



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΜΕΤΑΛΛΑΞΕΙΣ ΤΗΣ ΠΡΩΤΕΪΝΗΣ Notch3 ΣΤΗΝ ΑΣΘΕΝΕΙΑ CADASIL: 3D ΜΟΡΙΑΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΕΣ ΑΝΑΛΥΣΕΙΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αικατερίνη Π. Ιωαννίδου

Επιβλέπων: Γιώργος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

ΜΕΤΑΛΛΑΞΕΙΣ ΤΗΣ ΠΡΩΤΕΪΝΗΣ Notch3 ΣΤΗΝ ΑΣΘΕΝΕΙΑ CADASIL: 3D ΜΟΡΙΑΚΗ ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΕΣ ΑΝΑΛΥΣΕΙΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αικατερίνη Π. Ιωαννίδου

Επιβλέπων: Γιώργος Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 7^η Ιουλίου 2014.

.....
Γ. Ματσόπουλος
Επ. Καθηγητής Ε.Μ.Π.

.....
Δ. Κουτσούρης
Καθηγητής Ε.Μ.Π.

.....
Ν. Ουζούνγλου
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2014

.....

Αικατερίνη Π. Ιωαννίδου

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Αικατερίνη Π. Ιωαννίδου, 2014

Με επιφύλαξη παντός δικαιώματος, All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου

Περίληψη

Η νόσος CADASIL ανήκει στην ομάδα των σπάνιων ασθενειών. Είναι καλά τεκμηριωμένο ότι η πρωτεΐνη Notch3 είναι κυρίως υπεύθυνη για την ανάπτυξη της CADASIL. Εδώ, θα προσπαθήσουμε να ρίξουμε φως στον πραγματικό μοριακό μηχανισμό στον οποίο υποκείται η ασθένεια CADASIL μέσω των γνώσεων που έχουμε από την προκαταρκτική *in silico* και άλλες πρωτογενείς μελέτες για την πρωτεΐνη Notch3. Σήμερα, είναι γνωστή μια σειρά σημειακών μεταλλάξεων στην πρωτεΐνη Notch3 που προωθούν την CADASIL. Σε αυτή την κατεύθυνση θα διερευνήσουμε τη φύση, την έκταση, τη φυσικοχημική και δομική σημασία των μεταλλαγμένων ειδών, σε μια προσπάθεια να προσδιοριστεί ο υποκείμενος μηχανισμός και ρόλος της Notch3 και οι επιπτώσεις του στην μεταγωγή του σήματος μεταξύ των κυττάρων και στην ανάπτυξη καρκινικών κυττάρων.

Συνολικά, η *in silico* μελέτη μας αποκάλυψε έναν μάλλον πολύπλοκο μοριακό μηχανισμό της Notch3 σε δομικό επίπεδο, καθώς ανάλογα με τη φύση και τη θέση της κάθε μετάλλαξης, παρατηρήθηκε σημαντική απώλεια της δομής των β-φύλλων σε όλα τα *in silico* πρότυπα των μεταλλαγμένων/άγριου τύπου βιολογικών συστημάτων.

Η παρούσα διπλωματική εργασία αρχίζει με μια εισαγωγή γενικά για τις πρωτεΐνες, την κυτταρική σηματοδότηση και τον ρόλο της βιοπληροφορικής στις *in silico* μελέτες. Ακολουθεί μια θεωρητική ανασκόπηση της νόσου CADASIL, της πρωτεΐνης Notch3 που ενοχοποιείται για την εμφάνιση της νόσου και του μοριακού μηχανισμού της. Στη συνέχεια, παρουσιάζονται οι μέθοδοι που χρησιμοποιήθηκαν για τη μελέτη, καθώς και τα αποτελέσματα της κάθε μεθόδου συνοδευόμενα από σχόλια. Τέλος, αναφέρονται συνοπτικά τα συμπεράσματα που προέκυψαν στην παρούσα μελέτη και πιθανές μελλοντικές επεκτάσεις της.

Λέξεις κλειδιά

CADASIL, Notch3, 3D μοριακή μοντελοποίηση, μοριακή δυναμική, ανάλυση μεταλλάξεων, κυτταρική σηματοδότηση

Abstract

CADASIL disease belongs to the group of rare diseases. It is well established that Notch3 protein is primarily responsible for the development of the CADASIL. Herein, we attempt to shed light to the actual molecular mechanism underlying CADASIL via insights that we have from preliminary *in silico* and proteomics studies on the Notch3 protein. At the moment, we are aware of a series of Notch3 point mutations that promote the CADASIL. In this direction we investigate the nature, extent, physicochemical and structural significance of the mutant species in an effort to identify the underlying mechanism of Notch3 role and implications in cell signal transduction and cancer development.

Overall our *in silico* study has revealed a rather complex molecular mechanism of Notch3 on the structural level, as depending of the nature and position of each mutation, a consensus significant loss of beta-sheet structure is observed throughout all *in silico* modeled mutant/wild type biological systems.

The current thesis begins with a general introduction to proteins, cell signaling and bioinformatics' role in silico studies. It is followed by a theoretical review of CADASIL disease, Notch3 protein which is implicated in the disease and the molecular mechanism. Then, the methods used for the study are presented, as well as the results of each method accompanied by comments. Finally, there is a summarized reference to the conclusions drawn in this study and possible future extensions.

Key words

CADASIL, Notch3, 3D molecular modeling, molecular dynamics, mutation analysis, cell signaling

Ιατρικά Δεδομένα

Στα πλαίσια αυτής της διπλωματικής εργασίας, χρησιμοποιήθηκαν πραγματικά ιατρικά δεδομένα όσον αφορά στις μεταλλάξεις της πρωτεΐνης Notch3 που εξετάστηκαν. Σε αυτό το σημείο, θα ήθελα να εκφράσω τις ιδιαίτερες ευχαριστίες μου στον καθηγητή του Πανεπιστημίου του Ελσίνκι κ. Marc Baumann, για τη συμβολή του στην εργασία μου, με την αποστολή όχι μόνο των δεδομένων αυτών, αλλά και χρήσιμων πληροφοριών από την ιατρική του έρευνα στον συγκεκριμένο τομέα. Αξίζει να σημειωθεί εδώ, ότι πολλές από τις εν λόγω μεταλλάξεις δεν είχαν αναφερθεί προηγουμένως. Τα αποτελέσματα της εργασίας προέκυψαν από την μετα-ανάλυση επιλεγμένων ιατρικών δεδομένων. Η πλήρης λίστα των μεταλλάξεων που έλαβα παρατίθεται στο παράρτημα στο τέλος της εργασίας.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2013-2014 υπό την επίβλεψη του κ. Γεώργιου Ματσόπουλου, επ. καθηγητή του Ε.Μ.Π. της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, στον οποίο οφείλω ιδιαίτερες ευχαριστίες για την ανάθεσή της, δίνοντάς μου την ευκαιρία να ασχοληθώ ουσιαστικά με έναν τομέα με τόσες προοπτικές ανάπτυξης όπως της Βιοπληροφορικής και Ιατρικής Πληροφορικής.

Επίσης, θα ήθελα να ευχαριστήσω την Δρ. Σοφία Κοσσίδα, ερευνήτρια και επικεφαλής του τμήματος Βιοπληροφορικής και Ιατρικής Πληροφορικής του Ιδρύματος Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών (ΙΙΒΕΑΑ), καθώς και τον Δρ. Δημήτριο Βλαχάκη, μεταδιδακτορικό ερευνητή στο ίδιο τμήμα, για τη συνεχή επίβλεψή τους και την αμέριστη βοήθεια και καθοδήγησή που μου προσέφεραν κατά τη συνεργασία μου με το Ίδρυμα.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου, για τη βοήθεια και την στήριξή τους καθ' όλη τη διάρκεια των σπουδών μου.

Περιεχόμενα

1. Εισαγωγή	
1.1. Σκοπός και Σημασία της Διπλωματικής Εργασίας.....	-1-
1.2. Διάρθρωση Διπλωματικής Εργασίας.....	-2-
2. Ανασκόπηση των Βασικών Βιολογικών και Ιατρικών Εννοιών	
2.1. Πρωτεΐνες.....	-5-
2.1.1. Εισαγωγικά – Γενικά.....	-5-
2.1.2. Κυτταρική Σηματοδότηση (cell signaling) και Δέσμευση Συνδετών (ligand binding).....	-7-
2.1.3. Πρωτεΐνες και Βιοπληροφορική.....	-8-
2.2. Η ασθένεια CADASIL.....	-9-
2.2.1. Εισαγωγικά.....	-9-
2.2.2. Επιδημιολογία της CADASIL.....	-10-
2.2.3. Η Πρωτεΐνη Notch3 και οι Δομικές της Ιδιότητες.....	-11-
2.2.4. Μοριακός Μηχανισμός / Συσσώρευση Αποθέσεων GOM.....	-14-
2.2.5. Διάγνωση της ασθένειας.....	-16-
3. Μέθοδοι επεξεργασίας των ιατρικών δεδομένων και μελέτης της αλληλουχίας και της δομής της Notch3	
3.1. Βιοπληροφορική.....	-19-
3.2. Ιατρικά Δεδομένα.....	-20-
3.3. Στοίχιση/Ευθυγράμμιση Αλληλουχιών.....	-20-
3.3.1. Εισαγωγή.....	-20-
3.3.2. Αλγόριθμοι ClustalW και MUSCLE.....	-21-
3.4. Ομόλογη Μοντελοποίηση και Ελαχιστοποίηση Ενέργειας.....	-21-
3.5. Προσομοιώσεις Μοριακής Δυναμικής.....	-22-

4.	Αποτελέσματα επεξεργασίας των ιατρικών δεδομένων	
4.1.	Μοτίβο επαναλήψεων αμινοξέων (Amino Acid Repeats – AAR pattern) - Στοίχιση αλληλουχιών.....	-23-
4.1.1.	Περιγραφή του μοτίβου.....	-25-
4.2.	Χαρτογράφηση των μεταλλάξεων στην πρωτοταγή και δευτεροταγή δομή της Notch3.....	-27-
4.3.	3D μοριακή μοντελοποίηση της δομής της πρωτεΐνης Notch3.....	-29-
4.4.	Δομική ευθυγράμμιση και υπέρθεση της πρωτεΐνης αγρίου τύπου και της αντίστοιχης μεταλλαγμένης.....	-31-
4.4.1.	Εισαγωγή.....	-31-
4.4.2.	Παρουσίαση των αποτελεσμάτων σε επιλεγμένα μοντέλα....	-31-
4.5.	Τα Πεδία Δυνάμεων (Force Fields).....	-33-
4.5.1.	Γενικά.....	-33-
4.5.2	Τα πεδία δυναμικού AMBER, CHARMM και OPLS.....	-35-
4.6.	Μοριακή δυναμική.....	-37-
4.6.1.	Εισαγωγή.....	-37-
4.6.2.	Θεωρία και Μαθηματική προσέγγιση.....	-37-
4.6.2.1.	Γενικά.....	-37-
4.6.2.2.	Οι εξισώσεις NHA.....	-39-
4.6.2.3.	Οι εξισώσεις BER.....	-39-
4.6.2.4.	Ολονομικοί Περιορισμοί.....	-40-
4.6.2.5.	Διορθώσεις στις ορμές.....	-41-
4.6.3.	Περιορισμοί Μοριακής Δυναμικής.....	-41-
4.6.4.	Σκοπός.....	-42-
4.6.5.	Αποτελέσματα Μοριακής Δυναμικής.....	-42-

5. Συμπεράσματα – Μελλοντικές Επεκτάσεις	
5.1. Συμπεράσματα.....	-45-
5.2. Μελλοντικές Επεκτάσεις.....	-46-
Βιβλιογραφία.....	-47-
Παράρτημα.....	-51-

Κεφάλαιο 1

Εισαγωγή

1.1. Σκοπός και Σημασία της Διπλωματικής Εργασίας

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η διερεύνηση του ρόλου των πρόσφατα ανακαλυφθέντων θέσεων μεταλλάξεων της πρωτεΐνης Notch3 στην εκδήλωση του συνδρόμου CADASIL.

Η σημασία της εν λόγω εργασίας έγκειται στο ιδιαίτερο ενδιαφέρον που παρουσιάζει η επεξεργασία της ακατέργαστης ιατρικής πληροφορίας, με σκοπό να αποκτήσουν τα δεδομένα αυτά βιολογική σημασία. Ο απώτερος στόχος αυτής της ενέργειας είναι η οργάνωση της μετέπειτα έρευνας και η χάραξη μεθοδικής στρατηγικής για τον ορθολογικό σχεδιασμό φαρμάκων. Εδώ, ως ακατέργαστα ιατρικά δεδομένα

αναφερόμαστε στο σύνολο των μεταλλάξεων της πρωτεΐνης Notch3 που μας δόθηκαν. Πρόκειται για μια πρωτεΐνη που αποτελείται από μεγάλο αριθμό αμινοξέων. Επομένως, το να καταφέρουμε να περιορίσουμε την περιοχή ενδιαφέροντος του φαρμακολογικού σχεδιασμού σε ένα μόνο μέρος της είναι σπουδαίο εγχείρημα, καθώς θα περιορίσει τον χρόνο και το κόστος της διαδικασίας. Επιπλέον, μέσω αυτής της εργασίας προτείνεται μία γενικότερη μεθοδολογία για τη μελέτη μεταλλάξεων σε πρωτεΐνες.

Το δεύτερο σκέλος της σημασίας της διπλωματικής εργασίας αφορά τον τύπο της ασθένειας που μελετάται. Συγκεκριμένα, η νόσος CADASIL αποτελεί μια σπάνια ασθένεια. Ως γνωστόν, σε αυτού του είδους τις ασθένειες δεν επενδύεται επαρκής χρηματοδότηση για την εξέτασή τους. Ωστόσο, η νόσος αυτή είναι κληρονομική, δηλαδή παρουσιάζει προοπτικές για αύξηση της επιδημιολογίας της όσο δεν αντιμετωπίζεται. Μέσω αυτής της εργασίας, επιθυμούμε να θέσουμε ένα λιθαράκι σε αυτή την προσπάθεια για την επίτευξη του στόχου.

2.2. Διάρθρωση Διπλωματικής Εργασίας

Στη συνέχεια της διπλωματικής εργασίας, γίνεται αναφορά σε βασικές θεωρητικές έννοιες της βιολογίας και της ιατρικής που αφορούν το θέμα μας, αναλύονται οι μέθοδοι επεξεργασίας των ιατρικών δεδομένων που χρησιμοποιήθηκαν και παρουσιάζονται τα αποτελέσματα αυτής. Η παρούσα διπλωματική διαρθρώνεται στα κάτωθι κεφάλαια.

Στο κεφάλαιο 2, στο πρώτο μέρος, γίνεται μια θεωρητική γενική εισαγωγή στις πρωτεΐνες και στην κυτταρική σηματοδότηση (cell signaling), στην οποία οι πρωτεΐνες παίζουν κυρίαρχο ρόλο και η οποία αποτελεί κλειδί για την κατανόηση του ρόλου των μεταλλάξεων στην εμφάνιση της ασθένειας, αλλά και στη σχέση των πρωτεϊνών με τη βιοπληροφορική που είναι και ο τομέας άμεσου ενδιαφέροντος της εργασίας. Στο δεύτερο μέρος, παρουσιάζεται η ασθένεια CADASIL, η επιδημιολογία της, η διάγνωσή της, καθώς και ο ρόλος της πρωτεΐνης Notch3 στην εμφάνισή της και ο μοριακός μηχανισμός της.

Στο κεφάλαιο 3, αναλύονται οι μέθοδοι επεξεργασίας των ιατρικών δεδομένων και μελετάται η αλληλουχία αμινοξέων και η δομή της Notch3. Συγκεκριμένα, παρουσιάζονται οι αλγόριθμοι για τη στοίχιση/ ευθυγράμμιση της αλληλουχίας και αναλύονται οι μέθοδοι ομόλογης μοντελοποίησης και ελαχιστοποίησης ενέργειας, καθώς και προσομοίωσης της Μοριακής Δυναμικής.

Στο κεφάλαιο 4, παρουσιάζονται βήμα προς βήμα τα αποτελέσματα της επεξεργασίας και περιγράφεται αναλυτικά η διαδικασία που ακολουθήθηκε, από τη στοίχιση της αλληλουχίας, στην κατασκευή ορισμένου μοτίβου της πρωτοταγούς δομής μέχρι την τρισδιάστατη μοριακή μοντελοποίηση της δευτεροταγούς δομής. Ακολουθεί δομική ευθυγράμμιση και υπέρθεση της πρωτεΐνης αγρίου τύπου και της αντίστοιχης

μεταλλαγμένης, σε επιλεγμένες μεταλλάξεις, για άμεση σύγκριση των δύο μοντέλων. Το κεφάλαιο ολοκληρώνεται με αναλυτική αναφορά στην Μοριακή Δυναμική και παράθεση ενδεικτικών αποτελεσμάτων αυτής.

Στο κεφάλαιο 5, αναφέρονται συνοπτικά τα συμπεράσματα που προέκυψαν από την ανάλυση των αποτελεσμάτων του προηγούμενου κεφαλαίου, καθώς και οι μελλοντικές επεκτάσεις της έρευνας που έγινε για την παρούσα διπλωματική εργασία.

Κεφάλαιο 2

Ανασκόπηση των Βασικών Βιολογικών και Ιατρικών Εννοιών

2.1. Πρωτεΐνες

2.1.1. Εισαγωγικά - Γενικά

Το μεγαλύτερο μέρος αυτής της μελέτης βασίζεται στην επεξεργασία των πρωτεϊνών. Προκειμένου να επιτευχθεί βαθύτερη κατανόηση του θέματος, κρίνεται απαραίτητα μια συνοπτική εισαγωγή στις πρωτεΐνες και τις ιδιότητές τους. Το κεντρικό δόγμα της μοριακής βιολογίας αφορά τη λεπτομερή μεταφορά της γενετικής πληροφορίας.

Συγκεκριμένα, ο γενετικός κώδικας ρέει από το DNA στο RNA και αντίστροφα, μέσω των καθορισμένων διαδικασιών της αντιγραφής, μεταγραφής και αντίστροφης μεταγραφής (σε ορισμένους RNA ιούς) και εν συνεχεία, σε πρωτεΐνες με τη μορφή αλληλουχίας αμινοξέων μέσω της διαδικασίας της μετάφρασης.

Αυτά τα αμινοξέα συμπυκνώνονται και δημιουργούν αλυσίδες, συνδεδεμένα με δεσμούς πεπτιδίων. Οι αλυσίδες αυτές, μόνες τους ή συνδεδεμένες με άλλες, αποτελούν μεγάλα βιολογικά μόρια, ή αλλιώς μακρομόρια, τις πρωτεΐνες. Οι πρωτεΐνες είναι υπεύθυνες για έναν μεγάλο αριθμό λειτουργιών, όπως μεταβολικές αντιδράσεις και εξυπηρετούν ως μεταγραφικοί παράγοντες DNA, αλλά και στη μεταφορά των μορίων. Ως εκ τούτου, διαδραματίζουν ζωτικό ρόλο για κάθε ζωντανό οργανισμό.

Η διαφορά μεταξύ δύο πρωτεϊνών έγκειται κυρίως στις διαφορετικές αλληλουχίες των αμινοξέων τους, οι οποίες υπαγορεύονται από τις νουκλεοτιδικές αλληλουχίες των γονιδίων τους. Συνήθως, διαφορετική αλληλουχία αμινοξέων οδηγεί σε διαφορετική αναδίπλωση της πρωτεΐνης στην τρισδιάστατη δομή που καθορίζει τη δράση της. Στις περισσότερες περιπτώσεις, η 3-διάστατη δομή της κάθε πρωτεΐνης είναι μοναδική. Αυτό το μοναδικό σχήμα στο οποίο μία πρωτεΐνη διπλώνει φυσικά ονομάζεται φυσική διαμόρφωση. Στη βιοχημεία υπάρχουν τέσσερα διακριτά επίπεδα πρωτεϊνικής δομής. Οι τέσσερις τύποι αλληλεπιδράσεων που παρατηρούνται εντός μιας πρωτεΐνης είναι: η πρωτοταγής, δευτεροταγής, τριτοταγής και τεταρτοταγής δομή. Η πρωτοταγής δομή (primary structure) αναφέρεται στην γραμμική ακολουθία των αμινοξέων στην πολυπεπτιδική αλυσίδα. Η δευτεροταγής δομή (secondary structure) αναφέρεται στην τρισδιάστατη μορφή της διπλωμένης πρωτεΐνης και πιο συγκεκριμένα, στις περιοδικά επαναλαμβανόμενες τοπικές δομές οι οποίες σταθεροποιούνται με δεσμούς υδρογόνου. Οι πιο κοινές επαναλαμβανόμενες δομές είναι η α -έλικα, η β -πτυχωτή επιφάνεια (ή β φύλλο) και η β -στροφή. Στη συνέχεια, οι δομές αυτές συσκευάζονται μαζί κατά καθορισμένο τρόπο, προκειμένου να συντάξουν τη συνολική διαμόρφωση μίας μονήρους πολυπεπτιδικής αλυσίδας, ή αλλιώς την τριτοταγή δομή (tertiary structure) ενός πολυπεπτιδίου. Τέλος, ο τρόπος με τον οποίο ενώνονται μεταξύ τους αυτές οι πολυπεπτιδικές αλυσίδες καλείται και τεταρτοταγής δομή (quarternary structure). Σε αυτή τη μελέτη, θα ασχοληθούμε με τις δύο πρώτες δομές. [1]

Οι πρωτεΐνες παίζουν σημαντικό ρόλο στο εσωτερικό του κυττάρου, καθώς ασχολούνται με τις πληροφορίες στα γονίδια. Πολλές πρωτεΐνες είναι υπεύθυνες για τη διαδικασία της κυτταρικής σηματοδότησης και την μεταγωγή σήματος. Μερικές από αυτές είναι εξωκυτταρικές και επιτρέπουν την επικοινωνία μεταξύ των κυττάρων με την μετάδοση σήματος, ενώ άλλες είναι μεμβρανικές, που λειτουργούν ως υποδοχείς διασυνδέσεως ενός μορίου σηματοδότησης και προκαλούν μια ορισμένη βιοχημική αντίδραση στο κύτταρο. Οι τελευταίες συνήθως έχουν μία θέση συνδέσεως εκτεθειμένη στην επιφάνεια του κυττάρου και μια περιοχή τελεστή εντός του κυττάρου.

2.1.2. Κυτταρική Σηματοδότηση (cell signaling) και Δέσμευση Συνδετών (ligand binding)

Τα αντισώματα, γνωστά και ως ανοσοσφαιρίνες, είναι μεγάλου σχήματος Y πρωτεϊνικά μόρια. Αποτελούν μέρος του προσαρμοστικού ανοσοποιητικού συστήματος και η κύρια λειτουργία τους είναι να δεσμεύουν αντιγόνα, ή ξένες ουσίες στο σώμα, και να τις θέτουν ως στόχους (target) για καταστροφή. Συγκεκριμένα, το αντίσωμα αναγνωρίζει ένα μοναδικό κομμάτι του εισβολέα που ονομάζεται αντιγόνο. Κάθε κορυφή της δομής Y περιέχει μια δομή ανάλογη με κλειδαριά που είναι ειδική για ένα συγκεκριμένο σήμα σε ένα αντιγόνο επιτρέποντας στις δύο δομές να συνδέονται με ακρίβεια. Με την σύνδεση ένα αντίσωμα μπορεί να καταδείξει ένα μικρόβιο ή ένα μολυσμένο κύτταρο για επίθεση από άλλα κομμάτια του ανοσοποιητικού συστήματος, ή να εξουδετερώσει το στόχο του απευθείας.

Τα αντισώματα μπορούν να υπάρξουν σε δύο μορφές: μία διαλυτή μορφή που εκκρίνεται από το κύτταρο στο εξωκυτταρικό (extracellular) περιβάλλον και μια μεμβρανο-συνδετή μορφή που προσκολλάται στην επιφάνεια ενός B λεμφοκυττάρου και ονομάζεται υποδοχέας κυττάρου B (B cell receptor BCR). Λαμβάνοντας υπόψη ότι τα ένζυμα έχουν περιορισμένη συγγένεια δέσμευσης για τα υποστρώματα τους λόγω της ανάγκης για διεξαγωγή της αντίδρασης τους, τα αντισώματα δεν έχουν τέτοιους περιορισμούς. Η συγγένεια δέσμευσης ενός αντισώματος για τον στόχο του είναι πολύ υψηλή. [2]

Πολλές πρωτεΐνες μεταφοράς συνδετών δεσμεύουν συγκεκριμένα μικρά βιομόρια και τα μεταφέρουν σε άλλα σημεία του σώματος ενός πολυκύτταρου οργανισμού. Αυτές οι πρωτεΐνες πρέπει να έχουν υψηλή συγγένεια σύνδεσης, όταν ο συνδέτης τους παρουσιάζει υψηλές συγκεντρώσεις, αλλά πρέπει επίσης να απελευθερώνουν το συνδέτη όταν παρουσιάζει χαμηλές συγκεντρώσεις στους ιστούς-στόχους. Το πιο τυπικό παράδειγμα μιας δεσμευτικής πρωτεΐνης-συνδέτη (ligand-binding) είναι η αιμοσφαιρίνη, η οποία μεταφέρει οξυγόνο από τους πνεύμονες σε άλλα όργανα και ιστούς σε όλα τα σπονδυλόζωα και έχει στενά ομόλογα σε κάθε βιολογικό βασίλειο. [3] Οι λεκτίνες είναι δεσμευτικές πρωτεΐνες με εξαιρετική εξειδίκευση στους υδατάνθρακες. Τα μόριά τους έχουν τουλάχιστον δύο θέσεις δέσμευσης των υδατανθράκων. Οι λεκτίνες τυπικά εμπλέκονται βιολογικά φαινόμενα αναγνώρισης που περιλαμβάνουν κύτταρα και πρωτεΐνες. [4] Οι υποδοχείς και οι ορμόνες αποτελούν εξειδικευμένες πρωτεΐνες δέσμευσης. Οι διαμεμβρανικές πρωτεΐνες μπορούν επίσης να χρησιμεύσουν και ως πρωτεΐνες μεταφοράς συνδέτη (ligand transport) μεταβάλλοντας τη διαπερατότητα της κυτταρικής μεμβράνης σε μικρά μόρια και ιόντα. Η μεμβράνη έχει υδρόφοβο πυρήνα μέσω του οποίου πολωμένα ή φορτισμένα μόρια δεν μπορούν να διαχυθούν. Οι μεμβρανικές πρωτεΐνες περιέχουν εσωτερικά κανάλια που επιτρέπουν σε τέτοιου είδους μόρια να εισέλθουν και να εξέλθουν από το κύτταρο. Πολλές πρωτεΐνες ιοντικών καναλιών ειδικεύονται στην επιλογή ενός συγκεκριμένου ιόντος. Για παράδειγμα, τα κανάλια καλίου και νατρίου συχνά διακρίνονται για ένα μόνο από τα δύο ιόντα. [5]

2.1.3. Πρωτεΐνες και Βιοπληροφορική

Μια μεγάλη σειρά από υπολογιστικές μεθόδους έχει αναπτυχθεί για να αναλύσει τη δομή, τη λειτουργία και την εξέλιξη των πρωτεϊνών. Η ανάπτυξη τέτοιων εργαλείων έχει οδηγηθεί από τη μεγάλη ποσότητα γονιδιωματικών και πρωτεομικών δεδομένων που είναι διαθέσιμα για μια ποικιλία οργανισμών, συμπεριλαμβανομένων και του ανθρώπινου γονιδιώματος. Είναι απλά αδύνατο να μελετήσει κανείς όλες τις πρωτεΐνες πειραματικά. Ως εκ τούτου, μόνο λίγες από αυτές υποβάλλονται σε εργαστηριακά πειράματα, ενώ για τη μελέτη των υπόλοιπων παρόμοιων πρωτεϊνών χρησιμοποιούνται υπολογιστικά εργαλεία. Τέτοιες ομόλογες πρωτεΐνες μπορούν να ταυτοποιηθούν αποτελεσματικά σε μακρινής συγγένειας οργανισμούς με ευθυγράμμιση αλληλουχίας. Γονιδίωμα και γονιδιακές αλληλουχίες μπορούν να αναζητηθούν από μια ποικιλία εργαλείων για ορισμένες ιδιότητες. Πιο συγκεκριμένα, εργαλεία sequence profiling μπορούν να βρουν θέσεις ενζύμων περιορισμού, ανοικτά πλαίσια ανάγνωσης στις νουκλεοτιδικές αλληλουχίες, και να προβλέψουν δευτεροταγείς δομές. Επιπλέον, μπορούν να κατασκευαστούν φυλογενετικά δέντρα, ενώ έχουν αναπτυχθεί και εξελικτικές υποθέσεις σχετικά με την καταγωγή των σύγχρονων οργανισμών και τα γονίδια που εκφράζουν, με τη βοήθεια ειδικού λογισμικού όπως ο αλγόριθμος ClustalW. Το πεδίο της βιοπληροφορικής είναι πλέον απαραίτητο για την ανάλυση των γονιδίων και των πρωτεϊνών.

