικές και άλλες μη δραστικές. Ένα μόριο πρωτεΐνης αποτελείται από μία μακριά αλυσίδα αμινοξέων, στην οποία το κάθε αμινοξύ συνδέεται με το επόμενο με έναν ομοιοπολικό πεπτιδικό δεσμό. Το μέγεθος των πρωτεϊνών κυμαίνεται από 30 έως 10.000 και πλέον αμινοξέα, ενώ η μεγάλη πλειονότητα των πρωτεϊνών έχουν μήκος 50 έως 2.000 αμινοξέων. Η επαναληπτική αλληλουχία ατόμων κατά μήκος της αλυσίδας ονομάζεται πολυπεπτιδικός σκελετός. Στην επαναληπτική αυτή αλυσίδα συνδέονται οι πλευρικές αλυσίδες. Κάθε πρωτεΐνη έχει μοναδική αλληλουχία αμινοξέων, η οποία είναι ακριβώς ίδια σε όλα τα μόρια της πρωτεΐνης [1, 2].



Εικόνα 1-7. Δομή αμινοξέος. Κάθε αμινοξύ αποτελείται από 4 μέρη: ένα άτομο άνθρακα συνδεδεμένο με μία καρβοξυλική ομάδα (-COOH), μία αμινομάδα (-NH₃), ένα άτομο υδρογόνου (H) και μία πλευρική ομάδα R.



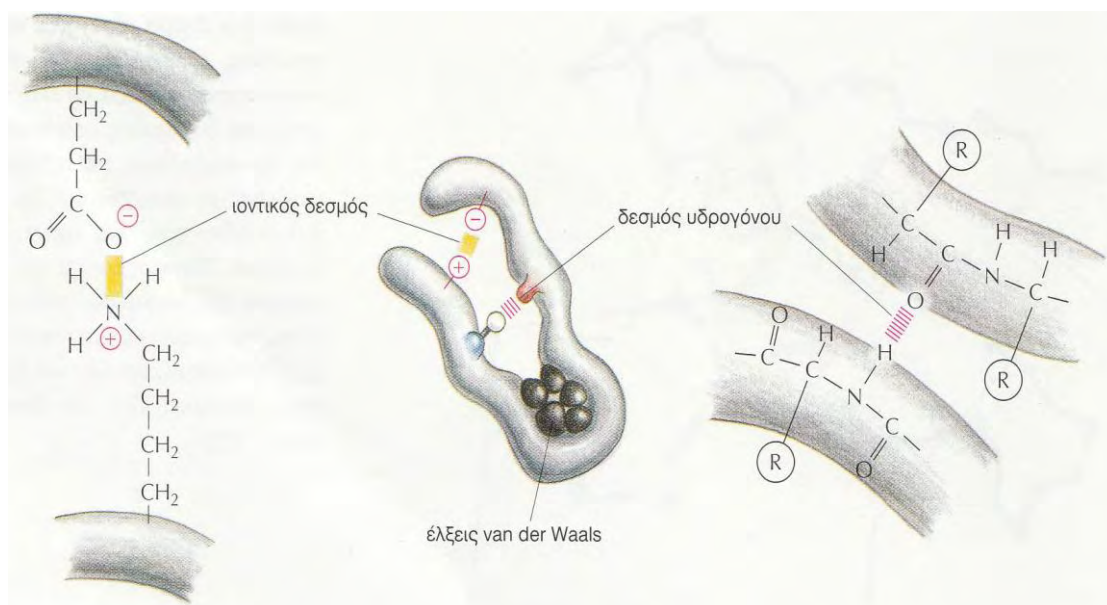
Εικόνα 1-8. Πολυπεπτιδική αλυσίδα που αποτελείται από τέσσερα αμινοξέα.

1.2.2 Αλληλεπιδράσεις και υπάρχοντες δεσμοί στις πρωτεΐνες

Κατά το σχηματισμό των πρωτεϊνών παρατηρούνται τόσο ομοιοπολικοί, όσο και κάποιοι μη ομοιοπολικοί δεσμοί [1, 2]. Πολλοί από τους ομοιοπολικούς δεσμούς σε μία μακριά αλυσίδα αμινοξέων επιτρέπουν στα άτομα που συνδέουν να περιστρέφονται ελεύθερα και εξαιτίας του γεγονότος αυτού, θεωρητικά, ο πολυπεπτιδικός σκελετός μπορεί να διπλωθεί με αμέτρητους τρόπους. Ωστόσο, η ύπαρξη διαφορετικών ειδών ασθενών μη ομοιοπολικών δεσμών, που σχηματίζονται τόσο από άτομα του πολυπεπτιδικού σκελετού όσο και από άτομα των πλευρικών αλυσίδων των αμινοξέων, περιορίζει κάθε διπλωμένη αλυσίδα. Μόνο οι μη ομοιοπολικοί δεσμοί είναι ασθενείς σε σύγκριση με τους ομοιοπολικούς δεσμούς. Έτσι χρειάζονται πολλοί μη ομοιοπολικοί δεσμοί ώστε να συγκρατηθούν δύο περιοχές μιας διπλωμένης πολυπεπτιδικής αλυσίδας. Επομένως η σταθερότητα κάθε διπλωμένης δομής επηρεάζεται από τη συνολική ισχύ πολλών τέτοιων δεσμών.

Υπάρχουν τέσσερα είδη τέτοιων ασθενών δεσμών. Αυτά περιλαμβάνουν τους δεσμούς υδρογόνου, τους ιοντικούς δεσμούς, τις έλξεις van der Waals και τις μη πολικές (υδρόφοβες) πλευρικές αλυσίδες μιας πρωτεΐνης. Οι πλευρικές αλυσίδες αποτελούν ένα σημαντικό παράγοντα που κατευθύνει το δίπλωμα ή πτύχωση κάθε πρωτεΐνης. Οι αλυσίδες που ανήκουν σε αμινοξέα όπως η φαινυλαλανίνη, η λευκίνη, η βαλίνη και η τρυπτοφάνη τείνουν να συναθροίζονται στο εσωτερικό του μορίου. Έτσι αποφεύγουν την επαφή με το νερό που τις περιβάλλει στο εσωτερικό ενός κυττάρου. Αντιθέτως, οι πολικές πλευρικές αλυσίδες (υδρόφιλες), οι αλυσίδες που

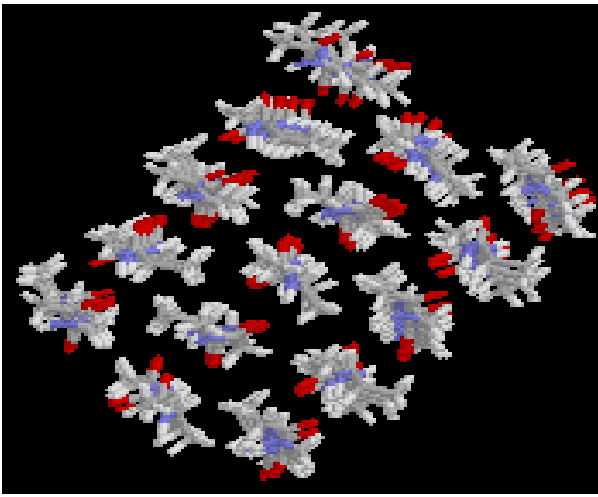
ανήκουν σε αμινοξέα όπως η αργινίνη, η γλουταμίνη και η ιστιδίνη τείνουν να διατάσσονται κοντά στην εξωτερική επιφάνεια της πρωτεΐνης όπου σχηματίζουν δεσμούς υδρογόνου με το νερό και με άλλα πολικά μόρια.



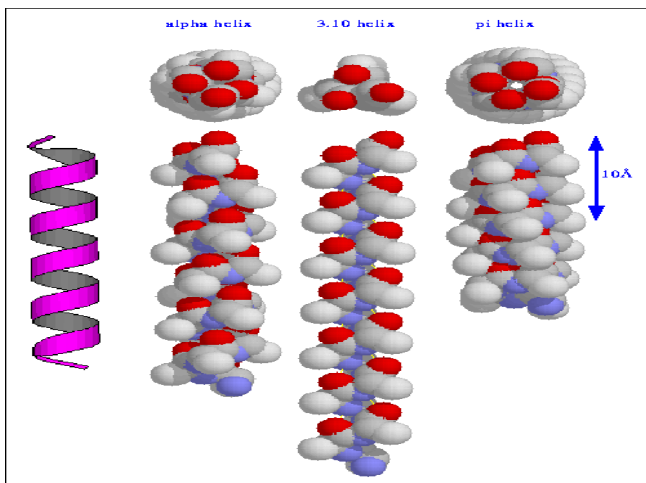
Εικόνα 1-9 . Τρία είδη μη ομοιοπολικών δεσμών που συμβάλλουν στην πτύχωση των πρωτεϊνών. Το R είναι ένα γενικό σύμβολο για τις πλευρικές αλυσίδες.

1.2.3 Επίπεδα οργάνωσης των πρωτεϊνών

Οι πρωτεΐνες είναι οι εργάτες της βιοχημείας, συμμετέχοντας σε όλες σχεδόν τις κυτταρικές διεργασίες. Η πρωτεϊνική δομή μπορεί να περιγραφεί σε τέσσερα επίπεδα (Εικόνα 1- 12). Η αλληλουχία των αμινοξέων μιας πρωτεΐνης αποτελεί την πρωτοταγή της δομή (primary structure). Τα τμήματα της πολυπεπτιδικής αλυσίδας που σχηματίζουν α-έλικες και β-πτυχωτές επιφάνειες συνιστούν τη δευτεροταγή δομή (secondary structure) μιας πρωτεΐνης.

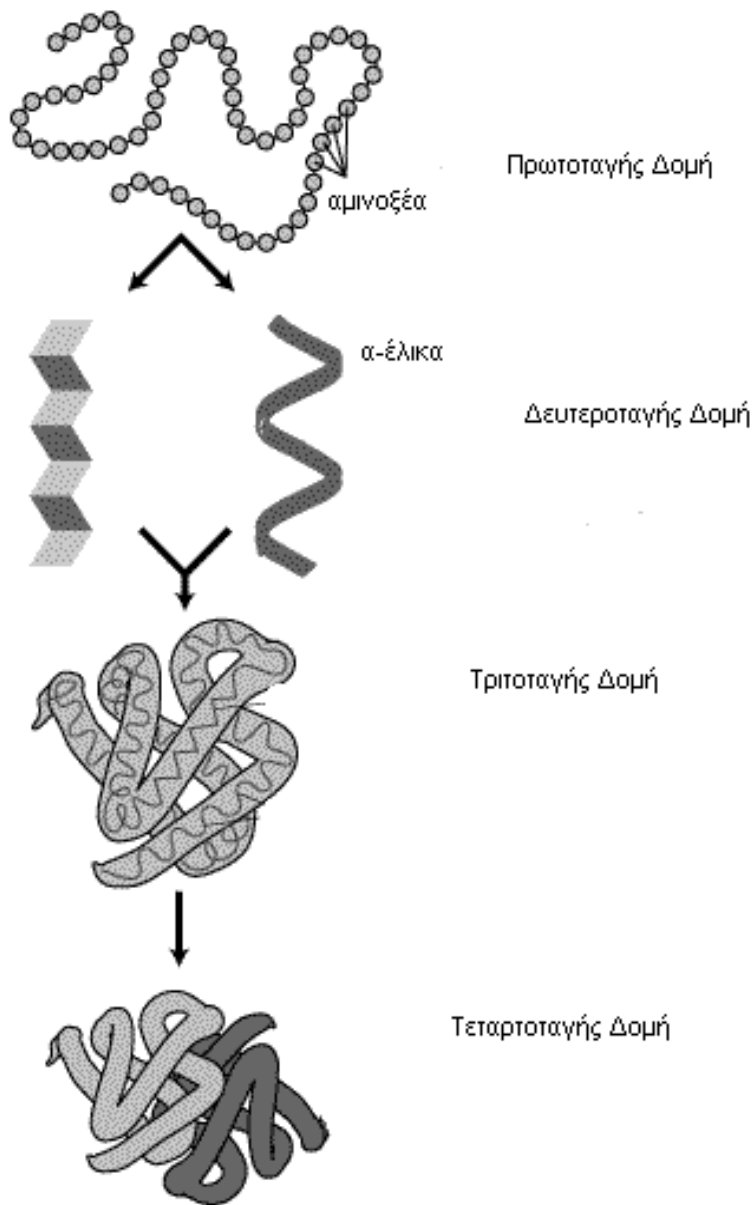


Εικόνα 1-10. Β-πτυχωτή επιφάνεια. Ανακαλύφθηκε στην πρωτεΐνη ινδοΐνη (fibroin), το κύριο συστατικό του μεταξιού.



Εικόνα 1-11. Παρουσίαση σχηματισμού α-έλικας.

Η τρισδιάστατη διαμόρφωση μιας πολυπεπτιδικής αλυσίδας αναφέρεται ως τριτοταγής δομή (tertiary structure) ενώ αν μία πρωτεΐνη σχηματίζεται ως σύμπλοκο δύο ή περισσότερων πολυπεπτιδίων τότε η πλήρης δομή ορίζεται ως τεταρτοταγής δομή (quaternary structure).



Εικόνα 1-12. Οι διαμορφώσεις της πρωτοταγούς, δευτεροταγούς, τριτοταγούς και τεταρτοταγούς δομής των πρωτεϊνών.

1.2.4. Οι πρωτεΐνες μπορούν να ταξινομηθούν σε οικογένειες

Πολλές πρωτεΐνες μπορούν να καταταγούν σε πρωτεϊνικές οικογένειες (protein families) [1, 2]. Αυτό συμβαίνει διότι κατά τη διάρκεια της εξέλιξης, μία πρωτεΐνη που απέκτησε μια συγκεκριμένη διαμόρφωση που της εξασφαλίζει σταθερή δομή με χρήσιμες ιδιότητες, είναι δυνατόν να τροποποιηθεί ελαφρά ώστε να

αποκτήσει την ικανότητα να επιτελεί νέες λειτουργίες. Κάθε μέλος μιας οικογένειας έχει τρισδιάστατη διαμόρφωση και παρεμφερή αλληλουχία αμινοξέων με όλα τα άλλα μέλη της οικογένειας. Χαρακτηριστικό παράδειγμα πρωτεϊνικής οικογένειας αποτελούν οι πρωτεάσες με σερίνη στο ενεργό κέντρο, μία οικογένεια πρωτεολυτικών ενζύμων. Στην οικογένεια αυτή ανήκουν ένζυμα που συμμετέχουν στη διαδικασία της πέψης (θρυψίνη, ελαστάση και χυμοθρυψίνη) καθώς και κάποιες από τις πρωτεάσες που σχετίζονται με την πήξη του αίματος. Όταν γίνεται σύγκριση σε δύο από αυτά τα ένζυμα παρατηρείται πως (i) υπάρχουν ομοιότητες ανάμεσα στις τρισδιάστατες διαμορφώσεις τους, (ii) υπάρχουν τμήματα της αλληλουχίας των αμινοξέων που είναι σχεδόν ταυτόσημα, (iii) έχουν διαφορετική ενζυμική ενεργότητα, (iv) οι περισσότερες από τις κάμπυες και στροφές στις πολυπεπτιδικές αλυσίδες τους είναι παρόμοιες και (v) το καθένα τους διασπά διαφορετικές πρωτεΐνες ή τους πεπτιδικούς δεσμούς ανάμεσα σε διαφορετικά είδη αμινοξέων. Καταλήγουμε λοιπόν στο συμπέρασμα πως το καθένα διεκπεραιώνει μία διαφορετική λειτουργία σε έναν οργανισμό.

1.3 Λειτουργίες των πρωτεϊνών και τρόπος δράσης τους

Στο δεύτερο μέρος του κεφαλαίου θα ασχοληθούμε με τον καθορισμό των λειτουργιών των πρωτεϊνών καθώς και με το βασικό ζήτημα του τρόπου με τον οποίο οι πρωτεΐνες επιτελούν τη λειτουργία τους.

1.3.1 Λειτουργίες των πρωτεϊνών

Σε κάθε πρωτεΐνη η ακολουθία των αμινοξέων καθορίζει την τρισδιάστατη δομή της πρωτεΐνης και η οποία με τη σειρά της καθορίζει τη λειτουργία της. Οι λειτουργίες των πρωτεϊνών ποικίλουν, συνεπώς οι πρωτεΐνες χαρακτηρίζονται ως:

- **Μεταφορικές πρωτεΐνες**

Λειτουργία τους είναι η μεταφορά μικρών μορίων ή ιόντων. Στην κυκλοφορία του αίματος, η λευκωματίνη του ορού μεταφέρει λιπίδια, η αιμοσφαιρίνη οξυγόνο και η τρανσφερρίνη σίδηρο. Πολλές πρωτεΐνες ενσωματωμένες σε διάφορες μεμβράνες του κυττάρου μεταφέρουν ιόντα ή μικρά μόρια διαμέσου των αντίστοιχων μεμβρανών.

- **Δομικές πρωτεΐνες**

Λειτουργία τους είναι η μηχανική στήριξη του κυττάρου και των ιστών. Στον εξωκυττάριο χώρο, το κολλαγόνο και η ελασίνη είναι κοινά συστατικά του εξωκυττάρου στρώματος και σχηματίζουν ίνες στους τένοντες και στους συνδέσμους.

- **Κινητήριες πρωτεΐνες**

Λειτουργία τους είναι η δημιουργία κινήσεων σε κύτταρα και ιστούς. Στα κύτταρα των γραμμωτών μυών η μυοσίνη παρέχει την κινητήρια δύναμη για τις κινήσεις των μυών. Η κινησίνη αλληλεπιδρά με τους μικροσωληνίσκους για να μετακινηθούν διάφορα οργανίδια μέσα στο κύτταρο. Η δυνεΐνη προσφέρει στους κροσσούς και τα μαστίγια των ευκαρυωτικών κυττάρων τη δυνατότητα να χτυπούν.

- **Αποθηκευτικές πρωτεΐνες**

Λειτουργία τους είναι η αποθήκευση μικρών μορίων ή ιόντων. Ο σίδηρος αποθηκεύεται στο ήπαρ συνδεδεμένος με τη μικρή πρωτεΐνη φερριτίνη. Η οβαλβουμίνη που βρίσκεται στο ασπράδι των αυγών χρησιμοποιείται ως πηγή αμινοξέων για τα αναπτυσσόμενα έμβρυα των πτηνών.

- **Σηματοδοτικές πρωτεΐνες**

Λειτουργία τους είναι η μεταφορά σημάτων από κύτταρο σε κύτταρο. Πολλές από τις ορμόνες και τους αυξητικούς παράγοντες που συντονίζουν τη φυσιολογική λειτουργία των ζώων είναι πρωτεΐνες. Έτσι, η ινσουλίνη είναι μία μικρή πρωτεΐνη που ελέγχει τα επίπεδα της γλυκόζης στο αίμα.

- **Πρωτεΐνες-Υποδοχείς**

Χρησιμοποιούνται από τα κύτταρα για την ανίχνευση σημάτων και τη μεταβίβασή τους στον απαντητικό μηχανισμό των κυττάρων. Η ροδοψίνη του αμφιβληστροειδούς ανιχνεύει το φως, ενώ ο υποδοχέας της ινσουλίνης επιτρέπει στα ηπατοκύτταρα να απαντούν στην ινσουλίνη προσλαμβάνοντας γλυκόζη.

- **Πρωτεΐνες Ρύθμισης Γονιδίων**

Συνδέονται με το DNA και ενεργοποιούν ή καταστέλλουν την έκφραση γονιδίων. Στα βακτήρια, ο καταστολέας της λακτόζης σιγεί τα γονίδια που κωδικοποιούν τα ένζυμα τα οποία αποδομούν το σάκχαρο λακτόζη. Πολλές ομοιωτικές πρωτεΐνες δρουν σαν γενετικοί διακόπτες για να ρυθμίσουν την ανάπτυξη διαφόρων πολυκύτταρων οργανισμών συμπεριλαμβανομένου και του ανθρώπου.

- **Πρωτεΐνες με Ειδικές Λειτουργίες**

Η λειτουργία των πρωτεϊνών αυτών είναι πολύ ποίκιλη. Οι οργανισμοί συνθέτουν πολλές πρωτεΐνες με εξαιρετικά εξειδικευμένες ιδιότητες. Οι αντιψυκτικές πρωτεΐνες των ψαριών της Αρκτικής και της Ανταρκτικής προφυλάσσουν το αίμα τους από το να πήξει ενώ μια πράσινη φθορίζουσα πρωτεΐνη στις μέδουσες εκπέμπει ένα πράσινο φως.

- **Ένζυμα**

Καταλύουν τη διάσπαση ή το σχηματισμό των ομοιοπολικών δεσμών. Τα ζωντανά κύτταρα περιέχουν χιλιάδες διαφορετικά ένζυμα, το καθένα από τα οποία καταλύει (επιταχύνει) μια συγκεκριμένη αντίδραση.

1.3.2 Τρόπος δράσης των πρωτεϊνών

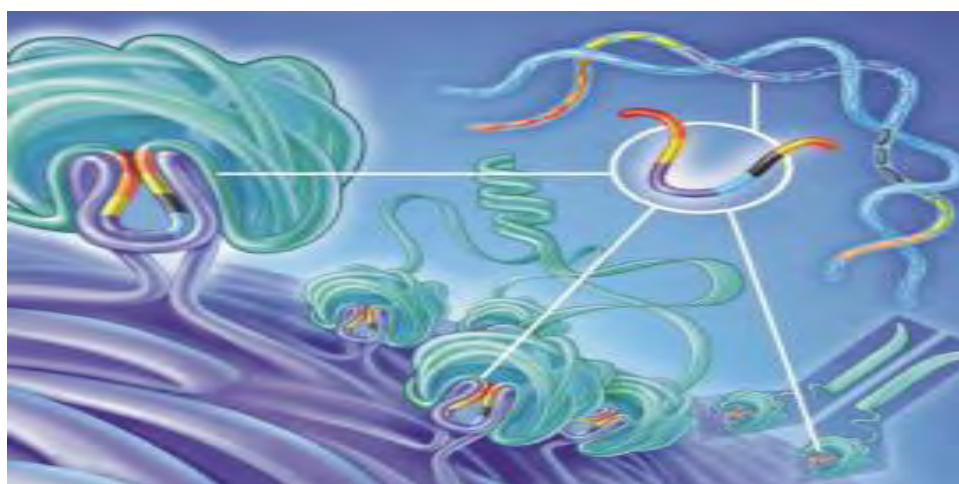
1.3.2.1 Ενζυμική δράση

Οι πρωτεΐνες ως μία τάξη μακρομορίων, καταλύουν πολύ αποτελεσματικά διάφορες χημικές αντιδράσεις λόγω της ικανότητάς τους να προσδένουν εξειδικευμένα μία ευρεία τάξη μορίων. Για το λόγο αυτό λειτουργούν και ως ένζυμα. Τα ένζυμα, οι καταλύτες των βιολογικών συστημάτων, είναι αξιοθαύμαστα μόρια που προσδιορίζουν τον τρόπο των χημικών μετασχηματισμών και μεσολαβούν στον μετασχηματισμό των διαφόρων μορφών ενέργειας. Κυριότερα χαρακτηριστικά των ενζύμων είναι η καταλυτική ισχύς και η εξειδίκευσή τους. Σχεδόν όλα τα γνωστά ένζυμα είναι πρωτεΐνες.

Η κατάλυση λαμβάνει χώρα σε μία ιδιαίτερη περιοχή του ενζύμου που ονομάζεται ενεργό κέντρο. Ωστόσο, οι πρωτεΐνες δεν έχουν το μονοπώλιο της κατάλυσης. Η ανακάλυψη των καταλυτικά ενεργών μορίων RNA παρέχει τις αναγκαίες ενδείξεις ότι το RNA ήταν ένας από τους πρώτους βιοκαταλύτες. Τα ένζυμα, χρησιμοποιώντας ολόκληρο το απόθεμα των ενδομοριακών τους δυνάμεων, φέρνουν μαζί τα υποστρώματα σε έναν άριστο προσανατολισμό που είναι η προετοιμασία για τον σχηματισμό και τη διάσπαση χημικών δεσμών. Τα ένζυμα καταλύουν αντιδράσεις με το να σταθεροποιούν τις μεταβατικές καταστάσεις, τις υψηλότερες ενεργειακά στερεοδιατάξεις στις πορείες των αντιδράσεων. Ένα ένζυμο με το να σταθεροποιεί επιλεκτικά μία μεταβατική κατάσταση, προσδιορίζει ποια από τις αντιδράσεις που μπορούν να πραγματοποιηθούν όντως λαμβάνει χώρα.

1.3.2.2 Αντισώματα

Όλες οι πρωτεΐνες για να διεκπεραιώσουν τις ποικίλες λειτουργίες τους πρέπει να συνδεθούν με συγκεκριμένους συνδέτες. Τα αντισώματα ή αλλιώς ανοσοσφαιρίνες, είναι πρωτεΐνες που παράγονται από το ανοσοποιητικό σύστημα σε απάντηση προς ξένα μόρια, όπως τα μόρια της επιφάνειας ενός παθογόνου μικροοργανισμού. Κάθε αντίσωμα συνδέεται εξαιρετικά ισχυρά με ένα ορισμένο μόριο-στόχο και με τον τρόπο αυτό είτε το αδρανοποιεί άμεσα είτε το σηματοδίδει για να επακολουθήσει η καταστροφή του.

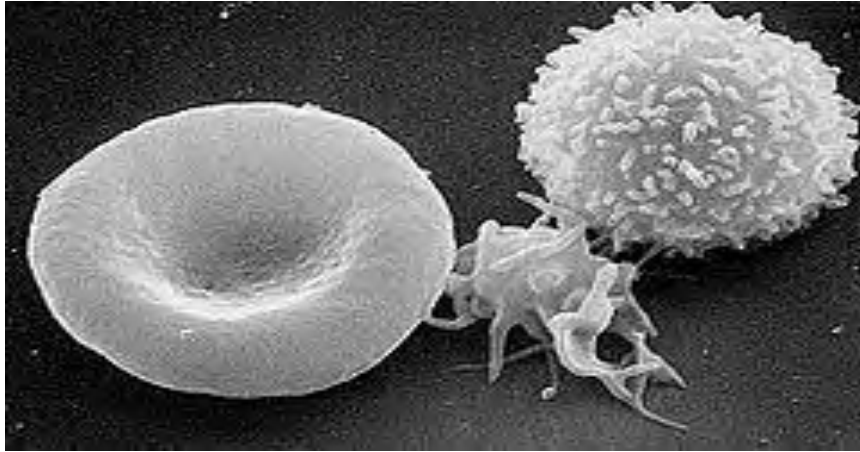


Εικόνα 1-13. Αντισώματα.

Κάθε αντισώμα αναγνωρίζει το στόχο του με αξιοσημείωτη εξειδίκευση. Τα αντισώματα είναι μόρια με σχήμα Y που φέρουν δύο ταυτόσημες θέσεις πρόσδεσης, καθεμία από τις οποίες είναι συμπληρωματική με ένα μικρό τμήμα της επιφάνειας του μορίου του αντιγόνου. Οι θέσεις πρόσδεσης με το αντιγόνο σχηματίζονται από αρκετούς βρόχους της πολυπεπτιδικής αλυσίδας οι οποίοι προεξέχουν από τα άκρα δύο πρωτεϊνικών περιοχών που βρίσκονται απέναντι η μία στην άλλη και σε μικρή απόσταση. Η αλληλουχία αμινοξέων σε αυτούς τους βρόχους μπορεί να μεταβληθεί από μεταλλάξεις χωρίς να τροποποιηθεί η βασική δομή του αντισώματος. Εξαιτίας αυτού, με τη μεταβολή μόνο του μήκους ή της αλληλουχίας των αμινοξέων των βρόχων μπορεί να δημιουργηθεί μια τεράστια ποικιλία θέσεων σύνδεσης για το αντιγόνο. Με τον τρόπο αυτό προκύπτει η μεγάλη ποικιλία των διαφορετικών αντισωμάτων ενός οργανισμού.

1.3.2.3 Μεταφορική δράση

Κατά την ισχυρή σύνδεσή τους με μικρά μόρια, οι πρωτεΐνες αποκτούν επιπρόσθετες λειτουργίες. Η αιμοσφαιρίνη είναι μία πρωτεΐνη που περιέχει ένα μη πρωτεϊνικό τμήμα. Ένα μόριο αιμοσφαιρίνης αποτελείται από τέσσερις ομάδες αίμης. Η αίμη συντίθεται από ένα σύνολο δακτυλίων στο κέντρο των οποίων βρίσκεται ένα άτομο σιδήρου και προσδίδει στην αιμοσφαιρίνη (και γενικότερα στο αίμα) το χαρακτηριστικό κόκκινο χρώμα. Η αίμη συνδέεται με το αέριο οξυγόνο μέσω του ατόμου του σιδήρου και με τον τρόπο αυτό η αιμοσφαιρίνη παραλαμβάνει οξυγόνο στους πνεύμονες και το απελευθερώνει στους ιστούς.

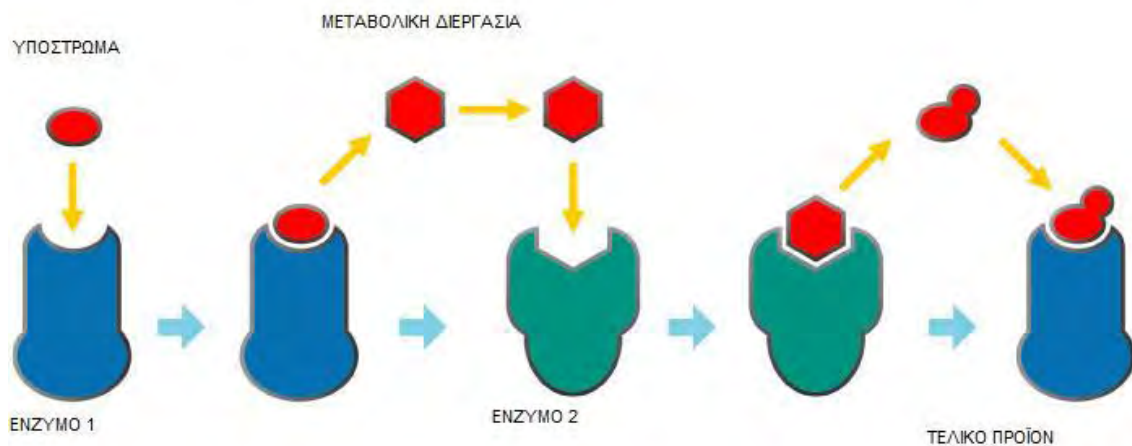


Εικόνα 1-14. Αναπαράσταση της πρωτεΐνης αιμοσφαιρίνης.

Ένα ακόμη χαρακτηριστικό παράδειγμα αποτελεί η ροδοψίνη, μία πορφυρή φωτοευαίσθητη πρωτεΐνη που δρα ως υποδοχέας σημάτων, η οποία παράγεται από τα ραδιοφόρα κύτταρα του αμφιβληστροειδούς. Η ροδοψίνη ανιχνεύει το φως μέσω της ρετινάλης, ενός μορίου ενσωματωμένου στην πρωτεΐνη. Όταν η ρετινάλη απορροφήσει ένα φωτόνιο, αλλάζει η δομή. Η μεταβολή αυτή ενισχύεται από την πρωτεΐνη για να πυροδοτήσει μια ακολουθία ενζυμικών αντιδράσεων που οδηγούν σε μεταφορά ενός ηλεκτρικού σήματος στον εγκέφαλο.

1.3.2.4 Ανάδρομη τροφοδότηση

Το πιο κοινό είδος ελέγχου στις χημικές αντιδράσεις συμβαίνει όταν μεταβάλλεται η ταχύτητα με την οποία το ένζυμο μετατρέπει τα υποστρώματά του σε προϊόντα. Κατά την αναστολή από ανάδρομη τροφοδότηση (feedback inhibition), ένα ένζυμο που δρα σε ένα πρώιμο στάδιο μιας οδού αντιδράσεων αναστέλλεται από το προϊόν μιας μετέπειτα αντίδρασης της ίδιας οδού. Έτσι, περιορίζεται η περαιτέρω είσοδος υποστρωμάτων στη συγκεκριμένη οδό κάθε φορά που συσσωρεύονται μεγάλες ποσότητες του τελικού προϊόντος μιας μεταβολικής οδού.



Εικόνα 1-15. Ανάδρομη τροφοδότηση.

1.3.2.5 Αλλοστερική δράση

Τα περισσότερα πρωτεϊνικά μόρια είναι αλλοστερικά (allosteric). Μπορούν δηλαδή, να προσλάβουν δύο ή περισσότερες ελαφρώς διαφορετικές διαμορφώσεις και ότι η ενεργότητά τους μπορεί να μεταβληθεί με μετάπτωση από μία διαμόρφωση σε μία άλλη. Αυτό ισχύει όχι μόνο για τα ένζυμα αλλά και για πολλές άλλες πρωτεΐνες όπως υποδοχείς, δομικές πρωτεΐνες και κινητήριες πρωτεΐνες. Η θέση σύνδεσης για έναν ή περισσότερους προσδέτες μπορεί να μεταβάλλεται εύκολα καθώς η κάθε διαμόρφωση έχει κάπως διαφορετικό επιφανειακό περίγραμμα. Κάθε προσδέτης ευνοεί τη διαμόρφωση με την οποία συνδέεται πιο ισχυρά και σε αρκετά μεγάλη συγκέντρωση, μπορεί να μετατρέψει τη διαμόρφωση της πρωτεΐνης από τη μία στην άλλη.

1.3.2.6 Κινητήριες πρωτεΐνες

Μία από τις λειτουργίες των πρωτεϊνών αφορά τη μετακίνηση άλλων μορίων. Οι πρωτεΐνες αυτές χαρακτηρίζονται ως κινητήριες και αναπτύσσουν δυνάμεις που ευθύνονται για τη συστολή των μυών και τις εντυπωσιακές κινήσεις των κυττάρων. Οι κινητήριες πρωτεΐνες ευθύνονται επίσης για τις ενδοκυττάριες μετακινήσεις, με τον τρόπο αυτό βοηθούν στη μετακίνηση των χρωμοσωμάτων προς τα αντίθετα άκρα

του κυττάρου κατά τη διάρκεια της μίτωσης, μετακινούν οργανίδια κατά μήκος “μοριακών διόδων” μέσα στο κύτταρο και τέλος μετακινούν ένζυμα κατά μήκος ενός κλώνου DNA κατά τη σύνθεση ενός νέου μορίου DNA.

2. ΜΕΤΑ-ΜΕΤΑΦΡΑΣΤΙΚΕΣ ΤΡΟΠΟΠΟΙΗΣΕΙΣ ΠΡΩΤΕΪΝΩΝ (PTM-Post Translational Modifications)

Οι πρωτεΐνες μετά τη μετάφρασή τους υφίστανται τροποποιήσεις γνωστές ως μετά-μεταφραστικές τροποποιήσεις των πρωτεϊνών. Οι μετά-μεταφραστικές τροποποιήσεις γίνονται κατά κύριο λόγο στο ενδοπλασματικό δίκτυο και στη συσκευή Golgi ακολουθώντας ορισμένες πολύπλοκες διεργασίες και συσχετίζονται με αλλαγές στερεοδιάταξης, με την πρόσδεση στη μεμβράνη καθώς και με τον τελικό προορισμό της πρωτεΐνης μέσα στο κύτταρο. Είναι δυνατό να περιλαμβάνουν τροποποίηση του αμινοτελικού άκρου, τροποποίηση μεμονωμένων αμινοξέων, πρωτεολυτική επεξεργασία και σχηματισμό δισουλφιδικών γεφυρών. Η μελέτη των τροποποιήσεων αυτών είναι σημαντική για την κατανόηση του μηχανισμού παθογένειας καθώς και για την τελική διαμόρφωση των πρωτεϊνών στο χώρο και για την λειτουργικότητά τους.

Μια από τις συνήθεις τροποποιήσεις του αμινοτελικού άκρου των πρωτεϊνών των ευκαρυωτικών κυττάρων είναι η απομάκρυνση του υπολείμματος μεθειονίνης με το οποίο αρχίζει η σύνθεση των πρωτεϊνών. Στα βακτηριακά κύτταρα, απομακρύνεται η φορμυλομάδα από τη μεθειονίνη και σε μερικές περιπτώσεις απομακρύνεται και το ίδιο το αμινοξύ. Εκτός από τις περιπτώσεις αυτές που είναι δυνατό να γίνουν στο αμινοτελικό άκρο της πρωτεΐνης, οποιοδήποτε άλλο αμινοξύ μέσα στην πρωτεΐνη μπορεί να τροποποιηθεί. Για παράδειγμα, οι πλευρικές ομάδες των αμινοξέων σερίνη, θρεονίνη και τυροσίνη μπορεί να συνδεθούν με φωσφορικές ομάδες. Αυτός ο τύπος τροποποίησης χρησιμοποιείται από το κύτταρο για να μεταδώσει μεταβολές στη λειτουργία ορισμένων ρυθμιστικών οδών που είναι αποτέλεσμα μεταβολών του περιβάλλοντός του.

Μερικά αμινοξέα, όπως η λυσίνη, τροποποιούνται με προσθήκη μιας μεθυλομάδας ενώ άλλα, όπως η κυστεΐνη, μπορούν να υποστούν προσθήκη ομάδων ισοπρενίου ή άλλων λιπιδίων για να διευκολύνουν τη σύνδεση των πρωτεϊνών με τις μεμβράνες του κυττάρου. Τα υπολείμματα κυστεΐνης σχηματίζουν επίσης ειδικές δισουλφιδικές γέφυρες που είναι σημαντικές για τη δομική ακεραιότητα ορισμένων πρωτεϊνών. Τέλος, πολλές πρωτεΐνες συντίθεται ως πρόδρομες μορφές (προ-πρωτεΐνες) που πρέπει να υποστούν πρωτεολυτική διάσπαση για να μετατραπούν στις αντίστοιχες δραστικές πρωτεΐνες. Η διάσπαση μιας προ-πρωτεΐνης προς τη βιολογικά

δραστική της μορφή επιτελείται συνήθως από μια ειδική πρωτεάση και αποτελεί μία ρυθμιζόμενη κυτταρική λειτουργία.

Μέρος 1^ο

2.1 Είδη Μετά-Μεταφραστικών Τροποποιήσεων

Οι επιστήμονες στηρίζονται στο γεγονός ότι η πρωτεΐνη δεν είναι μόνη της όταν εκτελεί των βιολογικό της ρόλο. Χρειάζεται να χρησιμοποιήσει τους ίδιους κυτταρικούς μηχανισμούς για τροποποίηση και ταξινόμηση όπως κάνουν και όλες οι πρωτεΐνες. Στις σημαντικότερες μορφές μετά-μεταφραστικών τροποποιήσεων περιλαμβάνονται οι ακόλουθες:

- ❖ φωσφορυλίωση (phosphorylation)
- ❖ γλυκοζυλίωση (glycosylation)
- ❖ μυριστοϋλίωση (myristoylation)
- ❖ σουλφιδιλίωση (ή σουλφίδωση) (sulfation)
- ❖ παλμιτοϋλίωση (palmitoylation)
- ❖ σουμοϋλίωση (sumoylation)
- ❖ ακετυλίωση (acetylation)
- ❖ γλυκίωση (glycation)

Ωστόσο, υπάρχουν και άλλες γνωστές μορφές μετά-μεταφραστικών τροποποιήσεων οι οποίες περιλαμβάνουν: λιπιδίωση (lipidation), ουβικουϊτινίωση (ubiquitination), πρενυλίωση (prenylation) καθώς και αποκοπή σήματος οδηγού. Πολλές από τις μετά-μεταφραστικές τροποποιήσεις χαρακτηρίζονται από μοτίβα, ενώ άλλες από πιο περίπλοκα πρότυπα συσχέτισης μεταξύ των αμινοξέων.

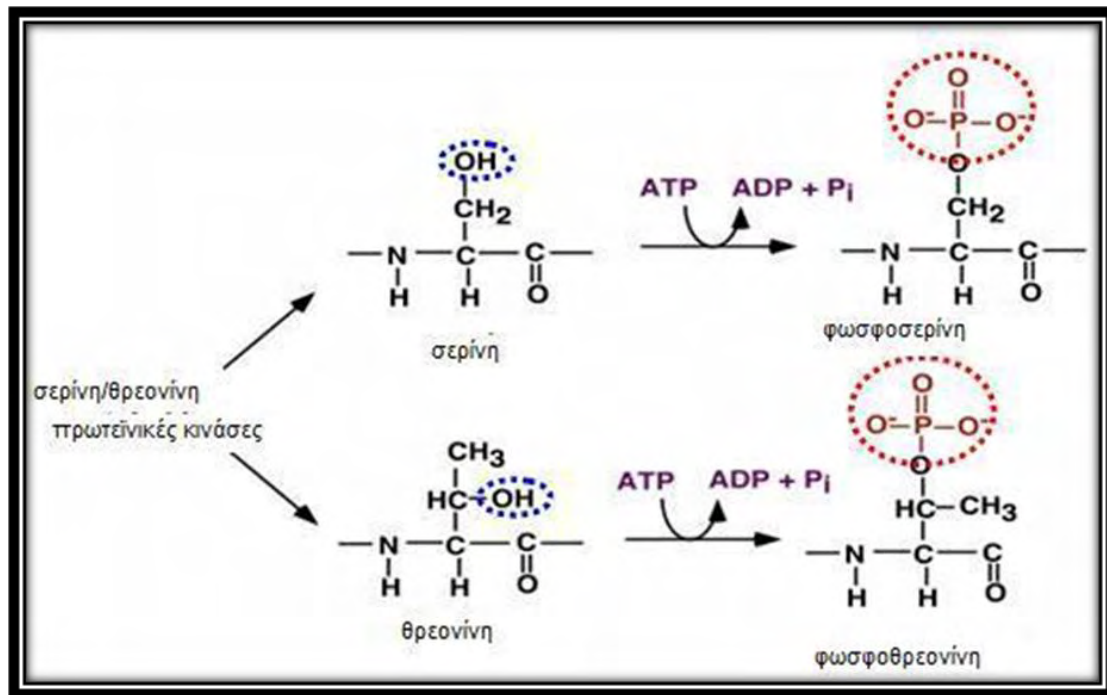
2.1.1 Φωσφορυλίωση

Όπως αναφέρθηκε, μία από τις σημαντικότερες μετά-μεταφραστικές τροποποιήσεις είναι η φωσφορυλίωση. Με τον όρο φωσφορυλίωση ορίζεται η μέθοδος που χρησιμοποιείται συχνά από τα ευκαρυωτικά κύτταρα για τη ρύθμιση της λειτουργίας μιας πρωτεΐνης και συνίσταται στην ομοιοπολική προσθήκη μιας φωσφορικής ομάδας στην πλευρική αλυσίδα ενός αμινοξέος. Είναι ένας παγκόσμιος μηχανισμός που περιλαμβάνει πολλαπλές λειτουργίες στα κύτταρα των Ευκαριωτών, των Προκαρυωτών και των Αρχαίων. Επειδή κάθε φωσφορική ομάδα φέρει δύο αρνητικά φορτία, η καταλυόμενη ενζυμικά προσθήκη μιας φωσφορικής ομάδας σε μία πρωτεΐνη μπορεί να επιφέρει μία σημαντική μεταβολή διαμόρφωσης, όπως για παράδειγμα, να προσελκύσει μια ομάδα από θετικά φορτισμένες πλευρικές αλυσίδες ορισμένων αμινοξέων. Η αφαίρεση της φωσφορικής ομάδας από κάποιο άλλο ένζυμο επαναφέρει την πρωτεΐνη στην αρχική της διαμόρφωση και αποκαθιστά την προηγούμενη ενεργότητά της.

Ο αντιστρεπτός αυτός μηχανισμός ελέγχει την ενεργότητα πολλών διαφορετικών πρωτεϊνών των ευκαρυωτικών κυττάρων. Σε οποιαδήποτε χρονική στιγμή περίπου πάνω από το 1/3 των 10.000 πρωτεϊνών ενός κοινού κυττάρου θηλαστικού πιστεύεται ότι είναι φωσφορυλιωμένες. Η προσθήκη και η αφαίρεση φωσφορικών ομάδων από ειδικές πρωτεΐνες (ένζυμα) συχνά συμβαίνει ως απάντηση στα σήματα που προδιαγράφουν κάποια μεταβολή στην κατάσταση του κυττάρου. Για παράδειγμα, η περίπλοκη σειρά των γεγονότων που συμβαίνουν κατά τη διαίρεση ενός ευκαρυωτικού κυττάρου προγραμματίζεται χρονικά με το μηχανισμό αυτό. Επίσης, πολλά από τα σήματα που παράγονται από ορμόνες και νευρομεταβιβαστές μεταφέρονται από την κυτταρική μεμβράνη στον πυρήνα με μια ακολουθία αντιδράσεων φωσφορυλίωσης των πρωτεϊνών.

Η φωσφορυλίωση μιας πρωτεΐνης γίνεται με ενζυμική μεταφορά της τελικής φωσφορικής ομάδας του ATP στο υδροξύλιο της πλευρικής αλυσίδας της σερίνης, της θρεονίνης ή της τυροσίνης μιας πρωτεΐνης. Η αντίδραση καταλύεται από μία πρωτεϊνική κινάση ενώ η αντίστροφη αντίδραση δηλαδή η αφαίρεση μιας φωσφορικής ομάδας (αποφωσφορυλίωση) από μια πρωτεϊνική φωσφατάση (Εικόνα 2-1). Τα κύτταρα περιέχουν εκατοντάδες διαφορετικές πρωτεϊνικές κινάσες, που καθεμία είναι υπεύθυνη για τη φωσφορυλίωση μιας διαφορετικής πρωτεΐνης ή μιας ομάδας πρωτεϊνών ενώ υπάρχουν και πολλές διαφορετικές πρωτεϊνικές φωσφατάσες. Ορισμένες από αυτές είναι πολύ ειδικές και αφαιρούν φωσφορικές ομάδες μόνο από

μία ή λίγες πρωτεΐνες ενώ αντίθετα άλλες δρουν σε μεγάλη ποικιλία πρωτεϊνών. Η κατάσταση φωσφορυλίωσης μιας πρωτεΐνης σε μια δεδομένη χρονική στιγμή, επομένως και η δραστηρότητά της, εξαρτάται από τη σχετική ενεργότητα των κινασών και φωσφατασών που επιδρούν στην πρωτεΐνη αυτή.



Εικόνα 2-1. Πρωτεϊνική φωσφορυλίωση. Γίνεται ενζυμική μεταφορά της τελικής φωσφορικής ομάδας του ATP στο υδροξύλιο της πλευρικής αλυσίδας της σερίνης, της θρεονίνης ή της τυροσίνης μιας πρωτεΐνης. Η αντίδραση καταλύεται από μία πρωτεϊνική κινάση ενώ η αντίστροφη αντίδραση δηλαδή η αφαίρεση μιας φωσφορικής ομάδας (αποφωσφορυλίωση) από μια πρωτεϊνική φωσφατάση.

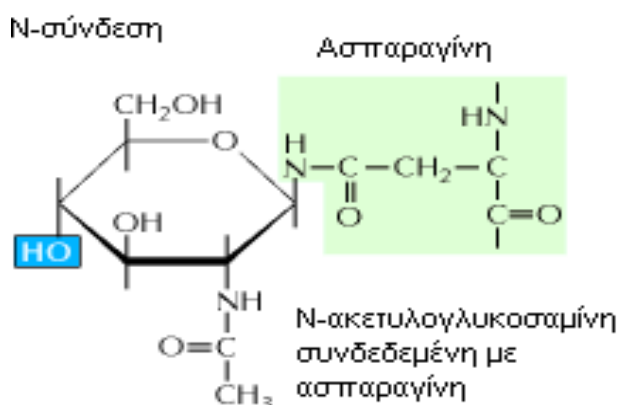
2.1.1.1 Η φωσφορυλίωση στα φυτά

Η φωσφορυλίωση είναι ένας βασικός ρυθμιστικός παράγοντας σε όλες τις πτυχές της βιολογίας των φυτών αλλά είναι δύσκολο, εάν όχι αδύνατο, για τους περισσότερους ερευνητές να προσδιορίσουν τις *in vivo* (σε ζωντανό οργανισμό) περιοχές φωσφορυλίωσης μέσα στις πρωτεΐνες του ενδιαφέροντός τους. Οι επιστήμονες έχουν αναπτύξει μια μεγάλης κλίμακας στρατηγική για την απομόνωση των φωσφοπεπτιδίων και τον προσδιορισμό τους από τη μαζική φασματομετρία [3].

