



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

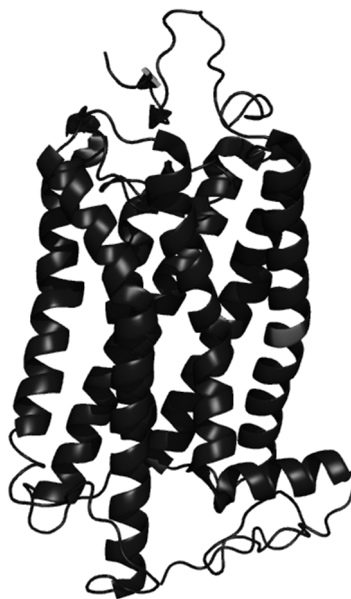
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΒΙΟΛΟΓΙΑΣ ΚΥΤΤΑΡΟΥ & ΒΙΟΦΥΣΙΚΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΜΕΛΕΤΕΣ ΣΕ ΣΥΖΕΥΓΜΕΝΟΥΣ ΜΕ G-ΠΡΩΤΕΪΝΕΣ
ΥΠΟΔΟΧΕΙΣ (GPCRs)**



ΜΑΡΓΑΡΙΤΑ Κ. ΘΕΟΔΩΡΟΠΟΥΛΟΥ

ΑΘΗΝΑ 2014

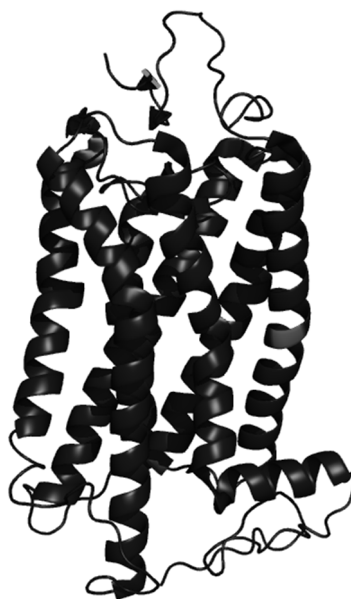


HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCES
FACULTY OF BIOLOGY
DEPARTMENT OF CELL BIOLOGY AND BIOPHYSICS

Ph. D THESIS

COMPUTATIONAL STUDIES ON G-PROTEIN COUPLED RECEPTORS (GPCRs)



MARGARITA C. THEODOROPOULOU
ATHENS 2014

Η έγκριση Διδακτορικής Διατριβής από τη Σχολή Θετικών Επιστημών του
Πανεπιστημίου Αθηνών δε δηλώνει αποδοχή των γνώμων του συγγραφέα.

N. 5343/1932 Άρθρο 202

Συμβουλευτική επιτροπή

- | | |
|---|---|
| 1. Καθηγητής Σταύρος Ι. Χαμόδρακας
(επιβλέπων) | Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών |
| 2. Επίκουρος Καθηγητής Παντελής Γ.
Μπάγκος | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 3. Επίκουρος Καθηγητής Βασίλειος
Προμπονάς | Τμήμα Βιολογικών Επιστημών,
Πανεπιστήμιο Κύπρου |

Εξεταστική επιτροπή

- | | |
|---|---|
| 1. Καθηγητής Σταύρος Ι. Χαμόδρακας
(επιβλέπων) | Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών |
| 2. Επίκουρος Καθηγητής Παντελής Γ.
Μπάγκος | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 3. Επίκουρος Καθηγητής Βασίλειος
Προμπονάς | Τμήμα Βιολογικών Επιστημών,
Πανεπιστήμιο Κύπρου |
| 4. Καθηγητής Κωνσταντίνος Βοργιάς | Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών |
| 5. Καθηγητής Ηλίας Ηλιόπουλος | Τμήμα Γεωπονικής Βιοτεχνολογίας,
Γεωπονικό Πανεπιστήμιο Αθηνών |
| 6. Επίκουρος Καθηγητής Δημήτριος
Στραβοπόδης | Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών |
| 7. Λέκτορας Βασιλική Οικονομίδου | Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών |

“There's no such thing as destiny. There are only different choices. Some choices are easy, some aren't. Those are the really important ones, the ones that define us as people.”

(Walter Sparrow, The Number 23 (2007))

Αφιερωμένο σε όλους όσους
ήταν πάντα εκεί...

Πρόλογος

Η διατριβή αυτή πραγματοποιήθηκε στον Τομέα Βιολογίας Κυττάρου και Βιοφυσικής, του Τμήματος Βιολογίας του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, υπό την επίβλεψη του Καθηγητή και Διευθυντή του Τομέα Σταύρου Χαμόδρακα και μέλη της Τριμελούς Συμβουλευτικής Επιτροπής τον Επίκουρο Καθηγητή Παντελή Μπάγκο και τον Επίκουρο Καθηγητή Βασίλειο Προμπονά.

Ευχαριστώ πολύ τον Καθηγητή μου, Σταύρο Χαμόδρακα, για την ευκαιρία που μου έδωσε να συνεργαστώ μαζί του, ξεκινώντας ως προπτυχιακή φοιτήτρια και συνεχίζοντας τη συνεργασία μας στα πλαίσια του Μ.Δ.Ε. «Βιοπληροφορική», και, τελικά, στα πλαίσια της παρούσας διδακτορικής διατριβής. Οφείλω ένα μεγάλο ευχαριστώ για την εμπιστοσύνη που μου έδειξε σε όλη τη διάρκεια της συνεργασίας μας, το χρόνο του που μου προσέφερε απλόχερα, τη δυνατότητα να γνωρίσω σημαντικούς επιστήμονες, αλλά, κυρίως, να συνεργάζομαι με μεταπτυχιακούς και προπτυχιακούς φοιτητές σε ένα ευχάριστο περιβάλλον. Τον ευχαριστώ ακόμα για την καθοδήγησή του και τις πολύτιμες συμβουλές του για θέματα επιστημονικά και μη, αλλά, κυρίως, γιατί, μέσα από τη δική του αγάπη για την έρευνα και την επιστήμη, με έκανε και μένα να τα αγαπήσω και να καλλιεργήσω τη σκέψη και την προσωπικότητά μου.

Ευχαριστώ θερμά τον Επίκουρο Καθηγητή Παντελή Μπάγκο, ο οποίος, ήδη από τη διπλωματική μου εργασία σε προπτυχιακό επίπεδο, ήταν πάντα εκεί με συμβουλές και καθοδήγηση σε διάφορα επιστημονικά και μη θέματα. Τον ευχαριστώ ιδιαιτέρως για το χρόνο του και τις συζητήσεις μας όλα αυτά τα χρόνια, αλλά και για τις γνώσεις που μου μετέδωσε.

Ένα μεγάλο ευχαριστώ, οφείλω επίσης στον Επίκουρο Καθηγητή Βασίλειο Προμπονά, ο οποίος ήταν πάντα διαθέσιμος και μου πρόσφερε σημαντικές συμβουλές και ιδέες. Οι τελευταίες ήρθαν συχνά σε δύσκολες πνευματικά περιόδους και μου έδωσαν νέα ώθηση.

Η παρουσία των αξιότιμων μελών της Επταμελούς Εξεταστικής Επιτροπής, του Καθηγητή Κωνσταντίνου Βοργιά, του Καθηγητή Ηλία Ηλιόπουλου, του Επίκουρου Καθηγητή Δημητρίου Στραβοπόδη και, τέλος, της Λέκτορος Βασιλικής Οικονομίδου, αλλά και ο πολύτιμος χρόνος που διέθεσαν, με τιμούν και τους ευχαριστώ ιδιαιτέρως.

Νιώθω την ανάγκη να ευχαριστήσω όλα τα μέλη της ερευνητικής ομάδας του Καθηγητή Σταύρου Χαμόδρακα, με τα οποία συνεργάστηκα άψογα όλα αυτά τα χρόνια. Οφείλω να αναφερθώ ονομαστικά στη Δρ. Ζωή Λίτου, η οποία σήμερα είναι φίλη μου, κυρίως. Όλα

αυτά τα χρόνια ήταν πάντα δίπλα μου για να κάνει την καθημερινότητα στο εργαστήριο και τις διάφορες διαδικασίες ευκολότερες. Την ευχαριστώ θερμά για τη συνεργασία μας, τις συμβουλές και τις συζητήσεις μας. Χωρίς την παρουσία της πολλά πράγματα θα ήταν ανέφικτα. Ένα θερμό ευχαριστώ οφείλω και στο συμφοιτητή και καλό μου φίλο κ. Γιώργο Τσαούση, ο οποίος όλα αυτά τα χρόνια ήταν, κυριολεκτικά, δίπλα μου. Μαζί ξεκινήσαμε όλη αυτή την περιπέτεια και νομίζω ότι, αν δεν τον είχα συνοδοιπόρο, δεν θα μπορούσα ποτέ να φτάσω εδώ. Η λογική και ο χαρακτήρας του με βοήθησαν να βρίσκω πάντα ισορροπία, ακόμα και στις πιο δύσκολες στιγμές.

Η Λέκτορας Βασιλική Οικονομίδου και ο Δρ. Νικόλαος Παπανδρέου είναι δύο άνθρωποι με τους οποίους συνεργάστηκα συχνά κυρίως σε θέματα που δεν αφορούν άμεσα την εργασία μου. Τους ευχαριστώ ιδιαίτερα γιατί μου πρόσφεραν απλόχερα τη βοήθειά τους και τις γνώσεις τους σε διάφορα ζητήματα, αλλά, κυρίως, γιατί συνέβαλαν ενεργά στην εύρυθμη λειτουργία του εργαστηρίου.

Σε αυτό το σημείο θέλω να ευχαριστήσω δύο πρώην μέλη του εργαστηρίου, τον κ. Κώστα Τσιρίγο και τον κ. Νικόλαο Πανούση. Τους ευχαριστώ ιδιαίτερα για τη συνεργασία που είχαμε και τη βοήθειά τους σε ό,τι χρειάστηκα, αλλά κυρίως γιατί είναι και αυτοί πλέον φίλοι μου.

Οι Υποψήφιοι Διδάκτορες Παρασκευή Τσιολάκη και Νικόλαος Λούρος είναι δύο ακόμη άνθρωποι, τους οποίους θα ήθελα να ευχαριστήσω, για το χρόνο που περνάμε καθημερινά μαζί. Ιδιαίτερες ευχαριστίες οφείλω στην Παρασκευή Τσιολάκη για το χρόνο που διέθεσε στην επιμέλεια της παρούσας διατριβής.

Οφείλω επίσης ευχαριστίες, σε πολλούς προπτυχιακούς και μεταπτυχιακούς διπλωματικούς φοιτητές με τους οποίους συνεργάστηκα. Στον Ευθύμιο Παπασπύρου, τον Χρήστο Σταμπολάκη, την Αλίνα Λογαρίδη, τη Βασιλική Κωστίου, τον Ηλία Λαβίδα, τη Μαρία Καλτρεμτζίου, τον Φώτη Μπαλτούμα, τη Ζωή Ιωαννίδου, τη Μαρία Κατσαντώνη, την Αναστασία Καλαβρυτινού, τη Δώρα Σερδάρη, τη Βάσια Χονδρού, καθώς και όλους τους υπόλοιπους φοιτητές του εργαστηρίου.

Ευχαριστίες πρέπει να απευθύνω στο Τμήμα Βιολογίας του Πανεπιστημίου Αθηνών και στον Τομέα Βιολογίας Κυττάρου και Βιοφυσικής και όλα τα μέλη του Τομέα (ΔΕΠ, ερευνητές και διοικητικό προσωπικό). Η εύρυθμη λειτουργία του Τομέα αποτέλεσε ένα καθοριστικής σημασίας στοιχείο για την απρόσκοπτη διεξαγωγή της διατριβής.

Τέλος, οφείλω ένα τεράστιο ευχαριστώ στην οικογένειά μου και τους φίλους μου και, ειδικά στον Άγγελο, για τη συμπαράσταση, την κατανόηση και την αγάπη τους, αλλά και γιατί βοήθησαν να γίνω ο άνθρωπος που είμαι σήμερα. Δεν υπάρχουν αρκετές λέξεις για να περιγράψω το πόσο σημαντικοί είστε για μένα.

Η παρούσα έρευνα έχει συγχρηματοδοτηθεί από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης - ΕΤΠΑ) και από εθνικούς πόρους μέσω του Επιχειρησιακού Προγράμματος "Ανταγωνιστικότητα & Επιχειρηματικότητα" του Εθνικού Στρατηγικού Πλαισίου Αναφοράς (ΕΣΠΑ) (Κωδικός Έργου 09ΣΥΝ-13-999).

Περιεχόμενα

Πρόλογος.....	I
Περιεχόμενα.....	V
Περίληψη.....	VII
Abstract	XIII
Μέθοδοι πρόγνωσης, Βάσεις δεδομένων, Δημοσιεύσεις	XVII
1. Εισαγωγή	1
1.1 Μεταγωγή σήματος	1
1.2 Συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs).....	6
1.2.1 Γενικά.....	6
1.2.2 Συστήματα ταξινόμησης των GPCRs	7
1.2.3 Δομικά χαρακτηριστικά των GPCRs	9
1.2.4 Λειτουργικά χαρακτηριστικά των GPCRs	13
1.2.5 Υποκαταστάτες (Ligands).....	16
1.2.6 Εξέλιξη των GPCRs.....	19
1.3 G-πρωτεΐνες.....	22
1.3.1 Γενικά.....	22
1.3.2 Ταξινόμηση των G-πρωτεϊνών	23
1.3.3 Δομή των G-πρωτεϊνών	27
1.3.4 Κύκλος ενεργοποίησης G-πρωτεϊνών και ο ρόλος των πρωτεϊνών RGS	31
1.3.5 Η διεπαφή μεταξύ GPCRs και G-πρωτεϊνών	33
1.3.6 Αλληλεπιδράσεις G-πρωτεϊνών με εκτελεστές.....	35
1.3.7 Η εξέλιξη των G-πρωτεϊνών	36
1.4 Βάσεις Δεδομένων και υπολογιστικά εργαλεία για GPCRs, G-πρωτεΐνες, και τους εκτελεστές τους.....	40
1.5 Στόχοι της Διδακτορικής Διατριβής	42
2. Υλικά και Μέθοδοι	45
2.1 Αλγόριθμοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών.....	45
2.1.1 Hidden Markov Models και profile HMMs.....	45
2.1.2 Η μέθοδος GPCRpipe.....	53
2.1.3 Η μέθοδος GprotPRED.....	57
2.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών.....	60
2.2.1 Η βάση δεδομένων Human-gpDB	60
2.2.2 Η βάση δεδομένων PLHG-DB	62
2.3 Μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες.....	64

2.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες.....	66
2.4.1 Μονονουκλεοτιδικοί πολυμορφισμοί (SNPs).....	66
2.4.2 Συλλογή δεδομένων.....	68
3. Αποτελέσματα και Συζήτηση	71
3.1 Αλγόριθμοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών	71
3.1.1 Η μέθοδος GPCRipe	71
3.1.2 Η μέθοδος GprotPRED	79
3.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών	86
3.2.1 Η βάση δεδομένων Human-gpDB.....	86
3.2.2 Η βάση δεδομένων PLHG-DB.....	89
3.3 Μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες	92
3.3.1 Αλληλεπικαλυπτόμενες περιοχές αλληλεπιδράσεων στην επιφάνεια των Gα	92
3.3.2 Μεταβολές στη δομή κατά την ενεργοποίηση των Gα	96
3.3.3 Δομική ετερογένεια μεταξύ των διαφορετικών οικογενειών των Gα	98
3.3.4 Η πιθανή επιρροή του ηλεκτροστατικού δυναμικού στις αλληλεπιδράσεις των G-πρωτεϊνών	98
3.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες.....	101
3.4.1 Στατιστική επεξεργασία των αποτελεσμάτων	101
3.4.2 Οπτικοποίηση SNPs σε επίπεδο δευτεροταγούς και τριτοταγούς δομής.....	104
4. Συμπεράσματα	107
4.1 Μέθοδοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών	107
4.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών	108
4.3 Οι αλληλεπιδράσεις των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες	109
4.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες.....	110
4.5 Μελλοντικές κατευθύνσεις	111
Βιβλιογραφία.....	113
Παράρτημα I	133
Δημοσιεύσεις	153

Περίληψη

Οι συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs) αποτελούν την πολυπληθέστερη και την πλέον ποικιλόμορφη υπεροικογένεια διαμεμβρανικών υποδοχέων στους ευκαρυωτικούς οργανισμούς. Οι περισσότερες λειτουργίες τους επιτελούνται μέσω μιας ειδικής ομάδας πρωτεϊνών που ονομάζονται G-πρωτεΐνες και λειτουργούν ως «διακόπτες» για τη μεταγωγή μηνυμάτων από τον εξωκυττάριο χώρο στο εσωτερικό του κυττάρου μέσω της αλληλεπίδρασής τους με τους GPCRs. Οι G-πρωτεΐνες στη συνέχεια αλληλεπιδρούν με διάφορα μόρια-εκτελεστές (effectors) με αποτέλεσμα άμεσες αλλαγές στις συγκεντρώσεις κυτταρικών μορίων-σηματοδοτών. Οι GPCRs παίζουν σημαντικό ρόλο στην υγεία και τις ασθένειες και, λόγω της εξαιρετικής τους σημασίας, αποτελούν βασικούς στόχους της φαρμακευτικής έρευνας.

Στα πλαίσια της παρούσας διατριβής μελετήθηκε εκτενώς, με υπολογιστικές μεθόδους, η πολύ ενδιαφέρουσα υπεροικογένεια των GPCRs, καθώς και οι G-πρωτεΐνες. Αναπτύχθηκαν δύο μέθοδοι: η μία αφορά τον εντοπισμό από την αλληλουχία και μόνο και το λεπτομερή σχολιασμό πιθανών συζευγμένων με G-πρωτεΐνες υποδοχέων (GPCRpipe), ενώ η μέθοδος GprotPRED έρχεται να συμπληρώσει την προηγούμενη μέθοδο, καθώς εντοπίζει και ταξινομεί σε οικογένειες πιθανές G-πρωτεΐνες.

Η μέθοδος GPCRpipe (<http://bioinformatics.biol.uoa.gr/GPCRpipe/>) χαρακτηρίζει πρωτεΐνες ως πιθανούς GPCRs με μόνη πληροφορία την αλληλουχία, και στηρίζεται σε ειδικά μοτίβα που περιγράφονται από pHMMs της βάσης Pfam (Punta, et al., 2012), καθώς και σε ένα HMM το οποίο έχει κατασκευαστεί ειδικά για τους GPCRs και επιτρέπει την πρόγνωση της ακριβούς τους τοπολογίας. Το μοντέλο αποτελεί τροποποίηση του HMM που έχει προταθεί από τη μέθοδο HMM-TM (Bagos et al, 2006) και επιτρέπει το αμινοτελικό άκρο να είναι μόνο στον εξωκυττάριο και το καρβόξυτελικό στον κυτταροπλασματικό χώρο, προτυποποιώντας με τον τρόπο αυτό την κοινή τοπολογία των συγκεκριμένων υποδοχέων. Η μέθοδος προσφέρει επιπλέον λεπτομερή σχολιασμό των πιθανών GPCRs και συγκεκριμένα: (α) την ειδικότητα “σύζευξης” του υποδοχέα με συγκεκριμένες οικογένειες G-πρωτεϊνών με τη χρήση του αλγορίθμου PRED-COUPLE2 (Sgourakis, et al., 2005), (β) την τοπολογία μέσω του HMM, της εμφάνισης συγκεκριμένων pHMMs της Pfam που εμφανίζονται ειδικά ενδοκυτταρικά ή εξωκυτταρικά, και της θέσης του πεπτιδίου οδηγητή μέσω του αλγορίθμου SignalP 4.0 (Petersen, et al., 2011) και (γ) την οικογένεια στην οποία

ανήκει ο υπό μελέτη υποδοχέας με βάση το χαρακτηριστικό ρHMM της Pfam, από τη βιβλιοθήκη των ειδικών για GPCRs profiles.

Η μέθοδος GPCRripe αποτελεί ένα αξιόπιστο εργαλείο στην ανίχνευση πιθανών συζευγμένων με G-πρωτεΐνες υποδοχέων σε νέα υπό μελέτη πρωτεώματα καθώς παρουσιάζει υψηλή ευαισθησία (97%) και ειδικότητα (99%), ενώ προσφέρει ταυτόχρονα και σημαντικές πληροφορίες για την πιθανή λειτουργία του υποδοχέα.

Η μέθοδος GprotPRED (<http://bioinformatics.biol.uoa.gr/GprotPRED/>) σχεδιάστηκε για να εντοπίζει με ταχύτητα και αξιοπιστία G-πρωτεΐνες με μόνη πληροφορία την αλληλουχία τους. Ειδικότερα, με τα ειδικά κατασκευασμένα ρHMMs που περιλαμβάνει, μπορεί να εντοπίσει Gα υπομονάδες και μάλιστα να τις ταξινομήσει στις τέσσερις βασικές οικογένειες που συναντώνται στα θηλαστικά (G_s , $G_{i/o}$, $G_{q/11}$, $G_{12/13}$), χωρίς όμως να αποκλείονται Gα υπομονάδες που δεν ανήκουν σε κάποια από αυτές τις οικογένειες (αυτό επιτυγχάνεται με τη χρήση του γενικού ρHMM της Pfam για την Gα υπομονάδα με κωδικό PF00503). Επιπλέον, με τα δύο ειδικά κατασκευασμένα, από την εργασία μας, ρHMMs για τη β και τη γ υπομονάδα, ο εντοπισμός G-πρωτεΐνων είναι ολοκληρωμένος. Η μέθοδος GprotPRED είναι μία αξιόπιστη μέθοδος, καθώς οι τιμές ειδικότητας και ευαισθησίας και για τα έξι ρHMMs είναι 100%.

Οι δύο μέθοδοι που αναπτύχθηκαν, αποτελούν χρήσιμα εργαλεία για τον αυτοματοποιημένο σχολιασμό νέων υπό μελέτη πρωτεωμάτων.

Ο τεράστιος όγκος δεδομένων που αφορούν τους συζευγμένους με G-πρωτεΐνες υποδοχείς οδήγησε στη δημιουργία μιας νέας βάσης δεδομένων. Η Human-gpDB είναι διαθέσιμη στον ιστότοπο http://bioinformatics.biol.uoa.gr/human_gpdb/ και περιλαμβάνει πληροφορίες για ανθρώπινους GPCRs, G-πρωτεΐνες και εκτελεστές και τις μεταξύ τους αλληλεπιδράσεις (Satagoram, et al., 2010). Περιλαμβάνει δεδομένα που ανασύρθηκαν από τη UniProt/SwissProt (UniProt, 2012), καθώς και τις αλληλεπιδράσεις μεταξύ των μορίων μέσω εκτενούς βιβλιογραφικής έρευνας. Η πληροφορία αυτή συνδυάζεται με δεδομένα για γνωστά φάρμακα που αφορούν τις τρεις κατηγορίες μορίων και όλα τα δεδομένα απεικονίζονται με τη βοήθεια προηγμένων τεχνικών απεικόνισης τόσο στις 2 (Medusa, (Hooper and Bork, 2005) όσο και στις 3 διαστάσεις (Arena3D, (Pavlopoulos, et al., 2008). Σκοπός της συγκεκριμένης βάσης δεδομένων είναι να συγκεντρώσει και να παραθέσει στο χρήστη όλη τη διαθέσιμη πληροφορία για το συγκεκριμένο θέμα, μιας και οι GPCRs και τα επαγόμενα από αυτούς μονοπάτια μεταγωγής σήματος είναι εξαιρετικής σημασίας για την

εύρυθμη λειτουργία του ανθρώπινου οργανισμού και σχετίζονται με πολυάριθμες ασθένειες.

Με αφορμή τις ήδη υπάρχουσες βάσεις δεδομένων του εργαστηρίου μας, grDB και Human-grDB, σχεδιάσαμε και υλοποιήσαμε μια νέα βάση δεδομένων που περιλαμβάνει πληροφορία για τους πεπτιδικούς υποκαταστάτες των ανθρώπινων GPCRs. Η PLHG-DB (http://bioinformatics.bio.uoa.gr/plhg_db/) περιλαμβάνει τους πεπτιδικούς υποκαταστάτες των GPCRs ταξινομημένους σε μια ιεραρχία οικογενειών, υπο-οικογενειών και τύπων, η οποία προέκυψε από εκτενή βιβλιογραφική έρευνα και βασίζεται κυρίως στη λειτουργία των πεπτιδίων αυτών. Πληροφορίες που αφορούν την ακολουθία του πεπτιδικού υποκαταστάτη, συχνά τη λειτουργία αυτού, το προπεπτίδιο από το οποίο προέρχεται, καθώς και πληροφορίες για την αλληλεπίδραση με τον υποδοχέα, όλα συνοδευόμενα από αντίστοιχες βιβλιογραφικές αναφορές, παρέχονται στο χρήστη. Μια επιπλέον δυνατότητα που προσφέρεται είναι η περιήγηση στα δεδομένα της βάσης, είτε ακολουθώντας την ταξινόμηση των πεπτιδικών υποκαταστατών, ή ακολουθώντας την ταξινόμηση των υποδοχέων με τους οποίους οι υποκαταστάτες αλληλεπιδρούν.

Επιθυμώντας να μελετήσουμε από δομική σκοπιά τη μεταγωγή σήματος, ύστερα από εκτεταμένη αναζήτηση στη βιβλιογραφία πάνω στις αλληλεπιδράσεις των G-πρωτεϊνών με GPCRs, εκτελεστές και πρωτεΐνες RGS, συλλέξαμε πληροφορία που αφορά τόσο τις προσδιορισμένες κρυσταλλογραφικά δομές, όσο και αποτελέσματα από πειράματα μεταλλαξογένεσης αλλά και *in silico* μελέτες. Παράλληλα, συγκεντρώθηκαν όλες οι κρυσταλλογραφικά προσδιορισμένες δομές των G-πρωτεϊνών και των συμπλόκων τους με άλλες πρωτεΐνες από την PDB. Χρησιμοποιώντας υπολογιστικές μεθόδους που εξειδικεύονται στην ανάλυση κρυσταλλικών δομών για την εύρεση αλληλεπιδράσεων, και σε συνδυασμό με τη βιβλιογραφική έρευνα, προσδιορίσαμε τις περιοχές και τα κατάλοιπα των Gα υπομονάδων που συμμετέχουν σε αλληλεπιδράσεις με GPCRs, εκτελεστές και πρωτεΐνες RGS.

Έχοντας εντοπίσει τις θέσεις αλληλεπιδράσεων στην επιφάνεια των Gα, στη συνέχεια προχωρήσαμε στη σύγκριση των α υπομονάδων. Αλληλουχίες των Gα με προσδιορισμένες κρυσταλλογραφικά δομές ανασύρθηκαν από την UniProt και στοιχήθηκαν για τη μελέτη της ετερογένειας της αμινοξικής ακολουθίας ανάμεσα στις διαφορετικές οικογένειες και υπο-οικογένειες των Gα. Οι διαθέσιμες δομές των Gα στις ενεργές, ανενεργές και μεταβατικές στερεοδιατάξεις τους συγκρίθηκαν αρχικά μέσω δομικών στοιχίσεων. Συγκρίσεις μέσω

δομικών στοιχίσεων πραγματοποιήθηκαν επιπρόσθετα ανάμεσα σε μέλη διαφορετικών οικογενειών των Gα. Επιπλέον, με τη χρήση κατάλληλου λογισμικού, υπολογίστηκε το ηλεκτροστατικό δυναμικό των επιφανειών διαφορετικών Gα υπομονάδων, εκτελεστών από σύμπλοκα με G-πρωτεΐνες και πρωτεϊνών RGS, ενώ συμπληρωματικά προσδιορίστηκαν τα επιφανειακά κατάλοιπα των Gα υπομονάδων. Τέλος, μια σειρά από κρυσταλλικές δομές συμπλόκων Gα – RGS και Gα – εκτελεστών μελετήθηκαν ενεργειακά, και προσδιορίστηκαν οι τιμές των διαμοριακών ενεργειών των αλληλεπιδράσεων μεταξύ τους.

Μελετώντας όλα τα προηγούμενα, αναγνωρίσαμε στην επιφάνεια των Gα υπομονάδων μια σειρά από περιοχές και κατάλοιπα που μπορούν σε πολλές περιπτώσεις να αλληλεπιδράσουν τόσο με GPCRs όσο και με εκτελεστές. Η σύγκριση ενεργών και ανενεργών Gα υπομονάδων δείχνει πως οι συγκεκριμένες περιοχές διαφοροποιούνται κατά την ενεργοποίηση των G-πρωτεϊνών, γεγονός που θα μπορούσε να επηρεάζει το ρόλο τους στις αλληλεπιδράσεις των Gα. Η σύγκριση της αλληλουχίας και της δομής αυτών των περιοχών ανάμεσα σε διαφορετικές Gα υπομονάδες υποδεικνύει πως η ετερογένειά τους μπορεί να επηρεάσει τις αλληλεπιδράσεις τους με εκτελεστές και ίσως με υποδοχείς. Τέλος, η αξιοσημείωτη ετερογένεια των ηλεκτροστατικών ιδιοτήτων που παρατηρείται όχι μόνο ανάμεσα στις διαφορετικές οικογένειες των Gα υπομονάδων αλλά ακόμη και ανάμεσα σε μέλη της ίδιας οικογένειας, η συμπληρωματικότητά τους με τις ηλεκτροστατικές ιδιότητες των αντίστοιχων αλληλεπιδρώντων με αυτές εκτελεστών και πρωτεϊνών RGS, αλλά και η ενεργειακή ανάλυση των συμπλόκων τους, προτείνουν πως το ηλεκτροστατικό δυναμικό είναι ίσως καθοριστικής σημασίας παράγοντας στις αλληλεπιδράσεις των G-πρωτεϊνών.