2.2. Η ασθένεια CADASIL

2.2.1. Εισαγωγικά

Η λέξη CADASIL είναι ένα αρκτικόλεξο που προκύπτει από τις λέξεις: Cerebral Autosomal Dominant Arteriopathy Subcortical Infarcts Leukoencephalopathy: εγκεφαλική αρτηριοπάθεια αυτοσωμικού επικρατούντος χαρακτήρα με υποφλοιώδη έμφρακτα και λευκοεγκεφαλοπάθεια. Πιο συγκεκριμένα, η λέξη ‘Cerebral’ (εγκεφαλικός) δηλώνει συσχέτιση με τον εγκέφαλο. ‘Autosomal Dominant’ (αυτοσωμικός κυρίαρχος) είναι ένας τύπος κληρονομικότητας όπου ένα μόνο αντίγραφο του παθολογικού γονιδίου είναι απαραίτητο για την ανάπτυξη της ασθένειας. ‘Arteriopathy’ (αρτηριοπάθεια) είναι μια ασθένεια των μικρών αρτηριών (των αιμοφόρων αγγείων που μεταφέρουν το αίμα από την καρδιά). ‘Subcortical’ (υποφλοιώδης) σημαίνει ότι σχετίζεται με μια συγκεκριμένη περιοχή του εγκεφάλου βαθιά που συμμετέχει στην υψηλότερη λειτουργία (π.χ., εθελοντική κινήσεις, συλλογιστική, μνήμη). ‘Infarcts’ (έμφρακτα) είναι η απώλεια εγκεφαλικού ιστού που προκαλείται από έλλειψη οξυγόνου στον εγκέφαλο, η οποία εμφανίζεται όταν η ροή του αίματος στις μικρές αρτηρίες είναι μπλοκαρισμένη ή ανώμαλη. Τέλος, “Leukoencephalopathy” (λευκοεγκεφαλοπάθεια) είναι η καταστροφή της μυελίνης, μιας λιπαρής ουσίας που καλύπτει και προστατεύει τις ίνες των νεύρων στο κεντρικό νευρικό σύστημα. Η ασθένεια CADASIL είναι ο επικρατέστερος τύπος κληρονομικής εγκεφαλικής αγγειοπάθειας (4) ή εγκεφαλικού επεισοδίου. Ως κύρια αιτία της θεωρούνται οι μεταλλάξεις του γονιδίου NOTCH3 στο χρωμόσωμα 19. Η νόσος ανήκει σε μια οικογένεια διαταραχών που ονομάζονται Λευκοδυστροφίες. [6] Πάνω από 30 χρόνια έχουν περάσει από την πρώτη φορά που ο όρος αυτός χρησιμοποιήθηκε για να περιγράψει την κατάσταση μιας σουηδικής οικογένειας. Ωστόσο, το ακρωνύμιο CADASIL δεν είχε εμφανιστεί μέχρι τις αρχές του 1990. [7, 8] Οι περισσότεροι ασθενείς με CADASIL έχουν ένα μέλος της οικογένειας με τη συγκεκριμένη διαταραχή. [9] Ωστόσο, επειδή η γενετική εξέταση για CADASIL δεν ήταν διαθέσιμη πριν από το 2000, πολλές περιπτώσεις διαγνώστηκαν λανθασμένα ως νόσος του Alzheimer, σκλήρυνση κατά πλάκας ή άλλων νευροεκφυλιστικών ασθενειών. [10]

Αυτή η κληρονομική πάθηση πλήττει κυρίως τα μικρά αιμοφόρα αγγεία στη λευκή ουσία του εγκεφάλου, προκαλώντας πολλαπλά κενотоπιώδη και υποφλοιώδη έμφρακτα της λευκής ουσίας. Αναφέρεται, επίσης, δυσανάλογος φλοιώδης υπομεταβολισμός. [10, 11] Τα κύτταρα των μυών που περιβάλλουν αυτά τα αιμοφόρα αγγεία, γνωστά και ως αγγειακά λεία μυϊκά κύτταρα (VSMCs – vascular smooth muscle cells) είναι παθολογικά και σταδιακά πεθαίνουν.

Ως αποτέλεσμα, η βλάβη των αιμοφόρων αγγείων (αρτηριοπάθεια) στον εγκέφαλο μπορεί να προκαλέσει ημικρανίες, οι οποίες συχνά ακολουθούνται από οπτικές παραισθήσεις ή αύρες, στην ηλικία των 30 [12], ή σπανιότερα, από επαναλαμβανόμενες επιληπτικές κρίσεις (5-10% των ασθενών). [1, 9, 12, 13] Η πάχυνση των τοιχωμάτων των κατεστραμμένων αιμοφόρων αγγείων εμποδίζει τη ροή του αίματος προς τον εγκέφαλο [10] και μπορεί να προκαλέσει το θάνατο σε περιοχές του ιστού (έμφρακτα). Ένα έμφρακτο στον εγκέφαλο μπορεί να οδηγήσει σε εγκεφαλικό επεισόδιο. Τα άτομα με CADASIL υποφέρουν από επαναλαμβανόμενα

εγκεφαλικά επεισόδια του υποφλοιώδους μυοκαρδίου [7, 14] σε όλη τη διάρκεια της ζωής τους, τα οποία με την πάροδο του χρόνου βλάπτουν τον εγκέφαλο. Εγκεφαλικά επεισόδια που εμφανίζονται στην περιοχή υποφλοιώδους του εγκεφάλου, η οποία εμπλέκεται στη συλλογιστική και στη μνήμη, μπορεί να προκαλέσει προβλήματα όρασης και γνωστική εξασθένηση. Η τελευταία εξελίσσεται σε δυσκολία συγκέντρωσης και προσοχής, μέχρι σε απώλεια της νοητικής λειτουργίας (υποφλοιώδεις άνοια) και σε ψυχιατρικά προβλήματα (10-20% των περιπτώσεων), όπως διαταραχή της διάθεσης, απάθεια, οξεία κατάθλιψη και αλλαγές στη συμπεριφορά και την προσωπικότητα. [1, 13] Η ψύχωση, η ψευδοπρομηκική παράλυση, η ακράτεια ούρων και το εστιακό νευρολογικό έλλειμμα είναι επίσης πιθανά συμπτώματα. [11, 15]

Μπορεί, ακόμη, να εμφανιστούν προβλήματα με την ισορροπία και το περπάτημα καθώς ο ασθενής πλησιάζει την ηλικία των εξήντα. Απώλεια της ανεξαρτησίας με κινητικά και γνωστικά μειονεκτήματα είναι επίσης συχνή μετά την ηλικία των 60. [12]

Η λευκοεγκεφαλοπάθεια διαγιγνώσκεται μέσω μαγνητικής τομογραφίας (MRI) ως διάχυτη λευκή ουσία. [9, 16]

Η έναρξη της εγκεφαλικής αγγειακής νόσου εμφανίζεται συνήθως στον ενήλικο στις ηλικίες μεταξύ 30 και 60, ενώ επαναλαμβανόμενα παροδικά ισχαιμικά επεισόδια (TIAs – transient ischemic) ή εγκεφαλικά επεισόδια, σε πολλαπλές αγγειακές εδάφη συμβεί μεταξύ 40 και 50 ετών. [11] Ωστόσο, η μαγνητική Τομογραφία είναι σε θέση να ανιχνεύουν σημάδια του τα χρόνια ασθένεια πριν από την κλινική εκδήλωση της νόσου. [15] Τα συμπτώματα και η έναρξη της νόσου ποικίλλουν ευρέως. Μερικά άτομα είναι πιθανόν να μην παρουσιάσουν συμπτώματα της ασθένειας μέχρι αργότερα στη ζωή.

Άλλες κλινικές ονομασίες για την ασθένεια CADASIL είναι: κληρονομική άνοια, πολλαπλού εμφράγματος, χρόνια οικογενής αγγειακή εγκεφαλοπάθεια, οικογενής διαταραχή με υποφλοιώδεις ισχαιμικά εγκεφαλικά επεισόδια, αγνώστου προελεύσεως έσω αρτηριοπάθεια και οικογενής νόσος του Binswanger.

2.2.2. Επιδημιολογία της CADASIL

Αν και η ασθένεια CADASIL περιγράφηκε για πρώτη φορά στην Ευρώπη, έχει παρατηρηθεί σε ντόπιους ασιατικούς, αφρικανικούς και αμερικανικούς πληθυσμούς καθώς και σε πληθυσμούς της Μέσης Ανατολής. [7, 12] Με την καλύτερη γνώση της κλινικής εικόνας και τις βελτιωμένες διαγνωστικές εξετάσεις, αυξάνεται σταθερά ο αριθμός των διαπιστωμένων με CADASIL οικογενειών. Μέχρι σήμερα, περίπου 400 οικογένειες έχουν αναφερθεί. [17] Πιο συγκεκριμένα, στη Γαλλία δεν έχει πραγματοποιηθεί καμία επιδημιολογική μελέτη για CADASIL. Σύμφωνα με μια μελέτη που διενεργήθηκε στα δυτικά της Σκωτίας το 2002, 22 ασθενείς καταμετρήθηκαν με CADASIL από 7 οικογένειες από έναν πληθυσμό 1,418,990. Η επικράτηση-βαθμός διάδοσης (prevalence) με οριστική διάγνωση είναι 1.98 (1.24–3.00)/100.000 (διάστημα αξιοπιστίας κατά 95%). Επιπλέον 37 άτομα (ασθενείς με προβλεπόμενη διάγνωση, που κινδυνεύουν να είναι φορείς του μεταλλαγμένου γονιδίου

NOTCH3) παρουσιάζουν επικράτηση πιθανής μετάλλαξης 4.14 (3.04–5.53) ανά 100.000 ενήλικες. [10] Τα στοιχεία αυτά είναι ταυτόσημα με εκείνα στη Φινλανδία, όπου η επικράτηση των μοριακά γενετικά επαληθευμένων ασθενών είναι περίπου 3, και όταν βασίζεται σε οικογενειακό ιστορικό είναι 4 ανά 100.000. Στο Ηνωμένο Βασίλειο, η επικράτηση έχει υπολογιστεί τουλάχιστον της τάξης του 1 ανά 100.000. Όπως φαίνεται, ο μεγαλύτερος αριθμός των ασθενών οικογενειών που έχουν μέχρι τώρα σημειωθεί είναι μεταξύ των Καυκασίων Ευρωπαίων, ενώ στις ΗΠΑ και τον Καναδά, οι περιπτώσεις που αναφέρθηκαν ήταν εκπληκτικά λίγες, λαμβάνοντας υπόψη το μέγεθος του πληθυσμού τους, το υψηλό επίπεδο ανάπτυξης της νευρολογίας και την υψηλή επιστημονική εκδοτική τους δραστηριότητα. [16] Για να είμαστε ακριβείς, η πρώτη βορειοαμερικανική οικογένεια με CADASIL που συνδέεται με το χρωμόσωμα 19 αναφέρθηκε το 1996. [18]

Η CADASIL φαίνεται να προσβάλλει ισομερώς τα δύο φύλα. Η έναρξη των κλινικών συμπτωμάτων συμβαίνει συνήθως μεταξύ των 40 και 50 ετών, με μια μέση ηλικία τα 46.1 χρόνια. Ωστόσο, έχει αναφερθεί συμπτωματική έναρξη και στην ηλικία των 8 ετών. Η ημικρανία με αύρα ως σύμπτωμα της ασθένειας εμφανίζεται στο ένα τρίτο περίπου των ασθενών και μπορεί να αρχίσει ήδη πριν από την ηλικία των 10 ετών. [7]

Η ηλικία κατά την οποία συμβαίνει το πρώτο εγκεφαλικό επεισόδιο ποικίλλει ευρέως, συνήθως από 25 έως 65 ετών. Ωστόσο, ο νεότερος ασθενής που έχει αναφερθεί με εγκεφαλικό επεισόδιο ήταν 11 χρονών. [19] Ένας στους εννιά (11%) ασθενείς ηλικίας 50 ετών ή λιγότερο με κενотоπιώδη έμφρακτα και λευκοαραιώση έπασχε από CADASIL, ενώ μεταξύ των ασθενών κάτω των 65 ετών, η συχνότητα ήταν 1 στους 48 ασθενείς. 2%, με εκτιμώμενη ευαισθησία γενετικής ανάλυσης 90%, με βάση τον προσυμπτωματικό έλεγχο για μεταλλάξεις στα εξόνια 3, 4, 5 και 6 από PCR ακολουθούμενο από ανάλυση για τον πολυμορφισμό διαμόρφωσης μονόκλωνης αλυσίδας. Είναι, ωστόσο, πιθανό η συχνότητα της νόσου να είναι ακόμη υποτιμημένη. [12] Παρόλα αυτά, για τα οριστικά στοιχεία επικράτησης πρέπει να προηγηθούν αναλύσεις σε αρκετά μεγάλους πληθυσμούς. [16]

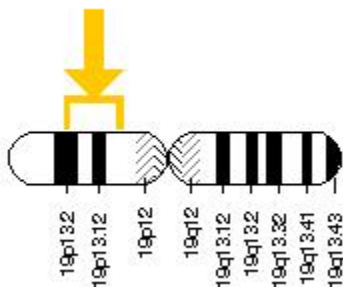
2.2.3. Η Πρωτεΐνη Notch3 και οι Δομικές της Ιδιότητες

Σε γενικές γραμμές, τα γονίδια Notch κωδικοποιούν υποδοχείς που πυροδοτούν μικρής εμβέλειας γεγονότα σηματοδότησης. Ένα πρωτότυπο γονίδιο Notch κωδικοποιεί έναν μονό διαμεμβρανικό υποδοχέα. [16, 20]

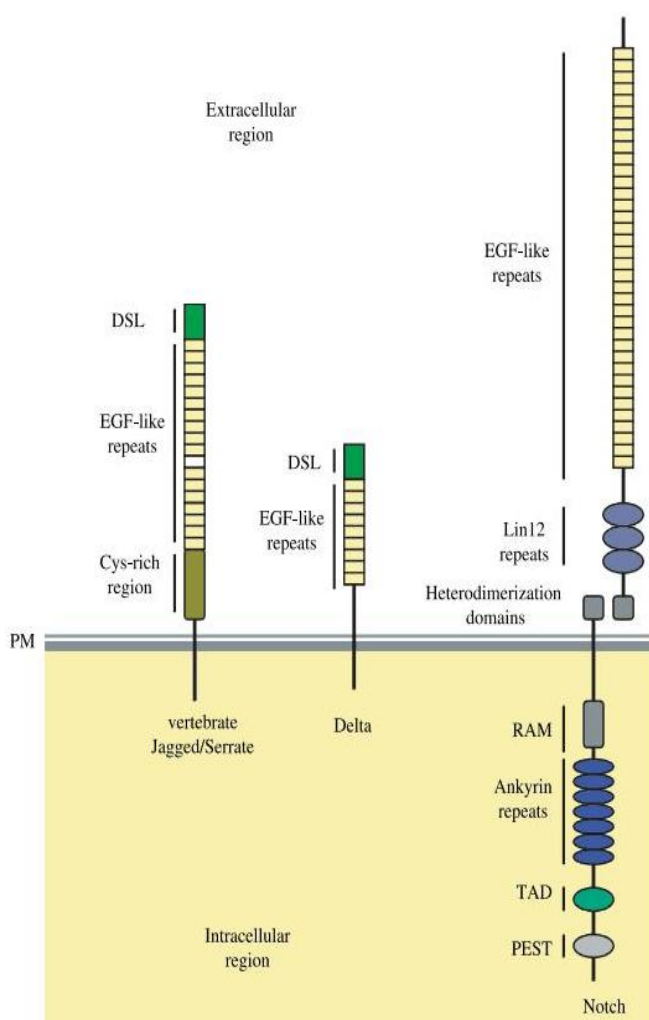
Συγκεκριμένα, το γονίδιο NOTCH3 εντοπίστηκε στις αρχές της δεκαετίας ως το τρίτο γονίδιο Notch των θηλαστικών και αρχικά αναφέρθηκε ότι εκφράζεται στο νευροεπιθήλιο. [21] Παρέχει οδηγίες για την παραγωγή της πρωτεΐνης-υποδοχέα Notch3, που βρίσκεται στην επιφάνεια των κυττάρων μυών που περιβάλλουν τα αιμοφόρα αγγεία (αγγειακά λεία μυϊκά κύτταρα-VSMC). [22] Όταν ορισμένα μόρια προσκολλώνται (δεσμεύουν) στους υποδοχείς Notch3, οι τελευταίοι αποστέλλουν μηνύματα προς τον πυρήνα του κυττάρου.

Τα σήματα αυτά στη συνέχεια ενεργοποιούν συγκεκριμένα γονίδια εντός των VSMCs. Οι υποδοχείς Notch3 διαδραματίζουν καίριο ρόλο στην βέλτιστη λειτουργικότητα, την επιβίωση και τη συντήρηση των μυϊκών κυττάρων στο αρτηριακό δίκτυο του εγκεφάλου. [23]

Το γονίδιο NOTCH3 βρίσκεται στον μικρό (p) βραχίονα του χρωμοσώματος 19 μεταξύ των θέσεων 13.2 και 13.1 (κυτταρογενετική θέση, σχήμα 2.1.). Ακριβέστερα, η μοριακή θέση του γονιδίου NOTCH3 είναι μεταξύ των ζευγών βάσεων 15.159.633 και 15.200.981 στο χρωμόσωμα 19. (NCBI, Gene ID: 4854, updated on 22-Jun-2014)



Σχήμα 2.1. Κυτταρογενετική θέση του γονιδίου NOTCH3 στο χρωμόσωμα 19. (εικόνα από <http://ghr.nlm.nih.gov/gene/NOTCH3>)

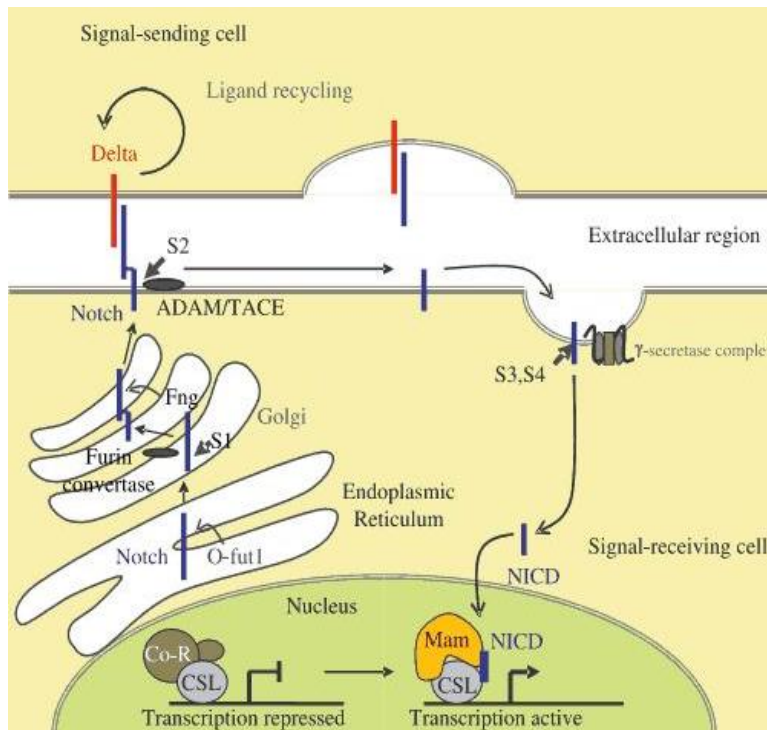


Η Notch αποτελείται στην εξωκυτταρική της περιοχή από μια διατηρημένη μικροσυστοιχία με έως 36 επαναλήψεις επιδερμικού αυξητικού παράγοντα (EGF) που συμβάλλουν στην αλληλεπίδραση του συνδέτη, και τρεις Lin-12-Notch (LN) επαναλήψεις, οι οποίες είναι παραμεμβρανικές επαναλήψεις υπεύθυνες για τις αλληλεπιδράσεις μεταξύ των εξωκυτταρικών και των μεμβρανο-συνδεδεμένων (membrane-tethering) ενδοκυτταρικών περιοχών. Η ενδοκυτταρική περιοχή της Notch αποτελείται από επτά επαναλήψεις ανκυρίνης που ορίζονται από σήματα πυρηνικού εντοπισμού (NLS), έναν τομέα PEST (προλίνη, γλουταμίνη, σερίνη, θρεονίνη) και ένα τομέα TAD (Transactivation Domain- πεδίο μετενεργοποίησης). Ο C-τερματικός τομέας ετεροδιμερισμού είναι μία υδρόφοβη περιοχή της εξωκυτταρικής Notch, ικανός να δημιουργήσει ένα σταθερό σύμπλοκο με την εξωκυτταρική περιοχή της διαμεμβρανικής Notch. (Σχήμα 2.2, αριστερά)

Σχήμα 2.2. Η δομή της πρωτεΐνης Notch και των συνδετών της. (Cell and molecular biology of Notch. Fiuza, U. M. & Arias, A. M.)

Η διάσπαση του υποδοχέα Notch συμβαίνει στην περιοχή S1 από μία κονβερτάση όμοια με φουρίνη στην trans-Golgi περιοχή, κατά τη διάρκεια της διαδικασίας εκκρίσεως. Η διαδικασία της διάσπασης, καθώς και η προκύπτουσα δομή είναι ζωτικής σημασίας για την λειτουργία της Notch στα θηλαστικά. (Σχήμα 2.3)

Στην επιφάνεια του κυττάρου, η Notch μπορεί να αλληλεπιδράσει με έναν από τους συνδέτες της Delta (DI) ή Serrate (Ser) που εκφράζονται σε ένα γειτονικό κύτταρο. Αυτή η αλληλεπίδραση ρίχνει την εκτοπεριοχή και εκθέτει τον εξωκυττάριο χώρο μεταλλοπρωτεάσης (περιοχή S2), ο οποίος γίνεται, έτσι, ευάλωτος σε διάσπαση από διαμεμβρανικές πρωτεάσες της οικογένειας ADAM / TAC. Στη συνέχεια, το υπόλοιπο μεμβρανο-προσδετό τμήμα της Notch υφίσταται διάσπαση στην περιοχή S2 και δύο επιπλέον ενδομεμβρανικές διασπάσεις (στις περιοχές S3/S4), λόγω της g-σεκρετάσης ενός πρωτεϊνικού συμπλόκου της μεμβράνης που περιέχει πρεσενιλίνη ως καταλυτικό συστατικό. Ως εκ τούτου, η ενδοκυτταρική περιοχή της Notch τελικά απελευθερώνεται και μετατοπίζεται στον πυρήνα όπου ρυθμίζει την έκφραση του γονιδίου, δρώντας ως συν-ενεργοποιητής του μεταγραφικού παράγοντα του γονιδίου Su(H) (Suppressor of Hairless). [20]

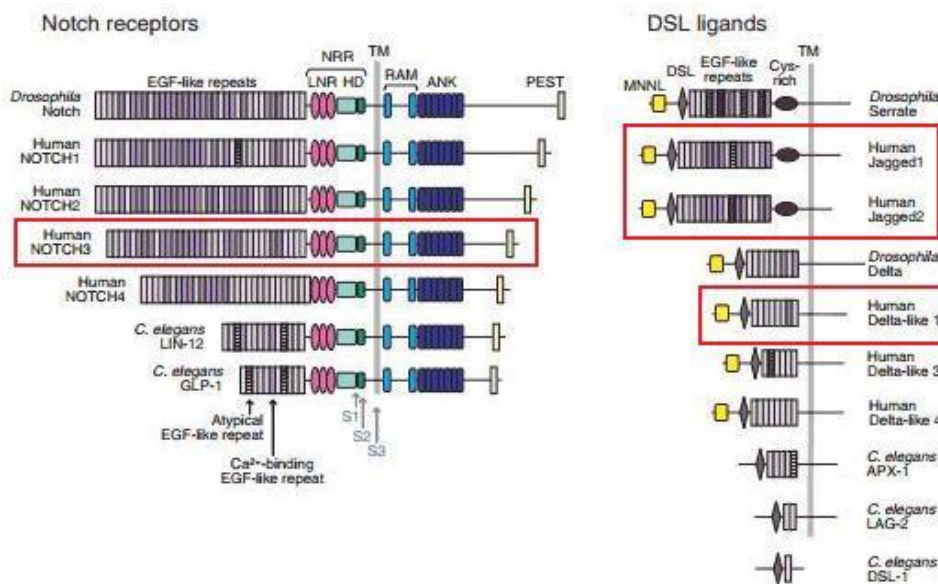


Σχήμα 2.3. Κανονικό σηματοδοτικό μονοπάτι της Notch (Cell and molecular biology of Notch. Fiuza, U. M. & Arias, A. M.)