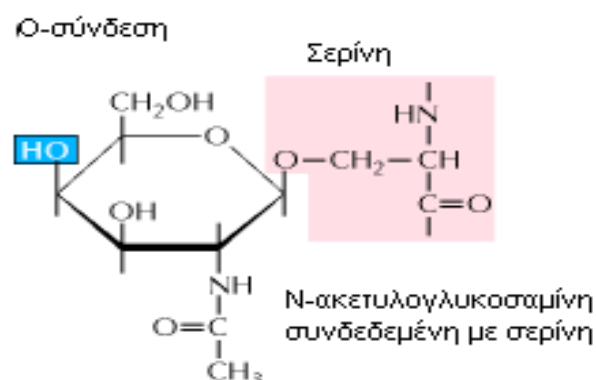
Έχουν προσδιοριστεί περισσότερες από 300 περιοχές φωσφορυλίωσης από τις μεμβρανικές πρωτεΐνες του πλάσματος της *Arabidopsis thaliana*. Χαρτογραφήθηκαν περισσότερες από 50 περιοχές στόχοι κινασών και αποκάλυψαν μια μη-αναμενόμενη πολυπλοκότητα.

2.1.2 Γλυκοζυλίωση

Μία ακόμη από τις πιο σημαντικές μετά-μεταφραστικές τροποποιήσεις των πρωτεϊνών αποτελεί η γλυκοζυλίωση, δηλαδή η ομοιοπολική σύνδεση των ολιγοσακχαριτών ώστε να επιλέξουν τα κατάλοιπα-στόχους ασπαραγίνης. Υπάρχουν δύο κύρια είδη γλυκοζυλίωσης, η O-γλυκοζυλίωση και η N-γλυκοζυλίωση (Εικόνες 2-2 και 2-3).



Εικόνα 2-2. Η N-γλυκοζυλίωση.



Εικόνα 2-3. Η O-γλυκοζυλίωση.

2.1.2.1 Ο-Γλυκοζυλίωση

Η πρωτεϊνική γλυκοζυλίωση με την Ο-συνδεδεμένη Ν-ακετυλογλυκοσαμίνη (O-GlcNAc) είναι μια αντιστρεπτή μετά-μεταφραστική τροποποίηση των σερίνων/θρεονινών στις πρωτεΐνες των μεταζώων και εμφανίζεται σε όμοια χρονικά διαστήματα όπως συμβαίνει και με την πρωτεϊνική φωσφορυλίωση. Τα επίπεδα αυτής της τροποποίησης ρυθμίζονται από δύο ένζυμα, την O-GlcNAc τρανσφεράση (OGT) και την O-GlcNAc υδρολάση (OGA). Παρόλο που έχουν μελετηθεί εκτενώς η βιοχημεία και οι λειτουργικές επιπτώσεις αυτών των O-GlcNAc ενζύμων, οι δομές και οι μοριακοί τους μηχανισμοί δεν έχουν γίνει κατανοητοί μέχρι σήμερα.

Η O-GlcNAc ανακαλύφθηκε περισσότερο από 20 έτη πριν [4], σπάζοντας το δόγμα ότι η πρωτεϊνική γλυκοζυλίωση ήταν περιορισμένη στο ενδοπλασματικό δίκτυο, τις συσκευές Golgi, την επιφάνεια των κυττάρων και στο εξωκυτταρικό περίβλημα [5]. Η O-GlcNAc βρέθηκε αντ'αυτού στο πυρηνόπλασμα, στο κυτταρόπλασμα και ως ένα ορισμένο βαθμό στα μιτοχόνδρια. Μέχρι στιγμής η O-GlcNAc έχει βρεθεί σε όλα τα μετάζωα που έχουν μελετηθεί. Υπάρχουν στοιχεία για σύνδεση της O-GlcNAc με νευροεκφυλιστικές ασθένειες, με τον καρκίνο και το διαβήτη [6]. Κατά συνέπεια, έχει αποδειχθεί πως είναι ουσιαστική για τη νευρωνική λειτουργία και την επιβίωση στα ποντίκια [7] και έχει προταθεί ένας συσχετισμός μεταξύ των παραλλαγών των επιπέδων της Ο-γλυκοζυλίωσης και της ασθένειας του Alzheimer [7, 8]. Υπάρχουν επίσης σημαντικά στοιχεία σχετικά με τη σύνδεση των αυξανόμενων επιπέδων της O-GlcNAc στην ανάπτυξη των αντιδράσεων της ινσουλίνης και στις διαβητικές επιπλοκές, όπως είναι η υπερλεπτιναιμία, η υπερτροφία και η αρτηριοσκλήρωση [6]. Σε μια πολύ πρόσφατη μελέτη, ο Yang [9] περιέγραψε πως το ανθρώπινο OGT (hOGT) εξασθενίζει την έκφραση των γονιδίων ινσουλίνης και προκαλεί την αντίδραση της ινσουλίνης και την δυσλιπιδαιμία μέσω της Ο-γλυκοζυλίωσης κατά τη βασική σηματοδότηση της κινάσης PKB.

Στην υπόθεση Yang παρουσιάζει ιδιαίτερο ενδιαφέρον η πρόταση ότι η O-GlcNAc παρουσιάζει αλληλεπίδραση με την πρωτεϊνική φωσφορυλίωση και σε ορισμένο βαθμό μπορεί να ελέγχει τα επίπεδά της [6]. Σε περισσότερο από το 50% των γνωστών περιοχών της O-GlcNAc συναντάται το μοτίβο Προλίνη-Βαλίνη-Σερίνη (Pro-Val-Ser), παρόμοιο με αυτό που αναγνωρίστηκε κοντά στις κατευθυνόμενες από προλίνη κινάσες [6] και σε μερικές περιπτώσεις η O-GlcNAc είναι αντίστοιχη με τη φωσφορυλίωση. Επίσης, η O-GlcNAc μπορεί να εμφανιστεί δίπλα στις θέσεις φωσφορυλίωσης επηρεάζοντας έτσι το βαθμό της φωσφορυλίωσης.

2.1.2.2 N-Γλυκοζυλίωση

Η πρωτεϊνική N-γλυκοζυλίωση είναι μία καλά διαταγμένη διαδικασία που εμπλέκει μια σειρά ενζύμων σχετικών με τη βιοσύνθεση των σακχάρων, τον ολιγομερισμό, τη δυνατότητα διακίνησης, μεταφοράς και αναδιαμόρφωσης.

Στους ζυμομήκυτες και στα ανώτερα ευκαρυωτικά κύτταρα, όπου η N-γλυκοζυλίωση έχει περιγραφεί καλύτερα, τα γονίδια γλυκοζυλίωσης που συνδέονται με ασπαραγίνη (ALG) κωδικοποιούν συγκεκριμένες γλυκοτρανσφεράσες που αντιδρούν για να συγκεντρώσουν τη διαλυτή ουσία των σακχάρων σε ένα επτασακχαρίδιο, σε ένα μεταφορέα στη μεμβράνη του ενδοπλασματικού δικτύου [10, 11]. Μόλις συγκεντρωθεί το επτασακχαρίδιο μεταφέρεται πέρα από τη μεμβράνη του ενδοπλασματικού δικτύου σε μια διαδικασία που μεσολαβείται στους ζυμομήκυτες από την Rft1 [12]. Κατά την άφιξη στη μεμβράνη του ενδοπλασματικού δικτύου, το επτασακχαρίδιο τροποποιείται από κάποιες πρόσθετες υπομονάδες σακχάρων. Στη συνέχεια αυτός ο ολιγοσακχαρίτης μεταφέρεται στα επιλεγμένα κατάλοιπα Asn από τα μοτίβα ακολουθίας N-X-S/T (όπου το X δεν είναι η Προλίνη). Ο δεμένος με την πρωτεΐνη ολιγοσακχαρίτης υφίσταται έπειτα περαιτέρω τροποποίηση στο ενδοπλασματικό δίκτυο και τη συσκευή Golgi.

2.1.2.3 Η Γλυκοζυλίωση στα Αρχαία

Ενώ οι διαβάσεις της N-γλυκοζυλίωσης στα ευκαρυωτικά κύτταρα και τα βακτήρια είναι σχετικά καλά καθορισμένες, λίγα είναι γνωστά για την παράλληλη διαδικασία στα αρχαία. Εντούτοις, έχουν γίνει σημαντικές πρόοδοι για την περιγραφή της διαδικασίας στα κύτταρα αυτά. Τέτοιες προσπάθειες έχουν παρουσιάσει, πως η πρωτεϊνική N-γλυκοζυλίωση στα αρχαία συνδυάζει ιδιαίτερες πτυχές των μεταβολικών μονοπατιών των ευκαρυωτών και των βακτηρίων μαζί με τα μοναδικά για αυτή τη μορφή ζωής γνωρίσματα. Για παράδειγμα, ενώ οι ολιγοσακχαρίτες των γλυκοπρωτεϊνών των αρχαίων περιλαμβάνουν σάκχαρα που διαμορφώνονται από τα μεταβολικά μονοπάτια των βακτηρίων, ο μεταφορέας λιπιδίων στον οποίο συγκεντρώνονται τέτοιοι ολιγοσακχαρίτες είναι ίδιος με αυτόν που χρησιμοποιείται στην ευκαρυωτική N-γλυκοζυλίωση. Αντιθέτως όμως, η μεταφορά των συγκεντρωμένων ολιγοσακχαριτών στους πρωτεϊνικούς στόχους τους παρουσιάζει τις συγκεκριμένες ιδιότητες των αρχαίων.

Πριν από 30 έτη όπου και περιγράφηκαν για πρώτη φορά, τα αρχαία χρησίμευσαν για να καταρριφθούν πολλά βιολογικά δόγματα. Ένα τέτοιο σχετικό παράδειγμα είναι η πρωτεϊνική N-γλυκοζυλίωση. Με την ανακάλυψη ότι η επιφάνεια του (S) στρώματος της γλυκοπρωτεΐνης, στον οργανισμό *Halobacterium salinarum*, περιλαμβάνει τμήματα πολυσακχαριτών που συνδέονται με επίλεκτα κατάλοιπα ασπαραγίνης [17], καταρρίφθηκε η άποψη πως η N-γλυκοζυλίωση αποτελεί μοναδικό γνώρισμα των ευκαρυωτικών κυττάρων. Αν και τα βακτήρια, όπως το *Campylobacter jejuni*, είχε αποδειχτεί πως εμφάνιζαν N-γλυκοζυλίωση [18], αυτή η μετά-μεταφραστική τροποποίηση είναι πολύ πιο επικρατούσα στα αρχαία, όπου εμφανίζεται με μια ευρύτερη ποικιλομορφία από σάκχαρα σύνδεσης και πολυσακχαρίτες από ο,τι σε καθεμία άλλη μορφή ζωής [19].

2.1.2.4 Η Γλυκοζυλίωση στα Βακτήρια

Σήμερα, είναι σαφές ότι και οι δύο μορφές της N-γλυκοζυλίωσης και της O-γλυκοζυλίωσης, που θεωρήθηκε αρχικά ότι περιορίζονται στους ευκαρυώτες, συναντώνται επίσης στα αρχαία και τα βακτήρια. Στην πραγματικότητα, η γλυκοζυλίωση στους προκαρυώτες εμφανίζεται να οδηγεί σε μια μεγαλύτερη ποικιλομορφία συνθέσεων και δομών γλυκιδίων από αυτές που υπάρχουν στα ευκαρυωτικά κύτταρα. Αυτήν την περίοδο, η προσοχή στρέφεται στην πρωτεϊνική γλυκοζυλίωση στα βακτήρια, πρώτιστα λόγω της αυξανόμενης συχνότητας με την οποία αυτή η μετά-μεταφραστική τροποποίηση “βλέπει” τα παθογόνα είδη [13]. Συνεπώς, η έρευνα που πραγματοποιήθηκε στα διάφορα εργαστήρια κατά τη διάρκεια της προηγούμενης δεκαετίας έχει χρησιμεύσει στο να αποκρυπτογραφήσει τη διάβαση της N-γλυκοζυλίωσης στο εντερικό παθογόνο *Campylobacter jejuni*, κάνοντάς το έτσι το πρώτο βακτήριο για το οποίο έχει περιγραφεί μια τέτοια πλήρης διάβαση [14]. Αντίθετα η O-γλυκοζυλίωση έχει αποδειχθεί πως εμφανίζεται μέσα σε διάφορα άλλα βακτηριακά παθογόνα, όπως τα *Neisseria gonorrhoeae* και *Helicobacter pylori* [13]. Υπό αυτήν τη μορφή, οι διαβάσεις της πρωτεϊνικής γλυκοζυλίωσης μέσα σε αυτά τα παθογόνα, ειδικά για αυτά που είναι υπεύθυνα για τη βιοσύνθεση σπανίων γλυκοπρωτεϊνικών σακχάρων που δεν υπάρχουν στους ανθρώπους, μπορούν να προσφέρουν στους νέους θεραπευτικούς στόχους.

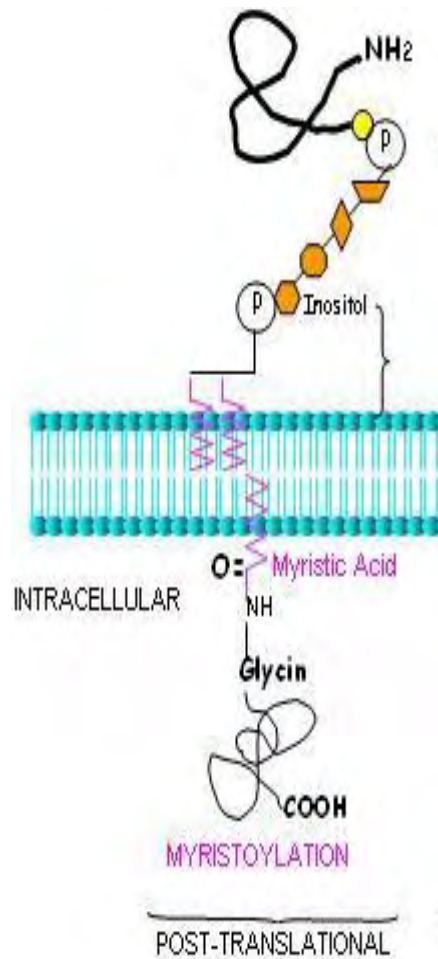
Τα βακτήρια είναι λοιπόν σε θέση να περιλαμβάνουν και τα δύο είδη N και O γλυκιδίων για την πρωτεϊνική σύνδεση. Τέτοια γλυκίδια μπορούν να περιλάβουν

σπάνιους μονοσακχαρίτες, που χρησιμεύουν συχνά ως ομάδες σύνδεσης υδατανθράκων-πεπτιδίων, σε αντίθεση με τα πιο συχνά εμφανιζόμενα σάκχαρα στις ευκαρυωτικές γλυκοπρωτεΐνες, όπως είναι η γλυκόζη, η γαλακτόζη, η N-ακετυλογαλακτοσαμίνη και η ξυλόζη.

2.1.3 Μυριστοϋλίωση

Η N-αμινοτελική N-μυριστοϋλίωση είναι μία μετά-μεταφραστική τροποποίηση λιπιδικής φύσης που εμφανίζεται σε ευκαρυώτες και πρωτεΐνες που προέρχονται από ιούς και στοχεύει τους οργανισμούς σε μεμβρανικές τοποθεσίες, ενώ αλλάζει την κυτταρική λειτουργία των τροποποιημένων πρωτεϊνών. Παίζει σημαντικό ρόλο στη σηματοδοτική μετατροπή της απόπτωσης καθώς και στην εξοκυτταρική εξαγωγή των πρωτεϊνών. Η myristoyl-CoA πρωτεΐνη N-μυριστοτρανσφεράση (NMT) αναγνωρίζει το μοτίβο ακολουθίας των κατάλληλων υποστρωμάτων μιας πρωτεΐνης στα N-αμινοτελικά άκρα και επιτίθεται στο λιπιδικό τμήμα του καταλοίπου της γλυκίνης [21]. Οι N-μυριστοπρωτεΐνες έχουν ξεχωριστές λειτουργίες και ενδοκυτταρικούς προορισμούς. Περιλαμβάνουν πρωτεΐνες που σχετίζονται με μία ευρεία ποικιλία διαδοχικών σημάτων γενετικών μεταλλαγών. Η N-μυριστοϋλίωση προάγει αδύναμες και αντιστρεπτές αλληλεπιδράσεις ανάμεσα σε πρωτεΐνες-μεμβράνες και πρωτεΐνες-πρωτεΐνες. Τυπικά, το μυριστικό οξύ δρα σε συνεργασία με άλλους μηχανισμούς ώστε να ελέγχει τη στόχευση και τη λειτουργία των πρωτεϊνών.

Μερικές πρωτεΐνες χρησιμοποιούν “μυριστοηλεκτροστατικούς διακόπτες” όπου η μεμβρανική σύνδεση προωθείται από μυριστικό οξύ και ηλεκτροστατικές αλληλεπιδράσεις ανάμεσα στις θετικά φορτισμένες συμπληρωματικές αλυσίδες των πρωτεϊνών και στα αρνητικά φορτισμένα μεμβρανικά φωσφολιπίδια.



Εικόνα 2-4. Η Μυριστοϋλίωση ως μία μετά-μεταφραστική τροποποίηση. Η myristoyl-CoA πρωτεΐνη N-μυριστοτρανσφεράση (NMT) αναγνωρίζει το μοτίβο ακολουθίας των κατάλληλων υποστρωμάτων μιας πρωτεΐνης στα N-τελικά και επιτίθεται στο λιπιδικό τμήμα του καταλοίπου γλυκίνης. Στη μυριστοϋλίωση σημαντική είναι η δράση του μυριστικού οξέος.

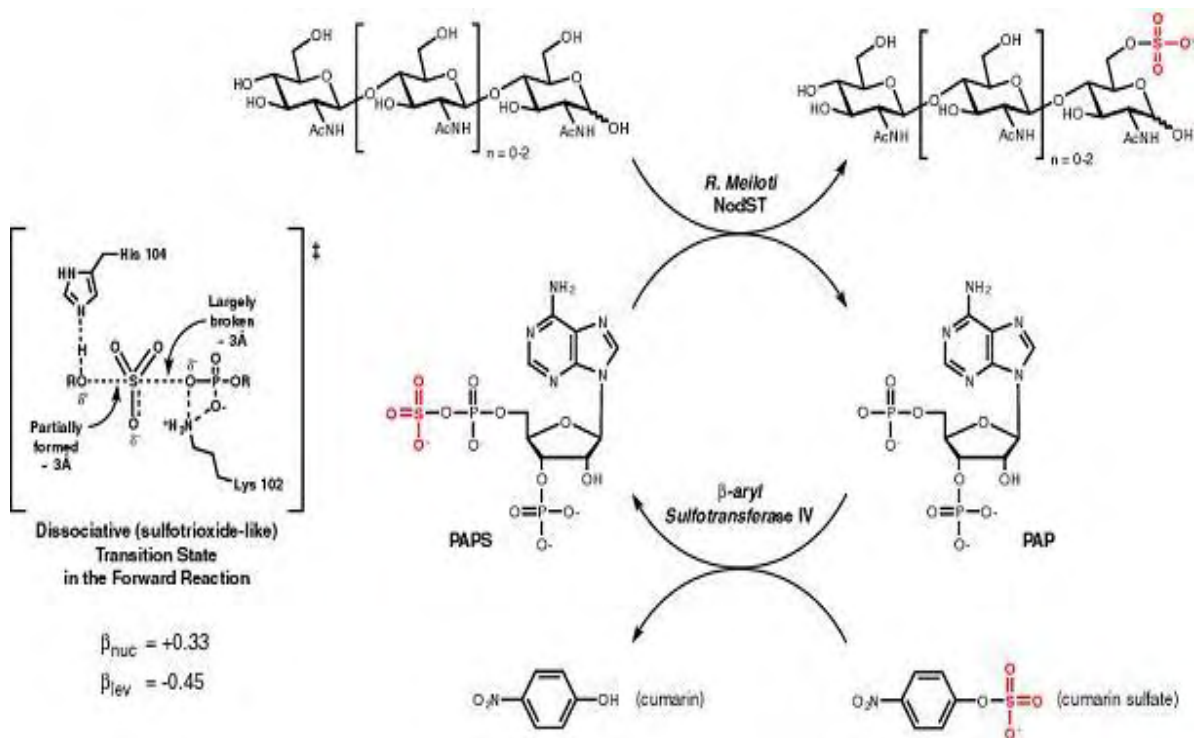
2.1.4 Σουλφίδωση

Η σουλφίδωση της τυροσίνης είναι μία μετά-μεταφραστική τροποποίηση που συμβαίνει σχεδόν αποκλειστικά σε κρυφές και διαμεμβρανικές πρωτεΐνες. Έχει αποδειχτεί πως πάνω από το 1% όλων των καταλοίπων τυροσίνης από όλη την πρωτεΐνη σε έναν οργανισμό μπορεί να υποστεί σουλφίδωση [22]. Επιπλέον, η σουλφίδωση της τυροσίνης είναι η πιο συχνή PTM στα κατάλοιπα τυροσίνης. Εκτός από την O-σουλφίδωση υπάρχουν η N-σουλφίδωση και η S-σουλφίδωση [27]. Η πρωτεϊνική O-σουλφίδωση της τυροσίνης ανακαλύφθηκε αρχικά από τον Bettelheim πριν από 40 χρόνια σε ένα πεπτίδιο που προέρχεται από το ινωδοπεπτίδιο B του

μοσχαριού [23]. Δύο δεκαετίες αργότερα, στα μέσα του 1980, ο Huttner ανέφερε πολλές άλλες πρωτεΐνες απομονωμένες από διαφορετικά κυτταρικά περιγράμματα χρησιμοποιώντας δείγματα ιστού διαφορετικής προέλευσης, οι οποίες είχαν υποστεί σουλφίδωση τυροσίνης, δείχνοντας έτσι πως η σουλφίδωση της τυροσίνης δεν είναι μία μοναδική PTM, αλλά συναντάται συχνά [24].

Οι Lee και Huttner ερεύνησαν την σουλφίδωση της τυροσίνης με μεγαλύτερη ακρίβεια και βρήκαν πως η τροποποίηση εμφανίζεται στο δίκτυο Golgi και έχει ως ενδιάμεσο μία μεμβράνη που καθορίζει το όριο της πρωτεΐνης [25]. Η χαρακτηριστική σουλφοτρανσφεράση της τυροσίνης (TPST, EC 2.8.2.20) έδειξε πως καταλύει τη μεταφορά του θεικού οξέος από τον γενικό δότη θεικού οξέος 3'-φωσφοαδενοσίνη-5'-φωσφοθειϊκό οξύ (PAPS) στην υδροξυλική ομάδα των καταλοίπων τυροσίνης που αποτελεί πρότυπο για τους εστέρες τυροσίνης Ο-θεικού οξέος και 3',5'-ADP. Η TPST καθορίστηκε από το επινεφρίδιο του μοσχαριού και τον προμήκη μυελό και χαρακτηρίστηκε το 1990 από τους Niehrs και Huttner [26].

Μέχρι τώρα έχουν αναγνωριστεί δύο διαφορετικές πρωτεΐνες σουλφοτρανσφεράσες της τυροσίνης, η TPST1 και η TPST2 [28, 29]. Η ανάλυση ακολουθιών έδειξε πως τα δύο αυτά ένζυμα είναι παρόμοια και πως περιλαμβάνουν μία περιοχή υψηλής διατήρησης σουλφοτρανσφεράσης [30]. Οι TPSTs βρέθηκαν επίσης σε πολλά σπονδυλωτά και ασπόνδυλα είδη. Ενδιαφέρον προκαλεί το γεγονός ότι η *Drosophila melanogaster* δεν έχει δεύτερο γονίδιο TPST τη στιγμή που σχεδόν όλα τα άλλα είδη έχουν δύο γονίδια TPST [31]. Έρευνες που έγιναν *in vitro* χρησιμοποίησαν συνθετικούς πεπτιδικούς υποδοχείς και υποδηλώνουν πως οι δύο TPST πρωτεΐνες παρουσιάζουν διαφορές στην επιλογή του υποστρώματος. Πεπτίδια που αναπαράγονται στο N-αμινοτελικό άκρο της II-επιλεκτικής γλυκοπρωτεΐνης ligand-1 σουλφιδώνονται από τα δύο συνένζυμα με αντίστοιχη αποδοτικότητα. Αντίθετα, πεπτίδια που αναπαράγονται σε θέσεις σουλφίδωσης στην ανθρώπινη αλυσίδα C4 και στο συνένζυμο της ηπαρίνης II σουλφιδώνονται πιο αποδοτικά από την TPST1 από ότι με την TPST2. Όλοι οι κυτταρικοί τύποι και τα κυτταρικά περιγράμματα των θηλαστικών εκφράζουν και τα δύο είδη TPST. Η σχετική αφθονία και η διανομή των δύο αυτών συνενζύμων δεν έχει μελετηθεί ακόμη εξαιτίας της έλλειψης των κατάλληλων αναλυτικών αντιδραστηρίων και εργαλείων.



Εικόνα 2-5. Σουλφίδωση της τυροσίνης. Η πρωτεΐνη σουλφοτρανσφεράση της τυροσίνης (TPST, EC 2.8.2.20) έδειξε πως καταλύει τη μεταφορά του θεικού οξέος από τον γενικό δότη θεικού οξέος 3'-φωσφοαδενοσίνη-5'-φωσφοθειικό οξύ (PAPS) στην υδροξυλική ομάδα των καταλοίπων τυροσίνης που αποτελεί πρότυπο για τους εστέρες τυροσίνης Ο-θεικού οξέος και 3',5'-ADP.

2.1.5 Παλμιτοϋλίωση

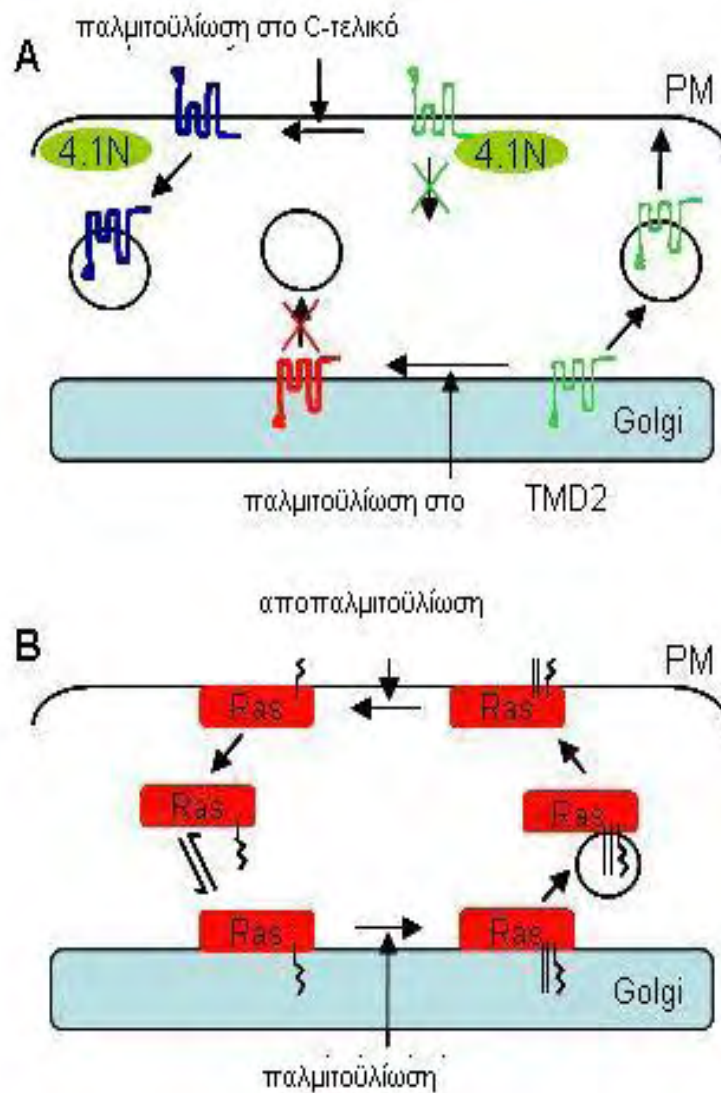
Η παλμιτοϋλίωση είναι μία σημαντική μετά-μεταφραστική τροποποίηση των λιπιδίων των πρωτεϊνών. Είναι μία ανιστρεπτή ομοιοπολική τροποποίηση που επιτρέπει την κυκλοφορία των πολλαπλών και σύνθετων κυτταρικών συστημάτων. Οι αναλύσεις των επιστημόνων στηρίζονται στην S-παλμιτοϋλίωση που αποτελεί ομοιοπολική επίθεση των λιπαρών οξέων, όπως είναι το παλμιτικό οξύ, στα κατάλοιπα κυστεΐνης των μεμβρανικών πρωτεϊνών. Η ακριβής λειτουργία της παλμιτοϋλίωσης στηρίζεται στη συγκεκριμένη πρωτεΐνη που ελέγχεται κάθε φορά. Η παλμιτοϋλίωση ενθαρρύνει την υδροφοβικότητα των πρωτεϊνών και συνεισφέρει στη μεμβρανική τους σύνδεση. Επίσης, εμφανίζεται να παίζει ένα σημαντικό ρόλο στην υποκυτταρική κυκλοφορία των πρωτεϊνών ανάμεσα στα διαμεμβρανικά τμήματα καθώς και στη διαμόρφωση των αλληλεπιδράσεων μεταξύ των πρωτεϊνών. Σε

αντίθεση με την πρενυλίωση (prenylation) και τη μυριστοϋλίωση (myristoylation), η παλμιτοϋλίωση είναι αντιστρεπτή. Αυτό επιτρέπει στο κύτταρο να καθορίζει δυναμικά τις τοποθεσίες συγκεκριμένων πρωτεϊνών.

Παράδειγμα πρωτεΐνης που υφίσταται παλμιτοϋλίωση είναι η αιμαγλουτινίνη (hemagglutinin), μία μεμβρανική γλυκοπρωτεΐνη που χρησιμοποιείται για πρόσδεση στους κυτταρικούς υποδοχείς. Επειδή η S-παλμιτοϋλίωση είναι δυναμική, η μετά-μεταφραστική διαδικασία, πιστεύεται πως χρησιμοποιείται από το κύτταρο για τη μεταβολή της υποκυτταρικής τοποθεσίας, τις αλληλεπιδράσεις μεταξύ των πρωτεϊνών ή για τις δεσμικές ικανότητες μιας πρωτεΐνης.

Η παλμιτοϋλίωση των διαμεμβρανικών πρωτεϊνών εμφανίζεται στα κατάλοιπα κυστεΐνης που βρίσκονται στην περιοχή του ορίου μεταξύ της διαμεμβρανικής περιοχής (TMR) και της κυτοπλασμικής περιοχής.

Με τη χρήση τεχνολογίας ανασυνδιασμένου DNA μπορούμε να αποδείξουμε ότι και οι δύο αυτές περιοχές και κυρίως τα κατάλοιπα γλυκίνης διαμορφώνουν την αποδοτικότητα της παλμιτοϋλίωσης. Όμοια με άλλες τροποποιήσεις λιπιδιακής φύσεως, η πρωτεϊνική παλμιτοϋλίωση χαρακτηρίζεται ως ενζυμική αντίδραση η οποία επιτυγχάνεται με τη μεσολάβηση μίας πρωτεΐνης γνωστής ως παλμιτοτρανσφεράση (PAT) ή με μία ακυλοτρανσφεράση.

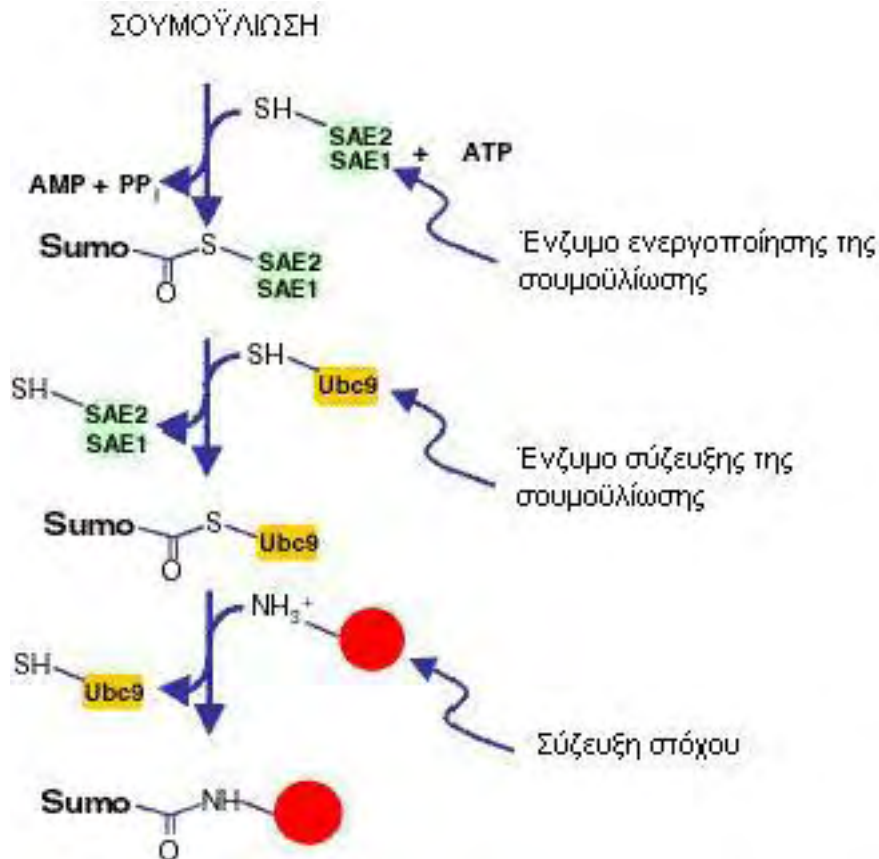


Εικόνα 2-6. Παλμιτοϋλίωση και Αποπαλμιτοϋλίωση.

2.1.6 Σουμοϋλίωση

Η σουμοϋλίωση αποτελεί μία αντιστρεπτή μετά-μεταφραστική τροποποίηση των πρωτεϊνών από μικρούς και ευρέως διαδεδομένους τροποποιητές (SUMOs) και είναι απαραίτητη για μία ποικιλία βιολογικών διαδικασιών, όπως η μεταγραφή [33, 34], ο μεταβολισμός του mRNA [35], η μεταγωγή σήματος [36] και πιθανόν στην αντίληψη του ήχου [37]. Έχει επίσης αναφερθεί πως η σουμοϋλίωση των πρωτεϊνών παίζει απαραίτητο ρόλο σε αρκετές ασθένειες και διαταραχές όπως είναι ο διαβήτης τύπου 1 (T1D) [38] και στη νόσο του Πάρκινσον (PD) [39]. Οι SUMO πρωτεΐνες υπάρχουν σε μεγάλη κλίμακα στους ευκαρυώτες, ενώ στα θηλαστικά αποτελούνται από τέσσερις συνιστώσες, τις SUMO-1, SUMO-2, SUMO-3 και SUMO-4 [40]. Στους

αναπτυσσόμενους ζυμομύκητες υπάρχει μόνο ένα γονίδιο σουμοϋλίωσης το SMT3, ενώ στα φυτά υπάρχουν τουλάχιστον οχτώ αντίστοιχα [41]. Η σουμοϋλίωση είναι ένα ασυνήθιστο φαινόμενο με αρκετά ευδιάκριτα χαρακτηριστικά.

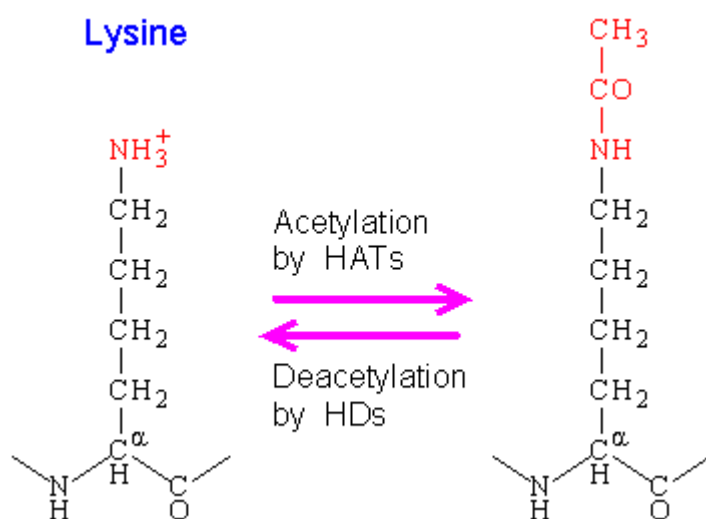


Εικόνα 2-7. Σουμοϋλίωση των πρωτεϊνών. Σε αρχικό στάδιο συμμετέχει το ένζυμο ενεργοποίησης της σουμοϋλίωσης ώστε να ξεκινήσει η διαδικασία. Στη συνέχεια λαμβάνει δράση το ένζυμο σύζευξης της σουμοϋλίωσης, καθορίζεται η σύζευξη του στόχου και επιτυγχάνεται η μετά-μεταφραστική τροποποίηση.

2.1.7 Ακετυλίωση

Η N-αμινοτελική ακετυλίωση είναι μια από τις πιο κοινές μετά-μεταφραστικές τροποποιήσεις που συναντώνται στους ευκαρυώτες, ενώ εμφανίζεται επίσης στα αρχαία και στα βακτήρια με μικρότερη συχνότητα [71]. Η ακετυλίωση των καταλοίπων λυσίνης στα N-αμινοτελικά των πρωτεϊνών μειώνει τους δεσμούς μεταξύ των ιστονών και του DNA. Το γεγονός αυτό διευκολύνει την πρόσβαση της

RNA πολυμεράσης και των μεταγραφικών παραγόντων στην περιοχή πρόσδεσης. Η μετά-μεταφραστική αυτή τροποποίηση εμφανίζεται κυρίως στις ευκαρυωτικές κυτταροπλασματικές πρωτεΐνες και στις περισσότερες περιπτώσεις ενισχύει τη διαδικασία της μεταγραφής ενώ η αντίστροφη της διαδικασία καταστέλλει τη μεταγραφή. Το ποσοστό της μετά-μεταφραστικής αυτής τροποποίησης υπολογίζεται μεταξύ 80% και 90% στα θηλαστικά και 50% στους ζυμομύκητες [86]. Η ιστονική ακετυλίωση καταλύεται από ακετυλοτρανσφεράσες των ιστονών (histone acetyltransferases-HATs), ενώ η ιστονική αποακετυλίωση καταλύεται από αποακετυλάσες των ιστονών (histone deacetylases-HDs ή HDACs). Από όλες τις Hats και HDs που έχουν αναγνωριστεί, η CBP/p300 θεωρείται πως είναι η πιο σημαντική καθώς μπορεί να αντιδρά με μεγάλο αριθμό μεταγραφικών παραγόντων.



Εικόνα 2-8. Ακετυλίωση (Acetylation) και Αποακετυλίωση (Deacetylation) των καταλοίπων λυσίνης.

2.1.8 Γλυκίωση

Η γλυκίωση (glycation) είναι μια μη ενζυμική διαδικασία στην οποία οι πρωτεΐνες αντιδρούν με τα μόρια μειώνοντας τα επίπεδα των σακχάρων. Με τον τρόπο αυτό εξασθενίζουν τη λειτουργία των πρωτεϊνών και αλλάζουν τα χαρακτηριστικά τους. Η συσσώρευση των παραγώγων της γλυκίωσης προκαλεί

παρενέργειες καθώς συμμετέχει στο διαβήτη και στη διαδικασία της γήρανσης. Τα όξινα αμινοξέα, κυρίως το γλουταμικό οξύ και τα κατάλοιπα της λυσίνης, καταλύουν τη γλυκίωση των γειτονικών λυσινών. Τα καταλυτικά όξινα αμινοξέα βρίσκονται κυρίως στο C-αμινοτελικό άκρο από την περιοχή της γλυκίωσης, ενώ τα βασικά κατάλοιπα της λυσίνης βρίσκονται κυρίως στο N-αμινοτελικό άκρο. Η αντίδραση εμφανίζεται μεταξύ της α αμινομάδας του N-αμινοτελικού αμινοξέος ή στις ε αμινομάδες των λυσινών και των αλδεύδων ή των κετονών από τα ανοιγμένα σάκχαρα [91]. Το πρώτο βήμα μιας αντίδρασης γλυκίωσης είναι ένας σχηματισμός από μια ασταθή βάση Schiff, η οποία ρυθμίζεται εκ νέου για να διαμορφώσει ένα περισσότερο σταθερό προϊόν Amadori. Έπειτα το προϊόν Amadori μπορεί αντιδράσει περαιτέρω για να δημιουργήσει μία μορφή από αμετάκλητα διασυνδεδεμένα προϊόντα, αποκαλούμενα προηγμένα τελικά προϊόντα της γλυκίωσης (Advanced Glycation End Products-AGEs).

Μέρος 2^ο

2.2 Τεχνικές πρόγνωσης μετά-μεταφραστικών τροποποιήσεων

Για την πρόγνωση των μετά-μεταφραστικών τροποποιήσεων των πρωτεϊνών έχουν αναπτυχθεί διάφορες μέθοδοι στηριζόμενες σε εφαρμογές Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks-ANNs), ελέγχου προτύπου (pattern search-PS) και Κρυφών Μαρκοβιανών Μοντέλων (Hidden Markov Models-HMMs). Στο 2^ο μέρος αυτού του κεφαλαίου γίνεται μία ανασκόπηση των μεθόδων αυτών καθώς και των διαθέσιμων αλγορίθμων για την υλοποίησή τους.

2.2.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANNs)

Τα Τεχνητά Νευρωνικά Δίκτυα είναι ένα μαθηματικό μοντέλο για την επεξεργασία πληροφορίας που προσεγγίζει την υπολογιστική και αναπαραστατική δυνατότητα μέσω συνάψεων. Το μοντέλο είναι εμπνευσμένο από τα βιοηλεκτρικά δίκτυα που δημιουργούνται στον εγκέφαλο ανάμεσα στους νευρώνες και στις συνάψεις (σημεία επαφής των νευρικών απολήξεων). Στο μαθηματικό μοντέλο των νευρωνικών δικτύων υπάρχουν κομβικά σημεία (nodes) στα οποία καταλήγουν συνδέσεις από άλλους κόμβους του δικτύου στις οποίες συνήθως αποδίδεται κάποιο βάρος. Πρακτικά ένα νευρωνικό δίκτυο βελτιστοποιεί μία συνάρτηση, σύμφωνα με

κάποιους περιορισμούς [101]. Οι νευρώνες είναι τα δομικά στοιχεία του δικτύου. Υπάρχουν δύο είδη νευρώνων, οι νευρώνες εισόδου και οι υπολογιστικοί νευρώνες. Οι νευρώνες εισόδου δεν υπολογίζουν τίποτα, μεσολαβούν ανάμεσα στις εισόδους του δικτύου και τους υπολογιστικούς νευρώνες. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν τις εισόδους τους με τα συνοπτικά βάρη και υπολογίζουν το άθροισμα του γινομένου. Το άθροισμα που προκύπτει είναι το όρισμα της συνάρτησης ενεργοποίησης. Επινοήθηκαν το 1943 από τον Αμερικανό νευροφυσιολόγο Γουόρεν ΜακΚούλοκ και το μαθηματικό Γουόλτερ Πιτς. Σήμερα βρίσκονται σε φάση ταχείας ανάπτυξης. Τα περιβάλλοντα στα οποία χρησιμοποιούνται ευρέως είναι αυτά της διαγνωστικής ιατρικής, της πρόληψης των φυσικών καταστροφών και της πρόβλεψης των εξελίξεων στο χρηματιστήριο, τομείς στους οποίους αποδείχθηκαν αρκετά αποτελεσματικά.

Ένα νευρωνικό δίκτυο δεν αναγνωρίζει γράμματα, γι' αυτό η ακολουθία του αμινοξέως και τα διάφορα χαρακτηριστικά πρέπει να μεταφραστούν σε αριθμούς. Αυτό ονομάζεται κωδικοποίηση και μπορεί να πραγματοποιηθεί με αρκετούς τρόπους. Κάθε αριθμός που παρουσιάζεται στο νευρωνικό δίκτυο αποτελεί αυτό που ονομάζομαι νευρώνα εισαγωγής. Σκοπός είναι να παρέχουμε στο δίκτυο όσο το δυνατόν περισσότερες πληροφορίες ενώ συγχρόνως να κρατάμε τον αριθμό των νευρών εισαγωγής όσο πιο χαμηλά γίνεται.

Είδη κωδικοποίησης που συναντώνται:

- Sparse (Αραιή) κωδικοποίηση [69] είναι ο συμβατικός τρόπος να μετατραπεί η ακολουθία αμινοξέως σε αριθμητική μορφή.
- Με την profile encoding (κωδικοποίηση σχεδιαγράμματος) η εισαγωγή για κάθε αμινοξύ αποτελείται από την αντίστοιχη σειρά στον πίνακα BLOSUM62 [70].
- Με την κωδικοποίηση PSI-BLAST η εισαγωγή για κάθε αμινοξύ αποτελείται από την αντίστοιχη σειρά στον πίνακα ακριβής θέσης που υπολογίζεται από τρεις κύκλους PSI-BLAST [64].
- Η κωδικοποίηση 5-letter alphabet (αλφάβητου 5 επιστολών) ήταν συμβατική αραιή κωδικοποίηση, αλλά με ένα μειωμένο αλφάβητο.

- Το 8-letter alphabet (αλφάβητο 8 επιστολών) είναι ένα άλλο μειωμένο αλφάβητο.
- Η σύνθεση αμινοξέος υπολογίστηκε για μια ακολουθία παραθύρου γύρω από κάθε ιδιαίτερη περιοχή.
- Η δυνατότητα πρόσβασης επιφάνειας προβλέφθηκε χρησιμοποιώντας ένα νευρωνικό δίκτυο με μία μέθοδο αποκαλούμενη surfg [68].
- Η δευτεροβάθμια δομή προβλέφθηκε χρησιμοποιώντας PSIPRED [72] με χρήση πινάκων ακριβής θέσης που υπολογίζονται από τρεις κύκλους PSIBLAST [64].
- Οι πρωτεϊνικοί περιορισμοί απόστασης προβλέφθηκαν χρησιμοποιώντας DistanceP [73].

2.2.2 Έλεγχος προτύπου (Pattern Search-PS)

Η πιο απλή προσέγγιση για την αναγνώριση των θέσεων των μετά-μεταφραστικών τροποποιήσεων βασίζεται στην εφαρμογή της κανονικής έκφρασης “έλεγχος προτύπου” (pattern search) [102]. Για τον έλεγχο αυτό οι επιστήμονες ανέπτυξαν το position-specific_scoring matrix (PSSM), που αποτελεί έναν μαθηματικό πίνακα των θέσεων σκοραρίσματος, για την πρόγνωση των θέσεων των μετά-μεταφραστικών τροποποιήσεων [77, 79, 80]. Ενώ η ικανότητα να ταξινομηθούν και να ξεχωριστούν οι τροποποιημένες ακολουθίες από τις μη τροποποιημένες εμπλουτίζεται με μεταφορά από τις απλές εκφράσεις στο PSSM, η ικανότητα να βρεθούν νέα ή αθέατα πρότυπα μετά-μεταφραστικών τροποποιήσεων υποβαθμίζεται εξαιτίας της στατικής συμπεριφοράς των PSSMs. Οι καταχωρήσεις και οι διαγραφές δημιουργούν ένα άλυτο πρόβλημα για αυτό το είδος προσέγγισης, επειδή οι μέθοδοι δεν είναι αρκετά δυναμικές ώστε να έχουν πρόσβαση σε διαγραφές και καταχωρήσεις.

2.2.3 Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models-HMMs)

2.2.3.1 Θεωρία

Τα Hidden Markov Models (HMMs) μελετήθηκαν από τον Markov και ήταν από τα πρώτα Μαρκοβιανά Μοντέλα που εφαρμόστηκαν με μεγάλη επιτυχία για την επίλυση βιολογικών προβλημάτων. Ιδιαίτερα δημοφιλή έγιναν τη δεκαετία του '70 όταν ο Baum τα εφάρμοσε επιτυχώς για αναγνώριση της ομιλίας με την ανάπτυξη ενός αποδοτικού αλγορίθμου εκπαίδευσης που βασίζονταν σε αυτά.

Η βασική ιδέα είναι ότι ένα HMM είναι μία γεννήτρια ακολουθίας. Κάθε HMM συνδέεται με μια πρώτης τάξεως Μαρκοβιανή Αλυσίδα (MC) με κατάλληλες πιθανότητες μετάβασης μεταξύ των καταστάσεων και με μία αρχική κατανομή [103]. Οι πιθανότητες εκπομπής εξαρτώνται μόνο από την παρούσα κατάσταση της MC, ανεξάρτητα από τις προηγούμενες καταστάσεις. Ξεκινώντας από μία αρχική κατάσταση με μία αρχική πιθανότητα δημιουργείται μία ακολουθία συμβόλων, καθώς πηγαίνοντας από τη μία κατάσταση στην άλλη σύμφωνα με τις πιθανότητες μετάβασης, παράγεται μία ακολουθία καταστάσεων. Η ακολουθία των συμβόλων είναι ορατή σε μας, ενώ οι καταστάσεις από τις οποίες περνά το μοντέλο όχι.