Στα τελικά στάδια της παρούσας διδακτορικής διατριβής, μελετήθηκαν οι παρερμηνεύσιμοι μονονουκλεοτιδικοί πολυμορφισμοί των ανθρώπινων συζευγμένων με G-πρωτεΐνες υποδοχέων της Κατηγορίας A και η συσχέτισή τους με ασθένειες. Συγκεντρώθηκαν 650 άνθρωποι, class A GPCRs, από τις βάσεις δεδομένων UniProtKB/SwissProt (UniProt, 2012) και RefSeq (Pruitt, et al., 2012) του NCBI και 21746 παρερμηνεύσιμοι SNPs που εντοπίστηκαν σε αυτούς μέσω των βάσεων dbSNP του NCBI (Sherry, et al., 2001), UniProtKB/SwissProt, SNPdbe (Schaefer, et al., 2012) και ClinVar (Riggs, et al., 2013). Εξαιρέθηκαν οι Olfactory υποδοχείς, λόγω μεγάλης ομοιογένειας θα αποτελούσαν «θόρυβο» σε μία στατιστική ανάλυση, και προέκυψε ένα σύνολο 283 GPCRs και 9662 SNPs που κατηγοριοποιήθηκαν με βάση τη θέση τους στην πρωτεϊνική ακολουθία και με βάση την τοπολογία της εκάστοτε πρωτεΐνης. Ύστερα από στατιστική μελέτη του ποσοστού των παρερμηνεύσιμων

μονονουκλεοτιδικών πολυμορφισμών, με κριτήριο την τοπολογία του υποδοχέα, εντοπίσαμε μία «προτίμηση» εμφάνισής τους σε συγκεκριμένες τοπολογικές περιοχές (ενδοκυττάριοι βρόχοι), ενώ σε κάποιες άλλες υπάρχει στατιστικά μειωμένη εμφάνιση αυτών (διαμεμβρανικά τμήματα). Θα είχε ενδιαφέρον η σύγκριση της υπερεμφάνισης ή της υποεμφάνισης παθολογικών SNPs σε συγκεκριμένες τοπολογικές περιοχές, όμως προς το παρόν το σύνολο των δεδομένων δεν επιτρέπει την εξαγωγή στατιστικά σημαντικών αποτελεσμάτων. Για 39 από τους υποδοχείς βρέθηκαν 441 παρερμηνεύσιμοι SNPs οι οποίοι σχετίζονται με ασθένειες. Σε αυτό το σύνολο έγινε οπτικοποίηση των SNPs (διάκριση των σχετιζόμενων με ασθένειες από τους μη σχετιζόμενους) σε διαγράμματα δευτεροταγούς δομής και σε τρισδιάστατα μοντέλα και παρατηρήθηκε σε συγκεκριμένους υποδοχείς μια «συγκέντρωση» των πολυμορφισμών σε ορισμένες περιοχές στο χώρο. Η παρατήρηση αυτή πρέπει να αξιολογηθεί και να ποσοτικοποιηθεί, γεγονός που θα αποτελέσει αντικείμενο περαιτέρω μελέτης.

Abstract

G-protein coupled receptors (GPCRs) are the largest and most diverse superfamily of transmembrane receptors in eukaryotes. Most GPCR functions are carried out by a special group of proteins called G-proteins. G-proteins act as "switches" for signal transduction from the extracellular space into the cell, through their interaction with GPCRs. G-proteins interact with various effector molecules to immediately change the concentration of cellular molecules, leading eventually to a wide range of cellular and physiological responses. GPCRs play important roles in health and disease and, due to their significance, are major pharmacological targets.

In the context of this thesis, using computational methods, we studied extensively, the very interesting superfamily of GPCRs and their partners, G-proteins. Two methods were developed: GPCRpipe detects, using solely the sequence, and annotates probable G-protein coupled receptors, while GprotPRED complements the aforementioned method by identifying and classifying probable G-proteins.

GPCRpipe (<http://bioinformatics.biol.uoa.gr/GPCRpipe/>) characterizes proteins as probable GPCRs, using only their sequence as input. It is based on specific patterns described by Pfam pHMMs (Punta, et al., 2012) and on a specially designed by our work GPCR specific HMM, which allows the prediction of their topology. This HMM is a modified version of the one used in the HMM-TM algorithm (Bagos et al, 2006) and, by allowing the N-terminus only on the extracellular and the C-terminus on the cytoplasmic space, it models effectively the common topology of GPCRs. Additionally, GPCRpipe implements further important information regarding the function of probable GPCRs: (a) coupling specificity of the receptor with specific G-proteins using the PRED-COUPLE2 algorithm (Sgourakis, et al., 2005), (b) topology provided by the HMM, the occurrence of Pfam profiles domains with known extra- or intra-cellular localization and the signal peptide's location provided by SignalP 4.0 (Petersen, et al., 2011) and (c) family annotation using the best hit pHMM from the Pfam library.

GPCRpipe is a reliable method for the discrimination of GPCRs in proteomes, with approximately 97% sensitivity and 99% specificity, while offering at the same time important information for the function of the receptor.

GprotPRED (<http://bioinformatics.biol.uoa.gr/GprotPRED/>) was designed to accurately detect G-proteins with solely their sequence as input. Using the specific pHMMs that were

built, it detects G α proteins, and classifies them in the four basic mammal families (G_s, G_{i/o}, G_{q/11}, G_{12/13}). With the use of the general G-alpha pHMM of Pfam (PF00503), it has the ability to also detect G α proteins not belonging in one of the four families mentioned above. Moreover, with two additional pHMMs for the β and γ subunits, identification of G-proteins is complete. GprotPRED is a reliable method with 100% sensitivity and 100% specificity for all six pHMMs.

Both GPCRpipe and GprotPRED are useful tools in the annotation of newly discovered proteomes.

The sheer volume of data on G-coupled protein receptors led to the creation of a new database. Human-gpDB is available at http://bioinformatics.biol.uoa.gr/human_gpdb/ and contains information regarding human GPCRs, G-proteins, effectors and their interactions (Satagopam, et al., 2010). All sequence data were retrieved from the UniProt/Swiss-Prot (UniProt_Consortium, 2010), while the interaction data were manually gathered through an extensive literature search. All data is then enriched with information about interacting drugs for all three types of molecules and are visualized using advanced visualization techniques both in 2D (Medusa, (Hooper and Bork, 2005) and 3D (Arena3D, (Pavlopoulos, et al., 2008). The goal of Human-gpDB is to make the exploration of information more efficient and the extraction of novel knowledge easier.

As an addition to the databases, previously developed, in our lab, gpDB and Human-gpDB, we designed and implemented a new database for the peptide ligands of human GPCRs. PLHG-DB (http://bioinformatics.bio.uoa.gr/plhg_db/) includes all peptide ligands classified on a hierarchy of families, subfamilies and types. This classification system resulted from an extensive literature search and is based on the function of the peptide ligands. Information on the sequence of the peptide ligand, its function, the precursor protein and the interaction with the respective receptor, all accompanied by the respective citations, are presented to the user. An additional feature of PLHG-DB is the ability to navigate through two different pathways. The one is following Peptide Ligands Classification and the other GPCRs classification that leads eventually to the peptide ligands interacting with each GPCR subfamily.

In an effort to study signal transduction from a structural perspective, we performed an extensive literature search on G α interactions with GPCRs and effectors, gathering information from crystal structures, mutagenesis experiments and computational studies.

We then retrieved all relevant structures from PDB and tried to identify the regions and specific amino acids of G α subunits that take part in the interaction with GPCRs, effectors and RGS proteins. To achieve this goal we used computational methods specialized in finding interactions in crystal structures and all available data from our literature research.

After identifying the interaction sites on the G α surface, we compared G α subunit sequences and structures. We aligned the sequences of all G α subunits with available structures, to study the diversity of amino acid sequence among the different families and subfamilies of G α 's. Also, structural alignments were made in order to compare the active, inactive and intermediate forms of G α subunits and, also, the members of different G α families. In addition, the electrostatic potential of different G α subunits, effectors and RGS proteins that interact with G α 's was calculated, while all surface amino acid residues of G α subunits were identified. Finally, a series of energy calculations on known G α – RGS and G α – effector complexes were conducted and the intermolecular energies were calculated.

We have identified certain surfaces of G α subunits that may, in many cases, participate in binding both receptors and effectors. The comparison between active and inactive G α subunits shows that these surfaces shift during activation and maybe this affects their role in G α interactions. The differences displayed in the sequence and structure of these sites may perhaps account for G α specificity towards their binding partners. Furthermore, the diversity in the electrostatic potential of G α surfaces, combined with observed electrostatic properties of various effectors and RGS structures, suggests that electrostatic complementarity is, most probably, an important factor in the regulation of effectors by G-proteins, as well as G α interactions with RGS proteins.

In the final stages of this thesis, we studied all missense SNPs on human, class A GPCRs and their associations with diseases. A set of 650 human class A GPCRs from UniProtKB/SwissProt (UniProt, 2012) and RefSeq (Pruitt, et al., 2012) databases along with 21746 missense SNPs through dbSNP (Sherry, et al., 2001), UniProtKB/SwissProt, SNPdbe (Schaefer, et al., 2012) and ClinVar (Riggs, et al., 2013) was created. Excluding the Olfactory receptors, a set of 283 GPCRs and 9662 SNPs was created and all SNPs were categorized based on the topological domain on which they appear. Studying the frequencies of occurrence of missense SNPs in each topological domain, a tendency for the SNPs to be more abundant in some domains (cytoplasmic loops), and less abundant in others (transmembrane segments) was observed. It would be of great interest to compare the abundance of

pathogenic SNPs in certain topological domains, yet the lack of sufficient data does not allow statistically significant results to be delivered. Consolidation of the data disclosed a total of 441 SNPs, located in 39 receptors, which were found to have a clinical impact or/and an association with disease. For these receptors, both secondary structure diagrams and 3D models were created and all SNPs were mapped on them. In some of the constructed models, pathogenic SNPs tend to accumulate in certain regions. This observation will be studied and further evaluated, in the near future.

Μέθοδοι πρόγνωσης, Βάσεις δεδομένων, Δημοσιεύσεις

Μέθοδοι πρόγνωσης που αναπτύχθηκαν κατά τη διάρκεια της διδακτορικής διατριβής:

- GPCRpipe (<http://bioinformatics.biol.uoa.gr/GPCRpipe>)
- GprotPRED (<http://bioinformatics.biol.uoa.gr/GprotPRED>)

Βάσεις δεδομένων που αναπτύχθηκαν κατά τη διάρκεια της διδακτορικής διατριβής:

- Human-gpDB (http://bioinformatics.biol.uoa.gr/human_gpdb)
- PLHG-DB (http://bioinformatics.biol.uoa.gr/plhg_db)

Αναφέρονται οι δημοσιεύσεις που έχουν προκύψει στα πλαίσια της παρούσας διδακτορικής διατριβής:

1. Satagopam* VP, Theodoropoulou* MC, Stampolakis CK, Pavlopoulos GA, Papandreou NC, Bagos PG, Schneider R, Hamodrakas SJ. **GPCRs, G-proteins, effectors and their interactions: human-gpDB, a database employing visualization tools and data integration techniques.** *Database (Oxford)* 2010;2010:baq019. (*equally contributing authors)
2. Baltoumas, F.A., Theodoropoulou, M.C., Hamodrakas, S.J. **Interactions of the α -subunits of heterotrimeric G-proteins with GPCRs, effectors and RGS proteins: A critical review and analysis of interacting surfaces, conformational shifts, structural diversity and electrostatic potentials.** *J. Struct. Biol.* (2013), <http://dx.doi.org/10.1016/j.jsb.2013.03.004>
3. Theodoropoulou* M.C., Tsaousis* G.N., Litou Z.I., Bagos P.G., Hamodrakas S.J., **GPCRpipe: A pipeline for the detection of G-protein coupled receptors in proteomes.** 2014 (submitted) (*equally contributing authors)
4. Theodoropoulou, M.C., Logaridi, A.S., Bagos, P.G., Hamodrakas, S.J., **PLHG-DB: a database of peptide ligands of human G-protein coupled receptors.** 2014 (final stages of preparation)
5. Theodoropoulou, M.C., Kostiou, V.D., Hamodrakas, S.J., **GprotPRED: Detection of the four distinct $G\alpha$ families, the $G\beta$ and the $G\gamma$ subunits of G-proteins using profile Hidden Markov Models (pHMMs).** 2014 (final stages of preparation)
6. Katsantoni, M.G., Theodoropoulou, M.C., Hamodrakas, S.J., **Exploring SNPs distribution and associations with disease: a study of human, class A GPCRs.** 2014 (final stages of preparation)
7. Chondrou, V.P., Kalavritinou, A.P., Serdari, T.G., Theodoropoulou, M.C., Hamodrakas, S.J., **Analysis of missense single nucleotide polymorphisms (SNPs) in human G-protein coupled receptors (GPCRs) with pathogenic and non-pathogenic phenotypes.** 2014 (in preparation)

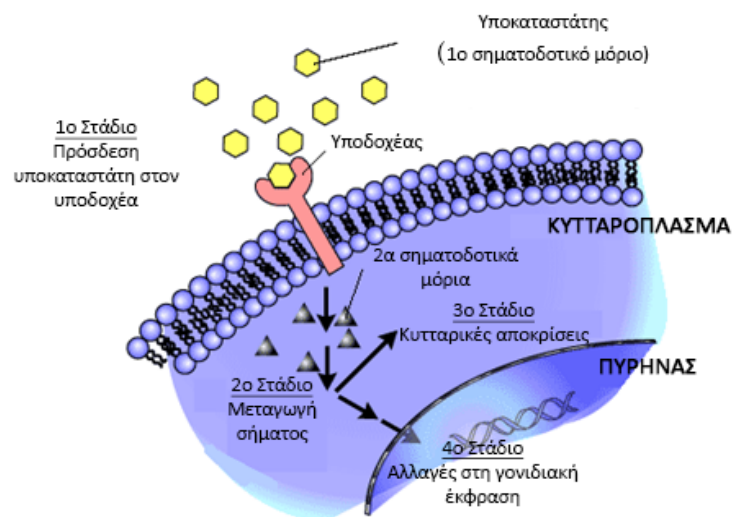
Άλλες δημοσιεύσεις:

1. Theodoropoulou, M.C., Bagos, P.G., Spyropoulos, I.C. and Hamodrakas, S.J., **gpDB: a database of GPCRs, G-proteins, effectors and their interactions.** Bioinformatics, 2008. 24(12): p. 1471-2.

1. Εισαγωγή

1.1 Μεταγωγή σήματος

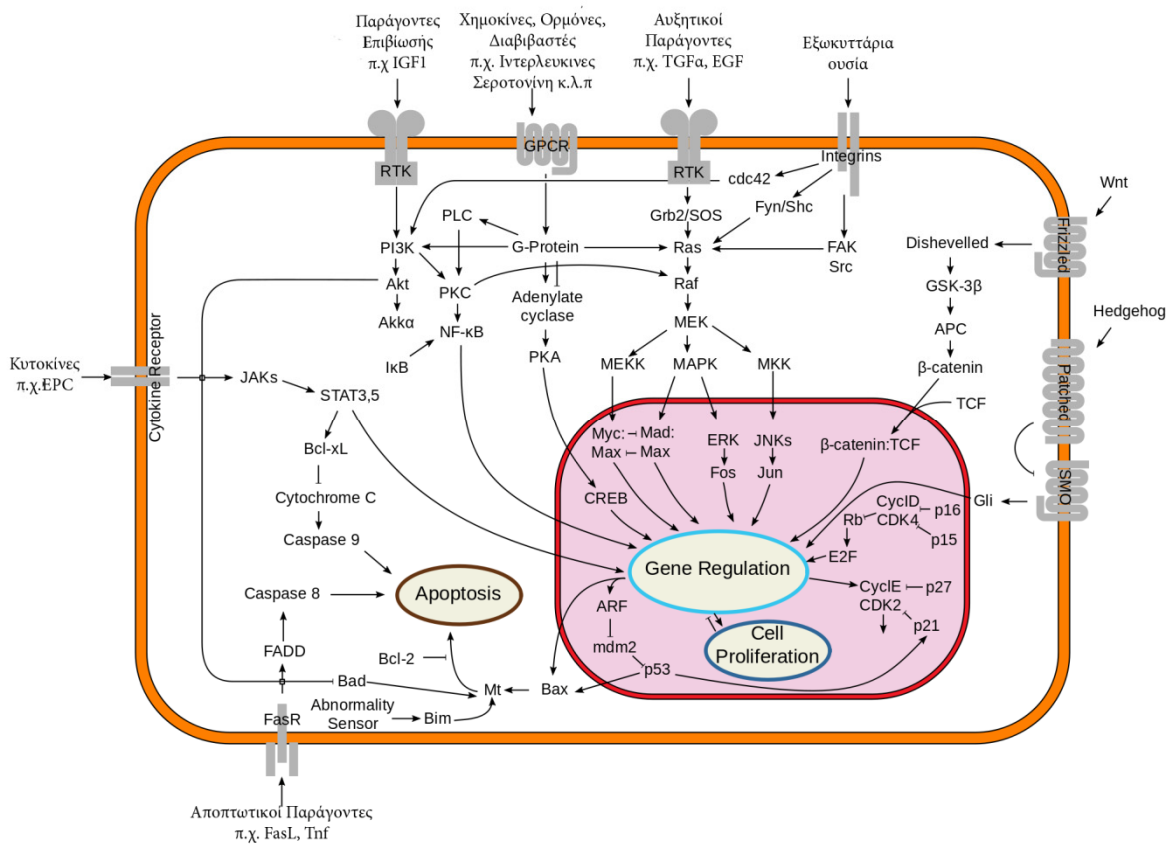
Η μεταγωγή σήματος περιλαμβάνει τις κυτταρικές εκείνες διεργασίες, με τις οποίες ερεθίσματα φυσικής ή χημικής μορφής μπορούν και επάγουν συγκεκριμένες κυτταρικές λειτουργίες, μέσω της ενεργοποίησης επιλεγμένων μοριακών μηχανισμών (Cooper GM, 2007). Οι πορείες μεταγωγής σήματος ακολουθούν μια ευρύτερη κοινή οδό, η οποία μπορεί να θεωρηθεί ως ένα μοριακό κύκλωμα (Berg JM, 2002). Ο μηχανισμός της μεταγωγής σήματος περιλαμβάνει διακριτά στάδια, που ρυθμίζουν εξειδικευμένες κυτταρικές λειτουργίες. Αλλοιώσεις στο μηχανισμό αυτό μπορούν να προκαλέσουν κυτταρικές δυσλειτουργίες και, πιθανώς, ανθρώπινες ασθένειες (Cooper GM, 2007). Στη διαδικασία αυτή πρωταρχικό ρόλο παίζουν υποδοχείς, οι οποίοι αναγνωρίζουν συγκεκριμένα εξωκυτταρικά ερεθίσματα. Ως απόκριση σε αυτά τα ερεθίσματα, οι υποδοχείς ρυθμίζουν τις δραστηριότητες διαφόρων ενδοκυττάρων ρυθμιστικών πρωτεϊνών, οι οποίες με τη σειρά τους επάγουν άλλες κυτταρικές αποκρίσεις (εικόνα 1.1) (Wauson, et al., 2013).



Εικόνα 1.1: Σχηματική αναπαράσταση της μεταγωγής σήματος.

Στους μονοκύτταρους οργανισμούς, η ποικιλία των μονοπατιών μεταγωγής σήματος που μπορεί να πραγματοποιήσει το κύτταρο επηρεάζει την ποικιλία των τρόπων με τους οποίους μπορεί να αντιδρά και να ανταποκρίνεται στο περιβάλλον του (Antunes and Ferreira, 2009).

Στους πολυκύτταρους οργανισμούς, πολυάριθμα μονοπάτια μεταγωγής σήματος απαιτούνται για το συντονισμό της συμπεριφοράς κάθε κυττάρου, ώστε να υποστηρίξει τη λειτουργία του οργανισμού ως όλου (Εικόνα 1.2).



Εικόνα 1.2: Γενική επισκόπηση των μονοπατιών μεταγωγής σήματος.

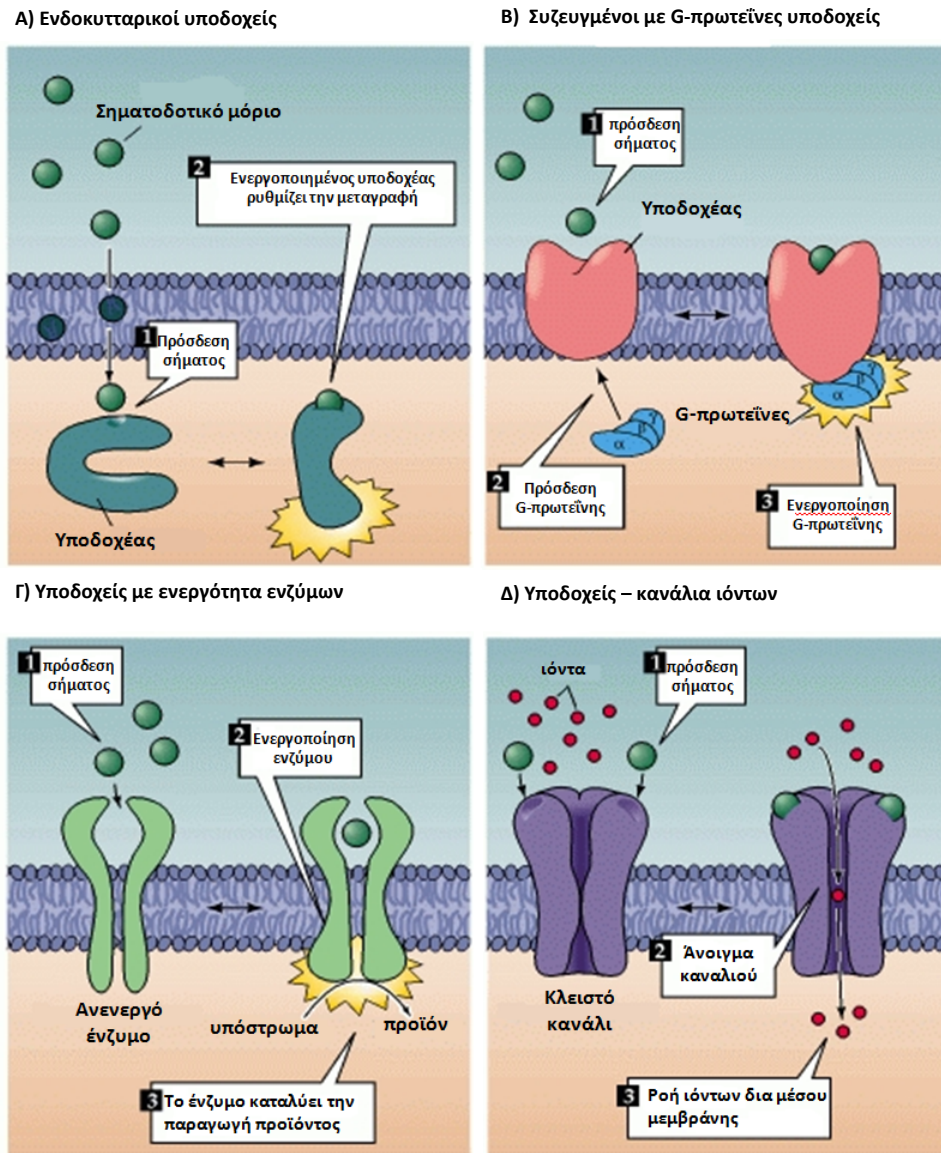
Τα περισσότερα μονοπάτια μεταγωγής σήματος ακολουθούν τη διαδικασία πρόσδεσης μορίων – σημάτων (υποκαταστάτες) σε κατάλληλους υποδοχείς, είτε στην κυτταρική μεμβράνη, είτε στο εσωτερικό του κυττάρου. Οι υποκαταστάτες μπορούν να είναι πρωτεΐνες και πεπτίδια, αμινοξέα, λιπίδια, βιταμίνες, σάκχαρα, ακόμα και ανόργανα μόρια και ιόντα ή/και φωτόνια. Ανάλογα με τη λειτουργία τους και τον τρόπο μεταβίβασης στον οργανισμό, ομαδοποιούνται σε ορμόνες (πρωτεϊνικές ή στεροειδείς), νευροδιαβιβαστές, παρακρινείς ή αυτοκρινείς παράγοντες, σε αυξητικούς παράγοντες, συστατικά της εξωκυττάριας ουσίας, ελεύθερες ρίζες και νευροτροφίνες. Εκτός από τα μόρια – σήματα του ίδιου του οργανισμού, οι μηχανισμοί μεταγωγής μπορούν να ενεργοποιηθούν και από ερεθίσματα του εξωτερικού περιβάλλοντος, που μπορεί να είναι χημικής φύσεως, όπως τα μόρια που προαναφέρθηκαν, ή και φυσικά ερεθίσματα όπως ακτινοβολία, μηχανική καταπόνηση κτλ. (Cooper GM, 2007; Vander AJ, 2001). Η απόκριση στο σήμα μπορεί να οδηγήσει σε μια σειρά από κυτταρικές αποκρίσεις,

όπως έκφραση ή καταστολή συγκεκριμένων γονιδίων, αλλαγές στο μεταβολισμό του κυττάρου, επαγωγή της μίτωσης, ακόμα και κυτταρικό θάνατο.

Η ειδικότητα της κυτταρικής απόκρισης στο σήμα καθορίζεται κυρίως από το είδος του πρωτεϊνικού υποδοχέα (receptor), που εκφράζεται στον κυτταρικό πληθυσμό – στόχο. Οι υποδοχείς είναι πρωτεΐνες που εντοπίζονται στη μεμβράνη, στο κυτταρόπλασμα ή στον πυρήνα του κυττάρου και μπορούν να συνδέονται με βιολογικά σήματα – υποκαταστάτες (ligands), προκαλώντας συνήθως τη διαδοχική ενεργοποίηση ειδικών ενζυμικών αντιδράσεων (Cooper GM, 2007; Vander AJ, 2001).

Οι υποδοχείς κατατάσσονται σε δύο μεγάλες κατηγορίες: τους ενδοκυττάρους (Εικόνα 1.3 Α) και τους υποδοχείς της κυτταρικής επιφάνειας (Εικόνα 1.3 Β-Δ). Οι ενδοκυτταρικοί υποδοχείς χωρίζονται περαιτέρω σε πυρηνικούς και κυτταροπλασματικούς και είναι πρωτεΐνες, οι οποίες εντοπίζονται στην πυρηνική μεμβράνη και το κυτταρόπλασμα αντίστοιχα. Οι υποδοχείς κυτταρικής επιφάνειας είναι διαμεμβρανικές πρωτεΐνες και αναγνωρίζουν την πλειοψηφία των εξωκυτταρικών σηματοδοτικών μορίων (Cooper GM, 2007).

Οι διαμεμβρανικοί υποδοχείς διαπερνούν την κυτταροπλασματική μεμβράνη, έχοντας τελικά ένα τμήμα τους στον εξωκυττάριο χώρο (εξωκυττάρια περιοχή) και ένα στο εσωτερικό του κυττάρου (ενδοκυττάρια περιοχή). Η μεταγωγή σήματος συμβαίνει ως αποτέλεσμα ενός διεγερτικού μορίου ή της πρόσδεσης ενός υποκαταστάτη στην εξωκυττάρια περιοχή. Ο υποκαταστάτης δεν διαπερνά την κυτταροπλασματική μεμβράνη πριν προσδεθεί στον υποδοχέα. Η πρόσδεση του υποκαταστάτη σε έναν επιφανειακό υποδοχέα πυροδοτεί μια σειρά γεγονότων στο εσωτερικό του κυττάρου, με διαφορετικούς τύπους υποδοχέων να οδηγούν σε διαφορετικές κυτταρικές αποκρίσεις. Οι υποδοχείς ανταποκρίνονται τυπικά στην πρόσδεση ενός μόνο συγκεκριμένου υποκαταστάτη. Μετά την πρόσδεση, ο υποκαταστάτης ξεκινά τη μεταγωγή του σήματος διαμέσου της μεμβράνης προκαλώντας, συνήθως, μια αλλαγή στη στερεοδιάταξη του υποδοχέα. Συχνά, αυτές οι αλλαγές στη στερεοδιάταξη έχουν ως αποτέλεσμα την ενεργοποίηση μιας ενζυμικής λειτουργίας του ίδιου του υποδοχέα ή την έκθεση μιας περιοχής πρόσδεσης για άλλες σηματοδοτικές πρωτεΐνες στο εσωτερικό του κυττάρου. Μόλις οι πρωτεΐνες προσδεθούν στον υποδοχέα, μπορεί να ενεργοποιηθούν και να μεταδώσουν το σήμα στο κυτταρόπλασμα (Cooper GM, 2007; Kristiansen, 2004; Oldham and Hamm, 2008; Vander AJ, 2001).



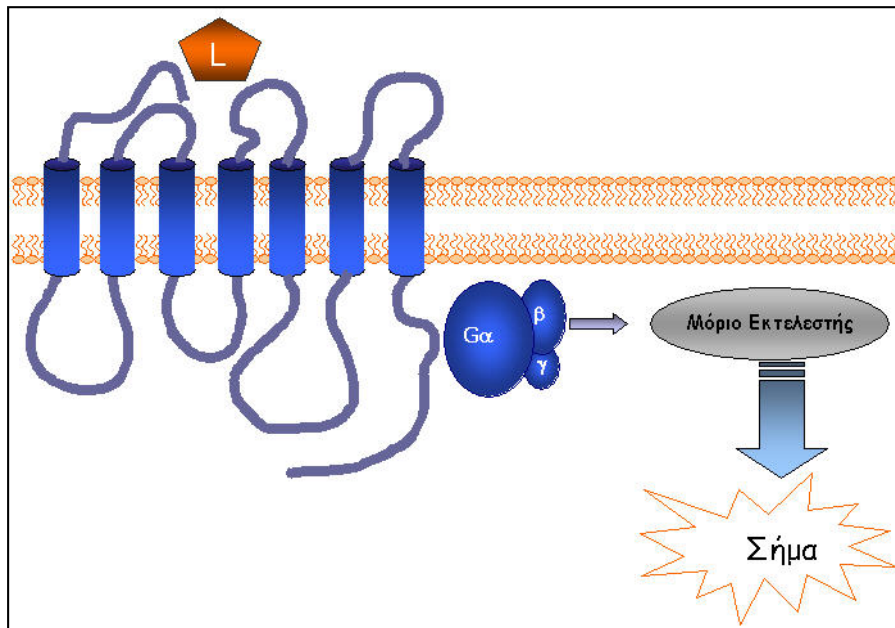
Εικόνα 1.3: Κατηγορίες υποδοχέων στα ευκαρυωτικά κύτταρα.