Η αρνητική ρυθμιστική περιοχή (NRR) του υποδοχέα Notch, η οποία παρεμβάλλεται μεταξύ της περιοχής πρόσδεσης του συνδέτη και της διαμεμβρανικής περιοχής, προστατεύει τον δομικό μηχανισμό που είναι απαραίτητος για να διατηρηθεί η αντίσταση μεταλλοπρωτεάσης απουσία συνδέτη. Η NRR περιλαμβάνει τρεις πλούσιες σε κυστεΐνη επαναλήψεις LIN-12-Notch (LNRs) και ένα πεδίο ετεροδιμερισμού (HD), το οποίο έρχεται ακριβώς πριν από την μεμβράνη και περιέχει τόσο την περιοχή διάσπασης της φουρίνης (S1) όσο και της μεταλλοπρωτεάσης (S2). Πολλά χρόνια μελετών παρείχαν οδήγησαν στο συμπέρασμα ότι η

NRR λειτουργεί ως «διακόπτη ενεργοποίησης» του υποδοχέα. Οι υποδοχείς στους οποίους λείπουν οι επαναλήψεις τύπου EGF είναι λειτουργικά ανενεργοί. Σε αντίθεση, η διαγραφή των ενοτήτων LNR ή οι σημειακές μεταλλάξεις των βασικών καταλοίπων στο εσωτερικό τους προκαλούν φαινότυπους με αυξημένο κύκλο λειτουργίας, καθώς και ανεξάρτητη από τον συνδέτη διάσπαση της μεταλλοπρωτεάσης. [24]

Η πρωτεΐνη Notch3 λειτουργεί ως υποδοχέας για τους μεμβρανο-συνδεόμενους συνδέτες Jagged1, Jagged2 και Delta1 (Σχήμα 2.4) με σκοπό να ελέγχει τον προσδιορισμό της κυτταρικής μοίρας, τόσο κατά τη διάρκεια της εμβρυϊκής ανάπτυξης όσο και κατά την ομοίωση των ενήλικων ιστών. Η επαφή μεταξύ των κυττάρων αποτελεί βασική προϋπόθεση για την κυτταρική σηματοδότηση. Το πιο σημαντικό χαρακτηριστικό της Notch3 είναι η λειτουργία της ως διαμεμβρανικός υποδοχέας και ως μεταγραφικός παράγοντας. Κατά την ενεργοποίηση του συνδέτη μέσω της ενδοκυτταρικής περιοχής της Notch (Notch Intracellular Domain-NICD), επηρεάζεται η εφαρμογή των προγραμμάτων διαφοροποίησης, πολλαπλασιασμού και απόπτωσης. [8, 20, 24]



Σχήμα 2.4. Οργάνωση της επικράτειας των υποδοχέων της Notch και των DSL συνδετών της. (The molecular logic of Notch signaling – a structural and biochemical perspective. Gordon, W. R., Arnett, K. L. & Blacklow, S. C.)

2.2.4. Μοριακός Μηχανισμός / Συσσώρευση Αποθέσεων GOM

Περισσότερες από 190 μεταλλάξεις [23] στο γονίδιο NOTCH3 (στο μικρό βραχίονα του χρωμοσώματος 19q12) έχει βρεθεί ότι προκαλούν CADASIL. Το γονίδιο κωδικοποιεί μια μεγάλη διαμεμβρανική (Notch3) πρωτεΐνη-υποδοχέα, η οποία κατέχει κρίσιμο ρόλο εντός του αγγειακού λείου μυϊκού σηματοδότηκού μονοπατιού. Αποτελείται από 33 εξόνια. Σε μοριακό επίπεδο, σχεδόν όλες οι παθογόνες μεταλλάξεις οδηγούν σε μονό αριθμό καταλοίπων κυστεΐνης (που μένουν ασύζευκτα, λόγω της προσθήκης ή διαγραφής ενός μορίου κυστεΐνης) σε μία από τις 34 επαναλήψεις τύπου αυξητικού επιδερμικού παράγοντα (EGF) στην εξωκυτταρική περιοχή (N3ECD). Αυτό έχει ως αποτέλεσμα κακή αναδίπλωση της πρωτεΐνης Notch3. Το γονίδιο NOTCH3 εκφράζεται αποκλειστικά σε αγγειακά λεία μυϊκά κύτταρα

(VSMC) και αυτή η λανθασμένη αναδίπλωση (misfolding) της πρωτεΐνης οδηγεί σε μια σταδιακή συσσώρευση N3ECD στην επιφάνεια των εκφυλιζόμενων VSMC. Εκτός από τη συσσώρευση N3ECD, η εναπόθεση κοκκώδους οσμιοφιλικού υλικού (granular osmiophilic material- GOM) στο τοίχωμα του αγγείου είναι ένα άλλο χαρακτηριστικό μορφολογικό χαρακτηριστικό. Έχουν αμφότερα περιγραφεί μόνο στην νόσο CADASIL και αποτελούν παθολογικά χαρακτηριστικά για την ασθένεια. [19, 23, 25]

Διάφορες νευροεκφυλιστικές ασθένειες έχουν παρουσιάσει το παθολογικό γνώρισμα-κλειδί της συσσωμάτωσης και τη συσσώρευση των ασυνήθιστα διπλωμένων πρωτεϊνών. Έχει αναφερθεί πρόσφατα ότι ο πολυμερισμός της Notch3 ενισχύεται σημαντικά από τις μεταλλάξεις που σχετίζονται με την ασθένεια CADASIL. Σε περίπτωση που οι αποθέσεις GOM περιέχουν N3ECD, τα ολιγομερή της μεταλλαγμένης πρωτεΐνης Notch3 που σχηματίζονται αυτόματα μπορεί να είναι τα στέγαστρα (lodgements) των GOM. Η αποικοδόμηση δεν επηρεάζει τα συσσωματώματα της μεταλλαγμένης Notch3 και, όπως και σε πολλές διαταραχές που σχετίζονται με τη συσσώρευση των ασυνήθιστα διπλωμένα πρωτεϊνών, ο σχηματισμός τους μπορεί να προκαλέσει κυτταρική δυσλειτουργία και τελικά κυτταρικό θάνατο. Ωστόσο, ο εκφυλισμός των VSMC και η απώλειά τους μπορεί να μην είναι οι μόνες αιτίες της καταστροφής της ακεραιότητας του τοιχώματος των μικρών αγγείων στην νόσο CADASIL. Άλλοι παράγοντες που συμβάλλουν σε αυτή τη διαδικασία είναι η υαλίνωση των αγγείων, η ίνωση με ισχυρή θετική ανοσοχρώση με κολλαγόνα III και IV και η μη φυσιολογική αύξηση του χώρου μεταξύ των κυττάρων στο τοίχωμα του αγγείου, η οποία παρατηρείται σε φωτονικό (οπτικό) ή ηλεκτρονικό μικροσκόπιο.

Αρκετές περιπτώσεις που εξετάστηκαν παρουσίασαν ετερογενή μικροσκοπική εικόνα των αποθέσεων GOM. Πολλά αιμοφόρα αγγεία βρέθηκαν με πολλές εξωκυτταρικές αποθέσεις GOM που ποικίλλουν σε μέγεθος, σχήμα και μορφολογία, καθώς και στη θέση τους στα αγγειακά τοιχώματα. Φάνηκε ότι το GOM μπορεί να σχηματιστεί στην επιφάνεια του VSMC στο τοίχωμα του αγγείου και να διεισδύσει από τα κύτταρα στο κάτω μέρος μεμβράνης και / ή στην εξωκυτταρική μήτρα.

Καθώς απομακρυνόμαστε από το κυτταρικό σώμα, το στρώμα των αποθέσεων καθίσταται σταδιακά πιο λεπτό και χαλαρό. Η αραίωση του κοκκώδους υλικού σε αποθέσεις GOM αυξάνεται ανάλογα με την απόσταση από το VSMC και ακολουθείται από χαρακτηριστικές αλλοιώσεις στο σχήμα τους, έχοντας ως αποτέλεσμα μια σταδιακή διάλυση των στεγαστρών των GOM. Σε αγγεία με σοβαρή βλάβη των VSMC, οι αποθέσεις GOM που βρίσκονται στην εξωτερική επιφάνεια των αρτηριών κατέδειξαν πυκνότητα ηλεκτρονίων παρόμοια με την πυκνότητα της βασικής μεμβράνης.

Αυτό το αποτέλεσμα θα μπορούσε να υποστηρίξει όχι μόνο την υπόθεσή μας σχετικά με τη σταδιακή κατάρρευση των αποθέσεων GOM, αλλά ρίχνει φως και στην καταγωγή τους. Βασίζομενοι στη διαπίστωση ότι οι αποθέσεις GOM βρίσκονται γύρω από τα εκφυλισμένα VSMC ή στις κοιλότητες των κυττάρων αυτών, αλλά όχι εντός των κυττάρων, καταλήγουμε ότι ο σχηματισμός τους λαμβάνει χώρα λογικά στην κυτταρική επιφάνεια. Η ίδια διαδικασία μπορεί να σχετίζεται (α) με τη δυσλειτουργία του ενζύμου TACE, που διαχωρίζει πρωτεολυτικά την N3ECD από τον υποδοχέα Notch3 μετά τη δέσμευση του συνδέτη ή (β) με

την διαταραγμένη ενδοκύτωση, η οποία, υπό κανονικές συνθήκες, αφαιρεί την N3ECD από την κυτταρική μεμβράνη. Μία άλλη πιθανότητα είναι ότι οι αποθέσεις GOM σχηματίζονται στο ενδοπλασματικό δίκτυο και απελευθερώνονται στον εξωκυτταρικό χώρο από τη διατάραξη των εκφυλισμένων κυττάρων που εκφράζουν τον υποδοχέα Notch3. Από την άλλη πλευρά, η παρουσία αποθέσεων GOM κοντά στα σχετικά καλά διατηρημένα VSMC είναι κατά αυτής της υπόθεσης. Η διάχυση των κοκκίων που σχηματίζουν GOM αποθέσεις μπορεί να αποτελεί έναν από τους παράγοντες που προκαλούν την πάχυνση του και μεταβολές στη βασική μεμβράνη και / ή εξωκυτταρική μήτρα. [25]

Ωστόσο, πρόσφατα, έχουν αναφερθεί και ορισμένες μη-Cys μεταλλάξεις. Αυτές οι μεταλλάξεις δεν ακολουθούν τη χαρακτηριστική παθολογία και το σχέδιο της νόσου. Περαιτέρω, έχει αναγνωριστεί ότι ορισμένες από τα μη-Cys μεταλλάξεις προκαλούν GOM ενώ μερικές άλλες όχι. [23]

2.2.5. Διάγνωση της ασθένειας

Σήμερα, η πλέον αξιόπιστη μέθοδος διάγνωσης της νόσου είναι η ανάλυση της αλληλουχίας του γονιδίου NOTCH3, με άλλα λόγια ο γενετικός έλεγχος. Αυτή η μέθοδος μπορεί να διαγνώσει το 95% των περιπτώσεων CADASIL με βεβαιότητα και περιλαμβάνει μια εξέταση αίματος σε ειδικευμένο εργαστήριο. Το αποτέλεσμα της εξέτασης κάνει τη διάγνωση της νόσου και σε άλλα μέλη της οικογένειας σχετικά εύκολη, καθώς στο πλαίσιο της ίδιας οικογένειας, όλοι όσοι επηρεάζονται από τη νόσο θα έχουν την ίδια μετάλλαξη NOTCH3. Ανίχνευση μιας μετάλλαξης θεωρείται 100% απόδειξη της ύπαρξης της νόσου.

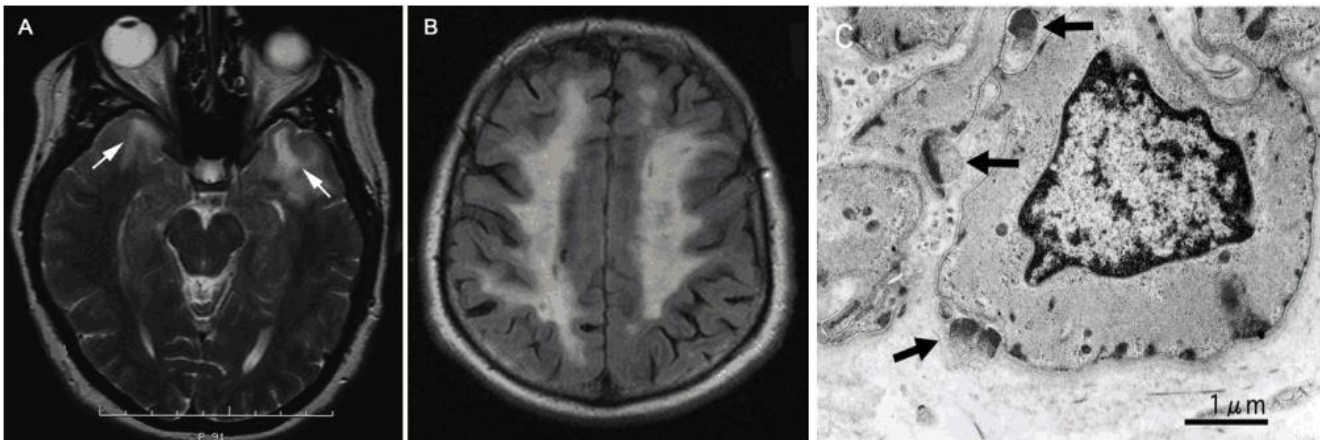
Η κατανομή των μεταλλάξεων στο γονίδιο διαφέρει μεταξύ των χωρών. Στη μελέτη επιπολασμού στο Ηνωμένο Βασίλειο, το 73% των μεταλλάξεων βρέθηκε να είναι στο εξόνιο 4. Η αναζήτηση για μεταλλάξεις σε όλα τα εξόνια θα ήταν κοπιώδης, πολύπλοκη και δαπανηρή. Το NOTCH3 είναι ένα μεγάλο γονίδιο και η δοκιμασία διαλογής (screening test) για όλα τα εξόνια είναι κλινικά ανέφικτη. Υπάρχουν, επίσης, σχετικές δημοσιονομικές επιπτώσεις. Για το λόγο αυτό, πολλά διαγνωστικά εργαστήρια δοκιμάζουν μόνο τις θέσεις στο γονίδιο που είναι πιο πιθανό να εμφανίζουν μετάλλαξη που ενοχοποιείται για CADASIL. Παρά το γεγονός ότι η επιλεκτική διαλογή για εξόνια θα ανιχνεύσει την πλειονότητα των περιπτώσεων και κάνει το γενετικό τεστ ευκολότερο και λιγότερο δαπανηρό, κάποιες μεταλλάξεις μπορεί ακόμα να χαθούν. Ως εκ τούτου, σε μεμονωμένες περιπτώσεις, μπορεί να είναι αναγκαία η αναζήτηση για μια μετάλλαξη σε ολόκληρο το γονίδιο. [8]

Πριν το γενετικό τεστ καταστεί διαθέσιμο, η διάγνωση της νόσου CADASIL γινόταν με βιοψία δέρματος. Με άλλα λόγια, η ανίχνευση του κοκκώδους οσμιοφιλικού υλικού (GOM) σε οδοντώσεις των VSMCs ή στον εξωκυτταρικό χώρο, σε στενή γειτνίαση με VSMCs (Σχήμα 2.5C) γινόταν με ηλεκτρονική μικροσκοπία (EM) στο δέρμα, τους μύς ή τα νεύρα. Η ακριβής σύνθεση του GOM δεν έχει εξακριβωθεί, αλλά μια μελέτη ανοσοχρυσού σε επίπεδο ηλεκτρονικού μικροσκοπίου έδειξε το N3ECD να είναι ένα συστατικό του GOM.

Η ανίχνευση των GOM μέσω EM προτιμάται ως μέθοδος στην ανάλυση βιοψιών δέρματος, αν ληφθεί υπόψη η υψηλή εξειδίκευση που παρουσιάζει και η ευαισθησία της. Δεν είναι ούτε

χρονοβόρα ούτε υπερβολικά ακριβή και είναι λιγότερο επεμβατική. Επιπροσθέτως, είναι αδύνατον να εκτιμηθεί το πόσο μακριά πρέπει να προχωρήσει κανείς με τις γενετικές αναλύσεις. Μολονότι οι αποθέσεις GOM δεν έχουν αναφερθεί σε καμία άλλη ασθένεια και η εξειδίκευση θεωρείται ότι είναι 100%, οι εκθέσεις σχετικά με την ευαισθησία της ανίχνευσης GOM σε βιοψία δέρματος των ασθενών με γενετικά επαληθευμένη CADASIL υπήρξαν αντιφατικές. Πράγματι, η παρουσία του υλικού μπορεί να διαγνώσει θετικά CADASIL. Ωστόσο, ένα αρνητικό αποτέλεσμα δεν σημαίνει απαραίτητα απουσία της νόσου. Παρ' όλα αυτά, η συσσώρευση λαμβάνει χώρα πολύ πριν από οποιαδήποτε συμπτώματα και ως εκ τούτου μπορεί να χρησιμοποιηθεί για τη διάγνωση άλλων μελών της οικογένειας. Επί του παρόντος, η βιοψία του δέρματος μπορεί να χρησιμοποιηθεί για να επιβεβαιώσει αμφίβολες περιπτώσεις. [22,26]

Εναλλακτικά, θα μπορούσε να χρησιμοποιηθεί η συμβατική μαγνητική τομογραφία (MRI) για τη διάγνωση της CADASIL, καθώς οι ασθενείς εμφανίζουν πολλαπλές συρρέουσες βλάβες της λευκής ουσίας σε διάφορα μεγέθη. Η MRI παρέχει ευρεία προβολή των περιοχών του εγκεφάλου που επηρεάζονται από τη νόσο, καθώς εμφανίζει μειωμένη ένταση σήματος στην ακολουθία T1 και αυξημένη στην T2 (Σχήμα 2.5A & B). Αυτά τα κλινικά και ακτινολογικά ευρήματα μπορεί να υποδεικνύουν έντονα προσβολή από τη νόσο CADASIL όταν συνοδεύονται από θετικό οικογενειακό ιστορικό, αλλά το καθένα από αυτά μεμονωμένα δεν είναι ειδικό για κατάδειξη της ασθένειας. Επομένως, η μαγνητική τομογραφία του εγκεφάλου δεν πρέπει να χρησιμοποιείται ως μοναδικό διαγνωστικό εργαλείο, αν και μπορεί να δείξει την εξέλιξη των μεταβολών της λευκής ουσίας ακόμη και δεκαετίες πριν από την έναρξη των συμπτωμάτων της νόσου. [15, 26]



Σχήμα 2.5. Διάγνωση της CADASIL με MRI.(A) Αυξημένη ένταση σήματος στον πρόσθιο κροταφικό λοβό σε ακολουθία T2, αμέσως μετά το πρώτο παροδικό ισχαιμικό επεισόδιο της ασθενούς. (B) FLAIR MRI που δείχνει εκταταμένη ένταση σήματος στην εγκεφαλική λευκή ουσία. (Diagnosing Vascular Dementia by Skin Biopsy - Uniqueness of CADASIL. Tikka, S., Mykkänen, K., Junna, M., Bergholm, R., Røyhönen, M., Baumann, M., Viitanen, M. & Kalimo, H.) (C) Τυπικές αποθέσεις GOM στην εξωτερική επιφάνεια λείου μυϊκού κυττάρου υποδεικνύονται από τα βέλη. (Notch3 ectodomain is a major component of granular osmiophilic material (GOM) in CADASIL. Ishiko, A., Shimizu, A., Nagata, E., Takahashi, K., Tabira, T. & Suzuki, N.)

Κεφάλαιο 3

Μέθοδοι επεξεργασίας των ιατρικών δεδομένων και μελέτης της αλληλουχίας και της δομής της Notch3

3.1. Βιοπληροφορική

Μια αναζήτηση BLAST, με την ακολουθία της πρωτεΐνης Notch3 πραγματοποιήθηκε στο Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής (European Bioinformatics Institute - EBI), προκειμένου να βρεθούν παρόμοιες ακολουθίες στη βάση δεδομένων Uniprot [27]. Όλες οι ακολουθίες με ένα E-value κατώτερο του $7.0E-4$ επιλέχθηκαν για ανάλυση μέσω του λογισμικού Multiple EM for Motif Elicitation (MEME). Η ανάλυση πραγματοποιήθηκε στο http://meme.sdsc.edu/meme4_6_1/ χρησιμοποιώντας προκαθορισμένες τιμές. Τα λογότυπα που παρατίθενται στην παρούσα εργασία για τα μοτίβα που βρέθηκαν δημιουργήθηκαν με τη βοήθεια του λογισμικού WebLogo. [28]

3.2. Ιατρικά Δεδομένα

Η συνολική λίστα με τις σημειακές μεταλλάξεις της πρωτεΐνης Notch3 δόθηκε από τον καθηγητή Δρ. Marc Baumann (Μονάδα Πρωτεϊνικής Χημείας, Ινστιτούτο Βιοϊατρικής/Ανατομίας, Πανεπιστήμιο του Ελσίνκι, Φινλανδία) και παρατίθεται στο παράρτημα στο τέλος της εργασίας. Ο ακόλουθος πίνακας (πίνακας 3.1) συνοψίζει τις σημειακές μεταλλάξεις που εξετάστηκαν στα πλαίσια αυτής της διπλωματικής εργασίας.

A	Arg107Trp	Καναδική περίπτωση, προηγουμένως μη αναφερθείσα (δεδομένα που δόθηκαν από τον καθηγητή Baumann, M.)
B	Ser497Leu	αναφέρεται στη βάση δεδομένων ESP (NHLBI GO Exome Sequencing Project) (στοιχεία που δόθηκαν από τον καθηγητή Baumann, M.)
C	Glu813Lys	αναφέρεται στη βάση δεδομένων ESP (NHLBI GO Exome Sequencing Project) (στοιχεία που δόθηκαν από τον καθηγητή Baumann, M.)
D	Ala1020Pro	Scheid, R. et al. 2008, Cysteine-sparing notch3 mutations: cadasil or cadasil variants? <i>Neurology</i> , 71 (10), p. 774-6.
E	Ser978Arg	Ferreira, S. et al. 2007, Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL). <i>Hum. Genet.</i> , 121 (5), p. 649-50.
F	His1133Glu	Καναδική περίπτωση, (δεδομένα που δόθηκαν από τον καθηγητή Baumann, M.)

Πίνακας 3.1. Σημειακές μεταλλάξεις της πρωτεΐνης Notch3 που εξετάζονται στην παρούσα μελέτη

3.3. Στοιχίση/Ευθυγράμμιση Αλληλουχιών

3.3.1. Εισαγωγή

Οι πολλαπλές ευθυγραμμίσεις των πρωτεϊνικών αλληλουχιών είναι σημαντικές σε πολλές εφαρμογές, συμπεριλαμβανομένου του υπολογισμού φυλογενετικών δέντρων, της πρόβλεψης της δομής και της αναγνώρισης κρίσιμων καταλοίπων. Η πιο φυσική διαμόρφωση του υπολογιστικού προβλήματος είναι να καθοριστεί ένα μοντέλο εξέλιξης της ακολουθίας, το οποίο αναθέτει πιθανότητες σε στοιχειώδεις αλλαγές της ακολουθίας και ψάχνει ένα πιο πιθανό κατευθυνόμενο γράφημα. Στο γράφημα αυτό, οι ακμές αντιπροσωπεύουν τις αλλαγές και οι τερματικοί κόμβοι είναι οι αλληλουχίες που μελετώνται. Πάγια μέθοδος για την εύρεση τέτοιου γραφήματος παραμένει άγνωστη.

Για την ευθυγράμμιση αλληλουχιών χρησιμοποιούνται κυρίως δύο αλγόριθμοι: ο ClustalW και ο MUSCLE, οι οποίοι περιγράφονται συνοπτικά στην συνέχεια.

3.3.2. Αλγόριθμοι ClustalW και MUSCLE

Το **ClustalW** είναι ένα ευρέως χρησιμοποιούμενο σύστημα για την ευθυγράμμιση οποιουδήποτε αριθμού ομόλογων νουκλεοτιδικών ή πρωτεϊνικών αλληλουχιών. Για την ευθυγράμμιση πολλαπλών αλληλουχιών, το ClustalW χρησιμοποιεί προοδευτικές μεθόδους ευθυγράμμισης. Σε αυτές, οι περισσότερο όμοιες αλληλουχίες, δηλαδή αυτές με την καλύτερη βαθμολογία ευθυγράμμισης (alignment score), ευθυγραμμίζονται πρώτες. Στη συνέχεια, προοδευτικά, οι πιο μακρινές ομάδες αλληλουχιών ευθυγραμμίζονται μέχρι να επιτευχθεί σφαιρική ευθυγράμμιση. Αυτή η ευριστική προσέγγιση είναι αναγκαία, επειδή η εύρεση της σφαιρικής βέλτιστης λύσης είναι απαγορευτική τόσο από άποψη απαιτήσεων μνήμης όσο και χρόνου.

Ο ClustalW αποδίδει πολύ καλά στην πράξη. Ο αλγόριθμος ξεκινάει υπολογίζοντας μια πρόχειρη μήτρα αποστάσεων μεταξύ κάθε ζεύγους αλληλουχιών βασιζόμενος στις βαθμολογίες ευθυγράμμισης του ζεύγους. Αυτές οι βαθμολογίες υπολογίζονται χρησιμοποιώντας τις παραμέτρους ευθυγράμμισης του ζεύγους για αλληλουχίες DNA και πρωτεϊνών. Στη συνέχεια, ο αλγόριθμος χρησιμοποιεί την μέθοδο ένωσης γειτόνων (neighbor-joining) με midpoint rooting (MPR) για να κατασκευάσει ένα δέντρο-οδηγό, που θα βοηθήσει στη σφαιρική ευθυγράμμιση. Εναλλακτικά, μπορεί να κατασκευαστεί και ένα δέντρο-οδηγός σε μορφή Newick. Το δέντρο-οδηγός χρησιμεύει ως ένα πρόχειρο πρότυπο για τους κλάδους που τείνουν να μοιράζονται χαρακτηριστικά εισαγωγής και διαγραφής.

Ο αλγόριθμος αυτός, γενικά, δίνει αποτέλεσμα κοντά στο βέλτιστο, ειδικά όταν το σύνολο δεδομένων περιλαμβάνει αλληλουχίες με ποικίλους βαθμούς απόκλισης. Άρα το δέντρο-οδηγός είναι λιγότερο ευαίσθητο στον θόρυβο. [29]

Το **MUSCLE** είναι ένα άλλο υπολογιστικό πρόγραμμα για την παραγωγή πολλαπλών ευθυγραμμίσεων πρωτεϊνικών αλληλουχιών. Χαρακτηριστικά του αλγορίθμου αποτελούν ο γρήγορος υπολογισμός αποστάσεων χρησιμοποιώντας την καταμέτρηση kmer, η σταδιακή ευθυγράμμιση αξιοποιώντας μια νέα λειτουργία προφίλ κατά την οποία καλούμε το προσδοκώμενο αποτέλεσμα του log, και η βελτιστοποίηση, χρησιμοποιώντας περιορισμένη στεγανοποίηση που εξαρτάται από το δέντρο. Ο αλγόριθμος MUSCLE σε σύγκριση με τον CLUSTALW επιτυγχάνει τον υψηλότερο, ή από κοινού υψηλότερο, βαθμό ακρίβειας και αποτελεί την ταχύτερη από τις δοκιμασμένες μεθόδους για μεγάλο αριθμό αλληλουχιών. [30]

3.4. Ομόλογη Μοντελοποίηση και Ελαχιστοποίηση Ενέργειας

Η ομόλογη μοντελοποίηση για την πρωτεΐνη Notch3 πραγματοποιήθηκε με Modeller. Η δομή της ανθρώπινης Notch1 EGFS δομής χρησιμοποιήθηκε ως πρότυπο για την παρούσα μελέτη (PDB entry: 2VJ3). [31] Περαιτέρω ελαχιστοποίηση έγινε χρησιμοποιώντας τα ενσωματωμένα στο Gromacs δυναμικά πεδία: AMBER, CHARMM27 και OPLS. Τα μοντέλα αξιολογήθηκαν δομικά χρησιμοποιώντας το βοηθητικό πρόγραμμα Procheck [32]. Οι ελαχιστοποιήσεις

ενέργειας χρησιμοποιήθηκαν για να αφαιρεθεί οποιαδήποτε γεωμετρική παραμόρφωση στο κάθε μοριακό σύστημα, χρησιμοποιώντας τα δυναμικά πεδία που αναφέρθηκαν όπως εφαρμόζονται στη σουίτα Gromacs, έκδοση 4.5.5. [33] Όλες οι προσομοιώσεις που σχετίζονται με το Gromacs πραγματοποιήθηκαν μέσω προηγούμενου γραφικού interface. [34] Μια έμμεση Γενικευμένη λύση (Generalized Born - GB) επιλέχθηκε σε αυτό το στάδιο, σε μια προσπάθεια να επιταχυνθεί η διαδικασία ελαχιστοποίησης της ενέργειας.