2.2.3.2 Ορισμός του μοντέλου

Ένα Hidden Markov Model, είναι ένα μοντέλο \mathbf{M} που περιέχει 3 στοιχεία Σ , \mathbf{Q} , θ .

$$\mathbf{M}=(\Sigma, \mathbf{Q}, \theta)$$

- Σ , το αλφάβητο των δυνατών ενδεχομένων
- \mathbf{Q} , το σύνολο των δυνατών καταστάσεων του μοντέλου (γονίδιο-όχι γονίδιο για το DNA, κλπ)
- θ , το σύνολο πιθανοτήτων που διέπουν το μοντέλο και μπορεί να είναι:

1. Πιθανότητες μεταβάσεως (transitions) από κατάσταση σε κατάσταση και
2. Πιθανότητες γεννήσεως (emissions) με τις οποίες παράγονται τα σύμβολα σε κάθε κατάσταση.

Πρέπει να τονιστεί ότι σε ένα HMM η μαρκοβιανή ιδιότητα ισχύει για τις καταστάσεις του μοντέλου (states) και όχι για τα σύμβολα.

2.2.3.3 Εφαρμογές των HMMs στη Βιολογία

Τα HMMs έχουν ευρεία εφαρμογή σε διαφορετικούς και ποικίλους τομείς της βιολογίας. Είναι απαραίτητα για την εύρεση πιθανών γονιδίων όπως είναι η ύπαρξη εσωνίων και εξωνίων, ενώ χρησιμοποιούνται και κατά την εύρεση ομόλογων οικογενειών ακολουθιών. Βασικός και απαραίτητος επίσης, είναι ο ρόλος τους κατά την πρόγνωση της δευτεροταγούς δομής των πρωτεϊνών, των διαμεμβρανικών τμημάτων των πρωτεϊνών καθώς και των πεπτιδίων οδηγητών.

Παρόλο που τα HMMs χρησιμοποιούνται ευρέως στη βιολογία τα τελευταία χρόνια και αποτελούν λύση για ένα μεγάλο φάσμα βιολογικών προβλημάτων με μεγάλα ποσοστά επιτυχίας, παρουσιάζουν ένα σημαντικό μειονέκτημα καθώς είναι ιδιαίτερα χρονοβόρα στη δημιουργία τους.

2.3 Μέθοδοι πρόγνωσης μετά-μεταφραστικών τροποποιήσεων

2.3.1 Πρόγνωση φωσφορυλίωσης

Οι εξελίξεις στη φασματογραφία μάζας (biological mass spectrometry) έδωσαν τη δυνατότητα αναγνώρισης εκατοντάδων θέσεων φωσφορυλίωσης των πρωτεϊνών [56]. Οι άνθρωποι έχουν κινάσες πρωτεϊνών που καθορίζουν την εξέλιξη, τη διαφορετικότητα και τη διακυτταρική κυκλοφορία [59] κάτι το οποίο δεν συναντάται στους ζυμομύκητες.

Για την πρόγνωση της φωσφορυλίωσης των πρωτεϊνών είναι διαθέσιμοι αρκετοί αλγόριθμοι. Σε αυτούς συμπεριλαμβάνονται οι Scansite 2.0 [48], Prosite [49], Netphos [50], Netphosk [51], GPS [52], Disphos [53], KinasePhos [54], PPSP [55].

Πίνακας 2-1. Ηλεκτρονικές διευθύνσεις των αλγορίθμων πρόγνωσης της φωσφορυλίωσης των πρωτεϊνών.

Scansite 2.0	http://scansite.mit.edu/
Prosite	http://www.expasy.org/prosite/
Netphos	http://www.cbs.dtu.dk/services/NetPhos/
NetphosK	http://www.cbs.dtu.dk/services/NetPhosK/
GPS	http://973-proteinweb.ustc.edu.cn/gps/gps_web/
Disphos	http://core.ist.temple.edu/pred/pred.html
KinasePhos	http://kinasephos.mbc.nctu.edu.tw/
PPSP	http://bioinformatics.lcd-ustc.org/PPSP/

Η πρόγνωση της φωσφορυλίωσης στηρίζεται στην εφαρμογή της τεχνικής των νευρωνικών δικτύων (Artificial Neural Networks-ANNs). Εργαστηριακές έρευνες που πραγματοποιήθηκαν σύγκριναν τους υπάρχοντες αλγόριθμους (KinasePhos, NetphosK, NetPhos, Scansite 2.0, Average και NetPhos Yeast) ως προς την ευαισθησία και την ακρίβειά τους. Ο NetPhos Yeast παρουσιάζει έναν υψηλό συντελεστή συσχέτισης MCC (Matthews correlation coefficient) της τάξης του 0.74, ευαισθησία τάξεως 0.84 και ακρίβεια της τάξης του 0.90 χρησιμοποιώντας κατώφλι με τιμή 0.5. Επίσης τα αποτελέσματα έδειξαν ότι οι μέθοδοι NetPhos, NetPhosK, KinasePhos και Scansite 2.0 μπορούν να εντοπίσουν το 82% των θέσεων φωσφορυλίωσης για τον άνθρωπο από τα δεδομένα που εξήχθησαν από τη βάση δεδομένων Swiss-Prot, κάτι το οποίο είναι συγκρίσιμο με τα αποτελέσματα του αλγορίθμου για τους ζυμομήκυτες NetPhos Yeast.

Enter the regular expression: "S/T/T" [Examples](#)

Select Database to Search: SWISS-PROT ▾

Only want human proteins? Or proteins under 50 kDa? You can use the following options to place additional restrictions on the proteins you want. Otherwise, just leave them blank. Alternatively, you can leave the regular expression field empty and search for matches to the criteria below instead.

Organism Class: All ▾ [Class descriptions](#)

Single species: [Examples](#)

Molecular weight range: to [Notes](#)

Isoelectric point range: to [Notes](#)

Phosphorylated sites: 1 ▾ [Notes](#)

Keyword search: [Examples](#)

Εικόνα 2-9. Απεικόνιση του αλγορίθμου Scansite για την πρόγνωση των θέσεων φωσφορυλίωσης.

NetPhos 2.0 Server

The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins. Kinase specific phosphorylation predictions are available at: <http://www.cbs.dtu.dk/services/NetPhosKf/>

[Instructions](#) [Output format](#) [Ph](#)

SUBMISSION

Paste a single sequence or several sequences in **FASTA** format into the field below:

Submit a file in **FASTA** format directly from your local disk:

Predict on: tyrosine serine threonine

Generate graphics

Εικόνα 2-10. Απεικόνιση του αλγορίθμου Netphos για την πρόγνωση των θέσεων φωσφορυλίωσης.

NetPhosK 1.0 Server

The NetPhosK 1.0 server produces neural network predictions of **kinase specific** eukaryotic protein phosphorylation sites. Currently NetPhosK covers the following kinases: PKA, PKC, PKG, CKII, Cdc2, CaM-II, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CK1, PKB, RSK, INSR, EGFR and Src.

Generic (non kinase specific) phosphorylation predictions are available at: <http://www.cbs.dtu.dk/services/NetPhos/>

Instructions Output format Article abstract NetPhosPhosphoBase

SUBMISSION

Paste a single sequence or several sequences in **FASTA** format into the field below:

Submit a file in **FASTA** format directly from your local disk:

Method to use:

- Prediction without filtering (fast)
- Prediction with ESS (Evolutionary Stable Sites) Filter (very slow)
- Kinase Landscapes (Graphics)

Threshold 0.50

Submit Clear fields

Εικόνα 2-11. Απεικόνιση του αλγορίθμου Netphosk για την πρόγνωση των θέσεων φωσφορυλίωσης.

Submission

Paste a single sequence or several sequences in **FASTA** format into the field below:

```
>DYRA_MOUSE
MHTGGGETSACKPSSVRLAPSFSAAGLQMAAQMPHSHQYSDRRQPSISDQQVSALEPYSDOIQQPL
```

Submit a file (< 2MB) in **FASTA** format directly from your local disk:

Browse...

Predict on: Serine(S) Threonine(T) Tyrosine(Y)

Kinase : No Specific

with Prediction Specificity 100%

by default HMM bit score

with HMM bit score >

Submit Clear fields

Εικόνα 2-12. Απεικόνιση του αλγορίθμου Kinasephos για την πρόγνωση των θέσεων φωσφορυλίωσης.

DISPHOS 1.3

Disorder-Enhanced Phosphorylation Sites Predictor

Paste your FASTA formatted protein sequence

Or select a file:

Default Predictor

Group Predictors

Group predictors are trained to minimize the number of total misclassified residues for a particular group (see [Usage](#) for detailed instructions).

Kingdoms and Viruses

Eukaryotes

Organism

- H. sapiens*
- M. musculus*
- R. norvegicus*
- C. elegans*
- S. cerevisiae*
- D. melanogaster*
- A. thaliana*

Functional Category

- regulation
- cancer
- cytoskeleton
- membrane
- ribosomal
- inhibitors
- transport
- kinases
- degradation
- biosynthesis
- metabolism
- GPCRs

Archaea

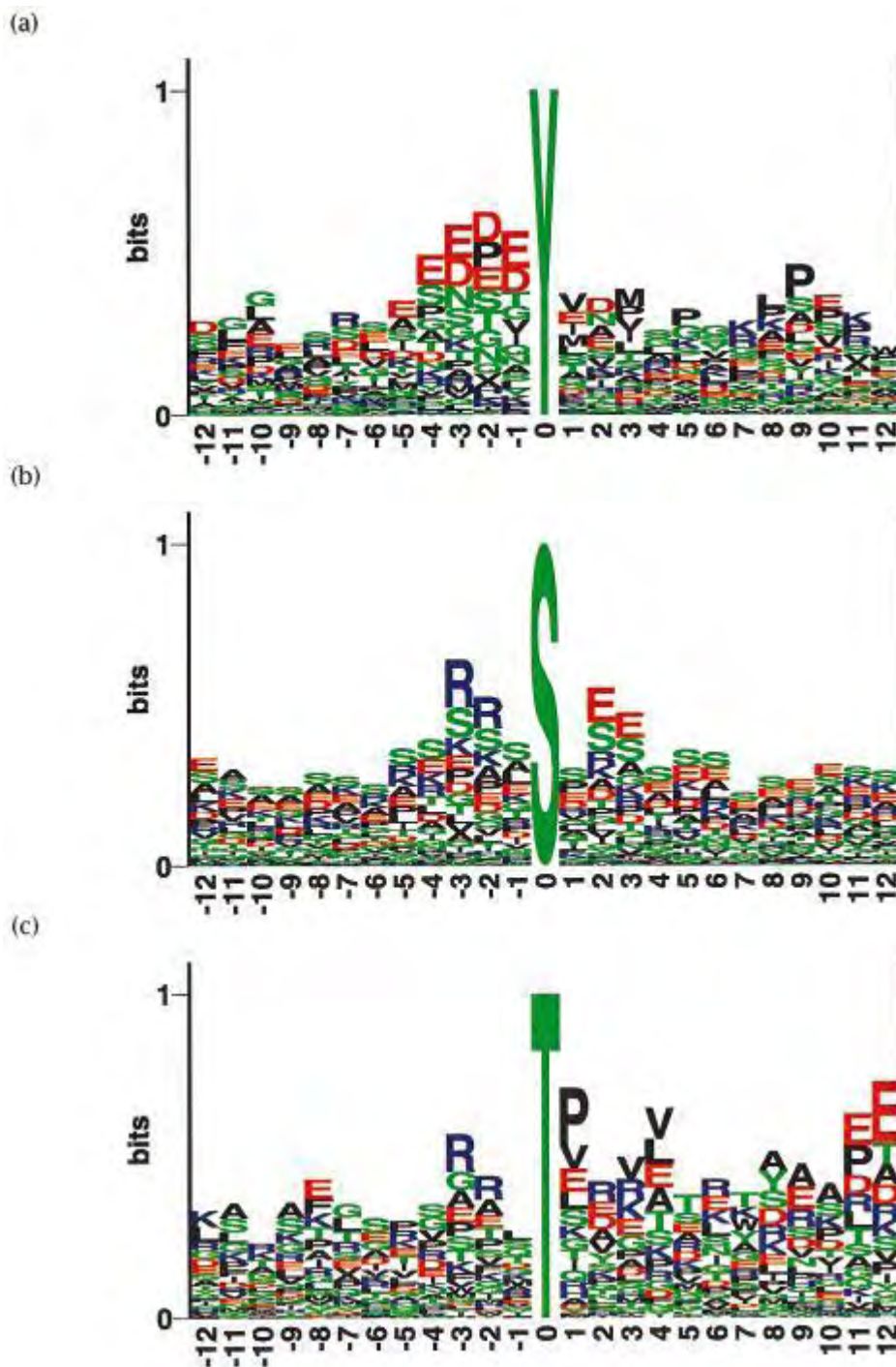
Bacteria

Viruses

Note: Due to available computational resources we can provide up to 100 predictions per IP address per day. The minimal length of the query sequence is limited to 25 residues; the maximal length is limited to 5,000 residues.

This site was developed and is maintained by Chad Haynes

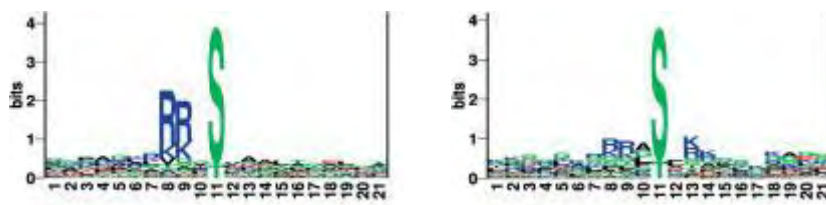
Εικόνα 2-13. Απεικόνιση του αλγορίθμου Disphos για την πρόγνωση των θέσεων φωσφορυλίωσης.



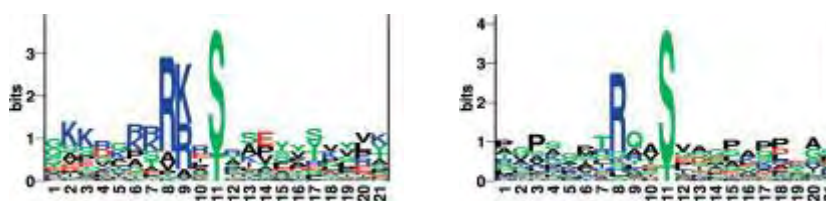
Εικόνα 2-14. Λογότυπα ακολουθίας για τις θέσεις φωσφορυλίωσης όπου, a) 210 θέσεις φωσφορυλίωσης τυροσίνης, b) 584 θέσεις φωσφορυλίωσης σερίνης και c) 108 θέσεις φωσφορυλίωσης θρεονίνης.

Η μέθοδος NetphosK, η οποία παρουσιάζεται αναλυτικά στο επόμενο κεφάλαιο, κάνει πρόγνωση της φωσφορυλίωσης αναγνωρίζοντας ειδικές κινάσες στόχους. Παρακάτω παρουσιάζονται τα λογότυπα για κάποιες από τις κινάσες αυτές όπως

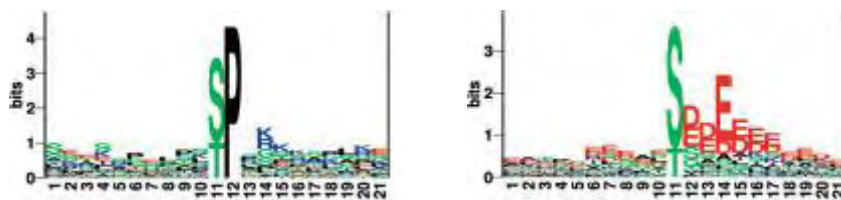
εξήχθησαν έπειτα από έρευνα που πραγματοποιήθηκε με χρήση της μεθόδου NetPhosK [108, 109].



Εικόνα 2-15. Λογότυπο ακολουθίας για την κινάση pka (αριστερά) και pkc (δεξιά).



Εικόνα 2-16. Λογότυπο ακολουθίας για την κινάση pkg (αριστερά) και CaM-II (δεξιά).



Εικόνα 2-17. Λογότυπο ακολουθίας για την κινάση cdc2 (αριστερά) και cki (δεξιά).

2.3.1.2 Πρόγνωση της φωσφορυλίωσης στους ζυμομήκυτες

Πρόσφατα, δύο εκτενής φωσφοπρωτεϊνικές έρευνες (phosphoproteomic studies) χαρτογράφησαν περισσότερες από 900 θέσεις φωσφορυλίωσης στους ζυμομήκυτες [57]. Παρόλο που αρκετές κινάσες έχουν ομόλογα στους ανθρώπους και αντιστρόφως, αρκετές από αυτές δεν χωρίζονται ανάμεσα σε αυτά τα είδη. Μία έρευνα πάνω στις κινάσες των πρωτεϊνών έδειξε ότι 32 κινάσες είναι μοναδικές στους ζυμομήκυτες και καλύπτουν λειτουργίες όπως τη σηματοδότηση του κυτταρικού τοιχώματος (cell wall signaling), τη ρύθμιση του κυτταρικού κύκλου (cell-cycle regulation) και τη μεταφορά των μικρομορίων (small molecule transport) [58].

Η μέθοδος NetPhosYeast, όπως περιγράφηκε από τους επιστήμονες, αναγνωρίζει το 84% των θέσεων φωσφορυλίωσης στους ζυμομύκητες, τη στιγμή που οι προαναφερθέντες αλγόριθμοι αναγνωρίζουν το 67% του συνόλου. Κατά τη μέθοδο αυτή παράγεται ένα θετικό σύνολο δεδομένων που περιλαμβάνει θέσεις σερίνης και θρεονίνης της φωσφορυλίωσης στους ζυμομύκητες. Περιλαμβάνει επίσης επεξεργασμένες θέσεις φωσφορυλίωσης που έχουν εξαχθεί από τη βάση Swiss-Prot, περιορίζοντας να μην περιλάβει τους τροποποιητές 'potential', 'probable', 'by similarity' ή 'autocatalysis' στον τομέα της περιγραφής. Για τη σύγκριση, οι επιστήμονες συνέλεξαν από τη βάση δεδομένων Swiss-Prot, 1696 ανθρώπινες θέσεις φωσφορυλίωσης σερίνης και θρεονίνης, μη περιλαμβάνοντας τις περιοχές με τροποποιητή 'potential', 'probable', 'by similarity' ή 'autocatalysis' στον τομέα της περιγραφής.

Στη μελέτη αυτή, χρησιμοποιήθηκαν Τεχνητά Νευρωνικά Δίκτυα (ANNs) τριών επιπέδων με πρόσθια τροφοδότηση [61]. Αντίθετα με τις προηγούμενες μεθόδους που προτάθηκαν για την πρόγνωση των θέσεων φωσφορυλίωσης, τα αμινοξέα κωδικοποιήθηκαν από τον πίνακα BLOSUM62 [60] για την επίτευξη μια γενικότερης φυσικοχημικής περιγραφής. Στον BLOSUM62 κάθε αμινοξύ αντιπροσωπεύεται με ένα αντίστοιχο διάνυσμα αριθμών που δείχνει την ποινική ρήτρα για την αντικατάσταση εκείνου του αμινοξέος με οποιοδήποτε άλλο από τα 19 άλλα αμινοξέα. Το μέγεθος των παραθύρων εισαγωγής στα Τεχνητά Νευρωνικά Δίκτυα (ANNs) για τις ακολουθίες και ο αριθμός των νευρώνων στο κρυμμένο στρώμα βελτιστοποιήθηκε στη συνέχεια σε κάθε ένα στη διαδικασία της cross-validation.

Οι ερευνητές έχουν καταλήξει στο συμπέρασμα ότι ο αλγόριθμος NetPhos Yeast είναι η βάση στην οποία θα στηρίζουν τη δουλειά τους σχετικά με την πρόγνωση της φωσφορυλίωσης των πρωτεϊνών. Καθώς όλο και περισσότερα δεδομένα θα είναι διαθέσιμα, η επόμενη γενιά πρόγνωσης φωσφορυλίωσης θα κινείται ανάμεσα στα είδη των κινάσεων με ακρίβεια.

Η μέθοδος NetPhosYeast είναι διαθέσιμη στην ηλεκτρονική διεύθυνση:

NetPhos Yeast	http://www.cbs.dtu.dk/services/NetPhosYeast/
---------------	---

NetPhosYeast 1.0 Server

NetPhosYeast 1.0 server predicts serine and threonine phosphorylation sites in yeast proteins. This service is closely related to [NetPhos](#) and [NetPhosK](#).



[Instructions](#) [Output format](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

Submit a file in [FASTA](#) format directly from your local disk:

Generate graphics

Εικόνα 2-18. Απεικόνιση του αλγορίθμου NetPhosYeast για την πρόγνωση των θέσεων φωσφορυλίωσης στους ζυμομήκυτες.

2.3.2 Πρόγνωση γλυκοζυλίωσης

2.3.2.1 Πρόγνωση της Ο-γλυκοζυλίωσης στα θηλαστικά

Επιστημονικές έρευνες παρουσίασαν μία μέθοδο πρόγνωσης της Ο-γλυκοζυλίωσης στα θηλαστικά. Συγκεκριμένα, από τη βάση O-GLYCBASE εξήχθησαν συνολικά 86 πρωτεΐνες θηλαστικών που ελέγχθησαν πειραματικά σε ζώντα οργανισμό για την εύρεση των θέσεων της O-GalNAc. Η βάση πρωτεϊνικών στοιχείων (Protein Data Bank-PDB), αναλύθηκε για δομικές πληροφορίες και από εκεί λήφθηκαν 12 γλυκοζυλιωμένες δομές. Χρησιμοποιήθηκε μια μέθοδος τεχνητών νευρωνικών δικτύων για την πρόβλεψη της θέσης στις περιοχές γλυκοζυλίωσης [62]. Το καλύτερο γενικό δίκτυο χρησιμοποιήθηκε ως είσοδος για τη σύνθεση αμινοξέος. Για να βελτιώσουν την πρόβλεψη στις απομονωμένες (ενιαίες) περιοχές, οι επιστήμονες έτρεξαν τα δίκτυα μόνο στις περιοχές αυτές. Η τελική μέθοδος που παρουσίασαν συνδυάζει τις προβλέψεις από το καλύτερο γενικό δίκτυο και το καλύτερο απομονωμένο δίκτυο περιοχών. Η μέθοδος αυτή πρόβλεψε σωστά τα γλυκοζυλιωμένα κατάλοιπα σε ποσοστό 76% και τα μη γλυκοζυλιωμένα σε ποσοστό 93%.

Χρησιμοποίησαν το πρόγραμμα GetStruct [63] με τις παραμέτρους προεπιλογής για να εξαγάγει δομικές πληροφορίες για τις περιοχές γλυκοζυλίωσης στα στοιχεία που συλλέχτηκαν από τη βάση δεδομένων PDB. Το GetStruct εκτέλεσε τις BLAST ευθυγραμμίσεις [64] από τις ακολουθίες στο σύνολο των στοιχείων μας έναντι των ακολουθιών της PDB. Εξετάστηκαν μόνο δομές με τουλάχιστον 90%

ποσοστό ταυτότητας ακολουθίας στις ακολουθίες εισαγωγής. Με μερικές αξιοσημείωτες εξαιρέσεις [65], μία σαφής σχέση της ακολουθίας αμινοξέων μεταξύ δύο πρωτεϊνών υπονοεί ότι έχουν παρόμοια δομή [66]. Επομένως, στα απαραίτητα επίπεδα ομοιότητας ακολουθίας (90% ή περισσότερο), οι δομές μπορούν να αναμένονται να είναι καλοί αντιπρόσωποι των δομών των γλυκοπρωτεϊνών. Ο αναφερόμενος εντοπισμός των περιοχών της Ο-γλυκοζυλίωσης είναι υποδειγμένος σχετικά με τη θέση τους στην ακολουθία εισαγωγής. Η πολλαπλή στοίχιση των ακολουθιών εκτελέστηκε με τη χρησιμοποίηση του αλγορίθμου CLUSTALW [67]. Η συντήρηση της ακολουθίας υπολογίστηκε σε ένα κατάλοιπο για τη βάση καταλοίπων. Η δυνατότητα πρόσβασης της επιφάνειας προβλέφθηκε χρησιμοποιώντας μία μέθοδο νευρωνικού δικτύου αποκαλούμενη Surfsg [68]. Η μέθοδος Surfsg δίνει το άμεσο αποτέλεσμα παραγωγής και ένα κανονικοποιημένο αποτέλεσμα. Και τα δύο είναι μεταξύ 0 και 1, με ένα αποτέλεσμα επάνω από 0.5 εάν το κατάλοιπο του αμινοξέος έχει προβλεφθεί για να θαφτεί και ένα αποτέλεσμα κάτω από 0.5 εάν προβλέπεται να είναι στην επιφάνεια που εκτίθεται. Τα κατάλοιπα σερίνης και θρεονίνης για το κανονικοποιημένο αποτέλεσμα είναι κάτω από 0.5 και θεωρείται πως βρίσκονται στην προβλεφθείσα επιφάνεια που εκτίθεται. Η μέθοδος NetOGlyc 3.1 μπορεί να προβλέψει τις θέσεις για τις νέες πρωτεΐνες χωρίς απώλεια απόδοσης. Η μέθοδος NetOGlyc 3.1 παρέχεται στην ηλεκτρονική διεύθυνση:

NetOGlyc 3.1	www.cbs.dtu.dk/services/netoglyc
--------------	--

NetOGlyc 3.1 Server

The NetOglyc server produces neural network predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins. This version replaces NetOglyc 2.0 (click [here](#) for that version).

View the [version history](#) of this server.

[Instructions](#) [Output format](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

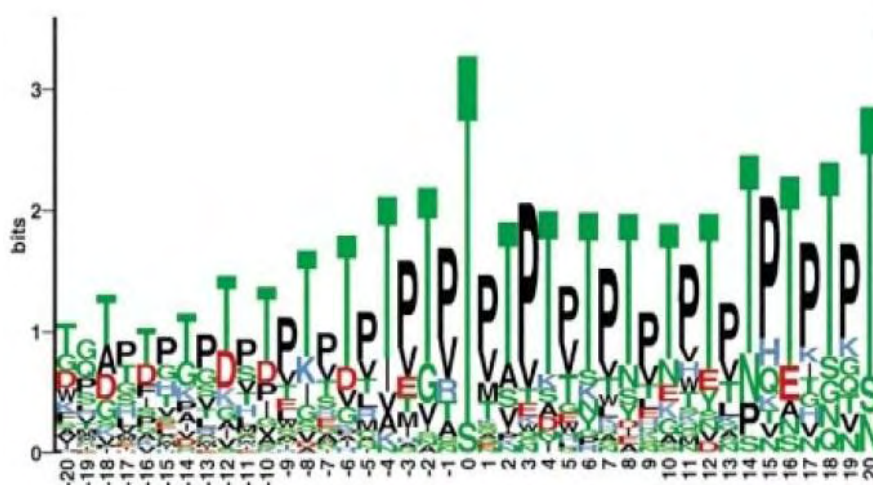
Submit a file in [FASTA](#) format directly from your local disk:

Generate graphics Run signal peptide check

Εικόνα 2-19. Απεικόνιση του αλγορίθμου NetOGlyc για την πρόγνωση των θέσεων γλυκοζυλίωσης.

2.3.2.2 Πρόγνωση της Ο-γλυκοζυλίωσης στο *Dictyostelium discoideum*

Το *Dictyostelium discoideum* έχει προταθεί ως ένα μοντέλο ευκαρυωτικού οργανισμού για τις έρευνες της γλυκοβιολογίας. Τα χαρακτηριστικά των θέσεων πρόσδεσης για τις N-ακετυλογλυκοσαμινυλοτρανσφεράσες στο *Dictyostelium discoideum*, τα οποία συνδέουν την GlcNac στα υδροξυλικά κατάλοιπα, είναι σε μεγάλο βαθμό άγνωστα. Αυτό κάνει έντονη την ανάγκη για την εξέλιξη μίας ακριβούς μεθόδου για την πρόγνωση των θέσεων Ο-γλυκοζυλίωσης στις εκκριτικές και μεμβρανικές πρωτεΐνες του *Dictyostelium discoideum*. Οι επιστήμονες χρησιμοποίησαν μία μέθοδο με χρήση των τεχνητών νευρωνικών δικτύων, καθώς τα δίκτυα αυτά αναγνωρίζουν την ομάδα ακολουθιών και την πρόσβαση στην επιφάνεια της πρωτεΐνης για τις 39 πειραματικές θέσεις Ο-γλυκοζυλίωσης που βρέθηκαν για το *Dictyostelium discoideum* έπειτα από πειράματα που έγιναν σε ζώντα οργανισμό.



Εικόνα 2-20. Λογότυπο ακολουθίας για την Ο-γλυκοζυλίωση στο *Dictyostelium discoideum*.

Μία cross-validation των δεδομένων έδειξε μία συσχέτιση στην οποία αναγνωρίζεται το 97% των γλυκοζυλιωμένων και μη γλυκοζυλιωμένων πρωτεϊνών. Με βάση τα πρόσφατα δεδομένα, παρατηρήθηκε μία άφθονη περιοδικότητα των δύο (θέσεις -3,-1,+1,+3 κτλ) στην εναλλαγή των καταλοίπων προλίνης με τα υδροξυλικά αμινοξέα κατά τη φορά αλλά και κατά την αντίθετη φορά του αποδέκτη θέσης. Η

μέθοδος χρησιμοποιήθηκε για ένα γρηγορότερο έλεγχο του κλάσματος του πρωτεόματος του *Dictyostelium* το οποίο διατίθεται στις δημοσίας χρήσης βάσεις δεδομένων και ένα ποσοστό αυτών της τάξης του 25-30% βρέθηκαν γλυκοζυλιωμένες. Ο έλεγχος αποκάλυψε θέσεις πρόσδεσης σε αρκετές πρωτεΐνες που ήταν πειραματικά αποδεδειγμένο πως υφίστανται Ο-γλυκοζυλίωση σε μη χαρτογραφημένες θέσεις. Η μέθοδος αυτή διατίθεται στην ηλεκτρονική διεύθυνση :

DictyOGlyc	http://www.cbs.dtu.dk/services/DictyOGlyc/
------------	---

Εικόνα 2-21. Απεικόνιση του αλγορίθμου DictyOGlyc για την πρόγνωση των θέσεων Ο-γλυκοζυλίωσης στο *Dictyostelium discoideum*.

2.3.3 Πρόγνωση μυριστοϋλίωσης

Η μόνη διαθέσιμη μέθοδος για πρόγνωση της Ν-τελικής Ν-μυριστοϋλίωσης είναι το αμινοοξικό πρότυπο ελέγχου που διατίθεται στη βάση δεδομένων PROSITE.

myristoylation	http://www.expasy.org/tools/myristoylator/
----------------	---

Myristoylator

Myristoylator predicts N-terminal myristoylation of proteins by neural networks. [Documentation / Reference / Data]
Only N-terminal glycines are myristoylated (leading methionines are cleaved prior to myristoylation).

Please enter one or more UniProtKB/Swiss-Prot protein identifiers (ID) (e.g. *ANX13_HUMAN*) or UniProt knowledgebase (Swiss-Prot/TrEMBL) accession numbers (AC) (e.g. *P27216*), separated by newlines. You can also insert a **single** sequence (minimal length: 17 aa with the N-terminal glycine).

Εικόνα 2-22. Απεικόνιση του προγράμματος Myristoylator για την πρόγνωση των θέσεων μυριστοϋλίωσης.

Ωστόσο δεν επαρκεί για προβλέψεις σε μεγάλα γονίδια καθώς παράγει μεγάλους αριθμούς από λανθασμένα θετικά και ακόμη και μερικές αρνητικές προβλέψεις. Ως αρχικό βήμα για μία νέα μέθοδο πρόγνωσης ήταν απαραίτητο για τους επιστήμονες να καθορίσουν ένα μοτίβο ακολουθίας που θα κωδικοποιείται για την N-τελική N-μυριστοϋλίωση. Στηριζόμενοι στην εις βάθος έρευνα της ποικιλίας αμινοξέων που υπάρχει στις ακολουθίες του υποστρώματος μιας πρωτεΐνης, στους δεσμούς που υπάρχουν στις X-ray δομές ή στα 3D ομόλογα μοντέλα των NTMs από διάφορες ομάδες ταξινόμησης και τέλος στα βιοχημικά δεδομένα που διατίθενται από την επιστημονική λογοτεχνία, οι επιστήμονες βρήκαν ενδείξεις πως τα 17 N-αμινοτελικά κατάλοιπα των πρωτεϊνών εκφράζουν διαφορετικούς τύπους περιορισμών. Αναγνωρίστηκαν 3 περιοχές μοτίβων: η περιοχή 1 που περιλαμβάνει τις θέσεις 1 έως 6 και αφορά τη σύνδεση των δεσμών, η περιοχή 2 που περιλαμβάνει τις θέσεις 7 έως 10 και αντιδρά με την επιφάνεια των NMTs στο ύψος της καταλυτικής κοιλότητας και η περιοχή 3 που περιλαμβάνει τις θέσεις 11 έως 17 [75].

2.3.3.1 Πρόγνωση μυριστοϋλίωσης στα φυτά

Όσον αφορά τα φυτά, έχει αναπτυχθεί μία νέα μέθοδος για την πρόγνωση της N-αμινοτελικής N-μυριστοϋλίωσης, η οποία έχει ελεγχθεί ως προς την ευαισθησία, την ακρίβεια και τη σταθερότητα. Σε σύγκριση με παλαιότερες μεθόδους, το νέο

μοντέλο έχει μεγαλύτερη ευαισθησία στο να ανιχνεύει τα σωστά θετικά και είναι πιο επιλεκτικό στο να αποφεύγει λανθασμένα θετικά. Τα αποτελέσματα των πρωτεϊνών που υφίστανται μυριστοϋλίωση από αυτές που δεν υφίστανται είναι πιο διάσπαρτα από ότι με άλλες μεθόδους, μειώνοντας έτσι την αμφιβολία και τον αριθμό των ενδιάμεσων ακολουθιών για τα μη πληροφοριακά αποτελέσματα. Το μοντέλο πρόγνωσης είναι διαθέσιμο στην ηλεκτρονική διεύθυνση :

Myristoylation in plants

<http://plantsp.sdsc.edu/myrist.html>.

Try a sample file first: yes no

Enter or paste sequence(s):

Upload a file:

Output: Formatted table Plain text

Sorting: By score None (input order)

Εικόνα 2-23. Απεικόνιση του προγράμματος για την πρόγνωση των θέσεων μυριστοϋλίωσης στα φυτά.

Η υψηλή απόδοση του μοντέλου αυτού οφείλεται στην επιλογή συγκεκριμένων δεδομένων από κάθε φυτικό οργανισμό και στο γεγονός ότι καλύπτει 266 παραδείγματα μοναδικών ακολουθιών από 40 διαφορετικά είδη. Χρησιμοποιούνται Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Model –HMM) για την εύρεση των scores πρόγνωσης καθώς και μία τιμή διακοπής κατωφλιού σχετικά με τη διάκριση θετικών και αρνητικών. Το νέο αυτό μοντέλο έχει χρησιμοποιηθεί για την πρόγνωση 589 φυτικών πρωτεϊνών που περιέχουν N-τελικά σήματα μυριστοϋλίωσης και για την ανάλυση των λειτουργικών οικογενειών στις οποίες εμφανίζονται αυτές οι πρωτεΐνες [76].

2.3.4 Πρόγνωση σουλφίδωσης

Η βιολογική λειτουργία της Ο-σουλφίδωσης της τυροσίνης δεν έχει κατανοηθεί αρκετά. Πρόσφατα, μόνο 275 πρωτεΐνες που υπέστησαν σουλφίδωση τυροσίνης έχουν αναφερθεί ή σχολιαστεί με βάση την ομοιότητα στη βάση δεδομένων UniProt.

UniProt	http://www.uniprot.org/
---------	---

Έχουν γίνει έλεγχοι από επιστημονικές ομάδες για το αν αυτή η μετά-μεταφραστική τροποποίηση χρησιμεύει σαν διαμορφωτής στις κρυφές αντιδράσεις πρωτεΐνης με πρωτεΐνη και σε αυτές που καθορίζουν τα όρια των μεμβρανών. Για τέτοιου είδους πρωτεΐνες η αναγνώριση της ομάδας σουλφίδωσης είναι ζωτικής σημασίας. Τα αποτελέσματα της σουλφίδωσης της τυροσίνης σχετίζονται με πολλές βιολογικές διαδικασίες οι οποίες περιλαμβάνουν τη ρύθμιση της αιμόστασης και τη μεταφορά και κίνηση των λευκοκυττάρων.

Αρχικά οι Rosenwuist και Nicholas χρησιμοποίησαν έλεγχο προτύπου (pattern search) για την πρωτεϊνική σουλφίδωση της τυροσίνης [77, 78]. Ανακάλυψαν έναν έλεγχο για τα όξινα αμινοξέα δίπλα στις τυροσίνες στόχους που είχαν ως αποτέλεσμα ένα πολύ καλό κριτήριο φιλτραρίσματος και απέκλειαν τις λανθασμένες θετικές προγνώσεις σουλφίδωσης. Στη συνέχεια ανέπτυξαν το position-specific scoring matrix (PSSM) που αποτελεί έναν μαθηματικό πίνακα των θέσεων σκοραρίσματος, για την πρόγνωση των θέσεων σουλφίδωσης [77, 79, 80]. Ωστόσο, οι καταχωρήσεις και οι διαγραφές δημιούργησαν ένα άλλο πρόβλημα για αυτό το είδος προσέγγισης, επειδή οι μέθοδοι δεν είναι αρκετά δυναμικές ώστε να έχουν πρόσβαση σε διαγραφές και καταχωρήσεις.

Για να αποφύγει αυτές τις ατέλειες, ο Monigatti, ανέπτυξε ένα Κρυφό Μαρκοβιανό Μοντέλο (Hidden Markov Model -HMM) βασισμένο στην πρόγνωση των θέσεων σουλφίδωσης [81]. Το πρόγραμμα αποτελείται από τρία διαφορετικά μοντέλα πινάκων που βασίζονται στη θέση της τυροσίνης που ερευνάται. Για μία τυροσίνη στόχο που τοποθετείται στη μέση της πρωτεϊνικής ακολουθίας, χρησιμοποιείται ένα εσωτερικό HMM. Αν οι θέσεις στόχοι βρίσκονται στα δύο τρίτα της ακολουθίας, χρησιμοποιείται είτε το N-τελικό HMM, είτε το C-τελικό HMM. Χρησιμοποιώντας ανεξάρτητους ελέγχους δοκιμών, η αναγνώριση των λανθασμένων θετικών αποτιμήθηκε κοντά στο 2% και η ευαισθησία στο 98%, από όλα τα

κατάλοιπα σουλφίδωσης της τυροσίνης. Το πρόγραμμα αυτό ονομάζεται ‘Sulfinator’ και είναι διαθέσιμο στην ηλεκτρονική διεύθυνση:

Sulfinator	http://www.expasy.ch/tools/sulfinator/
------------	---

The Sulfinator

The Sulfinator predicts tyrosine sulfation sites in protein sequences [Documentation / Reference]

Tyrosine sulfation is an important post-translational modification of proteins that go through the secretory pathway.

Please enter one or more Swiss-Prot protein identifiers (ID) (e.g. *FA5_HUMAN*) or Swiss-Prot/TrEMBL accession numbers (AC) (e.g. *P12259*), separated by newlines. You can also insert sequences in FASTA format.

Submit Reset

Εικόνα 2-24. Απεικόνιση του προγράμματος Sulfinator για την πρόγνωση των θέσεων σουλφίδωσης.



Εικόνα 2-25. Λογότυπο ακολουθία για τη σουλφίδωση.

Παρόλη την υψηλή ακρίβεια του Sulfinator, το πρόγραμμα αυτό δεν εγγυάται την ύπαρξη νέων θέσεων σουλφίδωσης της τυροσίνης σε όλες τις πρωτεϊνικές ακολουθίες. Από τη στιγμή που η σουλφίδωση της πρωτεΐνης είναι μία PTM που επηρεάζει τις κρυφές ή τις πρωτεΐνες που καθορίζουν τα όρια των μεμβρανών, το πρόγραμμα δεν μπορεί να λειτουργήσει για όλες τις ακολουθίες των πρωτεϊνών, για τις πρωτεΐνες των οποίων η κυτταρική τοποθεσία δεν είναι γνωστή. Υπάρχουν προγράμματα που προβλέπουν τις αξιοσημείωτες ακολουθίες και τις διαμεμβρανικές

έλικες, τα SignalIP και TMHMM αντίστοιχα και τα οποία είναι διαθέσιμα στις ηλεκτρονικές διευθύνσεις:

SignalIP	http://www.cbs.dtu.dk/services/SignalIP
TMHMM	http://www.cbs.dtu.dk/services/TMHMM/

2.3.5 Πρόγνωση παλμιτοϋλίωσης

Σχετικά με την πρόγνωση των θέσεων παλμιτοϋλίωσης οι επιστήμονες έχουν αναπτύξει δύο υπολογιστικά προγράμματα. Το ένα είναι το Clustering and Scoring Strategy for Palmitoylation site Prediction (CSS-Palm) [32] κατά το οποίο η πρόγνωση ενθαρρύνεται με υψηλά θετικά αποτελέσματα στον έλεγχο αξιοπιστίας Jack-Knife (ευαισθησία 82.16% και ακρίβεια 83.17% για score 2.6). Το δεύτερο πρόγραμμα ονομάζεται NBA-Palm και βασίζεται στον αλγόριθμο Naïve Bayes [82]. Τα πειραματικά δεδομένα εξάγονται από την PubMed και περιλαμβάνουν 245 θέσεις παλμιτοϋλίωσης από 105 ξεχωριστές πρωτεΐνες έπειτα από απαλοιφή. Το κατάλληλο μήκος του παραθύρου για ένα πιθανό πεπτίδιο παλμιτοϋλίωσης είναι 6. Για να αποτιμηθεί η κατάσταση της πρόγνωσης του NBA-Palm, διεξήχθησαν επιτυχώς 3-fold cross-validation, 8-fold cross-validation και αξιοπιστία Jack-Knife.

Η ορθότητα της πρόγνωσης φτάνει το 85.79% για την 3-fold cross-validation, το 86.72% για την 8-fold cross-validation και το 86.74% για την αξιοπιστία Jack-Knife. Το CSS-Palm αποτελεί ένα δυνατό και αποδοτικό εργαλείο ελέγχου των θέσεων παλμιτοϋλίωσης, ενώ η ακρίβεια του NBA-Palm είναι συγκρίσιμη με το εργαλείο CSS-Palm. Τα προγράμματα είναι διαθέσιμα στην ηλεκτρονική διεύθυνση:

CSS-Palm	http://bioinformatics.lcd-ustc.org/css_palm/
NBA-Palm	http://www.bioinfo.tsinghua.edu.cn/NBA-Palm

Enter your sequences in the TEXT BOX, and press "Submit" for the prediction results.

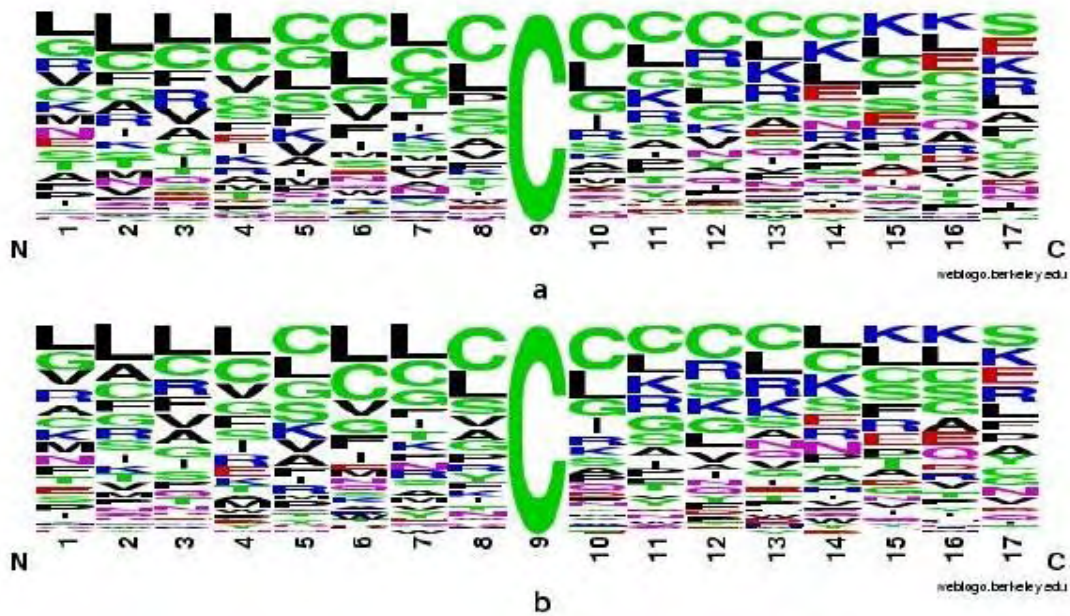
1. Enter the sequences:

All the **spaces, line breaks** will be automatically removed. You could input *raw sequence* for one protein or *multiple proteins' sequences in FASTA format*

```
>sp|Q02224|CENPE_HUMAN Centromeric protein E (CENP-E protein) - Homo sapiens [Human].
MAEEGAVAVCVRVRLPLNSREESLGETAQVYWKTDNNVIYQVDGSKSFNFDRVFHCNETTK
NVYEEIAAPIIDSAIOGYNGTIFAYGOTASGKTYTMMGSEDELGVIPRAIHDIPOKIKK'
PDREFLLRVSYHEIYNETITDLLCGTQKMKPLIREDVNRNVYVADLTEEVVYTSEMALK
WITKGEKSRHYGETKMNQSSRSHTIFRMILESREKGEPSNCEGSVKVSHLNLVDLAGSE
RAAQTCGAAGVRLKEGCNINRSLFILQQVIKKLSDGQVGGFINYRDKLTRLQNSLGGNP
KTRIICTITPVSFDELTLALQFASAKYMKNTPYVNEVSTDEALLKRYRKEIMDLKKQLE
```

Choose the cut-off value:

Εικόνα 2-26. Απεικόνιση του αλγορίθμου CSS-Palm για την πρόγνωση των θέσεων παλμιτοϋλίωσης.



Εικόνα 2-27. Λογότυπο ακολουθίας για την παλμιτοϋλίωση.

2.3.6 Πρόγνωσης σουμοϋλίωσης

Για την πρόγνωση της σουμοϋλίωσης είναι διαθέσιμο ένα υπολογιστικό σύστημα το SUMOsp-SUMOylation Sites Prediction, το οποίο βασίζεται σε δύο μεθόδους, τις GPS και MotifX, οι οποίες αρχικά είχαν σχεδιαστεί για την πρόγνωση των θέσεων φωσφορυλίωσης καθώς και για leave-one out validation και 5-fold cross validation. Οι τελευταίες δύο στρατηγικές για αναγνώριση προτύπων είναι σταθερές και ακριβείς για την πρόγνωση των θέσεων σουμοϋλίωσης. Η SUMOsp προσφέρει σχεδόν τόσο καλή πρόγνωση, όσο και το μοναδικό υπάρχον σύστημα το SUMOplot. Για να διευκολύνουν οι εφαρμογές αυτές άλλους χρήστες, έχει σχεδιαστεί ένας server διαδικτύου όπου η πρόσβαση είναι ελεύθερη και διατίθεται στην ηλεκτρονική διεύθυνση:

SUMOsp	http://bioinformatics.Icd-ustc.org/sumosp/
--------	---

Ερευνητές έψαξαν στην PubMed θέτοντας ως λέξεις κλειδιά τις ‘SUMO’ και ‘sumoylation’ και βρήκαν αναμφίβολα 239 πειραματικά αναγνωρισμένες θέσεις σουμοϋλίωσης σε 144 πρωτεΐνες (πριν τις 10 Δεκεμβρίου του 2005 είχαν εκδοθεί online 400 άρθρα έρευνας). Ανέκτησαν τις αρχικές ακολουθίες από την αρχική βάση δεδομένων της Swiss-Prot/TrEMBL.

Swiss-Prot/TrEMBL	http://cn.expacy.org
-------------------	---

Εξαιτίας των ανανεώσεων της βάσης δεδομένων, οι τοποθεσίες σουμοϋλίωσης που αναφέρθηκαν πιθανόν να άλλαξαν στις πρόσφατες αρχικές ακολουθίες, συνεπώς ελέγχθηκε χειροκίνητα η αξιοπιστία των δεδομένων προτού γίνουν οι αναλύσεις.

Οι επιστήμονες χρησιμοποίησαν δύο δυναμικές στρατηγικές πρόγνωσης, τη GPS [83, 84] και τη MotifX [95], για την πρόγνωση των θέσεων σουμοϋλίωσης. Η GPS αρχικά διαμέλισε τα δεδομένα των πεπτιδίων που συνορεύουν με τις 239 γνωστές θέσεις σουμοϋλίωσης, σε τρία συμπλέγματα. Για ένα δοθέν πεπτίδιο όπου συνορεύουν μία λυσίνη αμινοξέος και ένα από τα συμπλέγματα, η μέγιστη τιμή του score ανάμεσα από αυτό το πεπτίδιο και τα πεπτίδια του συμπλέγματος καθορίζεται ως το score του συμπλέγματος αυτού. Το GPS score αυτού του δοθέντος πεπτιδίου χαρακτηρίζεται ως το μέγιστο από όλα τα scores ανάμεσα στα πεπτίδια και τα συμπλέγματα. Για την τελική κρίση χρησιμοποιείται μία ειδική τιμή απομόνωσης.