Στους ευκαρυωτικούς οργανισμούς οι μεμβρανικοί υποδοχείς διακρίνονται σε 3 κύριες κατηγορίες με διακριτά δομικά και λειτουργικά χαρακτηριστικά (Alberts, et al., 2007; Vander AJ, 2001):

1. Τους υποδοχείς που είναι συζευγμένοι με G-πρωτεΐνες (GPCRs) (Εικόνα 1.3 B),
2. Τους υποδοχείς με ενεργότητα ενζύμων (Εικόνα 1.3 Γ), και
3. Τους υποδοχείς με δράση καναλιών ιόντων (Εικόνα 1.3 Δ)

Οι συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs) αποτελούν μια από τις μεγαλύτερες και την πλέον ποικιλόμορφη υπερ-οικογένεια υποδοχέων της κυτταρικής μεμβράνης στους ευκαρυωτικούς οργανισμούς (Lagerstrom and Schioth, 2008; Pierce, et al., 2002). Οι περισσότερες λειτουργίες των GPCRs επιτελούνται μέσω μιας ειδικής ομάδας πρωτεϊνών που

ονομάζονται G-πρωτεΐνες. Αυτές λειτουργούν ως «διακόπτες» για τη μεταγωγή μηνυμάτων από τον εξωκυττάριο χώρο στο εσωτερικό του κυττάρου μέσω της αλληλεπίδρασής τους με τους GPCRs (McCudden, et al., 2005; Oldham and Hamm, 2008). Οι G-πρωτεΐνες, στη συνέχεια, αλληλεπιδρούν με διάφορα μόρια-εκτελεστές (effectors) με αποτέλεσμα άμεσες αλλαγές στις συγκεντρώσεις κυτταρικών μορίων-σηματοδοτών (Εικόνα 1.4) (Kristiansen, 2004; Malbon, 2005; McCudden, et al., 2005; Oldham and Hamm, 2008).



Εικόνα 1.4: Σχηματική αναπαράσταση μεταγωγής σήματος μέσω G-πρωτεϊνών.

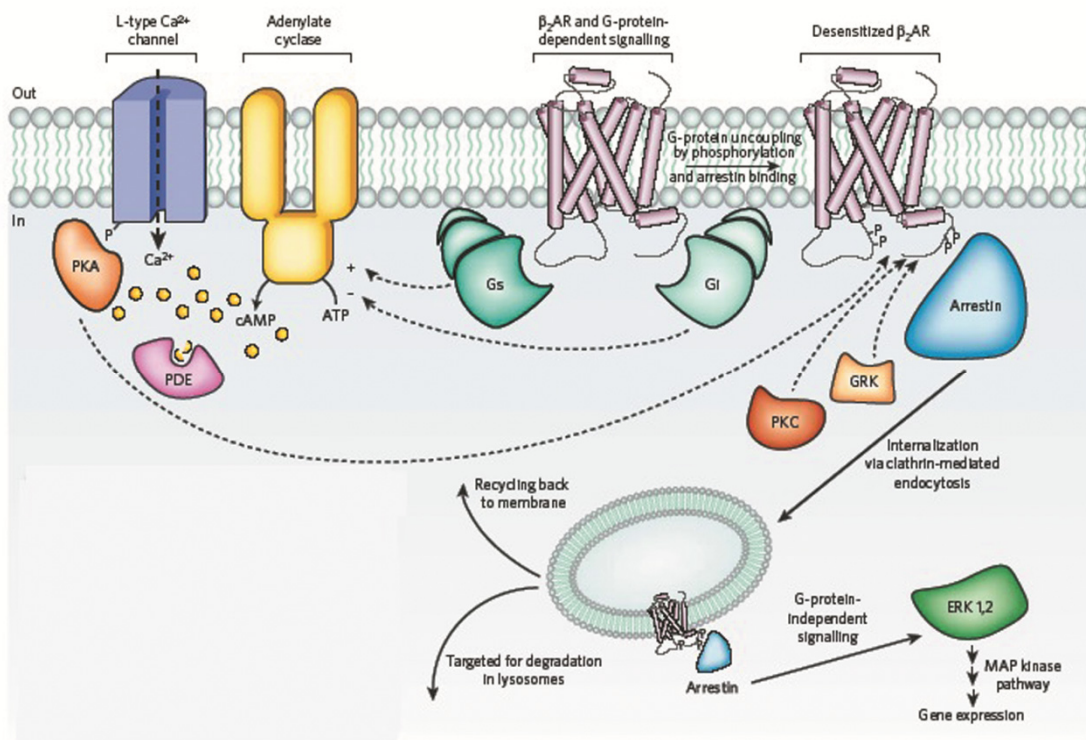
Εξαιτίας της λειτουργίας τους ως κόμβων εισόδου στα μονοπάτια μεταγωγής σήματος των ευκαρυωτικών κυττάρων, οι GPCRs παίζουν σημαντικό ρόλο σε διάφορες ασθένειες (Muller, 2000). Έτσι, εμπλέκονται σε μια μεγάλη ποικιλία παθολογικών καταστάσεων συμπεριλαμβανομένων της κυστικής ίνωσης, του καρκίνου και της μόλυνσης των κυττάρων του ξενιστή μέσω του ιού HIV.

Η δυνατότητα ρύθμισης των μονοπατιών μεταγωγής σήματος μέσω αυτών των μορίων-στόχων οδήγησε σε μια νέα εποχή στη φαρμακολογία. Μάλιστα η σημασία των GPCRs ως κύριων φαρμακευτικών στόχων δεν μπορεί να συγκριθεί με καμία άλλη οικογένεια πρωτεϊνών (Kenakin, 2004; Lin and Civelli, 2004). Αυτό καταδεικνύεται επιπλέον από το γεγονός ότι περίπου το 30% των συνταγογραφούμενων σήμερα φαρμάκων στοχεύει σε GPCRs (Siehler S, 2011).

1.2 Συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs)

1.2.1 Γενικά

Οι GPCRs, όπως έχει ήδη αναφερθεί, είναι διαμεμβρανικοί υποδοχείς που επιτελούν τις περισσότερες λειτουργίες τους μέσω μονοπατιών που εμπλέκουν τις G-πρωτεΐνες, γεγονός που έδωσε στους GPCRs το όνομά τους (Pierce, et al., 2002). Έχουν πολλά εναλλακτικά μονοπάτια μεταγωγής σήματος, αλληλεπιδρώντας άμεσα και με άλλες πρωτεΐνες, όπως οι κινάσες και οι αρρεστίνες (Εικόνα 1.5) (Rosenbaum, et al., 2009).



Εικόνα 1.5: Παραδείγματα των βασικών μονοπατιών μεταγωγής σήματος των GPCRs (Rosenbaum et al. 2009).

Το βασικό μοντέλο της μεταγωγής σήματος μέσω GPCRs προέρχεται από την ικανότητα αυτών των υποδοχέων να δρουν, ενεργοποιούμενοι από υποκαταστάτες, ως παράγοντες ανταλλαγής νουκλεοτιδίων γουανίνης (Guanine nucleotide Exchange Factors, GEFs) για τις ετεροτριμερείς G-πρωτεΐνες, οι οποίες μεταγουν το σήμα ενδοκυτταρικά μέσω της ενεργοποίησης μορίων εκτελεστών, όπως ενζύμων και καναλιών ιόντων. Αυτός είναι ο κυριότερος τρόπος για τη μεταγωγή σήματος μέσω GPCRs (Oldham and Hamm, 2008). Παρ' όλα αυτά, έχεις δειχθεί ότι οι GPCRs συμμετέχουν σε ποικίλες αλληλεπιδράσεις πρωτεϊνών-πρωτεϊνών, οι οποίες παράγουν ενδοκυτταρικά σήματα σε συνδυασμό με ή, ακόμα και, ανεξάρτητα από, την ενεργοποίηση των G-πρωτεϊνών. Πράγματι, πλήρεις αναλύσεις γονιδιωμάτων υποδεικνύουν ότι κάποιοι υποδοχείς που διαπερνούν επτά φορές τη μεμβράνη δεν ζευγαρώνουν με G-

πρωτεΐνες, αλλά, παρ' όλα αυτά, αποτελούν εξελικτικά κλαδιά του φυλογενετικού δέντρου των GPCRs (Fredriksson, et al., 2003).

Οι GPCRs αποτελούν τη μεγαλύτερη υπεροικογένεια μεμβρανικών υποδοχέων, αριθμώντας 800 γονίδια στον άνθρωπό, δηλαδή πάνω από το 1% του ανθρώπινου γονιδιώματος (Almen, et al., 2009). Εδώ και πάνω από μία δεκαετία γίνεται εκτενής μελέτη της ευρείας ποικιλότητας των GPCRs σε διάφορα είδη, προερχόμενα από τα φύλα των Χορδωτών, των Εχινόδερμων, των Αρθροπόδων, των Κνιδοζώων, των Πλακοζώων και των Πρωτοζώων (Bjarnadottir, et al., 2006; Fredriksson and Schioth, 2005; Kamesh, et al., 2008; Nordstrom, et al., 2008; Nordstrom, et al., 2011). Οι μελέτες αυτές έχουν δείξει ότι υπάρχουν είδη στα οποία το ποσοστό των γονιδίων που κωδικοποιούν GPCRs ξεπερνά κατά πολύ το 1%, με αντιπροσωπευτικό παράδειγμα τον *C. elegans* με ποσοστό περίπου 5% (Fredriksson and Schioth, 2005). Η εξελικτική ποικιλομορφία παράγει υποδοχείς που ανιχνεύουν μια εκπληκτική ποικιλία εξωκυττάρων σημάτων, ορμόνες και νευροδιαβιβαστές, ένα πολύ ετερόκλητο σύνολο ενδογενών υποκαταστατών, όπως αμίνες, πεπτίδια, αμινοξέα, νουκλεοσίδια, νουκλεοτίδια, φωσφολιπίδια, λιπαρά οξέα, και ιόντα ασβεστίου, καθώς και διάφορους εξωγενείς υποκαταστάτες, όπως παράγοντες οσμής και γεύσης, φερομόνες και φωτόνια (Kristiansen, 2004). Οι GPCRs συμμετέχουν στη νευροδιαβίβαση, στον άμεσο νευροενδοκρινικό έλεγχο της ομοιόστασης και της αναπαραγωγής, ρυθμίζουν την αιμοδυναμική και τον ενδιάμεσο μεταβολισμό και επηρεάζουν την αύξηση, τον πολλαπλασιασμό, τη διαφοροποίηση και το θάνατο πολλών κυτταρικών τύπων (Luttrell, 2006; Luttrell, 2008; Rosenbaum, et al., 2009).

1.2.2 Συστήματα ταξινόμησης των GPCRs

Ο αριθμός, αλλά, και η ποικιλία των συγκεκριμένων υποδοχέων έχει οδηγήσει συχνά σε προσπάθειες ταξινόμησής τους. Τα κριτήρια που χρησιμοποιούνται για την ταξινόμηση είναι ποικίλα και συχνά έγκεινται αποκλειστικά στον ερευνητή. Τα συστήματα ταξινόμησης βασίζονται στη δομή των GPCRs, στην εξελικτική τους καταγωγή, στην ομοιότητα σε επίπεδο αλληλουχίας, στους ενδογενείς υποκαταστάτες ή σε οποιοδήποτε άλλο χαρακτηριστικό. Κανένα όμως δεν έχει τύχει πλήρους αποδοχής από την επιστημονική κοινότητα (van der Horst, et al., 2010).

Το 1994 οι Attwood and Findlay, και στη συνέχεια ο Kolakowski, πρότειναν ένα σύστημα ταξινόμησης το οποίο είναι δημοφιλές ακόμα και σήμερα. Αυτό το σύστημα επιχειρεί να ταξινομήσει όλη την γκάμα των GPCRs ανεξαρτήτως φυλογενετικής προέλευσης, αλλά δεν

περιλαμβάνει οικογένειες υποδοχέων που ανακαλύφθηκαν αργότερα. Οι υποδοχείς ταξινομούνται σε έξι κατηγορίες (A, B, C, D, E και F) στο ανώτερο επίπεδο, οι οποίες βασίζονται στην ομολογία της αμινοξικής ακολουθίας (αλληλουχίας) και στην ομοιότητα της λειτουργίας, ενώ στη συνέχεια ταξινομούνται περαιτέρω σε οικογένειες (Attwood and Findlay, 1994; Kolakowski, 1994). Αναλυτικά:

- η Κατηγορία A των GPCRs ή, αλλιώς, υποδοχείς που μοιάζουν με Ροδοψίνη (rhodopsin-like GPCRs), αφορά υποδοχείς για πολλούς δομικά διαφορετικούς υποκαταστάτες, όπως βιογενείς αμίνες, νουκλεοτίδια, πεπτίδια, γλυκοπρωτεϊνικές ορμόνες κ.α.,
- η Κατηγορία B, η οποία ονομάζεται και υποδοχείς που μοιάζουν με Σεκρετίνη (secretin-like GPCRs), περιέχει καθαρά πεπτιδικούς υποδοχείς,
- η Κατηγορία C (Μεταβοτροπικοί υποδοχείς του Γλουταμικού), περιέχει υποδοχείς του γλουταμικού οξέος, υποδοχείς του γ-άμινο-βουτυρικού οξέος και κάποιους γευστικούς υποδοχείς,
- η Κατηγορία D, περιέχει υποδοχείς για φερομόνες των μυκήτων,
- η Κατηγορία E περιέχει τους υποδοχείς του κυκλικού AMP (cAMP), οι οποίοι έως τώρα έχουν εντοπιστεί μόνο στο μύκητα *Dictyostelium*,
- τέλος, υπάρχει και η Κατηγορία των Frizzled/Smoothened υποδοχέων, οι οποίοι συμμετέχουν, κυρίως, στο μονοπάτι μεταγωγής σήματος μέσω Wnt πρωτεϊνών.

Υπάρχουν επίσης, και άλλες Κατηγορίες, στις οποίες κατατάσσονται συνεχώς νέο-προσδιορισθέντες υποδοχείς, αλλά με μη ευρέως αποδεκτή ακόμα ονοματολογία. Αυτό το σύστημα είναι σχεδιασμένο να καλύπτει όλους τους GPCRs, τόσο στα Χορδωτά όσο και στα μη Χορδωτά.

Όμως, η καθολική κατάταξη των GPCRs δυσχεραίνεται από τις μεγάλες διαφορές στην αμινοξική ακολουθία (αλληλουχία) μεταξύ των GPCRs των θηλαστικών και αυτών των ασπονδύλων. Τα τελευταία χρόνια, ολοένα και περισσότερα δεδομένα, σε επίπεδο γονιδιώματος, έγιναν διαθέσιμα για μια μεγάλη ποικιλία ειδών, καθιστώντας δυνατή τη δημιουργία ενός μοντέλου για τη φυλογένεση των GPCRs. Οι Fredriksson et al., έπειτα από ανάλυση των χρωμοσωμικών θέσεων και των αποτυπωμάτων των ακολουθιών για μεγάλο αριθμό υποδοχέων, πρότειναν το σύστημα ταξινόμησης GRAFS, στο οποίο οι υποδοχείς του ανθρώπου ταξινομούνται σε 5 οικογένειες (Εικόνα 1.7): Glutamate, Rhodopsin, Adhesion, Frizzled/taste2 και Secretin (Fredriksson, et al., 2003; Schioth and Fredriksson, 2005), ταξινόμηση που μπορεί να χρησιμοποιηθεί συνολικά στα θηλαστικά (Bjarnadottir, et al., 2006;

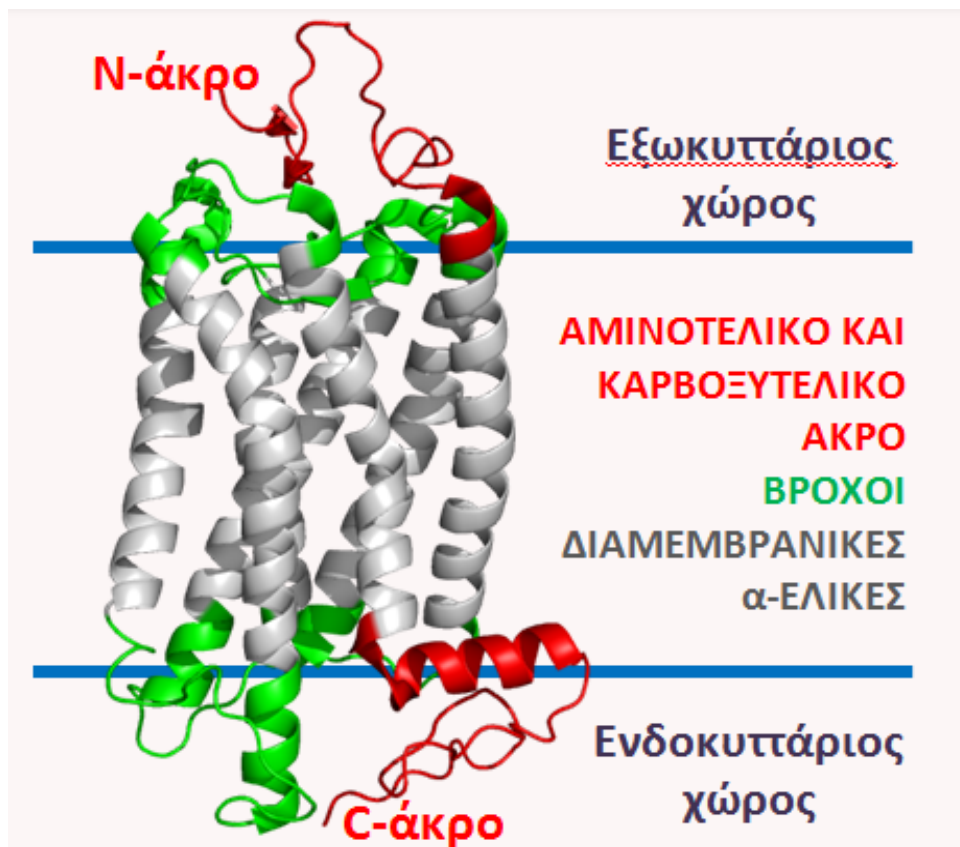
Fredriksson and Schioth, 2005). Η οικογένεια Rhodopsin, η μεγαλύτερη από τις πέντε οικογένειες, χωρίζεται περαιτέρω σε 13 υποκατηγορίες, ενώ προτείνεται και η ύπαρξη μιας έκτης οικογένειας, η οποία περιλαμβάνει όσους υποδοχείς δεν μπορούν να κατηγοριοποιηθούν σε καμία από τις υπόλοιπες οικογένειες (Other 7TM receptors) (Fredriksson, et al., 2003; Schioth and Fredriksson, 2005). Ο αριθμός των υποδοχέων που ανήκουν σε αυτήν την τελευταία οικογένεια είναι πολύ μικρός, καθώς με βάση την ομοιότητα σε επίπεδο αλληλουχίας, οι περισσότεροι υποδοχείς κατατάσσονται σε μία από τις πέντε βασικές οικογένειες. Η ταξινόμηση αυτή γίνεται αποκλειστικά με βάση φυλογενετικά κριτήρια, και πιο συγκεκριμένα, βασίζεται σε προσεκτικές στοιχίσεις της περιοχής των επτά διαμεμβρανικών α-ελίκων, η οποία είναι κοινή σε όλους τους GPCRs. Το σύστημα αυτό προτείνει ότι οι GPCRs προήλθαν από έναν κοινό πρόγονο, ενώ, η διαφοροποίησή τους προέκυψε από την ανακατάταξη των εξωνίων. Μέσω των φυλογενετικών αναλύσεων, αποκαλύφθηκαν μερικές μη αναμενόμενες σχέσεις, όπως η προτεινόμενη σχέση μεταξύ των Frizzled υποδοχέων, που γενικά δεν θεωρείται ότι μετάγουν σήματα μέσω G-πρωτεϊνών και της TAS2 υπο-οικογένειας των υποδοχέων γεύσης (Fredriksson, et al., 2003; Schioth and Fredriksson, 2005). Τέτοιες φυλογενετικές διασυνδέσεις υπονοούν ότι ο όρος «συζευγμένοι με G-πρωτεΐνες υποδοχείς» μπορεί να αποτελεί ένα μόνο τμήμα μιας υπεροικογένειας των υποδοχέων με 7 διαμεμβρανικά τμήματα, οι οποίοι χρησιμοποιούν ποικίλους μηχανισμούς μεταγωγής σήματος (Luttrell, 2006; Luttrell, 2008).

1.2.3 Δομικά χαρακτηριστικά των GPCRs

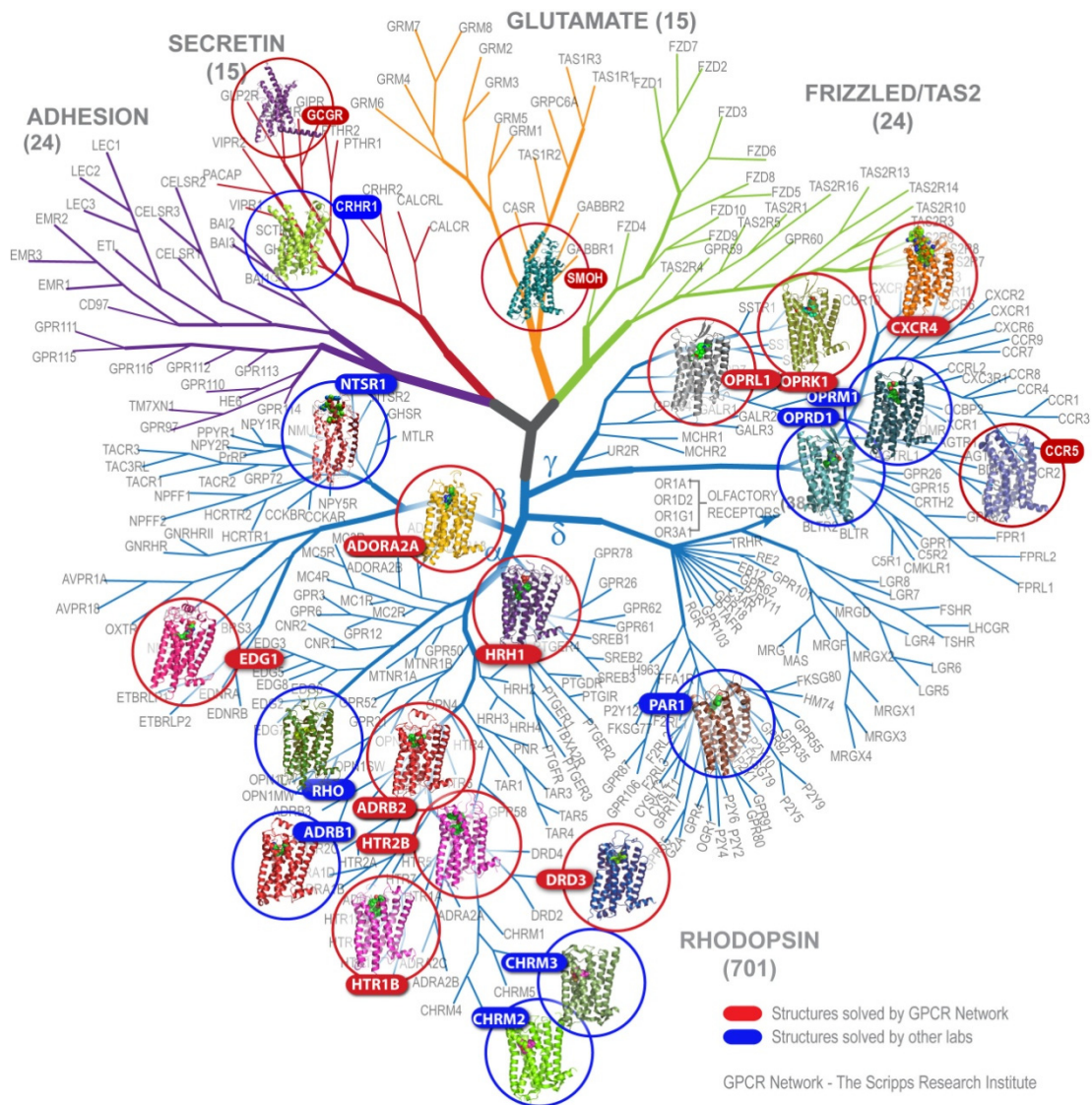
Όλοι οι GPCRs μοιράζονται κοινή τοπολογία αποτελούμενη από 7 διαμεμβρανικές α-έλικες, ένα εξωκυττάριο αμινοτελικό άκρο, ένα ενδοκυττάριο καρβοξυτελικό άκρο, καθώς και 3 βρόχους σε κάθε πλευρά της μεμβράνης (Εικόνα 1.6). Ένας τέταρτος κυτταροπλασματικός βρόχος δημιουργείται κάποιες φορές, όταν το καρβοξυτελικό τμήμα προσδένεται στη μεμβράνη μέσω παλμιτοϋλίωσης αμινοξικών καταλοίπων (Kristiansen, 2004).

Αυτή η κοινή τοπολογία επιβεβαιώνεται από τα συνεχώς αυξανόμενα δεδομένα που αφορούν την τρισδιάστατη δομή των GPCRs. Ο προσδιορισμός της πρώτης δομής ενός GPCR έγινε το 2000 και αφορούσε τη Ροδοψίνη του βοδιού (Palczewski, et al., 2000), όμως χρειάστηκε να περάσουν 7 ολόκληρα χρόνια για να επιτευχθεί ο προσδιορισμός της δομής του δεύτερου GPCR, γεγονός που οφείλεται στην εξαιρετική δυσκολία κρυστάλλωσής τους. Πέραν των γνωστών δυσκολιών που εμφανίζονται στην κρυστάλλωση μεμβρανικών πρωτεϊνών

γενικότερα και οφείλονται στις διαδικασίες διαλυτοποίησης και απομόνωσης τους (Tusnady, et al., 2004), ειδικά στους GPCRs εμφανίζονται δύο ακόμα σημαντικά προβλήματα: Τα χαμηλά επίπεδα έκφρασής τους στους ζωντανούς ιστούς δημιουργούν την ανάγκη για ειδικά συστήματα έκφρασης, που θα παράγουν σε μεγάλες συγκεντρώσεις μεμβρανικές πρωτεΐνες που διπλώνουν στο χώρο ακριβώς όπως και σε *in vivo* συνθήκες. Επιπλέον, εξαιρουμένης της Ροδοψίνης, όλοι οι υπόλοιποι GPCRs δεν έχουν θερμοδυναμική σταθερότητα και είναι ευάλωτοι σε πρωτεολυτικά ένζυμα λόγω των μη προστατευμένων εξωμεμβρικών περιοχών τους (Rosenbaum, et al., 2009). Αυτές οι δυσκολίες αντιμετωπίστηκαν επιτυχώς και έτσι από το 2007 μέχρι σήμερα έχουν λυθεί οι δομές 23 διαφορετικών υποδοχέων. Στην Εικόνα 1.7 παρουσιάζεται η ταξινόμηση των ανθρώπινων GPCRs βάση του συστήματος GRAFS και απεικονίζονται σε αυτό οι υποδοχείς για τους οποίους έχει προσδιοριστεί η δομή τους με κρυσταλλογραφία ακτίνων-Χ. Μόνο για έναν υποδοχέα ο προσδιορισμός της δομής έγινε με NMR (ανθρώπινος χημειοκινικός υποδοχέας CXCR1, PDB ID: 2LNL).



Εικόνα 1.6: Τρισδιάστατη δομή της Ροδοψίνης (PDBID: 1U19). Επισημαίνονται τα τοπολογικά χαρακτηριστικά του υποδοχέα.



Εικόνα 1.7: Κατάταξη των ανθρώπινων GPCRs κατά GRAFS. Παράλληλη απεικόνιση των υποδοχέων με κρυσταλλογραφικά προσδιορισμένη δομή. Προσαρμογή της αντίστοιχης εικόνας που είναι διαθέσιμη στη δημοσίευση των Stevens, et al. (Stevens, et al., 2013) και στην ιστοσελίδα GPCR Network (<http://gpcr.scripps.edu/>).

Οι δομές των περισσότερων από τους 24 υποδοχείς έχουν προσδιοριστεί επανειλημμένα, σε διαφορετική διακριτικότητα και ως σύμπλοκα με διάφορα μόρια. Τα επόμενα 5 χρόνια αναμένουμε 15-20 νέες δομές μέσα από τη μελέτη 100 υποψηφίων στόχων (Stevens, et al., 2013). Ο Πίνακας 1.1 αναφέρει τους 24 αυτούς υποδοχείς και δίνει πληροφορίες για την πρώτη δομή που προσδιορίστηκε. Στο Παράρτημα Ι δίνεται ο Πίνακας Ι.1 με το σύνολο των προσδιορισμένων δομών για τους GPCRs.