3.5. Προσομοιώσεις Μοριακής Δυναμικής

Μοριακά συστήματα της πρωτεΐνης Notch3 υποβλήθηκαν σε προσομοιώσεις Μοριακής Δυναμικής (Molecular Dynamics - MDs) χωρίς περιορισμούς χρησιμοποιώντας το Gromacs suite, έκδοση 4.5.5. [33] Τα MDs έλαβαν χώρα σε SPC νερό-διαλυτό, περιοδικό περιβάλλον. Τα μόρια του νερού προστέθηκαν χρησιμοποιώντας το πλαίσιο κόλουρο οκτάεδρο εκτεινόμενο 7Å από το κάθε άτομο. Τα μοριακά συστήματα εξουδετερώθηκαν με αντι-ιόντα, όπως απαιτείται. Για τους σκοπούς αυτής της μελέτης διενεργήθηκαν όλα τα MDs χρησιμοποιώντας το NVT σύνολο σε κανονικό περιβάλλον, σε 300 K, 1 atm και μέγεθος βήματος που ισούται με 2 femtoseconds, για συνολικό διάστημα προσομοίωσης 500 nanoseconds. Ένα σύνολο NVT απαιτεί ο αριθμός των ατόμων, ο όγκος και η θερμοκρασία να παραμένουν σταθερά κατά τη διάρκεια της προσομοίωσης.

Κεφάλαιο 4

Αποτελέσματα επεξεργασίας των ιατρικών δεδομένων

4.1. Μοτίβο επαναλήψεων αμινοξέων (Amino Acid Repeats – AAR pattern) - Στοιχισή αλληλουχιών

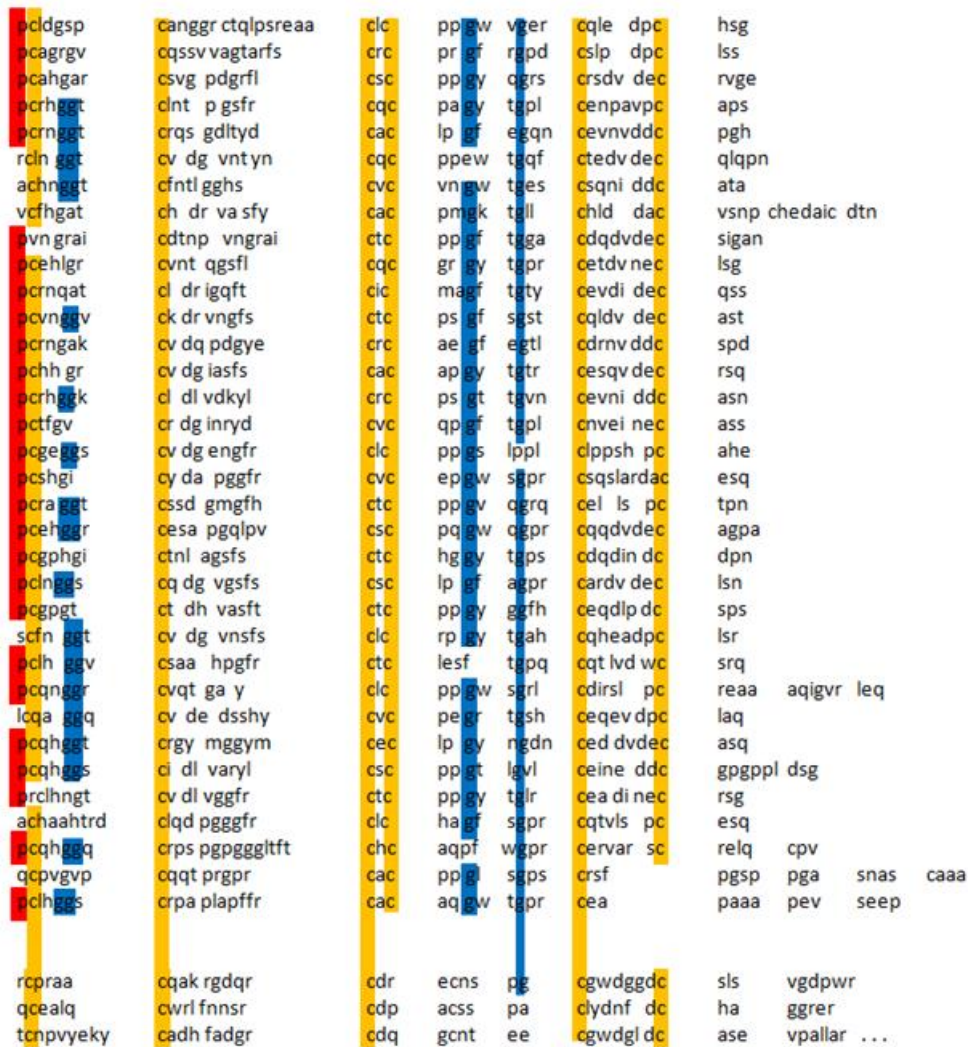
Οι επαναλήψεις των αμινοξέων (AAR) είναι άφθονες σε αλληλουχίες πρωτεϊνών και έχουν συγκεκριμένους ρόλους στη λειτουργία της πρωτεΐνης και την εξέλιξή της. Ως εκ τούτου, ο προσδιορισμός των AAR σε αλληλουχίες πρωτεϊνών είναι το πρώτο βήμα για την περαιτέρω διερεύνηση των βιολογικών τους λειτουργιών και του εξελικτικού τους μηχανισμού. [35]

Διαφορετικοί αλγόριθμοι που βασίζονται σε διαφορετικές στρατηγικές έχουν αναπτυχθεί για να αντιμετωπίσουν τα διαφορετικά μοτίβα επαναλήψεων. Στην περίπτωση αυτή, το μοτίβο αρχικά διαμορφώθηκε μέσω της παρατήρησης.

Η πρωτοταγής δομή (αλληλουχία αμινοξέων) της ανθρώπινης πρωτεΐνης Notch3, όπως λήφθηκε από τη βάση δεδομένων NCBI, αποτελείται από 2321 αμινοξέα. Μέσω της μεθόδου που αναπτύχθηκε στο προηγούμενο κεφάλαιο για την στοίχιση αλληλουχιών, προέκυψε ένα επαναλαμβανόμενο μοτίβο μεταξύ των 1461 αμινοξέων της ακολουθίας. Αυτά σχηματίζουν αρχικά 34 παρόμοιες επαναλήψεις και 3 περεταίρω, οι οποίες παρουσιάζουν μερική ομοιότητα. Από την τελευταία επανάληψη μέχρι το τέλος της ακολουθίας δεν σημειώνεται κανένα μοτίβο, ενώ ομοίως συμβαίνει και με τα πρώτα 41 αμινοξέα.

Η στοίχιση βασίστηκε στα αμινοξικά κατάλοιπα που παρουσιάζονται επισημασμένα στο σχήμα 4.1. Το επαναλαμβανόμενο μοτίβο είναι προφανές.

Στη συνέχεια, προκειμένου να κατασταθούν σαφέστερες οι συντηρημένες θέσεις των αμινοξικών καταλοίπων βάσει των οποίων σχηματίστηκε το συγκεκριμένο μοτίβο, επιδιώχθηκε μια οπτικοποίηση αυτών σε ένα συνοπτικό λογότυπο, όπως φαίνεται στο σχήμα 4.2. Στο λογότυπο αυτό φαίνεται σχηματικά το ποσοστό εμφάνισης του εκάστοτε καταλοίπου στη συγκεκριμένη θέση.



Σχήμα 4.1. Στοίχιση επαναλαμβανόμενου μοτίβου με επισημασμένα τα αμινοξικά κατάλοιπα στις συντηρημένες θέσεις. Με κίτρινο επισημαίνονται τα κατάλοιπα κυστεΐνης (C).

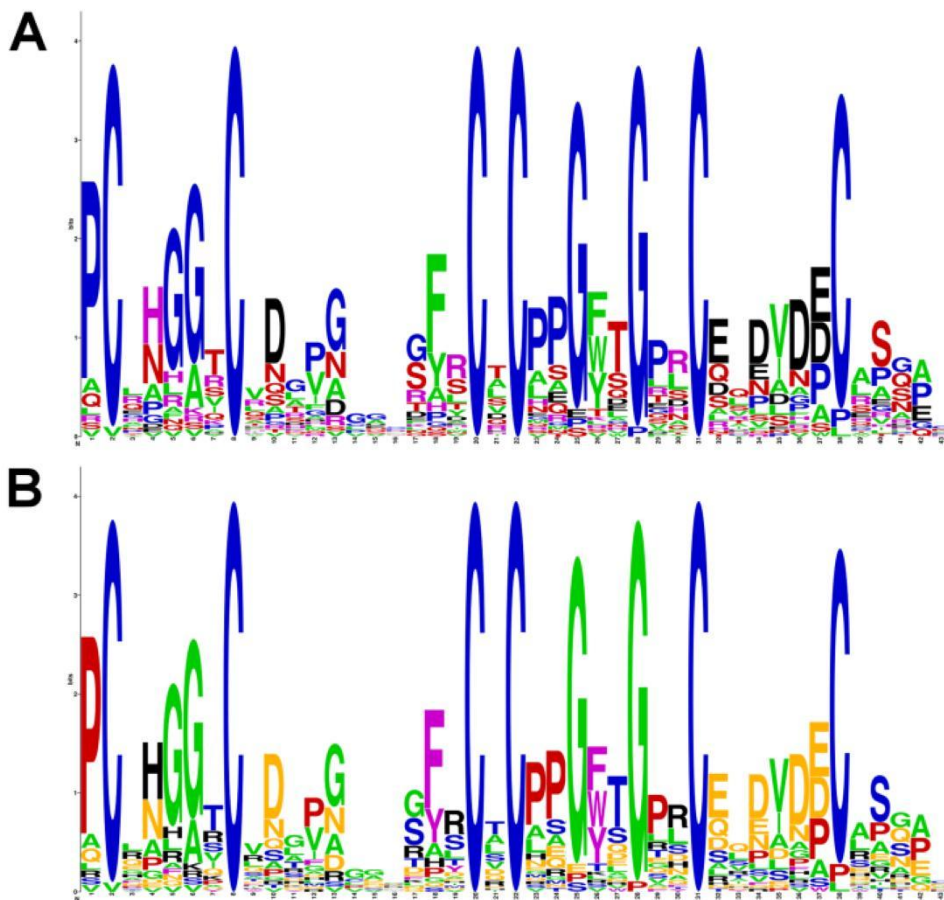


Σχήμα 4.2. Λογότυπο που αναπαριστά το επαναλαμβανόμενο μοτίβο με έμφαση στο ποσοστό εμφάνισης κάθε αμινοξικού καταλοίπου

4.1.1. Περιγραφή του μοτίβου

Όπως φαίνεται στο σχήμα 4.3.A (κατανομή χρωμάτων των αμινοξικών καταλοίπων με βάση την ταξινόμηση σύμφωνα με της δομή και χημικά χαρακτηριστικά) και στο σχήμα 4.3.B (κατανομή χρωμάτων των αμινοξικών καταλοίπων βασισμένη στις πλευρικές αλυσίδες, το ηλεκτρικό φορτίο και την υδροφοβικότητα), το επαναλαμβανόμενο μοτίβο αρχίζει με μια πολύ συντηρημένη προλίνη (θέση 1), που στη δευτεροβάθμια δομή εμφανίζεται σε διαμόρφωση υπερέλικας (coiled-coil) ακολουθούμενη από μία κυστεΐνη (Cys, θέση 2) σχεδόν συντηρημένη. Τα κατάλοιπα Cys είναι υπεύθυνα για το σχηματισμό των δισουλφιδικών δεσμών, που επηρεάζουν την σωστή αναδίπλωση και την βιολογική λειτουργία μιας πρωτεΐνης. Η υψηλή εμφάνιση των μεταλλάξεων στα Cys κατάλοιπα, τα οποία συντηρούνται ως επί το πλείστον, στην ακολουθία της Notch3 ευθύνονται για τη λανθασμένη αναδίπλωση και την απώλεια της λειτουργίας της πρωτεΐνης που μπορεί να οδηγήσει στη νόσο CADASIL. Η διαδοχή των δύο συνεχόμενων υψηλά συντηρημένων καταλοίπων γλυκίνης (Gly, θέσεις 5, 6) αποτελεί έναν ευέλικτο συνδέτη μεταξύ των δύο συντηρημένων καταλοίπων Cys (θέσεις 2 & 8) που παρέχουν σταθερότητα στην πρωτεϊνική δομή σε εκείνο το σημείο [36]. Η θέση του δεύτερου Gly (θέση 6) φαίνεται να είναι συντηρημένη από αλειφατικό. Τα κυριότερα αλειφατικά κατάλοιπα είναι η γλυκίνη και η αλανίνη (Ala). Ωστόσο, τα κατάλοιπα Ala δεν εμφανίζονται τόσο συχνά λόγω της ογκώδους πλευρικής αλυσίδας. Στις θέσεις 13-17, αν και δεν μπορούν να χαρακτηριστούν ως συντηρημένα, παρατηρούνται ως επί το πλείστον κατάλοιπα Gly, σχηματίζοντας και πάλι έναν ευέλικτο συνδέτη. Αξιοσημείωτη είναι η εμφάνιση των αρωματικών καταλοίπων (F, Y, W) (θέσεις 18 & 26) αμέσως μετά το Gly (θέσεις 17 & 25) πάνω στη β-πτυχωτή στη δευτεροταγή δομή της πρωτεΐνης. Τα αρωματικά κατάλοιπα παρέχουν ευνοϊκή ενέργεια αλληλεπίδρασης μέσω της σταυρωτής αντιστοίχισης με τη γλυκίνη, προκειμένου να μειωθεί η αποσταθεροποιητική επίδραση της γλυκίνης [37]. Οι θέσεις αυτές εμφανίζονται συντηρημένες και από υδροφοβικά κατάλοιπα (F, Y, A, W, Y). Το μοτίβο Cys-X-Cys (θέσεις 20-22) που ακολουθεί είναι εντελώς συντηρημένο και είναι μια ισχυρή αλυσίδα που είναι υπεύθυνη για την κατάλυση των οξειδοαναγωγικών αντιδράσεων [38]. Οι δύο συνεχόμενες προλίνες που ακολουθούν (θέσεις 23, 24), λόγω του κυκλικού χαρακτήρα τους, φαίνεται να βοηθούν στον σχηματισμό της β-στροφής στη δευτεροταγή δομή. Η ακολουθία αμινοξέων που ακολουθεί περιέχει δύο αρκετά συντηρημένα Gly (θέσεις 25, 28), που αποτελούν το μικρότερο αμινοξύ με την απλούστερη πλευρική αλυσίδα: ενός υδρογόνου. Μεταξύ αυτών των δύο Gly παρεμβάλλονται αρωματικά κατάλοιπα (F, Y, W) (θέση 26) που έχουν ογκώδεις πλευρικές αλυσίδες, και άλλα με πλευρικές αλυσίδες που περιέχουν υδροξύλιο

ή με όξινες (T, S, Q, D, E) (θέσεις 27) που είναι πιθανόν να συμμετάσχουν στους δεσμούς υδρογόνου (H-bonding). Η θέση 27 φαίνεται να είναι αρκετά συντηρημένη από αφόρτιστα κατάλοιπα (T, S, Q). Στις θέσεις 31-38, η ακολουθία ξεκινάει και τελειώνει με συντηρημένες cysteines με υδρόφιλα κατάλοιπα (E, D, N) στο μεταξύ τους, ενώ διακόπτεται μόνο από ένα αλειφατικό υδροφοβικό κατάλοιπο (V, I, A) στη θέση 35. Αυτή η περιοχή φαίνεται να εκτείνεται σε διαλύτη, λόγω του υδρόφιλου χαρακτήρα της. Τα κατάλοιπα Asp και Glu φαίνεται να αντικαθίσταται εξίσου το ένα από το άλλο σε αυτή την περιοχή, αφού είναι και τα δύο αρνητικά φορτισμένα. Η υψηλή περιεκτικότητα σε Gly και Pro στην πρωτεΐνη Notch3, καθώς και ο βαθμός συντήρησής τους στην ακολουθία, εξυπηρετούν κρίσιμο ρόλο για το σχηματισμό της πρωτεΐνης σε ελαστομερή ή αμυλοειδή ινίδια. Αυτό το εύρημα χάρει μεγάλης σημασίας αφού τα αμυλοειδή ινίδια είναι άρρηκτα συνδεδεμένα με ιστοεκφυλιστικές παθήσεις, όπως είναι και η CADASIL.



Σχήμα 4.3. Weblogo που αντιπροσωπεύουν τις συντηρημένες επαναλήψεις της πρωτεΐνης Notch3. (A) κατανομή χρωμάτων που βασίζεται στη δομή και τα χημικά χαρακτηριστικά των καταλοίπων (πράσινο: αλειφατικά, μπλε: περιέχουν υδροξύλιο ή θείο, κόκκινο: κυκλικά, μαύρο: βασικά, μωβ: αρωματικά, κίτρινο: όξινα & αμιδίου). (B) κατανομή χρωμάτων με βάση τις πλευρικές αλυσίδες, το φορτίο και την υδροφοβικότητα των καταλοίπων (μωβ: θετικά φορτισμένα, μαύρο: αρνητικά φορτισμένα, κόκκινο: αφόρτιστα, μπλε: ειδικές περιπτώσεις, πράσινο: υδρόφοβα)

4.2. Χαρτογράφηση των μεταλλάξεων στην πρωτοταγή και δευτεροταγή δομή της Notch3

Το σύνολο των αναφερόμενων μεταλλάξεων (βλ. Παράρτημα) στο γονίδιο NOTCH3 που λάβαμε από τον καθηγητή κ. Baumann ήταν 209, συμπεριλαμβανομένων και αυτών που δεν σχετίζονται με την ασθένεια CADASIL. Σημειώνοντας το σύνολο αυτών στην αλληλουχία της πρωτεΐνης (σχήμα 4.3), παρατηρήθηκε ότι η πλειοψηφία των μεταλλάξεων παρουσιάζεται σε συγκεκριμένες θέσεις (σε κατάλοιπα Cysteine, κυρίως). Συνεπώς, οι θέσεις αυτές μπορούν να θεωρηθούν στρατηγικές, αφού μετάλλαξή τους προκαλεί πιθανή μεταβολή στη λειτουργικότητα και τη δραστηριότητα της Notch3 και πιθανόν να ευθύνεται για την εμφάνιση της ασθένειας CADASIL.

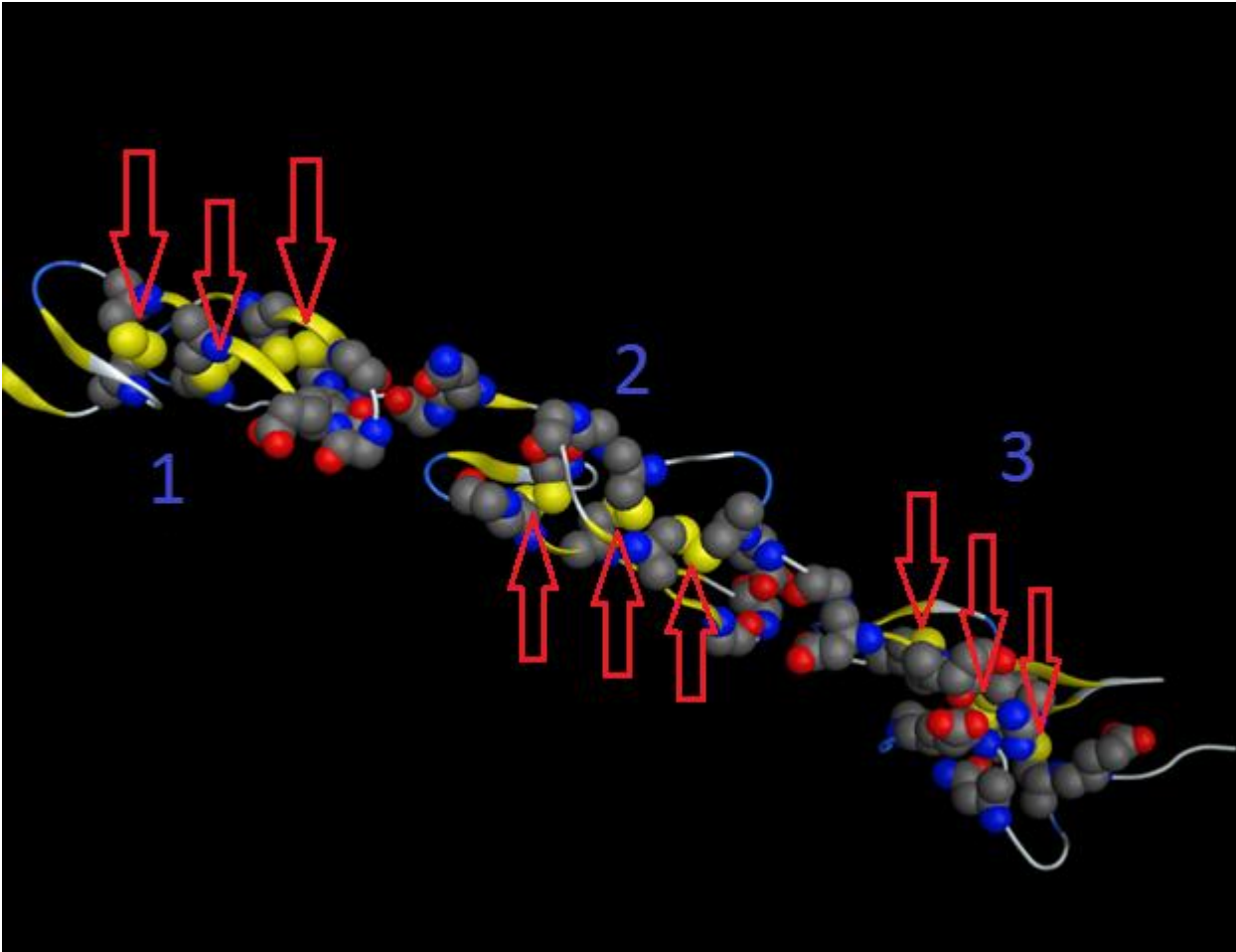
pcldgsp	cangr	ctqlpsreaa	clc	ppgw	vger	cqle dpc	hsg
pcagrsv	cqssv	vagtarfs	crc	prgf	rgpd	cslp dpc	lss
pcahgar	csvg	pdgrfl	csc	ppgy	qgrs	crsdvdec	rvge
pcrhggt	clnt	p gsfr	cqc	pagy	tgpl	cenpavpc	aps
pcrnngt	crqs	gdityd	cac	lpgf	egqn	cevnvddc	pgh
rclnngt	cv dg vnt yn		cac	ppew	tqaf	ctedvdec	qlqpn
achnggt	cfntl	gghs	cvc	vngw	tges	csqniddc	ata
vcfhgat	ch dr va sfy		cac	pmgk	tgll	chld dac	vsnp chedaic dtn
pvn grai	cdtn	pvngrai	ctc	ppgf	tgga	cdqdvdec	sigan
pcehlgr	cvnt	qgsfl	cqc	grgy	tgpr	cetdvdec	lsg
pcrnqat	cl dr igaft		clc	magf	tgty	cevdidec	qss
pcvnggv	ck dr vngfs		ctc	psgf	sgst	caqdvdec	ast
pcrngak	cv dq pdgye		crc	aegf	egtl	cdrnvddc	spd
pcchgr	cv dg iasfs		cac	apgy	tglr	cesqvdec	rsq
pcrhggk	cl dl vdkyl		crc	psgt	tgvn	cevniddc	asn
pcdfgv	cr dg inryd		cvc	qpgf	tgpl	cnveinec	ass
pcgeggs	cv dg engfr		clc	ppgs	lppi	clppshpc	ahc
pcshgi	cy da pggfr		cvc	epgw	sgpr	csqslardac	eq
pcraggt	cssd	gmgh	ctc	ppgv	qgrq	cel lspc	tpn
pcchggg	cesa	pgqlpv	csc	raqw	qgpq	q dec	agpa
pcgphgi	ctnl	agsfs	ctc	hggy	tgps	cdqindc	dpn
pclnngs	cq dg vgsfs		csc	lpgf	agpr	cardvdec	lsn
pcgpgt	ct dh vasft		ctc	ppgy	ggfh	ceqdlpdc	sps
scfnngt	cv dg vnsfs		clc	rpgy	tgah	cahea dpc	lsr
pcchggv	csaa	hpgfr	ctc	lesf	tgpq	cat lvdwc	srq
pcqnggr	cvqt	ga y	clc	ppgw	sgrl	cdirsl pc	reaa aqigvr leq
lcqaggq	cvde	dsshy	cvc	pegr	tgsh	ceqev dpc	laq
pcqhggt	crgy	mggym	cec	lpgy	ngdn	ced dvdec	asq
pcqhggg	cidl	varyl	csc	ppgt	lgvl	ceine ddc	gpgppl dsg
prclhngt	cvdl	vggfr	ctc	ppgy	tglr	cea dinec	rsg
achaahtrd	clqd	pgggr	clc	hagf	sgpr	catvls pc	esq
pcqhgga	crps	pgggltft	chc	aqpf	wgpr	cervar sc	relq cpv
qcrvnvp	cqqt	prgpr	cac	ppgl	sgps	crsf	pgsp pga snas caaa
pcchggg	crpa	plapffr	cac	aqgw	tgpr	cea	paaa pev seep

Σχήμα 4.4. Στοιχισμός επαναλαμβανόμενου μοτίβου με επισημασμένες τις θέσεις των μεταλλάξεων

Εν συνεχεία, πραγματοποιήθηκε χαρτογράφηση των μεταλλάξεων στη δευτεροταγή δομή της πρωτεΐνης (σχήμα 4.5). Οι θέσεις που απεικονίζονται με τη μορφή σφαιριδίων είναι οι θέσεις των αναφερόμενων μεταλλάξεων.

Όπως παρατηρείται, το επαναλαμβανόμενο μοτίβο είναι εμφανές και στη δευτεροταγή δομή, γεγονός που ενισχύει την προηγούμενη υπόθεση για τη σπουδαιότητα των συγκεκριμένων θέσεων στη λειτουργία της πρωτεΐνης.

Στο παρακάτω σχήμα, τα κόκκινα βέλη υποδεικνύουν 3 δισουλφιδικούς δεσμούς (κίτρινου χρώματος σφαιρίδια), οι οποίοι ακολουθούνται από 3 δεσμούς υδρογόνου (κόκκινου χρώματος σφαιρίδια). Το μοτίβο των μεταλλάξεων, όπως φαίνεται, εμφανίζεται 3 φορές.



Σχήμα 4.5. Επισήμανση των μεταλλάξεων στη δευτεροταγή 3D δομή της πρωτεΐνης. Τα κόκκινα βέλη υποδεικνύουν τους δισουλφιδικούς δεσμούς. Το θείο απεικονίζεται ως σφαιρίδια κίτρινου χρώματος, ενώ το υδρογόνο ως σφαιρίδια κόκκινου χρώματος.

4.3. 3D μοριακή μοντελοποίηση της δομής της πρωτεΐνης Notch3

Η περαιτέρω επεξεργασία των δεδομένων επικεντρώθηκε σκοπίμως σε 6 συγκεκριμένες μεταλλάξεις, οι οποίες φαίνεται να κατέχουν ιδιαίτερα στρατηγική θέση τόσο στην αλληλουχία όσο και στη δομή της πρωτεΐνης.