Η MotifX [87] παράγει ένα σετ από υψηλής ακρίβειας μοτίβα για τις θέσεις σουμουλίωσης όπως τα IKXEP, VKXE, IKXE, LKXE και KXE (το X μπορεί να είναι οποιοδήποτε αμινοξύ) και τα οποία μπορούν εύκολα να χρησιμοποιηθούν από τους χρήστες. Πράγματι, η MotifX μεγάλη υπολογιστική ικανότητα όταν συνδυάζεται με την GPS. Για παράδειγμα, ένας συνδυασμός της MotifX με την GPS προβλέπει τα πεπτίδια σουμουλίωσης σαν ένα θετικό ταίριασμα.

Έγινε σύγκριση της παρουσίασης πρόγνωσης της SUMOsp με το μοναδικό διαθέσιμο για χρήση εργαλείο SUMOplot. Η SUMOplot θεωρείται ως ένα εξαιρετικό προγραμματιστικό πρόγραμμα. Για να ελέγξουμε τη σταθερότητα της SUMOsp χρησιμοποιείται leave-one-out validation και 5-fold cross validation. Η SUMOsp παρέχει καλύτερη ευαισθησία ενώ κρατάει παρόμοια την ιδιαιτερότητα. Για να διευκρινιστεί πώς η SUMOsp είναι σε σχέση με το κατώφλι ανεξάρτητης εκτέλεσης, παρέχεται ένας αποδέκτης χαρακτηριστικών καμπυλών αυτοαξιοπιστίας (ROC). Για τις μη-κανονικές αληθινές θέσεις σουμουλίωσης, η SUMOsp μπορεί επίσης να παρέχει μία ικανοποιητική παράσταση πρόγνωσης. Το πρόγραμμα αυτό διατίθεται στην ηλεκτρονική διεύθυνση:

SUMOplot	http://www.abgent.com/doc/sumoplot
----------	---

Ο web server για τη SUMOsp έχει εξελιχθεί με έναν εύκολο τρόπο χρήσης. Ένας χρήστης μπορεί να επισκεφθεί τη SUMOsp στο <http://bioinformatics.lcd-ust.org/sumosp/prediction.php>, εισάγει την ακολουθία της πρωτεΐνης σε ανεπεξέργαστη διάταξη ή σε μορφή FASTA στο κουτί κειμένου και να τρέξει το πρόγραμμα πατώντας το κουμπί αποδοχής. Τα αποτελέσματα της πρόγνωσης θεωρούνται σαν πιθανές θέσεις πριν την πειραματική αξιοπιστία. Επίσης πατώντας τη λέξη στην πρόταση 'Download the TAB-delimited data file from here' ένας χρήστης μπορεί να πάρει αποτελέσματα πρόγνωσης σε tab-delimited plain text που μπορεί να χρησιμοποιήσει για περαιτέρω έρευνα.

✖ **SUMOsp: SUMOylation Sites Prediction**

Enter your sequences in the TEXT BOX, and press "Submit" for the prediction results.

Please input the sequences:

All the **spaces, line breaks** will be automatically removed. You could input *one primary sequence* or *multiple proteins' sequences in FASTA format* !

Cut-off score:

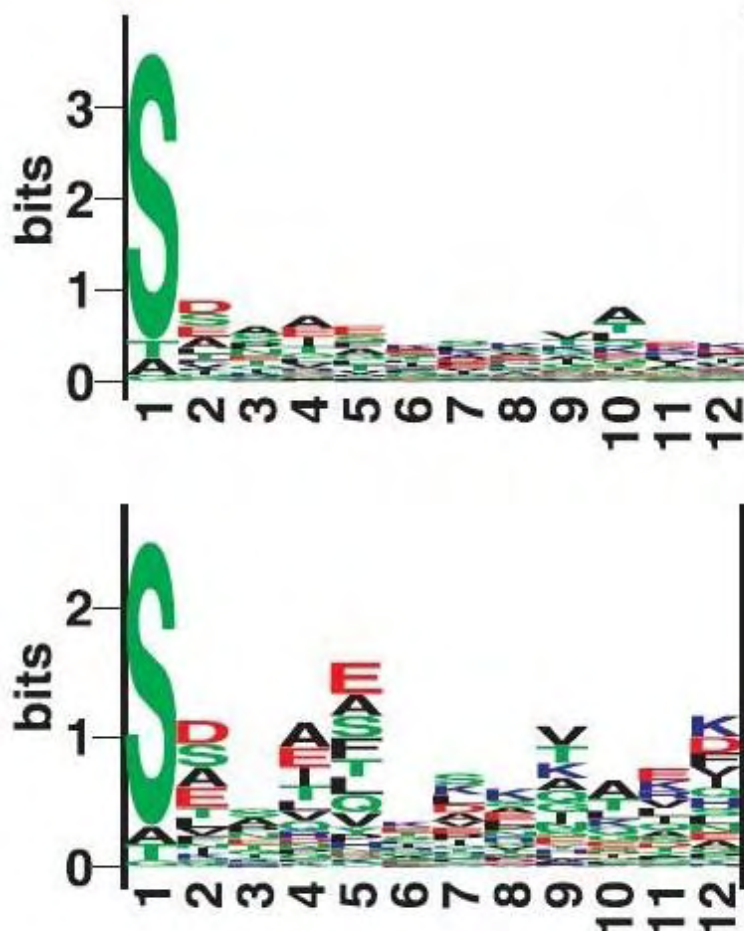
SUMOsp with a higher cut-off value will generate a shorter list of predicted sumoylated sites, which means lower sensitivity but higher specificity. Please refer to the manuscript for the calculated prediction performance of different cut-off scores.

Εικόνα 2-28. Η σελίδα πρόγνωσης του διαδικτυακού server SUMOsp.

2.3.7 Πρόγνωση των N-τελικών θέσεων ακετυλίωσης

Σχεδόν πριν από 20 έτη, έγινε μια προσπάθεια για να προβλεφθεί γενικά η N-αμινοτελική ακετυλίωση, στην οποία χρησιμοποιήθηκε ως εισαγόμενο στοιχείο σε ένα γραμμικό νευρωνικό δίκτυο [86] η δευτεροταγής δομή της προβλεφθείσας πρωτεΐνης. Η αξιολόγηση της απόδοσης ήταν αδύνατη εν τούτοις, καθώς το πρότυπο κατασκευάστηκε και δοκιμάστηκε στο ίδιο σύνολο δεδομένων. Με τους ζυμομύκητες να είναι από τους πιο λεπτομερώς μελετημένους ευκαρυώτες, έχει συσσωρευτεί μία ικανοποιητική ποσότητα στοιχείων που επιτρέπει την κατάρτιση μιας μεθόδου πρόβλεψης για την NatA ακετυλίωση. Οι ζυμομύκητες είναι αρκετά σημαντικοί για την κατασκευή μιας κεντρικής μεθόδου πρόγνωσης, αλλά πρόσθετα, η πρόγνωση φαίνεται να λαμβάνει συγκρίσιμες τιμές απόδοσης με τις πρωτεΐνες των θηλαστικών. Αυτό υποστηρίζει την ιδέα ότι τα συστήματα της N-τελικής ακετυλίωσης είναι παρόμοια σε όλους τους ευκαρυώτες. Σε μία επιστημονική έρευνα που πραγματοποιήθηκε οι ερευνητές εξέτασαν μόνο την NatA N-τελική ακετυλίωση και

όχι την ακετυλίωση στην ε-αμινομάδα των εσωτερικών καταλοίπων της λυσίνης από άλλες ακετυλοτρανσφεράσεις.



Εικόνα 2-29. Λογότυπο ακολουθίας του Shannon (πάνω) και Kullback-Leibler (κάτω) για τις θέσεις ακετυλίωσης.

Όλα τα δεδομένα της έρευνας εξήχθησαν από τον πίνακα 2 των Polevoda και Sherman [86] και ενώθηκαν με τα στοιχεία από τον πρωτεϊνικό χάρτη των ζυμομυκήτων (Yeast Protein Map-YPM) [87]. Οποιοσδήποτε ασυνέπειες ανάμεσα στα δεδομένα των δύο συνόλων αφαιρέθηκαν για να ληφθούν τα υψηλότερα πιθανά ποιοτικά στοιχεία. Επιπλέον, εξήχθησαν μόνο τα υποστρώματα που αναφέρθηκαν να έχουν υποστεί ακετυλίωση από την NatA, καθώς αυτή είναι η μόνη τρανσφεράση για την οποία μέχρι στιγμής έχει συσσωρευτεί ένα ικανοποιητικό ποσοστό στοιχείων.

Αυτό οδήγησε σε 61 θετικές και 76 αρνητικές ακολουθίες εκπαίδευσης. Οι ακολουθίες περικόφτηκαν στα 40 N-τερματικά τους κατάλοιπα και στη συνέχεια με ένα γειτονικό ομόλογο ενώθηκαν σε ένα δέντρο που παράγεται από μια πολλαπλάσια ευθυγράμμιση του αλγορίθμου ClustalW [89]. Στη συνέχεια εξετάστηκε ένα τεχνητό νευρωνικό δίκτυο χρησιμοποιώντας το τριπλάσιο cross-validation [90]. Χρησιμοποιώντας ένα παράθυρο δικτύου μεγέθους επτά αμινοξέων και οκτώ κρυμμένους νευρώνες οι επιστήμονες έλαβαν έναν συντελεστή συσχετισμού Matthews [74] της τάξης του 0.69 κατά τη χρησιμοποίηση ενός κατωφλιού της τάξης του 0.5. Αυτός ο συντελεστής συσχετισμού απεικονίζει μια ευαισθησία σε ποσοστό 75% και μια ειδικότητα σε ποσοστό 92%.

Η μέθοδος που παρουσιάστηκε προβλέπει τις θέσεις ακετυλίωσης της τρανσφεράσης NatA στους ζυμομήκυτες με την υψηλότερη επίδοση και επίσης μέχρι ένα σημείο και για τα θηλαστικά [100]. Πιστεύεται ότι η μέθοδος αυτή θα είναι ιδιαίτερα χρήσιμη στους ερευνητές που εργάζονται πάνω στην ακετυλίωση. Η κεντρική μέθοδος πρόγνωσης και οι πρόσθετες πληροφορίες είναι διαθέσιμες στην ηλεκτρονική διεύθυνση:

NetAcet	http://www.cbs.dtu.dk/services/NetAcet/
---------	---

NetAcet 1.0 Server

NetAcet 1.0 server predicts substrates of N-acetyltransferase A (NatA). The method was trained on yeast data but, as mentioned in the article describing the method, it obtains similar performance values on mammalian substrates acetylated by NatA orthologs.

[Background](#) [Instructions](#) [Output format](#) [Article abstract](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

Submit a file in [FASTA](#) format directly from your local disk:

Show input sequences Show individual network scores

Restrictions:
At most 2000 sequences and 200,000 amino acids per submission; each sequence not less than 40 and not more than 4,000 amino acids.

Confidentiality:
The sequences are kept confidential and will be deleted after processing.

Εικόνα 2-30. Απεικόνιση του αλγορίθμου NetAcet για την πρόγνωση των θέσεων ακετυλίωσης.

2.3.8 Πρόγνωση της πρωτεϊνικής γλυκίωσης στα θηλαστικά

Το πλαίσιο των λυσινών που υπόκεινται σε γλυκίωση απεικονίστηκε από το λογότυπο ακολουθίας [92]. Το λογότυπο ακολουθίας παρουσιάζει ποιά αμινοξέα είναι χαρακτηριστικά κοντά στις λυσίνες στην ακολουθία.

Εκτελέστηκαν αλγόριθμοι δικτύων χρησιμοποιώντας την ακολουθία που εισήχθη μόνη της και με την πρόσθετη εισαγωγή της σχετικής θέσης της λυσίνης στην ακολουθία. Παρατηρήθηκαν τα καλύτερα αποτελέσματα για την εισαγωγή αυτή που περιέλαβε τη σχετική θέση. Δοκιμάστηκαν όλοι οι συνδυασμοί των δικτύων για μεγέθη παραθύρων από 3 έως 51 καταλοίπων αμινοξέων καθώς και 0, 2, 3, ..., 20 κρυμμένοι νευρώνες. Για την αφαίρεση του θορύβου από το μοντέλο, οι επιστήμονες αποφάσισαν να προσθέσουν κενά στη μέση του παραθύρου αρχίζοντας με ένα κενό που κυμαίνεται από τις θέσεις - 3 έως +3. Η πρόγνωση έγινε με το συνδυασμό 60 τεχνητών νευρωνικών δικτύων. Ο συντελεστής συσχέτισης Matthews για την πρόγνωση είναι 0.58 κάτι το οποίο είναι αρκετά εντυπωσιακό δεδομένου ότι τα πειραματικά ποσοστά που είναι διαθέσιμα είναι σχετικά μικρής ποσότητας. Η μέθοδος αυτή είναι διαθέσιμη στην ηλεκτρονική διεύθυνση:

NetGlycate-1.0	www.cbs.dtu.dk/services/NetGlycate-1.0
----------------	--

NetGlycate 1.0 Server

NetGlycate 1.0 server predicts glycation of ε amino groups of lysines in mammalian proteins.
For a description of the data set used to develop the NetGlycate method see [here](#).

[Instructions](#) [Output format](#)

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

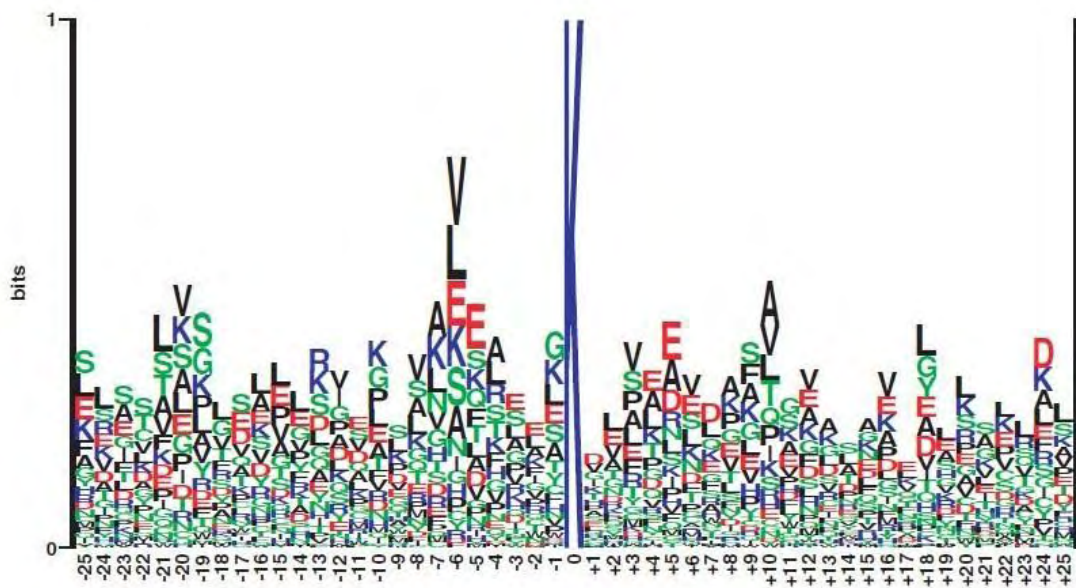
Submit a file in [FASTA](#) format directly from your local disk:

Generate graphics

Restrictions:
At most 2,000 sequences and 200,000 amino acids per submission; each sequence not more than 6,000 amino acids.

Confidentiality:
The sequences are kept confidential and will be deleted after processing.

Εικόνα 2-31. Απεικόνιση του αλγορίθμου NetGlycate για την πρόγνωση των θέσεων γλυκίωσης.



Εικόνα 2-32. Λογότυπο ακολουθίας των θέσεων πρωτεϊνικής γλυκίωσης στα θηλαστικά.

2.3.9 Πρόγνωση των περιοχών διαχωρισμού στις προ-πρωτεΐνες

Παρουσιάζεται μια υπολογιστική μέθοδος για την πρόβλεψη των περιοχών διαχωρισμού των προ-πρωτεϊνών βασισμένη στα τεχνητά νευρωνικά δίκτυα. Έχουν κατασκευαστεί δύο διαφορετικοί τύποι νευρωνικών δικτύων: ένα δίκτυο furin-specific με βάση τα πειραματικά αποτελέσματα που προέρχονται από τη βιβλιογραφία και ένα γενικό δίκτυο PC-specific που εξετάζει τα στοιχεία που προέρχονται από τη βάση δεδομένων Swiss-Prot. Η μέθοδος προβλέπει τις περιοχές διαχωρισμού στις ανεξάρτητες ακολουθίες με ευαισθησία της τάξης του 95% για το νευρικό δίκτυο furin και 62% για το γενικό δίκτυο PC. Η μέθοδος είναι διαθέσιμη στην ηλεκτρονική διεύθυνση:

ProP	http://www.cbs.dtu.dk/services/ProP
------	---

ProP 1.0 Server

ProP 1.0 server predicts arginine and lysine propeptide cleavage sites in eukaryotic protein sequences using an ensemble of neural networks. Furin-specific prediction is the default. It is also possible to perform a general proprotein convertase (PC) prediction.

For convenience, this server is integrated with the [SignalP](#) server predicting the presence and location of signal peptide cleavage sites.

Instructions	Output format	Article abstract
------------------------------	-------------------------------	----------------------------------

SUBMISSION

Paste a single sequence or several sequences in [FASTA](#) format into the field below:

Submit a file in [FASTA](#) format directly from your local disk:

<input checked="" type="checkbox"/> Generate graphics	<input checked="" type="checkbox"/> Include signal peptide prediction
<input type="checkbox"/> Verbose output	<input type="checkbox"/> General PC prediction

Εικόνα 2-33. Απεικόνιση του αλγορίθμου ProP για την πρόγνωση των θέσεων διαχωρισμού στις προ-πρωτεΐνες.

2.3.9.1 Λογότυπο ακολουθίας furin

Τα στοιχεία διαχωρισμού furin λήφθηκαν από τη λογοτεχνία [93] και αποτελούνται από 38 πρωτεΐνες, οι οποίες περιείχαν τις αναφερόμενες furin περιοχές, συμπεριλαμβανομένων των πρωτεϊνών από τα θηλαστικά και από παθογόνα βακτηρίδια και ιούς. Το γενικό σύνολο των στοιχείων του διαχωρισμού βασίστηκε στη βάση δεδομένων Swiss-Prot [94] και συμπεριλαμβάνει επίσης τα στοιχεία διαχωρισμού furin που περιγράφονται παραπάνω, συμπληρώνοντας συνολικά 235 περιοχές σε 227 πρωτεΐνες. Προκειμένου να αποφευχθεί ο έλεγχος στις ομόλογες ακολουθίες, κάθε σύνολο στοιχείων διαιρέθηκε σε τέσσερα μέρη περίπου ίσου μεγέθους βασισμένο στα φυλογενετικά δέντρα, τα οποία κατασκευάστηκαν χρησιμοποιώντας τις πολλαπλές στοιχίσεις και το γειτονικό αλγόριθμο του προγράμματος ClustalX [95].

Η αρχιτεκτονική των νευρωνικών δικτύων ήταν πλήρως συνδεδεμένη, με μηδενικό ή ένα στρώμα των δύο, τέσσερις, οκτώ ή δεκαέξι κρυμμένες μονάδες [96]. Τα νευρωνικά δίκτυα εκτελέστηκαν με back-propagation [97] και τα στοιχεία της ακολουθίας παρουσιάστηκαν στο δίκτυο χρησιμοποιώντας αραιά κωδικοποιημένα κινούμενα παράθυρα [98]. Κατά τη διάρκεια της εκπαίδευσης και δοκιμής χρησιμοποιήθηκαν μόνο οι θέσεις με τη λυσίνη ή την αργινίνη. Ο cross-validated συντελεστής συσχέτισης Matthews για το συνδυασμένο σύνολο υπολογίστηκε από το

συνολικό αριθμό των αληθών θετικών τα αληθή αρνητικά, τις σχολιασμένες περιοχές διαχωρισμού και τις σχολιασμένες περιοχές του μη διαχωρισμού.

Τα εξεταζόμενα δίκτυα παρέχουν αποτελέσματα μεταξύ του μηδενός και του ένα για κάθε κατάλοιπο αργινίνης (R) ή λυσίνης (L) σε μία ακολουθία αμινοξέως. Τα συνδυασμένα αποτελέσματα πρόβλεψης υπολογίστηκαν έπειτα ως ο μέσος όρος των τεσσάρων δικτύων. Για τα αποτελέσματα με βαθμό επάνω από 0.5 προβλέπεται ο διαχωρισμός των πεπτιδικών δεσμών στη C-αμινοτελική πλευρά του καταλοίπου του αμινοξέως. Για να επεξηγήσουν την απόδοση πρόβλεψης για τα κατώτατα όρια και τις ευαισθησίες πρόβλεψης, οι επιστήμονες παρουσίασαν τη λειτουργία χαρακτηριστικών καμπυλών (ROC), την ευαισθησία χάραξης στον x-άξονα και το ψευδές θετικό ποσοστό στον y-άξονα.



Εικόνα 2-34. Λογότυπα ακολουθιών των θέσεων διαχωρισμού των προ-πεπτιδίων γύρω από την περιοχή του P1.

2.3.9.2 Γενικό λογότυπο ακολουθίας PC

Το γενικό λογότυπο ακολουθίας PC είναι βασισμένο σε 234 από τις 235 περιοχές διαχωρισμού, οι οποίες επιλέχτηκαν έχοντας το P ή το K P1. Μια από τις 235 περιοχές διαχωρισμού βρίσκεται τρία μόνο κατάλοιπα από το C-αμινοτελικό άκρο της πρωτεΐνης και επομένως, για εκείνη την ενιαία περιοχή δεν ήταν δυνατό να εξαχθούν επτά C-αμινοτελικά κατάλοιπα, όπως απαιτείται για την παραγωγή του λογότυπου ακολουθίας στο σχήμα 1B. Η αργινίνη είναι το συχνότερο κατάλοιπο αμινοξέως P1, και αντιστοιχεί σε ποσοστό 92%. Για το P2 οι συχνότητες του P και

του K ήταν 22% και 43%, αντίστοιχα, ενώ η συχνότητα του P ήταν 50% P4. Μόνο έξι υπολείμματα λυσίνης βρέθηκαν σε αυτήν την θέση, που αντιστοιχεί σε ποσοστό 3%. Στη P1 θέση η συχνότητα της σερίνης ήταν 24%, ενώ η συχνότητα του υδροφοβικού, της αλειφατικής λευκίνης, της ισολευκίνης και της βαλίνης των αμινοξέων ήταν 17% στο σύνολο. Τα PC περιλαμβάνονται στην ενεργοποίηση μιας μεγάλης ποικιλίας πρωτεϊνών όπως τις ορμόνες, τα νευροπεπτίδια, την αύξηση των πεπτιδίων και τους παράγοντες διαφοροποίησης, τους παράγοντες προσκόλλησης, τους δέκτες, το αίμα και τους παράγοντες πήξης, τις πρωτεΐνες πλάσματος, τις εξωγενείς πρωτεΐνες όπως τις γλυκοπρωτεΐνες από τους μολυσματικούς ιούς (π.χ. ιός HIV-1 και influenza) και τις βακτηριακές τοξίνες (π.χ. διφθερίτιδα και τοξίνη του άνθρακα). Τα PC όχι μόνο καταλύουν την αφαίρεση των προ-περιοχών, αλλά περιλαμβάνουν επίσης τα μέσα για την επεξεργασία των πολυσύνθετων προδρόμων μορίων.

3. ΠΡΟΓΝΩΣΗ ΤΗΣ ΦΩΣΦΟΥΛΙΩΣΗΣ ΤΩΝ ΠΡΩΤΕΪΝΩΝ

3.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο κάναμε μία ανασκόπηση των ειδών των μετά-μεταφραστικών τροποποιήσεων και των μεθόδων πρόγνωσής τους. Στη συνέχεια θα ασχοληθούμε εκτενώς με τη μετά-μεταφραστική τροποποίηση της φωσφορυλίωσης στους ευκαρυωτικούς οργανισμούς, στα αρχαία και στα βακτήρια.

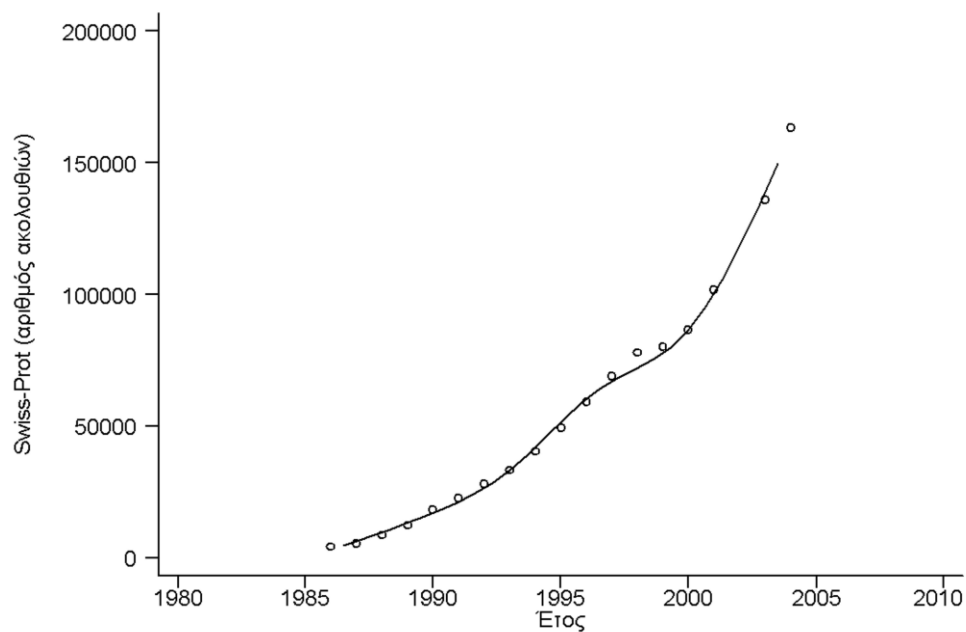
Για το σκοπό αυτό αναπτύξαμε ένα πρόγραμμα το οποίο βασίζεται στη γλώσσα προγραμματισμού perl και ελέγχοντας τις πρωτεΐνες των οργανισμών που υπόκεινται σε φωσφορυλίωση, μας δίνει τον αριθμό των διαφορετικών κινασών που υπάρχουν σε κάθε ένα από αυτούς. Οι οργανισμοί κατηγοριοποιούνται ως προς το είδος, το όνομα και το φύλο και τέλος γίνεται μία σύγκριση των συνολικών κινασών ανά κατηγορία με το φύλο. Τα αποτελέσματα της σύγκρισης αυτής μπορούν να δώσουν σημαντικές πληροφορίες για το μέγεθος της φωσφορυλίωσης μεταξύ ευκαρυωτών, αρχαίων και βακτηρίων.

3.2 Υλικά και μέθοδοι

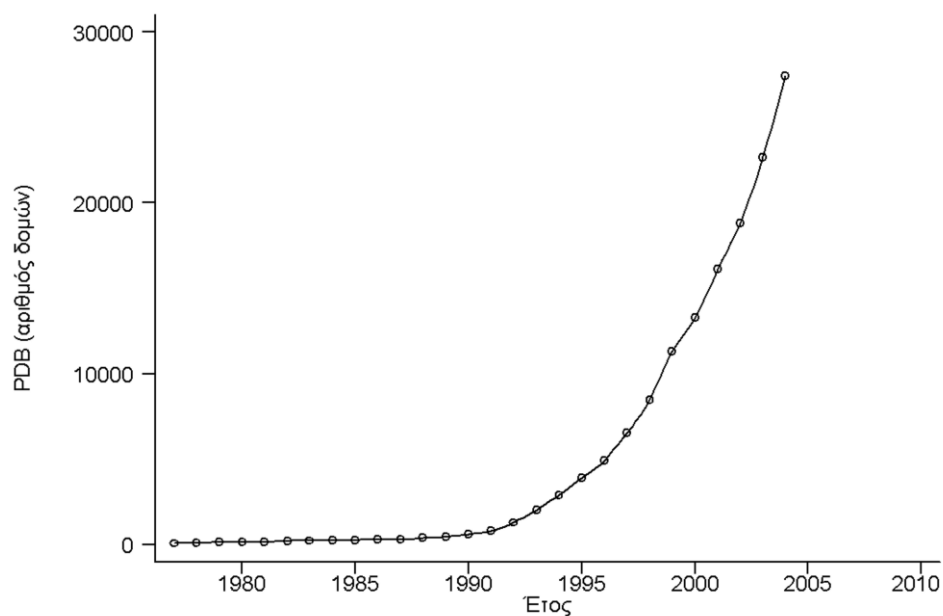
3.2.1 Συλλογή δεδομένων ακολουθιών

Για την εκπαίδευση μιας μεθόδου φωσφορυλίωσης των πρωτεϊνών είναι απαραίτητο να συγκροτήσουμε ένα σύνολο δεδομένων πρωτεϊνικών ακολουθιών για τις οποίες υπάρχουν αξιόπιστα δεδομένα. Το σύνολο των δεδομένων που χρησιμοποιήθηκε για εκπαίδευση, επιλέχτηκε από τη βιβλιογραφία, από έγκυρες πρωτεϊνικές βάσεις μέσω διαδικτύου. Συγκεκριμένα, εξετάστηκε η πρωτεϊνική βάση δεδομένων Swiss-prot η οποία περιέχει το 5% των πρωτεϊνών με γνωστή δομή. Από τη βάση αυτή επιλέχτηκε το αρχείο dbPTM2 που περιέχει τις πειραματικά ελεγμένες θέσεις φωσφορυλίωσης των πρωτεϊνών (<http://dbptm.mbc.nctu.edu.tw/download.php>). Στο σημείο αυτό χρησιμοποιήθηκε η μέθοδος KinasePhos, η οποία υιοθετείται για 20 είδη μετά-μεταφραστικών τροποποιήσεων και παρέχει περισσότερες από 30 γνωστές θέσεις φωσφορυλίωση (<http://kinasephos.mbc.nctu.edu.tw/download.html>). Η βάση dbPTM2 αποτελείται από ένα αρχείο το οποίο περιέχει όλες τις πειραματικά ελεγμένες θέσεις μετά-μεταφραστικών τροποποιήσεων καθώς και κάποιες μικρότερες ανά κατηγορία για cross-validation. Από τη λίστα των πρωτεϊνών επιλέχτηκε ένα παράδειγμα από κάθε κατηγορία ως προς έλεγχο, ενώ στη συνέχεια συλλέχτηκαν

όλες οι πρωτεΐνες από τη βάση δεδομένων Uniprot [104]. Έπειτα ελέγχθηκε ποιες από αυτές τις πρωτεΐνες έχουν λυμένη δομή στην PDB.



Εικόνα 3.1 Η εκθετική αύξηση των πρωτεϊνικών ακολουθιών οι οποίες είναι καταθετειμένες στην Swiss-Prot, από το 1986 έως το τέλος του 2004.



Εικόνα 3.2 Η εκθετική αύξηση των προσδιορισμένων πρωτεϊνικών δομών οι οποίες είναι καταθετειμένες στην PDB, από το 1977 έως το τέλος του 2004.

3.2.2 Χρήση του προγράμματος NetPhosK για μεγάλο αριθμό πρωτεϊνών

Από τις αντίστοιχες βάσεις δεδομένων της NCBI συλλέχτηκαν προς έλεγχο συνολικά 234 γονιδιώματα, συμπεριλαμβανομένου και του ανθρώπου. Συγκεκριμένα 35 γονιδιώματα για τους ευκαρυώτες, 28 για τα αρχαία και 171 για τους προκαρυωτικούς οργανισμούς. Έτσι το σύνολο των πρωτεϊνών προς ανάλυση ανέρχεται στο 1.020.130. Με τη βοήθεια του NetPhosK 1.0 Server, χρησιμοποιήσαμε το πρόγραμμα NetPhosK [105] και το τρέξαμε για μεγάλο αριθμό πρωτεϊνών σε περιβάλλον Linux. Η μέθοδος του NetPhosK βασίζεται στην εφαρμογή Τεχνητών Νευρωνικών Δικτύων (ANNs) τύπου standard feed-forward και αποτελεί μία προέκταση της γενικής μεθόδου NetPhos. Η εφαρμογή των ANNs κάνει πρόγνωση των θέσεων φωσφορυλίωσης σε ανεξάρτητες ακολουθίες με ευαισθησία που κυμαίνεται από το 69% έως και 96% [108, 109]. Στη μέθοδο NetPhos [108, 109], υπήρχαν τρία data sets για σερίνη-θρεονίνη-τυροσίνη και κάθε ένα από αυτά ήταν χωρισμένο σε πέντε μέρη για να εφαρμοστεί πολλαπλή στοίχιση. Η διαδικασία αυτή χρησιμοποίησε τέσσερα από τα πέντε υποσύνολα για εκπαίδευση, ενώ το τελευταίο υποσύνολο είχε χρησιμοποιηθεί για έλεγχο της απόδοσης. Το αποτέλεσμα ήταν πέντε διαφορετικά δίκτυα για κάθε ένα από τα τρία φωσφοκατάλοιπα. Έπειτα είχε εφαρμοστεί ο αλγόριθμος ClustalW, ενώ η πρόγνωση έγινε με χρήση του συντελεστή συσχέτισης Mathews.

Στη συνέχεια ήταν απαραίτητο να υλοποιήσουμε ένα πρόγραμμα το οποίο θα καλεί ένα αρχείο fasta-oneline, θα διαβάζει μία γραμμή και την επόμενη της και θα φτιάχνει ένα αρχείο μόνο με την πρωτεΐνη αυτή.

Το πρόγραμμα NetPhosK χρησιμοποιείται για την πρόγνωση των θέσεων φωσφορυλίωσης στις κινάσες των πρωτεϊνών και αποτελεί σημαντικό εργαλείο για την υλοποίηση της παρούσας πτυχιακής εργασίας. Περιλαμβάνει δύο μεθόδους εισαγωγής ακολουθιών προς εξέταση: i) ο χρήστης εισάγει με επικόλληση μία ή περισσότερες ακολουθίες σε μορφή FASTA στο πεδίο εισαγωγής ή ii) ο χρήστης “κατεβάζει” ένα φάκελο που είναι σε μορφή FASTA κατευθείαν από τον τοπικό δίσκο του υπολογιστή του. Ο φάκελος αυτός μπορεί να περιλαμβάνει πολλαπλές ακολουθίες. Οι ακολουθίες εισαγωγής πρέπει να ακολουθούν την εξής μορφή: 'acdefghiklmnpqrstvwyACDEFGHIKLMNPQRSTVWY'.

Οποιοσδήποτε άλλος χαρακτήρας γίνεται δεκτός, ωστόσο δεν λαμβάνεται υπόψη όταν γίνονται οι προβλέψεις. Ο server είναι προγραμματισμένος να εφαρμόζει πρόγνωση χωρίς τη χρήση φίλτρου. Μπορεί να χρησιμοποιηθεί το φίλτρο ESS

(Evolutionary Stable Sites) με μαρκάρισμα της επιλογής ‘ESS’. Επίσης με μαρκάρισμα της επιλογής ‘Kinase Landscapes’ ο server παράγει γραφήματα. Τέλος αφού ο χρήστης μαρκάρει τις κατάλληλες επιλογές, πατώντας στην επιλογή ‘Submit’, εμφανίζεται στην οθόνη του υπολογιστή η εξέλιξη της εργασίας και στη συνέχεια το αποτέλεσμα του server. Κατά τη διάρκεια της αναμονής ο χρήστης μπορεί να εισάγει την ηλεκτρονική του διεύθυνση και απλά να κλείσει το παράθυρο. Η εργασία θα συνεχίζει να εκτελείται και μόλις τερματιστεί ο χρήστης ενημερώνεται με e-mail. Το μήνυμα που θα λάβει θα περιλαμβάνει το URL στο οποίο θα έχουν αποθηκευτεί τα αποτελέσματα και τα οποία παραμένουν στον server για 24 ώρες έτσι ώστε να μπορεί ο χρήστης να τα συλλέξει.

Σε ένα αρχείο output με μορφή κειμένου αποθηκεύσαμε τα αποτελέσματα του NetPhosK, το μελετήσαμε και υλοποιήσαμε ένα πρόγραμμα το οποίο διαβάζει το αρχείο output και κρατάει το είδος της κινάσης. Ως αποτέλεσμα έχουμε τον αριθμό θέσεων φωσφορυλίωσης ανά πρωτεΐνη συνολικά και αριθμό ανά είδος κινάσης.

```
##gff-version 2
##source-version netphos-3.1b
##date 2008-04-18
##Type Protein aa
# seqname          source          feature          start    end    score  N/A    ?
# -----
aa                  netphos-3.1b  phos-PKA         16      16    0.621  . .    YES
aa                  netphos-3.1b  phos-PKC         26      26    0.530  . .    YES
aa                  netphos-3.1b  phos-CKII        31      31    0.526  . .    YES
aa                  netphos-3.1b  phos-unsp        43      43    0.906  . .    YES
aa                  netphos-3.1b  phos-DNAPK       51      51    0.532  . .    YES
aa                  netphos-3.1b  phos-PKA         59      59    0.668  . .    YES
aa                  netphos-3.1b  phos-unsp        62      62    0.656  . .    YES
aa                  netphos-3.1b  phos-unsp        64      64    0.799  . .    YES
aa                  netphos-3.1b  phos-unsp        95      95    0.715  . .    YES
aa                  netphos-3.1b  phos-PKC         97      97    0.524  . .    YES
aa                  netphos-3.1b  phos-PKA        113     113    0.635  . .    YES
aa                  netphos-3.1b  phos-unsp        130     130    0.870  . .    YES
aa                  netphos-3.1b  phos-unsp        131     131    0.983  . .    YES
aa                  netphos-3.1b  phos-unsp        151     151    0.526  . .    YES
aa                  netphos-3.1b  phos-PKC        156     156    0.580  . .    YES
aa                  netphos-3.1b  phos-CKII       169     169    0.532  . .    YES
aa                  netphos-3.1b  phos-PKC        195     195    0.689  . .    YES
aa                  netphos-3.1b  phos-unsp       197     197    0.574  . .    YES
aa                  netphos-3.1b  phos-unsp       200     200    0.739  . .    YES
aa                  netphos-3.1b  phos-unsp       213     213    0.828  . .    YES
aa                  netphos-3.1b  phos-PKC        214     214    0.839  . .    YES
```

aa	netphos-3.1b	phos-PKC	218	218	0.599	. .	YES
aa	netphos-3.1b	phos-SRC	281	281	0.526	. .	YES
aa	netphos-3.1b	phos-CKI	294	294	0.573	. .	YES
aa	netphos-3.1b	phos-CKII	300	300	0.545	. .	YES
aa	netphos-3.1b	phos-cdc2	302	302	0.524	. .	YES
aa	netphos-3.1b	phos-PKG	312	312	0.501	. .	YES
aa	netphos-3.1b	phos-unsp	319	319	0.698	. .	YES

Εικόνα 3.4 Ενδεικτικό παράδειγμα του αρχείου output.

Για την επιτυχή υλοποίηση του προγράμματός μας χρησιμοποιήθηκε το πρόγραμμα fasta2line σε περιβάλλον perl , το οποίο μετατρέπει οποιοδήποτε fasta με πολλές γραμμές σε fasta-oneline.

```

while(<>)
{ $tseqs.=$_; }
while( $tseqs=~/^>(.*?)$/mg )
{
    push @names, $1;
}
$tseqs=~s/^>.*$/ /mg;
$tseqs=~s/^ //s;
$tseqs=~s/\n//sg;
$nseq=0;
foreach $s ( split / /, $tseqs )
{
    $nseq++;
    print ">$names[$nseq-1]\n$s\n";
}

```

Εικόνα 3.5 Το αρχείο fasta2line.

3.2.3 Υλοποίηση προγράμματος σε περιβάλλον perl

Στις πρωτεΐνες που ελέγχθησαν βρέθηκαν 18 διαφορετικά είδη κινασών, τα οποία όπως καταγράφηκαν είναι τα εξής :

- cdc2
- unsp
- PKC
- PKG
- CKII

- CaM-II
- INSR
- GSK3
- EGFR
- PKA
- CKI
- SRC
- p38MAPK
- DNAPK
- cdk5
- ATM
- RSK
- PKB

Ως unsp (unspecified) χαρακτηρίζεται μία μη αναγνωρισμένη κινάση. Στη συνέχεια υλοποιήθηκε το πρόγραμμα program.pl σε περιβάλλον perl το οποίο παρουσιάζεται παρακάτω.

```

%word=("cdc2"=>"0",
      "unsp"=>"0",
      "PKC"=>"0",
      "GSK3"=>"0",
      "PKG"=>"0",
      "CKII"=>"0",
      "CaM-II"=>"0",
      "INSR"=>"0",
      "EGFR"=>"0",
      "PKA"=>"0",
      "CKI"=>"0",
      "SRC"=>"0",
      "p38MAPK"=>"0",
      "DNAPK"=>"0",
      "cdk5"=>"0",
      "ATM"=>"0",
      "RSK"=>"0",
      "PKB"=>"0",);

while (<>)
{
    if ($_~/^>(.*?) (\w+)\samino\sacids/)

```

```

{
print "$1\t$2\t";

    while (($x,$y)=each(%word))
    {
        print "$x is $y\t";
    }
    print "\n";

%word=("cdc2"=>"0",
"unsp"=>"0",
"PKC"=>"0",
"GSK3"=>"0",
"PKG"=>"0",
"CKII"=>"0",
"CaM-II"=>"0",
"INSR"=>"0",
"EGFR"=>"0",
"PKA"=>"0",
"CKI"=>"0",
"SRC"=>"0",
"p38MAPK"=>"0",
"DNAPK"=>"0",
"cdk5"=>"0",
"ATM"=>"0",
"RSK"=>"0",
"PKB"=>"0",);

}

if ($_ =~ /\.*\w+\s{3}[A-Z]{9}\s{3}[01]\.[0-9]{3}\s{3}(.*)\s+[YES]/)
{
    $a=$1;
    #print "$a\n";
    $word{"$a"}++;
}
}

```

Εικόνα 3.6 Το πρόγραμμα program.pl.

3.2.4 Ταξινόμηση των οργανισμών από την PubMed

Έπειτα από την υλοποίηση του προγράμματος στο περιβάλλον της perl ήταν απαραίτητος ο έλεγχος και η ταξινόμηση των οργανισμών που θα επιλεγθούν μεταξύ των ευκαρυωτών, των αρχαίων και των βακτηρίων. Για το λόγο αυτό χρησιμοποιήσαμε ως εργαλείο την κεντρική επιστημονική σελίδα της PubMed [106]. Συγκεκριμένα επιλέχτηκε η εντολή taxonomy θέτοντας κάθε φορά στο πλαίσιο αναζήτησης τις λέξεις eukaryota, archaea και bacteria για κάθε μία από τις κατηγορίες των ευκαρυωτών, των αρχαίων και των βακτηρίων αντίστοιχα. Επιλέγοντας κάθε μία από τις κατηγορίες αυτές μας εμφανίζει και το σύνολο των οργανισμών που περιλαμβάνονται στην κάθε κατηγορία. Έτσι επιλέξαμε ενδεικτικά έναν οργανισμό από το κάθε γένος, ως σύγκριση.