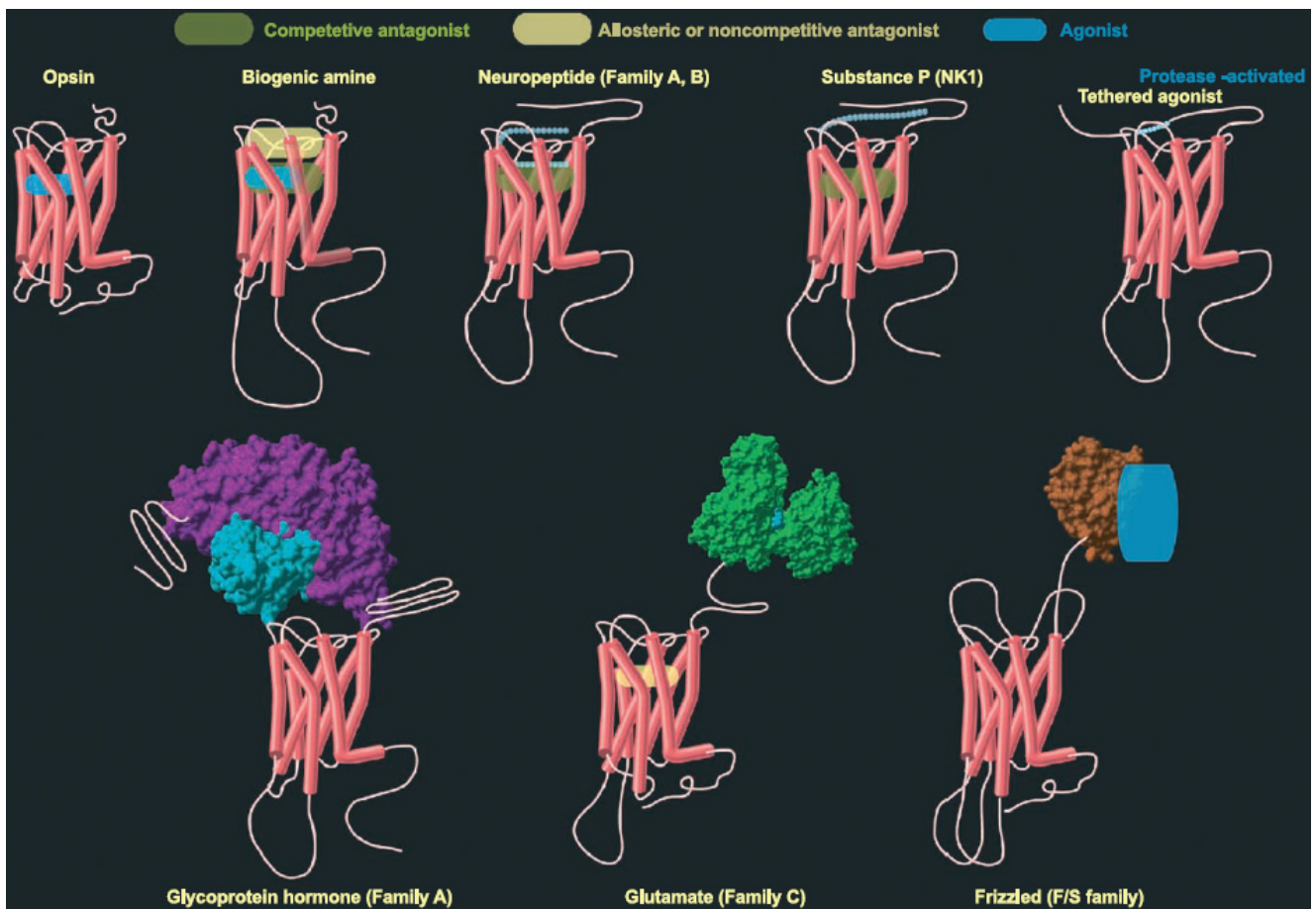
Πίνακας 1.1 Οι 24 GPCRs με προσδιορισμένη τρισδιάστατη δομή

Υποδοχέας	PDB	Resolution	Πρώτη δημοσίευση	Pubmed
Ροδοψίνη Βοδιού (Bovine Rhodopsin)	1F88	2.80 Å	4/8/2000	10926528
Ροδοψίνη καλαμαριού (Squid Rhodopsin)	2ZIY	3.70 Å	6/5/2008	18463093
Ανθρώπινος β2 αδρενεργικός υποδοχέας (Human β2 adrenergic Receptor)	2RH1	2.40 Å	30/10/2007	17962520
β1 αδρενεργικός υποδοχέας γαλοπούλας (Turkey β1 adrenergic Receptor)	2VT4	2.70 Å	24/6/2008	18594507
Ανθρώπινος A2α αδενোসινικός Υποδοχέας (Human A2α adenosine Receptor)	3EML	2.60 Å	14/10/2008	18832607
Ανθρώπινος CXCR4 χημειοκινικός Υποδοχέας (Human CXCR4 Chemokine Receptor)	3ODU	2.50 Å	27/10/2010	20929726
Ανθρώπινος D3 ντοπαμινικός Υποδοχέας (Human D3 Dopamine Receptor)	3PBL	2.89 Å	3/11/2010	21097933
Ανθρώπινος H1 ισταμινικός Υποδοχέας (Human H1 Histamine Receptor)	3RZE	3.10 Å	15/6/2011	21697825
Ανθρώπινος M2 μουσκαρινικός Υποδοχέας (Human M2 Muscarinic Receptor)	3UON	3.00 Å	1/2/2012	22278061
Ανθρώπινος λιπιδικός S1P1 υποδοχέας (Human Lipid S1P1 Receptor)	3V2W	3.35 Å	15/2/2012	22344443
Ανθρώπινος M3 μουσκαρινικός Υποδοχέας (Human M3 Muscarinic Receptor)	4DAJ	3.40 Å	22/2/2012	22358844
Ανθρώπινος κ οπιοειδής Υποδοχέας (Human κ oprioid receptor)	4DJH	2.90 Å	21/3/2012	22437504
Ανθρώπινος μ οπιοειδής Υποδοχέας (Human μ oprioid receptor)	4DKL	2.80 Å	21/3/2012	22437502
Ανθρώπινος N/OFQ οπιοειδής Υποδοχέας (Human N/OFQ oprioid Receptor)	4EA3	3.01 Å	25/4/2012	22596163
Ανθρώπινος δ οπιοειδής Υποδοχέας (Human δ oprioid receptor)	4EJ4	3.40 Å	16/5/2012	22596164
Ανθρώπινος CXCR1 χημειοκινικός Υποδοχέας (Human CXCR1 Chemokine Receptor)	2LNL	NMR	17/10/2012	23086146
NTS1 Νευροτενσινικός υποδοχέας Του αρουραίου (Rat NTS1 Neurotensin Receptor)	4GRV	2.8 Å	17/10/2012	23051748
PAR1 Ενεργοποιούμενος από πρωτεάση υποδοχέας 1 Του ανθρώπου (Human PAR1 Protease Activated Receptor)	3VW7	2.2 Å	12/12/2012	23222541
Ανθρώπινος 5-HT υποδοχέας 1B (Human 5-HT1B receptor)	4IAR	2.70 Å	13/3/2013	23519210
Ανθρώπινος 5-HT υποδοχέας 2B (Human 5-HT2B receptor)	4IB4	2.70 Å	13/3/2013	23519215
Ανθρώπινος Υποδοχέας Smoothened (Human Smoothened Receptor)	4JKV	2.45 Å	29/5/2013	23636324
Ανθρώπινος Υποδοχέας του παράγοντα απελευθέρωσης κορτικοτροπίνης Τύπου 1 (Human Corticotropin-releasing factor receptor 1)	4K5Y	2.98 Å	17/7/2013	23863939
Ανθρώπινος Υποδοχέας του Γλυκογόνου (Human Glucagon Receptor)	4L6R	3.30 Å	24/7/2013	23863937
Ανθρώπινος CCR5 χημειοκινικός Υποδοχέας (Human CCR5 Chemokine Receptor)	4MBS	2.71 Å	11/9/2013	24030490

1.2.4 Λειτουργικά χαρακτηριστικά των GPCRs

Τα διαμεμβρανικά τμήματα των διάφορων υποδοχέων παρουσιάζουν μεγαλύτερο βαθμό συντήρησης της αμινοξικής τους ακολουθίας, ενώ τα εξωκυττάρια και ενδοκυττάρια τμήματα εμφανίζουν μεγάλη μεταβλητότητα σε μέγεθος και πολυπλοκότητα. Οι εξωκυττάρια και οι διαμεμβρανικές περιοχές του υποδοχέα εμπλέκονται στην πρόσδεση του υποκαταστάτη, ενώ τα ενδοκυττάρια τμήματα είναι σημαντικά για τη μεταγωγή σήματος και την ανατροφοδοτούμενη διαφοροποίηση της λειτουργίας του υποδοχέα. Μία ή περισσότερες περιοχές για N-γλυκοζυλίωση υπάρχουν στο αμινοτελικό άκρο ή, λιγότερο συχνά, στους εξωκυττάρια βρόχους. Οι περισσότεροι υποδοχείς έχουν 2 συντηρημένες κυστεΐνες, που φτιάχνουν έναν δισουλφιδικό δεσμό, ο οποίος συνδέει τον πρώτο με το δεύτερο εξωκυττάριο βρόχο και είναι κρίσιμος για το φυσιολογικό δίπλωμα της πρωτεΐνης, καθώς και άλλα κατάλοιπα κυστεΐνης στο καρβοξυτελικό άκρο, που εξυπηρετούν ως θέσεις για παλμιτουλίωση. Αυτή η λιπιδική τροποποίηση οδηγεί στο σχηματισμό ενός πιθανού τέταρτου ενδοκυττάρια βρόχου (Luttrell, 2006; Luttrell, 2008). Τέλος, αξίζει να αναφερθεί, ότι, όπως σε όλες τις α-ελικοειδείς διαμεμβρανικές πρωτεΐνες, έτσι και στους GPCRs η ύπαρξη καταλοίπων προλίνης σε συγκεκριμένες θέσεις σε μια διαμεμβρανική α-έλικα επάγει το σχηματισμό κυρτώσεων (kinks). Η ύπαρξη τέτοιων καταλοίπων στους GPCRs, έχει προταθεί ότι εντοπίζεται σε διαφορετικές περιοχές, και, επομένως, υποδηλώνει διαφορετικό λειτουργικό ρόλο σε αυτές τις κυρτώσεις (Yohannan, et al., 2004).

Η θέση πρόσδεσης του υποκαταστάτη εξαρτάται από το είδος του (Εικόνα 1.8) (Kristiansen, 2004). Όταν ο υποκαταστάτης είναι πρωτεϊνικής φύσεως, αυτός αλληλεπιδρά με τον υποδοχέα κυρίως μέσω των μεγάλων περιοχών στον εξωκυττάρια χώρο. Διαφορετικές τέτοιες περιοχές είναι δυνατόν να εμπλέκονται στην αναγνώριση ακόμα και σχετικά όμοιων υποκαταστατών. Αντίθετα, υπάρχουν ενδείξεις ότι για τη δέσμευση και την αναγνώριση υποκαταστατών μικρών μοριακών βαρών, οι λειτουργικά απαραίτητες περιοχές εμφανίζονται, κυρίως, στα εξωτερικά όρια των διαμεμβρανικών α-ελίκων και, συγκεκριμένα, στην άνω εσωτερική κοιλότητα του δεματίου, στην περιοχή αλληλεπίδρασης των ελίκων. Τα αμινοξικά κατάλοιπα που πλαισιώνουν αυτόν το θύλακα διαφέρουν σημαντικά μεταξύ των υποδοχέων, παρέχοντας έτσι την εξειδίκευση σε κάθε υποδοχέα για τη σύνδεσή του μ' έναν συγκεκριμένο υποκαταστάτη (Hollmann, et al., 2005).



Εικόνα 1.8: Σχηματικά μοντέλα συμπλόκων υποδοχέων-υποκαταστατών για δομικά ποικίλους υποκαταστάτες που αλληλεπιδρούν με GPCRs (Kristiansen, 2004).

Οι υποδοχείς βρίσκονται σε μια ισορροπία μεταξύ δύο στερεοδιατάξεων: μια ανενεργό κατάσταση και μια ενεργό κατάσταση, κατά την οποία ο υποδοχέας έχει την ικανότητα να καταλύει την ανταλλαγή νουκλεοτιδίων στις G-πρωτεΐνες. Η σύνδεση του υποκαταστάτη με τον υποδοχέα προκαλεί την έναρξη της μεταγωγής σήματος, μετατοπίζοντας την ισορροπία προς την ενεργό στερεοδιάταξη (Kristiansen, 2004). Νεότερα κρυσταλλογραφικά δεδομένα, στα οποία περιλαμβάνονται οι προσδιορισμένες δομές των ενεργοποιημένων ενδιάμεσων της Ροδοψίνης, Οψίνης (Scheerer, et al., 2008) και Μεταροδοψίνης II (Choe, et al., 2011), καθώς και οι δομές των ενεργών στερεοδιατάξεων του β_2 αδρενεργικού υποδοχέα (Rasmussen, et al., 2011) και του $A_{2\alpha}$ αδενোসιλικού υποδοχέα (Xu, et al., 2011), δείχνουν πως οι GPCRs της Κατηγορίας A ακολουθούν ένα κοινό πρότυπο ενεργοποίησης σύμφωνα με το οποίο μετατοπίζεται κυρίως η 6^η διαμεμβρανική α -έλικα κατά 6-15Å. Παρατηρούνται επιπλέον μετατοπίσεις στην 3^η, 5^η και 7^η διαμεμβρανική περιοχή, όπου, συντηρημένα κατάλοιπα τυροσίνης στην 5^η και 7^η

έλικα και αργινίνης στην 3^η έλικα σταθεροποιούν τη δομή (Rasmussen, et al., 2011; Scheerer, et al., 2008). Οι μετατοπίσεις οδηγούν στο σχηματισμό μιας κοιλότητας στην κυτταροπλασματική περιοχή του υποδοχέα, στην οποία εισέρχονται και αλληλεπιδρούν οι G-πρωτεΐνες, οι GRKs (G-protein coupled receptor kinases) και οι αρρεστίνας (Huang and Tesmer, 2011; Rasmussen, et al., 2011).

Η παραδοσιακή οπτική για τους GPCRs θεωρούσε πως συμμετέχουν ως μονομερή στις διαδικασίες μεταγωγής σήματος. Ένας μοναδικός υποκαταστάτης ενεργοποιεί έναν μοναδικό υποδοχέα, οδηγώντας σε αλληλεπιδράσεις με μια G-πρωτεΐνη ή ένα μόριο εκτελεστή. Ωστόσο, η ανακάλυψη ότι ο υποδοχέας GABA_B είναι στην πραγματικότητα ένα υποχρεωτικό ετεροδιμερές (καθώς τα μονομερή του, GABA_{B1} και GABA_{B2} δεν είναι λειτουργικά από μόνα τους, παρά μόνο όταν σχηματίζουν σύμπλοκο) σε συνδυασμό με διάφορες άλλες μελέτες μεταφοράς ενέργειας (energy transfer), ανοσοκατακρήμνισης κτλ., τα τελευταία χρόνια έχει αρχίσει να γίνεται περισσότερο αποδεκτή η άποψη ότι οι GPCRs σχηματίζουν ομο- και ετερο-ολιγομερή που λειτουργούν συνεργατικά – χωρίς φυσικά αυτό να αποκλείει την ύπαρξη πλήρως λειτουργικών μονομερών υποδοχέων (Pin, et al., 2007; Skrabanek, et al., 2007).

Έχουν περιγραφεί λειτουργικές αλληλεπιδράσεις μεταξύ των πρωτομερών ενός ομο- / ετεροδιμερούς ή ολιγομερούς, όπου η πρόσδεση ενός υποκαταστάτη σε έναν από τους υποδοχείς του συμπλέγματος μπορεί να επηρεάσει τους υπόλοιπους (ενεργοποιώντας ή καταστέλλοντας τη λειτουργία τους). Διάφορες, όμως, μελέτες έχουν δώσει στοιχεία για λειτουργίες που απαντώνται αποκλειστικά στα ετερο-ολιγομερή και όχι στα πρωτομερή που τα αποτελούν, καθώς και συσχετισμούς με φαρμακολογικές αποκρίσεις και εμπλοκές σε ασθένειες, στοιχεία που οδηγούν στο συμπέρασμα ότι, σε κάποιες περιπτώσεις, τα ολιγομερή θα πρέπει να αντιμετωπίζονται ως ξεχωριστοί υποδοχείς με συγκεκριμένη λειτουργικότητα (Skrabanek, et al., 2007).

1.2.5 Υποκαταστάτες (Ligands)

Σύμφωνα με τα στάδια της μεταγωγής σήματος, για να υπάρξει μεταφορά του εξωκυττάριου σήματος, χρειάζεται συνήθως η ύπαρξη ενός υποκαταστάτη, ο οποίος προσδένεται στον υποδοχέα. Όπως έχει αναφερθεί και σε προηγούμενες ενότητες, οι υποκαταστάτες που προσδένονται στην υπεροικογένεια των GPCRs, χαρακτηρίζονται από μεγάλη ετερογένεια, και μπορεί να είναι είτε φυσικής μορφής, όπως φωτόνια και παράγοντες οσμής και γεύσης είτε χημικής, όπως ορμόνες, παράγοντες αύξησης, συστατικά της εξωκυττάριας ουσίας, κυτταροκίνες, χημειοκίνες, νευροδιαβιβαστές, νευροτροφίνες και ενεργές ελεύθερες ρίζες.

Υποδοχείς ομοιάζοντες με Ροδοψίνη (Κατηγορία A) και οι υποκαταστάτες τους

Η Κατηγορία A των GPCRs, υποδοχείς ομοιάζοντες με τη Ροδοψίνη (Rhodopsin-like), συνιστούν τη μεγαλύτερη διακριτή οικογένεια GPCRs. Οι συγκεκριμένοι υποδοχείς ενεργοποιούνται από ένα μεγάλο εύρος υποκαταστατών και για το λόγο αυτό η συγκεκριμένη κατηγορία μπορεί περαιτέρω να χωριστεί σε 6 κατηγορίες υποδοχέων ανάλογα με τους υποκαταστάτες τους (Kristiansen, 2004). Η θέση πρόσδεσης του υποκαταστάτη διαφοροποιείται ανάλογα με τη φύση και το μέγεθός του (Εικόνα 1.8).

- **Οψίνες:** Σε αντίθεση με άλλους υποδοχείς της Κατηγορίας A, οι οψίνες ενεργοποιούνται από φωτόνια που προκαλούν μία στερεοδιαταξική αλλαγή σε έναν ομοιοπολικά συνδεδεμένο υποκαταστάτη/χρωμοφόρο. Το χρωμοφόρο (11-cis-ρετινάλη) είναι ομοιοπολικά συνδεδεμένο στον υποδοχέα και η ενεργοποίηση συμβαίνει ως απόκριση στην απορρόφηση φωτός (Filipek, et al., 2003; Menon, et al., 2001).
- **Υποδοχείς για οσμές:** Οι οσφρητικοί υποδοχείς αποτελούν την πολυπληθέστερη κατηγορία της Κατηγορίας A (περί τα 500 μέλη στον άνθρωπο) και προσδένουν μικρά μόρια, τις οσμές. Συγκεκριμένα αμινοξικά κατάλοιπα του 2^{ου} και του 7^{ου} διαμεμβρανικού τμήματος καθώς και του 2^{ου} εξωκυττάριου βρόχου είναι σημαντικοί παράγοντες για την εξειδικευμένη πρόσδεση οσμών (Man, et al., 2004).
- **Υποδοχείς για μικρούς ενδογενείς υποκαταστάτες:** Οι υποδοχείς αυτοί ενεργοποιούνται από μη ομοιοπολική σύνδεση συγκεκριμένων αγωνιστών στο άνω μέρος των 7 διαμεμβρανικών α-ελίκων. Οι υποκαταστάτες φαίνεται να

“βυθίζονται” μεταξύ των ελίκων (Tota and Strader, 1990), με τις θέσεις πρόσδεσης να βρίσκονται βαθιά μεταξύ της 3^{ης}, 4^{ης}, 5^{ης}, 6^{ης} και 7^{ης} έλικας. Βιογενείς αμίνες (Liarakis, et al., 2000; Strader, et al., 1989; Strader, et al., 1987), αδενοσίνη (Jiang, et al., 1997), νουκλεοτίδια (Jiang, et al., 1997), η ορμόνη απελευθέρωσης της θυρεοτροπίνης (TRH) (Perlman, et al., 1996), fMLP (Miettinen, et al., 1997), PAF (Ishii, et al., 1997), μελατονίνη (Kokkola, et al., 2003) και συστατικά που μοιάζουν με λιπίδια (Song and Bonner, 1996), όπως η ανανδαμίνη, αποτελούν υποκαταστάτες που έχει βρεθεί ότι προσδένονται σε αυτές τις θέσεις.

- **Υποδοχείς πεπτιδίων:** Αποτελούν τη μεγαλύτερη υποομάδα υποδοχέων και μεσολαβούν σε σημαντικούς φυσιολογικούς ρόλους των ενδογενών πεπτιδίων που δρουν ως νευροδιαβιβαστές, ορμόνες και παρακρινείς παράγοντες. Σε αντίθεση με την προηγούμενη κατηγορία, στις θέσεις πρόσδεσης αυτών των υποδοχέων φαίνεται να παίζουν σημαντικό ρόλο και οι εξωκυττάριες περιοχές στην πρόσδεση μεγαλύτερων σε μέγεθος υποκαταστατών (Gether, 2000). Για την πλειοψηφία των πεπτιδίων των υποδοχέων που έχουν μελετηθεί, οι πεπτιδοκοί υποκαταστάτες φαίνεται να αλληλεπιδρούν απευθείας με κατάλοιπα του αμινοτελικού άκρου και των εξωκυττάριων βρόχων (Gether, 2000). Σε αυτούς περιλαμβάνονται υποδοχείς της αγγειοτενσίνης II, του νευροπεπτιδίου Υ, χημειοκινών, των βασοπρεσίνης/οξυτοκίνης, της ορμόνης απελευθέρωσης γοναδοτροπίνης (GnRH), της TRH, του παράγοντα συμπληρώματος 5a, του fMLP, της σωματοστατίνης, οπιοειδών, της βραδυκινίνης, των χολεκυστοκινίνης/γαστρίνης και της νευροτενσίνης (Gether, 2000).
- **Υποδοχείς με επαναλήψεις του πλούσιου σε λευκίνη μοτίβου (leucine-rich repeat motif - LRR motif):** Οι υποδοχείς της ωοθυλακιοτρόπου ορμόνης (LH), της ωχρινοτρόπου ορμόνης (LH), της χοριονικής γοναδοτροπίνης (CG), της θυρεοειδούς ορμόνης (TSH), αλλά και οι υποδοχείς που περιέχουν επαναλήψεις του μοτίβου LRR (LGR) είναι μοναδικοί, καθώς έχουν μεγάλο αμινοτελικό άκρο (με μήκη που ποικίλουν από <350 ως >500) και περιλαμβάνουν 9 ατελείς επαναλήψεις του μοτίβου LRR. Υπάρχουν πολλά σημεία επαφής μεταξύ υποκαταστάτη και υποδοχέα μεταξύ του αμινοτελικού άκρου και του 1^{ου} και 2^{ου} εξωκυττάριου βρόχου (Nishi, et al., 2002; Wonerow, et al., 2001).

- **Υποδοχείς ενεργοποιούμενοι από πρωτεάσες:** Η ενεργοποίηση των PAR (protease-activating receptors), PAR₁ και PAR₃ (θρομβίνη), PAR₂ (θρυψίνη) και PAR₄ (θρομβίνη και θρυψίνη) εμπλέκει έναν πρωτεολυτικό μηχανισμό, όπου οι πρωτεάσες “κόβουν” τον υποδοχέα για να δημιουργήσουν ένα νέο αμινοτελικό άκρο (Macfarlane, et al., 2001; Vu, et al., 1991). Το νεοεκτεθειμένο αμινοτελικό άκρο λειτουργεί στη συνέχεια ως υποκαταστάτης, ο οποίος προσδένεται και ενεργοποιεί τους υποδοχείς. Υπάρχουν ενδείξεις ότι η σύνδεση γίνεται στο 2^ο εξωκυττάριο βρόχο (Macfarlane, et al., 2001).

Υποδοχείς ομοιάζοντες με Σεκρετίνη (Κατηγορία Β) και οι υποκαταστάτες τους

Η Ομάδα Β των GPCRs (Secretin-like), σε αντίθεση με την Α, περιέχει μικρότερο αριθμό υποδοχέων και δεν χαρακτηρίζεται από ετερογένεια υποκαταστατών. Όλα τα μέλη έχουν μία περιοχή πρόσδεσης ορμονών (hormone binding domain - HBD) στο αμινοτελικό τους άκρο, ενώ η θέση πρόσδεσης του υποκαταστάτη είναι μη συνεχής και αποτελείται από σημεία «επαφής» που ανήκουν στο αμινοτελικό άκρο και στους τρεις εξωκυττάριους βρόχους (Unson, 2002). Ένα συντηρημένο ασπαρτικό του 1^{ου} εξωκυττάριου βρόχου στα όρια της 2^{ης} διαμεμβρανικής έλικας εμφανίζεται σε όλα τα μέλη της Κατηγορίας Β, ενώ του ασπαρτικού προηγείται ένα, επίσης, συντηρημένο βασικό κατάλοιπο, με εξαίρεση τους υποδοχείς του Γλουκαγόνου (Langer, et al., 2003). Τα φορτισμένα αυτά κατάλοιπα αποτελούν «κλειδιά» για την πρόσδεση του υποκαταστάτη και την ενεργοποίηση του υποδοχέα (Langer, et al., 2003). Την Κατηγορία Β των GPCRs συνιστούν οι υποδοχείς της Καλσιτονίνης (Calcitonin), των Παραγόντων Απελευθέρωσης Κορτικοτροπινών (Corticotropin-Releasing Factor), του Αγγειοδραστικού Εντερικού Πολυπεπτιδίου (VIP) και του Πολυπεπτιδίου της Αδενυλικής Κυκλάσης της Υπόφυσης (PACAP), του Γλουκαγόνου (Glucagon) και των Παραθυροειδών Ορμονών (Parathyroid Hormone). Οι υποκαταστάτες, που αλληλεπιδρούν με αυτούς τους υποδοχείς είναι πεπτίδια, που ανήκουν στις οικογένειες Γλουκαγόνου/Σεκρετίνης (Glucagon/Secretin), του Παράγοντα Απελευθέρωσης των Κορτικοτροπινών (Corticotropin-Releasing Factor), η οποία περιλαμβάνει τον Παράγοντα Απελευθέρωσης Κορτικοτροπινών (Corticotropin -Releasing Factor) και τις Urocortins, Καλσιτονίνη και την οικογένεια των Παραθυροειδών Ορμονών.

Μεταβοτροπικοί Υποδοχείς του Γλουταμικού (Κατηγορία C) και οι υποκαταστάτες τους

Σε αυτή την κατηγορία ανήκουν οι μεταβοτροπικοί υποδοχείς του Γλουταμικού, του κύριου διεγερτικού νευροδιαβιβαστή στο κεντρικό νευρικό σύστημα, καθώς και υποδοχείς για το γ-αμινοβουτυρικό οξύ, υποδοχείς Ca^{2+} , γευστικοί υποδοχείς του γλυκού και των αμινοξέων, υποδοχείς για ορισμένες φερομόνες, καθώς και οσμές στα ψάρια (Pin, et al., 2003). Τα περισσότερα μέλη της Κατηγορίας φέρουν μεγάλο εξωκυττάριο τμήμα, το οποίο μπορεί να χωριστεί σε 2 αυτοτελή στοιχεία (domains): 1. Μια περιοχή “αχιβάδας - εντομοπαγίδας” (Venus Flytrap module – VFTM), η οποία εμπλέκεται στην πρόσδεση του αγωνιστή και εμφανίζει ομοιότητα σε επίπεδο αμινοξικής ακολουθίας με βακτηριακές πρωτεΐνες του περιπλασμικού χώρου που προσδένουν αμινοξέα, και 2. Μια περιοχή πλούσια σε κυστεΐνη (cystein-rich domain - CRD) που διασυνδέει VFTM με την περιοχή των 7 ελίκων στους περισσότερους υποδοχείς της Κατηγορίας C (Pin, et al., 2003).

Υποδοχείς Frizzled/Smoothened (Κατηγορία F) και οι υποκαταστάτες τους

Η Κατηγορία των Frizzled/Smoothened, όπως δηλώνει και το όνομά τους περιλαμβάνει τις Frizzled και Smoothened οικογένειες GPCR. Οι Frizzled υποδοχείς αποτελούν στόχους για μια σειρά από μεγάλες πρωτεΐνες, ονόματι WNTs και για τις Norrin και R-spondins οικογένειες πεπτιδικών υποκαταστατών (Tamai, et al., 2000). Από την πλευρά του, ο Smoothened (Smo), είναι ένας ιδιαίτερος υποδοχέας, ο οποίος δεν ενεργοποιείται άμεσα από τους υποκαταστάτες του, που είναι τα πεπτίδια της οικογένειας των Hedgehog (Hh) πρωτεϊνών, αλλά οι υποκαταστάτες αυτοί αρχικά ενεργοποιούν έναν άλλο υποδοχέα ονόματι Patched (Ptc) και στη συνέχεια ακολουθούν τα επόμενα βήματα μεταγωγής σήματος, που εμπλέκουν και τον Smo (Bürglin, 2008).

1.2.6 Εξέλιξη των GPCRs

Η εξελικτική επιτυχία της υπεροικογένειας των GPCRs αντικατοπτρίζεται, τόσο από την παρουσία των υποδοχέων σε σχεδόν όλους τους ευκαρυωτικούς οργανισμούς, όσο και από την αφθονία τους στα θηλαστικά (Εικόνα 1.9). Οι υποδοχείς που μοιάζουν με Ροδοψίνη αποτελούν τη μεγαλύτερη οικογένεια στα Χορδωτά. Εξαιτίας της πολύ μικρής ομολογίας σε επίπεδο ακολουθίας μεταξύ των Κατηγοριών A (ομοιάζοντες με Ροδοψίνη), B (ομοιάζοντες με Σεκρετίνη) και C (Μεταβοτροπικοί Υποδοχείς του Γλουταμικού), η

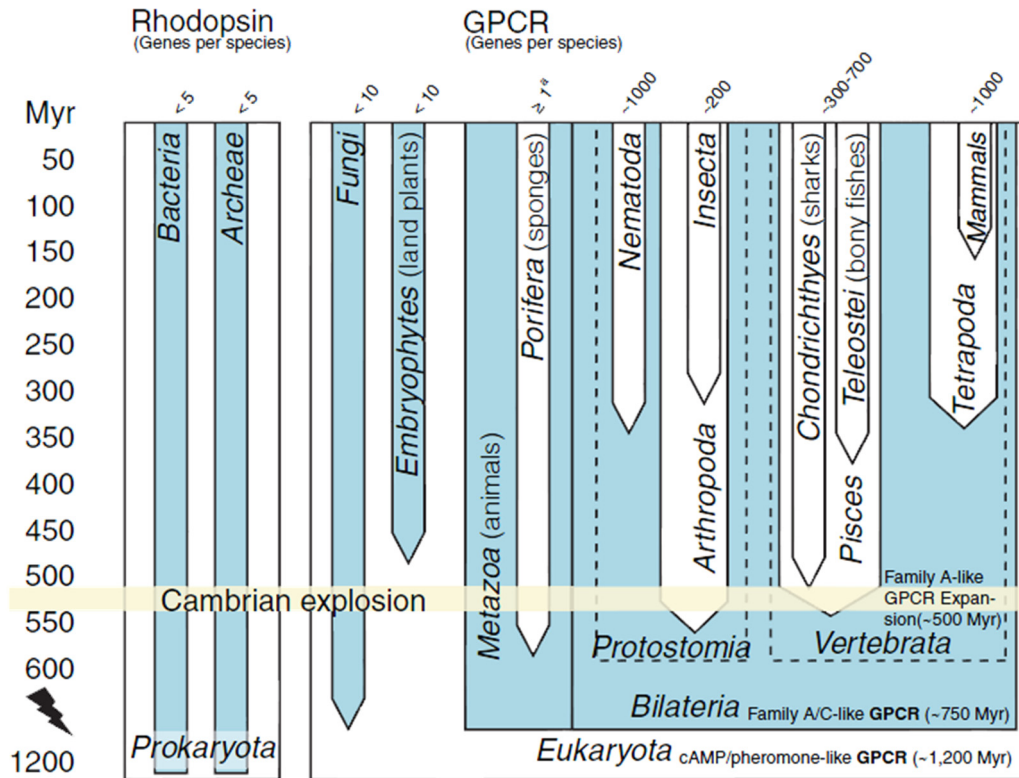
εξελικτική προέλευση των GPCRs και η καταγωγή τους εξακολουθούν να αποτελούν μείζον θέμα αντιπαράθεσης (Rompler, et al., 2007).