Οι μεταλλάξεις αυτές είναι οι παρακάτω:

1. Arg107Trp
2. Ser497Leu
3. Glu813Lys
4. Ala1020Pro
5. Ser978Arg
6. His1133Glu

Προκειμένου να κατανοήσουμε τις συνέπειες για την πρωτεΐνη Notch3 κατά την εισαγωγή των αλλαγών αμινοξέων της, πήραμε μια προσεγγιστική ομόλογη μοντελοποίηση, δεδομένου ότι δεν έχει προσδιοριστεί η δομή της συγκεκριμένης πρωτεΐνης. Χρησιμοποιώντας το Modeller, η Notch3 διαμορφώθηκε και προσδιορίστηκε η *in silico* 3D δομή της. Επιπλέον, το αποτέλεσμα από τις εισαγόμενες μεταλλάξεις θα μπορούσε να προβλεφθεί και μέσω προσομοιώσεων μοριακής δυναμικής χρησιμοποιώντας το GROMACS. Το Modeller προέβλεψε μια σειρά από δύο αντι-παράλληλα δίκτυα β-φύλλων τοποθετημένα σε επαναλήψεις σε όλη την πρωτεΐνη Notch3. Το ένα ζεύγος είναι βραχύ και αποτελείται από 5 έως 6 αμινοξικά κατάλοιπα, ενώ το άλλο είναι μεγαλύτερο καθώς αποτελείται από σχεδόν 10 έως 12 κατάλοιπα. Τα β-φύλλα συγκρατούνται μεταξύ τους με β-στροφές και τμήματα σπείρας (coil fragments) που περιέχουν κυρίως μη υδρόφοβα αμινοξέα. Μια σειρά από στρατηγικά τοποθετημένα κατάλοιπα κυστεΐνης είναι υπεύθυνα για τη δομική ακεραιότητα της Notch3 μέσω δισουλφιδικών γεφύρων.

Η μετάλλαξη **Arg107Trp** (R107W) προκάλεσε την μερική απώλεια της δομής ενός από τα β-φύλλα στο μακρύ ζεύγος της αντι-παράλληλης χωροδιαμόρφωσης της Notch3. Η θέση του καταλοίπου 107 ήταν πιθανότατα κρίσιμη για την σωστή διαμόρφωση της πρώτης αντιπαράλληλης δέσμης φύλλου-βρόχου-φύλλου (sheet-loop-sheet), αφού διαμορφώθηκε να βρίσκεται ακριβώς στη μέση του βρόχου σύνδεσης. Η ογκώδης και συζευγμένη πλευρική αλυσίδα της τρυπτοφάνης (Trp, W) ωθείται μακριά από τον πυρήνα της πρωτεΐνης. Ως εκ τούτου, βρέθηκε να διεκδικεί το διαθέσιμο χώρο στο ίδιο επίπεδο που ορίζουν τα δύο β-φύλλα. Ως αποτέλεσμα, η «ραχοκοκαλιά» της δομής (backbone) της πρωτεΐνης γύρω από τη θέση 107 συστρέφεται και μετατρέπεται σε σπείρα, έχοντας χάσει κάθε ίχνος της αρχικής επιπεδότητας β-φύλλου.

Η μετάλλαξη **Ser497Leu** (S497L) προβλεπόταν να οδηγήσει σε μερική απώλεια του πρώτου μακρού ζεύγους των σχηματισμών β-φύλλου, καθώς αλλάζει σημαντικά την αδόμητη διαμόρφωση έλικας, η οποία παρουσιάζεται σχεδόν κάθετη στο επίπεδο του β-φύλλου. Ως αποτέλεσμα το δίκτυο των δεσμών υδρογόνου στο κοντινό ζεύγος αντι-παράλληλων β-φύλλων έχει χαθεί, το διάκενο μεταξύ των δευτερευόντων στοιχείων γίνεται ελαφρώς μεγαλύτερο και

αναπόφευκτα ένα από τα β-φύλλα καταρρέει και χάνει περισσότερο από το ήμισυ της δομής του.

Η ίδια δομική αρχή ισχύει και για την μετάλλαξη **Glu813Lys** (E813K). Το αρνητικώς φορτισμένο κατάλοιπο γλουταμινικού οξέως (Glu, E) προβλέπεται να εγκαταστήσει ενεργά δεσμούς υδρογόνου μεταξύ των καταλοίπων που βρίσκονται σε στενή γειτνίαση σταθεροποιώντας το κοντινό ζεύγος β-φύλλων. Η μετάλλαξη E813K εισάγει ένα ογκώδες κατάλοιπο λυσίνης (Lys, K) στην αρχική θέση του γλουταμινικού οξέως που διαμορφώθηκε για τη θραύση της επαφής μεταξύ των δύο άνω β-φύλλων. Ο προφανής λόγος πίσω από αυτή την αλλαγή είναι η φυσικοχημική μεταβολή της πλευρικής αλυσίδας από αρνητικά σε θετικά φορτισμένη στην περίπτωση της λυσίνης και το γεγονός ότι η λυσίνη είναι ελαφρώς μεγαλύτερη σε μέγεθος από το κατάλοιπο γλουταμινικού οξέως. Συνεπώς, ο διαθέσιμος χώρος δεν είναι αρκετός για να φιλοξενήσει ένα αμινοξύ μεγαλύτερο από το γλουταμινικό οξύ.

Η μετάλλαξη **Ala1020Pro** (A1020P) οδηγεί σε σημαντική δομική απώλεια στην κοντινή διάταξη β-φύλλου της Notch3. Παρ'ότι το κατάλοιπο 1020P –προλίνη (Pro, P) στη θέση 1020- βρίσκεται σε δομική διαμόρφωση coil-coil, οι φ/ψ δίεδρες γωνίες των εμπλεκόμενων αμινοξέων είναι πολύ κοντά στις α-ελικοειδείς περιοχές του διαγράμματος Ramachandran. Η προλίνη δεν μπορεί να δημιουργήσει δεσμό υδρογόνου μέσω του αμινικού καταλοίπου της «ραχοκοκαλιάς» της, δεδομένου ότι δεν έχει αμίδιο του υδρογόνου. Ωστόσο, παρόλο που η προλίνη παρουσιάζει πολύ μικρή τάση για σχηματισμό α-έλικας, αρκετά συχνά εμφανίζεται ως το πρώτο ή το τελευταίο κατάλοιπο μιας α-έλικας, προφανώς λόγω της δομική ακαμψίας της αφού η πλευρική αλυσίδα της σχηματίζει δακτύλιο. Η αλανίνη (Ala, A) από την άλλη μεριά ανήκει στην ομάδα αμινοξέων των καταλοίπων με χαμηλή τάση σχηματισμού έλικας. Σε αυτή την περίπτωση φαίνεται ότι η προλίνη προσπαθεί να σταθεροποιήσει μια σχεδόν α-ελικοειδή τοπική διαμόρφωση για το μικρό τέντωμα του βρόχου ώστε να τοποθετηθεί επάνω. Συνεπώς, η διαρθρωτική αλλαγή είναι πάρα πολύ μεγάλη, με αποτέλεσμα τα κοντινά β-φύλλα να μην διατηρούνται και να καταρρέουν, χάνοντας έτσι το μεγαλύτερο μέρος της δομής τους.

Η μετάλλαξη **Ser978Arg** (S978R) οδηγεί επίσης σε απώλεια της δομής β-φύλλου της Notch3. Η 978R -αργινίνη (Arg, R) στη θέση 978- είχε προβλεφθεί να είναι εκτός του σχηματισμού του β-φύλλου αφού δεν υπάρχει διαθέσιμος χώρος για να φιλοξενηθεί. Ωστόσο, η μικροσκοπική πλευρική αλυσίδα του αρχικού καταλοίπου σερίνης (Ser, S) ήταν αρκετά μικρή για να επιτρέψει στο βρόχο να λυγίσει προς τα πίσω περαιτέρω υπέρ του σχηματισμού β-φύλλου. Ως αποτέλεσμα της μετάλλαξης του καταλοίπου S σε R, η γεωμετρία της φ/ψ δίεδρης γωνίας των κοντινών καταλοίπων αλλάζει και μερικά κατάλοιπα του πρώτου σχηματισμού β-φύλλου παίρνουν μέρος στο σχηματισμό δομής σπείρας (coil structure).

Η μετάλλαξη **His1133Glu** (H1133E) βρίσκεται σε διαμόρφωση β-στροφής που σταθεροποιεί το βραχύ ζεύγος των αντι-παράλληλων β-φύλλων. Το αρνητικά φορτισμένο Γλουταμινικό οξύ έλκεται προς το εσωτερικό τμήμα του πυρήνα της Notch3 μέσω νέων, τυχαία εγκατεστημένων δεσμών υδρογόνου σε κοντινά κατάλοιπα. Το θετικό φορτίο του ιμιδαζολικού δακτυλίου της ιστοιδίνης (His, H) ήταν τοποθετημένο σε ένα ουδέτερο ροταμερές που επέτρεψε στα δύο φύλλα βήτα να αλληλεπιδράσουν μέσω δεσμών υδρογόνου. Η κλίση της πλευρικής αλυσίδας

του γλουταμικού οξέος και η διαδοχική αναδιάταξη της «ραχοκοκαλίας» της πρωτεΐνης τοπικά, απαγορεύει το σχηματισμό ενός από τα δύο β-φύλλα εντελώς. Ως αποτέλεσμα, η πρωτεΐνη Notch3 καταλήγει πολύ αποδιοργανωμένη και τοπικά σχεδόν μετουσιώνεται καθώς χάνει εξ ολοκλήρου ένα ζωτικό στοιχείο της δευτεροταγούς δομής της.

4.4. Δομική ευθυγράμμιση και υπέρθεση της πρωτεΐνης αγρίου τύπου και της αντίστοιχης μεταλλαγμένης

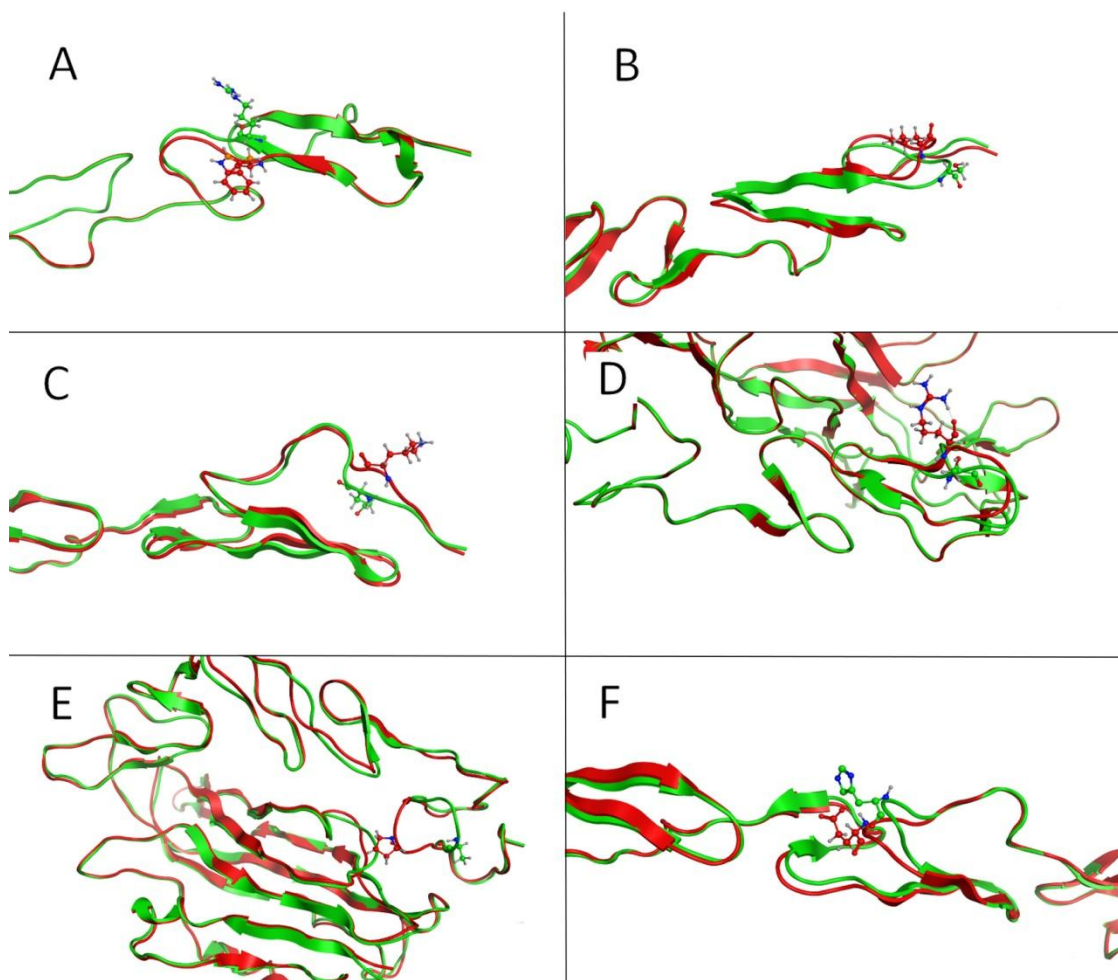
4.4.1. Εισαγωγή

Γενικά, τα αποτελέσματα που λαμβάνουμε από τη δομική ευθυγράμμιση των δύο δομών (αγρίου τύπου και μεταλλαγμένης πρωτεΐνης) είναι μια υπέρθεση των ατομικών συστημάτων συντεταγμένων και μια ελάχιστη μέση τετραγωνική απόκλιση (RMSD) μεταξύ των δομών. Η μέση τετραγωνική απόκλιση των δύο ευθυγραμμισμένων δομών υποδεικνύει τον βαθμό διαφορετικότητας της μίας από την άλλη. Η δομική ευθυγράμμιση, ωστόσο, μπορεί να περιπλέκεται, λόγω ύπαρξης πολλαπλών πρωτεϊνικών τομέων σε μία ή περισσότερες από τις δομές εισόδου. Αυτό συμβαίνει, γιατί πιθανές αλλαγές στον σχετικό προσανατολισμό των τομέων μεταξύ δύο δομών μπορεί να μεγαλώσουν τεχνητά την τιμή της RMSD.

Η βασικότερη πιθανή σύγκριση μεταξύ πρωτεϊνικών δομών δεν προσπαθεί να ευθυγραμμίσει τις δομές εισόδου και απαιτεί μια προ-υπολογισμένη ευθυγράμμιση ως είσοδο για να ορίσει ποιο από τα αμινοξικά κατάλοιπα της αλληλουχίας θα ληφθεί υπόψιν στον υπολογισμό της RMSD τιμής. Η δομική υπέρθεση συνήθως χρησιμοποιείται για να συγκριθούν πολλαπλές διαμορφώσεις της ίδιας πρωτεΐνης (όπως εδώ, αγρίου τύπου και μεταλλαγμένης), όπου η ευθυγράμμιση δεν κρίνεται αναγκαία, αφού οι αλληλουχίες είναι παρόμοιες. Επιπλέον, μπορεί να χρησιμοποιηθεί για να αξιολογηθεί η ποιότητα των ευθυγραμμίσεων που παράγονται χρησιμοποιώντας μόνο την πληροφορία της αλληλουχίας μεταξύ δύο ή περισσότερων πρωτεϊνών, των οποίων οι δομές παραμένουν άγνωστες. Αυτή η μέθοδος παραδοσιακά χρησιμοποιεί έναν απλό αλγόριθμο ελαχίστων τετραγώνων, ο οποίος ελαχιστοποιεί το άθροισμα όλων των τετραγωνισμένων αποστάσεων μεταξύ των δομών που λαμβάνουν μέρος στην υπέρθεση. Πρόσφατα, η ακρίβεια της μεθόδου έχει αυξηθεί με τη χρήση μεθόδων μέγιστης πιθανοφάνειας (maximum likelihood) και Μπαγесиανών (Bayesian) μεθόδων. [39, 40, 41, 42]

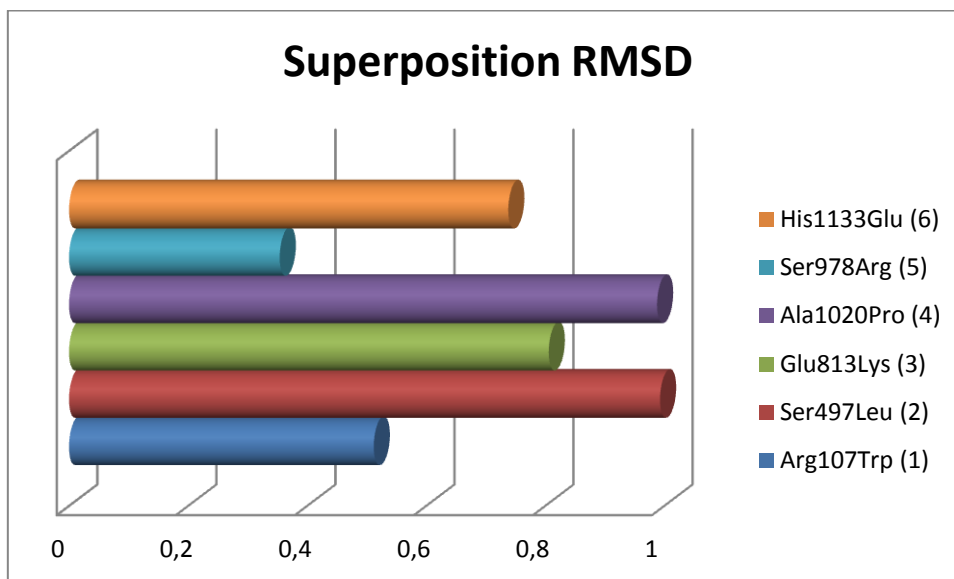
4.4.2. Παρουσίαση των αποτελεσμάτων σε επιλεγμένα μοντέλα

Ακολουθούν, οι εικόνες (4.6. A – F) που προέκυψαν από τη δομική υπέρθεση της πρωτεΐνης Notch3 αγρίου τύπου και της αντίστοιχης μεταλλαγμένης για κάθε μια από τις παραπάνω 6 μεταλλάξεις. Η πράσινη «αλυσίδα» αντιστοιχεί στην αγρίου τύπου, ενώ η κόκκινη στην αντίστοιχη μεταλλαγμένη.



Σχήμα 4.6. Δομική υπέρθεση της αγρίου τύπου και της μεταλλαγμένης πρωτεΐνης Notch3. Ο χρωματικός κώδικας είναι ο εξής: πράσινο – αγρίου τύπου, κόκκινο – μεταλλαγμένη. **(A)** Arg107Trp, **(B)** Ser497Leu, **(C)** Glu813Lys, **(D)** Ser978Arg, **(E)** Ala1020Pro, **(F)** His1133Glu.

Στην παρακάτω γραφική παράσταση (σχήμα 4.7), παρουσιάζονται γραφικά οι τιμές για τη μέση τετραγωνική απόκλιση (RMSD) που προέκυψαν από την υπέρθεση καθεμιάς από τις 6 μεταλλάξεις.



Σχήμα 4.7. Γραφική αναπαράσταση των τιμών για τη μέση τετραγωνική απόκλιση (RMSD) που προέκυψε από την υπέρθεση καθεμιας από τις μεταλλάξεις. Οι εν λόγω τιμές είναι οι ακόλουθες: (1) 0,512, (2) 0,993, (3) 0,806, (4) 0,988, (5) 0,354, (6) 0,738

4.5. Τα Πεδία Δυνάμεων (Force Fields)

4.5.1. Γενικά

Στη βιοπληροφορική, η γνώση για τη συμπεριφορά των μορίων είναι απαραίτητη. Κάθε μόριο χαρακτηρίζεται από επιμέρους χαρακτηριστικά, από τη συνολική δομή του, όλες τις γωνίες δεσμών μεταξύ των ατόμων, τις ατομικές ακτίνες και την κατανομή φορτίου, μέχρι την ικανότητά του να αλληλεπιδρά με άλλα μόρια ή χημικές ενώσεις. Οι αλληλεπιδράσεις μεταξύ των μορίων χαρακτηρίζονται από αρκετές αναλυτικές λειτουργίες, ζωτικής σημασίας για τις εφαρμογές των υπολογιστικών μοντέλων. Επί του παρόντος, ο χαρακτηρισμός της αναλυτικής δυναμικής ενεργειακής επιφάνειας είναι ένα από τα σημαντικότερα βήματα για την αποκάλυψη των μυστικών του νουκλεϊκού οξέος: η αλληλεπίδραση πρωτεϊνών, η πρόσδεση του συνδέτη με πρωτεΐνες, οι ενζυματικές αντιδράσεις, ο χαρακτηρισμός των αλληλεπιδράσεων μεταξύ των μορίων ή η συμπεριφορά των πρωτεϊνών σε διαλύτες. Οι ιδιότητες μεταξύ δύο μορίων είναι το βασικότερο πρόβλημα. Οι κύριοι ρόλοι του πεδίου δυνάμεων είναι να περιγράψει εύλογα τις γενικές ιδιότητες των μορίων, από πλευράς μοριακής γεωμετρίας, διαμορφωτικής (conformational) και στερεοϊσομετρικής ενέργειας, στρεπτικών εμποδίων (torsional barriers), στρεπτικής παραμόρφωσης και ενέργειας των διαμοριακών αλληλεπιδράσεων, να αξιολογήσει τη γεωμετρία μεταξύ των αλληλεπιδρώντων μορίων και να εκτιμήσει την συχνότητα δόνησης και τη θερμότητα σχηματισμού. Μια προσέγγιση για να αποκαλυφθούν όλες αυτές οι ιδιότητες είναι να χρησιμοποιηθούν προσομοιώσεις ή πειράματα σε ηλεκτρονικό υπολογιστή, τα οποία παρέχουν αποκλειστική εικόνα σχετικά με την αλληλεπίδραση της ζωής σε μοριακό επίπεδο. Σε γενικές γραμμές, υπάρχουν δύο τύποι προσομοιώσεων σε υπολογιστή: (i) Monte Carlo, όπου πολλά μόρια ή ιόντα βρίσκονται σε ένα κουτί και ένα τυχαίο σωματίδιο κινείται σε διαφορετικές τυχαίες

θέσεις (Allen & Tildesley, 1987), και (ii) η μοριακή δυναμική MD, η οποία παρέχει πληροφορίες σχετικά με την πραγματική κίνηση των μορίων με βάση το πεδίο δυνάμεων. Εδώ, ο υπολογιστής υπολογίζει τη δύναμη του κάθε μορίου και στη συνέχεια προσδιορίζει την κίνηση του μορίου ως ανταπόκριση του πεδίου. (Leach, 2001). Ακόμη και αν γενική κατήχηση του πεδίου δυνάμεων δεν υπάρχει, μπορούμε να πούμε ότι το πεδίο δυνάμεων είναι ο λειτουργικός συνδυασμός ενός μαθηματικού τύπου και των συναφών παραμέτρων που οδηγεί σε συνάρτηση της αναλυτικής δυναμικής ενέργειας, δηλαδή της ενέργειας της πρωτεΐνης ως μια συνάρτηση των ατομικών παραμέτρων της.

Το πεδίο δυνάμεων αντιπροσωπεύει δύο ομάδες ιδιοτήτων της μοριακής δομής: (i) τις δεσμικές αλληλεπιδράσεις (bonded interactions), οι οποίες χαρακτηρίζουν το μήκος των δεσμών, την κάμψη των γωνιών σθένους, και την περιστροφή των δίδρων και (ii) τις μη-δεσμικές αλληλεπιδράσεις (nonbonded interactions) που αξιολογούν ηλεκτροστατικά δεδομένα, την απαγορευτική αρχή του Pauli, και τη διασπορά. Οι παράμετροι για τον υπολογισμό του πεδίου δυνάμεων μπορούν να ληφθούν από ab initio ημιεμπειρικούς κβαντομηχανικούς υπολογισμούς ή από τις πειραματικές μεθόδους περίθλασης ακτίνων χ και ηλεκτρονίων και φασματοσκοπία NMR ή Ράμαν και νευρώνων. Παρά τα αρκετά πεδία δυνάμεων των πρωτεϊνών, τα περισσότερα από αυτά χρησιμοποιούν απλή συνάρτηση δυναμικής ενέργειας:

(i) Για δεσμικές αλληλεπιδράσεις:

$$V(r) = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi [\cos(n\phi + \delta) + 1] \\ + \sum_{\text{unbonded pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right]$$

(ii) Για μη-δεσμικές αλληλεπιδράσεις:

$$V(r) = \sum_{\text{nonbonded pairs}} \left(\epsilon_{ij} \left[\left(\frac{R_{\min, ij}}{r_{ij}} \right)^{12} - 2 \times \left(\frac{R_{\min, ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right)$$

όπου b είναι η διατομική απόσταση, k_b και b_0 οι παράγοντες ακαμψίας και ισορροπίας του μήκους του δεσμού, το θ χαρακτηρίζει τη γωνία που σχηματίζεται από τα διανύσματα των δύο δεσμών, και τα μεγέθη θ_0 και k_θ εκτιμούν την ακαμψία και την ισορροπία της γεωμετρίας της γωνίας. Το στρεπτικό δυναμικό στην ισορροπία χαρακτηρίζεται από συνάρτηση συνημιτόνου, όπου f είναι η στρεπτική γωνία, d είναι η φάση και n είναι το δυναμικό του δίδρου. Το τελευταίο μέρος της εξίσωσης περιλαμβάνει την ηλεκτροστατική αλληλεπίδραση, όπου ο παράγοντας ϵ_{ij} είναι μια παράμετρος που βασίζεται στα δύο άτομα που αλληλεπιδρούν i και j , q_i και q_j περιγράφοντας το πραγματικό φορτίο στα i και j άτομα, και $R_{\min, ij}$ ορίζοντας την απόσταση στην οποία η ενέργεια της εξίσωσης Lennard-Jones είναι στο ελάχιστο. Αυτό το τμήμα περιλαμβάνει επίσης την κάμψη των γωνιών μεταξύ των ατόμων A και C (Guvench & MacKerell, 2008).

Η συνολική ενέργεια από τις δεσμικές και μη-δεσμικές αλληλεπιδράσεις υπολογίζεται από τον τύπο:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} + E_{\text{other}}$$

όπου E_{other} περιλαμβάνει τις απωθητικές δυνάμεις, τις δυνάμεις van der Waals, καθώς και τη δύναμη Coulomb.

Τα πιο δημοφιλή πεδία δυνάμεων είναι τα: AMBER, DREIDING, CHARMM, UFF, Groms, OPLS και COMPASS, τα οποία διαφέρουν σε διάφορες ιδιότητες. Για παράδειγμα, τα DREIDING και UFF περιέχουν παραμέτρους όλων των ατόμων του περιοδικού πίνακα, τα CHARMM, AMBER, και Groms χρησιμοποιούνται κυρίως για τις προσομοιώσεις των βιομορίων, ενώ η συμπυκνωμένη ύλη ανιχνεύεται από τα OPLS και COMPASS.

Στα πλαίσια αυτής της διπλωματικής χρησιμοποιήθηκαν τα πεδία δυνάμεων: AMBER, CHARMM και OPLS, τα οποία περιγράφονται αναλυτικά παρακάτω.