Όσον αφορά τους ευκαρυώτες οι οργανισμοί που επιλέχτηκαν συμπεριλαμβανομένου και του ανθρώπου, είναι οι εξής :

- *Anopheles_gambiae*
- *Arabidopsis_thaliana*
- *Ashbya_gossypii*
- *Aspergillus_fumigatus*
- *Aspergillus_niger*
- *Bos_taurus*
- *Caenorhabditis_briggsae*
- *Caenorhabditis_elegans*
- *Candida_glabrata*
- *Cryptosporidium_parvum*
- *Danio_rerio*
- *Debaryomyces_hansenii*
- *Dictyostelium_discoideum*
- *Drosophila_melanogaster*
- *Encephalitozoon_cuniculi*
- *Gallus_gallus*
- *Guillardia_theta*
- *Homo_sapiens*
- *Kluyveromyces_lactis*
- *Leishmania_braziliensis*
- *Leishmania_infantum*

- *Leishmania_major*
- *Oryza_sativa*
- *Ostreococcus_lucimarinus*
- *Ostreococcus_tauri*
- *Pichia_stipitis*
- *Plasmodium_falciparum*
- *Plasmodium_knowlesi*
- *Plasmodium_vivax*
- *Rattus_norvegicus*
- *Saccharomyces_cerevisiae*
- *Schizosaccharomyces_pombe*
- *Theileria_annulata*
- *Toxoplasma_gondii*
- *Yarrowia_lipolytica*

Για τα αρχαία επιλέχτηκαν:

Στην κατηγορία των Crenarchaeota:

- *Aeropyrum_pernix*
- *Ignicoccus_hospitalis_KIN4_I*
- *Staphylothermus_marinus_F1*
- *Hyperthermus_butylicus*
- *Metallosphaera_sedula_DSM_5348*
- *Sulfolobus_solfataricus*
- *Thermofilum_pendens_Hrk_5*
- *Caldivirga_maquilingensis_IC-167*
- *Pyrobaculum_aerophilum*

Στην κατηγορία Euryarchaeota:

- *Halobacterium_sp*
- *Natronomonas_pharaonis*
- *Methanobrevibacter_smithii_ATCC_35061*
- *Methanosphaera_stadtmanae*
- *Methanothermobacter_thermautotrophicus*
- *Methanocaldococcus_jannaschii*

- *Methanococcus_aeolicus_Nankai-3*
- *Methanocorpusculum_labreanum_Z*
- *Methanoculleus_marisnigri_JR1*
- *Methanospirillum_hungatei_JF-1*
- *Methanosaeta_thermophila_PT*
- *Methanococcoides_burtonii_DSM_6242*
- *Methanosarcina_acetivorans_C2A*
- *Methanopyrus_kandleri_AV19*
- *Pyrococcus_abyssi*
- *Thermococcus_kodakarensis_KOD1*
- *Picrophilus_torridus_DSM_9790*
- *Thermoplasma_volcanium*

Στην κατηγορία Nanoarchaeota:

- *Nanoarchaeum_equitans*

Για τις δύο κατηγορίες των βακτηρίων επιλέχτηκαν:

Για την κατηγορία των gram-positive:

- *Acidobacteria_bacterium_Ellin34*
- *Acidothermus_cellulolyticus_11B*
- *Alkaliphilus_metalliredigens_QYMF*
- *Bacillus_subtilis*
- *Bifidobacterium_longum*
- *Caldicellulosiruptor_saccharolyticus_DSM_8903*
- *Carboxydothemus_hydrogenoformans_Z-2901*
- *Clavibacter_michiganensis_NCPPB_382*
- *Clostridium_tetani_E88*
- *Corynebacterium_diphtheriae*
- *Dehalococcoides_ethenogenes_195*
- *Deinococcus_radiodurans*
- *Desulfitobacterium_hafniense_Y51*
- *Desulfotomaculum_reducens_MI-1*
- *Frankia_alni_ACN14a*
- *Lactococcus_lactis*

- *Leifsonia_xyli*_CTCB0
- *Listeria_monocytogenes*
- *Mesoplasma_florum*_L1
- *Moorella_thermoacetica*_ATCC_39073
- *Mycobacterium_bovis*
- *Mycoplasma_genitalium*
- *Oceanobacillus_iheyensis*
- *Oenococcus_oeni*_PSU-1
- *Onion_yellow*s_phytoplasma
- *Pediococcus_pentosaceus*_ATCC_25745
- *Pelotomaculum_thermopropionicum*_SI
- *Propionibacterium_acnes*_KPA171202
- *Renibacterium_salmoninarum*_ATCC_33209
- *Rhodococcus_RHA1*
- *Rubrobacter_xylanophilus*_DSM_9941
- *Saccharopolyspora_erythraea*_NRRL_2338
- *Salinispora_tropica*_CNB-440
- *Staphylococcus_aureus*_COL
- *Streptococcus_mutans*
- *Symbiobacterium_thermophilum*_IAM14863
- *Syntrophomonas_wolfei*_Goettingen
- *Thermoanaerobacter_tengcongensis*
- *Tropheryma_whipplei*_Twist
- *Ureaplasma_urealyticum*

Για την κατηγορία των gram-negative:

- *Acaryochloris_marina*_MBIC11017
- *Acidiphilium_cryptum*_JF-5
- *Acidovorax_avenae_citrulli*_AAC00-1
- *Acinetobacter_baumannii*
- *Actinobacillus_succinogenes*_130Z
- *Aeromonas_salmonicida*_A449
- *Agrobacterium_tumefaciens*_C58_Cereon
- *Anabaena_variabilis*_ATCC_29413

- *Anaplasma marginale*_St_Maries
- *Aquifex aeolicus*
- *Azorhizobium caulinodans*_ORS_571
- *Bacteroides fragilis*_NCTC_9434
- *Bacteroides thetaiotaomicron*_VPI-5482
- *Bartonella quintana*_Toulouse
- *Bdellovibrio bacteriovorus*
- *Bordetella pertussis*
- *Borrelia burgdorferi*
- *Bradyrhizobium BTAi1*
- *Brucella abortus*_9-941
- *Brucella melitensis*
- *Buchnera aphidicola*_APS
- *Burkholderia cepacia*_AMMD
- *Campylobacter jejuni*
- *Candidatus Blochmannia floridanus*
- *Candidatus Sulcia muelleri*_GWSS
- *Caulobacter crescentus*
- *Chlamydia muridarum*
- *Chlamydia trachomatis*
- *Chlamydophila caviae*
- *Chlamydophila pneumoniae*_AR39
- *Chlorobium chlorochromatii*_CaD3
- *Chromobacterium violaceum*
- *Chromohalobacter salexigens*_DSM_3043
- *Citrobacter koseri*_ATCC_BAA-895
- *Coxiella burnetii*
- *Dechloromonas aromatica*_RCB
- *Delftia acidovorans*_SPH-1
- *Desulfovibrio vulgaris*_DP4
- *Ehrlichia ruminantium*_Gardel
- *Erwinia carotovora atroseptica*_SCRI1043
- *Erythrobacter litoralis*_HTCC2594
- *Escherichia coli*_O157H7

- *Fervidobacterium_nodosum*_Rt17-B1
- *Flavobacterium_johnsoniae*_UW101
- *Francisella_tularensis_tularensis*
- *Geobacter_sulfurreducens*
- *Gloeobacter_violaceus*
- *Gluconacetobacter_diazotrophicus_PA1_5*
- *Gramella_forsetii_KT0803*
- *Granulobacter_bethesdensis_CGDNIH1*
- *Haemophilus_influenzae*
- *Hahella_chejuensis_KCTC_2396*
- *Halorhodospira_halophila_SL1*
- *Helicobacter_pylori_J99*
- *Herminiimonas_arsenicoxydans*
- *Idiomarina_loihiensis_L2TR*
- *Janthinobacterium_Marseille*
- *Klebsiella_pneumoniae_MGH_78578*
- *Legionella_pneumophila_Corby*
- *Leptospira_interrogans_serovar_Lai*
- *Leptothrix_cholodnii_SP_6*
- *Magnetococcus_MC-1*
- *Mannheimia_succiniciproducens_MBEL55E*
- *Marinobacter_aquaeolei_VT8*
- *Marinomonas_MWYL1*
- *Mesorhizobium_BNC1*
- *Methylibium_petroleiphilum_PM1*
- *Methylobacillus_flagellatus_KT*
- *Methylobacterium_extorquens_PA1*
- *Methylococcus_capsulatus_Bath*
- *Microcystis_aeruginosa_NIES_843*
- *Microcystis_aeruginosa_NIES_843*
- *Neisseria_meningitidis_MC58*
- *Neorickettsia_sennetsu_Miyayama*
- *Nitratiruptor_SB155-2*
- *Nitrobacter_hamburgensis_X14*

- *Nitrosococcus_oceani*_ATCC_19707
- *Nitrosomonas_europaea*
- *Nitrospira_multiformis*_ATCC_25196
- *Nostoc_sp*
- *Parachlamydia_sp_UWE25*
- *Paracoccus_denitrificans*_PD1222
- *Pasteurella_multocida*
- *Pelobacter_carbinolicus*
- *Pelodictyon_luteolum*_DSM_273
- *Petrotoga_mobilis*_SJ95
- *Photobacterium_profundum*_SS9
- *Photorhabdus_luminescens*
- *Polynucleobacter_necessarius*_STIR1
- *Porphyromonas_gingivalis*_W83
- *Prochlorococcus_marinus*_AS9601
- *Prosthecochloris_vibrioformis*_DSM_265
- *Pseudoalteromonas_atlantica*_T6c
- *Pseudomonas_aeruginosa*
- *Psychrobacter_arcticum*_273-4
- *Ralstonia_eutropha*_H16
- *Rhizobium_leguminosarum_bv_viciae*_3841
- *Rhodobacter_sphaeroides*_ATCC_17025
- *Rhodoferax_ferrireducens*_DSM_15236
- *Rhodopseudomonas_palustris*_BisA53
- *Rhodospirillum_rubrum*_ATCC_11170
- *Rickettsia_prowazekii*
- *Roseobacter_denitrificans*_OCh_114
- *Salmonella_typhi*
- *Serratia_proteamaculans*_568
- *Shewanella_oneidensis*
- *Shigella_sonnei*_Ss046
- *Silicibacter_pomeroyi*_DSS-3
- *Sinorhizobium_meliloti*
- *Sodalis_glossinidius_morsitans*

- *Sorangium_cellulosum__So_ce_56_*
- *Sphingomonas_wittichii_RW1*
- *Sphingopyxis_alaskensis_RB2256*
- *Sulfurovum_NBC37-1*
- *Synechococcus_elongatus_PCC_6301*
- *Synechocystis_PCC_6803*
- *Syntrophobacter_fumaroxidans_MPOB*
- *Thermosipho_melanesiensis_BI429*
- *Thermotoga_maritima*
- *Thermus_thermophilus_HB27*
- *Thiomicrospira_crunogena_XCL-2*
- *Thiomicrospira_denitrificans_ATCC_33889*
- *Treponema_pallidum*
- *Vibrio_cholerae*
- *Wigglesworthia_brevipalpis*
- *Wolbachia_endosymbiont_of_Drosophila_melanogaster*
- *Wolinella_succinogenes*
- *Xanthomonas_campestris*
- *Xylella_fastidiosa*
- *Yersinia_enterocolitica_8081*
- *Zymomonas_mobilis_ZM4*

Τα δεδομένα για να εξεταστούν επιτυχώς έπρεπε να είναι σε μορφή FASTA. Για το λόγο αυτό, έχοντας συγκεντρώσει τους οργανισμούς που θα εξεταστούν, για κάθε ένα από αυτούς τρέξαμε το πρόγραμμα σε περιβάλλον perl και αποθηκεύσαμε τα αποτελέσματα σε αρχεία μορφής κειμένου με κατάληξη txt. Στη συνέχεια τοποθετήσαμε όλα τα αρχεία στο ίδιο directory και στην γραμμή εντολών του προγράμματος δώσαμε την εντολή `cat *.txt > results`, ώστε να εμφανίζονται όλα τα αποτελέσματα των πρωτεϊνών σε ένα ενιαίο αρχείο για κάθε κατηγορία. Στα τελικά αυτά αρχεία παρουσιάζονται το όνομα της πρωτεΐνης, το μήκος της, οι διαφορετικές κινάσες και ο αριθμός εμφάνισής τους σε κάθε πρωτεΐνη, το όνομα του οργανισμού (organism name), η κατηγορία (eukarya, bacteria, archaea), το φύλο (phylum), το είδος (kingdom-για κάποιες από τις πρωτεΐνες αυτές) και μία επιπλέον κατηγορία που

αφορά τα βακτήρια, το `gram`, η οποία διαμορφώνεται σε `gram_negative` και `gram_positive`.

3.2.5 Εξαγωγή τελικών αποτελεσμάτων με χρήση του προγράμματος Stata

Αφού «τρέξαμε» τις πρωτεΐνες ήταν απαραίτητο να ενώσουμε τα αρχεία μας σε ένα μεγάλο αρχείο το οποίο θα περιελάμβανε τα συνολικά αποτελέσματα. Για το λόγο αυτό χρησιμοποιήσαμε το πρόγραμμα μετά-ανάλυσης Stata. Στο πρόγραμμα αυτό εισήγαμε προς επεξεργασία ένα συνολικό αρχείο μεγάλου μεγέθους το οποίο περιελάμβανε όλες τις πρωτεΐνες που ελέχθησαν και επεξεργάστηκαν με το πρόγραμμα `program.pl`, για τους ευκαρυώτες, τα αρχαία και τα βακτήρια και το οποίο ονομάσαμε `genomes.dta`. Κατά το άνοιγμα του προγράμματος Stata, χρησιμοποιήθηκε η εντολή `set mem 300m` έτσι ώστε να υπάρχει μεγάλη και επαρκής μνήμη στο πρόγραμμα για την ανάλυση του μεγάλου όγκου των δεδομένων.

Χρησιμοποιήσαμε Poisson Regression, έτσι ώστε να αντιστοιχίσουμε τη μέγιστη πιθανότητα εμφάνισης των κινασών μέσα στο μοντέλο. Πιο αναλυτικά, κάναμε συγκρίσεις στα δεδομένα και εξετάσαμε το ποσοστό κάθε κινάσης σε κάθε μία από τις κατηγορίες phylum και σε σχέση με το phylum “CHORDATA” που αντιστοιχεί στον άνθρωπο, καθώς και το σύνολο των κινασών που υπάρχουν σε αρχαία και βακτήρια έχοντας ως σημείο αναφοράς τους ευκαρυώτες.

3.2.6 Παλινδρόμηση Poisson (Poisson Regression)

Στην στατιστική, η παλινδρόμηση Poisson είναι μια μορφή ανάλυσης που χρησιμοποιείται για να διαμορφώσει τα στοιχεία αρίθμησης και τους πίνακες πιθανότητας. Η παλινδρόμηση Poisson υποθέτει ότι η εξαρτημένη μεταβλητή Y έχει μια κατανομή Poisson και πως ο λογάριθμος της αναμενόμενης τιμής της μπορεί να διαμορφωθεί από έναν γραμμικό συνδυασμό άγνωστων παραμέτρων [107]. Ένα πρότυπο της παλινδρόμησης Poisson είναι μερικές φορές γνωστό ως λογαριθμικό-γραμμικό πρότυπο (*log-linear model*), ειδικά όταν χρησιμοποιείται για να διαμορφώσει τους πίνακες των πιθανοτήτων. Στην πιο απλή περίπτωση με μία ανεξάρτητη μεταβλητή X , το μοντέλο παίρνει τη μορφή:

$$\log(E(Y))=a+bx+\log(L)$$

$$\Leftrightarrow \log(E(Y))-\log(L)=a+bx$$

$$\Leftrightarrow a+bx=\log\left[\frac{E(Y)}{L}\right]$$

Εάν οι Y_i είναι ανεξάρτητες παρατηρήσεις με τις αντίστοιχες τιμές x_i της μεταβλητής πρόβλεψης, τότε το a και το b μπορούν να υπολογιστούν από τη μέγιστη πιθανότητα, εάν ο αριθμός των ευδιάκριτων x τιμών είναι τουλάχιστον 2. Τα μοντέλα οπισθοδρόμησης Poisson είναι γενικευμένα γραμμικά μοντέλα που χρησιμοποιούν λογάριθμο ως (κανονική) λειτουργία των συνδέσεων και τη λειτουργία της κατανομής Poisson. Η παλινδρόμηση Poisson είναι κατάλληλη όταν η εξαρτώμενη μεταβλητή αποτελεί αρίθμηση, για μία σειρά γεγονότων όπως είναι η άφιξη ενός τηλεφωνήματος σε ένα κέντρο κλήσης. Τα γεγονότα πρέπει να είναι ανεξάρτητα υπό την έννοια ότι η άφιξη μιας κλήσης δεν θα κάνει μία άλλη κλήση λιγότερο ή περισσότερο πιθανή. Η παλινδρόμηση Poisson είναι επίσης κατάλληλη για τα στοιχεία σε μορφή ποσοστού, όπου το ποσοστό είναι η αρίθμηση των γεγονότων που εμφανίζονται σε μια ιδιαίτερη μονάδα παρατήρησης. Για παράδειγμα, οι βιολόγοι μπορούν να μετρήσουν τον αριθμό των διαφόρων ειδών δέντρων σε ένα δάσος και το ποσοστό που θα βρουν θα ήταν ο αριθμός των διαφόρων ειδών ανά τετραγωνικό χιλιόμετρο. Οι δημογράφοι μπορούν να διαμορφώσουν τα ποσοστά θανάτου στις γεωγραφικές περιοχές ως μία αρίθμηση των θανάτων που διαιρούνται με βάση τα στοιχεία άτομα-ανά έτη. Γενικότερα, τα ποσοστά ενός γεγονότος μπορούν να υπολογιστούν ως γεγονότα ανά μονάδα χρόνου, ο οποίος επιτρέπει στο παράθυρο παρατήρησης να ποικίλει για κάθε μονάδα. Στην παλινδρόμηση Poisson αυτό αντιμετωπίζεται ως ένα offset, όπου η μεταβλητή έκθεσης εισάγεται στη δεξιά πλευρά της εξίσωσης, αλλά με μια παράμετρο εκτίμησης που περιορίζεται στο 1.

Στην πτυχιακή αυτή εργασία εφαρμόσαμε παλινδρόμηση Poisson για να συγκρίνουμε το σύνολο των θέσεων φωσφορυλίωσης (sum) για τις κατηγορίες των Αρχαίων και των Βακτηρίων, με μέτρο σύγκρισης τους Ευκαρυώτες, καθώς και για τη σύγκριση του αριθμού των θέσεων φωσφορυλίωσης στα φύλλα των Ευκαρυωτών, Βακτηρίων και Αρχαίων έχοντας ως σημείο σύγκρισης το φύλο CHORDATA, που αντιστοιχεί στον ανθρώπινο οργανισμό (human).

3.3 Εφαρμογή της Poisson Regression

Αρχικά εκτελέσαμε παλινδρόμηση Poisson ώστε να συγκρίνουμε τα ποσοστά των θέσεων φωσφορυλίωσης για τα Βακτήρια και τα Αρχαία έχοντας ως μέτρο σύγκρισης τους Ευκαρυώτες, καθώς είναι ευρέως διαδεδομένο ότι η φωσφορυλίωση είναι μία μετά-μεταφραστική τροποποίηση που υφίσταται κατά κύριο λόγο στους ευκαρυωτικούς οργανισμούς. Χρησιμοποιήσαμε την εντολή `sort category phylum organism` η οποία ταξινομεί τις πρωτεΐνες ως προς κατηγορία (category), φύλο (phylum) και οργανισμό (organism). Στη συνέχεια δώσαμε την εντολή `char category[omit] "EUKARYA"` ώστε η σύγκριση να γίνει με σημείο αναφοράς τους Ευκαρυώτες και τέλος την εντολή παλινδρόμησης `xi:poisson sum i.category, exposure (length) irr.`

3.4 Αποτελέσματα

Κατά την εφαρμογή της εντολής παλινδρόμησης Poisson του αθροίσματος (sum) με την κατηγορία (category), όπως παρουσιάζεται στα παρακάτω αποτελέσματα (Πίνακας 3.1), ορίζεται ως category_1 να αντιστοιχεί στα Αρχαία, ως category_2 στα Βακτήρια και ως category_3 στους Ευκαρυώτες. Παρατηρείται πως τα Αρχαία διαθέτουν στο σύνολό τους 20% μικρότερο ποσοστό θέσεων φωσφορυλίωσης από τους Ευκαρυώτες, ενώ το ποσοστό για τα Βακτήρια είναι περίπου 25% μικρότερο από ότι στους Ευκαρυωτικούς οργανισμούς. Τα αποτελέσματα αυτά ήταν αναμενόμενα καθώς γνωρίζουμε πως η φωσφορυλίωση είναι μία μετά-μεταφραστική τροποποίηση που συναντάται κυρίως στους ευκαρυώτες και συνεπώς οι οργανισμοί που ανήκουν στην κατηγορία αυτή θα διαθέτουν τον μεγαλύτερο αριθμό θέσεων φωσφορυλίωσης.

Πίνακας 3.1: Αποτελέσματα εφαρμογής παλινδρόμησης του αθροίσματος με την κατηγορία

```
. xi:poisson sum i.category, exposure (length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood =  -18214576
Iteration 1:    log likelihood =  -18214554
Iteration 2:    log likelihood =  -18214554
```


Poisson regression	Number of obs	=	1020130
LR chi2(2)	=	703338.81	
Prob > chi2	=	0.0000	
Log likelihood = -18214554	Pseudo R2	=	0.0189

sum	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----					
_Icategory_1	.8000999	.0007108	-251.03	0.000	.7987079 .8014943
_Icategory_2	.746649	.0002636	-827.65	0.000	.7461326 .7471657
length (exposure)					

Στη συνέχεια θέλαμε να πραγματοποιήσουμε έλεγχο του συνόλου των κινασών ως προς το φύλο. Για να πραγματοποιηθεί αυτό στο πρόγραμμα Stata, έπρεπε να οριστούν με κάποιο αποδεκτό από το πρόγραμμα τρόπο τα διαφορετικά είδη των φύλων που υπάρχουν. Αυτό έγινε εφικτό με μία σειρά εντολών οι οποίες είναι:

```
. gen cat_ph=0
. replace cat_ph=1 if phylum=="CYANOBACTERIA"

. replace cat_ph=2 if phylum=="PROTEOBACTERIA"

. replace cat_ph=3 if phylum=="PROTEOBACTERIA-BETA"

. replace cat_ph=4 if phylum=="PROTEOBACTERIA-GAMMA"

. replace cat_ph=5 if phylum=="PROTEOBACTERIA-ALPHA"

. replace cat_ph=6 if phylum=="AQUIFICAE"

. replace cat_ph=7 if phylum=="BACTEROIDETES"

. replace cat_ph=8 if phylum=="PROTEOBACTERIA-DELTA"

. replace cat_ph=9 if phylum=="PROTEOBACTERIA-EPSILON"

. replace cat_ph=10 if phylum=="CHLAMYDIAE"

. replace cat_ph=11 if phylum=="CHLOROBII"

. replace cat_ph=12 if phylum=="THERMOTOGAE"

. replace cat_ph=13 if phylum=="SPIROCHAETES"
```

```

. replace cat_ph=14 if phylum=="THERMI"

. replace cat_ph=15 if phylum=="ACTINOBACTERIA"

. replace cat_ph=16 if phylum=="FIRMICUTES"

. replace cat_ph=17 if phylum=="CHLOROFLEXI"

. replace cat_ph=18 if phylum=="DEINOCOCCI"

. replace cat_ph=19 if phylum=="TENERICUTES"

. replace cat_ph=20 if phylum=="ACIDOBACTERIA"

. replace cat_ph=21 if phylum=="ARTHROPODA"

. replace cat_ph=22 if phylum=="STREPTOPHYTA"

. replace cat_ph=23 if phylum=="ASCOMYCOTA"

. replace cat_ph=24 if phylum=="NEMATODA"

. replace cat_ph=25 if phylum=="APICOMPLEXA"

. replace cat_ph=26 if phylum=="MICROSPORIDIA"

. replace cat_ph=27 if phylum=="CHLOROPHYTA"

. replace cat_ph=28 if phylum=="CRENARCHAEOTA"

. replace cat_ph=29 if phylum=="EURYARCHAEOTA"

. replace cat_ph=30 if phylum=="NANOARCHAEOTA"

```

Έπειτα εκτελέστηκε η παλινδρόμηση έχοντας ως σημείο σύγκρισης των φύλο CHORDATA που αντιστοιχεί στον άνθρωπο. Όπως παρουσιάζεται στον πίνακα αποτελεσμάτων (Πίνακας 3.2), τα cat_ph από 1 έως 20 αντιστοιχούν σε βακτήρια, από 21 έως 27 σε Ευκαρυώτες και από 28 έως 30 σε Αρχαία. Παρατηρείται πως για τα φύλα που αντιστοιχούν σε ευκαρυώτες υπάρχουν μικρότερες αποκλίσεις σε σχέση με τον άνθρωπο. Συγκεκριμένα για τα φύλα ARTHROPODA και ASCOMYCOTA παρατηρείται πως έχουν 1% και 4% αντίστοιχα, μεγαλύτερο συνολικό αριθμό θέσεων φωσφορυλίωσης από ότι ο άνθρωπος. Τα φύλα CRENARCHAEOTA, EURYARCHAEOTA και NANOARCHAEOTA που ανήκουν στην κατηγορία των αρχαίων, εμφανίζουν 23%, 19% και 28% μικρότερο αριθμό θέσεων φωσφορυλίωσης

σε σύγκριση με τον άνθρωπο. Για την κατηγορία των βακτηρίων, παρατηρείται πως τα ποσοστά των θέσεων φωσφορυλίωσης είναι στο σύνολό τους μικρότερα από ότι στον άνθρωπο, ενώ στην κατηγορία αυτή σημαντικό ενδιαφέρον παρουσιάζουν τα φύλα PROTEOBACTERIA, PROTEOBACTERIA BETA, PROTEOBACTERIA-ALPHA, AQUIFICAE, THERMI και DEINOCOCI καθώς τα ποσοστά που εμφανίζουν είναι μικρότερα και από αυτά των αρχαίων, ενώ κανείς θα περίμενε πως τα αρχαία θα ήταν αυτά με τις λιγότερες θέσεις φωσφορυλίωσης.

Πίνακας 3.2: Αποτελέσματα παλινδρόμησης Poisson με σημείο σύγκρισης τον άνθρωπο

```
. xi:poisson sum i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

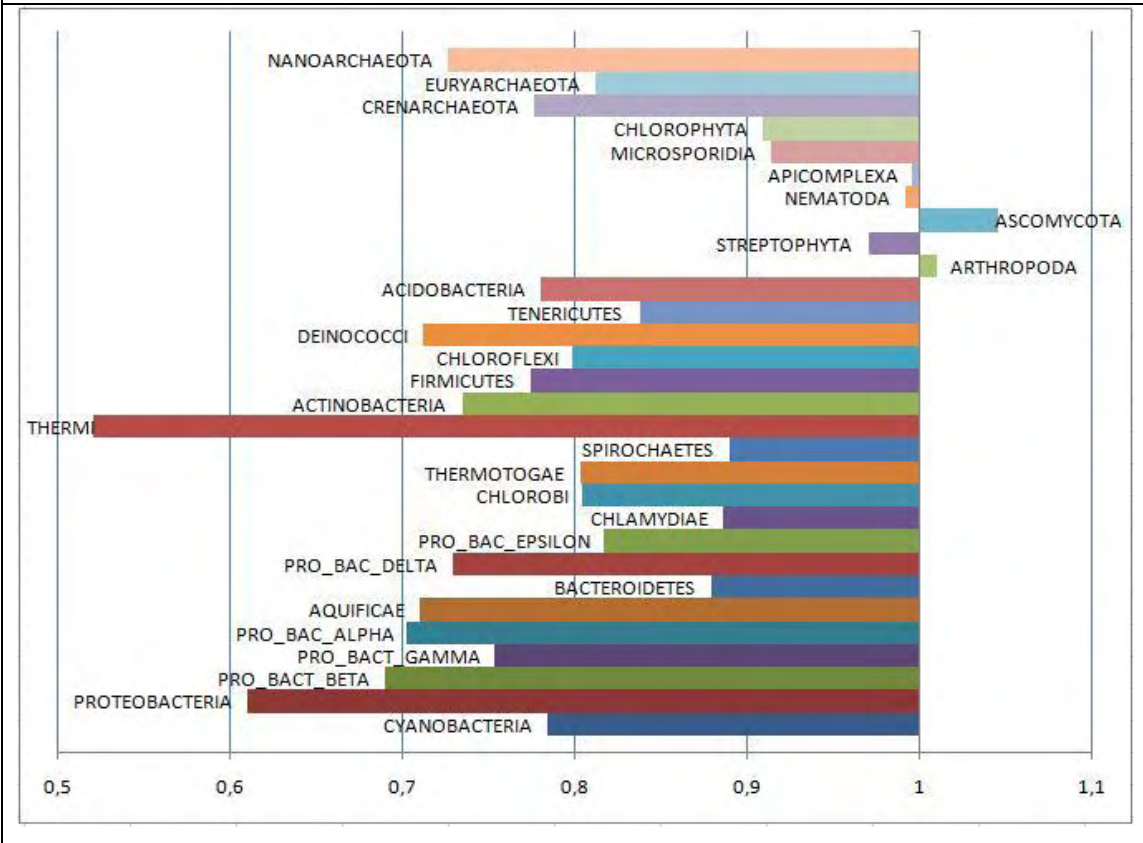
Iteration 0:   log likelihood =  -18176160
Iteration 1:   log likelihood =  -18175425
Iteration 2:   log likelihood =  -18175425

Poisson regression                                Number of obs   =   1020130
LR chi2(30)    =   781597.18
Prob > chi2    =         0.0000
Log likelihood =  -18175425                        Pseudo R2      =         0.0210

-----+-----
sum |          IRR   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   .7841511   .0009003  -211.78  0.000   .7823885   .7859176
_Icat_ph_2 |   .6101407   .0024122  -124.97  0.000   .6054312   .6148868
_Icat_ph_3 |   .6897187   .0006315  -405.75  0.000   .6884821   .6909574
_Icat_ph_4 |   .7538657   .0005033  -423.20  0.000   .7528799   .7548528
_Icat_ph_5 |   .7029623   .0005231  -473.68  0.000   .7019379   .7039882
_Icat_ph_6 |   .7098774   .0038021   -63.98  0.000   .7024643   .7173687
_Icat_ph_7 |   .8794006   .0011546   -97.88  0.000   .8771406   .8816665
_Icat_ph_8 |   .7290595   .0008127  -283.49  0.000   .7274685   .7306541
_Icat_ph_9 |   .8166007   .0015009  -110.23  0.000   .8136642   .8195478
_Icat_ph_10 |  .8859878   .0021147   -50.72  0.000   .8818528   .8901422
_Icat_ph_11 |  .8046788   .0020309   -86.10  0.000   .8007082   .8086691
_Icat_ph_12 |  .8032425   .0018339   -95.96  0.000   .799656   .806845
_Icat_ph_13 |  .8897541   .0022039   -47.16  0.000   .885445   .8940843
_Icat_ph_14 |  .5208706   .0029163  -116.50  0.000   .515186   .5266178
_Icat_ph_15 |  .7354534   .0006771  -333.72  0.000   .7341274   .7367818
_Icat_ph_16 |  .7747674   .000718   -275.36  0.000   .7733614   .776176
_Icat_ph_17 |  .7984042   .0042243   -42.55  0.000   .7901675   .8067269
_Icat_ph_18 |  .7123837   .002777   -87.00  0.000   .7069617   .7178473
_Icat_ph_19 |  .8377941   .0031156   -47.59  0.000   .8317098   .8439229
```

_Icat_ph_20		.780564	.0013107	-147.54	0.000	.7779993	.7831371
_Icat_ph_21		1.010232	.0008326	12.35	0.000	1.008601	1.011865
_Icat_ph_22		.9702405	.000711	-41.23	0.000	.968848	.9716349
_Icat_ph_23		1.045196	.000684	67.55	0.000	1.043856	1.046537
_Icat_ph_24		.9913104	.0008325	-10.39	0.000	.9896801	.9929434
_Icat_ph_25		.9956121	.0008931	-4.90	0.000	.9938632	.9973641
_Icat_ph_26		.9141087	.0036141	-22.71	0.000	.9070525	.9212198
_Icat_ph_27		.9090939	.0012932	-67.00	0.000	.9065627	.9116321
_Icat_ph_28		.7767916	.0012542	-156.44	0.000	.7743374	.7792537
_Icat_ph_29		.8121381	.0008892	-190.04	0.000	.8103971	.8138829
_Icat_ph_30		.7267452	.0068507	-33.86	0.000	.7134413	.7402971
length		(exposure)					

Εικόνα 3.1: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson με σημείο σύγκρισης τον άνθρωπο



Στη συνέχεια εκτελέστηκε παλινδρόμηση Poisson συγκρίνοντας το φύλο με κάθε μία κίνηση ξεχωριστά. Τα αποτελέσματα που εξήχθησαν για κάθε μία από τις κινήσεις παρουσιάζονται παρακάτω:

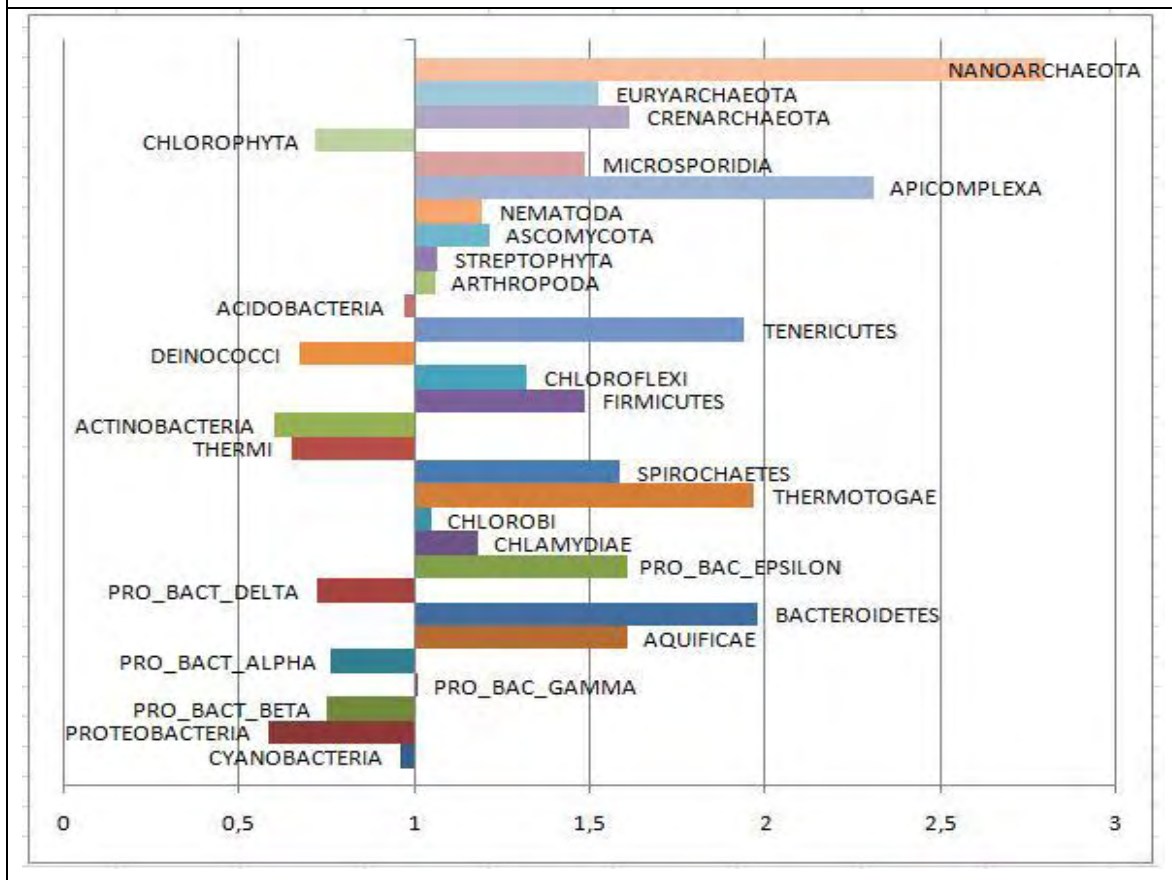
- **insr**

Για την κινάση insr (Πίνακας 3.3 παραρτήματος) οι σημαντικότερες διαφορές που παρατηρούνται για τους ευκαρυώτες είναι για τα φύλα APICOMPLEXA και CHLOROPHYTA, καθώς το πρώτο έχει αποτελέσματα της τάξης του 2,31 ενώ το δεύτερο περιλαμβάνει 28% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τον άνθρωπο. Τα βακτήρια εμφανίζουν υψηλές θέσεις για τα φύλα THERMOTOGAE και BACTEROIDETES, της τάξης του 96% και 97% αντίστοιχα. Ωστόσο, εμφανίζουν και πολύ μικρό αριθμό θέσεων για τα φύλα PROTEOBACTERIA, ACTINOBACTERIA, THERMI και DEINOCOCCI της τάξης του 42%, 40%, 35% και 33% αντίστοιχα. Αξιοσημείωτες είναι οι υψηλές τιμές που παρατηρούνται για τα αρχαία. Συγκεκριμένα για τα φύλα EURYARCHAEOTA, CRENARCHAEOTA και NANOARCHAEOTA το ποσοστό θέσεων είναι της τάξης του 52%, 39% και 2,79 αντίστοιχα μεγαλύτερο από ότι στον άνθρωπο. Παρατηρείται πως για το φύλο NANOARCHAEOTA τα αρχαία εμφανίζουν τον υψηλότερο συνολικά αριθμό θέσεων από τους ευκαρυώτες και τα βακτήρια. Ο μικρότερος συνολικά αριθμός θέσεων ανήκει στα βακτήρια και συγκεκριμένα στο φύλο PROTEOBACTERIA με ποσοστό 42% λιγότερες θέσεις από ότι ο άνθρωπος.

- **egfr**

Για την κινάση egfr (Πίνακας 3.4 παραρτήματος) και για τα τρία φύλα των αρχαίων υπάρχουν υψηλές θέσεις φωσφορυλίωσης, έχοντας για το φύλο NANOARCHAEOTA τον μεγαλύτερο αριθμό θέσεων που παρατηρείται για το σύνολο των πρωτεϊνών, της τάξεως του 88%. Για τους ευκαρυώτες υπάρχουν οι μικρότερες τιμές στα φύλα CHLOROPHYTA και STREPTOPHYTA που αντιστοιχούν σε ποσοστά 17% και 7% αντίστοιχα, ενώ το μεγαλύτερο ποσοστό θέσεων συναντάται στο φύλο APICOMPLEXA με τιμή 22%. Ο μεγαλύτερος αριθμός θέσεων για τα βακτήρια παρατηρείται στο φύλο BACTEROIDETES με ποσοστό 55%, ενώ ο μικρότερος συνολικός αριθμός όλων των πρωτεϊνών εμφανίζεται στο φύλο PROTEOBACTERIA των βακτηρίων με τιμές 29%.

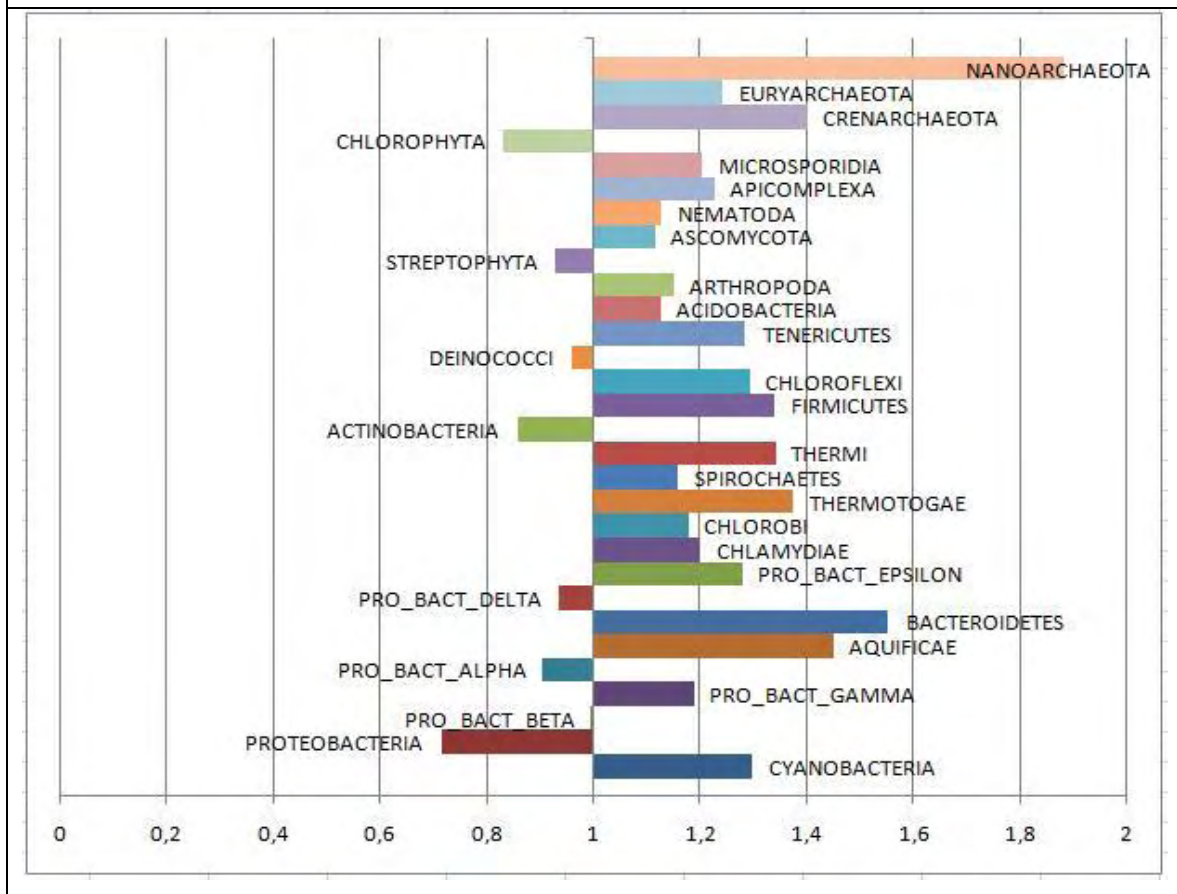
Εικόνα 3.2: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση insr



- **cki**

Για την κινάση cki (Πίνακας 3.5 παραρτήματος) τα αρχαία εμφανίζουν μικρότερο αριθμό θέσεων φωσφορυλίωσης από τον άνθρωπο, στο σύνολό τους. Το γεγονός αυτό παρατηρείται και στους ευκαρυώτες με μόνες εξαιρέσεις τα φύλα ARTHROPODA και ASCOMYCOTA που παρουσιάζουν 3% περισσότερες θέσεις. Στα βακτήρια υψηλό ποσοστό θέσεων εμφανίζεται μόνο για το φύλο CHLOROBI ποσοστό το οποίο είναι ίσο με το ποσοστό στον άνθρωπο, ενώ για το φύλο THERMI εμφανίζεται το μικρότερο συνολικό ποσοστό της τάξης του 49%.

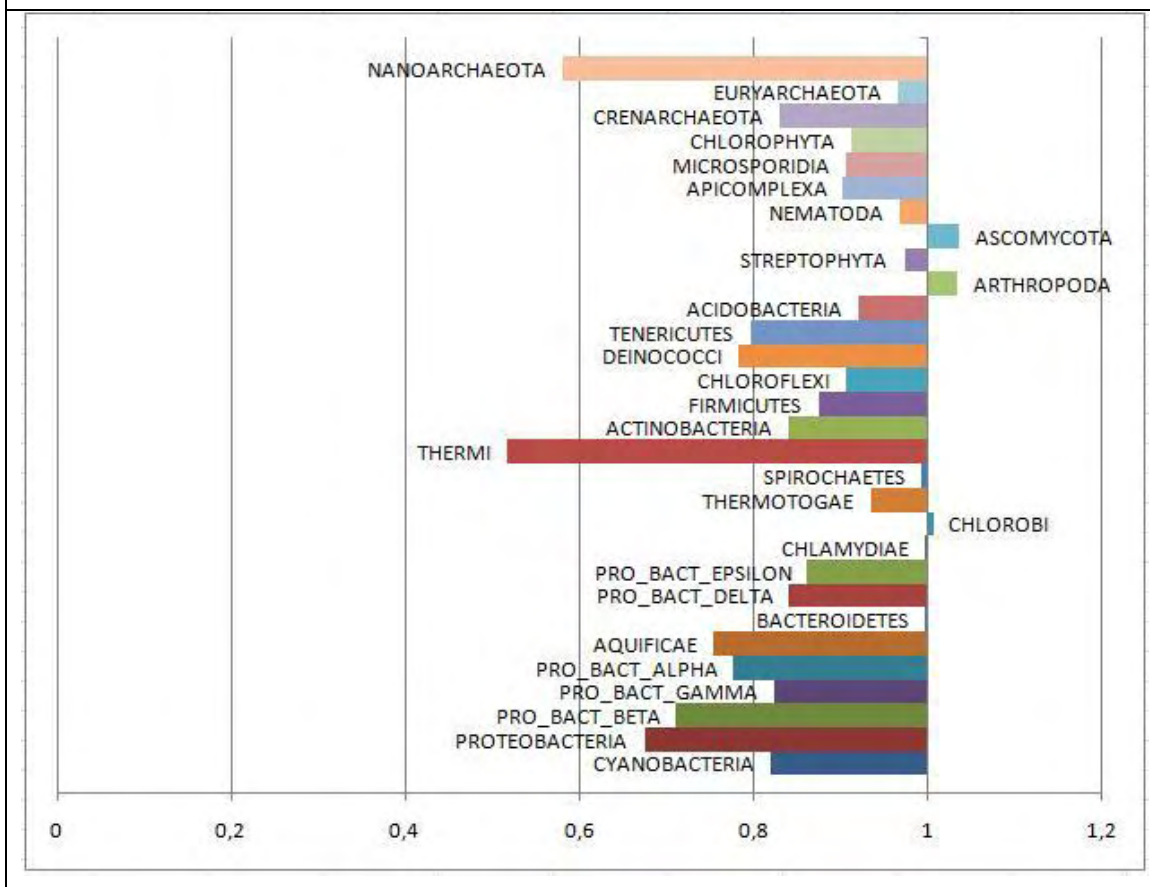
Εικόνα 3.3: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση egfr



- **ckii**

Για την κινάση ckii (Πίνακας 3.6 παραρτήματος) οι περισσότερες θέσεις φωσφορυλίωσης παρατηρούνται στους ευκαρυώτες στο φύλο APICOMPLEXA με ποσοστό 30% περισσότερο από ότι στον άνθρωπο και ακολουθεί το φύλο EURYARCHAEOTA των αρχαίων με ποσοστό 27%. Οι περισσότερες συγκριτικά θέσεις όσον αφορά τα βακτήρια βρίσκονται στο φύλο BACTEROIDETES με ποσοστό 18%. Τα μικρότερα ποσοστά θέσεων παρατηρούνται στα βακτήρια. Συγκεκριμένα τα φύλα THERMI και PROTEOBACTERIA έχουν 52% και 43% λιγότερες θέσεις, σε σχέση με τον άνθρωπο, αντίστοιχα.

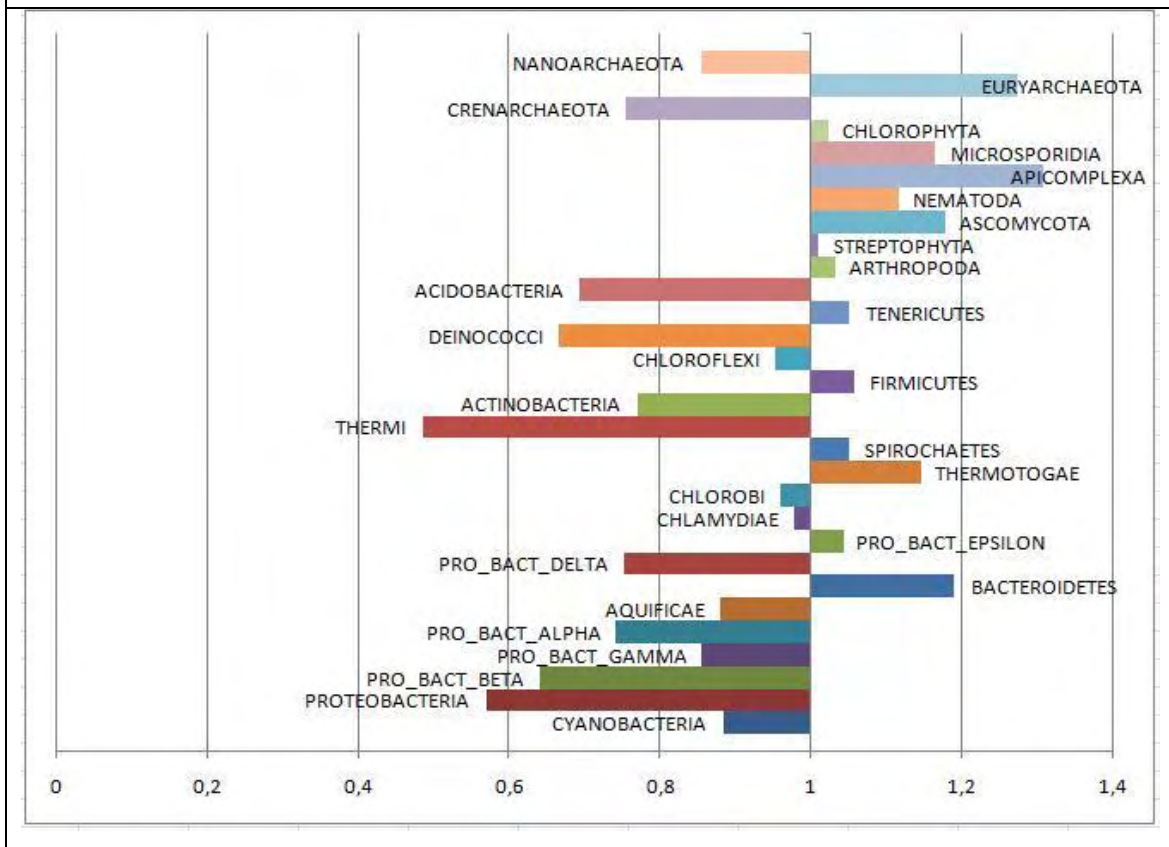
Εικόνα 3.4: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση cki



- **pka**

Για την κινάση pka (Πίνακας 3.7 παραρτήματος), τα βακτήρια για το φύλο CHLAMYDIAE παρουσιάζουν τον μεγαλύτερο αριθμό θέσεων φωσφορυλίωσης με ποσοστό 19% υψηλότερο από τον αριθμό των θέσεων στον άνθρωπο. Επίσης για το φύλο SPIROCHAETES έχουν τη δεύτερη σε σειρά μεγαλύτερη συγκέντρωση θέσεων με τιμή 15%. Ωστόσο τα βακτήρια για το φύλο THERMI παρουσιάζουν και το μικρότερο αριθμό θέσεων της τάξης του 26% μικρότερο από ότι στον οργανισμό HUMAN. Τα αρχαία για τα φύλα EURYARCHAEOTA και NANOARCHAEOTA έχουν λιγότερες θέσεις της τάξης του 15% περίπου, ενώ για το φύλο CRENARCHAEOTA έχουν 5% περισσότερες θέσεις από τον άνθρωπο. Οι ευκαρυώτες για τα φύλα MICROSPORIDIA και STREPTOPHYTA έχουν περισσότερες θέσεις σε ποσοστό 9% και 4% αντίστοιχα, ενώ στο φύλο CHLOROPHYTA παρατηρείται πως έχουν 18% λιγότερες θέσεις.

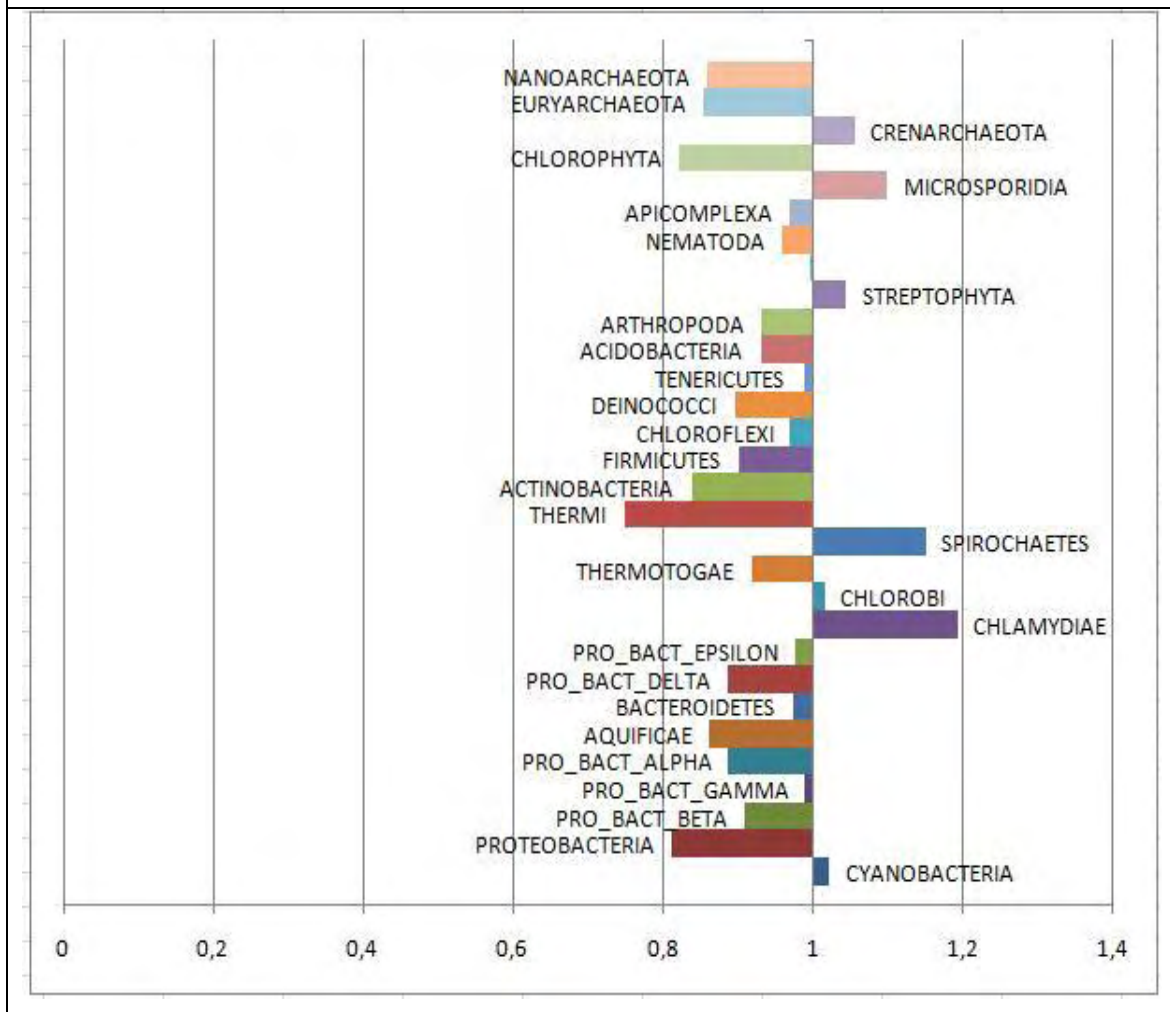
Εικόνα 3.5: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση ckii



- **pkg**

Για την κινάση pkg (Πίνακας 3.8 παραρτήματος), οι περισσότερες θέσεις φωσφορυλίωσης εμφανίζονται στα φύλα NEMATODA και ARTHROPODA των ευκαρυωτών με ποσοστά 4% και 3% αντίστοιχα, ενώ οι λιγότερες θέσεις, σε σχέση με τον άνθρωπο, εμφανίζονται για τα βακτήρια στο φύλο THERMI με ποσοστό 56%, καθώς και στα αρχαία στα φύλα CRENARCHAEOTA, EURYARCHAEOTA και NANOARCHAEOTA με ποσοστά 29%, 25% και 28% αντίστοιχα.