Πρωτεΐνες με 7 διαμεμβρανικές α -έλικες είναι παρούσες και στους προκαρυωτικούς οργανισμούς, π.χ. βακτηριοροδοψίνες. Τα κοινά δομικά και λειτουργικά χαρακτηριστικά μεταξύ προκαρυωτικών και ευκαρυωτικών ροδοψινών προτείνουν μία κοινή καταγωγή. Όμως, παρά τις ομοιότητες αυτές, συγκρίσεις των αμινοξικών ακολουθιών τους δε μπορούν να δώσουν ικανοποιητική απόδειξη για εξελικτική σχέση μεταξύ προκαρυωτικών ροδοψινών και GPCRs των ευκαρυωτικών (Sorpa, 1994).

Η μεταγωγή σήματος μέσω G-πρωτεϊνών είναι καθοριστικής σημασίας για το χαρακτηρισμό των GPCRs. Πολλές ομάδες σκωλήκων (π.χ. νηματώδεις και τρηματώδεις) αποτελούν τις εξελικτικά αρχαιότερες υπάρχουσες γενεαλογίες στις οποίες η φωτούποδοχή γίνεται μέσω ροδοψίνης και G-πρωτεϊνών (Hoffmann, et al., 2001). Αυτοί οι οργανισμοί, όμως, εμφανίστηκαν αρκετά μετά την εμφάνιση των GPCRs και των G-πρωτεϊνών στην εξέλιξη. Δομικά και λειτουργικά δεδομένα δείχνουν ξεκάθαρα ότι η μεταγωγή σήματος μέσω GPCRs και G-πρωτεϊνών είναι παρούσα σε μύκητες (Elion, 2000), φυτά (Fujisawa, et al., 2001) και μονοκύτταρους οργανισμούς, όπως το *Dictyostelium discoideum* (Kim, et al., 1996) (π.χ. υποδοχείς φερομονών στους μύκητες, υποδοχείς που μοιάζουν με τους υποδοχείς cAMP και υποδοχείς που μοιάζουν με τους μεταβοτροπικούς υποδοχείς του γλουταμικού). GPCRs που σχετίζονται δομικά με τους υποδοχείς που μοιάζουν με αυτούς του cAMP και με τους υποδοχείς προσκόλλησης (Adhesion) έχουν βρεθεί σε φυτικά γονιδιώματα, ενώ υποδοχείς που μοιάζουν με τους μεταβοτροπικούς υποδοχείς του γλουταμικού έχουν βρεθεί στο *Dictyostelium discoideum* και στο σπόγγο *Geodia cydonium* (Pin, et al., 2003), τα οποία απέκλιναν πριν από 600 εκατομμύρια χρόνια. Σημειώνεται ότι οι σπόγγοι (*Porifera*) αποτελούν τα πιο πρωτόγονα μετάζωα. Αντίθετα, υποδοχείς που μοιάζουν με Ροδοψίνη δεν φαίνεται να εκτείνονται φυλογενετικά πριν από την εμφάνιση (580-800 εκατομμύρια χρόνια πριν) των Αμφιπλεύρων, όπως εντόμων, μαλακίων, νηματωδών και τρηματωδών (Benton and Ayala, 2003; Feng, et al., 1997; Peterson and Butterfield, 2005).

Δεδομένης της εμφάνισης της «σύζευξης» μεταξύ G-πρωτεϊνών και πρωτεϊνών με 7 διαμεμβρανικά τμήματα πριν τη διαίρεση φυτών/μυκήτων/ζώων (~ 1,2 δισεκατομμύρια χρόνια πριν) και της πρώτης εμφάνισης των υποδοχέων που μοιάζουν με Ροδοψίνη νωρίς στην εξέλιξη των μετazών, φαίνεται ότι το φωτοαισθητήριο σύστημα που βασίζεται στη

ρετινάλη είναι αποτέλεσμα συγκλίνουσας εξέλιξης (εκ νέου ανακάλυψη). Οι υποδοχείς της Σεροτονίνης φαίνεται να είναι οι παλαιότεροι υποδοχείς της Κατηγορίας A μιας και εμφανίζονται σε πολλούς νηματώδεις (Tierney, 2001) και πλανάρια (Saitoh, et al., 1997).



Εικόνα 1.9: Σχηματική αναπαράσταση της εξέλιξης των GPCRs (Rompler, et al., 2007)

Συμπερασματικά, οι υποδοχείς με 7 διαμεμβρανικά τμήματα εμφανίστηκαν πολύ νωρίς κατά την εξέλιξη των ευκαρυωτικών οργανισμών και, πολύ σύντομα, έγιναν οι βασικές και πλέον ευπροσάρμοστες συνιστώσες της μεταγωγής σήματος μέσω G-πρωτεϊνών. Οι GPCRs, σε αντίθεση με τα υπόλοιπα μόρια που συμμετέχουν σε αυτόν τον τύπο μεταγωγής σήματος, εξελίχθηκαν σε μια μεγάλη ποικιλία οικογενειών και, πλέον, αριθμούν εκατοντάδες μέλη στα γονιδιώματα των Χορδωτών. Η μεγάλη ποικιλομορφία των GPCRs, σε συνδυασμό με τη μεγάλη συντήρηση που εμφανίζουν οι G-πρωτεΐνες, φαίνεται να οδηγούν στην ποικιλία και το πλήθος, αλλά και την ακρίβεια του μηχανισμού, στα επαγόμενα μονοπάτια μεταγωγής σήματος που παρατηρούνται σήμερα. Φαίνεται ότι η δυναμική εξέλιξη των GPCRs συνεχίζεται ακόμη, και τα πρόσφατα αποτελέσματα της διαδικασίας αυτής μπορούν να γίνουν ορατά με χρήση σύγχρονων μεθόδων γενετικής πληθυσμού (Strotmann, et al., 2011).

1.3 G-πρωτεΐνες

1.3.1 Γενικά

Οι ετεροτριμερείς G-πρωτεΐνες παρέχουν τον ενδιάμεσο σύνδεσμο μεταξύ των GPCRs της κυτταροπλασματικής μεμβράνης και των μορίων εκτελεστών, τα οποία μεταβιβάζουν την πληροφορία για την παρουσία εξωτερικού ερεθίσματος στο εσωτερικό του κυττάρου (Luttrell, 2006). Ουσιαστικά, οι G-πρωτεΐνες αποτελούν «μοριακούς διακόπτες», οι οποίοι ενεργοποιούν ενδοκυτταρικούς σηματοδοτικούς καταρράκτες ως απόκριση στην ενεργοποίηση των GPCRs από εξωκυτταρικά ερεθίσματα. Γι' αυτό το λόγο, οι πρωτεΐνες αυτές διαδραματίζουν ζωτικής σημασίας ρόλο στον καθορισμό της ειδικότητας και των προσωρινών χαρακτηριστικών της κυτταρικής απόκρισης (Oldham and Hamm, 2008).

Οι G-πρωτεΐνες είναι μια οικογένεια GTPασών, οι οποίες μοιράζονται κοινή δομή, αποτελούμενη από διαφορετικές υπομονάδες. Είναι ετεροτριμερείς και αποτελούνται από α, β και γ υπομονάδες. Η λειτουργία τους ως «μοριακών διακοπών» εξαρτάται από την ικανότητα της Gα υπομονάδας να εναλλάσσεται μεταξύ δύο στερεοδιατάξεων: την ανενεργό, GDP-συνδεδεμένη, στερεοδιάταξη, η οποία προετοιμάζει την αλληλεπίδραση με έναν ανενεργό υποδοχέα και την ενεργό, GTP-συνδεδεμένη, στερεοδιάταξη, η οποία μπορεί να διαφοροποιήσει τη λειτουργία μορίων εκτελεστών (βλέπε κύκλο ενεργοποίησης-απενεργοποίησης G-πρωτεϊνών στη συνέχεια), (Oldham and Hamm, 2008). Κατά την παρουσία του Gβγ ετεροδιμερούς και λόγω της αλληλεπίδρασης με λιπαρά οξέα, η Gα υπομονάδα παραμένει ισχυρά συνδεδεμένη στην πλασματική μεμβράνη. Το Gβγ ετεροδιμερές είναι υπεύθυνο για τη συσχέτιση και την ειδική αναγνώριση της G-πρωτεΐνης με τον υποδοχέα (Hollmann, et al., 2005).

Η ενεργοποίηση των G-πρωτεϊνών οδηγεί στο διαχωρισμό της Gα υπομονάδας από το Gβγ ετεροδιμερές, τα οποία και τα δύο μπορούν να αλληλεπιδράσουν με μια ποικιλία μορίων εκτελεστών. Επειδή η πρόσδεση στο ίδιο μόριο εκτελεστή μπορεί να έχει συνεργατικό, προσθετικό ή ακόμα και αντίθετο αποτέλεσμα, οι κυτταρικές αποκρίσεις στην ενεργοποίηση από GPCRs μπορεί να είναι πολύ περίπλοκες (Hollmann, et al., 2005).

Σήμερα, ένας τεράστιος όγκος δομικών πληροφοριών είναι διαθέσιμος για τις πρωτεΐνες αυτές, προερχόμενος από μελέτες κυρίως με χρήση κρυσταλλογραφίας ακτίνων-Χ. Οι πληροφορίες αυτές κατέστησαν δυνατή την κατανόηση των αλλαγών στις στερεοδιατάξεις των α υπομονάδων με τη διαμεσολάβηση του GTP, την αλληλεπίδραση με τα μόρια εκτελεστές και το μηχανισμό υδρόλυσης του GTP (Oldham and Hamm, 2006; Sprang, 1997). Αντίθετα, λίγα

είναι γνωστά για την αλληλεπίδραση μεταξύ του υποδοχέα και της G-πρωτεΐνης, για τον τρόπο με τον οποίο αυτή η αλληλεπίδραση οδηγεί στην απελευθέρωση του GDP και για τις συνέπειες της πρόσδεσης του GTP σε ένα «άδειο» σύμπλεγμα υποδοχέα – G-πρωτεΐνης (Oldham and Hamm, 2008).

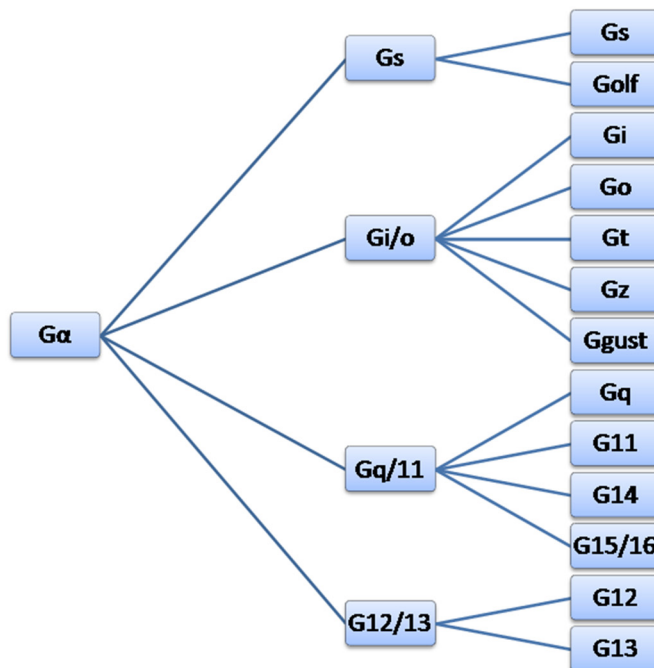
1.3.2 Ταξινόμηση των G-πρωτεϊνών

Σε αντίθεση με την τεράστια ποικιλομορφία των GPCRs σε επίπεδο αμινοξικής ακολουθίας, υπάρχουν συγκριτικά λίγα γονίδια που κωδικοποιούν για την κάθε υπομονάδα των ετεροτριμερών G-πρωτεϊνών (Downes and Gautam, 1999). Υπάρχουν 16 γνωστά γονίδια σήμερα, στα θηλαστικά, που κωδικοποιούν την G_α υπομονάδα, από τα οποία με εναλλακτικό μάτισμα (alternative splicing) εκφράζονται 21 διακριτές G_α υπομονάδες (Cabrera-Vera, et al., 2003; Downes and Gautam, 1999).

Η ονοματολογία και η λειτουργικότητα των G-πρωτεϊνών καθορίζεται από τις α υπομονάδες τους, οι οποίες ταξινομούνται σε τέσσερις οικογένειες ανάλογα με την επί τις εκατό ομολογία των αμινοξικών τους καταλοίπων: G_{α_s} , $G_{\alpha_{i/o}}$, $G_{\alpha_{q/11}}$ και $G_{\alpha_{12/13}}$ (Cabrera-Vera, et al., 2003).

Κάθε μια από τις τέσσερις οικογένειες αποτελείται από αντίστοιχες υπο-οικογένειες (Εικόνα 1.10) (Cabrera-Vera, et al., 2003).

- Η οικογένεια G_{α_s} (stimulatory, ενεργοποίηση αδενυλικών κυκλασών) από τις υπο-οικογένειες G_{α_s} και $G_{\alpha_{olf}}$ (olfactory, συνδέονται με υποδοχείς όσφρησης)
- Η οικογένεια $G_{\alpha_{i/o}}$ (inhibitory/other, αναστολή αδενυλικών κυκλασών και άλλες λειτουργίες) από τις υπο-οικογένειες G_{α_i} , G_{α_o} , G_{α_z} (η λειτουργία της δεν είναι απόλυτα γνωστή), G_{α_t} (transducin, μεταφορά οπτικών σημάτων σε συνδυασμό με τη ροδοψίνη στον αμφιβληστροειδή) και $G_{\alpha_{gust}}$ (gustatory, συνδέονται με υποδοχείς γεύσης)
- Η οικογένεια $G_{\alpha_{q/11}}$ από τις υπο-οικογένειες G_{α_q} (ενεργοποίηση φωσφολιπασών C), $G_{\alpha_{11}}$, $G_{\alpha_{14}}$ και $G_{\alpha_{15/16}}$
- Η οικογένεια $G_{\alpha_{12/13}}$ (σημαντικές για τη ρύθμιση του κυτταροσκελετού, της κυτταρικής σύνδεσης και γενικά των διαδικασιών που σχετίζονται με κίνηση) από τις υπο-οικογένειες $G_{\alpha_{12}}$ και $G_{\alpha_{13}}$



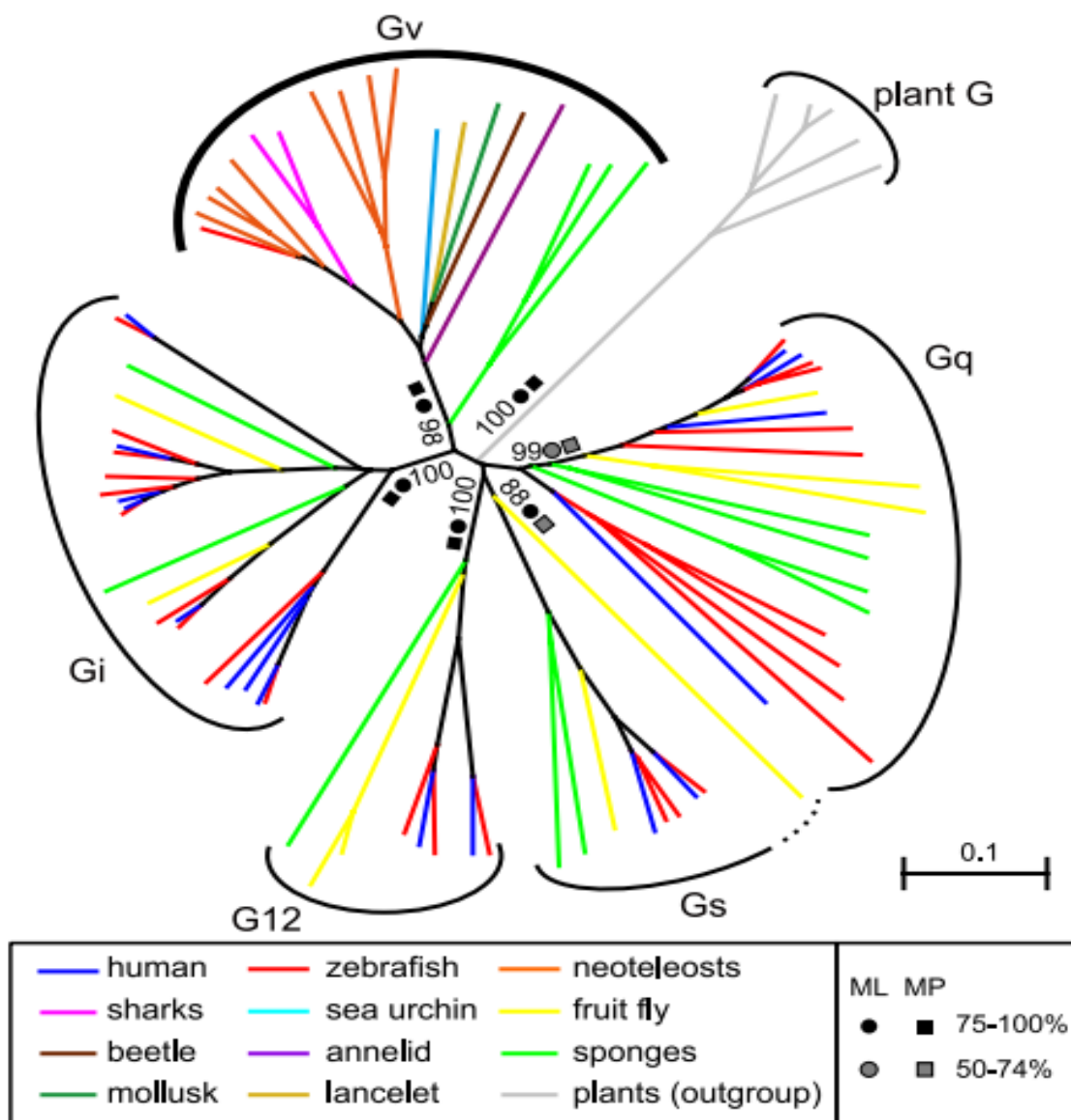
Εικόνα 1.10: Ιεραρχική κατάταξη των G-πρωτεϊνών στα θηλαστικά.

Άλλοι τύποι οικογενειών Gα

Gfa

Μελέτες πάνω στον οργανισμό *Drosophila* κατέδειξαν την ύπαρξη Gα υπομονάδων, η έκφραση των οποίων πραγματοποιούνταν με έναν τρόπο, χωρικά και χρονικά, αυστηρό κατά τη διάρκεια της εμβρυογένεσης του οργανισμού αυτού (Εικόνα 1.11). Σε μεταγενέστερα πειράματα απομονώθηκε μία τέτοιου είδους υπομονάδα, έχοντας 30-38% ομολογία με τις ήδη υπάρχουσες οικογένειες (Quan, et al., 1993). Ο νέος αυτός τύπος Gα αποτελεί μία ξεχωριστή οικογένεια Gα πρωτεϊνών στον οργανισμό *Drosophila* και ονομάστηκε **Gfa**. Τα μετάγραφα της οικογένειας αυτής εκφράζονται κυρίως κατά το εμβρυϊκό, το προνυμφικό στάδιο και στις αρχές του σταδίου της νύμφης και μόνο σε χαμηλά επίπεδα σε ενήλικες μύγες.

γνωστές οικογένειες (Oka, et al., 2009). Παρουσιάζει μεγαλύτερη ομοιότητα με την οικογένεια Gα_i, έχοντας μαζί της απόκλιση συγκρίσιμη με αυτή των Gα_s και Gα₁₂ (Εικόνα 1.12).



Εικόνα 1.12: Φυλογενετική ανάλυση Gγ πρωτεϊνών (Oka, et al., 2009): φυλογενετικό δέντρο (αλγόριθμος NJ) Gα πρωτεϊνών: 16 ανθρώπου (μπλε), 26 ψαριού-ζέβρας (κόκκινο), 9 μύγας (κίτρινο), 2 καρχαρία (ροζ), 8 νεοτελεόστεοι (πορτοκαλί), 1 κεφαλοχορδωτά (ώχρα), 1 αχινού (γαλάζιο), 1 δακτυλοειδή σκώληκα (μωβ) 1 οστρακόδερμου (σκούρο πράσινο), 1 σκαθαριού (καφέ), και 12 σπόγγων (πράσινο). Οι Gα 5 φυτών θεωρούνται ως εξω-ομάδα (outgroup) (γκρι). Η διακεκομμένη καμπύλη, αντιστοιχεί στην Gβ της μύγας και προσωρινά ανατίθεται ως Gα_s (Quan, et al., 1993). Το δέντρο περιλαμβάνει και τις 5 οικογένειες.

Αντίστοιχα, γονιδιακές μελέτες προτείνουν την ομαδοποίηση των Gβ υπομονάδων σε δύο μεγάλες οικογένειες (Downes and Gautam, 1999):

- Gβ₁ – like: περιλαμβάνει τις Gβ₁ ως Gβ₄ των θηλαστικών,

- Gβ₅ – like: περιλαμβάνει την Gβ₅ των θηλαστικών και την ανθρώπινη Gβ_{5L} που προκύπτει από εναλλακτικό μάτισμα τους.

Τέλος, οι 12 Gγ υπομονάδες ομαδοποιούνται σε τέσσερις οικογένειες, τις Group I – γ₁ (Gγ₁, Gγ_c και Gγ₁₁), Group II – γ₂ (Gγ₂, Gγ₃, Gγ₄), Group III – γ₇ (Gγ₇, Gγ₁₂) και Group IV – γ₅ (Gγ₅, Gγ₈, Gγ₁₀) (Downes and Gautam, 1999).

Αν και οι περισσότερες Gβ υπομονάδες μπορούν να αλληλεπιδράσουν με τις περισσότερες Gγ υπομονάδες, δεν παρατηρούνται όλοι οι πιθανοί συνδυασμοί τους. Επιπλέον, πολλά ετεροδιμερή Gβγ έχουν βρεθεί να αλληλεπιδρούν με την ίδια ισομορφή Gα, κάτι που σημαίνει ότι η διαφορική έκφραση ή η υποκυτταρική κατανομή είναι σημαντική στη ρύθμιση των οδών σηματοδότησης. Επιπρόσθετα, τα δεδομένα δείχνουν πως, ενώ η Gγ δεν φαίνεται να συμμετέχει σε αλληλεπιδράσεις με άλλες πρωτεΐνες πέρα από την Gβ, επηρεάζει την επιλεκτικότητα του ετεροδιμερούς απέναντι τόσο στους εκτελεστές όσο και στους υποδοχείς που προσδένει.

1.3.3 Δομή των G-πρωτεϊνών

Μια σειρά από κρυσταλλικές δομές, τόσο για τις υπομονάδες Gα, όσο και για τα ετεροδιμερή Gβγ υπάρχουν διαθέσιμες στην PDB, προσφέροντας αρκετές πληροφορίες για τα δομικά χαρακτηριστικά των G-πρωτεϊνών και για τη φύση διάφορων αλληλεπιδράσεών τους.

Gα υπομονάδα

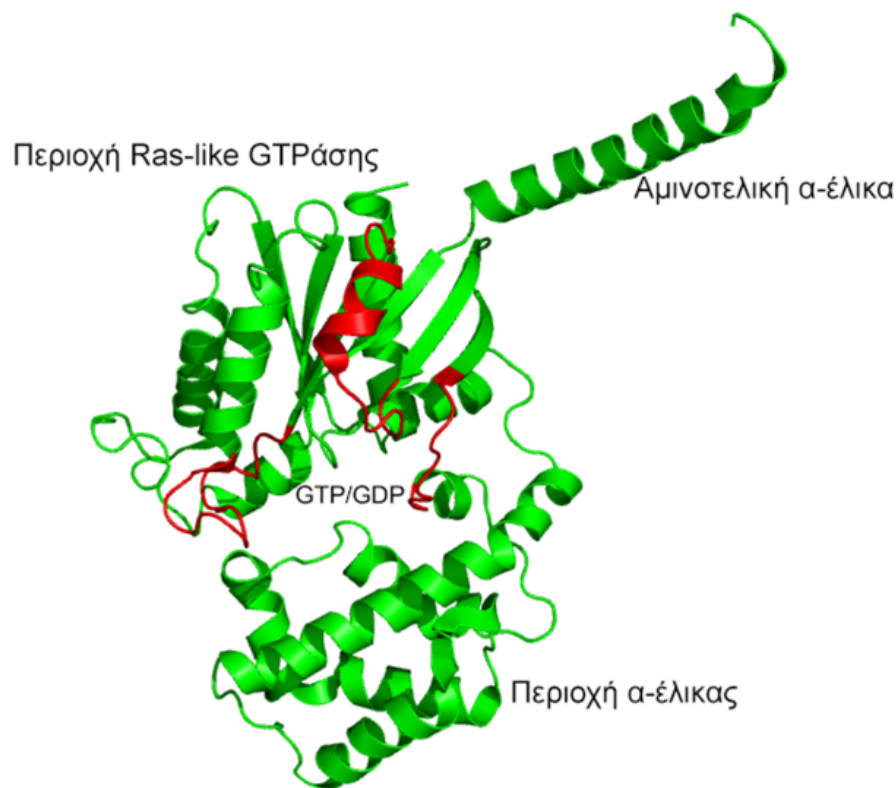
Η Gα υπομονάδα έχει μήκος 350-390 αμινοξικά κατάλοιπα. Αποτελείται από τρεις αυτοτελείς δομικές περιοχές/στοιχεία (domains)

1. Την αμινοτελική α-έλικα
2. Την περιοχή της Ras-like GTPάσης
3. Την περιοχή της α-έλικας

Το νουκλεοτίδιο φαίνεται ότι συνδέεται σε μια κοιλότητα μεταξύ της περιοχής της GTPάσης και της περιοχής της α-έλικας (Εικόνα 1.13) (Cabrera-Vera, et al., 2003; Hollmann, et al., 2005; Lambright, et al., 1996).

Η περιοχή της Ras-like GTPάσης αποτελείται από έξι α-έλικες που περιβάλλουν μία β-πτυχωτή επιφάνεια/φύλλο η οποία αποτελείται από έξι β-κλώνους. Η περιοχή της α-έλικας αποτελείται από πέντε α-έλικες που περιβάλλουν μια μεγαλύτερη α-έλικα. Η αμινοτελική περιοχή

αποτελείται από μία α έλικα (Cabrerera-Vera, et al., 2003; Hollmann, et al., 2005; Lambright, et al., 1996).

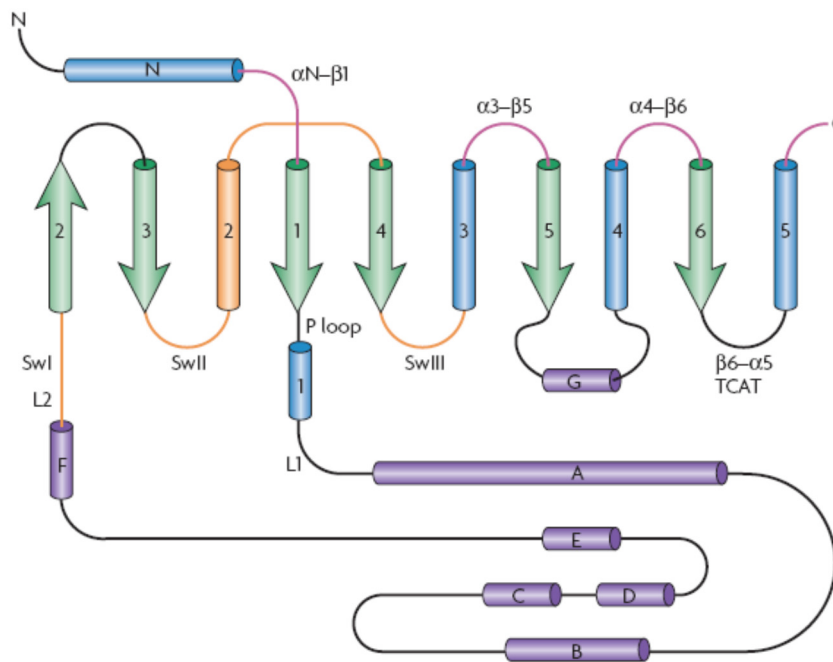


Εικόνα 1.13: Τρισδιάστατη δομή της $G\alpha$ υπομονάδας (PDBID:1GP2). Διακρίνονται τα τρία αυτοτελή δομικά στοιχεία (domains). Με **κόκκινο** είναι χρωματισμένοι οι διακόπτες I, II και III (switches).

Οι $G\alpha$ υπομονάδες εμφανίζουν ομολογία σε επίπεδο αμινοξικής ακολουθίας σε ποσοστό της τάξης του 40%, το οποίο αντιστοιχεί, κατά κύριο λόγο, στις περιοχές της πρωτεΐνης που σχηματίζουν τη κοιλότητα πρόσδεσης του νουκλεοτιδίου της γουανίνης. Η περιοχή της Ras-like GTPάσης είναι συντηρημένη σε όλα τα μέλη της υπεροικογένειας των G-πρωτεϊνών, συμπεριλαμβανομένων των μονομερών G-πρωτεϊνών και των παραγόντων επιμήκυνσης. Αυτή η περιοχή υδρολύει το GTP και παρέχει τις επιφάνειες πρόσδεσης για το $G\beta\gamma$ ετεροδιμερές, τους GPCRs και τα μόρια εκτελεστές. Επιπλέον, αυτή η περιοχή περιλαμβάνει και τρεις εύκαμπτους βρόχους, που ονομάζονται διακόπτες I, II και III (switches) (Εικόνες 1.13 και 1.14), στις οποίες έχουν παρατηρηθεί σημαντικές δομικές διαφορές μεταξύ της στερεοδιάταξης της, GDP-συνδεδεμένης, $G\alpha$ υπομονάδας (Lambright, et al., 1994; Mixon, et al., 1995) και της στερεοδιάταξης της, GTP-συνδεδεμένης, $G\alpha$ υπομονάδας (Coleman, et al., 1994; Noel, et al., 1993). Η περιοχή α -έλικας είναι μοναδική για της $G\alpha$ πρωτεΐνες και σχηματίζει ένα κάλυμμα

πάνω από τη κοιλότητα πρόσδεσης του νουκλεοτιδίου, «θάβοντας» τα συνδεδεμένα νουκλεοτίδια στον πυρήνα της πρωτεΐνης (Oldham and Hamm, 2008). Μεγαλύτερη απόκλιση σε επίπεδο ακολουθίας εμφανίζεται στο αμινοτελικό άκρο, το οποίο είναι υπεύθυνο για τη σύνδεση με το Gβγ ετεροδιμερές, και στο καρβοξυτελικό άκρο, το οποίο συμμετέχει στη σύνδεση, τόσο με τους υποδοχείς, όσο και με τα μόρια εκτελεστές (Luttrell, 2006).