4.5.2 Τα πεδία δυναμικού AMBER, CHARMM και OPLS

Το **AMBER** (Assisted Model Building with Energy Refinement) περιλαμβάνει διάφορα πεδία δυνάμεων κλασικής μοριακής μηχανικής και περιγράφει τις δομικές και δυναμικές ιδιότητες των βιομορίων, κανονικών και μη-κανονικών νουκλεϊκών οξέων και πρωτεϊνών στο νερό, καθώς επίσης και πολύπλοκες δομικές αλλαγές, όπως την μετάβαση A→B σε DNA διπλής και τριπλής έλικας και την προσομοίωση του DNA σε ακραίες συνθήκες (Cheatham & Kollman, 1996; Miller & Kollman, 1997; Rueda, Kalko, Luque, και Orozco, 2003). Η πιο σημαντική διαφορά μεταξύ του AMBER και των άλλων πεδίων δυνάμεων είναι η ενιαία ατομική αναπαράσταση, πράγμα που σημαίνει ότι μια αναπαράσταση όλων των ατόμων σε μη πολικά άτομα υδρογόνου δεν εκπροσωπείται ρητά, αλλά συγχωνεύεται στην περιγραφή των βαρέων ατόμων με τα οποία είναι συνδεδεμένα. Επιπλέον, το AMBER περιλαμβάνει έναν όρο δεσμού-υδρογόνου που αυξάνει το μέγεθος της ενέργειας του δεσμού-υδρογόνου που προέρχεται από την αλληλεπίδραση διπόλου-διπόλου των ομάδων δότη και δέκτη. Δεύτερη διαφοροποίηση είναι η χρήση των γενικών στρεπτικών παραμέτρων, οι οποίες εξαρτώνται αποκλειστικά από τους τύπους των δύο ατόμων που περιλαμβάνουν το κεντρικό δεσμό και όχι από τους τύπους των τερματικών ατόμων όπως στα άλλα πεδία δυνάμεων. Οι όροι στρέψης είναι ακατάλληλοι για να διατηρηθεί η στερεοχημεία στα στροφικά κέντρα (Leach, 2001).

Η εξίσωση του πεδίου δυνάμεων AMBER είναι πολύ παρόμοια με τη γενική εξίσωση του πεδίου δυνάμεων και περιλαμβάνει όλες τις ιδιότητες, τους δεσμούς, τις γωνίες, τις στρέψεις, τις ηλεκτροστατικές αλληλεπιδράσεις και τις δυνάμεις van der Waals:

$$V = \sum_{\text{bonds}} K_2 (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsion}} \frac{V_n}{2} (1 + \cos [n\phi - \gamma]) + \sum_{\text{nonbonded}} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum_{\text{nonbonded}} \frac{q_i q_j}{\epsilon R_{ij}}$$

Το AMBER94, που χρησιμοποιήθηκε στα πλαίσια αυτής της εργασίας, είναι ένα απλό πεδίο δυνάμεων που βασίζεται στο πεδίο Weiner το οποίο χρησιμοποιείται για τη μελέτη των οργανικών μορίων, των πεπτιδίων και των νουκλεϊκών οξέων. Το πεδίο δυνάμεων παραμετροποιείται με απλές ενώσεις όπως το αιθάνιο, αλλά για ειδικές παραμέτρους, θα πρέπει να προστεθούν πιο πολύπλοκα σωματίδια. Το AMBER94 έχει επίσης “helix-friendly” χαρακτήρα.

Το πεδίο δυνάμεων **CHARMM** (Chemistry at HARvard Macromolecular Mechanics) χρησιμοποιείται για να εκτελεί τον υπολογισμό της αλληλεπίδρασης και των ενεργειών διαμόρφωσης, των τοπικών ελαχίστων, της γεωμετρίας των σωματιδίων, της εξαρτώμενης από το χρόνο δυναμικής συμπεριφοράς, των εμποδίων περιστροφής, των συχνοτήτων δόνησης και της ελεύθερης ενέργειας του ευρέος φάσματος ή των σωματιδίων, από τα μικρά μόρια μέχρι τα μεγάλα βιολογικά μακρομόρια, συμπεριλαμβανομένων των λιπιδίων και των υδατανθράκων. Το CHARMM χρησιμοποιεί δύο προσεγγίσεις για τη μοριακή προσομοίωση και μοντελοποίηση, (i) των κλασικών ενεργειακών συναρτήσεων, συμπεριλαμβανομένων των εμπειρικών και ημιεμπειρικών συναρτήσεων και (ii) των συναρτήσεων κβαντομηχανικής ενέργειας, οι οποίες είναι ημιεμπειρικές ή *ab initio*.

Η παραμετροποίηση του CHARMM επιτρέπει τη μελέτη οργανικών ενώσεων φαρμακευτικού ενδιαφέροντος καθώς και τη χρήση εναλλακτικών μεθόδων θεραπείας των μη-δεσμικών αλληλεπιδράσεων. Αυτές οι μέθοδοι ομαλοποιούν τη μεταβολή στην ενέργεια και στο δυναμικό στην απόσταση αποκοπής με σκοπό τη μείωση των σφαλμάτων στην μελετηθείσα περιοχή. Μπορούν να χρησιμοποιηθούν σε ηλεκτροστατικές και Lennard-Jones αλληλεπιδράσεις κατά την μεταφορά και την μεταγωγή ενέργειας (Brooks et al., 1983). Η σημασία των μεθόδων μεταφοράς και μεταγωγής έγκυται στον καθορισμό των τροχιών των ατόμων κυρίως για τα πολύ φορτισμένα βιομόρια.

Το CHARMM22 χρησιμοποιεί μία ευέλικτη και ευρείας κλίμακας συνάρτηση:

$$\begin{aligned}
 V = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi - \delta)] \\
 & + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_\mu (\mu - \mu_0)^2 \\
 & + \sum_{\text{nonbonded}} \in \left(\left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - 2 \times \left(\frac{R_{\min,ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{s r_{ij}} \right)
 \end{aligned}$$

όπου το k_b είναι μια σταθερά για τη δύναμη του δεσμού, το $(b - b_0)$ δηλώνει τη μορφή ισορροπίας όπου το άτομο έχει μετακινηθεί, k_θ είναι η σταθερά δύναμης της γωνίας, η τιμή του $(\theta - \theta_0)$ προσδιορίζει τη γωνία από την ισορροπία μεταξύ των τριών συνδεδεμένων ατόμων, k_ω είναι η σταθερά της δύναμης στρέψης, n είναι η πολυπλοκότητα της συνάρτησης, ϕ είναι η γωνία στρέψης, και δ είναι η μετατόπιση φάσης. Ο τέταρτος όρος της εξίσωσης χαρακτηρίζει το καταχρηστικό κλάσμα που είναι έξω από το επίπεδο κάμψης. Εδώ, k_ω είναι η σταθερά δύναμης και τα $\omega - \omega_0$

ορίζουν την γωνία έξω από το επίπεδο. Στο τμήμα Urey-Bradley (όρος που υπολογίζει την κάμψη της γωνίας χρησιμοποιώντας 1,3-μη δεσμικές αλληλεπιδράσεις), k_{μ} είναι η αντίστοιχη σταθερά δύναμης και το μ - $\mu 0$ είναι η απόσταση μεταξύ των 1,3-ατόμων στο αρμονικό δυναμικό. Στο τελευταίο τμήμα, ο όρος των μη δεσμικών αλληλεπιδράσεων μεταξύ των ζευγών των ατόμων περιγράφηκε στο προηγούμενο κείμενο. Για τον ενδελεχή χαρακτηρισμό των ενδομοριακών και διαμοριακών ενεργειών, η διηλεκτρική σταθερά ϵ έχει οριστεί ίση με τη μονάδα σε αυτή τη συνάρτηση δυναμικής ενέργειας, κυρίως για τις ομάδες πεπτιδίων (MacKerell et al., 1998).

Το πεδίο δυνάμεων **OPLS-AA** χρησιμοποιείται για όλα τα άτομα. Κυρίως βρίσκει εφαρμογή σε πρωτεΐνες και μικρά οργανικά μόρια. Τα μερικά φορτία βασίζονται σε υποδιαιρέσεις των φορτίων των δεσμών οι οποίες αναπαράγουν τα αρχικά φορτία. Τα πολικά υδρογόνα έχουν μηδενικές ακτίνες van der Waals. [43]

4.6. Μοριακή δυναμική

4.6.1. Εισαγωγή

Η προσομοίωση της μοριακής δυναμικής αντιμετωπίζει το πρόβλημα της κλασικής αριθμητικής επίλυσης των εξισώσεων της κίνησης για ένα σύστημα N ατόμων σε μία προσπάθεια να μελετηθεί ένα θερμοδυναμικό σύνολο ή τροχιά, υπό καθορισμένες θερμοδυναμικές συνθήκες (π.χ. σταθερή θερμοκρασία ή σταθερή πίεση). Μια τέτοια τροχιά είναι σημαντική για δύο λόγους: Πρώτον, παρέχει πληροφορίες για τη διαμόρφωση ή τη δομή κάθε ατόμου, από το οποίο μπορούν στη συνέχεια να προσδιοριστούν οι θερμοδυναμικές ιδιότητες του συστήματος. Δεύτερον, η τροχιά αντιπροσωπεύει μια εξερεύνηση ή αναζήτηση του χώρου διαμόρφωσης που είναι διαθέσιμος για ένα συγκεκριμένο σύστημα. Η αρχή που διέπει την αναζήτηση διαμόρφωσης χρησιμοποιώντας δυναμική είναι ότι τα άτομα σε μια προσομοίωση δυναμικής θα ψάξουν τελικά ολόκληρο το χώρο διαμόρφωσής τους. Το προφανές πρόβλημα με αυτή την προσέγγιση είναι ότι ο χώρος αναζήτησης μπορεί να είναι τεράστιος και ως εκ τούτου μπορεί να χρειαστούν πολλές χιλιάδες, εκατομμύρια ή ακόμη και δισεκατομμύρια βήματα για την επαρκή και ακριβή μελέτη του χώρου. Κατά συνέπεια, πολλές ερευνητικές προσπάθειες έχουν δαπανηθεί για την εκπόνηση μεθόδων με σκοπό την επιτάχυνση της έρευνας, όχι όμως εις βάρος της ακρίβειας. [44, 45]

4.6.2. Θεωρία και Μαθηματική προσέγγιση

4.6.2.1. Γενικά

Η προσομοίωση των μοριακών συστημάτων περιγράφεται καλά στο [Allen 1987] και είναι πέρα από το πεδίο αυτού του εγγράφου η ανάλυση όλης της έρευνας πάνω στη μοριακή δυναμική. Αντ' αυτού, χρησιμότερη θα ήταν μια πιο λακωνική περιγραφή των μεθόδων που χρησιμοποιούνται για την επίλυση των εξισώσεων κίνησης. Στις περιγραφές των μεθόδων, θα γίνει χρήση των σημειογραφιών που φαίνονται στον πίνακα 4.1.:

Σύμβολο	Περιγραφή
g	Βαθμοί ελευθερίας του ατομικού συστήματος
k	Σταθερά Boltzmann
K	Κινητική ενέργεια των ατόμων: $2K = \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}$
\mathbf{M}	Διαγώνια μήτρα ατομικής μάζας
Q_T	Θερμοστατικής παράμετρος μάζας
Q_P	Βαροστατική παράμετρος μάζας
P	Το άθροισμα των καταστάσεων της μακροσκοπικής πίεσης του συστήματος
\mathbf{p}	Οι ορμές των ατόμων στον πραγματικό χώρο. Η επιτάχυνσή είναι: $\mathbf{M}^{-1} \mathbf{p}$
\mathbf{q}	Οι συντεταγμένες των ατόμων σε χώρο κλίμακας (scale space): $\mathbf{q} = V^{-1/3} \mathbf{r}$
\mathbf{r}	Οι συντεταγμένες των ατόμων στον πραγματικό χώρο: $\mathbf{r} = V^{1/3} \mathbf{q}$
T	Το άθροισμα των καταστάσεων της μακροσκοπικής θερμοκρασίας του συστήματος
\mathbf{u}	Οι ορμές των ατόμων σε χώρο κλίμακας (scale space) (συζυγής του \mathbf{q})
U	Ατομική δυναμική ενέργεια, συνάρτηση του \mathbf{r}
V	Στιγμιαίος όγκος του συστήματος
W	Θεώρημα Virial; $W = \mathbf{r}^T \nabla U(\mathbf{r})$

Πίνακας 4.1. Σημειογραφία που χρησιμοποιείται στις μεθόδους μοριακής δυναμικής

Οι εξισώσεις κίνησης Nosé-Poincaré-Andersen (NPA) [Sturgeon 2002] [Bond 1999] προκύπτουν από την Χαμιλτονιανή NPT (σταθερός αριθμός ατόμων, σταθερή πίεση, σταθερή θερμοκρασία) συνόλου. Επομένως, μπορούν να προσομοιωθούν και τα θεωρητικά σωστά στατιστικά σύνολα NVE (σταθερός αριθμός ατόμων, σταθερός όγκος, σταθερή ολική ενέργεια), NVT (σταθερός αριθμός ατόμων, σταθερός όγκος, σταθερή θερμοκρασία), NPH (σταθερός αριθμός ατόμων, σταθερή πίεση, σταθερή ενθαλπία) και NPT. Επιπλέον, οι συμπλεκτικοί ολοκληρωτές φαίνονται πιο σταθεροί με Χαμιλτονιανά συστήματα από ότι με μη-Χαμιλτονιανά. Η NPA Χαμιλτονιανή είναι:

$$H(t) = s[\tilde{H}(t) - \tilde{H}(0)]$$

$$\tilde{H}(t) = \frac{\mathbf{u} \cdot \mathbf{M}^{-1} \mathbf{u}}{2s^2 V^{2/3}} + \frac{u^2}{2Q_T} + \frac{v^2}{2Q_P} + U(V^{1/3} \mathbf{q}) + gkT \log s + PV$$

Όπου g είναι ο βαθμός ελευθερίας του ατομικού συστήματος συν 1 αν το V δεν είναι σταθερό. Η συντεταγμένη s είναι μια συντεταγμένη της κλίμακας του χρόνου που χρησιμοποιείται για να ενισχύσει τη σταθερή θερμοκρασία. Η ποσότητα H είναι συντηρημένη. Οι εξισώσεις κίνησης μπορούν να προκύψουν από διαφορίση του H , λαμβάνοντας υπόψιν τις συζυγείς συντεταγμένες και τις ορμές, όπως φαίνεται παρακάτω:

$$\begin{aligned} \dot{\mathbf{q}} &= s^{-1} V^{-2/3} \mathbf{M}^{-1} \mathbf{u} & \dot{\mathbf{u}} &= -s V^{-1/3} \nabla U(\mathbf{r}) & \mathbf{r} &= V^{1/3} \mathbf{q} \\ \dot{s} &= su/Q_T & \dot{u} &= 2K - gkT - \tilde{H}(t) + \tilde{H}(0) & 2K &= \mathbf{u} \cdot \mathbf{M}^{-1} \mathbf{u} s^{-2} V^{-2/3} \\ \dot{V} &= sv/Q_P & \dot{v} &= s [(2K - W)/3V - P] & W &= \mathbf{r} \cdot \nabla U(\mathbf{r}) \end{aligned}$$

Για να επιλύσουμε αυτές τις εξισώσεις, ο καταλληλότερος τρόπος είναι ο συμπλεκτικός, χρόνο-αντιστρέψιμος Γενικευμένος Αλγόριθμος Leapfrog [Sun 1993]. Μια Χαμιλτονιανή $H(\mathbf{p}, \mathbf{q})$, όπου το \mathbf{q} είναι οι συντεταγμένες και το \mathbf{p} η συζυγής ορμή σε χρόνο t , μπορεί να ολοκληρωθεί με βήμα h ως εξής:

$$\begin{aligned} \mathbf{p}_{t+h/2} &= \mathbf{p}_t - \frac{h}{2} \frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_{t+h/2}, \mathbf{q}_t) \\ \mathbf{q}_{t+h} &= \mathbf{q}_t + \frac{h}{2} \left[\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_{t+h/2}, \mathbf{q}_t) + \frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_{t+h/2}, \mathbf{q}_{t+h}) \right] \\ \mathbf{p}_{t+h} &= \mathbf{p}_{t+h/2} - \frac{h}{2} \frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_{t+h/2}, \mathbf{q}_{t+h}) \end{aligned}$$

Που αποτελεί μια συνένωση της Συμπλεκτικής μεθόδου Euler και της συζυγούς της (αντιστρόφως ανάλογη του χρόνου). [44]

4.6.2.2. Οι εξισώσεις NHA

Οι εξισώσεις κίνησης Nosé-Hoover-Andersen (NHA) δίνονται στο [Hoover 1985] και έχουν στόχο την ανίχνευση του NPT συνόλου. Ωστόσο, κρατώντας σταθερά ορισμένα μεγέθη, αυτές οι εξισώσεις μπορούν να παράγουν οποιαδήποτε από τα NVE, NVT, NPH και NPT σύνολα. Οι εξισώσεις κίνησης Hoover είναι:

$$\begin{aligned} \dot{\mathbf{q}} &= V^{-1/3} \mathbf{M}^{-1} \mathbf{p} & \dot{\mathbf{p}} &= -\nabla U(\mathbf{r}) - (\dot{\chi} + \dot{\eta}) \mathbf{p} & \dot{\mathbf{r}} &= V^{1/3} \mathbf{q} \\ & & \dot{\eta} &= (2K - gkT) Q_T^{-1} & 2K &= \mathbf{p} \cdot \mathbf{M}^{-1} \mathbf{p} \\ \dot{V} &= 3V \dot{\chi} & \dot{\chi} &= [(2K - W)/3 - PV] Q_P^{-1} & W &= \mathbf{r} \cdot \nabla U(\mathbf{r}) \end{aligned}$$

Όπου το g είναι ο αριθμός των ατομικών βαθμών ελευθερίας συν 1 εάν η πίεση διατηρείται σταθερή. Η χρονική συντεταγμένη δεν είναι ομοιογενής στις εξισώσεις κίνησης NHA. Δηλαδή, οι ποσότητες που μετρώνται σε ομοιόμορφα διαστήματα κατά τη διάρκεια της αριθμητικής ολοκλήρωσης δεν αντιστοιχούν σε δείγματα πραγματικού χρόνου, αλλά σε δειγματοληψία σε κάποια χρονική κλίμακα. Στις περισσότερες περιπτώσεις, αυτή η μη-ομοιομορφία καθιστά αναγκαία την εξαγωγή ενός μέσου όρου για αρκετά μεγάλες προσομοιώσεις. Αν και δεν είναι Χαμιλτονιανές, οι εξισώσεις NHA διατηρούν την ακόλουθη ποσότητα:

$$H(t) = \frac{1}{2} \mathbf{p} \cdot \mathbf{M}^{-1} \mathbf{p} + \frac{1}{2} Q_T \dot{\eta}^2 + \frac{1}{2} Q_P \dot{V}^2 + U(V^{1/3} \mathbf{q}) + gkT\eta + PV$$

Οι εξισώσεις NHA μπορούν να επιλυθούν με τεχνικές χωρισμού τελεστών που διαδίδουν η και χ ορμές στα βήματα εξωτερικής διάδοσης (κρατώντας σταθερές τις υπόλοιπες ποσότητες) και έναν Γενικευμένο ολοκληρωτή Leapfrog για το Χαμιλτονιανό υπόλοιπο. [44]

4.6.2.3. Οι εξισώσεις BER

Οι εξισώσεις Berendsen (BER) είναι NVE εξισώσεις κίνησης επαυξημένες με τον θερμοστάτη και βαροστάτη Berendsen. Οι εξισώσεις κίνησης και η αντίστοιχη Χαμιλτονιανή είναι:

$$\dot{\mathbf{r}} = \mathbf{M}^{-1}\mathbf{p} \quad \dot{\mathbf{p}} = -\nabla U(\mathbf{r}) \quad H(t) = \frac{1}{2}\mathbf{p} \cdot \mathbf{M}^{-1}\mathbf{p} + U(\mathbf{r})$$

Σε κάθε χρονικό βήμα, οι ορμές και οι συντεταγμένες υπολογίζονται υπό κλίμακα σύμφωνα με τη διαδικασία Berendsen [Berendsen 1984] για την επιβολή σταθερής θερμοκρασίας και/ ή πίεσης.

$$\dot{\mathbf{r}} \leftarrow \dot{\mathbf{r}} \left[1 + \frac{h}{\tau} \left(\frac{gkT}{2K} - 1 \right) \right]^{1/2} \quad \mathbf{r} \leftarrow \mathbf{r} \left[1 - \frac{h}{\gamma} \left(P - \frac{2K - W}{3V} \right) \right]^{1/3}$$

Αυτές οι εξισώσεις είναι περισσότερο χρήσιμες για την ισορροπία, δεδομένου ότι δεν παράγουν τυπικά σωστά NVT ή NPT σύνολα. [44]

4.6.2.4. Ολονομικοί Περιορισμοί

Όταν τα μήκη των δεσμών είναι περιορισμένα, η συνάρτηση δυναμικού U αυξάνεται με μια περιοριστική συνάρτηση πολλαπλασιασμένη με τον τελεστή Lagrange, $\mathbf{G}(\mathbf{r})$, η οποία επιβάλλει τους περιορισμούς. Χρησιμοποιούμε τη διατύπωση LINCS [Hess 1997]

$$\mathbf{G} = (g_1, \dots, g_k) \quad g_i(\mathbf{r}) = |\mathbf{r}_{a_i} - \mathbf{r}_{b_i}| - d_i$$

Όπου d_i είναι το μήκος στο οποίο ο δεσμός μεταξύ των ατόμων a_i και b_i πρέπει να περιορίζεται. Με άλλα λόγια, $\mathbf{G}=0$ για κάθε t .

Ορίζουμε $\mathbf{B}_{ij} = dg_i/dx_j$, όπου ο x_j είναι η j^{th} συντεταγμένη σε έναν φορέα 3N.

Λόγω της ιδιαίτερης μορφής των συναρτήσεων περιορισμού, έχουμε ότι: $\mathbf{G}(\mathbf{r}) = \mathbf{B}\mathbf{r} - \mathbf{d}$.

Σύμφωνα με τις περιορισμένες εξισώσεις της κίνησης, οι χρονικές παράγωγοι των συναρτήσεων περιορισμού είναι μηδέν.

$$\mathbf{G}(\mathbf{r}) = 0 \quad \dot{\mathbf{G}}(\mathbf{r}) = \frac{\partial \mathbf{G}}{\partial \mathbf{r}} \frac{\partial \mathbf{r}}{\partial t} = \mathbf{B}\dot{\mathbf{r}} = 0 \quad \ddot{\mathbf{G}}(\mathbf{r}) = \mathbf{B}\ddot{\mathbf{r}} + \dot{\mathbf{B}}\dot{\mathbf{r}} = \mathbf{B}\ddot{\mathbf{r}} = 0$$

Σε κάθε δεδομένη στιγμή, οι πολλαπλασιαστές Lagrange \mathbf{y} , τροποποιούν τη δυνατότητα επιβολής των περιορισμών.

$$U(\mathbf{r}) - \mathbf{y} \cdot \mathbf{G}(\mathbf{r}) \quad \mathbf{y} = (\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{M}^{-1}\nabla U(\mathbf{r})$$

και οι δυνάμεις περιορισμού δίνονται από το $\mathbf{B}^T\mathbf{y}$. Έστω $\mathbf{Z} = (\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T)^{-1}$ η οποία είναι ένας τετραγωνικός πίνακας με $1/m_a + 1/m_b$ στη διαγώνιο. Τα στοιχεία εκτός της διαγωνίου είναι μη μηδενικά μόνο όταν δύο περιορισμοί συνδέονται. Σε αυτή την περίπτωση, το στοιχείο έχει τιμή $\cos x / m_c$ όπου m_c είναι η μάζα του ατόμου που συνδέει τους δύο περιορισμένους δεσμούς και το x είναι η γωνία μεταξύ των δεσμών. Αν ορίσουμε τη διαγώνια μήτρα \mathbf{S} έτσι ώστε $\mathbf{S}_i = \text{diag} (\dots, \text{sqrt}(1/m_a + 1/m_b), \dots)$, έχουμε:

$$\mathbf{Z} = \mathbf{S}\mathbf{S}^{-1}\mathbf{Z}\mathbf{S}^{-1}\mathbf{S} = \mathbf{S}(\mathbf{I} - \mathbf{A})\mathbf{S}$$

Ο πίνακας \mathbf{A} είναι συμμετρικός, αραιός και με μηδενική διαγώνιο. Με εξαίρεση τις περιπτώσεις υψηλά τριγωνοποιημένων περιορισμών, όλες οι ιδιοτιμές του είναι μικρότερες από 1. Ως εκ τούτου, μια δυναμική σειρά μπορεί να χρησιμοποιηθεί για να αντιστραφεί ο \mathbf{A} . Με αυτό τον τρόπο, απαιτούνται μόνο πολλαπλασιασμοί πίνακα επί φορέα [Hess 1997], φτάνοντας, έτσι, στην εύρεση ταχύτερου και ακριβέστερου περιορισμού-λύση από ότι ο SHAKE [Ryckaert 1977]. [44]

4.6.2.5. Διορθώσεις στις ορμές

Κατά την ολοκλήρωση των εξισώσεων κίνησης, προκύπτουν αναπόφευκτα αριθμητικά λάθη, οδηγώντας σε μη-συντήρηση της Χαμιλτονιανής. Όταν το πεδίο δυνάμεων είναι συντηρητικό, δηλαδή δεν υπάρχουν σταθερά άτομα ή εξωτερικές δυνάμεις, η συνολική γραμμική και γωνιακή ορμή πρέπει να διατηρείται. Παρόλα αυτά, αν τα αριθμητικά λάθη προκαλούν ανακριβείς διατηρήσεις αυτών των ποσοτήτων, τότε αυτά τα λάθη θα πρέπει να διατηρούνται σε μεγάλο βαθμό στην επόμενη επανάληψη. Για αυτό τον λόγο, είναι σημαντικό να αφαιρούμε τη γραμμική και τη γωνιακή ορμή σε κάθε βήμα. [44, 45]

4.6.3. Περιορισμοί Μοριακής Δυναμικής

Προκειμένου να ολοκληρωθεί η αναφορά στη Μοριακή Δυναμική, θα περιγραφούν σε αυτό το σημείο οι περιορισμοί. Το ζήτημα της χρονικής κλίμακας είναι ο πιο προφανής περιορισμός, με κλίμακες που κυμαίνονται από μερικά picoseconds μέχρι μερικά nanoseconds το πολύ, ανάλογα με το επίπεδο στο οποίο μοντελοποιείται το σύστημα (πρώτες αρχές ή εμπειρικό πεδίο δυναμικού). Άμεση Μοριακή Δυναμική δεν επιφέρει αποτέλεσμα σε περιπτώσεις όπως προβλήματα αργής διάχυσης σε στερεά, δυναμική σε υαλώδη ή πολυμερή συστήματα ή αναδίπλωση πρωτεϊνών. Αυτό συμβαίνει γιατί πολλές διεργασίες χημικού ή φυσικού ενδιαφέροντος πραγματοποιούνται σε χρονικές κλίμακες πολλών τάξεων μεγέθους μεγαλύτερες. Η υπερδυναμική (Hyperdynamics, Voter, 1997), η επιταχυνόμενη δυναμική ανάλογα με τη θερμοκρασία (temperature-accelerated dynamics) και η μέθοδος της παράλληλης ρέπλικας (parallel-replica method) είναι μερικές από τις τεχνικές του Voter και άλλων που προσπαθούν να αντιμετωπίσουν αυτό το πρόβλημα.