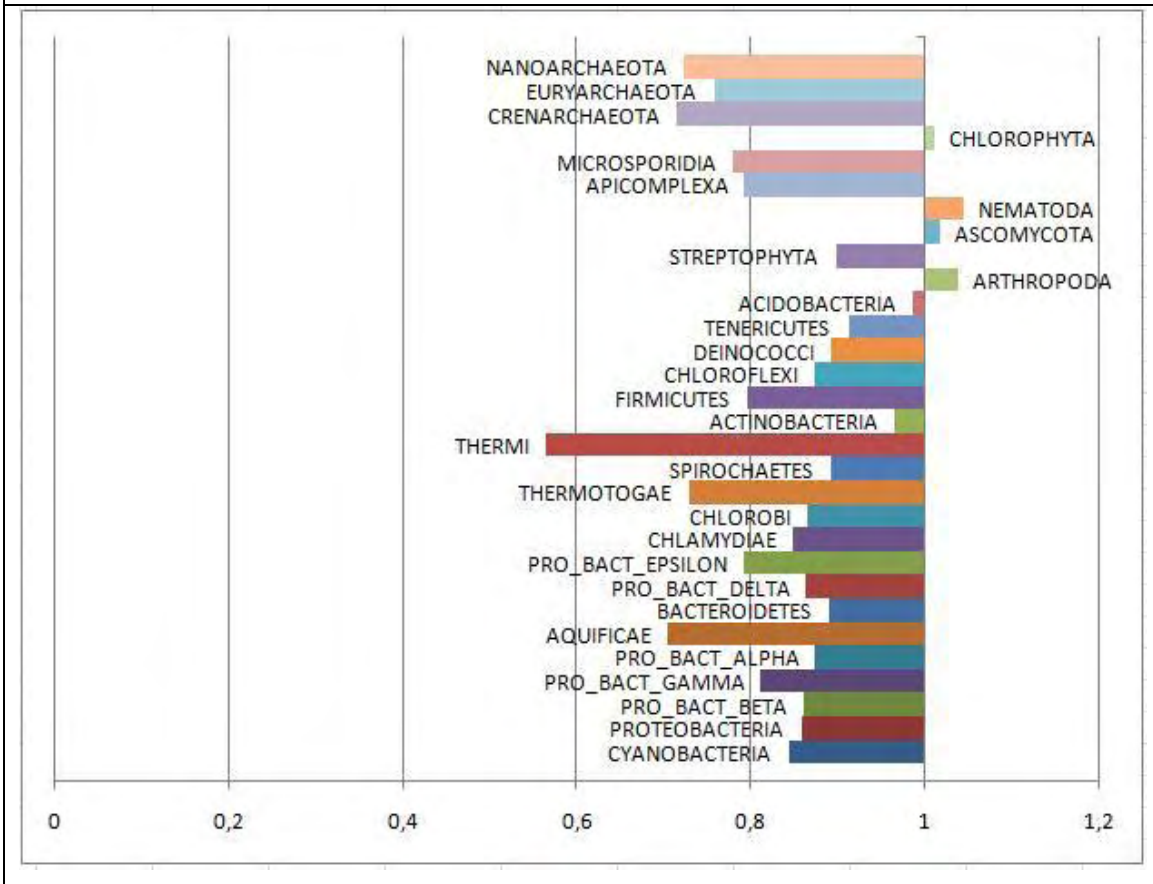
Εικόνα 3.6: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση ρka



- **ρka**

Για την κινάση ρka (Πίνακας 3.9 παραρτήματος), τα βακτήρια εμφανίζουν τον μεγαλύτερο αριθμό θέσεων φωσφορυλίωσης για το φύλο TENERICUTES και φτάνει σε ποσοστό 14%, ενώ στο φύλο THERMI εμφανίζουν τον μικρότερο αριθμό θέσεων από το σύνολο των πρωτεϊνών σε ποσοστό 57%. Οι ευκαρυώτες εμφανίζουν υψηλό αριθμό θέσεων στα φύλα ARTHROPODA, ASCOMYCOTA και APICOMPLEXA με τιμές 2%, 7% και 2% αντίστοιχα, ενώ στο φύλο MICROSPORIDIA έχουν τις λιγότερες θέσεις, από το σύνολο των ευκαρυωτών, με ποσοστό 13% λιγότερο από ότι στον άνθρωπο. Τα αρχαία εμφανίζουν χαμηλό αριθμό θέσεων με μικρότερο αυτό που αντιστοιχεί στο φύλο NANOARCHAEOTA, με τιμή 24%.

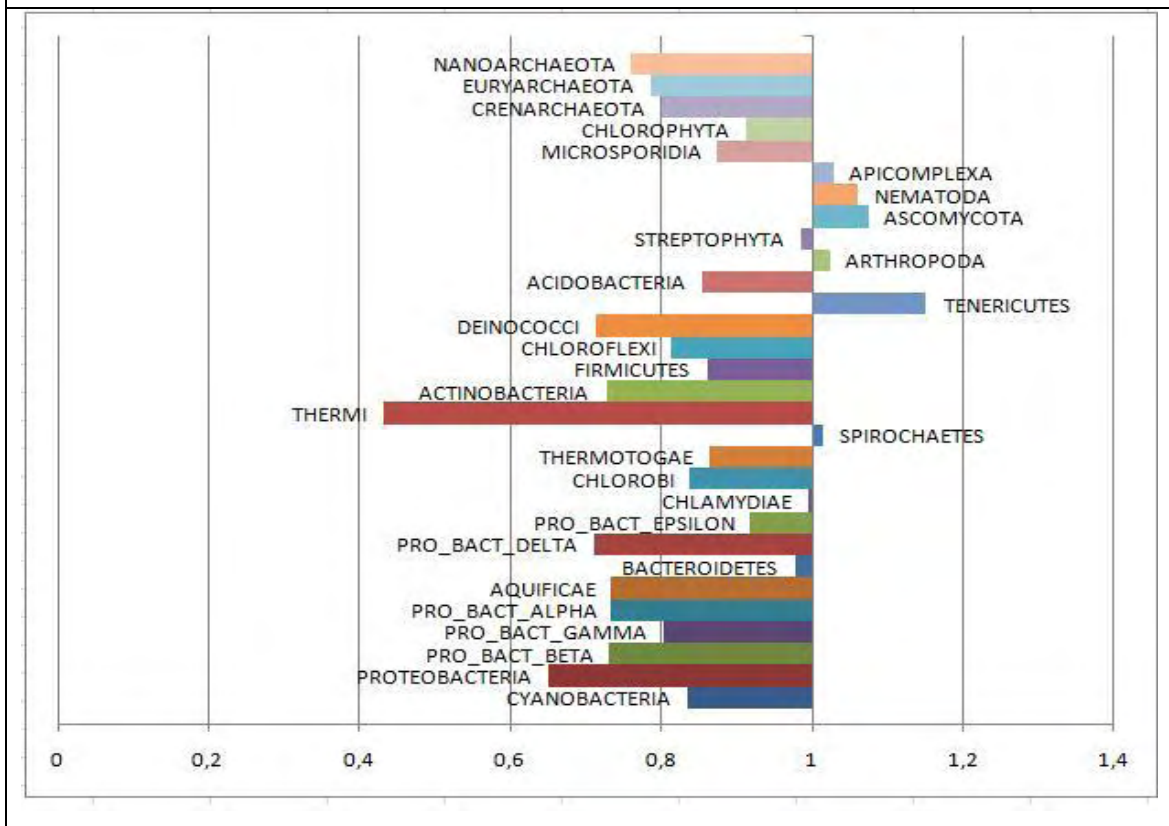
Εικόνα 3.7: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση ρkb



- **ρkb**

Για την κινάση ρkb (Πίνακας 3.10 παραρτήματος), το υψηλότερο ποσοστό θέσεων φωσφορυλίωσης παρουσιάζεται στο φύλο CHLOROPHYTA των ευκαρυωτών με ποσοστό 84%. Τα βακτήρια εμφανίζουν μεγάλο αριθμό θέσεων για το φύλο ACTINOBACTERI της τάξης του 40%, ενώ εμφανίζουν μικρό αριθμό θέσεων στα φύλα PROTEOBACTERIA_EPSILON, THERMOTOGAE και TENERICUTES της τάξης του 64%, 68% και 70% αντίστοιχα. Τα αρχαία εμφανίζουν τον μικρότερο αριθμό θέσεων από το σύνολο των πρωτεϊνών, στο φύλο NANOARCHAEOTA με ποσοστό 73%.

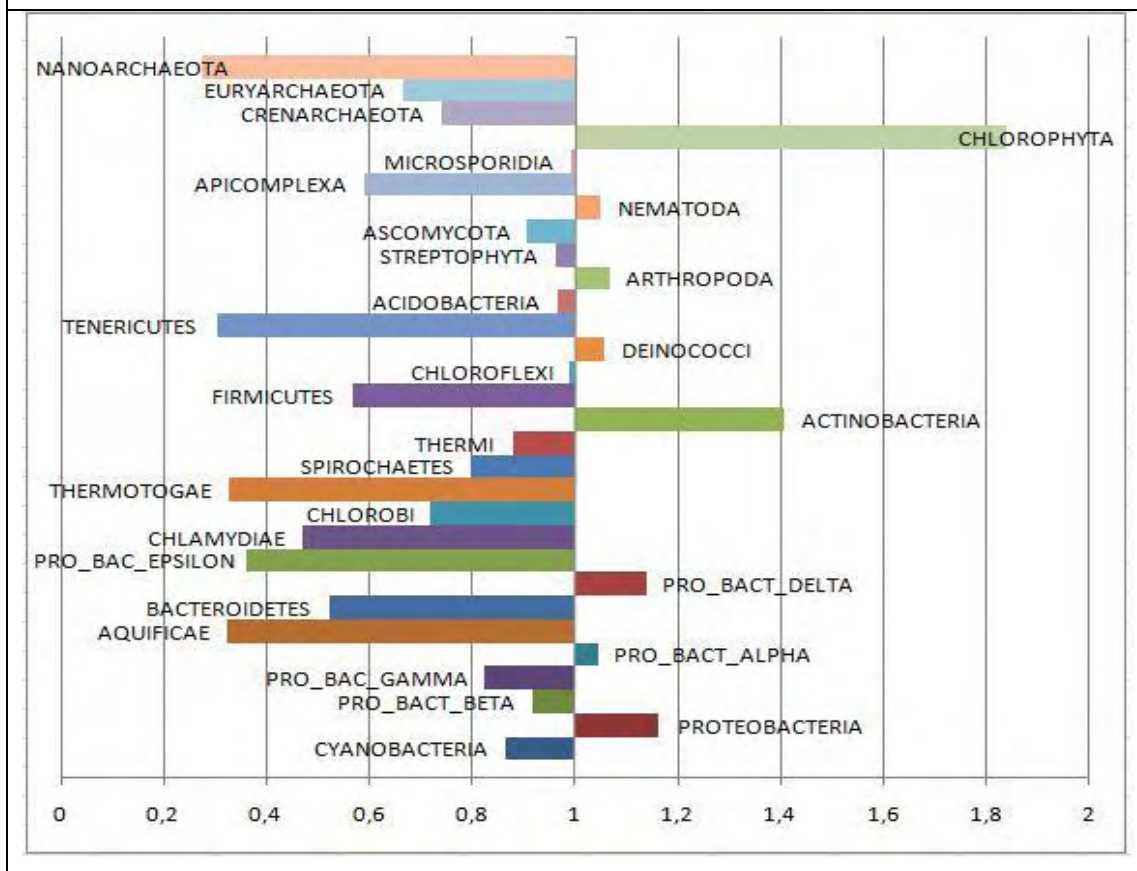
Εικόνα 3.8: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση ρκκ



- **dnark**

Για την κινάση dnark (Πίνακας 3.11 παραρτήματος), από το σύνολο των πρωτεϊνών μόνο για το φύλο CYANOBACTERIA των βακτηρίων παρατηρείται μεγαλύτερος αριθμός θέσεων φωσφορυλίωσης της τάξης του 6%, ενώ για τα φύλα AQUIFICAE και THERMI παρατηρούνται οι μικρότερες τιμές στο σύνολο σε ποσοστά 62% και 63% λιγότερο από τον άνθρωπο. Τα αρχαία εμφανίζουν πολύ μικρό αριθμό θέσεων για τα φύλα CRENARCHAEOTA, EURYARCHAEOTA και NANOARCHAEOTA σε ποσοστά 47% για τα φύλα CRENARCHAEOTA και EURYARCHAEOTA και 54% για το φύλο NANOARCHAEOTA. Όλοι οι ευκαρυώτες έχουν μικρότερες τιμές θέσεων με τις χαμηλότερες από αυτές να αντιστοιχούν στα φύλα MICROSPORIDIA και CHLOROPHYTA και να έχουν ποσοστά 44% και 42% αντίστοιχα.

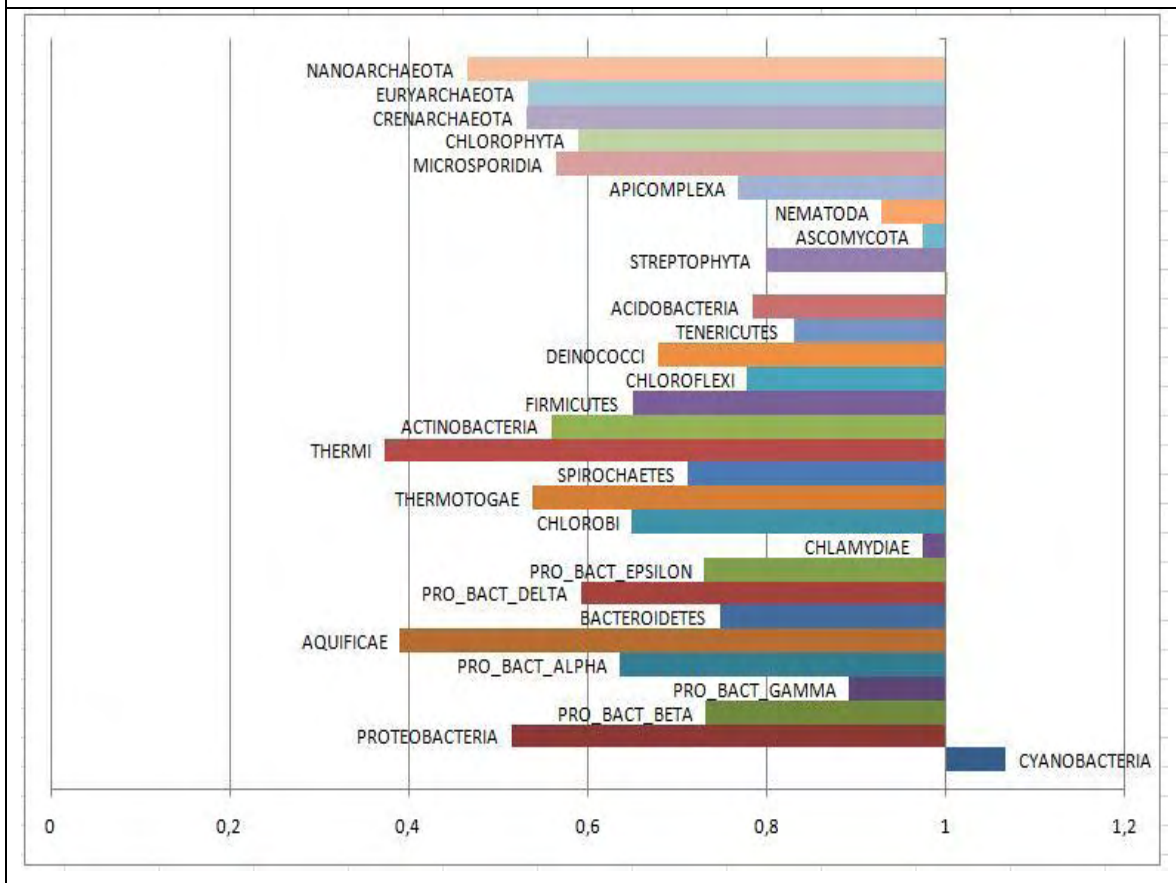
Εικόνα 3.9: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση pkb



- **gsk3**

Για την κινάση gsk (Πίνακας 3.12 παραρτήματος), από το σύνολο των πρωτεϊνών παρατηρείται πως όλες έχουν λιγότερο αριθμό θέσεων φωσφορυλίωσης από ότι ο άνθρωπος. Η χαμηλότερη τιμή εμφανίζεται στο φύλο TENERICUTES των βακτηρίων με ποσοστό 63%.

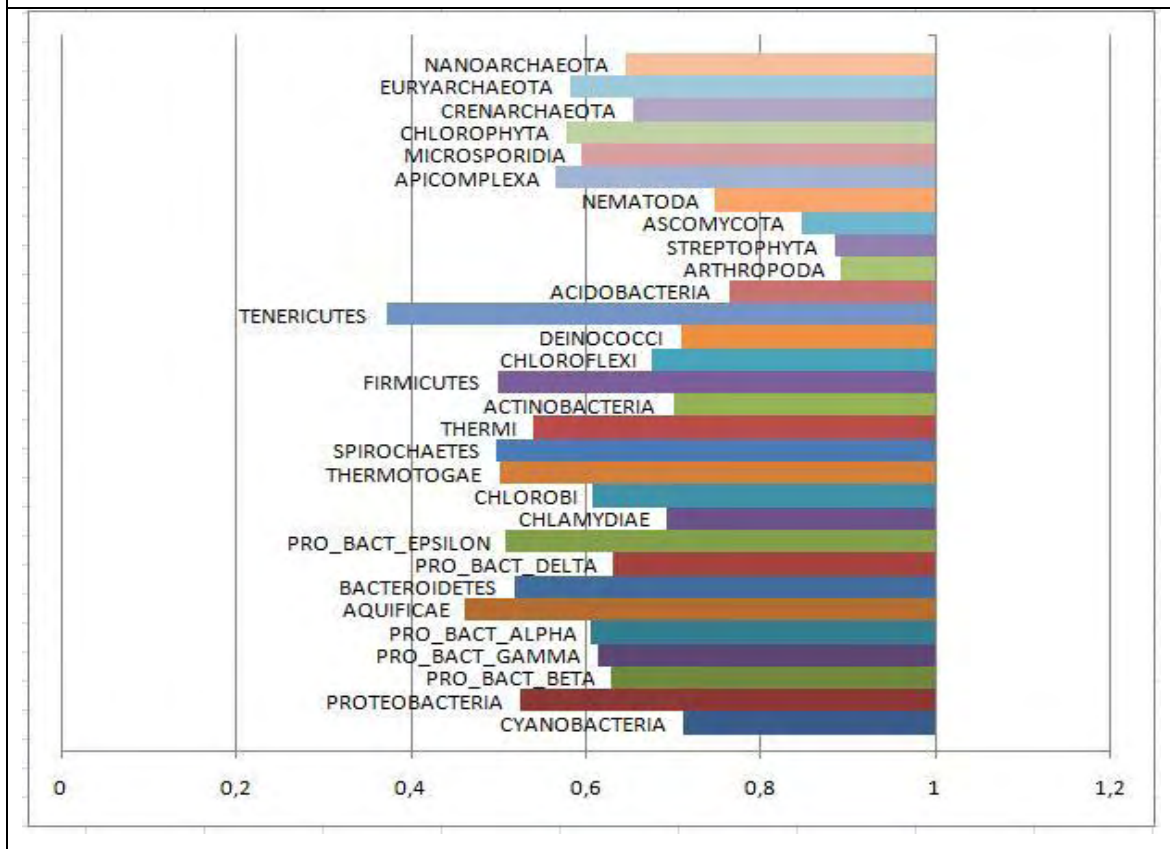
Εικόνα 3.10: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση dnark



- **src**

Για την κινάση src (Πίνακας 3.13 παραρτήματος), όλοι οι ευκαρυώτες που εξετάζονται έχουν μεγαλύτερο αριθμό θέσεων φωσφορυλίωσης από ότι ο άνθρωπος. Μεγάλο ενδιαφέρον παρατηρείται στα αρχαία καθώς έχουν τις υψηλότερες τιμές από όλες τις πρωτεΐνες. Συγκεκριμένα για τα φύλα CRENARCHAEOTA, EURYARCHAEOTA και NANOARCHAEOTA η εμφάνιση των θέσεων είναι σε ποσοστά 83%, 65% και 221% αντίστοιχα. Τα βακτήρια εμφανίζουν τον μεγαλύτερο αριθμό θέσεων στο φύλο AQUIFICAE με ποσοστό 77% και το μικρότερο ποσοστό συνολικά στο φύλο PROTEOBACTERIA της τάξης του 11%.

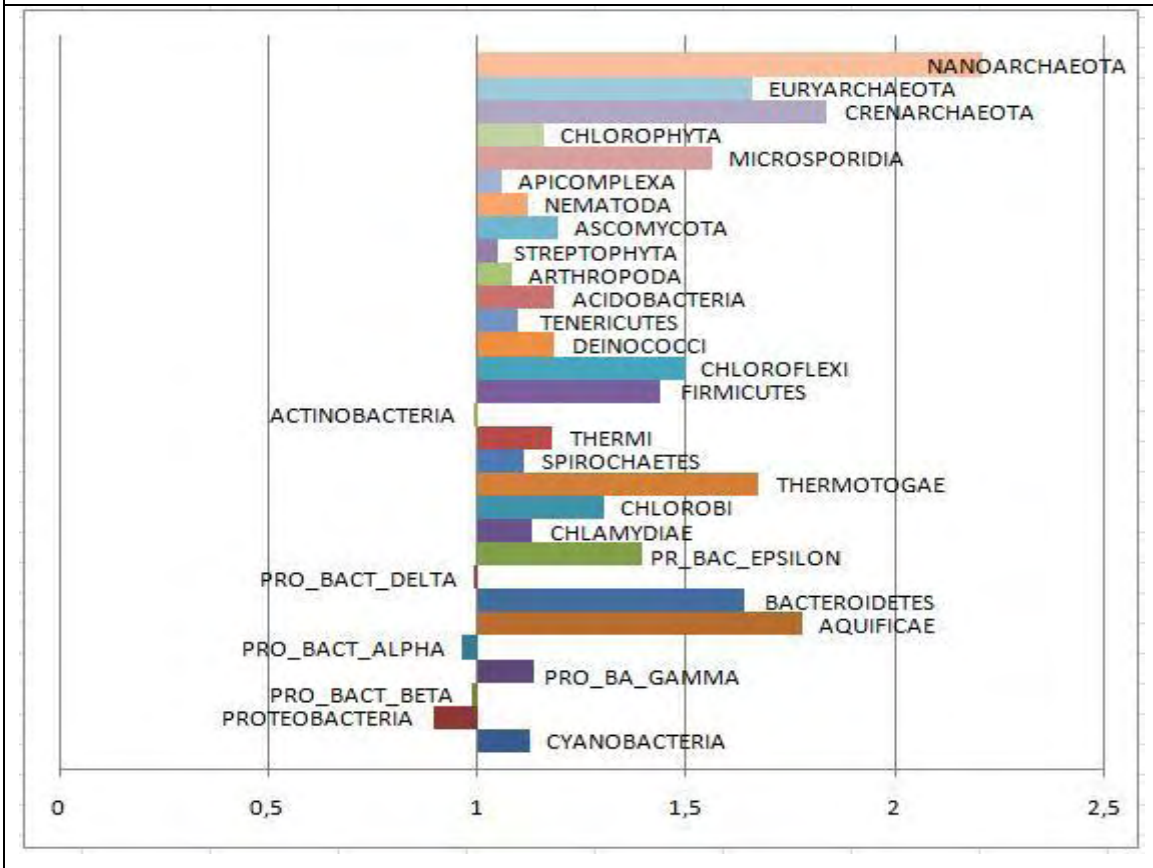
Εικόνα 3.11: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση gsk3



- **CaM-II**

Για την κινάση CaM-II (Πίνακας 3.14 παραρτήματος), για όλο το σύνολο των πρωτεϊνών παρατηρείται πως εμφανίζονται μικρότερες θέσεις φωσφορυλίωσης από ότι στον άνθρωπο. Για τα βακτήρια η μικρότερη τιμή εμφανίζεται στο φύλο AQUIFICAE με ποσοστό 88%, για τους ευκαρυώτες στο φύλο APICOMPLEXA με ποσοστό 77%, ενώ η συνολική μικρότερη τιμή εμφανίζεται στα αρχαία στο φύλο NANOARCHAEOTA.

Εικόνα 3.12: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση src



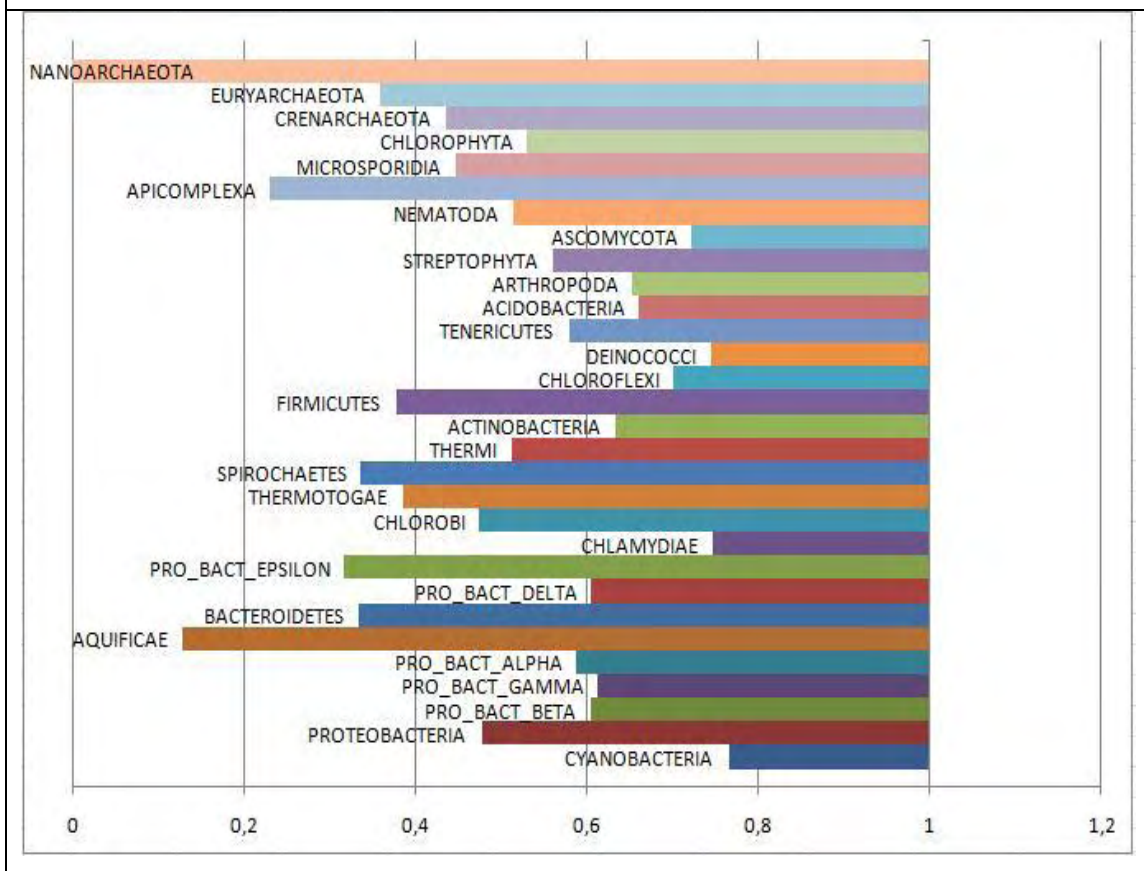
- **unsp**

Για την κινάση unsp (Πίνακας 3.15 παραρτήματος), για τα αρχαία και τα βακτήρια παρατηρείται πως εμφανίζουν λιγότερες θέσεις φωσφορυλίωσης από ότι ο άνθρωπος, με το φύλο THERMI των βακτηρίων να εμφανίζει την μικρότερη τιμή από το σύνολο των εξεταζόμενων πρωτεϊνών σε ποσοστό 49%. Οι ευκαρυώτες εμφανίζουν το μεγαλύτερο αριθμό θέσεων συνολικά στο φύλο ASCOMYCOTA με ποσοστό 4%.

- **cdc2**

Για την κινάση cdc2 (Πίνακας 3.16 παραρτήματος), τόσο οι ευκαρυώτες, όσο και τα αρχαία και τα βακτήρια παρατηρείται πως εμφανίζουν λιγότερες θέσεις φωσφορυλίωσης από ότι ο άνθρωπος, με το φύλο NANOARCHAEOTA των αρχαίων να εμφανίζει την μικρότερη τιμή από το σύνολο των εξεταζόμενων πρωτεϊνών σε ποσοστό 60%.

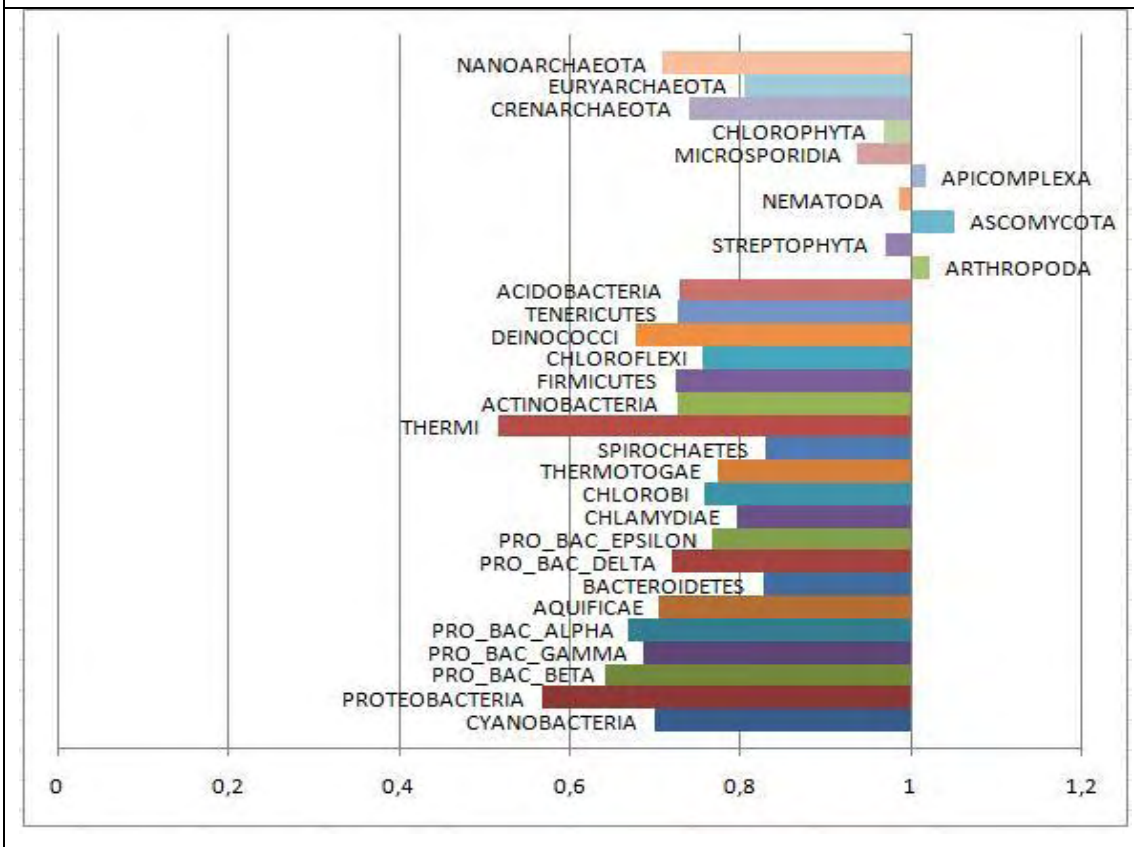
Εικόνα 3.13: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση CaM-II



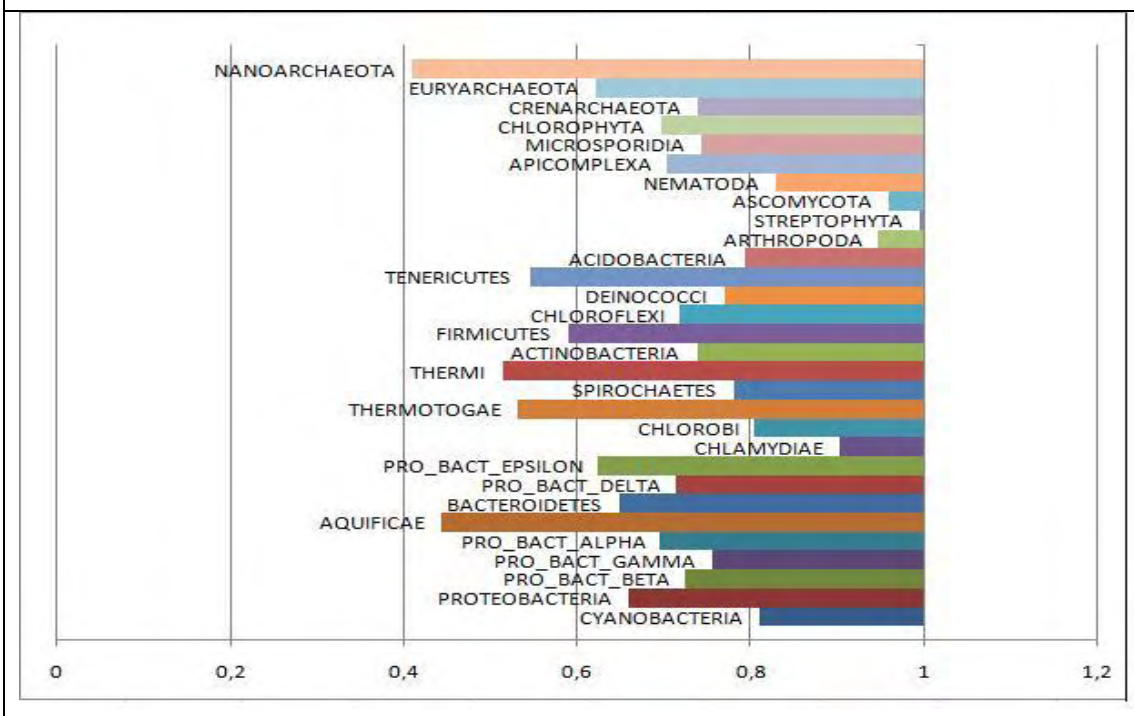
- **p38MAPK**

Για την κινάση p38MAPK (Πίνακας 3.17 παραρτήματος), για τα αρχαία και τους ευκαρυώτες παρατηρείται πως εμφανίζουν λιγότερες θέσεις φωσφορυλίωσης από ότι ο άνθρωπος, με το φύλο NANOARCHAEOTA των αρχαίων να εμφανίζει την μικρότερη τιμή από το σύνολο των εξεταζόμενων πρωτεϊνών σε ποσοστό 50%, ενώ ακολουθούν τα φύλα APICOMPLEXA και MICROSPORIDIA των ευκαρυωτών με ποσοστά 47% και 48% αντίστοιχα. Τα βακτήρια εμφανίζουν το μεγαλύτερο αριθμό θέσεων συνολικά στο φύλο DEINOCOCCI με ποσοστό 2%, ενώ ακολουθεί το φύλο CYANOBACTERIA με ποσοστό 1%.

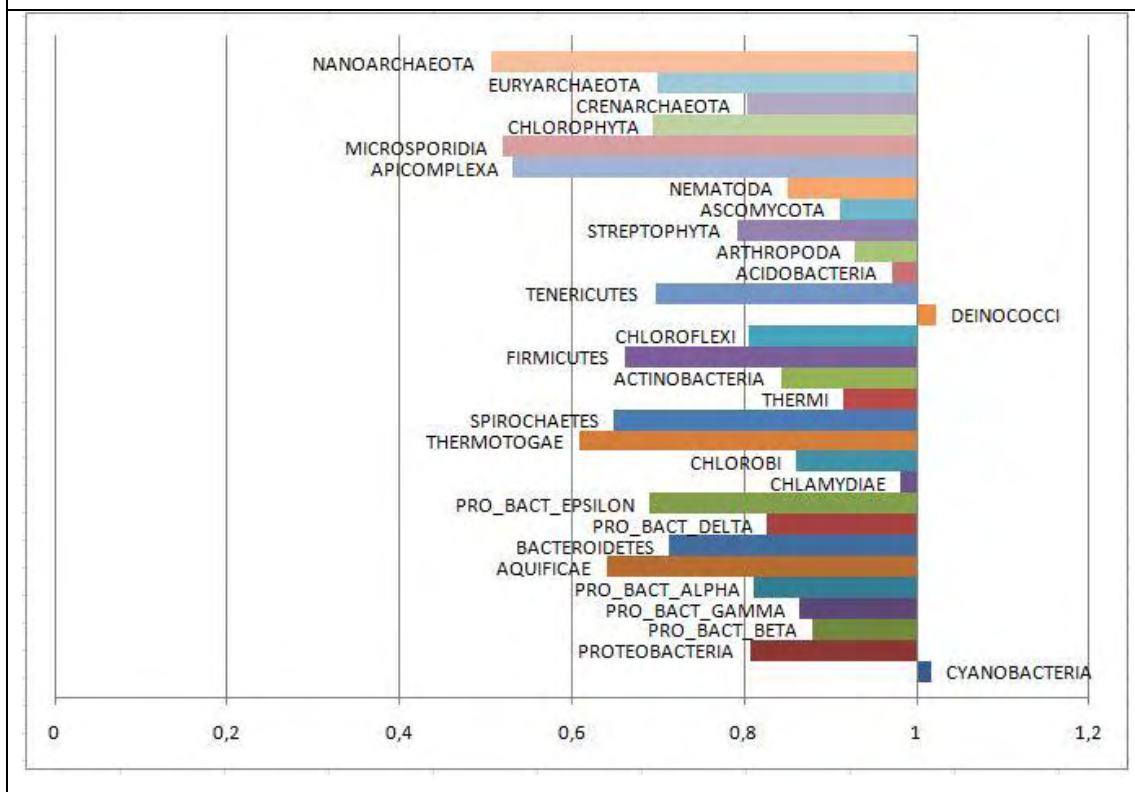
Εικόνα 3.14: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση unsp



Εικόνα 3.15: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση cdc2



Εικόνα 3.16: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση p38MAPK



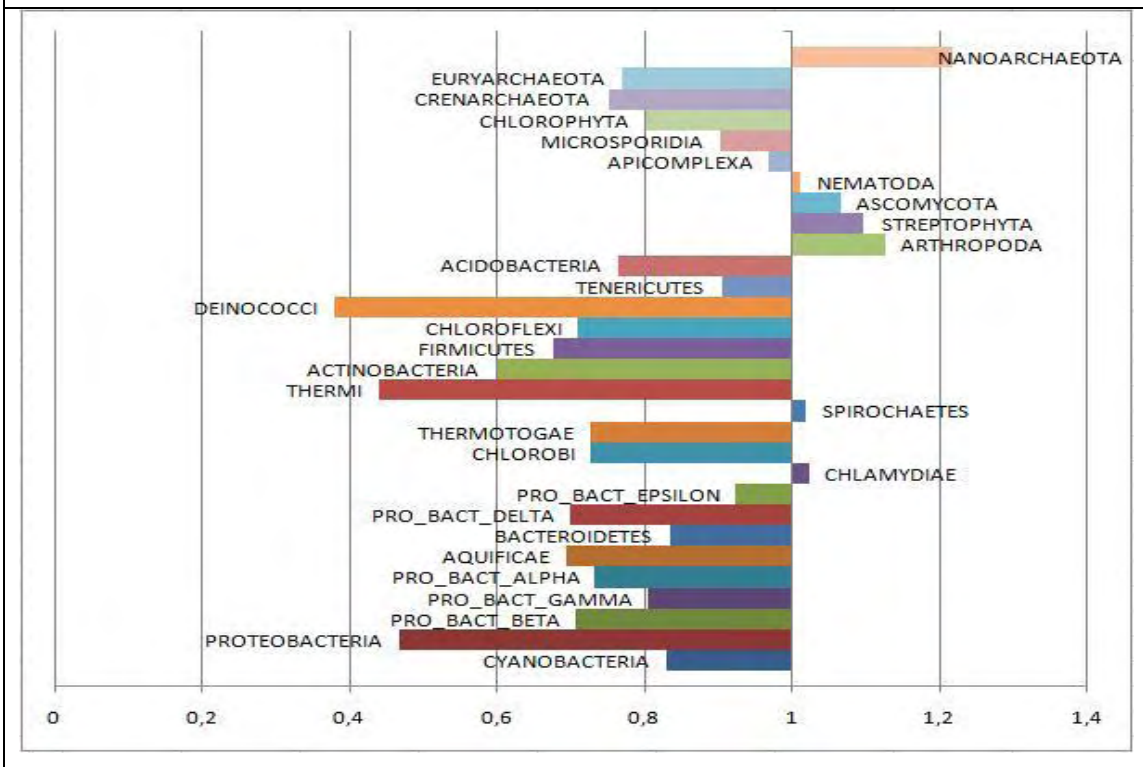
- **rsk**

Για την κινάση rsk (Πίνακας 3.18 παραρτήματος), για το φύλο NANOARCHAEOTA των αρχαίων παρατηρείται ο μεγαλύτερος συνολικός αριθμός θέσεων φωσφορυλίωσης σε ποσοστό 21%, ενώ στο φύλο DEINOCOCCI των βακτηρίων ο μικρότερος συνολικός αριθμός θέσεων σε ποσοστό 63%.

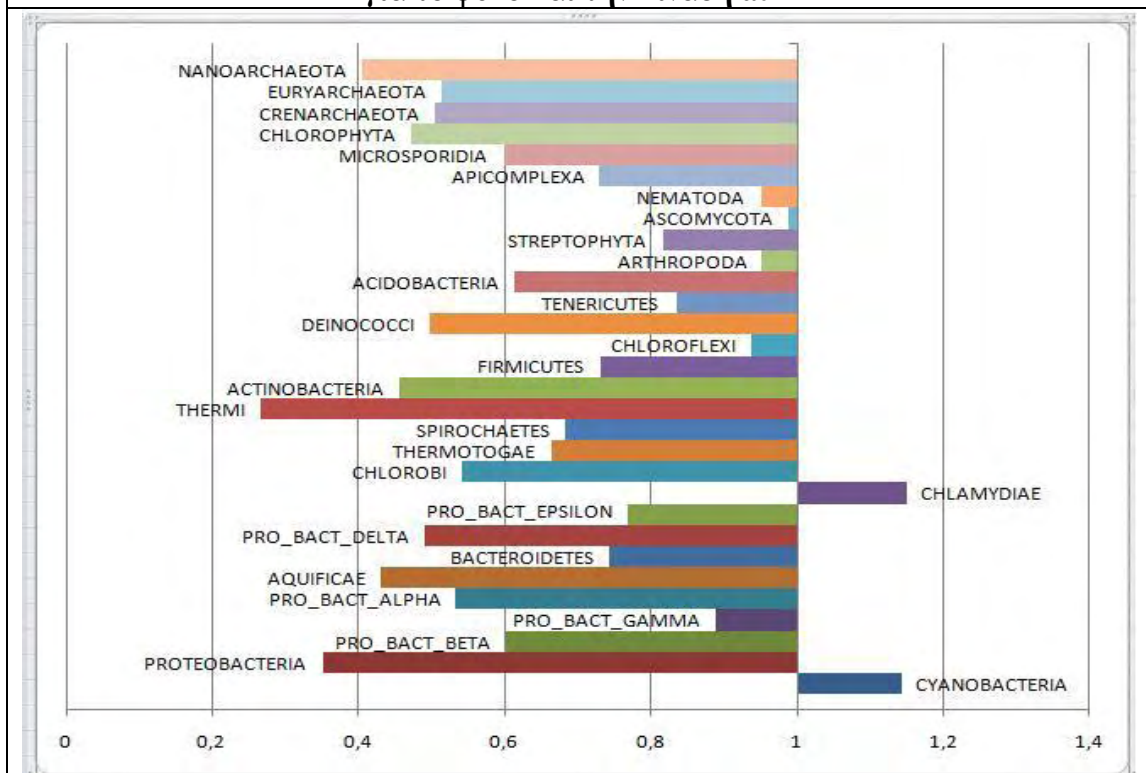
- **atm**

Για την κινάση atm (Πίνακας 3.19 παραρτήματος), για τα αρχαία και τους ευκαρυώτες παρατηρείται πως εμφανίζουν λιγότερες θέσεις φωσφορυλίωσης από ότι ο άνθρωπος. Η μεγαλύτερη τιμή των θέσεων συνολικά, εμφανίζεται στο φύλο CHLAMYDIAE των βακτηρίων σε ποσοστό 15%, ενώ ακολουθεί το φύλο CYANOBACTERIA των βακτηρίων με ποσοστό 14%. Τα βακτήρια επίσης εμφανίζουν και το μικρότερο αριθμό θέσεων συνολικά στο φύλο THERMI σε ποσοστό 74%.

Εικόνα 3.17: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση rsk



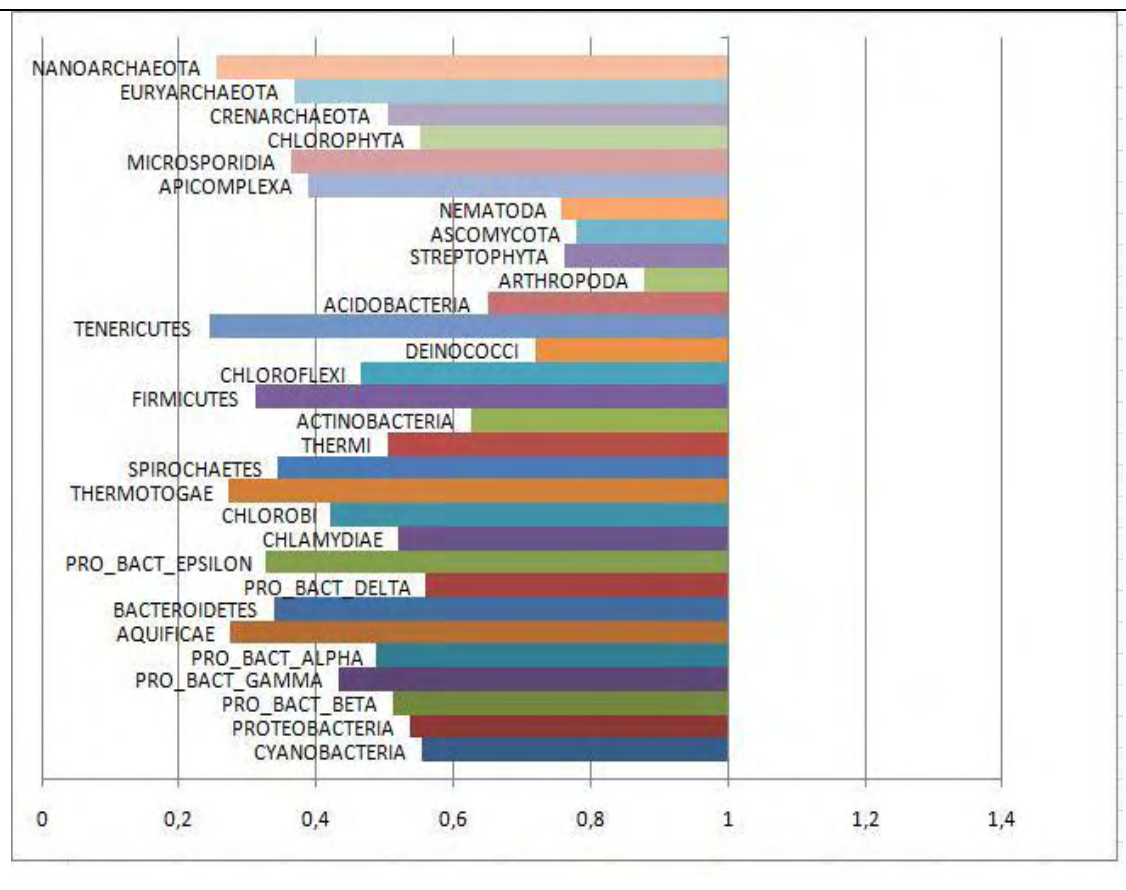
Εικόνα 3.18: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση atm



- **cdk5**

Για την κινάση cdk5 (Πίνακας 3.20 παραρτήματος), παρατηρείται πως όλο το σύνολο των πρωτεϊνών εμφανίζει λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τον άνθρωπο. Συγκεκριμένα το φύλο TENERICUTES των βακτηρίων εμφανίζει τον μικρότερο συνολικό αριθμό θέσεων με ποσοστό 76% και ακολουθεί το φύλο NANOARCHAEOTA των αρχαίων με ποσοστό 75%.