Όλες οι Gα υπομονάδες, εκτός της Gα_t, τροποποιούνται μετά-μεταγραφικά με την προσθήκη ενός παλμιτικού οξέος στο αμινοτελικό άκρο. Στα μέλη της Gα_{i/o} οικογένειας στο αμινοτελικό άκρο προστίθεται επίσης και μυριστικό οξύ. Σε αντίθεση με τη μυριστοϋλίωση, η παλμιτοϋλίωση είναι μια αντιστρεπτή τροποποίηση. Κατά την πρόσδεση του GTP και τον αποχωρισμό του Gβγ ετεροδιμερούς, το παλμιτικό οξύ αποκόπτεται. Επανασύνδεση της Gα υπομονάδας και του Gβγ ετεροδιμερούς προωθεί την επαναπαλμιτοϋλίωση. Αυτές οι τροποποιήσεις ρυθμίζουν τον εντοπισμό στη μεμβράνη και τις αλληλεπιδράσεις πρωτεϊνών-πρωτεϊνών (Chen and Manning, 2001; Smotrys and Linder, 2004).

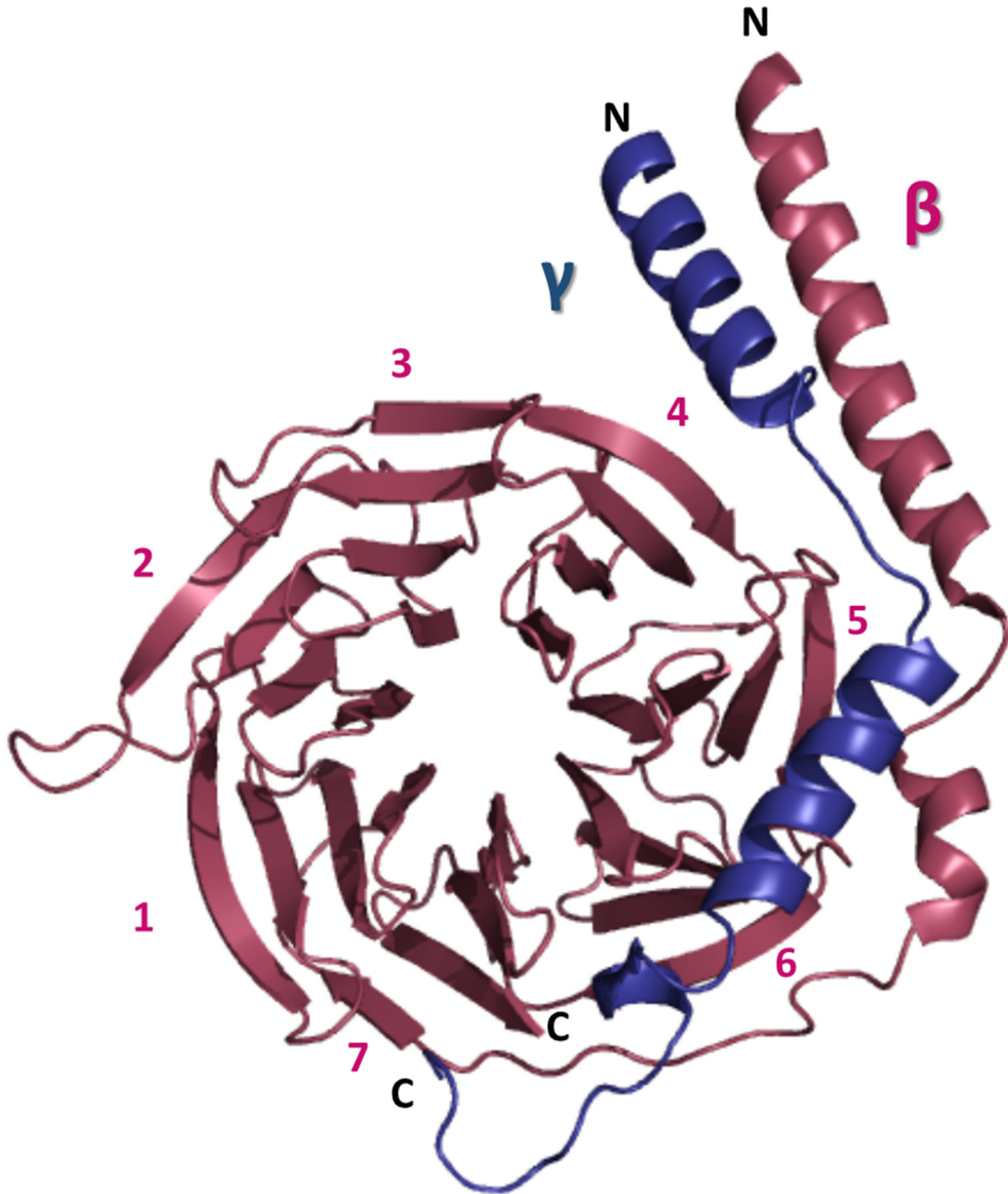


Εικόνα 1.14: Σχηματική αναπαράσταση της δευτεροταγούς δομής της Gα υπομονάδας. Οι α έλικες δείχνονται με κυλίνδρους και τα β φύλλα με βέλη. Η περιοχή της α έλικας αποτελείται από 7 α έλικες (μωβ). Οι σύνδεσμοι L1 και L2 συνδέουν την περιοχή α έλικας με την περιοχή GTPάσης. Οι περιοχές διακόπτες (πορτοκαλί) και οι περιοχές επαφής με τον υποδοχέα (ροζ) είναι τονισμένες. Το μοτίβο TCAT (υπεύθυνο για την πρόσδεση του νουκλεοτιδίου) στο β6-α5 βρόχο βρίσκεται απέναντι από, το υπεύθυνο για την πρόσδεση στον υποδοχέα, καρβοξυτελικό άκρο που βρίσκεται στην α5 έλικα. Ο υπεύθυνος για την πρόσδεση του φωσφο-νουκλεοτιδίου βρόχος (P) συνδέει το β1 φύλλο με την α1 έλικα (Oldham and Hamm, 2008).

βγ ετεροδιμερές

Οι υπομονάδες β και γ βρίσκονται πάντα ενωμένες ως ετεροδιμερές, σχηματίζοντας μια λειτουργική μονάδα, η οποία μπορεί να διαχωριστεί μόνο κάτω από συνθήκες αποδιάταξης (Εικόνα 1.15) (Schmidt, et al., 1992). Το δίπλωμα της Gβ υπομονάδας είναι χαρακτηριστικό και ονομάζεται “7 bladed β propeller” (Εικόνα 1.15), αποτελούμενο από επτά WD40 επαναλήψεις, ενώ παρουσιάζει και μία ιδιαίτερα συντηρημένη αμινοτελική περιοχή (50-83%), η οποία σχηματίζει μια υπερέλικα α-ελίκων (coiled-coil) με το αμινοτελικό άκρο της Gγ υπομονάδας (Cabrera-Vera, et al., 2003; Oldham and Hamm, 2006; Oldham and Hamm, 2008). Το καρβοξυτελικό άκρο της Gγ υπομονάδας προσδένεται στις «λεπίδες» 5 και 6 της Gβ υπομονάδας (Sondek, et al., 1996; Wall, et al., 1995). Όλες οι Gγ υπομονάδες επιδέχονται μετα-μεταφραστική ισοπρενυλίωση στο καρβοξυτελικό τους άκρο και, συγκεκριμένα, στο συντηρημένο Cys-Ala-Ala-X μοτίβο, ώστε να εξυπηρετείται η σύνδεση του ετεροδιμερούς με τη μεμβράνη (Zhang and Casey, 1996).

Οι δομές ετεροτριμερών G-πρωτεϊνών που έχουν λυθεί κρυσταλλογραφικά (Lambright, et al., 1994; Nishimura, et al., 2010; Wall, et al., 1995) δείχνουν δύο περιοχές αλληλεπίδρασης ανάμεσα στην Gα υπομονάδα και το ετεροδιμερές Gβγ. Η πρωταρχική επιφάνεια αλληλεπίδρασης στην GDP-συνδεδεμένη Gα υπομονάδα είναι μια υδρόφοβη κοιλότητα σχηματιζόμενη από τις περιοχές Switch I και II, με μικρότερη συνεισφορά της αμινοτελικής έλικας, ενώ από το Gβγ συμμετέχουν οι κορυφές των λεπίδων WD40. Τα κρυσταλλογραφικά δεδομένα δεν υποδεικνύουν κάποια άμεση αλληλεπίδραση της Gα υπομονάδας με την Gγ, αν και τα ακυλιωμένα αμινοτελικά και καρβοξυτελικά άκρα της Gα και της Gγ, αντίστοιχα, βρίσκονται αρκετά κοντά και πιθανόν να εισχωρούν συνεργατικά στην πλασματική μεμβράνη. Οι συγκεκριμένες μετατροπές αυξάνουν, επίσης, τη συγγένεια της Gα με το διμερές Gβγ, κάτι που μπορεί να είναι ανεξάρτητο από τη μεμβράνη (Oldham and Hamm, 2008).

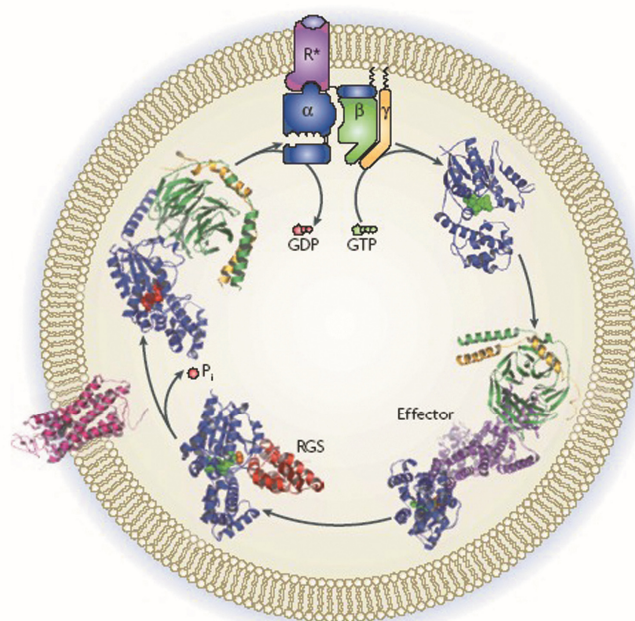


Εικόνα 1.15: Τρισδιάστατη δομή του Gβγ ετεροδιμερούς (PDBID:1GP2). Επισημαίνονται τα αμινοτελικά και καρβοξυτελικά άκρα των δύο υπομονάδων, καθώς και οι «λεπίδες» της Gβ υπομονάδας.

1.3.4 Κύκλος ενεργοποίησης G-πρωτεϊνών και ο ρόλος των πρωτεϊνών RGS

Όλες οι ετεροτριμερείς G-πρωτεΐνες ακολουθούν τον ίδιο κύκλο ενεργοποίησης-απενεργοποίησης (Εικόνα 1.16). Όταν το GDP συνδέεται με την Gα υπομονάδα επιτρέπει τη σύνδεση της με το Gβγ ετεροδιμερές και την απενεργοποίηση της ετεροτριμερούς πρωτεΐνης.

Η πρόσδεση του ετεροδιμερούς Gβγ στην Gα υπομονάδα ενισχύει την πρόσδεση της στον GPCR. Με τη σύνδεση του υποκαταστάτη, ο GPCR υπόκειται σε στερεοδιαταξική αλλαγή, αυξάνοντας τη συγγένεια του για τη G-πρωτεΐνη. Αυτό έχει ως αποτέλεσμα την απομάκρυνση του GDP από την Gα υπομονάδα και την αντικατάστασή του από GTP. Η αντικατάσταση αυτή ελαττώνει τη συγγένεια της Gα υπομονάδας για το Gβγ ετεροδιμερές και την αποδιάταξη του ετεροτριμερούς. Η Gα υπομονάδα και το ετεροδιμερές Gβγ στη συνέχεια ενεργοποιούν αντίστοιχα μόρια εκτελεστές. Η δράση της Gα υπομονάδας ως GTPάσης έχει ως αποτέλεσμα τη διάσπαση του GTP σε GDP και την επανασύνδεση και απενεργοποίηση του ετεροτριμερούς (Hollmann, et al., 2005; Kristiansen, 2004; Oldham and Hamm, 2008).



Εικόνα 1.16: Ο κύκλος ενεργοποίησης των G-πρωτεϊνών (Oldham and Hamm, 2006).

Συχνά σε αυτό το μηχανισμό συμμετέχουν συμπληρωματικά ειδικές πρωτεΐνες με την ιδιότητα να επιταχύνουν την υδρόλυση του GTP (Oldham and Hamm, 2008). Πρόκειται για τις πρωτεΐνες RGS (Regulators of G-protein Signalling), οι οποίες επάγουν τη δραστηριότητα GTPάσης, επιταχύνοντας την υδρόλυση του GTP και την απενεργοποίηση της Gα. Αυτή η λειτουργία επιτελείται από μια δομικά αυτοτελή περιοχή (περιοχή RGS) αποτελούμενη από 7 ως 9 α-έλικες, που εντοπίζεται μία ή περισσότερες φορές σε κάθε πρωτεΐνη. Η οικογένεια των RGS χωρίζεται σε 8 υπο-οικογένειες. Οι τέσσερις πρώτες (R4, R7, RZ και R12) περιλαμβάνουν πρωτεΐνες που έχουν ως κύρια δράση τους τη ρύθμιση της ενεργότητας GTPάσης των Gα. Έχουν βρεθεί επίσης τουλάχιστον τρεις πρωτεΐνες RGS που σχηματίζουν σύμπλοκα με την υπομονάδα Gβ, εμποδίζοντας την πρόσδεση του ετεροδιμερούς με

εκτελεστές. Οι υπόλοιπες τέσσερις υπο-οικογένειες περιλαμβάνουν πρωτεΐνες που παρουσιάζουν δομικές περιοχές παρόμοιες με την RGS, χωρίς όμως να επιδεικνύουν απαραίτητα ενεργότητα GAP (GTPase-activating protein).

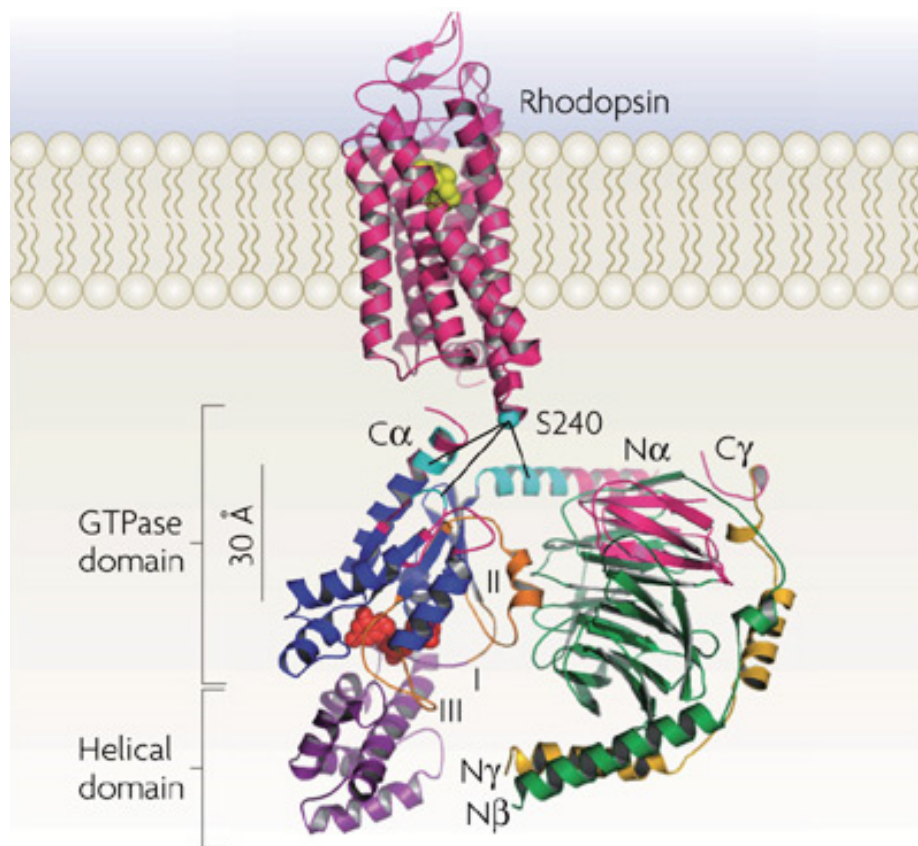
Δεδομένα από μελέτες cross-linking αλλά και από κρυσταλλικές δομές (Slep, et al., 2001; Soundararajan, et al., 2008) υποδεικνύουν ότι σε κάποιες περιπτώσεις οι πρωτεΐνες RGS προσδένονται στην επιφάνεια των Gα ταυτόχρονα με μόρια εκτελεστές. Επιπλέον, κάποιοι εκτελεστές όπως η Φωσφολιπάση C και διάφοροι RhoGEFs (παράγοντες ανταλλαγής νουκλεοτιδίων για πρωτεΐνες Rho) παρουσιάζουν παρόμοια με RGS ενεργότητα προς τις Gα (Chen, et al., 2005; Waldo, et al., 2010).

1.3.5 Η διεπαφή μεταξύ GPCRs και G-πρωτεϊνών

Πολυάριθμες μελέτες έχουν αναγνωρίσει διάφορες περιοχές των G-πρωτεϊνών που συνεισφέρουν στην επιφάνεια αλληλεπίδρασης με τους υποδοχείς (Εικόνες 1.14 και 1.17). Η ενεργοποίηση του υποδοχέα ανοίγει μια κοιλότητα στην κυτταροπλασματική πλευρά για το καρβοξυτελικό άκρο της Gα υπομονάδας (Janz and Farrens, 2004), το οποίο είναι το καλύτερα χαρακτηρισμένο σημείο «επαφής» στις G-πρωτεΐνες. Ένα πεπτιδίο που αντιστοιχεί στα 11 αμινοξικά κατάλοιπα του καρβοξυτελικού άκρου της Gα_t δεν έχει οργανωμένη δομή, αλλά μπορεί να προσδεθεί σε ενεργοποιημένους υποδοχείς ροδοψίνης μέσω επαγόμενης προσαρμογής (Dratz, et al., 1993; Kisselev, et al., 1998; Koenig, et al., 2002), σταθεροποιεί τη μεταροδοψίνη II (ενεργή στερεοδιάταξη της ροδοψίνης που σχηματίζεται μετά το φωτοϊσομερισμό του αμφιβληστροειδούς και καταλύει την ανταλλαγή νουκλεοτιδίου στην Gα_t) και ανταγωνίζεται με το ετεροτριμερές G_t για την πρόσδεση στον υποδοχέα (Aris, et al., 2001; Hamm, et al., 1988). Τα αμινοξικά κατάλοιπα που είναι σημαντικά για την αλληλεπίδραση με την κοιλότητα του υποδοχέα που ανοίγει με την ενεργοποίησή του έχουν αποκαλυφθεί με τη χρήση μεταλλαγμένων πεπτιδίων που αντιστοιχούν στο καρβοξυτελικό άκρο της Gα υπομονάδας.

Πολλά κατάλοιπα του α4-β6 βρόχου εμπλέκονται επίσης στην επαφή με τον υποδοχέα, όπως έχει δειχθεί με πληθώρα πειραμάτων (Bae, et al., 1997; Bae, et al., 1999; Cai, et al., 2001; Lichtarge, et al., 1996; Mazzoni and Hamm, 1996; Onrust, et al., 1997). Όταν πέντε κατάλοιπα στον α3-β5 βρόχο της Gα_s αντικατασταθούν με τα ομόλογα αμινοξικά κατάλοιπα της Gα_{i2}, η πρόσδεση του GTP με τη μεσολάβηση του β2 αδρενεργικού υποδοχέα μειώνεται, ενώ η συγγένεια με τον υποδοχέα αυξάνεται (Grishina and Berlot, 2000). Η αμινοτελική α έλικα της

Γα υπομονάδας αλληλεπιδρά επίσης με τον υποδοχέα, όπως έχει δειχθεί με διάφορα πειράματα μεταλλαξογένεσης, με χρήση χμαιρών Γα υπομονάδων κ.α. (Ho and Wong, 2000; Itoh, et al., 2001; Kisselev, et al., 1998; Onrust, et al., 1997; Taylor, et al., 1994).



Εικόνα 1.17: Διεπαφή ενός GPCR με μια G-πρωτεΐνη (Oldham and Hamm, 2008).

Το Gβγ ετεροδιμερές προσδέεται και αυτό στον υποδοχέα και είναι απαραίτητο για τη σταθεροποίηση της διεπαφής υποδοχέα – Γα υπομονάδας. Τα 60 καρβοξυτελικά κατάλοιπα της Gβ υπομονάδας μπορούν να διασταυρώνονται σε ένα πεπτίδιο που αντιστοιχεί στον τρίτο ενδοκυττάριο βρόχο του α₂ αδρενεργικού υποδοχέα (Taylor, et al., 1994; Taylor, et al., 1996).

Με διάφορα πειράματα μεταλλαξογένεσης της Gβ υπομονάδας αναγνωρίστηκαν μεταλλαγμένες Gβ υπομονάδες που σχηματίζουν φυσιολογικά ετεροτριμερή, αλλά εμφανίζουν ελαττωματική, μέσω ροδοψίνης, πρόσδεση GTPγS (GTPγS, guanosine 5'-O-[gamma-thio]triphosphate), κάτι που υποδεικνύει, είτε άμεση, είτε έμμεση συμμετοχή του Gβγ ετεροδιμερούς σε αυτή τη διαδικασία, μέσω της σταθεροποίησης της Γα υπομονάδας (Ford, et al., 1998).

Ένα πρενυλιωμένο καρβοξυτελικό δωδεκαπεπτίδιο της Gγ₁ σταθεροποιεί τη μεταροδοψίνη II (Kisselev, et al., 1995). Αυτό το πεπτίδιο της Gγ υπομονάδας σχηματίζει μια αμφιπαθική έλικα

κατά την αλληλεπίδραση με την ενεργοποιημένη ροδοψίνη. Ομοίως, μια πρενυλιωμένη G_s υπομονάδα αναστέλλει την ενεργοποίηση της G-πρωτεΐνης από το μ₂ μουσκαρινικό υποδοχέα (Azria, et al., 1999).

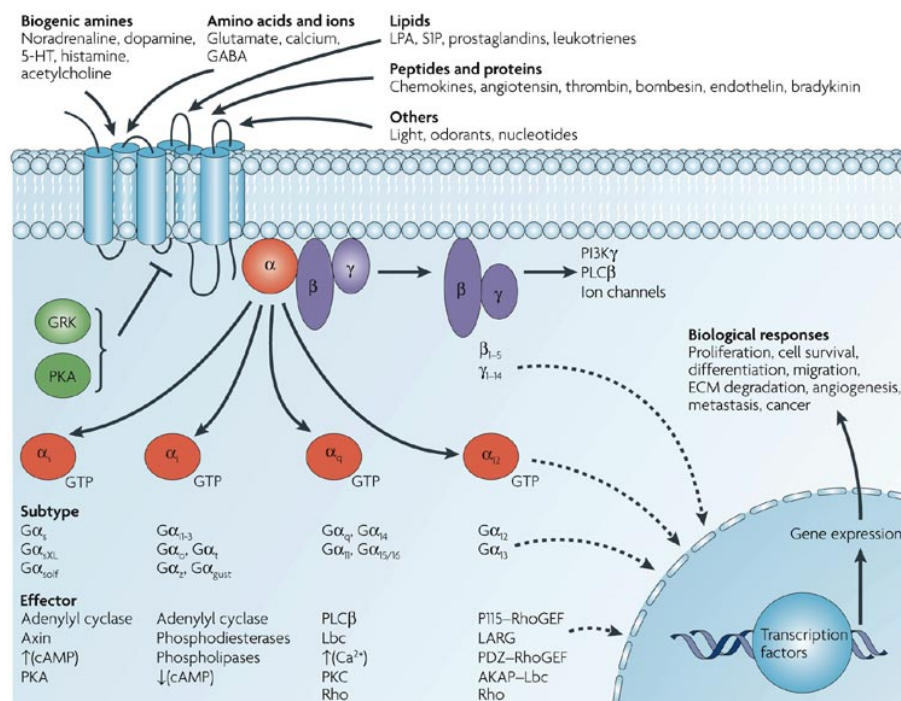
Συλλογικά, όλες αυτές οι μελέτες έχουν εντοπίσει μια πιθανή διεπαφή για την πρόσδεση του υποδοχέα στα ετεροτριμερή των G-πρωτεϊνών. Είναι ενδιαφέρον το γεγονός ότι αυτό το μοριακό αποτύπωμα μοιάζει πολύ μεγάλο για να σχηματίζεται από έναν και μόνο υποδοχέα τη φορά. Αυτή η παρατήρηση υποδηλώνει πιθανή συμμετοχή διμερών υποδοχέων ή/και μια σειρά διαδοχικών αλληλεπιδράσεων που οδηγούν στην ενεργοποίηση των G-πρωτεϊνών.

Έχουν προταθεί δύο μοντέλα που εξηγούν το πώς οι G-πρωτεΐνες έρχονται σε επαφή με ενεργοποιημένους υποδοχείς. Στο μοντέλο “σύζευξης μέσω σύγκρουσης” (collision coupling) αυτές οι αλληλεπιδράσεις προκύπτουν ως αποτέλεσμα ελεύθερης πλευρικής διάχυσης μέσα στην πλασματική μεμβράνη, όπου οι G-πρωτεΐνες αλληλεπιδρούν μόνο με ενεργοποιημένους υποδοχείς. Το εναλλακτικό μοντέλο προτείνει ότι οι G-πρωτεΐνες μπορούν να αλληλεπιδράσουν με τους υποδοχείς πριν την πρόσδεση του υποκαταστάτη, δηλαδή είναι “εκ των προτέρων συζευγμένες” (precoupled). Το δεύτερο μοντέλο υποστηρίζεται από πολλά *in vitro* και *in situ* πειραματικά δεδομένα, και, γενικά, είναι αρκετά ελκυστικό, εφόσον εξηγεί, εν μέρει, την ειδικότητα στις αλληλεπιδράσεις υποδοχέων – G-πρωτεϊνών, ωστόσο, δεν μπορεί να εξηγήσει ικανοποιητικά το φαινόμενο ενίσχυσης του σήματος (Oldham and Hamm, 2008).

1.3.6 Αλληλεπιδράσεις G-πρωτεϊνών με εκτελεστές

Οι εκτελεστές είναι μια ετερογενής ομάδα πρωτεϊνών που, μέσω των αλληλεπιδράσεών τους με τις G-πρωτεΐνες, δρουν ως φορείς δεύτερου μηνύματος ή οδηγούν απ’ ευθείας σε κάποια κυτταρική και φυσιολογική απόκριση (Εικόνα 1.18) (Kristiansen, 2004). Το πρώτο μόριο εκτελεστής που μελετήθηκε εκτενώς είναι η αδενυλική κυκλάση, ένα ένζυμο που παράγει κυκλικό AMP. Η αποσαφήνιση της δράσης του συγκεκριμένου ενζύμου ήταν που οδήγησε τελικά στη μελέτη των διαμεμβρανικών μηχανισμών μεταγωγής σήματος και στην ανακάλυψη των G-πρωτεϊνών (Woehler and Ponimaskin, 2009). Σε αντίθεση με τις G-πρωτεΐνες που είναι συγκεκριμένες σε αριθμό, και τους GPCRs που ακολουθούν συγκεκριμένο πρότυπο δομής και λειτουργίας, οι εκτελεστές είναι μια μεγάλη ομάδα πρωτεϊνών με μεγάλη ποικιλία στις δομές και τις λειτουργίες τους, από κινάσες και στοιχεία

του κυτταροσκελετού ως και ιοντικά κανάλια. Κατά συνέπεια, είναι δύσκολη η διαδικασία ταξινόμησής τους.



Εικόνα 1.18: Αλληλεπιδράσεις GPCRs, G-πρωτεϊνών και εκτελεστών (Dorsam and Gutkind, 2007).

Η κλασική οπτική των σηματοδοτικών μονοπατιών των GPCRs πρότεινε πως ένας υποδοχέας αλληλεπιδρά με ένα ετεροτριμερές G-πρωτεϊνών, οδηγώντας στην ενεργοποίηση ενός εκτελεστή και σε μια συγκεκριμένη κυτταρική απόκριση. Ωστόσο, έχει γίνει πλέον αποδεκτό πως ένας GPCR μπορεί να αλληλεπιδράσει με περισσότερες από μία διαφορετικές υπομονάδες $G\alpha$, και αντίστοιχα ετεροδιμερή $G\beta\gamma$, ενώ διάφορες μελέτες τα τελευταία χρόνια έχουν δείξει πως μια οικογένεια $G\alpha$ ή ετεροδιμερές $G\beta\gamma$ μπορεί να αλληλεπιδράσει με περισσότερους από έναν διαφορετικούς εκτελεστές, και αντίστροφα ένας εκτελεστής μπορεί να ρυθμιστεί από περισσότερες της μίας διαφορετικές G-πρωτεΐνες (Woehler and Ronimaskin, 2009).