Ο υπολογισμός της ενέργειας και των απαραίτητων δυνάμεων για την εκτέλεση των προσομοιώσεων Μοριακής Δυναμικής είναι μία από τις αιτίες για τις οποίες η κλίμακα του χρόνου που μπορεί να καλυφθεί είναι περιορισμένη. Έχει αναφερθεί ότι η δυνατότητα πρόβλεψης από ειδικά σχεδιασμένα νευρωνικά δίκτυα για τις ενέργειες και τις δυνάμεις ενός συγκεκριμένου συστήματος μπορεί να βελτιωθεί τροφοδοτώντας το νευρωνικό δίκτυο με αρκετά δεδομένα που έχουν ληφθεί από προσομοιώσεις. Με αυτή τη μέθοδο, μειώνεται το κόστος εκτέλεσης προσομοιώσεων ακριβείας, με αποτέλεσμα να παρατείνεται το μέγεθος των προσβάσιμων χρονικών κλιμάκων. Ωστόσο, η τεχνική αυτή είναι ακόμα αρκετά νέα και οι ικανότητές της χρήζουν αποδείξεως.

Η Μοριακή Δυναμική είναι παρόμοια τεχνική δειγματοληψίας με την Προσομοίωση

Monte Carlo (MC), αλλά παρέχει επίσης δυναμικές πληροφορίες, με επιπλέον κόστος υπολογισμού των δυνάμεων. Ιδιαίτερη προσοχή απαιτείται στις περιπτώσεις των συστημάτων με περίπλοκα επίπεδα ενέργειας, καθώς είναι πιο προβληματικά να δειγματοστούν επαρκώς και επιπλέον, η εξερεύνηση του χώρου διαμόρφωσής τους από τη δυναμική είναι εξαιρετικά χρονοβόρα. Μια βελτίωση αυτού του προβλήματος προτείνεται από τον Parrinello και τους συνεργάτες του, μια μέθοδος γνωστή ως μεταδυναμική (metadynamics, Laio, Bernard, Chiarotti, Scandolo, & Tosatti, 2000). Στη μεταδυναμική, μια τροχιά προσομοίωσης δημιουργείται για να εξαλείψει την πιθανότητα να επανεξετάσει το ίδιο μέρος του τρισδιάστατου χώρου διαμόρφωσης δύο φορές. Έτσι, μπορούμε σταδιακά να καλύψουμε τα ελάχιστα της δυναμικής ενέργειας με τη διευκόλυνση της διαφυγής του συστήματος από τέτοιου είδους «παγίδες» και με τη βελτίωση της δειγματοληψίας του χώρου διαμόρφωσης (Martona'k et al., 2005).

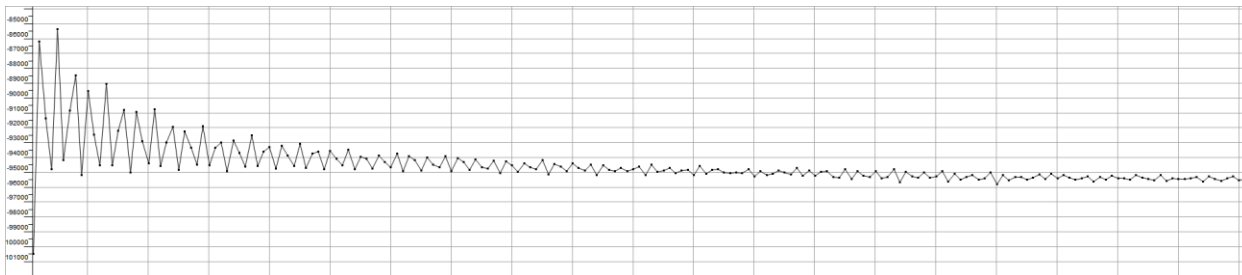
Το ζήτημα κλιμάκων ποικίλου μήκους είναι άλλη μια πρόκληση για τη Μοριακή Δυναμική, όπου σε πολλά συστήματα καλύπτεται από τα φαινόμενα υπό παρακολούθηση. Το πιο χαρακτηριστικό παράδειγμα είναι η διάδοση των ρωγμών. Στη διάδοση ρωγμών, κάποιο υλικό επιβαρύνεται σε οριακό σημείο όταν σχηματιστεί μία ρωγμή. Κοντά σε αυτό το άκρο ρωγμής, οι χημικοί δεσμοί σπάνε, και τα άτομα υφίστανται έντονες ανακατατάξεις. Πιο μακριά από αυτό το άκρο ρωγμής, το υλικό μπορεί να παραμορφωθεί χωρίς απαραίτητα το σπάσιμο των δεσμών και ακόμη πιο μακριά, η ατομική θέση αποκλίνει από τις θέσεις ενός τέλει κρυστάλλου. Για να μοντελοποιήσουμε τα συστήματα σε αυτή την περίπτωση, χρειαζόμαστε μεγάλα κύτταρα προσομοίωσης, τα οποία μπορούν να αυξηθούν έως και περισσότερο από έξι τάξεις μεγέθους. Για την αντιμετώπιση αυτών των προβλημάτων αποτελεσματικά, πρέπει να χρησιμοποιούνται διαφορετικές κλίμακες μήκους σε διαφορετικά θεωρητικά επίπεδα. Για παράδειγμα, σε μια πολύ μικρή περιοχή γύρω από την άκρη της ρωγμής, θα εφαρμοστεί ένας κβαντο-μηχανικός αλγόριθμος και όχι το καθολικό πεδίο δυνάμεων Newton που εφαρμόζεται στο πλήρες σύστημα. Στη συνέχεια, για τις κλίμακες μεγάλου μήκους, μπορούμε να χρησιμοποιήσουμε μια περιγραφή της μηχανικής του συνεχούς. Παρόμοια είναι η περίπτωση των ενζυματικών αντιδράσεων. [43]

4.6.4. Σκοπός

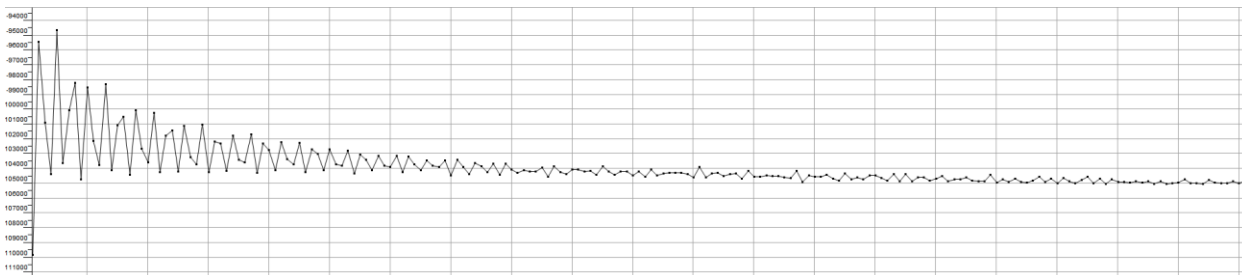
Προκειμένου να μελετηθούν η ελαστικότητα και οι κινητικές ιδιότητες της πρωτεΐνης, απαραίτητη κρίνεται η μελέτη της μοριακής δυναμικής της. Μέσω αυτής, καθίσταται δυνατή η παρατήρηση αλλοιώσεων στο σχήμα ή στο μέγεθός της πρωτεΐνης, κατά τη μετάλλαξη της, καθώς και η επισήμανση τυχόν μεταβολών στις φυσικοχημικές και ηλεκτροστατικές της ιδιότητες. Οποιοδήποτε είδους τέτοιες αλλαγές στη διαμόρφωση ή τις μοριακές αλληλεπιδράσεις επιφέρουν δυσλειτουργία της πρωτεΐνης.

4.6.5. Αποτελέσματα Μοριακής Δυναμικής

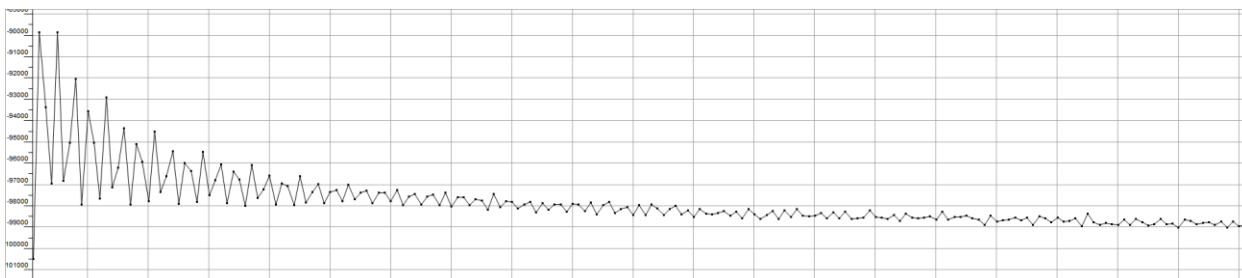
Παρακάτω (σχήματα 4.8-4.10) φαίνονται, ενδεικτικά, τα αποτελέσματα της Μοριακής Δυναμικής (molecular trajectories) για τη μετάλλαξη Glu813Lys για τα τρία πεδία δυνάμεων που περιγράφηκαν παραπάνω.



Σχήμα 4.8. Molecular trajectory με χρήση πεδίου δυνάμεων AMBER



Σχήμα 4.9. Molecular trajectory με χρήση πεδίου δυνάμεων CHARMM



Σχήμα 4.10. Molecular trajectory με χρήση πεδίου δυνάμεων OPLS

Κεφάλαιο 5

Συμπεράσματα – Μελλοντικές Επεκτάσεις

5.1. Συμπεράσματα

Κάνοντας χρήση των μεθόδων που περιγράφηκαν στο κεφάλαιο 3 και λαμβάνοντας τα αποτελέσματα που παρουσιάστηκαν στο κεφάλαιο 4, επιτύχαμε διαδοχικά:

- Την προσομοίωση της τρισδιάστατης δομής της Notch3 με τη μέθοδο της ομόλογης μοντελοποίησης
- Την χαρτογράφηση των μεταλλάξεων τόσο στην αλληλουχία όσο και στη δομή της πρωτεΐνης. Δηλαδή, κάναμε χρήση της ακατέργαστης πληροφορίας – πραγματικών ιατρικών δεδομένων που λάβαμε από τον καθηγητή K. Baumann.
- Την ανίχνευση επαναλαμβανόμενων μοτίβων, όχι μόνο μέσα στην αλληλουχία των αμινοξέων της Notch3, αλλά και στη δευτεροταγή δομή της, ούτως ώστε να παρατηρήσουμε, στη συνέχεια, την κατανομή των μεταλλάξεων σε κάθε μοτίβο.

- Την επισήμανση στρατηγικών θέσεων-κλειδιών, στις οποίες συγκεντρώνεται το μεγαλύτερο ποσοστό των μεταλλάξεων. Οι θέσεις αυτές μπορούν να αποτελέσουν πιθανούς μελλοντικούς φαρμακολογικούς στόχους για την αντιμετώπιση της ασθένειας CADASIL στο μέλλον.

5.2. Μελλοντικές Επεκτάσεις

Στα πλαίσια αυτής της διπλωματικής εργασίας, σκιαγραφήθηκε ένας μεθοδολογικός τρόπος μελέτης της αλληλουχίας και της δομής μιας πρωτεΐνης. Ο τρόπος αυτός μπορεί να χρησιμοποιηθεί μελλοντικά σε περιπτώσεις όπου έχουν αναφερθεί μεταλλάξεις σε συγκεκριμένες θέσεις μίας πρωτεΐνης, προκειμένου να εντοπιστούν περιορισμένου αριθμού στρατηγικές θέσεις, που πιθανόν να διαδραματίζουν κρίσιμο ρόλο στη λειτουργικότητά της.

Συνεπώς, προτείνεται να κατευθυνθεί, σε πρώτη φάση τουλάχιστον, η φαρμακολογική έρευνα στοχευμένα στις θέσεις αυτές, χωρίς να αναλώνεται σε ολόκληρη την ακολουθία της πρωτεΐνης, η οποία μπορεί να αποτελείται από μεγάλο αριθμό αμινοξέων.

Επιπλέον, εδώ επιλέχθηκε και μελετήθηκε δείγμα 6 μεταλλάξεων της Notch3 που πιθανόν να ευθύνονται για την εμφάνιση της CADASIL. Στο Παράρτημα που ακολουθεί στο τέλος της εργασίας υπάρχει ακόμη μεγάλο πλήθος μεταλλάξεων που προτείνεται να μελετηθούν στο μέλλον.

Βιβλιογραφία

- [1] Μοριακή Βιολογία του Γονιδίου, σελ. 101-102, Utopia.
- [2] van Holde and Mathews, pp. 247–50.
- [3] van Holde and Mathews, pp. 220–29.
- [4] Rüdiger H, Siebert HC, Solís D, Jiménez-Barbero J, Romero A, von der Lieth CW, Diaz-Mariño T, Gabius HJ (2000). "Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets". *Current Medicinal Chemistry* 7 (4): 389–416. doi:10.2174/0929867003375164. PMID 10702616.
- [5] Branden and Tooze, pp. 232–34.
- [6] Joutel A, Corpechot C, Ducros A, et al. (October 1996). "Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia". *Nature* 383 (6602): 707–10. doi:10.1038/383707a0. PMID 8878478.
- [7] Sourander P, Walinder J. Hereditary multi-infarct dementia. Morphological and clinical studies of a new disease. *Acta Neuropathol.* Aug 31 1977;39(3):247-54.
- [8] Tournier-Lasserre E, Joutel A, Melki J, Weissenbach J, Lathrop GM, Chabriat H, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy maps to chromosome 19q12. *Nat Genet.* Mar 1993;3(3):256-9.
- [9] Joutel, A. et al. (2000). De novo mutation in the NOTCH3 gene causing CADASIL. *Ann. Neurol.*, 47(3), 388-91.
- [10] Razvi, S. S. M., Davidson, R., Bone, I., et al (2005) The prevalence of CADASIL in the West of Scotland. *Journal of Neurology, Neurosurgery and Psychiatry*, 76, 739–741
- [11] Tatsch, K. et al. (2003). Cortical hypometabolism and crossed cerebellar diaschisis suggest subcortically induced disconnection in CADASIL: an 18F-FDG PET study. *J. Nucl. Med.*, 44(6), 862-9.
- [12] Velizarova, R. et al. (2011). Focal epilepsy as first symptom in CADASIL. *Seizure*, 20(6), 502-4. doi: 10.1016.
- [13] Buffon, F. et al. (2006). Cognitive profile in CADASIL. *J Neurol Neurosurg Psychiatry.*, 77(2), 175-80.

- [14] Saskia AJ Lesnik Oberstein, MD, PhD, Elles MJ Boon, PhD, and Gisela M Terwindt, MD, PhD, CADASIL, GeneReviews, Initial Posting: March 15, 2000; Last Update: June 28, 2012.
- [15] Chabriat H, Vahedi K, Iba-Zizen MT, et al; Clinical spectrum of CADASIL: a study of 7 families. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. *Lancet*. 1995 Oct 7;346(8980):934-9.
- [16] Ueda, M. et al. (2009). Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). [Article in Japanese]. *Rinsho Byori.*, 57(3), 242-51.
- [17] Bersano, A. et al. (2012). Considerations on a mutation in the NOTCH3 gene sparing a cysteine residue: a rare polymorphism rather than a CADASIL variant. *Funct Neurol.*, 27(4), 247-52.
- [18] Bonthius DJ, Mathews KD, Adams HP. CADASIL in a North American family: linkage to chromosome 19. *Neurology* 1996;46:A211-2.
- [19] Saara Tikka, Kati Mykkänen, Maija Junna, Robert Bergholm, Minna Pöyhönen, Marc Baumann, Matti Viitanen and Hannu Kalimo, Diagnosing Vascular Dementia by Skin Biopsy - Uniqueness of CADASIL
- [20] Ulla-Maj Fiuza, Alfonso Martinez Arias, Cell and molecular biology of Notch, *Journal of Endocrinology* (2007) 194, 459–474, 2007 Society for Endocrinology, Printed in Great Britain
- [21] Bellavia, D. et al. (2008). Notch3: from subtle structural differences to functional diversity. *Oncogene.*, 27(38), 5092-8. doi: 10.1038/onc.2008.230.
- [22] Tikka, S. et al. (2008). Congruence between NOTCH3 mutations and GOM in 131 CADASIL patients. *Brain*, 132(4), 933-939.
- [23] Sophia Kossida, Molecular modelling study of Notch 3 provides insights on CADASIL, Fondation Santé Research Grant Proposal in Biomedical Sciences, 2013
- [24] Wendy R. Gordon, Kelly L. Arnett and Stephen C. Blacklow, The molecular logic of Notch signaling – a structural and biochemical perspective , *Journal of Cell Science* 121, 3109-3119 Published by The Company of Biologists 2008
- [25] Eliza Lewandowska, Dorota Dziewulska, Magdalena Parys, Elzbieta Pasennik, *Folia Neuropathol* 2011; 49 (3): 174-180, Ultrastructure of granular osmiophilic material deposits (GOM) in arterioles of CADASIL patients
- [26] Akira Ishiko, Atsushi Shimizu, Eiichiro Nagata, Keikichi Takahashi, Takeshi Tabira, Norihiro Suzuki, Notch3 ectodomain is a major component of granular osmiophilic material (GOM) in CADASIL, Springer-Verlag 2006

- [27] Consortium, T.U. (2011) Ongoing and future developments at the Universal Protein Resource. *Nucleic acids research*, 39, D214-D219.
- [28] Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, 14, 1188-1190.
- [29] Higgins D., Thompson J., Gibson T. Thompson J. D., Higgins D. G., Gibson T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids. Res.* 22:4673-4680.
- [30] Robert C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, 1792±1797 *Nucleic Acids Research*, 2004, Vol. 32, No. 5 DOI: 10.1093/nar/gkh340
- [31] Cordle, J. et al. (2008) A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nat Struct Mol Biol.*, 15(8), 849-57. doi: 10.1038/nsmb.1457.
- [32] Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8, 477-486.
- [33] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4, 435-447; Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D., et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29, 845-854
- [34] Sellis, D., Vlachakis, D., and Vlassi, M. (2009) Gromita: a fully integrated graphical user interface to gromacs 4. *Bioinformatics and biology insights*, 3, 99-102.
- [35] Hong Luo and Harm Nijveen, Understanding and identifying amino acid repeats, *Briefings in Bioinformatics Advance Access published February 14, 2013*, doi:10.1093/bib/bbt003
- [36] Robinson, CR. And Sauer, RT. (1998) Optimizint the stability of single-chain proteins by linker length and composition mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.*, 95(11), 5929-34.
- [37] Merkel, JS. and Regan, L. (1998) Aromatic rescue of glycine in beta sheets. *Fold. Des.*, 3(6), 449-55.
- [38] Zhang, T. et al. (2008) Accurate sequence-based prediction of catalytic residues. *Bioinformatics*, 24(20), 2329-38. doi: 10.1093/bioinformatics/btn433.

- [39] Zhang Y, Skolnick J. (2005). "The protein structure prediction problem could be solved using the current PDB library". *Proc Natl Acad Sci USA* 102 (4): 1029–34. doi:10.1073/pnas.0407152101
- [40] Martin ACR (1982). "Rapid Comparison of Protein Structures". *Acta Cryst A* 38 (6): 871–873. doi:10.1107/S0567739482001806.
- [41] Theobald DL, Wuttke DS (2006). "Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem". *Proceedings of the National Academy of Sciences* 103 (49): 18521–18527. doi:10.1073/pnas.0508445103.
- [42] Theobald DL, Wuttke DS (2006). THESEUS: Maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* 22 (17): 2171 – 2172. doi: 10.1093/bioinformatics/btl332. PMC 2584349. PMID 16777907.
- [43] Vlachakis D, Bencurova E, Papangelopoulos N, Kossida S (2014). Current State-of-the-Art Molecular Dynamics Methods and Applications. *Adv Protein Chem Struct Biol.*;94:269-313. doi: 10.1016/B978-0-12-800168-4.00007-X.
- [44] MOE 2010/help/Molecular Dynamics
- [45] Α. Κουκούλας, (2012) Υπολογιστικές προσομοιώσεις μοριακής δυναμικής σε τήγματα πολυμερικών δακτυλίων του αιθυλενοξειδίου, μεταπτυχιακή διπλωματική εργασία, Τμήμα Χημικών Μηχανικών, Πανεπιστήμιο Πατρών

ΠΑΡΑΡΤΗΜΑ

(Εδώ παρατίθεται η λίστα των συνολικών μεταλλάξεων της Notch3, όπως αυτή στάλθηκε από τον καθηγητή κ. Baumann)

Reported *NOTCH3* mutations. 22.3.2011

Nro	DNA reported*	Coding sequence†	Exon /intron	Mutations per exon (%)	Amino acid change	EGF repeat	Reference‡	Notices
1	205T>G	c.127T>G	exon 2	11/203 (5,4)	p.Cys43Gly	1	[1, 2]	
2	G206T	c.128G>T			p.Cys43Phe	1	[3]	
3	TGT>GGT	c.145T>G			p.Cys49Gly	1	[4]	
4	c.145T>C	c.145T>C			p.Cys49Arg	1	[81]	
5	224TGT>TAT	c.146G>A			p.Cys49Tyr	1	[5]	
6	224G>T	c.146G>T			p.Cys49Phe	1	[1]	
7	c.157G>T	c.157G>T			p.Gly53Cys	1	[81]	
8	238C>T	c.160C>T			p.Arg54Cys	1	[6]	
9	257C>G	c.179C>G			p.Ser60Cys	1	[1]	
10	272G>C	c.194G>C			p.Cys65Ser	1	[1]	
11	TGC-TAC	c.194G>A			p.Cys65Tyr	1	[7]	
12	c.199T>A	c.199T>A	exon 3	21/203 (10.3)	p.Cys67Ser	1	[8] ‡	
13	G>A	c.200G>A			p.Cys67Tyr	1	[9]	
14	TGG>TGT	c.213G>T			p.Trp71Cys	1	[10]	
15	224G>C	c.224G>C			p.Arg75Pro	1	[11]	non-cysteine
16	Tgt 304 Cgt	c.226T>C			p.Cys76Arg	1	[12]	
17	c.226_234del	c.226_234del			p.Cys76_Leu78del	1	[81]	
18	306T>G	c.228T>G			p.Cys76Trp	1	[1]	
19	309_326del	c.231_248del			p.Gln77_Cys82del	1 + 2	[1]	deletion
20	317-331 del	c.239_253del			p.Asp80_Ser84del	2	[13]	deletion
21	337T>C	c.259T>C			p.Cys87Arg	2	[1]	
22	338G>A	c.260G>A			p.Cys87Tyr	2	[1]	
23	341del12	c.263_274del		p.Ala88_Gly 91del	2	[14]	non-cysteine, deletion	
24	346CGT>TGT	c.268C>T		p.Arg90Cys	2	[5]		
25	c.357insTGC	c.277_279dup		p.Cys93dup	2	[15]	duplication	
26	tGc 356 tAc	c.278G>A		p.Cys93Tyr	2	[12]		
27	356TGC>TTC	c.278G>T		p.Cys93Phe	2	[13]		
28	396C>G	c.318C>G		p.Cys106Trp	2	[1]		
29	322C>T	c.322T>C		p.Cys108Arg	2	[16]		
30	401G>A	c.323G>A		p.Cys108Tyr	2	[1]		
31	TGC>TGG	c.324C>G		p.Cys108Trp	2	[17]		
32	406CGT>TGT	c.328C>T		p.Arg110Cys	2	[5]		
33	A>G in 3' splice site of exon 4	c.341-2A>G	intron 3	1/203 (0.5)	p.Gly114_Pro120del	2	[18]	splice site mutation, deletion
34	-	c.350G>T	exon 4	67/203 (33.0)	p.Cys117Phe	2	[19]	
35	c.349T>C	c.349T>C			p.Cys117Arg	2	[81]	
36	G428A	c.350G>A			p.Cys117Tyr	2	[80]	
37	TCC>TGC	c.353C>G			p.Ser118Cys	3	[20]	
38	446TGC>TTC	c.368G>T			p.Cys123Phe	3	[13]	
39	446G>A	c.368G>A			p.Cys123Tyr	3	[6]	
40	del TTGTG	c.381_385del			p.Cys128Pro fsX32	2	[21]	frame shift
41	TGT>GGT	c.382T>G			p.Cys128Gly	3	[22]	de novo
42	tGt 461 tAt	c.383G>A			p.Cys128Tyr	3	[12]	
43	c.383G>T	c.383G>T			p.Cys128Phe	3	[23]	

44	cGGT>TGT	c.391G>T	p.Gly131Cys	3	[24]	
45	475CGC>TGC	c.397C>T	p.Arg133Cys	3	[5, 25]	homozygous
46	-	c.402C>G	p.Cys134Trp	3	[26]	
47	499CGC>TGC	c.421C>T	p.Arg141Cys	3	[5]	
48	tTc 503 tGc	c.425T>G	p.Phe142Cys	3	[12]	
49	509TGC>TCC	c.431G>C	p.Cys144Ser	3	[13]	
50	509TGC>TAC	c.431G>A	p.Cys144Tyr	3	[13]	
51	g509G>T	c.431G>T	p.Cys144Phe	3	[27]	
52	512C>G	c.434C>G	p.Ser145Cys	3	[1]	
53	514TGC>CGC	c.436T>C	p.Cys146Arg	3	[5]	
54	G437A	c.437G>A	p.Cys146Tyr	3	[28]	
55	523G>T	c.445G>T	p.Gly149Cys	3	[1]	
56	527TAC>TGC	c.449A>G	p.Tyr150Cys	3	[13]	
57	C529G	c.451C>G	p.Gln151Glu		[80]	non-cysteine
58	535CGC>TGC	c.457C>T	p.Arg153Cys	3	[5]	
59	537-545	c.459_467del	p.Arg153_Cys155del	3	[13]	deletion
60	542G>C	c.464G>C	p.Cys155Ser	3	[1]	
61	c.464G>A	c.464G>A	p.Cys155Tyr	3	[23]	
62	562T>A	c.484T>A	p.Cys162Ser	4	[6]	
63	484T>C	c.484T>C	p.Cys162Arg	4	[29]	
64	-	c.486C>G	p.Cys162Trp	4	[30]	
65	G571T	c.493G>T	p.Gly165Cys		[80]	
66	CGC>TGC	c.505C>T	p.Arg169Cys	4	[10]	
67	c.509A>G	c.509A>G	p.His170Cys	4	[81]	
68	A587T	c.509A>T	p.His170Arg		[80]	non-cysteine
69	589GGT>TGT	c.511G>T	p.Gly171Cys	4	[5]	
70	-	c.520T>C	p.Cys174Arg	4	[31]	
71	c.520T>A	c.520T>A	p.Cys174Ser	4	[32] ‡	
72	599G>T	c.521G>T	p.Cys174Phe	4	[33]	
73	-	c.521G>A	p.Cys174Tyr	4	[19]	
74	617C>G	c.539C>G	p.Ser180Cys	4	[6]	
75	F181C	c.542T>G	p.Phe181Cys		[82]	
76	CGC>TGC	c.544C>T	p.Arg182Cys	4	[10, 34]	de novo
77	-	c.547T>C	p.Cys183Arg	4	[19]	
78	625TGC>AGC	c.547T>A	p.Cys183Ser	4	[13]	
79	626G>T	c.548G>T	p.Cys183Phe	4	[1]	
80	631TGT>CGT	c.553T>C	p.Cys185Arg	4	[5]	
81	-	c.553T>G	p.Cys185Gly	4	[26]	
82	A644G	c.566A>G	p.Tyr189Cys	4	[3]	
83	24 bp del ex4 codons 192-200	c.574_597del	p. Pro192_Val200del		[79]	deletion
84	Tgt 658 Cgt	c.580T>C	p.Cys194Arg	4	[12]	
85	T658A	c.580T>A	p.Cys194Ser	4	[35]	
86	tGt 659 tTt	c.581G>T	p.Cys194Phe	4	[13]	
87	659G>A	c.581G>A	p.Cys194Tyr	4	[6]	
88	c.T601C	c.601T>C	p.Cys201Arg	5	[36]	
89	680G>A	c.602G>A	p.Cys201Tyr	5	[1]	
90	-	c.616T>C	p.Cys206Arg	5	[37]	
91	695G>A	c.617G>A	p.Cys206Tyr	5	[6]	
92	697CGT>TGT	c.619C>T	p.Arg207Cys	5	[38]	
93	712TGC>AGC	c.634T>A	p.Cys212Ser	5	[5]	
94	(c713G>A)	c.635G>A	p.Cys212Tyr		[83]	
95	716G>A	c.638G>A	p.Arg213Lys	5	[39]	non-cysteine
96	TAC>TGC	c.659A>G	p.Tyr220Cys	5	[40]	
97	742TGT>GGT	c.664T>G	p.Cys222Gly	5	[5]	
98	tGt 743 tAt	c.665G>A	p.Cys222Tyr	5	[41]	