Εικόνα 3.19: Γραφική απεικόνιση των αποτελεσμάτων παλινδρόμησης Poisson για το φύλο και την κινάση cdk5



Κατά το τελευταίο μέρος της μετά-ανάλυσης, έγινε Poisson Regression, κατά την οποία συγκρίνεται η κάθε κινάση και το ποσοστό εμφάνισής της σε κάθε μία από τις κατηγορίες Αρχαία, Βακτήρια και Ευκαρυώτες, έχοντας ως σημείο σύγκρισης την κατηγορία των Ευκαρυωτών. Τα αποτελέσματα που εξήχθησαν παραθέτονται παρακάτω:

- **insr**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση insr (Πίνακας 3.21) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 33% περισσότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 16% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.21: Αποτελέσματα παλινδρόμησης Poisson για την κινάση insr σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. char category[omit] "EUKARYA"

. xi:poisson insr i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -920496.44
Iteration 1:   log likelihood = -920496.24
Iteration 2:   log likelihood = -920496.24

Poisson regression                               Number of obs   =   1020130
                                                    LR chi2(2)      =   5346.00
                                                    Prob > chi2     =   0.0000
Log likelihood = -920496.24                       Pseudo R2      =   0.0029
```

insr	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
_Icategory_1	1.337834	.0091095	42.74	0.000	1.320098 1.355808
_Icategory_2	.846559	.002797	-50.42	0.000	.8410946 .8520589
length	(exposure)				

- **egfr**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση egfr (Πίνακας 3.22) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 23% περισσότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 4% περισσότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.22: Αποτελέσματα παλινδρόμησης Poisson για την κινάση egfr σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson egfr i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -846540.84
Iteration 1:    log likelihood = -846540.74
Iteration 2:    log likelihood = -846540.74

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =     746.70
                                                    Prob > chi2     =     0.0000
Log likelihood = -846540.74                      Pseudo R2      =     0.0004

-----
          egfr |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |    1.233904   .0096474    26.88  0.000     1.21514    1.252958
_Icategory_2 |    1.046989   .0036253    13.26  0.000     1.039908    1.054119
      length | (exposure)

```

- **cki**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση cki (Πίνακας 3.23) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 8% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 17% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

- **ckii**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση ckii (Πίνακας 3.24) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 3% περισσότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 23% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.23: Αποτελέσματα παλινδρόμησης Poisson για την κινάση cki σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson cki i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -1256946.4
Iteration 1:    log likelihood = -1256946.1
Iteration 2:    log likelihood = -1256946.1

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =    5143.68
                                                    Prob > chi2     =    0.0000
Log likelihood = -1256946.1                        Pseudo R2       =    0.0020

```

cki	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
_Icategory_1	.9277445	.005816	-11.96	0.000	.9164152	.9392139
_Icategory_2	.8309717	.0021513	-71.52	0.000	.8267659	.8351989
length (exposure)						

- **pka**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση pka (Πίνακας 3.25) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 8% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 6% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.24: Αποτελέσματα παλινδρόμησης Poisson για την κινάση ckii σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson ckii i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -2370201.6
Iteration 1:   log likelihood = -2370200.1
Iteration 2:   log likelihood = -2370200.1

Poisson regression                               Number of obs   =   1020130
                                                    LR chi2(2)      =   28332.10
                                                    Prob > chi2     =   0.0000
Log likelihood = -2370200.1                       Pseudo R2      =   0.0059

-----+-----
          ckii |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   1.035055    .00362      9.85  0.000    1.027985    1.042175
_Icategory_2 |   .7782887   .0012039   -162.04  0.000    .7759327    .7806519
length | (exposure)
```

Πίνακας 3.25: Αποτελέσματα παλινδρόμησης Poisson για την κινάση pka σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson pka i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -2470627.1
Iteration 1:   log likelihood = -2470626.4
Iteration 2:   log likelihood = -2470626.4

Poisson regression                               Number of obs   =   1020130
                                                    LR chi2(2)      =   1912.14
                                                    Prob > chi2     =   0.0000
Log likelihood = -2470626.4                       Pseudo R2      =   0.0004

-----+-----
          pka |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .9279233    .0032966   -21.06  0.000    .9214846    .934407
_Icategory_2 |   .9426144    .0013373   -41.66  0.000    .939997    .9452391
length | (exposure)
```

- **pkg**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση pkg (Πίνακας 3.26) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 25% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 13% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.26: Αποτελέσματα παλινδρόμησης Poisson για την κινάση pkg σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson pkg i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -959347.77
Iteration 1:    log likelihood = -959347.55
Iteration 2:    log likelihood = -959347.55

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =    2608.63
                                                    Prob > chi2     =    0.0000
Log likelihood = -959347.55                       Pseudo R2       =    0.0014

```

pkg	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
_Icategory_1	.7596786	.0063492	-32.89	0.000	.7473357 .7722252
_Icategory_2	.8731788	.0027096	-43.70	0.000	.8678842 .8785057
length (exposure)					

- **pkc**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση pkc (Πίνακας 3.27) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως τόσο στα αρχαία όσο και στα βακτήρια υπάρχουν 23% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες.

- **pkb**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση pkb (Πίνακας 3.28) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 31% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 8% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.27: Αποτελέσματα παλινδρόμησης Poisson για την κινάση pkb σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες						
. xi:poisson pkb i.category, exposure(length) irr						
i.category	_Icategory_1-3		(_Icategory_3 for cat~y==EUKARYA omitted)			
Iteration 0: log likelihood = -5098228						
Iteration 1: log likelihood = -5098224.2						
Iteration 2: log likelihood = -5098224.2						
Poisson regression			Number of obs = 1020130			
			LR chi2(2) = 104601.78			
			Prob > chi2 = 0.0000			
Log likelihood = -5098224.2			Pseudo R2 = 0.0102			

	pkb	IRR	Std. Err.	z	P> z	[95% Conf. Interval]

	_Icategory_1	.7762437	.0016133	-121.87	0.000	.7730881 .7794121
	_Icategory_2	.7759883	.0006249	-314.94	0.000	.7747645 .777214
	length (exposure)					

- **dnapk**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση dnapk (Πίνακας 3.29) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 43% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 22% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.28: Αποτελέσματα παλινδρόμησης Poisson για την κινάση pkb σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson pkb i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -157266.6
Iteration 1:   log likelihood = -157266.59
Iteration 2:   log likelihood = -157266.59

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =    179.18
                                                    Prob > chi2     =    0.0000
Log likelihood = -157266.59                       Pseudo R2       =    0.0006
```

pkb	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
_Icategory_1	.697255	.0216998	-11.59	0.000	.6559954 .7411097
_Icategory_2	.9194354	.0100662	-7.67	0.000	.8999161 .939378
length	(exposure)				

Πίνακας 3.29: Αποτελέσματα παλινδρόμησης Poisson για την κινάση dnapk σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson dnapk i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -1458178.7
Iteration 1:   log likelihood = -1458177.9
Iteration 2:   log likelihood = -1458177.9

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =   16242.41
                                                    Prob > chi2     =    0.0000
Log likelihood = -1458177.9                       Pseudo R2       =    0.0055
```

dnapk	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
_Icategory_1	.5729232	.0038621	-82.63	0.000	.5654034 .5805431
_Icategory_2	.7898626	.0017788	-104.75	0.000	.7863838 .7933567
length	(exposure)				

- **gsk3**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση gsk3 (Πίνακας 3.30) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 31% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 30% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.30: Αποτελέσματα παλινδρόμησης Poisson για την κινάση gsk3 σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson gsk3 i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -275036.24
Iteration 1:    log likelihood = -275036.19
Iteration 2:    log likelihood = -275036.19

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =    2287.22
                                                    Prob > chi2     =    0.0000
Log likelihood = -275036.19                       Pseudo R2      =    0.0041
-----
          gsk3 |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .6910858   .0140673   -18.15  0.000   .6640571   .7192146
_Icategory_2 |   .7007791   .0053907   -46.22  0.000   .6902928   .7114247
      length | (exposure)

```

- **src**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση src (Πίνακας 3.31) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 60% περισσότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 5% περισσότερες θέσεις σε σχέση με τους ευκαρυώτες.

- **CaM-II**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση CaM-II (Πίνακας 3.32) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 49% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 23% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.31: Αποτελέσματα παλινδρόμησης Poisson για την κινάση src σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson src i.category, exposure(length) irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -288609.54
Iteration 1:    log likelihood = -288609.52

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =     868.06
                                                    Prob > chi2     =     0.0000
Log likelihood = -288609.52                       Pseudo R2      =     0.0015
```

src	IRR	Std. Err.	z	P> z	[95% Conf. Interval]
_Icategory_1	1.603233	.0242188	31.25	0.000	1.55646 1.65141
_Icategory_2	1.053298	.0079034	6.92	0.000	1.037921 1.068902
length (exposure)					

- **unsp**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση unsp (Πίνακας 3.33) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 22% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 31% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.32: Αποτελέσματα παλινδρόμησης Poisson για την κινάση CaM-II σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson camii i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -26356.711
Iteration 1:   log likelihood = -26356.704
Iteration 2:   log likelihood = -26356.704

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =    96.05
                                                    Prob > chi2     =    0.0000
Log likelihood = -26356.704                       Pseudo R2       =    0.0018

-----+-----
          camii |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .5159754   .0522575    -6.53  0.000   .4230777   .6292712
_Icategory_2 |   .7780895   .0251514    -7.76  0.000   .7303227   .8289805
length | (exposure)
```

Πίνακας 3.33: Αποτελέσματα παλινδρόμησης Poisson για την κινάση unsp σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson unsp i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -11171977
Iteration 1:   log likelihood = -11171964
Iteration 2:   log likelihood = -11171964

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =  558730.46
                                                    Prob > chi2     =    0.0000
Log likelihood = -11171964                       Pseudo R2       =    0.0244

-----+-----
          unsp |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .7801594   .0009532  -203.18  0.000   .7782933   .78203
_Icategory_2 |   .696859   .0003412  -737.71  0.000   .6961906   .697528
length | (exposure)
```

- **cdc2**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση cdc2 (Πίνακας 3.34) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 31% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 24% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.34: Αποτελέσματα παλινδρόμησης Poisson για την κινάση cdc2 σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson cdc2 i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:    log likelihood = -2638615.4
Iteration 1:    log likelihood = -2638613.9
Iteration 2:    log likelihood = -2638613.9

Poisson regression                                Number of obs   =    1020130
                                                    LR chi2(2)      =    41782.22
                                                    Prob > chi2     =    0.0000
Log likelihood = -2638613.9                       Pseudo R2      =    0.0079
-----
          cdc2 |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .6994028   .0026503   -94.35  0.000   .6942275   .7046166
_Icategory_2 |   .7623717   .0010714  -193.07  0.000   .7602748   .7644744
      length | (exposure)

```

- **p38MAPK**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση p38MAPK (Πίνακας 3.35) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 18% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 7% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

- **rsk**

Πραγματοποιείται έλεγχος κατανομής Poisson συγκρίνοντας την κινάση rsk (Πίνακας 3.36) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 26% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 28% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.35: Αποτελέσματα παλινδρόμησης Poisson για την κινάση p38MAPK σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες						
<code>. xi:poisson p38mapk i.category,exposure(length)irr</code>						
<code>i.category _Icategory_1-3 (_Icategory_3 for cat~y==EUKARYA omitted)</code>						
Iteration 0: log likelihood = -832219.8						
Iteration 1: log likelihood = -832219.66						
Iteration 2: log likelihood = -832219.66						
Poisson regression			Number of obs = 1020130			
			LR chi2(2) = 719.72			
			Prob > chi2 = 0.0000			
Log likelihood = -832219.66			Pseudo R2 = 0.0004			

p38mapk	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
_Icategory_1	.8219028	.0075393	-21.38	0.000	.807258	.8368132
_Icategory_2	.9350686	.0032447	-19.35	0.000	.9287307	.9414496
length (exposure)						

- **atm**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση atm (Πίνακας 3.37) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 45% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 26% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.36: Αποτελέσματα παλινδρόμησης Poisson για την κινάση rsk σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

xi:poisson rsk i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -136888.93
Iteration 1:   log likelihood = -136888.91
Iteration 2:   log likelihood = -136888.91

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =    767.87
                                                    Prob > chi2     =    0.0000
Log likelihood = -136888.91                       Pseudo R2       =    0.0028

-----
          rsk |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .7458184   .0231848    -9.43  0.000   .7017337   .7926726
_Icategory_2 |   .7209302   .0087051   -27.10  0.000   .7040687   .7381955
length | (exposure)

```

Πίνακας 3.37: Αποτελέσματα παλινδρόμησης Poisson για την κινάση atm σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```

. xi:poisson atm i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood = -508984.45
Iteration 1:   log likelihood = -508984.27
Iteration 2:   log likelihood = -508984.27

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =   4563.30
                                                    Prob > chi2     =    0.0000
Log likelihood = -508984.27                       Pseudo R2       =    0.0045

-----
          atm |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .5521367   .0082621   -39.69  0.000   .5361784   .5685699
_Icategory_2 |   .7473677   .0037293   -58.36  0.000   .7400941   .7547127
length | (exposure)

```

- **cdk5**

Πραγματοποιείται έλεγχος κατανομής Poisson, συγκρίνοντας την κινάση cdk5 (Πίνακας 3.38) στις κατηγορίες των αρχαίων και των βακτηρίων, με τους ευκαρυώτες. Παρατηρείται πως στα αρχαία υπάρχουν 51% λιγότερες θέσεις φωσφορυλίωσης σε σχέση με τους ευκαρυώτες, ενώ στα βακτήρια υπάρχουν 43% λιγότερες θέσεις σε σχέση με τους ευκαρυώτες.

Πίνακας 3.38: Αποτελέσματα παλινδρόμησης Poisson για την κινάση cdk5 σε σύγκριση των αρχαίων και των βακτηρίων με τους ευκαρυώτες

```
. xi:poisson cdk5 i.category,exposure(length)irr
i.category      _Icategory_1-3      (_Icategory_3 for cat~y==EUKARYA omitted)

Iteration 0:   log likelihood =   -797798
Iteration 1:   log likelihood = -797797.63
Iteration 2:   log likelihood = -797797.63

Poisson regression                                Number of obs   =   1020130
                                                    LR chi2(2)      =   25522.92
                                                    Prob > chi2     =   0.0000
Log likelihood = -797797.63                       Pseudo R2      =   0.0157

-----
          cdk5 |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icategory_1 |   .4934327   .0053788   -64.80  0.000   .4830024   .5040883
_Icategory_2 |   .571241    .0021355  -149.79  0.000   .5670709   .5754418
      length | (exposure)
```

4.ΣΥΖΗΤΗΣΗ-ΣΥΜΠΕΡΑΣΜΑΤΑ

Οι πρωτεΐνες αποτελούν το μεγαλύτερο μέρος της ξηρής μάζας ενός κυττάρου και επιτελούν όλες σχεδόν τις κυτταρικές λειτουργίες. Εξετάζοντας και αναλύοντας την ηλεκτρική ή βιοχημική δραστηριότητα ενός κυττάρου, στην ουσία παρατηρούμε τις πρωτεΐνες οι οποίες εξυπηρετούν βασικές λειτουργίες σε όλες σχεδόν τις βιολογικές εργασίες. Η δυνατότητα των πρωτεϊνών να επιτελούν τόσες πολλές και διαφορετικές λειτουργίες προκύπτει από τον τεράστιο αριθμό διαφορετικών τρισδιάστατων στερεοδομών που μπορεί να προσλάβουν: η λειτουργία ακολουθεί τη δομή.

Στην παρούσα πτυχιακή εργασία επιλέξαμε να ασχοληθούμε με μία κατηγορία τροποποιήσεων τις οποίες οι πρωτεΐνες υφίστανται μετά τη μετάφρασή τους και είναι γνωστές ως Μετά-Μεταφραστικές Τροποποιήσεις Πρωτεϊνών. Οι Μετά-Μεταφραστικές Τροποποιήσεις γίνονται κατά κύριο λόγο στο ενδοπλασματικό δίκτυο και στη συσκευή Golgi ακολουθώντας ορισμένες πολύπλοκες διεργασίες και συσχετίζονται με αλλαγές στερεοδιάταξης, με την πρόσδεση στη μεμβράνη καθώς και με τον τελικό προορισμό μέσα στο κύτταρο. Στα πλαίσια του θέματος που επιλέξαμε κάναμε αρχικά μία ανασκόπηση στα κυριότερα είδη των Μετά-Μεταφραστικών Τροποποιήσεων που συναντώνται και στη συνέχεια συγκεντρώσαμε το ενδιαφέρον μας στη Φωσφορυλίωση των πρωτεϊνών.

Είναι ευρέως αποδεδειγμένο πως η Φωσφορυλίωση έχει αναπτυχθεί ιδιαίτερα για τους ευκαρυωτικούς οργανισμούς και πως οι μέθοδοι πρόγνωσης αυτής της μετά-μεταφραστικής τροποποίησης είναι προσαρμοσμένες για τους ευκαρυώτες. Ωστόσο υπάρχουν ενδείξεις πως η Φωσφορυλίωση δεν αποτελεί αποκλειστικό χαρακτηριστικό των ευκαρυωτών καθώς εμφανίζεται και σε προκαρυωτικούς οργανισμούς καθώς και στα αρχαία με μικρότερη συχνότητα. Για το λόγο αυτό εφαρμόσαμε την υπάρχουσα μέθοδο πρόγνωσης Φωσφορυλίωσης στα Βακτήρια και τα Αρχαία. Δηλαδή, για τη μέθοδο που εφαρμόσαμε είχαν χρησιμοποιηθεί δεδομένα από ευκαρυωτικούς οργανισμούς. Συλλέχτηκε ένας επαρκής αντιπροσωπευτικός αριθμός οργανισμών που ανήκουν στις κατηγορίες των ευκαρυωτών, των βακτηρίων και των αρχαίων και στη συνέχεια οι οργανισμοί αυτοί ταξινομήθηκαν ως προς την κατηγορία και το φύλο. Η εξέταση της φωσφορυλίωσης θα βασίζονταν στις 18 διαφορετικές κινάσες που υπάρχουν στο σύνολο των πρωτεϊνών που εξετάζονται. Έτσι αναπτύχθηκε ένα πρόγραμμα σε περιβάλλον προγραμματισμού perl το οποίο

μετρούσε τον αριθμό εμφάνισης κάθε μίας από τις 18 κινάσες σε κάθε οργανισμό και στη συνέχεια τα αποτελέσματα εξετάστηκαν με τη χρήση του προγράμματος μετά-ανάλυσης Stata. Εφαρμόσαμε παλινδρόμηση Poisson συγκρίνοντας κάθε μία από τις διαφορετικές 18 κινάσες για κάθε ένα από τα 30 φύλα που εξετάζονται έχοντας ως σημείο σύγκρισης το φύλο chordata που αντιστοιχεί στον ανθρώπινο οργανισμό.

Διαπιστώθηκε πως για τις κινάσες CaM-II, cdk5, cdc2 και gsk3 οι θέσεις φωσφορυλίωσης τόσο στους ευκαρυωτικούς οργανισμούς όσο και στα αρχαία και τα βακτήρια, από το σύνολο των οργανισμών που εξετάστηκαν, είναι μικρότερες σε ποσοστό σε σχέση με τις θέσεις φωσφορυλίωσης στον άνθρωπο. Το αποτέλεσμα αυτό είναι αναμενόμενο καθώς η μέθοδος εκπαιδεύτηκε σε ευκαρυωτικούς οργανισμούς.

Όσον αφορά τα αρχαία και τα βακτήρια, διαπιστώθηκε πως για κάποιες κινάσες εμφανίζουν περισσότερες θέσεις φωσφορυλίωσης από τους ευκαρυώτες. Συγκεκριμένα τα αρχαία, για τις κινάσες insr, src και egfr εμφανίζουν μεγαλύτερα ποσοστά θέσεων φωσφορυλίωσης από τους ευκαρυωτικούς οργανισμούς. Σημαντικές διαφορές παρατηρούνται και στα βακτήρια. Συγκεκριμένα στην κινάση pka, παρατηρούνται για τα φύλα CHLAMYDIAE και SPIROCHAETES μεγαλύτερα ποσοστά θέσεων φωσφορυλίωσης από τον άνθρωπο της τάξης του 19% και 15% αντίστοιχα, τη στιγμή που τα υψηλότερα ποσοστά των ευκαρυωτών είναι 4% και 9% για τα φύλα STREPTOPHYTA και MICROSPORIDIA αντίστοιχα. Ομοίως στην κινάση dnark, η μεγαλύτερη τιμή των ευκαρυωτών είναι περίπου ίση με τον άνθρωπο, τη στιγμή που τα βακτήρια εμφανίζουν την υψηλότερη συγκέντρωση θέσεων της τάξης του 6,6% για το φύλο CYANOBACTERIA. Τέλος για την κινάση p38MAPK, από όλο το σύνολο των δεδομένων μεγαλύτερες τιμές θέσεων φωσφορυλίωσης από τον άνθρωπο εμφανίζουν μόνο τα βακτήρια για τα φύλα CYANOBACTERIA και DEINOCOCCI της τάξης του 1,7% και 2,2% αντίστοιχα. Το αποτέλεσμα αυτό είναι αξιοσημείωτο καθώς θα περιμέναμε σε όλους τους μη ευκαρυωτικούς οργανισμούς τα ποσοστά να ήταν μικρότερα. Το ίδιο αξιοσημείωτο είναι και το αποτέλεσμα που προκύπτει για την κινάση cki και συγκεκριμένα στο φύλο CHLOROBΙ όπου το ποσοστό εμφάνισης θέσεων φωσφορυλίωσης συμπίπτει με αυτό του ανθρώπινου οργανισμού.

Τα αποτελέσματα που εξήχθησαν πρέπει να αποτελέσουν έναυσμα για περαιτέρω μελέτη καθώς καταλήγουμε στο συμπέρασμα πως είναι απαραίτητη η δημιουργία μίας απλής και γρήγορης μεθόδου που θα δίνει με ακρίβεια τον αριθμό θέσεων φωσφορυλίωσης όχι μόνο στους ευκαρυωτικούς οργανισμούς, αλλά και

στους οργανισμούς των αρχαίων και των βακτηρίων, καθώς όπως διαπιστώθηκε η φωσφορυλίωση είναι μία μετά-μεταφραστική τροποποίηση που δεν υφίστασται κατά κύριο λόγο στους ευκαρυώτες.

5.ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1]. Berg JM, Tymoczko JL, Stryer L: ΒΙΟΧΗΜΕΙΑ ΤΟΜΟΣ Ι.
- [2]. Alberts B, Bray D, Johnson A, Lewis J, Raff M, Roberts K, Walter P: Βασικές Αρχές Κυτταρικής Βιολογίας.
- [3]. Nuhse TS, Stensballe A, Jensen ON, Peck SC: Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol. Cell Proteomics* 2003, 2: 1234–1243.
- [4]. Torres CR, Hart GW: Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes— evidence for O-linked GlcNAc. *J Biol Chem* 1984, 259: 3308-3317.
- [5]. Hart GW, Haltiwanger RS, Holt GD, Kelly WG: Nucleoplasmic and cytoplasmic glycoproteins. *Ciba Found Symp* 1989, 145: 102-12.
- [6]. Hart GW, Housley MP, Slawson C: Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 2007, 446: 1017-1022.
- [7]. O'Donnell N, Zachara NE, Hart GW, Marth JD: Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* 2004, 24: 1680-1690.
- [8]. Liu K, Paterson AJ, Zhang F, McAndrew J, Fukuchi K, Wyss JM, Peng L, Hu Y, Kudlow JE: Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism. *J Neurochem* 2004 89: 1044-1055.
- [9]. Yang X, Ongusaha PP, Miles PD, Havstad JC, Zhang F, So WV, Kudlow JE, Michell RH, Olefsky JM, Field SJ, Evans RM: Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* 2008 451: 964-969.

- [10]. Helenius A, Aebi M: Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* 2004 73: 1019–1049.
- [11]. Weerapana E, Imperiali B: Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 2006 16: 91R–101R.
- [12]. Helenius J, Ng DT, Marolda CL, Walter P, Valvano MA, Aebi M: Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* 2002 415: 447–450.
- [13]. Szymanski CM, Wren BW: Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* 2005 3: 225-236.
- [14]. Weerepana E, Imperiali B: Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* 2006 16: 91R-101R.
- [15]. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW et al.: N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 2002 298: 1790-1793.
- [16]. Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C et al.: Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc Natl Acad Sci U S A* 2006, 103: 7088-7093.
- [17]. Mescher MF, Strominger JL: Purification and characterization of a prokaryotic glucoprotein from the cell envelope of *Halobacterium salinarium*. *J Biol Chem* 1976, 10: 2005–2014.
- [18]. Szymanski CM, Wren BW: Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* 2005, 3: 225–237.
- [19]. Eichler J, Adams MW: Posttranslational protein modification in Archaea. *Microbiol Mol Biol* 2005, 69: 393–425.

- [20]. Lechner J, Wieland F: Structure and biosynthesis of prokaryotic glycoproteins. *Annu Rev Biochem* 1989, 58: 173–194.
- [21]. Maurer-Stroh S, Eisenhaber B, Eisenhaber F: N-terminal N-Myristoylation of Proteins: Refinement of the Sequence Motif and its Taxon-specific Differences. *J. Mol. Biol* 2001 317: 523-540.
- [22]. Baeuerle PA, Huttner WB: Tyrosine sulfation of yolk proteins 1, 2, and 3 in *Drosophila melanogaster*. *J. Biol. Chem.* 1985, 260: 6434–6439.
- [23]. Bettelheim FR: Tyrosine-O-sulfate in a peptide from fibrinogen. *J. Am. Chem. Soc.* 1954, 76: 2838–2839.
- [24]. Huttner WB: Sulphation of tyrosine residues—A widespread modification of proteins. *Nature* 1982, 299: 273–276.
- [25]. Lee RW, Huttner WB: Tyrosine-O-sulfated proteins of PC12 pheochromocytoma cells and their sulfation by a tyrosylprotein sulfotransferase. *J. Biol. Chem.* 1983, 258: 11326–11334.
- [26]. Niehrs C, Huttner WB: Purification and characterization of tyrosylprotein Sulfotransferase. *EMBO J.* 1990, 9: 35–42.
- [27]. Huxtable RJ: *Biochemistry of Sulfur*. Plenum, New York, 1986.
- [28]. Beisswanger R, Corbeil D, Vannier C, Thiele C, Dohrmann U, Kellner R, Ashman K, Niehrs C, Huttner WB: Existence of distinct tyrosylprotein sulfotransferase genes: molecular characterization of tyrosylprotein sulfotransferase-2. *Proc. Natl. Acad. Sci. U. S. A.* 1998, 95: 11134–11139.
- [29]. Ouyang YB, Moore KL: Molecular cloning and expression of human and mouse tyrosylprotein sulfotransferase-2 and a tyrosylprotein sulfotransferase homologue in *Caenorhabditis elegans*. *J. Biol. Chem.* 1998, 273: 24770–24774.

- [30]. Marchler-Bauer A, Bryant SH: CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 2004, 32: W327–W331.
- [31]. Moore KL: The biology and enzymology of protein tyrosine O-sulfation. *J. Biol. Chem.* 2003, 278: 24243–24246.
- [32]. Zhou F, Xue Y, Yao X, and Xu Y: CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). *Bioinformatics* 2005, 894-896.
- [33]. Gregoire S, Yang XJ: Association with class IIa histone deacetylases upregulates the sumoylation of MEF2 transcription factors. *Mol. Cell Biol.* 2005, 25: 2273–2287.
- [34]. Girdwood DW, Tatham MH, Hay RT: SUMO and transcriptional regulation. *Semin. Cell Dev. Biol.* 2004, 15: 201–210.
- [35]. Li T, Evdokimov E, Shen RF, Chao CC, Tekle E, Wang T, Stadtman ER, Yang DC, Chock PB: Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. *Proc. Natl Acad. Sci. USA*, 2004, 101: 8551–8556.
- [36]. Liang M, Melchior F, Feng XH, Lin X: Regulation of Smad4 sumoylation and transforming growth factor-beta signaling by protein inhibitor of activated STAT1. *J. Biol. Chem.* 2004, 279: 22857–22865.
- [37]. Zhou F, Xue Y, Lu H, Chen G, Yao X: A genome-wide analysis of sumoylation-related biological processes and functions in human nucleus. *FEBS Lett.* 2005, 579: 3369–3375.
- [38]. Li M, Guo D, Isales CM, Eizirik DL, Atkinson M, She JX, Wang CY: SUMO wrestling with type 1 diabetes. *J. Mol. Med.* 2005, 83: 504–513.

- [39]. Shinbo Y, Niki T, Taira T, Ooe H, Takahashi-Niki K, Maita C, Seino C, Iguchi-Ariga SM, Ariga H: Proper SUMO-1 conjugation is essential to DJ-1 to exert its full activities. *Cell Death Differ.* 2005, 13: 96–108.
- [40]. Hay RT: SUMO: a history of modification. *Mol. Cell.* 2005, 18: 1–12.
- [41]. Kurepa J, Walker JM, Smalle J, Gosink MM, Davis SJ, Durham TL, Sung DY, Vierstra RD: The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J. Biol. Chem.* 2003, 278: 6862–6872.
- [42]. Johnson ES: Protein modification by SUMO. *Annu. Rev. Biochem.* 2004, 73: 355–382.
- [43]. Melchior F, Schergaut M, Pichler A: SUMO: ligases, isopeptidases and nuclear pores. *Trends Biochem. Sci.* 2003, 28: 612–618.
- [44]. Denison C, Rudner AD, Gerber SA, Bakalarski CE, Moazed D, Gygi SP: A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol Cell Proteomics* 2004, 4: 246–254.
- [45]. Harder Z, Zunino R, McBride H: Sumo1 conjugates mitochondrial substrates and participates in mitochondrial fission. *Curr. Biol.* 2004, 14: 340–345.
- [46]. Gocke CB, Yu H, Kang J: Systematic identification and analysis of mammalian small ubiquitin-like modifier substrates. *J. Biol. Chem.* 2005, 280: 5004–5012.
- [47]. Wykoff DD, O’Shea EK: Identification of sumoylated proteins by systematic immunoprecipitation of the budding yeast proteome. *Mol. Cell Proteomics* 2005, 4: 73–83.
- [48]. Obenauer J C, et al.: Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* 2003, 31: 3635-41.

- [49]. Sigrist C J, et al. :PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform* 2002, 3: 265-74.
- [50]. Blom N, et al. :Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999, 294: 1351-62.
- [51]. Blom N, et al.:Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 2004, 4: 1633-49.
- [52]. Xue Y, et al. : GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res* 2005, 33: 184-7.
- [53]. Iakoucheva LM, et al. :The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004, 32: 1037-49.
- [54]. Huang HD, et al. : KinasePhos: a web tool for identifying protein kinasespecific phosphorylation sites. *Nucleic Acids Res* 2005, 33: 226-9.
- [55]. Xue Y, et al. : PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics* 2006, 7: 163.
- [56]. Jensen ON: Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* 2006, 7: 391-403.
- [57]. Ficarro S B, et al. :Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 2002, 20: 301-5.
Gruhler A, et al. : Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 2005, 4: 310-27.
- [58]. Ball CA, et al. : Integrating functional genomic information into the *Saccharomyces* genome database. *Nucleic Acids Res* 2000, 28: 77-80.
- [59]. Manning G, et al. : Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 2002, 27: 514-20.

- [60]. Nielsen M, et al. : Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003, 12: 1007-17.
- [61]. Qian N, et al. : Predicting the secondary structure of globular proteins using neural network models. *J Mol Biol* 1988, 202: 865-84.
- [62]. Julenius K, Molgaard A, Gupta R, Brunak S: Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites *Glycobiology* 2004, 15: 2: 153-164
- [63]. www.cbs.dtu.dk/services/getstruct
- [64]. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997, 25: 3389–3402.
- [65]. Dalal S, Balasubramanian S, Regan L: Protein alchemy: changing beta-sheet into alpha-helix. *Nat. Struct. Biol.* 1997, 4: 548–552.
- [66]. Chothia C, Lesk AM : Relationship between the divergence of sequence and structure in proteins. *EMBO J.* 1986, 5: 823–827.
- [67]. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22: 4673–4680.
- [68]. Hansen JE, Lund O, Gooley AA, Williams KL, Brunak S: NetOGlyc. Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* 1998, 15: 115–130.
- [69]. Hertz J, Krogh A, Palmer R: Introduction to the theory of neural computation. Redwood City, CA: Addison-Wesley. 1991.

- [70]. Henikoff S, Henikoff JG: Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 1992, 89: 10915–10919.
- [71]. Li S, Li H, Li M, Shyr Y, Xie L, Li Y: Improved prediction of lysine acetylation by support vector machines. Bio-X Life Science Research Center and School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai 2002 China.
- [72]. Jones DT: Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 1999, 292: 195–202.
- McGuffin, L.J., Bryson K, Jones DT: The PSIPRED protein structure prediction server. Bioinformatics 2000, 16: 404–405.
- [73]. Gorodkin J, Lund O, Andersen CA, Brunak S: Using sequence motifs for enhanced neural network prediction of protein distance constraints. In Lengauer T, Schneider R, Bork P, Brutlag D, Glasgow J, Mewes HW, Zimmer R. Proceedings of the Seventh International Conference for Molecular Biology. 1999, 95–105.
- [74]. Matthews BW: Comparizon of the predicted and observed secondary structure of T4 phage lysozyme Biochim.Biophys.Acta 1975, 405: 442-451
- [75]. Maurer-Stroh S, Eisenhaber B, Eisenhaber F: N-terminal N-Myristoylation of Proteins: Refinement of the Sequence Motif and its Taxon-specific Differences, J.Mol.Biol. 2002, 317: 523-540
- [76]. Podell S, Gribskov M: Predicting n-terminal myristoylation sites in plant proteins, BMC Genomics 2004, 5:37.
- [77]. Nicholas HBJr, Chan SS, Rosenquist GL: Reevaluation of the determinants of tyrosine sulfation, Endocrine 1999, 11: 285–292.
- [78]. Rosenquist GL, Nicholas HBJr: Analysis of sequence requirements for protein tyrosine sulfation, Protein Sci. 1993, 2: 215–222.

- [79]. Pauwels S, Najdovski T, Dimaline R, Lee CM, Deschodt- Lanckman M: Degradation of human gastrin and CCK by endopeptidase 24.11: differential behaviour of the sulphated and unsulphated peptides, *Biochim. Biophys. Acta* 1989, 996: 82–88.
- [80]. Yu KM, Liu J, Moy R, Lin HC, Nicholas HBJr, Rosenquist GL: Prediction of tyrosine sulfation in seven-transmembrane peptide receptors, *Endocrine* 2002, 19: 333–338.
- [81]. Lin HC, Tsai K, Chang BL, Liu J, Young M, Hsu W, Louie S, Nicholas HBJr, Rosenquist GL: Prediction of tyrosine sulfation sites in animal viruses, *Biochem. Biophys. Res. Commun.* 2003, 312: 1154–1158.
- [82]. Xue Y, Chen H, Jin C, Sun Z, Yao X: NBA-Palm:prediction of palmitoylation site implemented in Naïve Bayes algorithm, *BMC Bioinformatics* 2006, 7:458.
- [83]. Xue Y, Zhou F, Zhu M, Ahmed K, Chen G, Yao X: GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.* 2005, 33: W184–W187.
- [84]. Zhou FF, Xue Y, Chen GL, Yao X: GPS: a novel group-based phosphorylation predicting and scoring method. *Biochem. Biophys. Res. Commun.* 2004, 325: 1443–1448.
- [85]. Schwartz D, Gygi SP: An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale datasets. *Nat. Biotechnol.* 2005, 23: 1391–1398.
- [86]. Polevoda B, Sherman F: N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J. Mol. Biol.* 2003, 325: 595–622.

[87]. Augen J, Wold F: How much sequence information is needed for the regulation of amino-terminal acetylation of eukaryotic proteins? *Trends Biochem. Sci.* 1986, 11: 494–497.

[88]. Perrot M, Sagliocco F, Mini T, Monribot C, Schneider U, Shevchenko A, Mann M, Jenö P, Boucherie H: Two-dimensional gel protein database of *Saccharomyces cerevisiae* (update 1999). *Electrophoresis* 1999, 20: 2280–2298.

[89]. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22:4673–4680.

[90]. Blom N, Hansen J, Blaas D, Brunak S: Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci.* 1996, 5: 2203–2216.

[91]. Bunn HF, Shapiro R, McManus M, Garrick L, McDonald MJ, Gallop PM, Gabbay KH: Structural heterogeneity of human hemoglobin A due to nonenzymatic glycosylation. *J. Biol. Chem.*, 1979, 254: 3892–3898.

[92]. Schneider TD, Stephens RM: Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990, 18: 6097–6100.

[93]. Nakayama K: *Biochem. J.* 1997, 327: 625–635.

Lehmann M, Andre F, Bellan C, Remacle-Bonnet M, Garrouste F, Parat F, Lissitsky JC, Marvaldi J, Pommier G: *Endocrinology* 1998, 139:3763–3771.

[94]. Bairoch A, Apweiler R: *Nucleic Acids Res.* 2000, 28: 45–48.

[95]. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: *Nucleic Acids Res.* 1997, 25: 4876–4882.

- [96]. Baldi P, Brunak S: The Machine Learning Approach, Bioinformatics: 2nd edn. MIT Press, Cambridge, MA.
- [97]. Rumelhart DE, Hinton GE, Williams RJ: In Rumelhart,D.E., McClelland,J.L. and the PDP Research Group (eds): Parallel Distributed Processing: Explorations in the Microstructure of Cognition. MIT Press, Cambridge, MA, 1986, 1: 318-362.
- [98]. Qian N, Sejnowski TJ: J. Mol. Biol. 1988, 202: 865-884.
Brunak S, Engelbrecht J, Knudsen S: J. Mol. Biol. 1991, 220: 49-65.
- [99]. Crick F: Central Dogma of Molecular Biology, Nature 1970, 227: 561-563.
- [100]. Basu A, Rose LK, Zhang J, Beavis CR, Ueberheide B, Garcia AB, Chait B, Zhao Y, Hunt FD, Segal E, Allis CD, Hake BS: Proteome-wide prediction of acetylation substrates. PNAS 2009, 106: 13785-13790.
- [101]. Kamp RG, Savenije HHG: Optimising training data for ANNs with Genetic Algorithms. Hydrol.Earth Syst.Sci. 2006, 10: 603-608.
- [102]. Gan Y, Guan J, Zhou S: A pattern-based nearest neighbor search approach for promoter prediction using DNA structural profiles. Bioinformatics 2009, 25(16): 2006-2012.
- [103]. Ghahramani Z, Jordan IM: Factorial Hidden Markov Models. 1997.
- [104]. <http://www.uniprot.org/>
- [105]. <http://www.cbs.dtu.dk/services/NetPhosk/>
- [106]. <http://www.ncbi.nlm.nih.gov/pubmed>
- [107]. Santos J, Neves M: A local maximum likelihood estimator for Poisson regression. Mertika 2008, 68: 257-270.

[108]. Blom N, Sicheritz-Pontin T, Gupta R, Gammeltoft S, Brunak S: Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 2004, 4: 1633-1649.

[109]. Blom N, Gammeltoft S, Brunak S: Sequence and Structure-based Prediction of Eukaryotic Protein Phosphorylation Sites. *J. Mol. Biol.* 1999, 294: 1351-1362.

6.ΠΑΡΑΡΤΗΜΑ

Παρουσιάζονται οι πίνακες των αποτελεσμάτων παλινδρόμησης Poisson που εκτελέστηκε για κάθε μία από τις 18 διαφορετικές κινάσες. Σημειώνεται πως οι κινάσες με Icat_ph_1 έως και Icat_ph_20 αντιστοιχούν σε οργανισμούς των βακτηρίων, οι κινάσεις με Icat_ph_21 έως και Icat_ph_27 αντιστοιχούν σε ευκαρυώτες και τέλος οι κινάσες με Icat_ph_28 έως και Icat_ph_30 αντιστοιχούν στα αρχαία.

Πίνακας 3.3: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση insr

```

. xi:poisson insr i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -906235.06
Iteration 1:  log likelihood = -902912.44
Iteration 2:  log likelihood = -902892.38
Iteration 3:  log likelihood = -902892.37

Poisson regression                               Number of obs   =    1020130
LR chi2(30)    =    40553.74
Prob > chi2    =     0.0000
Log likelihood = -902892.37                       Pseudo R2      =     0.0220

```

insr	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
_Icat_ph_1	.9626695	.0105704	-3.46	0.001	.9421733	.9836115
_Icat_ph_2	.5827183	.0246841	-12.75	0.000	.5362924	.6331633
_Icat_ph_3	.7506198	.0069487	-30.99	0.000	.7371234	.7643632
_Icat_ph_4	1.011866	.00642	1.86	0.063	.9993607	1.024527
_Icat_ph_5	.7619834	.0057623	-35.95	0.000	.7507728	.7733614
_Icat_ph_6	1.605037	.0600564	12.65	0.000	1.491541	1.727169
_Icat_ph_7	1.978024	.0190144	70.96	0.000	1.941105	2.015645
_Icat_ph_8	.7214277	.0084618	-27.84	0.000	.7050321	.7382046
_Icat_ph_9	1.609893	.0225021	34.07	0.000	1.566388	1.654606
_Icat_ph_10	1.181646	.0256793	7.68	0.000	1.132372	1.233063
_Icat_ph_11	1.049627	.0243716	2.09	0.037	1.00293	1.098498
_Icat_ph_12	1.967742	.0306297	43.49	0.000	1.908616	2.028701
_Icat_ph_13	1.586501	.0310901	23.55	0.000	1.526721	1.648622
_Icat_ph_14	.650836	.0341596	-8.18	0.000	.587213	.7213525
_Icat_ph_15	.6026015	.0063251	-48.25	0.000	.5903311	.6151268
_Icat_ph_16	1.483798	.0111403	52.56	0.000	1.462123	1.505794
_Icat_ph_17	1.322266	.057027	6.48	0.000	1.215089	1.438897
_Icat_ph_18	.671869	.0282372	-9.46	0.000	.6187432	.7295562
_Icat_ph_19	1.94156	.0499971	25.77	0.000	1.846	2.042068
_Icat_ph_20	.9703549	.0153951	-1.90	0.058	.9406453	1.001003
_Icat_ph_21	1.063243	.0089935	7.25	0.000	1.045762	1.081017
_Icat_ph_22	1.065954	.0079022	8.62	0.000	1.050578	1.081556
_Icat_ph_23	1.212761	.0079114	29.57	0.000	1.197354	1.228367

_Icat_ph_24		1.192368	.0097486	21.52	0.000	1.173414	1.211629
_Icat_ph_25		2.310984	.0157717	122.74	0.000	2.280278	2.342103
_Icat_ph_26		1.485258	.0483805	12.14	0.000	1.393398	1.583174
_Icat_ph_27		.7203516	.0119715	-19.74	0.000	.6972659	.7442016
_Icat_ph_28		1.613434	.0194587	39.66	0.000	1.575743	1.652026
_Icat_ph_29		1.522295	.0133723	47.84	0.000	1.49631	1.548731
_Icat_ph_30		2.793648	.1409802	20.36	0.000	2.530558	3.084091
length		(exposure)					

Πίνακας 3.4: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση egfr

```

. xi:poisson egfr i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -843241.61
Iteration 1:  log likelihood = -843230.35
Iteration 2:  log likelihood = -843230.35

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   7367.48
Prob > chi2    =   0.0000
Log likelihood = -843230.35                       Pseudo R2      =   0.0043

```

egfr	IRR	Std. Err.	z	P> z	[95% Conf. Interval]		
_Icat_ph_1		1.299041	.0131769	25.79	0.000	1.27347	1.325126
_Icat_ph_2		.7151444	.0287536	-8.34	0.000	.6609517	.7737805
_Icat_ph_3		.9967071	.0086498	-0.38	0.704	.9798971	1.013805
_Icat_ph_4		1.190909	.0075544	27.54	0.000	1.176195	1.205808
_Icat_ph_5		.9033545	.0067497	-13.60	0.000	.8902218	.916681
_Icat_ph_6		1.449862	.0599308	8.99	0.000	1.337032	1.572213
_Icat_ph_7		1.552805	.0173858	39.30	0.000	1.5191	1.587257
_Icat_ph_8		.9348678	.0102811	-6.12	0.000	.9149327	.9552372
_Icat_ph_9		1.2787	.0208998	15.04	0.000	1.238386	1.320326
_Icat_ph_10		1.201448	.0272085	8.10	0.000	1.149286	1.255976
_Icat_ph_11		1.178955	.0271803	7.14	0.000	1.126868	1.23345
_Icat_ph_12		1.373473	.0266293	16.37	0.000	1.32226	1.42667
_Icat_ph_13		1.158358	.0277597	6.13	0.000	1.105208	1.214064
_Icat_ph_14		1.342006	.0516778	7.64	0.000	1.244447	1.447213
_Icat_ph_15		.8586154	.0081579	-16.04	0.000	.8427742	.8747543
_Icat_ph_16		1.338009	.0109666	35.53	0.000	1.316687	1.359677
_Icat_ph_17		1.294959	.05928	5.65	0.000	1.183832	1.416518
_Icat_ph_18		.9582884	.0354903	-1.15	0.250	.8911934	1.030435

_Icat_ph_19		1.283064	.0425318	7.52	0.000	1.202353	1.369192
_Icat_ph_20		1.125579	.0175011	7.61	0.000	1.091795	1.160408
_Icat_ph_21		1.150858	.009914	16.31	0.000	1.13159	1.170454
_Icat_ph_22		.9300085	.0076175	-8.86	0.000	.9151977	.945059
_Icat_ph_23		1.115394	.0078565	15.50	0.000	1.100102	1.1309
_Icat_ph_24		1.126895	.0098941	13.61	0.000	1.107669	1.146455
_Icat_ph_25		1.228513	.0111299	22.72	0.000	1.206891	1.250522
_Icat_ph_26		1.203573	.0456896	4.88	0.000	1.117274	1.296539
_Icat_ph_27		.8308595	.0135625	-11.35	0.000	.8046982	.8578714
_Icat_ph_28		1.400111	.0189117	24.92	0.000	1.363531	1.437672
_Icat_ph_29		1.242016	.0124591	21.61	0.000	1.217835	1.266677
_Icat_ph_30		1.880975	.1214061	9.79	0.000	1.657459	2.134633
length		(exposure)					

Πίνακας 3.5: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση cki

```
. xi:poisson cki i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood =   -1255581
Iteration 1:   log likelihood = -1255577.4
Iteration 2:   log likelihood = -1255577.4