1.3.7 Η εξέλιξη των G-πρωτεϊνών

Η δομή των $G\alpha$ ανέπτυξε πολλές, ειδικές περιοχές αλληλεπίδρασης, όπως εκείνες για το $G\beta\gamma$ ετεροδιμερές, τον υποδοχέα και τα μόρια εκτελεστές, πολλές από τις οποίες αλληλοεπικαλύπτονταν μερικώς ή εντελώς. Λόγω αυτής της πολυπλοκότητας στη δομή της $G\alpha$ υπομονάδας, ακόμα και μία απλή μεταλλαγή θα συνοδευόταν πιθανότατα από ποικίλες συνέπειες (Temple, et al., 2010). Όπως ήδη έχει αναφερθεί, συγκεκριμένα η $G\alpha$ υπομονάδα

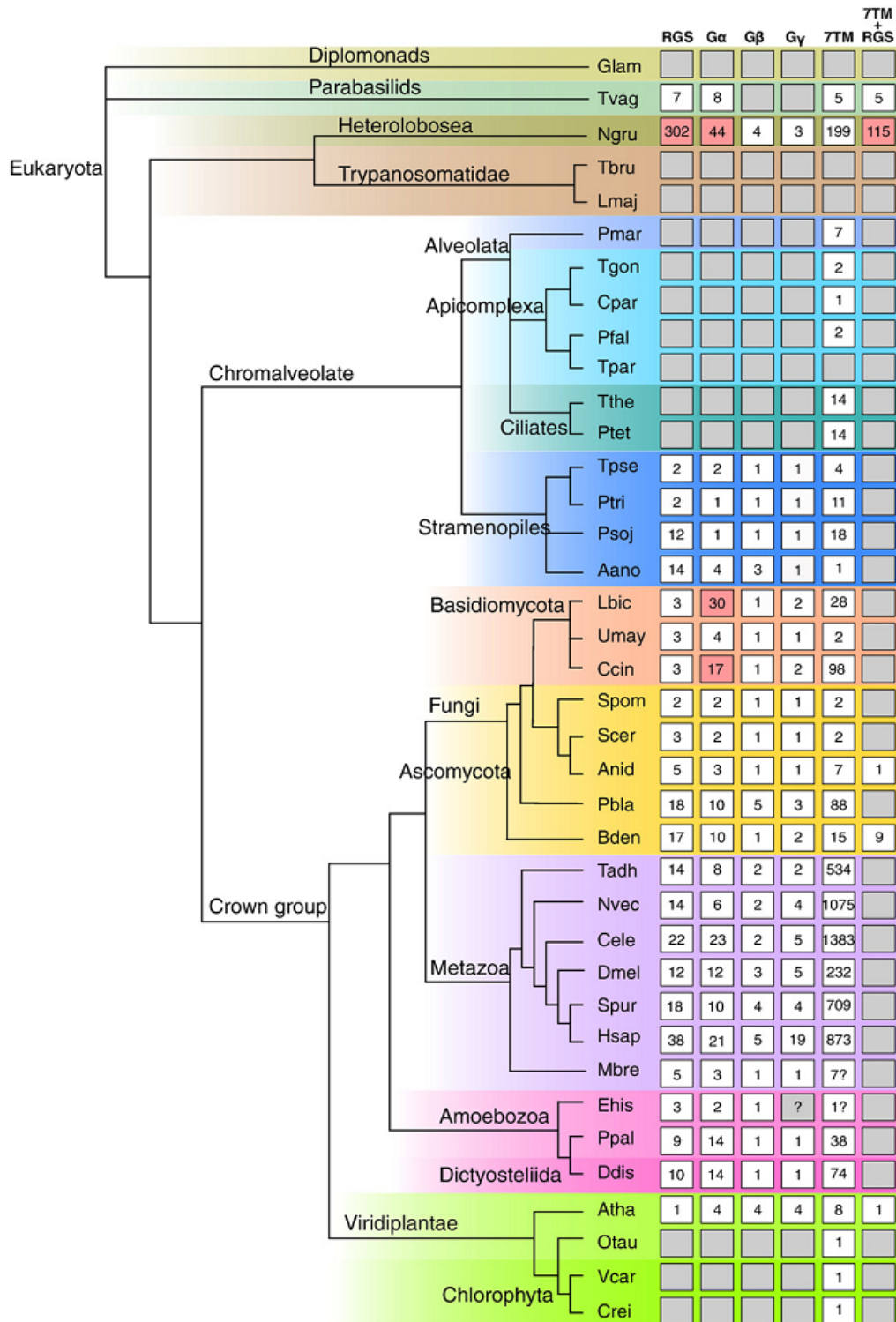
είναι υψηλά συντηρημένη εξαιτίας των πολλών αλληλεπιδράσεων στις οποίες συμμετέχει. Διάφοροι οργανισμοί περιέχουν διαφορετικά σύνολα αλληλεπιδρώντων με Gα πρωτεϊνών, γεγονός που υποδηλώνει ξεκάθαρα πως αλλαγές στην ακολουθία και τη λειτουργικότητα των Gα αποκτήθηκαν με το πέρασμα του χρόνου.

Είναι λοιπόν εύλογο το ερώτημα, δεδομένων των μεγάλων περιορισμών στην εξέλιξη των Gα εξαιτίας των πολλών τους αλληλεπιδράσεων, με ποιο τρόπο τελικά εξελίχθηκαν από μία προγονική υπομονάδα στις τέσσερις κύριες οικογένειες στα θηλαστικά ($G_{i/o}$, G_q , G_s , G_{12}) και σε πολλές υπο-οικογένειες, καθεμία με διακριτές λειτουργικότητες. Ο γονιδιακός διπλασιασμός συνέβαλε καθοριστικά στη διαφοροποίηση των οικογενειών, αλλά πώς αποκτήθηκαν με τον τρόπο αυτό λειτουργικότητες χαρακτηριστικές για κάθε οικογένεια είναι ασαφές (Conant and Wolfe, 2008). Έναν παράγοντα σύγχυσης αποτελεί η σποραδική εμφάνιση αλληλεπιδρώντων με Gα πρωτεϊνών σε όλη την εξέλιξη (Temple, et al., 2010).

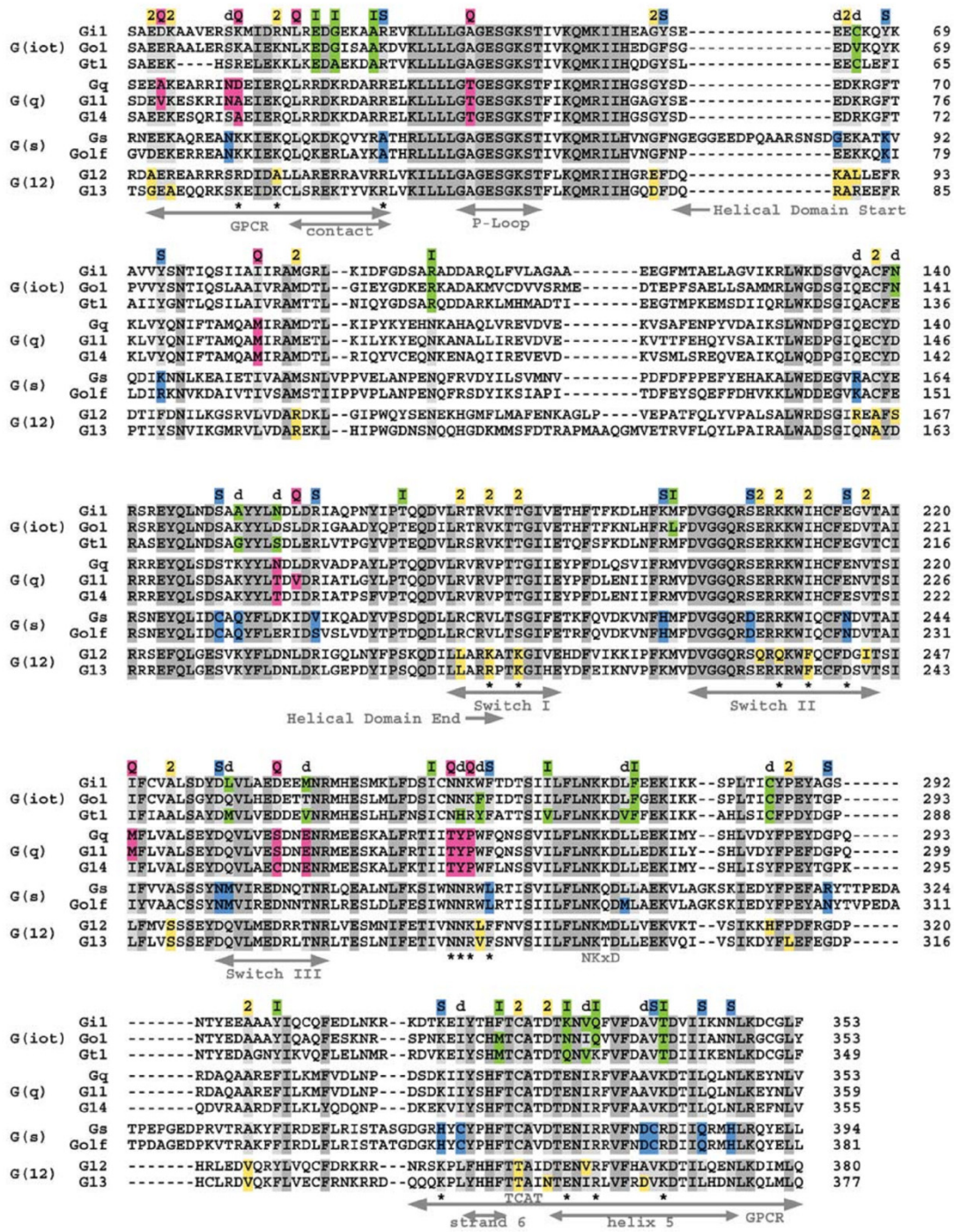
Η ανάλυση των Gα φυλετικών προτύπων υποδεικνύει μία σποραδική αλλά ευρεία παρουσία τους στους ευκαρυωτικούς οργανισμούς (Εικόνα 1.19). Όλα τα μέλη του ευκαρυωτικού κλάδου, συμπεριλαμβανομένων μεταζώων, μυκήτων, αμοιβάδων και φυτών κωδικοποιούν 1-30 παράλογα γονίδια για G-πρωτεΐνες (Aubry and Firtel, 1998; Bastiani and Mendel, 2006; Janetopoulos, et al., 2001; Johnston, et al., 2008; Zhu, et al., 2009). Ο μικρότερος αριθμός αντιπροσώπων παρατηρείται στα φυτά, τα οποία συνήθως κωδικοποιούν για μία τυπική Gα υπομονάδα, ενώ, εκτός από τα ζώα και τους μυξομήκυτες (slime molds), παρατηρήθηκε μεγάλος αριθμός παραλόγων σε μανιτάρια (*Laccaria*) και σε άλλα ήδη μυκήτων.

Λαμβάνοντας υπόψη μας όλα τα παραπάνω, είναι εύκολο να αντιληφθούμε ότι, όπως οι ευκαρυωτικοί οργανισμοί διαφοροποιήθηκαν και απέκτησαν πιο σύνθετες λειτουργίες, ομοίως εξελίχθηκαν και οι G-πρωτεΐνες. Και οι τέσσερις οικογένειες Gα δημιουργήθηκαν νωρίς κατά την εξέλιξη των μεταζώων. Η σημερινή μορφή των Gα προέκυψε έπειτα από πολλαπλούς κύκλους γονιδιακών διπλασιασμών και διαφοροποιήσεων (Wilkie, et al., 1992). Εικάζεται ότι ένα μικρό σύνολο θέσεων στην ακολουθία τους, χαρακτηριστικών για την κάθε οικογένεια, έχει την εξελικτική δυνατότητα να της προσδίδει ξεχωριστή λειτουργικότητα με το μικρότερο εξελικτικό κόστος. Μελέτες λοιπόν έχουν προτείνει πως συγκεκριμένες αλλαγές στην αμινοξική ακολουθία που συνέβησαν νωρίς κατά την απόκτηση χαρακτηριστικής για κάθε οικογένεια λειτουργικότητας, προέκυψαν από τροποποιήσεις της προγονικής λειτουργίας. Οι περισσότερες χαρακτηριστικές για κάθε οικογένεια θέσεις (class-distinctive sites) εντοπίζονται σε περιοχές που είναι ήδη συντηρημένες καθότι είναι

λειτουργικά σημαντικές (π.χ. switches). Τροποποιήσεις στις περιοχές αυτές, σε, παραδόξως, λίγα κατάλοιπα θα μπορούσαν να εμπλουτίσουν την ήδη υπάρχουσα λειτουργία συνεισφέροντας ταυτόχρονα στη διαφοροποίηση (Εικόνα 1.20) (Temple, et al., 2010).



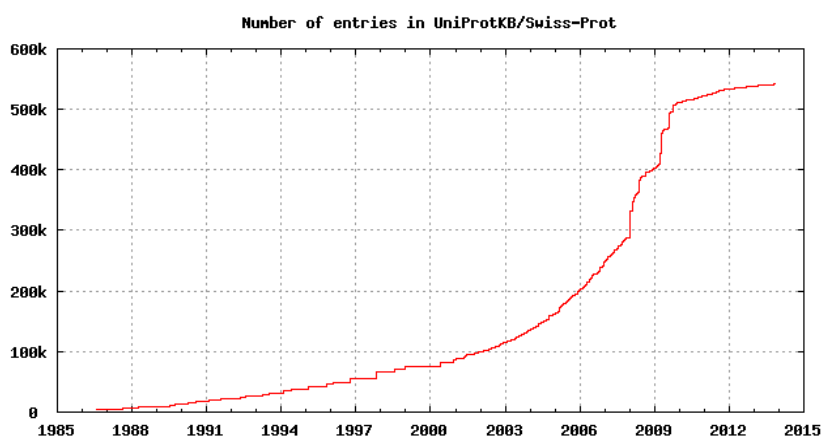
Εικόνα 1.19: Το ευκαρυωτικό φυλογενετικό δέντρο με τους αντίστοιχους αριθμούς πρωτεϊνών για κάθε οργανισμό (Anantharaman, et al., 2011).



Εικόνα 1.20: Στοιχισμός ανθρώπινων Ga υπομονάδων τονίζοντας κατάλοιπα που συντηρούνται σε όλες τις οικογένειες, καθώς και αυτά που διαχωρίζουν τη μία οικογένεια από τις άλλες (class-distinctive sites). Με γκρι χρώμα σημαίνονται τα κατάλοιπα που είναι συντηρημένα μεταξύ των τεσσάρων οικογενειών, ενώ αυτά τα class-distinctive sites είναι όμοια στις 3 από τις 4 οικογένειες. Το πράσινο αντιστοιχεί στην G₁₀ οικογένεια, το ματζέντα στην G_q, το μπλε στην G_s και το πορτοκαλί στην G₁₂ (Temple, et al., 2010).

1.4 Βάσεις Δεδομένων και υπολογιστικά εργαλεία για GPCRs, G-πρωτεΐνες, και τους εκτελεστές τους

Ο ολοένα αυξανόμενος αριθμός γονιδιωμάτων που προσδιορίζονται στις μέρες μας, έχει ως αποτέλεσμα τη συσσώρευση τεράστιου όγκου πληροφοριών για την αλληλουχία χιλιάδων πρωτεϊνών (Pagani, et al., 2012). Παρά τη συσσώρευση πληροφοριών, ο προσδιορισμός της λειτουργίας των πρωτεϊνών δεν είναι τόσο εύκολος, και έτσι περίπου το 50% των πρωτεϊνών έχουν άγνωστη λειτουργία (von Heijne, 1999). Το πρόβλημα προσδιορισμού της λειτουργίας των πρωτεϊνών εντείνεται, καθώς οι ακολουθίες (γονιδίων και πρωτεϊνών) που βρίσκονται κατατεθειμένες στις δημόσιες βάσεις δεδομένων αυξάνονται με γεωμετρικό ρυθμό (Boeckmann, et al., 2003; UniProt, 2013) (Εικόνα 1.21), όπως αντίστοιχα και οι πλήρως προσδιορισμένες πρωτεϊνικές δομές, αλλά ο συνολικός αριθμός των διαθέσιμων δομών είναι πολύ μικρότερος (Berman, et al., 2000).



Εικόνα 1.21: Αύξηση του κατατεθειμένων εγγραφών στη βάση πρωτεϊνικών ακολουθιών UniProtKB/Swiss-Prot (UniProt, 2013).

Ένα από τα κεντρικά δόγματα της βιοπληροφορικής είναι ότι πάντα υπάρχει ισχυρή συσχέτιση μεταξύ της αμινοξικής ακολουθίας (αλληλουχίας) μιας πρωτεΐνης με τη δομή και τη λειτουργία της. Με βάση αυτό, συχνά η λειτουργία μιας νέα πρωτεΐνης μπορεί να καθοριστεί από την ομοιότητα της αλληλουχίας της με μια πρωτεΐνη γνωστής λειτουργίας. Στην περίπτωση των GPCRs, παρά τη σημαντική ομοιότητα σε επίπεδο δομής, υπάρχει μικρή συντήρηση σε επίπεδο αμινοξικής ακολουθίας, όπως προαναφέρθηκε, υποδοχείς που προσδένουν τους ίδιους υποκαταστάτες μπορεί να αλληλεπιδρούν με διαφορετικές G-πρωτεΐνες, ενώ, αντίστοιχα, υποδοχείς που αλληλεπιδρούν με την ίδια G-πρωτεΐνη μπορεί να προσδένουν διαφορετικούς υποκαταστάτες (Yabuki, et al., 2005). Έχει, επίσης, αποδειχτεί ότι υποδοχείς που αλληλεπιδρούν με την ίδια G-πρωτεΐνη και τον ίδιο

υποκαταστάτη μπορεί να έχουν μικρότερη του 25% ομοιότητα σε επίπεδο αλληλουχίας. Οι ιδιαιτερότητες αυτές, καθώς και η σημασία των GPCRs ως φυσιολογικών παραγόντων και ως κύριων φαρμακευτικών στόχων, καθιστούν τη μελέτη τους, ήδη πάνω από μία δεκαετία, ένα πεδίο μεγάλου ενδιαφέροντος και για τη βιοπληροφορική.

Συχνά, έχουν γίνει διάφορες προσπάθειες συγκέντρωσης και οργάνωσης των διαθέσιμων πληροφοριών και δεδομένων που αφορούν τους GPCRs σε βάσεις δεδομένων. Καθεμία από τις διάφορες βάσεις δεδομένων που υπάρχουν στο διαδίκτυο έχει σχεδιαστεί ώστε να εξυπηρετεί διαφορετικό σκοπό. Έτσι, η GPCRDB περιλαμβάνει δεδομένα, για όλες τις κατηγορίες των GPCRs, τα οποία αφορούν την ακολουθία, τις σταθερές πρόσδεσης των υποκαταστών και τις θέσεις των παρατηρηθέντων μεταλλάξεων (Horn, et al., 2003). Η IUPHAR-DB διαθέτει ειδικό τομέα για τους GPCRs, περιλαμβάνοντας φαρμακολογικές, χημικές, γενετικές, λειτουργικές και ανατομικές πληροφορίες για “μη αισθητήριους υποδοχείς” στον άνθρωπο, τον αρουραίο και τον ποντικό, αποσκοπώντας στο να συγκεντρώσει δεδομένα για όλους τους υποδοχείς που μπορεί να αποτελέσουν στόχο φαρμάκων (Sharman, et al., 2012). Η SEVENS περιέχει δεδομένα για τα γονίδια των GPCRs, που έχουν αναγνωρισθεί σε γονιδιώματα 32 οργανισμών (Ono, et al., 2005; Suwa and Ono, 2009). Η GPCR-OKB παρέχει υπολογιστικά και πειραματικά δεδομένα σχετικά με τον ολιγομερισμό των GPCRs, τις αλληλεπιδράσεις μεταξύ των μονομερών τους και τις πιθανές εμπλοκές τους σε φυσιολογικές λειτουργίες και ασθένειες (Khelashvili, et al., 2010; Skrabanek, et al., 2007). Με τον ολιγομερισμό ασχολείται και η βάση GRIP παρέχοντας δεδομένα από υπολογιστικές προγνώσεις σχετικά με τις αλληλεπιδράσεις (Nemoto, et al., 2009). Η GPCRRD συγκεντρώνει στοιχεία από προσδιορισμένες δομές ολόκληρων υποδοχέων ή τμημάτων τους, καθώς και από πειράματα μεταλλαξογένεσης, δοκιμασίες μεταφοράς ενέργειας (energy transfer), αλλά, και, θεωρητικές προγνώσεις, με σκοπό την παροχή πληροφοριών που μπορούν να βοηθήσουν στον κρυσταλλογραφικό προσδιορισμό δομών υποδοχέων (Zhang and Zhang, 2010). Η gpDB είναι μια δημόσια διαθέσιμη σχεσιακή βάση δεδομένων που περιέχει πληροφορίες για τους GPCRs, τις G-πρωτεΐνες, τους εκτελεστές, και τις μεταξύ τους αλληλεπιδράσεις (Elefsinioti, et al., 2004; Theodoropoulou, et al., 2008), η οποία έχει αναπτυχθεί στο εργαστήριό μας.

Η έλλειψη ομοιότητας σε επίπεδο αλληλουχίας καθιστά τον εντοπισμό GPCRs από την αλληλουχία και μόνο δύσκολο, ειδικά στην περίπτωση νέων μελών της υπερικογένειας. Έχουν προταθεί κατά καιρούς διάφορες μέθοδοι για την επίτευξη αυτού του στόχου, οι

οποίες ποικίλουν από συστήματα που βασίζονται σε μοτίβα (motif-based) μέχρι και προσεγγίσεις που χρησιμοποιούν μηχανική μάθηση, καθώς και μια ποικιλία τεχνικών που δεν βασίζονται σε στοιχίσεις αλληλουχιών, αλλά στις φυσικοχημικές ιδιότητες των αμινοξικών ακολουθιών (Davies, et al., 2007). Το GPCRHMM είναι εργαλείο που χρησιμοποιεί profile HMMs για να εντοπίσει πιθανούς GPCRs σε πρωτεώματα (Wistrand, et al., 2006). Ο GRIP Server αποτελεί μέρος της βάσης δεδομένων GRIP και επιχειρεί να προβλέψει τις αλληλεπιδράσεις ανάμεσα σε υποδοχείς που δομούν ένα ολιγομερές GPCRs, χρησιμοποιώντας ως πρότυπα τις προσδιορισμένες δομές για το σχεδιασμό των μοντέλων των υποδοχέων (Nemoto, et al., 2009). Στο εργαστήριό μας, τέλος, έχουν αναπτυχθεί δύο εργαλεία: το PRED-GPCR που αφορά την κατάταξη των GPCRs σε οικογένειες (Papasaikas, et al., 2003; Papasaikas, et al., 2004), ενώ το PRED-COUPLE2 επιχειρεί την πρόβλεψη αλληλεπιδράσεων ανάμεσα σε GPCRs και οικογένειες των G-πρωτεϊνών (Sgourakis, et al., 2005; Sgourakis, et al., 2005).

1.5 Στόχοι της Διδακτορικής Διατριβής

Έχοντας ολοκληρώσει την ανασκόπηση που αφορά τη λειτουργία και τη δομή των συζευγμένων με G-πρωτεΐνες υποδοχέων (GPCRs) και των G-πρωτεϊνών, που αποτελούν σημαντικό «σύντροφό» τους για την πλειοψηφία των μονοπατιών μεταγωγής σήματος μέσω GPCRs, θα προσπαθήσουμε στη συνέχεια να οριοθετήσουμε τους στόχους της παρούσας διδακτορικής διατριβής.

Όπως ήδη αναφέραμε, ο αυξανόμενος αριθμός αλληλουχιών πρωτεϊνών που προκύπτει από τον προσδιορισμό ολοένα και περισσότερων γονιδιωμάτων, δημιουργεί την ανάγκη για αυτοματοποιημένους τρόπους προσδιορισμού της πιθανής λειτουργίας των πρωτεϊνών αυτών. Καθώς οι GPCRs αποτελούν σημαντική και πολυπληθή κατηγορία υποδοχέων στους ευκαρυωτικούς οργανισμούς, ο εντοπισμός τους στα νεοπροσδιοριζόμενα πρωτεώματα είναι κομβικής σημασίας. Έτσι, ο πρώτος στόχος που τέθηκε ήταν η ανάπτυξη υπολογιστικής μεθοδολογίας/αλγορίθμου που:

- ❖ **θα εντοπίζει και θα ταξινομεί σε οικογένειες πρωτεΐνες ως πιθανούς συζευγμένους με G-πρωτεΐνες υποδοχείς (GPCRs), από την αμινοξική τους ακολουθία και μόνο**

Για να κινηθούμε με αξιοπιστία σε αυτή την κατεύθυνση, επιλέξαμε να ασχοληθούμε με μια ειδική κατηγορία στοχαστικών μοντέλων, τα Μαρκοβιανά μοντέλα εξάρτησης (Hidden Markov Models, profile Hidden Markov Models) τα οποία έχουν δείξει αξιόλογη επιτυχία στο

παρελθόν σε αντίστοιχες προσπάθειες προτυποποίησης περιοχών των πρωτεϊνών, καθώς και να συνδυάσουμε υπάρχοντες αλγορίθμους με στόχο να εμπλουτίσουμε το αποτέλεσμα με επιπλέον σημαντικές πληροφορίες για τη λειτουργία των προβλεφθέντων GPCRs.

Επιπλέον, εξαιτίας του σημαντικού ρόλου των G-πρωτεϊνών στη λειτουργία των GPCRs, θεωρήθηκε σκόπιμο να τεθεί ως στόχος και η ανάπτυξη υπολογιστικής μεθοδολογίας/αλγορίθμου που:

- ❖ **θα εντοπίζει και θα ταξινομεί σε οικογένειες** πρωτεΐνες ως πιθανές **G-πρωτεΐνες**, από την αμινοξική τους ακολουθία και μόνο

Προς αυτή την κατεύθυνση χρησιμοποιήσαμε profile Hidden Markov Models, τα οποία άλλωστε είναι ειδικά σχεδιασμένα για να εντοπίζουν αν μία πρωτεΐνη ανήκει σε μία οικογένεια πρωτεϊνών, βασιζόμενα σε μία πολλαπλή στοίχιση αλληλουχιών από καλά χαρακτηρισμένα μέλη της.

Η πληθώρα διαθέσιμων αλληλουχιών, αλλά και βιβλιογραφίας που αφορά τους GPCRs και τις λειτουργίες τους, δημιουργεί την ανάγκη για συγκέντρωση και οργάνωση όλων αυτών των πληροφοριών, με σκοπό πάντα την καλύτερη κατανόηση του αντικειμένου που μελετάται, αλλά και την πιο εύχρηστη και αποτελεσματική πρόσβαση σε όλο αυτόν τον όγκο δεδομένων. Έτσι, ως επόμενος στόχος τέθηκε ο σχεδιασμός και η κατασκευή:

- ❖ **μιας βάσης δεδομένων για τους ανθρώπινους GPCRs, G-πρωτεΐνες, μόρια εκτελεστές και τις αλληλεπιδράσεις μεταξύ τους, και**
- ❖ **μιας βάσης δεδομένων για τους πεπτιδικούς υποκαταστάτες των GPCRs του ανθρώπου**, ως επέκταση της προαναφερθείσας βάσης δεδομένων.

Η μελέτη τόσο των δομικών και λειτουργικών χαρακτηριστικών όσο και του τρόπου αλληλεπίδρασης των υποδοχέων, των G-πρωτεϊνών, των εκτελεστών τους και των RGS πρωτεϊνών, μπορεί να αποτελέσει καθοριστικό ρόλο στην αποσαφήνιση των σηματοδοτικών μονοπατιών τους και των μηχανισμών μεταγωγής σήματος και ρύθμισης των λειτουργιών του οργανισμού γενικότερα. Με δεδομένα τα παραπάνω, τέθηκε ως στόχος:

- ❖ **η υπολογιστική μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με υποδοχείς, εκτελεστές και RGS πρωτεΐνες** και η σχέση τους με τα δομικά χαρακτηριστικά των υπομονάδων Gα στις διάφορες στερεοδιατάξεις τους.

Αντικείμενο μελέτης αποτέλεσαν τόσο τα υπάρχοντα κρυσταλλογραφικά δεδομένα όσο και οι ενδείξεις από βιοχημικές αλλά και υπολογιστικές μελέτες.

Η εμπλοκή των GPCRs σε ποικίλες ασθένειες και παθολογίες μας οδήγησε στα τελικά στάδια της διδακτορικής διατριβής να θέσουμε έναν επιπλέον στόχο που αφορά:

- ❖ **μελέτη και ανάλυση των μονονουκλεοτιδικών πολυμορφισμών (SNPs) στους ανθρώπινους συζευγμένους με G-πρωτεΐνες υποδοχείς (GPCRs) και τη συσχέτισή τους με ασθένειες.**

2. Υλικά και Μέθοδοι

2.1 Αλγόριθμοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών

2.1.1 Hidden Markov Models και profile HMMs

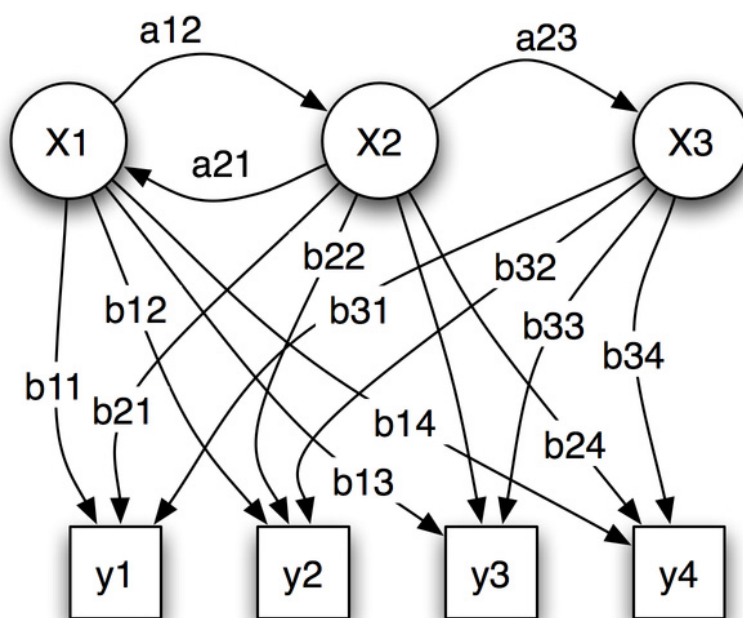
Οι πρωτεΐνες, τα διάφορα μόρια RNA και άλλα στοιχεία των πρωτεωμάτων και των γονιδιωμάτων μπορούν συνήθως να κατηγοριοποιηθούν σε οικογένειες συγγενικών αλληλουχιών και δομών (Henikoff, et al., 1997). Διαφορετικά κατάλοιπα, σε μια λειτουργική αλληλουχία, υπόκεινται σε διαφορετικές επιλεκτικές πιέσεις. Πολλαπλές στοιχίσεις μιας οικογένειας ακολουθιών αποκαλύπτουν αυτό το χαρακτηριστικό στο πρότυπο συντήρησής τους. Κάποιες θέσεις στην αλληλουχία είναι πιο συντηρημένες σε σχέση με άλλες, ενώ κάποιες περιοχές μιας πολλαπλής στοίχισης φαίνεται να «ανέχονται» παρεμβολές ή εξαλείψεις περισσότερο από άλλες.