99	c.665G>C	c.665G>C			p.Cys222Ser	5	[81]	
100	749TGT>TAT	c.671G>A			p.Cys224Tyr	5	[5]	
101	-	c.697T>A	exon 5	16/203 (7,9)	p.Cys233Ser	5	[26]	
102	T697C	c.697T>C			p.Cys233Arg		[76]	
103	776G>A	c.698G>A			p.Cys233Tyr	5	[1]	
104	G777T	c.699T>G			p.Cys233Trp	5	[3]	
105	-	c.709G>A			p.Val237Met	5	[42]	non-cysteine
106	del 45bp	c.714_758del			p.Asp239_Asp253del	6	[43]	deletion
107	797G>C	c.719G>C			p.Cys240Ser	6	[1]	
108	Tgt 811 Cgt	c.733T>C			p.Cys245Arg	6	[1]	
109	-	c.733T>A			p.Cys245Ser	6	[44]	
110	Tgc 829 Cgc	c.751T>C			p.Cys251Arg	6	[12]	
111	T829A	c.751T>A			p.Cys251Ser	6	[3]	
112	TGC>GGC	c.751T>G			p.Cys251Gly	6	[45]	
113	c.752G>A	c.752G>A			p.Cys251Tyr	6	[46] ‡	
114	851TAT>TGT	c.773A>G			p.Tyr258Cys	6	[5]	
115	856T>G				p.Cys260Gly		[77]	
116	857G>A	c.779G>A			p.Cys260Tyr	6	[1]	
117	TGC>TTC	c.812G>C	exon 6	7/203 (3.4)	p.Cys271Phe	6	[47]	
118	973A>T	c.895A>T			p.Ser299Cys	7	[48]	
119	1033_1034GC>TG	c.955_956delinsTG			p.Ala319Cys	8	[1]	
120	CGC>TGC	c.994C>T			p.Arg332Cys	8	[49]	
121	1082C>G	c.1004C>G			p.Ser335Cys	8	[1]	
122	A1088G	c.1010A>G			p.Tyr337Cys	8	[3]	
123	T1090C	c.1012T>C			p.Cys338Arg	8	[50]	
124	TGT->TGG	c.1098T>G	exon 7	6/203 (2.9)	p.Cys366Trp	9	[51]	
125		c.1135T>C			p.Cys379Arg		[57]	
126	1214G>C	c.1136G>C			p.Cys379Ser	9	[1]	
127	G1222T	c.1144G>T			p.Gly382Cys	9	[3]	
128	1241TGT>TAT	c.1163G>A			p.Cys388Tyr	9	[52]	
129	1261T>C	c.1183T>C			p.Cys395Arg	10	[1]	
130	C419R	c.1255T>C	exon 8	17/203 (8.4)	p.Cys419Arg		[79]	
131	-	c.1258G>T			p.Gly420Cys	10	[26]	
132	1339C>T	c.1261C>T			p.Arg421Cys	10	[1]	
133	C1279T	c.1279C>T			p.Arg427Cys		[76]	
134	T1360C	c.1282T>C			p.Cys428Arg	10	[50]	
135	-	c.1283G>C			p.Cys428Ser	10	[26]	
136	1361G>A	c.1283G>A			p.Cys428Tyr	10	[1]	
137	c.1300_1308dup	c.1300_1308dup			p.Glu434_Leu436dup	11	[53] ‡	duplication
138	T1381C	c.1303T>C			p.Cys435Arg	11	[3]	
139	Tgc 1396 Ggc	c.1318T>G			p.Cys440Gly	11	[12]	
140	1396T>C	c.1318T>C			p.Cys440Arg	11	[1]	
141	T1396A	c.1318T>A			p.Cys440Ser	11	[54]	
142	-	c.1337G>T			p.Cys446Phe	11	[30]	
143	1415G>C	c.1337G>C			p.Cys446Ser	11	[1]	
144	Cgc 1423 Tgc	c.1345C>T			p.Arg449Cys	11	[55]	
145	T1441C	c.1363T>C			p.Cys455Arg	11	[56]	
146	G1370T	c.1370G>T			p.Cys457Ser		[76]	
147	A1473G	c.1394A>T	exon 9	6/203 (3.0)	p.Tyr465Cys	11	[3]	
148	1397G>A	c.1397G>A			p.Cys466Tyr		[84]	
149	c.1450T>G	c.1450T>G			p.Cys484Gly	12	[57]	
150	1529G>A	c.1451G>A			p.Cys484Tyr	12	[1]	
151	1529G>T	c.1451G>T			p.Cys484Phe	12	[2]	
152	1562G>A	c.1484G>A			p.Cys495Tyr	12	[1]	

153	C504R	c.1510T>C	exon 10	6/203 (3.0)	p.Cys504Arg		[78]	
154	1609T>C	c.1531T>C			p.Cys511Arg	13	[1]	
155	TGC>TAC	c.1532G>A			p.Cys511Tyr	13	[58]	
156	G1660T	c.1582G>T			p.Gly528Cys	13	[50]	
157	TGC>TCC	c.1592G>C			p.Cys531Ser	13	[59]	
158	cCGC-TGC	c.1594C>T			p.Arg532Cys	13	[60]	
159	TGT>TAT	c.1625G>A	exon 11	13/203 (6.4)	p.Cys542Tyr	13	[10]	
160	1708CGC>TGC	c.1630C>T			p.Arg544Cys	14	[38]	
161	T1723C	c.1645T>C			p.Cys549Arg	14	[3]	
162	1724G>A	c.1646G>A			p.Cys549Tyr	14	[1]	
163	CGC>TGC	c.1672C>T			p.Arg558Cys	14	[10]	
164	TGT>TAT	c.1703G>A			p.Cys568Tyr	14	[61]	
165	TAC>TGC	c.1721A>G			p.Tyr574Cys	14	[62]	
166	ACA>GCA	c.1729A>G			p.Thr577Ala	14	[63]	non-cysteine
167	CGC>TGC	c.1732C>T			p.Arg578Cys	14	[10]	
168	1759C>T	c.1759C>T			p.Arg587Cys	15	[64]	
169	c.1771T>C	c.1771T>C			p.Cys591Arg	15	[65]	
170	G1774T	c.1774G>T			p.Arg592Cys		[76]	
171	1897C>T	c.1819C>T			p.Arg607Cys	15	[6]	
172	c.1918C>T	c.1918C>T	exon 12	1/203 (0.5)	p.Arg640Cys	16	[23]	
173	c.1999G>T	c.1999G>T	exon 13	1/203 (0.5)	p.Gly667Cys	17	[23]	
174	2260CGC>TGC	c.2182C>T	exon 14	1/203 (0.5)	p.Arg728Cys	18	[5]	
175	2402G>C	c.2324G>C	exon 15	1/203 (0.5)	p.Cys775Ser	20	[2]	
176	1279G>T	c.2411-1G>T	intron 15	1/203 (0.5)	p.Pro805_Asn856del	21,22,23	[66]	splice site mutation
177	Ggc 2935 Tgc	c.2857G>T	exon 18	7/203 (3.4)	p.Gly953Cys	24	[12]	
178	cTGC>AGC	c.2929T>A			p.Cys977Ser	25	[20]	
179	cAGC-CGC	c.2932A>C			p.Ser978Arg	25	[67]	non-cysteine
180	3029T>G	c.2951T>G			p.Phe984Cys	25	[6]	
181	3031CGC>TGC	c.2953C>T			p.Arg985Cys	25	[5]	
182	2963G>A	c.2963G>A			p.Cys988Tyr	25	[64]	
183	gTGC>GGC	c.2989T>G			p.Cys997Gly	25	[68]	
184	-	c.3011A>C	exon 19	7/203 (3.4)	p.Cys1004Tyr	26	[69]	
185	3094CGC>GCT	c.3016C>T			p.Arg1006Cys	26	[5]	
186	3121TGC>CGC	c.3043T>C			p.Cys1015Arg	26	[38]	
187	tAt 3141 tGt	c.3062A>T			p.Tyr1021Cys	26	[41]	
188	c.3058G>C	c.3058G>C			p.Ala1020Pro	26	[70]	non-cysteine
189	c.3084G>T	c.3084G>T			p.Trp1028Cys	26	[71]	
190	3169CGC>TGC	c.3091C>T			p.Arg1031Cys	26	[5]	
191	Ggt 3250 Tgt	c.3172G>T	exon 20	5/203 (2.5)	p.Gly1058Cys	27	[12, 26]	
192	c.3206A>G	c.3206A>G			p.Tyr1069Cys	27	[46] ‡	
193	-	c.3226C>T			p.Arg1076Cys	27	[30]	
194	c.3292A>T	c.3292A>T			p.Thr1098Ser	28	[81]	
195	TGC>TAC	c.3296G>A			p.Cys1099Tyr	28	[72]	
196	c.3471C>G	c.3471C>G	exon 21	1/203 (0.5)	p.Cys1131Trp	29	[73]	
197	3769CGT>TGT	c.3691C>T	exon 22	1/203 (0.5)	p.Arg1231Cys	31	[5]	
198		c.3750C>G	exon 23	3/203 (1.5)	p.Cys1250Trp		[57]	
199	TGC>CGC	c.3781T>C			p.Cys1261Arg	32	[10]	
200	3860G>A	c.3782G>A			p.Cys1261Tyr	32	[1]	
201	c.3944G>A	c.3944G>A	Exon 24	1/203 (0.5)	p.Cys1315Tyr		[85]	
202	c.4544T>C	c.4544T>C	exon 25	1/203 (0.5)	p.Leu1515Pro	-	[74]	activating, not CADASIL-related
203	GCT>ACT	c.4822G>A	exon 26	1/203 (0.5)	p.Ala1608Thr	-	[10]	non-cysteine

The reported mutations in *NOTCH3* gene are listed in the table. Mutations reported not to associate with CADASIL are marked in ‘Notices’ column.

* Description of mutation at the DNA level as reported in original publication (see reference).

† Mutation description at the DNA level in coding sequence of *NOTCH3*. Nucleotide number starting from the A of the AGT translation initiating methionine[75].

‡ Previously unpublished mutation.

References:

1. Opherkerk, C., et al., *Long-term prognosis and causes of death in CADASIL: a retrospective study in 411 patients*. Brain, 2004. **127**(Pt 11): p. 2533-9.
2. Peters, N., et al., *Spectrum of mutations in biopsy-proven CADASIL: implications for diagnostic strategies*. Arch Neurol, 2005. **62**(7): p. 1091-4.
3. Lesnik Oberstein, S.A.J., *Diagnostic strategies in CADASIL*. Neurology, 2003. **60**(12): p. 2020; author reply 2020.
4. Oki, K., et al., *Novel mutation of the Notch3 gene in a Japanese patient with CADASIL*. Eur J Neurol, 2007. **14**(4): p. 464-6.
5. Joutel, A., et al., *Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients*. Lancet, 1997. **350**(9090): p. 1511-5.
6. Escary, J.L., et al., *Evaluation of DHPLC analysis in mutational scanning of Notch3, a gene with a high G-C content*. Hum Mutat, 2000. **16**(6): p. 518-26.
7. Bianchi, S., et al., *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: CADASIL, exon 2 mutation*. Hum Genet, 2007. **122**(5): p. 558.
8. Bondeson, M.-L., *Previously unpublished data*. 2007, Department of Genetics and Pathology, University of Uppsala: Uppsala, Sweden.
9. Moon, S.Y., et al., *A novel mutation (C67Y) in the NOTCH3 gene in a Korean CADASIL patient*. J Korean Med Sci, 2003. **18**(1): p. 141-4.
10. Joutel, A., et al., *Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia*. Nature, 1996. **383**(6602): p. 707-10.
11. Kim, Y., et al., *Characteristics of CADASIL in Korea: a novel cysteine-sparing Notch3 mutation*. Neurology, 2006. **66**(10): p. 1511-6.
12. Kalimo, H., et al., *CADASIL: a common form of hereditary arteriopathy causing brain infarcts and dementia*. Brain Pathol, 2002. **12**(3): p. 371-84.
13. Dichgans, M., et al., *Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3 EGF-like repeat domains*. Eur J Hum Genet, 2000. **8**(4): p. 280-5.
14. Mazzei, R., et al., *A novel Notch3 gene mutation not involving a cysteine residue in an Italian family with CADASIL*. Neurology, 2004. **63**(3): p. 561-4.
15. Mazzei, R., et al., *First evidence of a pathogenic insertion in the NOTCH3 gene causing CADASIL*. J Neurol Neurosurg Psychiatry, 2008. **79**(1): p. 108-10.
16. Wang, Z.X., et al., *[NOTCH3 gene mutations in four Chinese families with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy]*. Zhonghua Yi Xue Za Zhi, 2004. **84**(14): p. 1175-80.
17. Rojas-Marcos, I., et al., *Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2004. **115**(2): p. 175.
18. Joutel, A., et al., *Splice site mutation causing a seven amino acid Notch3 in-frame deletion in CADASIL*. Neurology, 2000. **54**(9): p. 1874-5.
19. Dichgans, M., et al., *Quantitative MRI in CADASIL: correlation with disability and cognitive performance*. Neurology, 1999. **52**(7): p. 1361-7.
20. Lee, Y.C., et al., *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: two novel mutations in the NOTCH3 gene in Chinese*. J Neurol Sci, 2006. **246**(1-2): p. 111-5.

21. Dotti, M.T., et al., *A novel NOTCH3 frameshift deletion and mitochondrial abnormalities in a patient with CADASIL*. Arch Neurol, 2004. **61**(6): p. 942-5.
22. Coto, E., et al., *A new de novo Notch3 mutation causing CADASIL*. Eur J Neurol, 2006. **13**(6): p. 628-31.
23. Fokkema, I.F., J.T. den Dunnen, and P.E. Taschner, *LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach*. Hum Mutat, 2005. **26**(2): p. 63-8.
24. Ungaro, C., et al., *Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2008. **123**: p. 554.
25. Tuominen, S., et al., *Phenotype of a homozygous CADASIL patient in comparison to 9 age-matched heterozygous patients with the same R133C Notch3 mutation*. Stroke, 2001. **32**(8): p. 1767-74.
26. Joutel, A., et al., *Skin biopsy immunostaining with a Notch3 monoclonal antibody for CADASIL diagnosis*. Lancet, 2001. **358**(9298): p. 2049-51.
27. Grigg, R., et al., *Identification of a novel mutation C144F in the Notch3 gene in an Australian CADASIL pedigree*. Hum Mutat, 2000. **16**(5): p. 449-50.
28. Malandrini, A., et al., *Asymptomatic cores and paracrystalline mitochondrial inclusions in CADASIL*. Neurology, 2002. **59**(4): p. 617-20.
29. Andreadou, E., G. Papadimas, and C. Sfagos, *A novel heterozygous mutation in the NOTCH3 gene causing CADASIL*. Swiss Med Wkly, 2008. **138**(41-42): p. 614-7.
30. Lesnik Oberstein, S.A., et al., *Cerebral microbleeds in CADASIL*. Neurology, 2001. **57**(6): p. 1066-70.
31. Santa, Y., et al., *Genetic, clinical and pathological studies of CADASIL in Japan: a partial contribution of Notch3 mutations and implications of smooth muscle cell degeneration for the pathogenesis*. J Neurol Sci, 2003. **212**(1-2): p. 79-84.
32. Huoponen, K., *Previously unpublished data*. 2003, DNA Diagnostic Laboratory, Department of Medical Genetics, University of Turku: Turku, Finland.
33. Kotorii, S., et al., *Mutations of the notch3 gene in non-caucasian patients with suspected CADASIL syndrome*. Dement Geriatr Cogn Disord, 2001. **12**(3): p. 185-93.
34. Joutel, A., et al., *De novo mutation in the Notch3 gene causing CADASIL*. Ann Neurol, 2000. **47**(3): p. 388-91.
35. Markus, H.S., et al., *Diagnostic strategies in CADASIL*. Neurology, 2002. **59**(8): p. 1134-8.
36. Uyguner, Z.O., et al., *The R110C mutation in Notch3 causes variable clinical features in two Turkish families with CADASIL syndrome*. J Neurol Sci, 2006. **246**(1-2): p. 123-30.
37. Matsumoto, H., et al., *[A case of early stage CADASIL showing only dizziness and vertigo with a novel mutation of Notch 3 gene]*. Rinsho Shinkeigaku, 2005. **45**(1): p. 27-31.
38. Lesnik Oberstein, S.A., et al., *Diagnostic Notch3 sequence analysis in CADASIL: three new mutations in Dutch patients*. Dutch CADASIL Research Group. Neurology, 1999. **52**(9): p. 1913-5.
39. Okeda, R., K. Arima, and M. Kawai, *Arterial changes in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) in relation to pathogenesis of diffuse myelin loss of cerebral white matter: examination of cerebral medullary arteries by reconstruction of serial sections of an autopsy case*. Stroke, 2002. **33**(11): p. 2565-9.
40. Rojas-Marcos, I., et al., *Gene symbol: NOTCH3*. Hum Genet, 2007. **120**(6): p. 917.
41. Meeks, J.J., et al., *Two novel mutations in Notch3 in North American CADASIL patients*. Stroke, 1999. **30**: p. 250.
42. Uchino, M., et al., *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) and CADASIL-like disorders in Japan*. Ann N Y Acad Sci, 2002. **977**: p. 273-8.

43. Dichgans, M., J. Herzog, and T. Gasser, *NOTCH3 mutation involving three cysteine residues in a family with typical CADASIL*. Neurology, 2001. **57**(9): p. 1714-7.
44. Razvi, S.S., et al., *The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland*. J Neurol Neurosurg Psychiatry, 2005. **76**(5): p. 739-41.
45. Vikelis, M., J. Papatriantafyllou, and C.E. Krageorgiou, *A novel CADASIL-causing mutation in a stroke patient*. Swiss Med Wkly, 2007. **137**(21-22): p. 323-5.
46. Mykkänen, K., *Previously unpublished data*. 2007, Department of Medical Genetics, University of Turku: Turku, Finland.
47. Au, K.M., et al., *Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2006. **119**(6): p. 677.
48. Golomb, M.R., et al., *Recurrent hemiplegia, normal MRI, and NOTCH3 mutation in a 14-year-old: is this early CADASIL?* Neurology, 2004. **62**(12): p. 2331-2.
49. Oliveri, R.L., et al., *A novel mutation in the Notch3 gene in an Italian family with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: genetic and magnetic resonance spectroscopic findings*. Arch Neurol, 2001. **58**(9): p. 1418-22.
50. Dotti, M.T., et al., *The spectrum of Notch3 mutations in 28 Italian CADASIL families*. J Neurol Neurosurg Psychiatry, 2005. **76**(5): p. 736-8.
51. Pradotto, L., et al., *Sporadic vascular dementia as clinical presentation of a new missense mutation within exon 7 of NOTCH3 gene*. J Neurol Sci, 2008. **271**(1-2): p. 207-10.
52. Ishida, C., et al., *CADASIL with a novel mutation in exon 7 of NOTCH3 (C388Y)*. Intern Med, 2006. **45**(16): p. 981-5.
53. Laitinen, V., *Previously unpublished data*. 2008, DNA Diagnostic Laboratory, Department of Medical Genetics, University of Turku: Turku, Finland.
54. Federico, A., S. Bianchi, and M.T. Dotti, *The spectrum of mutations for CADASIL diagnosis*. Neurol Sci, 2005. **26**(2): p. 117-24.
55. Thomas, N.J., et al., *Hereditary vascular dementia linked to notch 3 mutations. CADASIL in British families*. Ann N Y Acad Sci, 2000. **903**: p. 293-8.
56. Arboleda-Velasquez, J.F., et al., *C455R notch3 mutation in a Colombian CADASIL kindred with early onset of stroke*. Neurology, 2002. **59**(2): p. 277-9.
57. Leiden University Medical Center. *Submitted to: LOVD - Leiden Open Variation Database*. 2008 [cited; Available from: http://chromium.liacs.nl/LOVD2/home.php?select_db=NOTCH3].
58. Bianchi, S., et al., *Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2005. **118**(3-4): p. 535.
59. Mazzei, R., et al., *Gene symbol: NOTCH3*. Hum Genet, 2007. **121**(2): p. 295.
60. Bianchi, S., et al., *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2007. **122**(5): p. 558-559.
61. Ferreira, S., R.S. Silva, and J.P. Oliveira, *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL)*. Hum Genet, 2007. **121**(5): p. 649.
62. Mazzei, R., et al., *Gene symbol: NOTCH3*. Hum Genet, 2007. **121**(2): p. 296.
63. Ferreira, S., F. Malheiro, and J.P. Oliveira, *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant aretriopathy with subcortical infarcts and leukoencephalopathy*. Hum Genet, 2007. **121**(5): p. 651-2.
64. Kim, Y., et al., *Two novel mutations of the NOTCH3 gene in Korean patients with CADASIL*. Mutat Res, 2006. **593**(1-2): p. 116-20.
65. Pérez-Tur, J. *Submitted to: LOVD -Leiden Open Variation Database*. 2008 [cited; Available from: http://chromium.liacs.nl/LOVD2/home.php?select_db=NOTCH3].
66. Saiki, S., et al., *Varicose veins associated with CADASIL result from a novel mutation*

- in the Notch3 gene*. Neurology, 2006. **67**(2): p. 337-9.
67. Ferreira, S., et al., *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL)*. Hum Genet, 2007. **121**(5): p. 649-50.
 68. Ungaro, C., et al., *Gene symbol: NOTCH3. Disease: CADASIL*. Hum Genet, 2008. **123**: p. 555.
 69. Guerrot, D., et al., *Nephroangiosclerosis in cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy: is NOTCH3 mutation the common culprit?* Am J Kidney Dis, 2008. **52**(2): p. 340-5.
 70. Scheid, R., et al., *Cysteine-sparing notch3 mutations: cadasil or cadasil variants?* Neurology, 2008. **71**(10): p. 774-6.
 71. Viana-Babstista, M., *The spectrum of NOTCH3 mutations in portugese patients with CADASIL: implications for diagnostic strategies*. Cerebrovasc Dis, 2007. **23**(suppl 2): p. 75.
 72. Ferreira, S., C. Costa, and J.P. Oliveira, *Novel human pathological mutations. Gene symbol: NOTCH3. Disease: cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL)*. Hum Genet, 2007. **121**(5): p. 649.
 73. Pescini, F., et al., *A pathogenic mutation on exon 21 of the NOTCH3 gene causing CADASIL in an octogenarian paucisymptomatic patient*. J Neurol Sci, 2008. **267**(1-2): p. 170-3.
 74. Fouillade, C., et al., *Activating NOTCH3 mutation in a patient with small-vessel-disease of the brain*. Hum Mutat, 2008. **29**(3): p. 452.
 75. den Dunnen, J.T. and S.E. Antonarakis, *Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion*. Hum Mutat, 2000. **15**(1): p. 7-12.
 76. Adib-Samii, P., Brice, G., Martin, R.J. and Markus, H.S. *Clinical spectrum of CADASIL and the effect of cardiovascular risk factors on phenotype: study in 200 consecutively recruited individuals*. Stroke. 2010. **41**(4): p.630-4.
 77. De Silva, K.R., Gamage, R., Dunuwille, J., Gunarathna, D., Sirisena, D., Weerasinghe, A., Amarasinghe, P.H., Hosomi, A. and Mizuno, T. *Cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL): a patient from Sri Lanka*. J Clin Neurosci. 2009. **16**(11): p.1492-3.
 78. Lee, Y.C., Liu, C.S., Chang, M.H., Lin, K.P., Fuh, J.L., Lu, Y.C., Liu, Y.F. and Soong, B.W. *Population-specific spectrum of NOTCH3 mutations, MRI features and founder effect of CADASIL in Chinese*. J Neurol. 2009. **256**(2): p.249-55.
 79. Adachi, Y., Doi, K. and Nakashima, K. *CLINICAL FEATURES OF 15 JAPANESE PATIENTS WITH GENETICALLY PROVEN CADASIL*. 2009. Poster, Sixth International Congress on Vascular Dementia.
 80. Ampuero, I., Alegre-Abarrategui, J., Rodal, I., España, A., Ros, R., Sendón, J.L., Galloway, E.G., Cervelló, A., Caminero, A.B., Zabala, A., Erro, E., Jarauta, F., Morlán, L., López-Valdés, E., Aladro, Y., Seijo, M., Rivas, G.G., Muñoz, D.G. and de Yébenes, J.G.. *On the diagnosis of CADASIL*. J Alzheimers Dis. 2009. **17**(4):p.787-94.
 81. Wang, Z., Yuan, Y., Zhang, W., Lv, H., Hong, D., Chen, B., Liu, Y., Luan, X. and Xie S, Wu S. *NOTCH3 mutations and clinical features in 33 mainland Chinese families with CADASIL*. J.Neurol Neurosurg Psychiatry. 2010
 82. Granild-Jensen, J., Jensen U. B., Schwartz M. and Hansen U. S. *Cerebral autosomal*

dominant arteriopathy with subcortical infarcts and leukoencephalopathy resulting in stroke in an 11-year-old male. Developmental Medicine & Child Neurology. 2009. 51:754–757

83. Bentley, P., Wang, T., Malik, O., Nicholas, R., Ban, M., Sawcer, S. and Sharma, P. *CADASIL with cord involvement associated with a novel and atypical NOTCH3 mutation.* J Neurol Neurosurg Psychiatry. 2011 Jan 8. [Epub ahead of print]
84. Rice, C.M., McGuone, D., Kurian, K.M., Love, S., Renowden, S.A., Giffin, N.J. *Autopsy-confirmed, co-existent CADASIL and multiple system atrophy.* Parkinsonism Relat Disord. 2011 Feb 10. [Epub ahead of print]
85. Valenti, R., Bianchi, S., Pescini, F., D'Eramo, C., Inzitari, D., Dotti, M.T., Pantoni, L. *First report of a pathogenic mutation on exon 24 of the NOTCH3 gene in a CADASIL family.* J Neurol. 2011 Mar 16. [Epub ahead of print]

(Επιπλέον, στάλθηκαν ξεχωριστά οι παρακάτω μεταλλάξεις, που παρατίθενται μαζί με τις αντίστοιχες σημειώσεις του κ. Baumann.)

Polymorphic are:

p.Ser497Leu

p.His1133Gln

Uncommon variants are:

p.Glu813Lys (a rare one but the variant has an "rs" number and it has been reported in the ESP database)

p.Ala1057Val (see the previous ones)

Previously unreported (could perhaps be pathogenic):

p.Arg107Trp

p.Arg107Trp and p.His1133Gln are Canadian cases. We have a tiny little bit of DNA of one of them (p.Arg107Trp). We have preserved the sample for possible further studies. According to Hannu's (a male Finnish first-name) email, this patient is GOM negative. The other variant is a polymorphism (the mutation may have been named incorrectly in its origin as mutation 3477C>A cannot trigger change H1133E but change H1133Q instead).