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =    7881.14
Prob > chi2    =    0.0000
Log likelihood = -1255577.4                       Pseudo R2      =    0.0031
```

cki	IRR	Std. Err.	z	P> z	[95% Conf. Interval]		
_Icat_ph_1		.8202011	.0069468	-23.40	0.000	.8066979	.8339303
_Icat_ph_2		.6747276	.019101	-13.90	0.000	.63831	.713223
_Icat_ph_3		.7108363	.0048369	-50.16	0.000	.701419	.7203799
_Icat_ph_4		.8224606	.0040117	-40.07	0.000	.8146354	.8303611
_Icat_ph_5		.7761815	.0041897	-46.94	0.000	.7680131	.7844367
_Icat_ph_6		.7531599	.0294807	-7.24	0.000	.6975396	.8132152
_Icat_ph_7		.9967458	.0093007	-0.35	0.727	.9786825	1.015142
_Icat_ph_8		.8405595	.0066229	-22.04	0.000	.8276786	.8536409
_Icat_ph_9		.8614322	.0116157	-11.06	0.000	.838964	.8845022
_Icat_ph_10		.9959112	.0168997	-0.24	0.809	.9633331	1.029591
_Icat_ph_11		1.006615	.0171422	0.39	0.699	.9735716	1.04078
_Icat_ph_12		.9347261	.0149222	-4.23	0.000	.905932	.9644354
_Icat_ph_13		.9923053	.0175403	-0.44	0.662	.9585157	1.027286
_Icat_ph_14		.5166271	.0218594	-15.61	0.000	.4755118	.5612975
_Icat_ph_15		.8408533	.0055123	-26.44	0.000	.8301185	.8517269

_Icat_ph_16		.8745818	.0058004	-20.21	0.000	.8632868	.8860247
_Icat_ph_17		.9059081	.0338788	-2.64	0.008	.841882	.9748034
_Icat_ph_18		.7827084	.021918	-8.75	0.000	.7409074	.8268677
_Icat_ph_19		.7973797	.0228718	-7.89	0.000	.7537885	.8434916
_Icat_ph_20		.9195032	.010753	-7.18	0.000	.8986673	.9408221
_Icat_ph_21		1.034056	.0063555	5.45	0.000	1.021674	1.046588
_Icat_ph_22		.973849	.0053635	-4.81	0.000	.9633932	.9844183
_Icat_ph_23		1.036435	.0051193	7.25	0.000	1.02645	1.046518
_Icat_ph_24		.9680372	.0061776	-5.09	0.000	.9560048	.980221
_Icat_ph_25		.9015001	.0063446	-14.73	0.000	.8891503	.9140215
_Icat_ph_26		.9058861	.0270783	-3.31	0.001	.8543384	.9605439
_Icat_ph_27		.9112788	.009746	-8.69	0.000	.8923758	.9305823
_Icat_ph_28		.8285985	.0097664	-15.95	0.000	.8096762	.8479631
_Icat_ph_29		.966277	.0073791	-4.49	0.000	.9519218	.9808486
_Icat_ph_30		.5807743	.0460868	-6.85	0.000	.4971198	.678506
length		(exposure)					

Πίνακας 3.6: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση ckii

```

. xi:poisson ckii i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -2353638.6
Iteration 1:  log likelihood = -2353308.1
Iteration 2:  log likelihood = -2353307.6
Iteration 3:  log likelihood = -2353307.6

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   62117.24
Prob > chi2    =   0.0000
Log likelihood = -2353307.6                       Pseudo R2      =   0.0130

```

ckii	IRR	Std. Err.	z	P> z	[95% Conf. Interval]		
_Icat_ph_1		.8845065	.0044171	-24.58	0.000	.8758914	.8932063
_Icat_ph_2		.5705343	.0107012	-29.92	0.000	.5499412	.5918985
_Icat_ph_3		.6406868	.0027766	-102.73	0.000	.6352678	.6461519
_Icat_ph_4		.8547957	.0025095	-53.44	0.000	.8498912	.8597284
_Icat_ph_5		.7417189	.0024823	-89.28	0.000	.7368696	.7466002
_Icat_ph_6		.8805017	.0194433	-5.76	0.000	.8432065	.9194465
_Icat_ph_7		1.189516	.006248	33.04	0.000	1.177333	1.201826
_Icat_ph_8		.751357	.0037924	-56.64	0.000	.7439608	.7588268
_Icat_ph_9		1.042589	.0078263	5.56	0.000	1.027362	1.058042
_Icat_ph_10		.9770718	.0102034	-2.22	0.026	.9572767	.9972762
_Icat_ph_11		.9591215	.0101958	-3.93	0.000	.9393448	.9793146
_Icat_ph_12		1.144964	.010104	15.34	0.000	1.125331	1.16494

_Icat_ph_13		1.050911	.0110139	4.74	0.000	1.029545	1.072721
_Icat_ph_14		.4857853	.0129223	-27.14	0.000	.4611069	.5117844
_Icat_ph_15		.7697384	.003191	-63.13	0.000	.7635095	.7760181
_Icat_ph_16		1.056356	.0039574	14.63	0.000	1.048628	1.064141
_Icat_ph_17		.9517581	.0211755	-2.22	0.026	.9111467	.9941795
_Icat_ph_18		.6664288	.012322	-21.95	0.000	.6427104	.6910225
_Icat_ph_19		1.049441	.016022	3.16	0.002	1.018504	1.081318
_Icat_ph_20		.6925525	.0056499	-45.03	0.000	.6815669	.7037151
_Icat_ph_21		1.032198	.0038708	8.45	0.000	1.024639	1.039813
_Icat_ph_22		1.009349	.0033448	2.81	0.005	1.002814	1.015925
_Icat_ph_23		1.176792	.0033968	56.40	0.000	1.170153	1.183468
_Icat_ph_24		1.114993	.0041004	29.60	0.000	1.106986	1.123059
_Icat_ph_25		1.306797	.0048187	72.57	0.000	1.297387	1.316276
_Icat_ph_26		1.164792	.0187443	9.48	0.000	1.128627	1.202116
_Icat_ph_27		1.022039	.0063189	3.53	0.000	1.009729	1.034499
_Icat_ph_28		.7546794	.0056687	-37.47	0.000	.7436504	.765872
_Icat_ph_29		1.271848	.0052692	58.04	0.000	1.261562	1.282218
_Icat_ph_30		.8540996	.0340861	-3.95	0.000	.7898382	.9235894
length		(exposure)					

Πίνακας 3.7: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση pka

```
. xi:poisson pka i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood = -2466808.4
Iteration 1:   log likelihood = -2466792.8
Iteration 2:   log likelihood = -2466792.8

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   9579.36
Prob > chi2    =   0.0000
Log likelihood = -2466792.8                       Pseudo R2      =   0.0019
```

pka	IRR	Std. Err.	z	P> z	[95% Conf. Interval]		
_Icat_ph_1		1.020684	.0044383	4.71	0.000	1.012022	1.02942
_Icat_ph_2		.8121575	.0118582	-14.25	0.000	.7892453	.8357348
_Icat_ph_3		.9074069	.0031601	-27.90	0.000	.9012343	.9136217
_Icat_ph_4		.989596	.0025687	-4.03	0.000	.9845743	.9946434
_Icat_ph_5		.8858628	.0025774	-41.65	0.000	.8808256	.8909289
_Icat_ph_6		.860728	.0178222	-7.24	0.000	.8264964	.8963773
_Icat_ph_7		.9739553	.0051919	-4.95	0.000	.9638323	.9841846
_Icat_ph_8		.8867693	.0038586	-27.62	0.000	.8792387	.8943644
_Icat_ph_9		.9749583	.0070053	-3.53	0.000	.9613244	.9887855

_Icat_ph_10		1.191873	.0104788	19.96	0.000	1.171511	1.212589
_Icat_ph_11		1.014599	.0097299	1.51	0.131	.9957069	1.03385
_Icat_ph_12		.9175859	.0083554	-9.45	0.000	.9013548	.9341092
_Icat_ph_13		1.150246	.0106965	15.05	0.000	1.129471	1.171403
_Icat_ph_14		.7483681	.014887	-14.57	0.000	.7197516	.7781223
_Icat_ph_15		.8376431	.0031091	-47.73	0.000	.8315716	.843759
_Icat_ph_16		.900857	.0033367	-28.19	0.000	.894341	.9074206
_Icat_ph_17		.9683333	.0198043	-1.57	0.116	.9302854	1.007937
_Icat_ph_18		.8948518	.0132575	-7.50	0.000	.8692412	.921217
_Icat_ph_19		.9891354	.0144167	-0.75	0.454	.961279	1.017799
_Icat_ph_20		.9300318	.0061154	-11.03	0.000	.9181228	.9420952
_Icat_ph_21		.9318627	.0033758	-19.48	0.000	.9252697	.9385027
_Icat_ph_22		1.04413	.0031683	14.23	0.000	1.037939	1.050359
_Icat_ph_23		.9955544	.0028184	-1.57	0.116	.9900457	1.001094
_Icat_ph_24		.9576862	.0034699	-11.93	0.000	.9509094	.9645113
_Icat_ph_25		.9685498	.0037392	-8.28	0.000	.9612488	.9759062
_Icat_ph_26		1.098202	.0168695	6.10	0.000	1.065631	1.131768
_Icat_ph_27		.8207835	.0052111	-31.11	0.000	.8106333	.8310608
_Icat_ph_28		1.055055	.0062772	9.01	0.000	1.042823	1.06743
_Icat_ph_29		.8541037	.0038914	-34.61	0.000	.8465107	.8617649
_Icat_ph_30		.8590978	.0316957	-4.12	0.000	.7991684	.9235214
length		(exposure)					

Πίνακας 3.8: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

pkg

```
. xi:poisson pkg i.cat_ph,exposure(length) irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood =  -958303.7
Iteration 1:   log likelihood =  -958302.15
Iteration 2:   log likelihood =  -958302.15

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   4699.43
Prob > chi2    =   0.0000
Log likelihood = -958302.15                       Pseudo R2      =   0.0024

-----+-----
pkg |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   .8431062   .0085151   -16.90  0.000   .8265811   .8599617
_Icat_ph_2 |   .8576032   .0260297    -5.06  0.000   .8080737   .9101685
_Icat_ph_3 |   .8601543   .006535   -19.83  0.000   .8474408   .8730586
_Icat_ph_4 |   .8102995   .0047952   -35.55  0.000   .8009554   .8197526
_Icat_ph_5 |   .8728527   .0054516   -21.77  0.000   .862233   .8836033
```


_Icat_ph_6		.7043624	.0344101	-7.17	0.000	.640048	.7751394
_Icat_ph_7		.8896528	.0105562	-9.85	0.000	.8692017	.910585
_Icat_ph_8		.8633956	.0081165	-15.62	0.000	.8476332	.8794512
_Icat_ph_9		.7925184	.0134265	-13.73	0.000	.766635	.8192757
_Icat_ph_10		.8474274	.0187807	-7.47	0.000	.8114059	.885048
_Icat_ph_11		.8653053	.01915	-6.54	0.000	.8285744	.9036645
_Icat_ph_12		.7300312	.0158662	-14.48	0.000	.699587	.7618003
_Icat_ph_13		.8908788	.0200372	-5.14	0.000	.8524596	.9310295
_Icat_ph_14		.5649723	.0276005	-11.69	0.000	.5133854	.6217429
_Icat_ph_15		.9656396	.0072258	-4.67	0.000	.9515805	.9799063
_Icat_ph_16		.7949076	.0066216	-27.55	0.000	.7820348	.8079923
_Icat_ph_17		.8721267	.0401232	-2.97	0.003	.7969279	.9544213
_Icat_ph_18		.892638	.0282752	-3.59	0.000	.8389048	.9498129
_Icat_ph_19		.9119638	.0295472	-2.84	0.004	.8558528	.9717536
_Icat_ph_20		.9859382	.0134689	-1.04	0.300	.9598899	1.012693
_Icat_ph_21		1.037068	.0076856	4.91	0.000	1.022113	1.052241
_Icat_ph_22		.898441	.006157	-15.63	0.000	.8864541	.9105899
_Icat_ph_23		1.016304	.0061006	2.69	0.007	1.004417	1.028331
_Icat_ph_24		1.043023	.0077977	5.63	0.000	1.027851	1.058418
_Icat_ph_25		.7918423	.0071093	-26.00	0.000	.7780302	.8058997
_Icat_ph_26		.7798951	.0303113	-6.40	0.000	.7226924	.8416256
_Icat_ph_27		1.009945	.012431	0.80	0.421	.9858722	1.034606
_Icat_ph_28		.7144385	.0109052	-22.03	0.000	.6933813	.7361351
_Icat_ph_29		.7577621	.0077743	-27.04	0.000	.742677	.7731536
_Icat_ph_30		.7239022	.0621218	-3.77	0.000	.6118344	.8564972
length		(exposure)					

**Πίνακας 3.9: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση
pkc**

```
. xi:poisson pkc i.cat_ph,exposure(length) irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -5082032.7
Iteration 1:  log likelihood = -5081880.6
Iteration 2:  log likelihood = -5081880.6

Poisson regression                               Number of obs   =   1020130
LR chi2(30)   =  137288.93
Prob > chi2   =    0.0000
Log likelihood = -5081880.6                       Pseudo R2      =    0.0133

-----+-----
pkc |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   .8339049   .0021677   -69.87  0.000   .829667   .8381644
```

_Icat_ph_2		.6500487	.0057958	-48.31	0.000	.6387878	.6615081
_Icat_ph_3		.7309562	.0015201	-150.70	0.000	.7279829	.7339417
_Icat_ph_4		.8026657	.0012205	-144.56	0.000	.8002771	.8050615
_Icat_ph_5		.7332187	.0012497	-182.06	0.000	.7307733	.7356722
_Icat_ph_6		.7335123	.0089949	-25.27	0.000	.7160929	.7513556
_Icat_ph_7		.9776119	.0028455	-7.78	0.000	.9720506	.9832049
_Icat_ph_8		.7103229	.0018641	-130.33	0.000	.7066787	.7139859
_Icat_ph_9		.9180268	.0037128	-21.15	0.000	.9107787	.9253326
_Icat_ph_10		.9947084	.005222	-1.01	0.312	.9845259	1.004996
_Icat_ph_11		.8372091	.0048228	-30.84	0.000	.8278098	.8467152
_Icat_ph_12		.8645076	.004432	-28.40	0.000	.8558645	.873238
_Icat_ph_13		1.015516	.005488	2.85	0.004	1.004817	1.02633
_Icat_ph_14		.4321753	.0061795	-58.67	0.000	.4202319	.4444582
_Icat_ph_15		.7272434	.0015656	-147.95	0.000	.7241814	.7303184
_Icat_ph_16		.8615478	.0017777	-72.22	0.000	.8580706	.865039
_Icat_ph_17		.8123168	.0099163	-17.03	0.000	.7931118	.8319868
_Icat_ph_18		.7135776	.0064678	-37.23	0.000	.7010128	.7263675
_Icat_ph_19		1.149579	.0085088	18.83	0.000	1.133023	1.166377
_Icat_ph_20		.8543484	.0031983	-42.05	0.000	.8481027	.86064
_Icat_ph_21		1.0246	.0019542	12.74	0.000	1.020777	1.028437
_Icat_ph_22		.9850676	.0016703	-8.87	0.000	.9817993	.9883469
_Icat_ph_23		1.075856	.0016224	48.49	0.000	1.072681	1.079041
_Icat_ph_24		1.06134	.0020183	31.31	0.000	1.057392	1.065303
_Icat_ph_25		1.029901	.00212	14.31	0.000	1.025754	1.034065
_Icat_ph_26		.8742242	.0082235	-14.29	0.000	.8582541	.8904914
_Icat_ph_27		.9112892	.0030134	-28.09	0.000	.9054021	.9172145
_Icat_ph_28		.8003594	.0029649	-60.11	0.000	.7945693	.8061918
_Icat_ph_29		.7865532	.0020328	-92.90	0.000	.782579	.7905476
_Icat_ph_30		.7605377	.0163096	-12.76	0.000	.729234	.7931852
length		(exposure)					

Πίνακας 3.10: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

pkb

```
. xi:poisson pkb i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood = -156267.34
Iteration 1:   log likelihood = -156265.57
Iteration 2:   log likelihood = -156265.57

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   2181.21
Prob > chi2    =   0.0000
Log likelihood = -156265.57                       Pseudo R2      =   0.0069
```

pkb	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
_Icat_ph_1	.8641603	.0309363	-4.08	0.000	.8056046	.9269722
_Icat_ph_2	1.162503	.108841	1.61	0.108	.9676071	1.396654
_Icat_ph_3	.9157115	.0243161	-3.32	0.001	.8692718	.9646322
_Icat_ph_4	.8244137	.0173819	-9.16	0.000	.79104	.8591953
_Icat_ph_5	1.045811	.0219903	2.13	0.033	1.003587	1.089812
_Icat_ph_6	.3222949	.0833067	-4.38	0.000	.1941937	.534899
_Icat_ph_7	.5207443	.0284563	-11.94	0.000	.4678538	.579614
_Icat_ph_8	1.13906	.0340845	4.35	0.000	1.074177	1.207862
_Icat_ph_9	.3619921	.0321688	-11.43	0.000	.3041279	.4308657
_Icat_ph_10	.4699965	.0498647	-7.12	0.000	.3817558	.5786336
_Icat_ph_11	.7178012	.0623811	-3.82	0.000	.6053824	.8510961
_Icat_ph_12	.3285061	.0378896	-9.65	0.000	.2620395	.4118322
_Icat_ph_13	.7978182	.0678728	-2.66	0.008	.6752889	.94258
_Icat_ph_14	.8789468	.1235321	-0.92	0.359	.6673135	1.157698
_Icat_ph_15	1.404815	.0326359	14.63	0.000	1.342284	1.470259
_Icat_ph_16	.5686239	.0195988	-16.38	0.000	.5314798	.608364
_Icat_ph_17	.9903361	.1532775	-0.06	0.950	.7312054	1.3413
_Icat_ph_18	1.057658	.1104121	0.54	0.591	.861957	1.297791
_Icat_ph_19	.3040478	.0609196	-5.94	0.000	.2053021	.4502879
_Icat_ph_20	.967841	.0477961	-0.66	0.508	.8785533	1.066203
_Icat_ph_21	1.068777	.028048	2.53	0.011	1.015194	1.125189
_Icat_ph_22	.9630911	.0230416	-1.57	0.116	.9189728	1.009327
_Icat_ph_23	.9067793	.0202984	-4.37	0.000	.8678553	.947449
_Icat_ph_24	1.049858	.028054	1.82	0.069	.9962884	1.106308
_Icat_ph_25	.5891502	.0215777	-14.45	0.000	.548341	.6329965
_Icat_ph_26	.9910639	.1225748	-0.07	0.942	.7777232	1.262927
_Icat_ph_27	1.840152	.0619496	18.11	0.000	1.722652	1.965666
_Icat_ph_28	.7403562	.0398176	-5.59	0.000	.6662874	.8226589
_Icat_ph_29	.6665042	.0259592	-10.42	0.000	.6175187	.7193756
_Icat_ph_30	.2734503	.1367601	-2.59	0.010	.102605	.7287658
length (exposure)						

Πίνακας 3.11: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση dnapk

```
. xi:poisson dnapk i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood = -1448212.8
Iteration 1:   log likelihood = -1448206.3
Iteration 2:   log likelihood = -1448206.3

Poisson regression                               Number of obs   =   1020130
```

```

LR chi2(30)      =    36185.56
Prob > chi2     =         0.0000
Log likelihood = -1448206.3          Pseudo R2      =         0.0123

```

```

-----
dnapk |          IRR   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
_Icat_ph_1 |    1.066497   .0066791    10.28  0.000    1.053486    1.079668
_Icat_ph_2 |    .5151022   .0138328   -24.70  0.000    .4886915    .5429401
_Icat_ph_3 |    .7308931   .0040795   -56.17  0.000    .722941    .7389325
_Icat_ph_4 |    .8919946   .0035142   -29.01  0.000    .8851334    .8989089
_Icat_ph_5 |    .6349008   .0030671   -94.04  0.000    .6289177    .6409409
_Icat_ph_6 |    .3897349   .0175759   -20.89  0.000    .3567653    .4257512
_Icat_ph_7 |    .7476037   .006609    -32.90  0.000    .7347619    .76067
_Icat_ph_8 |    .5922281   .004527    -68.53  0.000    .5834216    .6011676
_Icat_ph_9 |    .7300472   .0088443   -25.97  0.000    .7129169    .7475891
_Icat_ph_10 |   .9741392   .013865    -1.84  0.066    .9473398    1.001697
_Icat_ph_11 |   .6494855   .0113715   -24.65  0.000    .6275758    .6721601
_Icat_ph_12 |   .5376942   .0093304   -35.76  0.000    .5197144    .556296
_Icat_ph_13 |   .7117785   .0122836   -19.70  0.000    .6881057    .7362657
_Icat_ph_14 |   .3734811   .0154132   -23.87  0.000    .3444612    .4049458
_Icat_ph_15 |   .5597471   .0036176   -89.78  0.000    .5527015    .5668826
_Icat_ph_16 |   .650847    .0040573   -68.89  0.000    .6429432    .6588481
_Icat_ph_17 |   .7773243   .0260307    -7.52  0.000    .7279434    .830055
_Icat_ph_18 |   .6786939   .0169248   -15.54  0.000    .6463196    .7126899
_Icat_ph_19 |   .8300486   .0193675    -7.98  0.000    .7929439    .8688895
_Icat_ph_20 |   .7842241   .0082059   -23.23  0.000    .7683047    .8004733
_Icat_ph_21 |   1.000881   .0051713     0.17  0.865    .9907963    1.011068
_Icat_ph_22 |   .7981654   .0039351   -45.73  0.000    .7904898    .8059156
_Icat_ph_23 |   .9740751   .0040786    -6.27  0.000    .966114    .9821019
_Icat_ph_24 |   .9283017   .0050001   -13.81  0.000    .9185532    .9381537
_Icat_ph_25 |   .7673597   .0048003   -42.33  0.000    .7580088    .7768261
_Icat_ph_26 |   .5646373   .0177103   -18.22  0.000    .5309711    .6004381
_Icat_ph_27 |   .5889311   .0064229   -48.55  0.000    .576476    .6016552
_Icat_ph_28 |   .5312482   .0064237   -52.31  0.000    .518806    .5439887
_Icat_ph_29 |   .5332371   .004411    -76.01  0.000    .5246615    .5419529
_Icat_ph_30 |   .4652871   .0342255   -10.40  0.000    .4028177    .5374443
length | (exposure)

```

Πίνακας 3.12: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση gsk3

```

. xi:poisson gsk3 i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30    (naturally coded; _Icat_ph_0 omitted)

Iteration 0:    log likelihood = -274324.47
Iteration 1:    log likelihood = -274323.49

```


Πίνακας 3.13: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

src								
. xi:poisson src i.cat_ph,exposure(length)irr								
i.cat_ph _Icat_ph_0-30 (naturally coded; _Icat_ph_0 omitted)								
Iteration 0: log likelihood = -288016.6								
Iteration 1: log likelihood = -288016.34								
Iteration 2: log likelihood = -288016.34								
Poisson regression								
					Number of obs	=	1020130	
LR chi2(30)	=	2054.43						
Prob > chi2	=	0.0000						
Log likelihood = -288016.34						Pseudo R2	=	0.0036

src	IRR	Std. Err.	z	P> z	[95% Conf. Interval]			
-----+-----								
_Icat_ph_1	1.125843	.0265851	5.02	0.000	1.074925	1.179173		
_Icat_ph_2	.8952166	.0705721	-1.40	0.160	.7670538	1.044793		
_Icat_ph_3	.9862542	.0188291	-0.72	0.468	.9500318	1.023858		
_Icat_ph_4	1.136879	.0160189	9.10	0.000	1.105912	1.168713		
_Icat_ph_5	.9628277	.0154151	-2.37	0.018	.9330838	.9935198		
_Icat_ph_6	1.778046	.1455621	7.03	0.000	1.514461	2.087506		
_Icat_ph_7	1.640145	.0392917	20.65	0.000	1.564915	1.718992		
_Icat_ph_8	.9927885	.0233101	-0.31	0.758	.9481368	1.039543		
_Icat_ph_9	1.396367	.0479852	9.72	0.000	1.305415	1.493656		
_Icat_ph_10	1.130456	.0577666	2.40	0.016	1.022721	1.24954		
_Icat_ph_11	1.305428	.0627656	5.54	0.000	1.188028	1.434429		
_Icat_ph_12	1.672393	.0646855	13.30	0.000	1.550299	1.804104		
_Icat_ph_13	1.110126	.059502	1.95	0.051	.9994213	1.233094		
_Icat_ph_14	1.180611	.106121	1.85	0.065	.9899093	1.408051		
_Icat_ph_15	.9915825	.0194781	-0.43	0.667	.9541317	1.030503		
_Icat_ph_16	1.43821	.0251388	20.79	0.000	1.389774	1.488335		
_Icat_ph_17	1.49898	.1398178	4.34	0.000	1.248532	1.799666		
_Icat_ph_18	1.184187	.0865571	2.31	0.021	1.02613	1.36659		
_Icat_ph_19	1.099733	.0861744	1.21	0.225	.9431646	1.282291		
_Icat_ph_20	1.184898	.0394106	5.10	0.000	1.110119	1.264715		
_Icat_ph_21	1.081995	.0209189	4.08	0.000	1.041761	1.123782		
_Icat_ph_22	1.047715	.0179883	2.71	0.007	1.013045	1.083571		
_Icat_ph_23	1.193369	.0180107	11.71	0.000	1.158586	1.229197		
_Icat_ph_24	1.119289	.0215876	5.84	0.000	1.077768	1.16241		
_Icat_ph_25	1.057445	.0223022	2.65	0.008	1.014624	1.102072		
_Icat_ph_26	1.563565	.1142873	6.11	0.000	1.354871	1.804403		
_Icat_ph_27	1.15893	.0355214	4.81	0.000	1.091359	1.230684		

_Icat_ph_28	1.835851	.0481102	23.18	0.000	1.743938	1.932609
_Icat_ph_29	1.656467	.0323782	25.82	0.000	1.594207	1.721158
_Icat_ph_30	2.210268	.2884271	6.08	0.000	1.711465	2.854445
length (exposure)						

Πίνακας 3.14: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση CaM-II

```

. xi:poisson camii i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood =  -26189.22
Iteration 1:  log likelihood =  -26188.681
Iteration 2:  log likelihood =  -26188.615
Iteration 3:  log likelihood =  -26188.604
Iteration 4:  log likelihood =  -26188.601
Iteration 5:  log likelihood =   -26188.6
Iteration 6:  log likelihood =   -26188.6
Iteration 7:  log likelihood =   -26188.6

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =          432.26
Prob > chi2    =          0.0000
Log likelihood =   -26188.6                       Pseudo R2      =          0.0082
-----
camii |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   .7656447   .0705806    -2.90   0.004     .6390865   .9172652
_Icat_ph_2 |   .4771618   .1692861    -2.09   0.037     .2380558   .9564287
_Icat_ph_3 |   .6041983   .0464682    -6.55   0.000     .5196544   .702497
_Icat_ph_4 |   .6124133   .0349798    -8.58   0.000     .5475524   .6849573
_Icat_ph_5 |   .5876893   .0372108    -8.40   0.000     .5191014   .6653396
_Icat_ph_6 |   .127876    .1279343    -2.06   0.040     .017997    .9086127
_Icat_ph_7 |   .3345898   .0551619    -6.64   0.000     .2422033   .4622164
_Icat_ph_8 |   .6052045   .0582629    -5.22   0.000     .5011377   .7308819
_Icat_ph_9 |   .3173275   .0733956    -4.96   0.000     .2016652   .4993264
_Icat_ph_10 |  .7459437   .1538373    -1.42   0.155     .4979219   1.117509
_Icat_ph_11 |  .474686    .1233559    -2.87   0.004     .2852371   .7899631
_Icat_ph_12 |  .3858945   .1002814    -3.66   0.000     .2318831   .6421968
_Icat_ph_13 |  .3367489   .1069519    -3.43   0.001     .1807022   .6275509
_Icat_ph_14 |  .512867    .2298578    -1.49   0.136     .2130647   1.23452
_Icat_ph_15 |  .633257    .0493486    -5.86   0.000     .54356     .7377557
_Icat_ph_16 |  .3776051   .0381552    -9.64   0.000     .3097619   .4603072
_Icat_ph_17 |  .7016937   .3144858    -0.79   0.429     .2915111   1.68904
_Icat_ph_18 |  .7445358   .2255552    -0.97   0.330     .4111651   1.348202
_Icat_ph_19 |  .5790695   .2054401    -1.54   0.124     .288898    1.160691
_Icat_ph_20 |  .6605786   .0954184    -2.87   0.004     .4977041   .8767541
_Icat_ph_21 |  .6534006   .0512632    -5.42   0.000     .5602704   .7620113

```

_Icat_ph_22		.5607161	.040555	-8.00	0.000	.4866066	.6461123
_Icat_ph_23		.7221524	.0428689	-5.48	0.000	.6428346	.811257
_Icat_ph_24		.5153761	.0454027	-7.52	0.000	.4336473	.6125083
_Icat_ph_25		.2306907	.0318371	-10.63	0.000	.1760182	.3023449
_Icat_ph_26		.4468593	.2002741	-1.80	0.072	.1856427	1.075632
_Icat_ph_27		.5302882	.0773435	-4.35	0.000	.3984397	.7057668
_Icat_ph_28		.4358031	.0737552	-4.91	0.000	.3127758	.607222
_Icat_ph_29		.3591411	.0454508	-8.09	0.000	.2802477	.460244
_Icat_ph_30		2.98e-06	.0011017	-0.03	0.973	0	.
length		(exposure)					

Πίνακας 3.15: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

unsp							
<pre>. xi:poisson unsp i.cat_ph,exposure(length) irr</pre>							
<pre>i.cat_ph _Icat_ph_0-30 (naturally coded; _Icat_ph_0 omitted)</pre>							
 Iteration 0: log likelihood = -11153800							
Iteration 1: log likelihood = -11153205							
Iteration 2: log likelihood = -11153205							
 Poisson regression							
					Number of obs	=	1020130
LR chi2(30)		=	596248.02				
Prob > chi2		=	0.0000				
Log likelihood = -11153205			Pseudo R2		=	0.0260	

unsp		IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
 -----+-----							
_Icat_ph_1		.7001845	.0011526	-216.51	0.000	.697929 .7024472	
_Icat_ph_2		.5681818	.0031705	-101.31	0.000	.5620017 .5744299	
_Icat_ph_3		.641392	.0008249	-345.33	0.000	.6397773 .6430107	
_Icat_ph_4		.6860615	.0006451	-400.74	0.000	.6847984 .6873269	
_Icat_ph_5		.6679466	.0006904	-390.45	0.000	.6665949 .669301	
_Icat_ph_6		.7047415	.0051613	-47.78	0.000	.6946978 .7149303	
_Icat_ph_7		.8278903	.0015227	-102.69	0.000	.8249111 .8308802	
_Icat_ph_8		.7194125	.0010991	-215.56	0.000	.7172616 .7215699	
_Icat_ph_9		.7669313	.0019793	-102.82	0.000	.7630617 .7708205	
_Icat_ph_10		.7953304	.0027263	-66.80	0.000	.7900048 .8006919	
_Icat_ph_11		.7582584	.0026843	-78.17	0.000	.7530156 .7635378	
_Icat_ph_12		.7727318	.0024495	-81.33	0.000	.7679458 .7775477	
_Icat_ph_13		.8291834	.0028965	-53.62	0.000	.8235258 .8348799	
_Icat_ph_14		.5159697	.0039544	-86.34	0.000	.5082771 .5237787	
_Icat_ph_15		.7261836	.0009158	-253.71	0.000	.7243909 .7279807	
_Icat_ph_16		.724826	.0009413	-247.81	0.000	.7229834 .7266733	
_Icat_ph_17		.7563861	.0056012	-37.70	0.000	.7454873 .7674443	
_Icat_ph_18		.6768418	.003687	-71.65	0.000	.6696539 .6841069	

_Icat_ph_19		.7262573	.0039496	-58.81	0.000	.7185574	.7340397
_Icat_ph_20		.727971	.0017217	-134.25	0.000	.7246044	.7313532
_Icat_ph_21		1.020546	.0011414	18.18	0.000	1.018312	1.022786
_Icat_ph_22		.9696111	.0009682	-30.90	0.000	.9677153	.9715107
_Icat_ph_23		1.049381	.0009343	54.14	0.000	1.047551	1.051213
_Icat_ph_24		.9863551	.0011308	-11.98	0.000	.9841412	.9885739
_Icat_ph_25		1.016654	.0012318	13.63	0.000	1.014243	1.019071
_Icat_ph_26		.9367998	.0049853	-12.27	0.000	.9270796	.946622
_Icat_ph_27		.967212	.0018213	-17.70	0.000	.9636488	.9707883
_Icat_ph_28		.7404196	.0016662	-133.55	0.000	.7371611	.7436924
_Icat_ph_29		.8049307	.0012055	-144.89	0.000	.8025714	.8072969
_Icat_ph_30		.7086707	.0092167	-26.48	0.000	.6908346	.7269672
length		(exposure)					

Πίνακας 3.16: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση cdc2

```

. xi:poisson cdc2 i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30          (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -2627551.8
Iteration 1:  log likelihood =  -2627542
Iteration 2:  log likelihood =  -2627542

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   63926.11
Prob > chi2    =   0.0000
Log likelihood =  -2627542                       Pseudo R2      =   0.0120
-----
cdc2 |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   .8112784   .0035562   -47.71  0.000   .8043383   .8182784
_Icat_ph_2 |   .6594985   .0097248   -28.23  0.000   .6407111   .6788368
_Icat_ph_3 |   .7248707   .0025198   -92.56  0.000   .7199488   .7298262
_Icat_ph_4 |   .7554463   .0019537  -108.44  0.000   .7516268   .7592851
_Icat_ph_5 |   .6958915   .0020148  -125.22  0.000   .6919538   .6998517
_Icat_ph_6 |   .4424927   .0116258   -31.03  0.000   .4202834   .4658757
_Icat_ph_7 |   .6481268   .0038019   -73.93  0.000   .6407178   .6556214
_Icat_ph_8 |   .7148772   .0031161   -77.00  0.000   .7087958   .7210108
_Icat_ph_9 |   .6239587   .0050614   -58.15  0.000   .6141169   .6339582
_Icat_ph_10 |  .9025628   .0082754   -11.18  0.000   .8864883   .9189288
_Icat_ph_11 |  .8043333   .0078707   -22.25  0.000   .789054   .8199084
_Icat_ph_12 |  .5320815   .0057602   -58.28  0.000   .5209106   .543492
_Icat_ph_13 |  .7801668   .0079888   -24.24  0.000   .7646652   .7959828
_Icat_ph_14 |  .5139457   .0112288   -30.47  0.000   .4924022   .5364317
_Icat_ph_15 |  .7387657   .0026317   -84.99  0.000   .7336256   .7439419

```

_Icat_ph_16	.5904393	.0023832	-130.54	0.000	.5857867	.5951288
_Icat_ph_17	.7185399	.0155307	-15.29	0.000	.6887361	.7496334
_Icat_ph_18	.7700103	.0111955	-17.98	0.000	.7483772	.7922687
_Icat_ph_19	.5455713	.0097291	-33.98	0.000	.5268319	.5649772
_Icat_ph_20	.7927669	.0051221	-35.94	0.000	.782791	.8028699
_Icat_ph_21	.9478773	.0031069	-16.33	0.000	.9418073	.9539864
_Icat_ph_22	.9958404	.0028014	-1.48	0.138	.9903648	1.001346
_Icat_ph_23	.958923	.0025057	-16.05	0.000	.9540245	.9638467
_Icat_ph_24	.8288155	.0029033	-53.60	0.000	.8231446	.8345255
_Icat_ph_25	.7033179	.0028361	-87.28	0.000	.6977813	.7088985
_Icat_ph_26	.7432092	.012622	-17.47	0.000	.7188776	.7683642
_Icat_ph_27	.6969325	.0043545	-57.79	0.000	.68845	.7055196
_Icat_ph_28	.7398485	.0047387	-47.04	0.000	.7306189	.7491947
_Icat_ph_29	.6209839	.0029733	-99.51	0.000	.6151835	.626839
_Icat_ph_30	.408865	.0199119	-18.36	0.000	.3716431	.4498148
length (exposure)						

Πίνακας 3.17: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση p38MAPK

```

. xi:poisson p38mapk i.cat_ph,exposure(length) irr
i.cat_ph          _Icat_ph_0-30          (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -828984.54
Iteration 1:  log likelihood = -828982.86
Iteration 2:  log likelihood = -828982.86

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   7193.33
Prob > chi2    =   0.0000
Log likelihood = -828982.86                       Pseudo R2      =   0.0043

-----+-----
p38mapk |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |  1.017063   .0102462     1.68  0.093     .9971776   1.037345
_Icat_ph_2 |  .8067269   .0273377    -6.34  0.000     .7548865   .8621273
_Icat_ph_3 |   .87887    .0071171   -15.82  0.000     .864927    .8930379
_Icat_ph_4 |  .8643058   .0054222   -23.25  0.000     .8537437   .8749987
_Icat_ph_5 |  .8113081   .005635    -30.11  0.000     .8003386   .822428
_Icat_ph_6 |  .6400201   .035524    -8.04  0.000     .5740478   .7135742
_Icat_ph_7 |  .7130482   .0101554   -23.75  0.000     .6934192   .7332329
_Icat_ph_8 |  .825523    .0085731   -18.46  0.000     .8088898   .8424982
_Icat_ph_9 |  .6894687   .0135311   -18.95  0.000     .6634519   .7165058
_Icat_ph_10 | .9805699   .0219237    -0.88  0.380     .9385281   1.024495
_Icat_ph_11 | .8594346   .0206722    -6.30  0.000     .819858    .9009217

```

_Icat_ph_12		.6082254	.0156566	-19.32	0.000	.5783003	.639699
_Icat_ph_13		.6488483	.0184675	-15.20	0.000	.6136437	.6860725
_Icat_ph_14		.91492	.0381033	-2.14	0.033	.8432056	.9927337
_Icat_ph_15		.8423734	.0072159	-20.02	0.000	.8283485	.8566358
_Icat_ph_16		.6618356	.0064552	-42.32	0.000	.6493039	.6746092
_Icat_ph_17		.8057856	.0417694	-4.17	0.000	.7279404	.8919555
_Icat_ph_18		1.02252	.0328089	0.69	0.488	.9601962	1.088889
_Icat_ph_19		.6967561	.0279413	-9.01	0.000	.6440891	.7537297
_Icat_ph_20		.9713528	.0144755	-1.95	0.051	.9433917	1.000143
_Icat_ph_21		.9286935	.0077932	-8.82	0.000	.913544	.9440942
_Icat_ph_22		.7909462	.0061671	-30.08	0.000	.7789508	.8031264
_Icat_ph_23		.9104856	.0061519	-13.88	0.000	.8985076	.9226234
_Icat_ph_24		.8506923	.0074848	-18.38	0.000	.8361482	.8654894
_Icat_ph_25		.530593	.0061564	-54.62	0.000	.5186628	.5427977
_Icat_ph_26		.5200098	.0267782	-12.70	0.000	.4700872	.5752341
_Icat_ph_27		.6943202	.0110343	-22.96	0.000	.6730268	.7162873
_Icat_ph_28		.8022985	.0125557	-14.08	0.000	.7780634	.8272885
_Icat_ph_29		.6985542	.0080518	-31.12	0.000	.6829499	.7145151
_Icat_ph_30		.5059271	.0562443	-6.13	0.000	.4068733	.6290957
length		(exposure)					

Πίνακας 3.18: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

rsk

```
. xi:poisson rsk i.cat_ph,exposure(length)irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -136765.1
Iteration 1:  log likelihood = -136764.61
Iteration 2:  log likelihood = -136764.61

Poisson regression                                Number of obs   =   1020130
LR chi2(30)    =   1016.48
Prob > chi2    =   0.0000
Log likelihood = -136764.61                       Pseudo R2      =   0.0037
```

rsk	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
_Icat_ph_1		.828742	.0318715	-4.88	0.000	.7685712 .8936236
_Icat_ph_2		.4680058	.0724593	-4.90	0.000	.3455122 .6339268
_Icat_ph_3		.7069112	.0219931	-11.15	0.000	.6650935 .7513581
_Icat_ph_4		.8052723	.0180458	-9.66	0.000	.7706687 .8414296
_Icat_ph_5		.7322978	.0184235	-12.38	0.000	.6970642 .7693123
_Icat_ph_6		.6928279	.1289558	-1.97	0.049	.4810515 .997836
_Icat_ph_7		.8356251	.0385655	-3.89	0.000	.763356 .9147362

_Icat_ph_8		.6992945	.027265	-9.17	0.000	.6478468	.7548278
_Icat_ph_9		.9235636	.0549489	-1.34	0.181	.8219079	1.037792
_Icat_ph_10		1.021953	.0781195	0.28	0.776	.8797589	1.187129
_Icat_ph_11		.7271792	.0662155	-3.50	0.000	.608321	.8692609
_Icat_ph_12		.725725	.0597743	-3.89	0.000	.6175369	.852867
_Icat_ph_13		1.019215	.0811177	0.24	0.811	.8720071	1.191274
_Icat_ph_14		.4407438	.0920719	-3.92	0.000	.2926639	.663748
_Icat_ph_15		.5995424	.0206753	-14.84	0.000	.5603587	.641466
_Icat_ph_16		.6751067	.0227746	-11.65	0.000	.631913	.7212528
_Icat_ph_17		.7078855	.1365293	-1.79	0.073	.4850567	1.033079
_Icat_ph_18		.3793581	.0694286	-5.30	0.000	.2650123	.5430412
_Icat_ph_19		.9060234	.1112855	-0.80	0.422	.7121764	1.152633
_Icat_ph_20		.7651475	.0445318	-4.60	0.000	.6826608	.8576012
_Icat_ph_21		1.126372	.0305342	4.39	0.000	1.068088	1.187836
_Icat_ph_22		1.096935	.0263832	3.85	0.000	1.046425	1.149884
_Icat_ph_23		1.06494	.0237746	2.82	0.005	1.019347	1.112572
_Icat_ph_24		1.01063	.0289136	0.37	0.712	.9555203	1.068919
_Icat_ph_25		.9685101	.0301724	-1.03	0.304	.9111425	1.02949
_Icat_ph_26		.9016073	.123227	-0.76	0.449	.6897308	1.178569
_Icat_ph_27		.8006338	.0414998	-4.29	0.000	.723291	.8862471
_Icat_ph_28		.750833	.0422974	-5.09	0.000	.6723444	.8384843
_Icat_ph_29		.768534	.0296083	-6.83	0.000	.7126397	.8288123
_Icat_ph_30		1.216114	.304421	0.78	0.434	.744559	1.986321
length		(exposure)					

Πίνακας 3.19: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση

atm

```
. xi:poisson atm i.cat_ph,exposure(length)irr
i.cat_ph          _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:  log likelihood = -505296.64
Iteration 1:  log likelihood = -505294.96
Iteration 2:  log likelihood = -505294.96

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   11941.93
Prob > chi2    =     0.0000
Log likelihood = -505294.96                       Pseudo R2      =     0.0117

-----+-----
atm |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
_Icat_ph_1 |   1.142719   .0151167    10.08  0.000   1.113471   1.172734
_Icat_ph_2 |   .3508999   .0248134   -14.81  0.000   .3054863   .4030648
_Icat_ph_3 |   .6019887   .0079331   -38.51  0.000   .5866391   .6177399
```

_Icat_ph_4		.8877323	.0076221	-13.87	0.000	.8729183	.9027977
_Icat_ph_5		.5319398	.0059943	-56.02	0.000	.52032	.5438192
_Icat_ph_6		.4304378	.0401969	-9.03	0.000	.358443	.5168931
_Icat_ph_7		.7434963	.0143385	-15.37	0.000	.715918	.7721371
_Icat_ph_8		.4898864	.0088879	-39.33	0.000	.4727726	.5076197
_Icat_ph_9		.7674675	.01975	-10.28	0.000	.7297182	.8071697
_Icat_ph_10		1.149874	.0328558	4.89	0.000	1.087248	1.216108
_Icat_ph_11		.5409218	.0225482	-14.74	0.000	.4984852	.586971
_Icat_ph_12		.6633708	.0225972	-12.05	0.000	.6205272	.7091724
_Icat_ph_13		.6820907	.0261552	-9.98	0.000	.6327065	.7353294
_Icat_ph_14		.2671992	.0283549	-12.44	0.000	.2170234	.3289756
_Icat_ph_15		.456715	.0070261	-50.94	0.000	.4431496	.4706957
_Icat_ph_16		.7312922	.0094374	-24.25	0.000	.7130273	.7500251
_Icat_ph_17		.9365249	.0622015	-0.99	0.323	.8222139	1.066728
_Icat_ph_18		.4972641	.0314865	-11.03	0.000	.4392274	.5629694
_Icat_ph_19		.8347301	.0422622	-3.57	0.000	.755875	.9218116
_Icat_ph_20		.6129118	.0157007	-19.11	0.000	.5828987	.6444703
_Icat_ph_21		.9508452	.0109126	-4.39	0.000	.9296957	.9724758
_Icat_ph_22		.8164266	.0086818	-19.07	0.000	.7995867	.8336211
_Icat_ph_23		.9881342	.0089581	-1.32	0.188	.9707316	1.005849
_Icat_ph_24		.9502196	.0110315	-4.40	0.000	.9288424	.9720888
_Icat_ph_25		.7288497	.0101447	-22.72	0.000	.7092352	.7490066
_Icat_ph_26		.5990242	.0396991	-7.73	0.000	.5260569	.6821126
_Icat_ph_27		.4719655	.0124552	-28.45	0.000	.4481743	.4970197
_Icat_ph_28		.5041914	.0136042	-25.38	0.000	.4782205	.5315728
_Icat_ph_29		.5144333	.0094142	-36.32	0.000	.4963088	.5332197
_Icat_ph_30		.4048726	.0694649	-5.27	0.000	.2892516	.5667102
length		(exposure)					

Πίνακας 3.20: Αποτελέσματα παλινδρόμησης Poisson για το φύλο και την κινάση cdk5

```

. xi:poisson cdk5 i.cat_ph,exposure(length) irr
i.cat_ph      _Icat_ph_0-30      (naturally coded; _Icat_ph_0 omitted)

Iteration 0:   log likelihood = -790133.05
Iteration 1:   log likelihood = -790112.29
Iteration 2:   log likelihood = -790112.29

Poisson regression                               Number of obs   =   1020130
LR chi2(30)    =   40893.61
Prob > chi2    =   0.0000
Log likelihood = -790112.29                       Pseudo R2      =   0.0252

-----+-----
cdk5 |          IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----

```

_Icat_ph_1		.5534001	.0066209	-49.45	0.000	.5405744	.5665302
_Icat_ph_2		.5365142	.0201002	-16.62	0.000	.4985301	.5773923
_Icat_ph_3		.5117393	.0047328	-72.44	0.000	.5025468	.5210999
_Icat_ph_4		.4328685	.0031612	-114.66	0.000	.4267168	.4391089
_Icat_ph_5		.4878739	.0037246	-94.01	0.000	.4806281	.495229
_Icat_ph_6		.2739544	.0209692	-16.92	0.000	.2357898	.3182962
_Icat_ph_7		.3394142	.0062199	-58.96	0.000	.3274398	.3518266
_Icat_ph_8		.5599215	.0062506	-51.95	0.000	.5478036	.5723074
_Icat_ph_9		.3274494	.0083471	-43.80	0.000	.3114913	.344225
_Icat_ph_10		.519432	.014324	-23.75	0.000	.4921026	.5482791
_Icat_ph_11		.4206361	.0129876	-28.05	0.000	.3959357	.4468773
_Icat_ph_12		.272337	.0094114	-37.64	0.000	.2545017	.2914221
_Icat_ph_13		.3446892	.0121124	-30.31	0.000	.3217484	.3692656
_Icat_ph_14		.5037316	.0254962	-13.55	0.000	.4561586	.556266
_Icat_ph_15		.6262481	.0054885	-53.40	0.000	.6155828	.6370983
_Icat_ph_16		.3126579	.0038576	-94.23	0.000	.3051878	.3203109
_Icat_ph_17		.4641496	.0286072	-12.45	0.000	.4113346	.5237459
_Icat_ph_18		.7190826	.0248065	-9.56	0.000	.67207	.7693839
_Icat_ph_19		.2448404	.0149223	-23.09	0.000	.2172726	.2759061
_Icat_ph_20		.6498132	.010589	-26.45	0.000	.6293872	.6709022
_Icat_ph_21		.8769881	.0068052	-16.92	0.000	.8637511	.890428
_Icat_ph_22		.761756	.0054452	-38.07	0.000	.751158	.7725035
_Icat_ph_23		.7800377	.0050356	-38.48	0.000	.7702303	.78997
_Icat_ph_24		.7580435	.0063227	-33.21	0.000	.745752	.7705375
_Icat_ph_25		.3876325	.0046899	-78.33	0.000	.3785485	.3969345
_Icat_ph_26		.363883	.0202201	-18.19	0.000	.3263342	.4057523
_Icat_ph_27		.5526214	.0088448	-37.06	0.000	.535555	.5702317
_Icat_ph_28		.5038018	.0088968	-38.82	0.000	.4866627	.5215444
_Icat_ph_29		.3692111	.0051619	-71.27	0.000	.3592313	.3794681
_Icat_ph_30		.2548468	.0360505	-9.66	0.000	.1931383	.3362713
length		(exposure)					