Διαισθητικά, είναι επιθυμητό να χρησιμοποιηθούν πληροφορίες ειδικές για κάθε θέση για τις πολλαπλές στοιχίσεις, όταν γίνονται αναζητήσεις σε βάσεις δεδομένων για ομόλογες ακολουθίες. Οι μέθοδοι που χρησιμοποιούν profiles για την κατασκευή μοντέλων βαθμολόγησης που είναι ειδικά για κάθε θέση δημιουργήθηκαν για να επιτευχθεί ο σκοπός αυτός (Barton, 1990; Henikoff, 1996; Taylor, 1986). Τα τελευταία χρόνια αυτές οι μέθοδοι χρησιμοποιούνται όλο και συχνότερα, συμπληρώνοντας, βελτιώνοντας και ίσως μελλοντικά αντικαθιστώντας τις κλασσικές μεθόδους αναζήτησης σε βάσεις δεδομένων ακολουθιών που βασίζονται στη στοίχιση ζευγαριών ακολουθιών, όπως το BLAST (Altschul, et al., 1990; Altschul, et al., 1997) και το FASTA (Pearson and Lipman, 1988).

Τα Hidden Markov Models εισήχθησαν στο χώρο της Υπολογιστικής Βιολογίας στο τέλος της δεκαετίας τους '80 (Churchill, 1989) και η χρήση profile μοντέλων στα μέσα της δεκαετίας του '90 (Krogh, et al., 1994). Στην παρούσα διδακτορική διατριβή χρησιμοποιήσαμε τόσο HMM όσο και profile HMMs, για την υλοποίηση των αλγορίθμων εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών, και, γι' αυτό το λόγο, θα αναφερθούν οι βασικές αρχές της θεωρίας των HMMs και, στη συνέχεια, η ειδική περίπτωση των pHMMs.

2.1.1.1 Hidden Markov Models (HMMs)

Ένα HMM είναι ένα στατιστικό μοντέλο στο οποίο το σύστημα, το οποίο προτυποποιείται, θεωρείται μια αλυσίδα Markov με μη παρατηρούμενες καταστάσεις. Σε ένα μοντέλο Markov, η κατάσταση είναι απευθείας ορατή από τον παρατηρητή και γι' αυτό οι πιθανότητες μετάβασης μεταξύ καταστάσεων είναι οι μόνες παράμετροι. Σε ένα HMM, η κατάσταση δεν είναι απευθείας ορατή, αλλά το σύμβολο, που εξαρτάται από την κατάσταση, είναι ορατό. Κάθε κατάσταση έχει μια κατανομή πιθανοτήτων για τα πιθανά σύμβολα. Συνεπώς η ακολουθία συμβόλων που παράγεται από ένα HMM δίνει πληροφορίες για την ακολουθία καταστάσεων. Σε ένα HMM, αλυσίδα Markov συνιστά η ακολουθία καταστάσεων και όχι η ακολουθία συμβόλων (Eddy, 2004; Krogh, 1994).



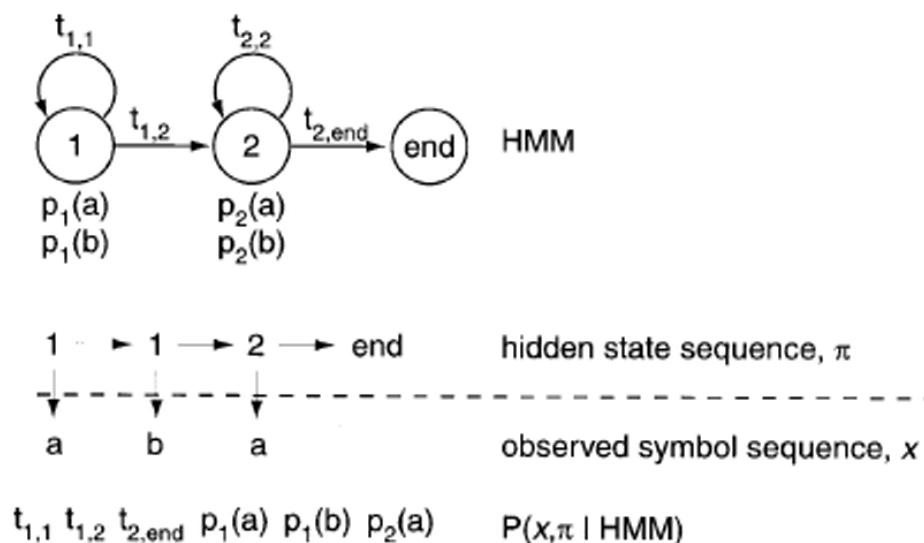
Εικόνα 2.1: Ευριστικές παράμετροι ενός HMM. Με x συμβολίζονται οι καταστάσεις, με y τα σύμβολα (πιθανές παρατηρήσεις), με a οι πιθανότητες μετάβασης μεταξύ καταστάσεων και με b οι πιθανότητες εκπομπής συγκεκριμένων συμβόλων, ως αποτέλεσμα συγκεκριμένων καταστάσεων.

Είναι σημαντικό να γίνει κατανοητή η σχετική απλότητα και η γενικότητα της θεωρίας των HMMs. Ένα HMM περιγράφει μια κατανομή πιθανοτήτων για έναν δυνητικά άπειρο αριθμό ακολουθιών. Επειδή το άθροισμα των πιθανοτήτων σε μια κατανομή πρέπει να είναι ίσο με τη μονάδα, οι «βαθμολογίες» που αναθέτει ένα HMM στις ακολουθίες είναι περιορισμένες. Η πιθανότητα μιας ακολουθίας δεν μπορεί να αυξηθεί χωρίς να μειωθεί η πιθανότητα μίας ή περισσότερων άλλων ακολουθιών. Αυτό ο θεμελιώδης περιορισμός της ευριστικής

προτυποποίησης (Jaynes, 1995) επιτρέπει στις παραμέτρους (Εικόνα 2.1) ενός HMM να έχουν υπολογίσιμες βέλτιστες τιμές. Η ερμηνεία μιας στοίχισης των παρατηρούμενων πρωτεϊνών ή DNA ακολουθιών ως μια ακολουθία κρυφών καταστάσεων είναι σαν να σημαίνεται η ακολουθία με σχετικές βιολογικές πληροφορίες (Eddy, 1998).

Ένα παράδειγμα για ένα απλό HMM, το οποίο προτυποποιεί ακολουθίες που αποτελούνται από 2 γράμματα (a, b), δείχνεται στην Εικόνα 2.2. Αυτό το HMM είναι ένα κατάλληλο μοντέλο για ένα πρόβλημα στο οποίο θεωρούμε ότι οι ακολουθίες ξεκίνησαν με μία σύνθεση καταλοίπων (ίσως πλούσια σε a), και στη συνέχεια στράφηκαν σε μια διαφορετική σύνθεση καταλοίπων (ίσως πλούσια σε b). Το μοντέλο αποτελείται από δύο καταστάσεις που συνδέονται με μεταβάσεις. Κάθε κατάσταση έχει μια κατανομή πιθανοτήτων εκπομπής για τη «γέννηση» ενός συμβόλου σε ένα αλφάβητο. Είναι σκόπιμο να βρεθεί ένα HMM, το οποίο να «γεννά» ακολουθίες. Ξεκινώντας από μια αρχική κατάσταση, επιλέγουμε μια νέα κατάσταση με κάποια πιθανότητα μετάβασης ($t_{1,1}$, $t_{1,2}$). Στη συνέχεια παράγουμε ένα κατάλοιπο με μια πιθανότητα εκπομπής ειδική για τη συγκεκριμένη κατάσταση (π.χ. $p_1(a)$). Επαναλαμβάνουμε τη διαδικασία μετάβασης/εκπομπής μέχρις ότου να φτάσουμε στην τελική κατάσταση. Στο τέλος αυτής της διαδικασίας, έχουμε μια κρυφή ακολουθία καταστάσεων και μια ακολουθία συμβόλων, την οποία και παρατηρούμε (Eddy, 1998).

Όταν ένα HMM σχεδιαστεί, άσχετα με την πολυπλοκότητά του, οι ίδιοι βασικοί προγραμματιστικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν για να στοιχηθούν και να βαθμονομηθούν οι ακολουθίες με βάση το μοντέλο (Durbin, et al., 1998). Οι παράμετροι ενός HMM μπορούν να οριστούν με δυο τρόπους. Ένα HMM μπορεί να εκπαιδευτεί από αρχικά μη στοιχισμένες ακολουθίες ή εναλλακτικά, το μοντέλο να κατασκευαστεί από προ-στοιχισμένες ακολουθίες (όπου η ακολουθίες καταστάσεων θεωρούνται γνωστές). Στη δεύτερη περίπτωση, η εκτίμηση των παραμέτρων είναι απλά ζήτημα της μετατροπής των παρατηρηθέντων μετρήσεων των εκπομπών των συμβόλων και των μεταβάσεων μεταξύ καταστάσεων σε πιθανότητες. Οι αλγόριθμοι εκπαίδευσης είναι σημαντικοί, γιατί μπορεί να μην είναι γνωστή μια αποδεκτή στοίχιση για τις ακολουθίες ενδιαφέροντος. Όλοι οι αλγόριθμοι εκπαίδευσης βελτιστοποιούν τοπικά και αυτός είναι ο λόγος που είναι σημαντικό τα HMMs να κατασκευάζονται από προ-στοιχισμένα δεδομένα, όταν αυτό είναι δυνατόν. Σε αντίθεση με την εκτίμηση των παραμέτρων, ο σχεδιασμός μιας κατάλληλης αρχιτεκτονικής για ένα HMM είναι συνήθως ζήτημα σχεδίασης με το χέρι (Eddy, 1998).



Εικόνα 2.2: Μοντέλο HMM, που προτυποποιεί ακολουθίες αποτελούμενες από a και b ως δύο περιοχές με πιθανώς διαφορετική σύνθεση καταλοίπων (Eddy, 1998). Οι κύκλοι αντιπροσωπεύουν τις καταστάσεις και τα βέλη συμβολίζουν τις μεταβάσεις μεταξύ καταστάσεων. Απεικονίζεται, επίσης, μια ακολουθία καταστάσεων, ακολουθούμενη από μια πιθανή ακολουθία συμβόλων. Η συνδυασμένη πιθανότητα της ακολουθίας συμβόλων και της ακολουθίας καταστάσεων είναι αποτέλεσμα όλων των πιθανοτήτων μετάβασης και εκπομπής. Η ίδια ακολουθία συμβόλων θα μπορούσε να προκύπτει από μια διαφορετική ακολουθία καταστάσεων, αν και πιθανότατα με μια τελείως διαφορετική πιθανότητα. Αυτή είναι η διάκριση μεταξύ των HMMs και ενός κανονικού Markov μοντέλου: στο HMM, η ακολουθία καταστάσεων δεν καθορίζεται μοναδικά από την παρατηρηθείσα ακολουθία συμβόλων, αλλά πρέπει να συνάγεται ευριστικά απ' αυτή.

Πλέον τα ευριστικά μοντέλα είναι ολοένα και αυξανόμενης σημασίας σε βιολογικές αναλύσεις, ειδικά σε περίπλοκα προβλήματα με πολλές παραμέτρους. Εξαιτίας του γεγονότος ότι πολλά προβλήματα στην υπολογιστική βιολογία περιορίζονται σε κάποιου είδους ανάλυση γραμμικών «ακολουθιών», τα ευριστικά μοντέλα που βασίζονται σε HMMs έχουν εφαρμοστεί σε πολλά προβλήματα.

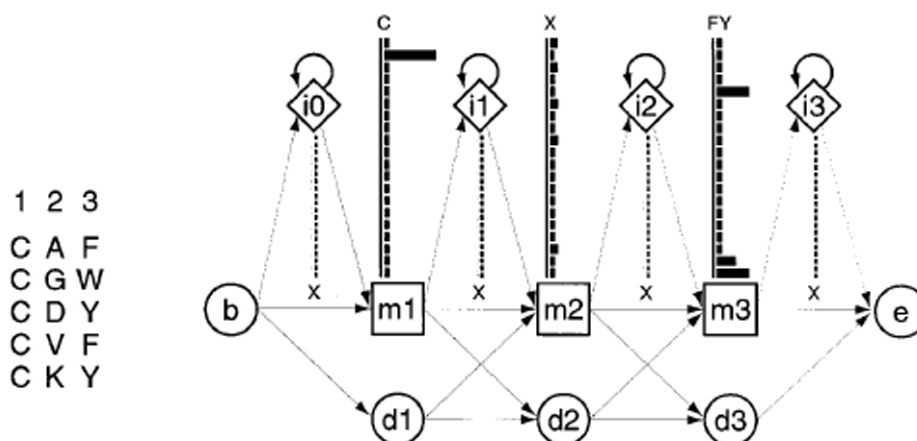
2.1.1.2 Profile Hidden Markov Models (pHMMs)

Τα profile HMMs είναι στατιστικά μοντέλα πολλαπλών στοιχίσεων ακολουθιών. Ενσωματώνουν, ειδική για κάθε θέση (position-specific), πληροφορία για το πόσο συντηρημένη είναι η κάθε στήλη της στοίχισης, καθώς και για το ποια αμινοξικά κατάλοιπα είναι πιθανά να εμφανιστούν. Ο Andrew Krogh και οι συνεργάτες του εισήγαγαν τα pHMMs στην Υπολογιστική Βιολογία (Krogh, et al., 1994), υιοθετώντας τεχνικές HMM που χρησιμοποιούνταν για χρόνια σε προγράμματα αναγνώρισης φωνής. Όπως ήδη αναφέραμε,

τα HMMs ήδη χρησιμοποιούνταν στη βιολογία πριν τη δουλειά του Krogh, ιδιαίτερα από τον Gary Churchill (Churchill, 1989), αλλά το άρθρο του Krogh είχε τεράστιο αντίκτυπο, γιατί η τεχνολογία των HMMs ήταν καλά προσαρμοσμένη στις δημοφιλείς μεθόδους με χρήση profiles για την αναζήτηση σε βάσεις δεδομένων, χρησιμοποιώντας πολλαπλές στοιχίσεις ακολουθιών αντί για μια μόνο ακολουθία ως επερώτηση (Eddy, 2006).

Τα profiles εισήχθησαν από τον Gribskov και συνεργάτες (Gribskov, et al., 1990; Gribskov, et al., 1987), ενώ πολλές άλλες ομάδες παρουσίαζαν παρόμοιες προσεγγίσεις την ίδια χρονική περίοδο (Barton, 1990; Bashford, et al., 1987; Taylor, 1986). Όλες οι μέθοδοι που χρησιμοποιούν profiles είναι στατιστικές περιγραφές μιας συναινετικής ακολουθίας που περιγράφει μια πολλαπλή στοίχιση. Χρησιμοποιούν, ειδική για κάθε θέση, βαθμονόμηση για τα αμινοξικά κατάλοιπα (ή τα νουκλεοτίδια) και, ειδικές για θέση, ποινές για την εκκίνηση και την επέκταση μιας εισαγωγής ή μιας διαγραφής. Η «παραδοσιακή» στοίχιση ζευγαριών ακολουθιών (BLAST (Altschul, et al., 1990), FASTA (Pearson and Lipman, 1988), αλγόριθμος Smith/Waterman (Smith and Waterman, 1981)) χρησιμοποιεί, ανεξάρτητες της θέσης, παραμέτρους βαθμονόμησης. Αυτή η ιδιότητα των profiles περιέχει σημαντικές πληροφορίες για το βαθμό συντήρησης σε διάφορες θέσεις της πολλαπλής στοίχισης, και τους διάφορους βαθμούς με τους οποίους τα κενά και οι εισαγωγές επιτρέπονται (Eddy, 2006).

Η Εικόνα 2.3 παρουσιάζει ένα παράδειγμα ενός rHMM, από το οποίο είναι δυνατόν να γίνει κατανοητή η αρχιτεκτονική του HMM, η οποία είναι ιδανική για να παρουσιάζει πολλαπλές στοιχίσεις ακολουθιών. Για κάθε συναινετική στήλη της πολλαπλής στοίχισης, μια κατάσταση «ταύτισης» προτυποποιεί την κατανομή των καταλοίπων που επιτρέπονται στη στήλη. Μια κατάσταση «εισαγωγής» και μια κατάσταση «διαγραφής» σε κάθε στήλη επιτρέπουν την εισαγωγή ενός ή περισσότερων καταλοίπων μεταξύ της συγκεκριμένης στήλης και της επόμενης, ή τη διαγραφή ενός καταλοίπου της συναινετικής ακολουθίας. Τα rHMMs είναι έντονα γραμμικά μοντέλα με κατεύθυνση από αριστερά προς τα δεξιά, σε αντίθεση με τη γενική περίπτωση των HMMs (Eddy, 1998).



Εικόνα 2.3: Ένα μικρό ρHMM που παρουσιάζει την πολλαπλή στοίχιση 5 ακολουθιών με τρεις συναινετικές στήλες (Eddy, 1998). Οι τρεις στήλες προτυποποιούνται από τρεις καταστάσεις «ταύτισης» (τετράγωνα, m1-m3), καθεμία από τις οποίες φέρει 20 πιθανότητες εκπομπής καταλοίπων, που απεικονίζονται με τις μαύρες ράβδους. Οι καταστάσεις «εισαγωγής» (ρόμβοι, i0-i3) φέρουν επίσης 20 πιθανότητες εκπομπής η καθεμία. Οι καταστάσεις «διαγραφής» (κύκλοι, d1-d3) είναι «σιωπηλές» καταστάσεις που δεν έχουν πιθανότητες εκπομπής. Μια κατάσταση έναρξης και μία λήξης περιλαμβάνονται, ενώ οι πιθανότητες μετάβασης μεταξύ των καταστάσεων συμβολίζονται με βέλη.

Με την εισαγωγή των διαφορετικών καταστάσεων ταύτισης και εισαγωγής, γίνεται μια σημαντική τομή με τις κλασικές μεθόδους στοίχισης, οι οποίες καθώς δεν προϋποθέτουν ένα μοντέλο δεν διαχωρίζουν τις πληροφοριακές θέσεις στη στοίχιση από τις απλές τυχαίες εισαγωγές. Επιπλέον, για πρώτη φορά οι ποινές για την εισαγωγή κενών (gap penalties), δεν τίθενται εκ των προτέρων αλλά εκτιμώνται από τα δεδομένα και αναπαρίστανται με καθαρά πιθανοθεωρητικό τρόπο, αποκλείοντας την υποκειμενική παρέμβαση. Έτσι, με τέτοια μοντέλα, είμαστε σε θέση να πραγματοποιήσουμε ιδιαίτερα ευαίσθητες αναζητήσεις και να εντοπίσουμε απομακρυσμένες ομολογίες (remote homologies), τις οποίες οι παραδοσιακοί αλγόριθμοι στοίχισης δεν θα μπορούσαν να εντοπίσουν. Η χρησιμότητα των ρHMM, ξεκινάει από δημιουργία πολλαπλών στοίχισεων, οι οποίες σε πολλές περιπτώσεις είναι εφάμιλλες αντίστοιχων δομικών πολλαπλών στοίχισεων (Eddy, 1995) και φτάνει στη δημιουργία μοντέλων μέσω των οποίων μπορεί να γίνει ευαίσθητη αναζήτηση απομακρυσμένων ομολογιών (Krogh, et al., 1994). Ιδιαίτερα με την τελευταία μέθοδο και τη δημιουργία όλο και περισσότερων μοντέλων για την ταξινόμηση πρωτεϊνικών οικογενειών, έχουν κατασκευαστεί ειδικές βάσεις δεδομένων όπως η PFAM (Punta, et al., 2012), μέσω της οποίας ταξινομείται μεγάλο μέρος των άγνωστης λειτουργίας πρωτεϊνών που προσδιορίζονται καθημερινά, π.χ. με την αποκωδικοποίηση των γονιδιωμάτων.

Οι εφαρμογές των ρHMMs είναι ποικίλες και πολύ χρήσιμες στη βιολογική έρευνα. Τα ρHMMs μπορούν να χρησιμοποιηθούν, όπως ήδη αναφέραμε, για να περιγράψουν μια πρωτεϊνική οικογένεια μέσω της δημιουργίας της συναινετικής ακολουθίας που την αντιπροσωπεύει και να εντοπίσουν απομακρυσμένες ομολογίες. Μπορούν επίσης να χρησιμοποιηθούν για τον αυτόματο σχολιασμό των δομικών αυτοτελών στοιχείων των πρωτεϊνών, μέσω των χαρακτηρισμένων πολλαπλών στοιχίσεων και των ρHMMs γνωστών αυτοτελών στοιχείων πρωτεϊνών που διατίθενται σε διάφορες βάσεις δεδομένων, όπως η Pfam. Τέλος, μια άλλη εφαρμογή είναι η αυτοματοποιημένη κατασκευή και συντήρηση μεγάλων βάσεων δεδομένων πολλαπλών στοιχίσεων.

2.1.1.3 Το πακέτο λογισμικού HMMER

Στην παρούσα μελέτη, όπου γίνεται χρήση profile Hidden Markov Models, η υλοποίηση των μοντέλων πραγματοποιείται με χρήση του γνωστού πακέτου λογισμικού HMMER (Eddy, 2011). Το πακέτο αυτό, είναι μια συλλογή προγραμμάτων, διανεμόμενο ελεύθερα με την άδεια 'ανοικτού κώδικα' GPL (GNU Public License), η οποία επιτρέπει ελεύθερη πρόσβαση στον πηγαίο κώδικα, και το οποίο έχει αποδειχθεί το καλύτερο ίσως πακέτο τέτοιου είδους με πολλές εφαρμογές σε μεγάλο εύρος βιολογικών δεδομένων. Το HMMER, στην τελευταία έκδοση του (HMMERv3) περιέχει τα παρακάτω προγράμματα:

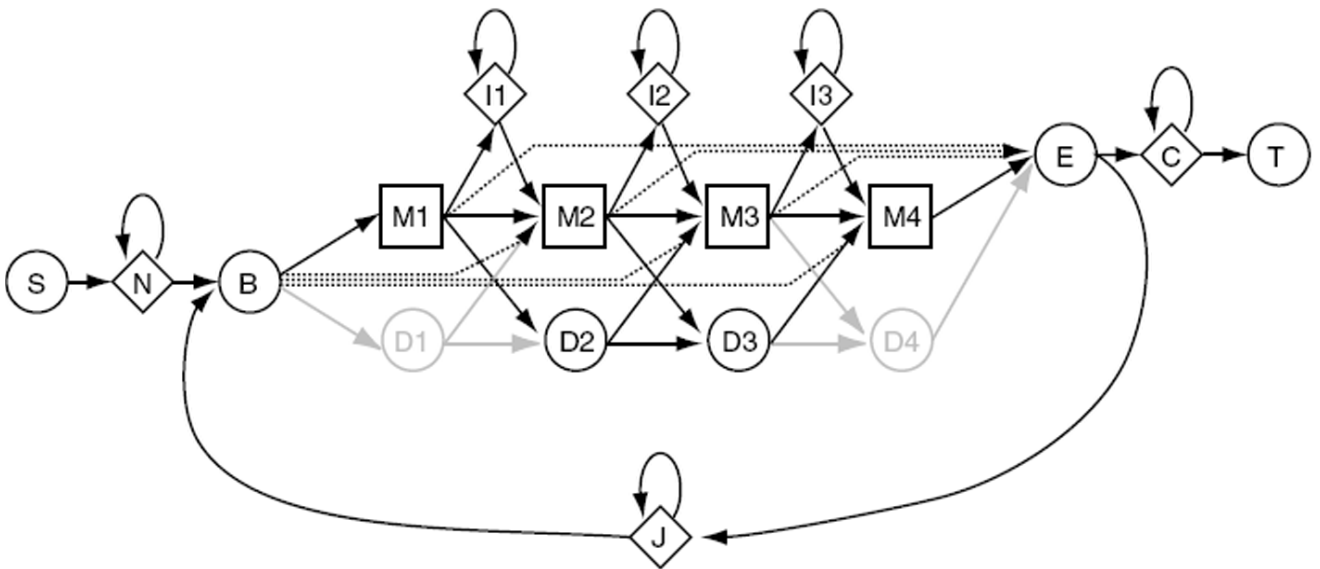
- *hmmsearch*: Πρόγραμμα το οποίο, πραγματοποιεί αναζητήσεις ενός μοντέλου HMM έναντι μιας βάσης ακολουθιών πρωτεϊνών.
- *hmmScan*: Πρόγραμμα με το οποίο πραγματοποιούνται αναζητήσεις μιας ή περισσότερων ακολουθιών έναντι μιας βάσης δεδομένων από μοντέλα HMM. Πρέπει να τονιστεί εδώ, ότι αν έχουμε μία ακολουθία και ένα HMM, τα δυο παραπάνω προγράμματα επιστρέφουν ακριβώς το ίδιο αποτέλεσμα. Αν διαφέρουν, είτε οι ακολουθίες είτε τα μοντέλα, τότε δίνουν άλλο αποτέλεσμα, λόγω του διαφορετικού τρόπου υπολογισμού της στατιστικής σημαντικότητας.
- *hmmemit*: Πρόγραμμα, με το οποίο 'εκπέμπεται' η καλύτερη (ανάλογα με τον ορισμό) ακολουθία η οποία θα μπορούσε να παραχθεί από το μοντέλο.
- *hmmBuild*: Πρόγραμμα με χρήση του οποίου, ξεκινώντας από μία αρχική πολλαπλή στοιχίση, κατασκευάζεται ένα μοντέλο HMM το οποίο να την περιγράφει.

- *hmmalign*: Πρόγραμμα με το οποίο μια σειρά ακολουθιών οι οποίες προέρχονται από ένα HMM, στοιχίζονται σε μια πολλαπλή στοίχιση. Η πολλαπλή στοίχιση, επιτυγχάνεται μέσω διαδοχικών στοιχίσεων των ακολουθιών με το μοντέλο.
- *hmmpress*: Πρόγραμμα το οποίο χρησιμοποιεί μία βάση από pHMMs (π.χ.) σε κανονική μορφοποίηση (HMMER3 format) και τη μετατρέπει σε συμπιεσμένα δυαδικά δεδομένα για το *hmmsearch* με σκοπό τη γρηγορότερη αναζήτηση.

Επιπλέον, υπάρχουν και μια σειρά από βοηθητικά προγράμματα, για μετατροπή αρχείων σε διάφορες μορφές, ευρετηριοποίηση (indexing) των βάσεων δεδομένων για γρηγορότερη αναζήτηση, ρουτίνες για πολύ-επεξεργασία κλπ. Τα παραπάνω προγράμματα, περιέχουν μια σειρά από βελτιστοποιήσεις με σκοπό την επιτάχυνση των διαδικασιών. Για παράδειγμα, υπάρχουν βελτιστοποιήσεις για την ταχύτητα κατά την εκτέλεση των αλγορίθμων με το μη υπολογισμό των προηγούμενων καταστάσεων, βελτιστοποιήσεις στον υπολογισμό των κατανομών συχνοτήτων των αμινοξικών καταλοίπων του μηδενικού (null) μοντέλου με την εισαγωγή μίξεων από εκ των προτέρων κατανομές, διαφορικό στάθμισμα των ακολουθιών με διαφορετικό βαθμό ομοιότητας, και μια σειρά από βελτιστοποιήσεις στη δομή του μοντέλου. Το τελευταίο, είναι πολύ σημαντικό, καθώς με αυτή τη διαδικασία, ο χρήστης δεν ασχολείται καθόλου με τη δομή και το μέγεθος του μοντέλου.

Το HMMER, χρησιμοποιεί ένα γενικό πρότυπο (template), για όλα τα μοντέλα που κτίζονται με αυτό, και το οποίο ονομάζεται Plan 7 architecture, η οποία φαίνεται παρακάτω, στην Εικόνα 2.4.

Με επαναλήψεις αυτού του βασικού μοτίβου, επιτυγχάνεται, αναλόγως και του μήκους της πολλαπλής στοίχισης, η τελική στερεοδιάταξη του μοντέλου. Το παραπάνω μοντέλο, διαφέρει από την τυπική έκδοση του profile Hidden Markov Model, που είδαμε παραπάνω, στο ότι δεν επιτρέπει μεταβάσεις από μια κατάσταση εισαγωγής κενού (I) σε κατάσταση απαλοιφής (D) και το αντίστροφο. Το πρόγραμμα, είναι διαθέσιμο στην ηλεκτρονική διεύθυνση: <http://hmmer.janelia.org/>.



Εικόνα 2.4: Η αρχιτεκτονική του μοντέλου Plan 7.

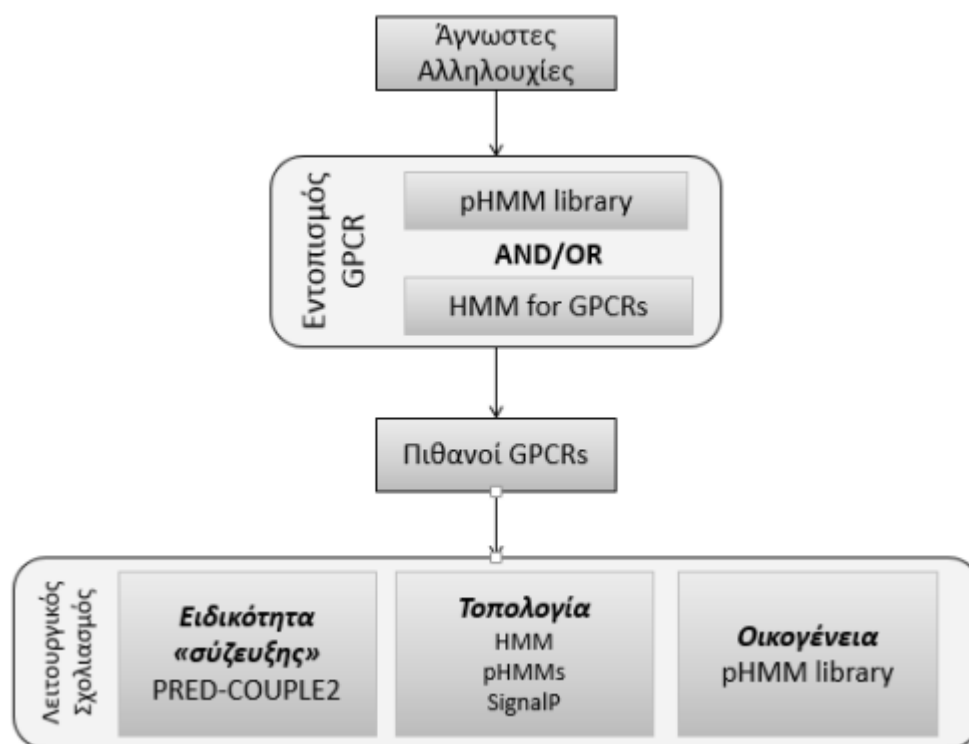
2.1.2 Η μέθοδος GPCRripe

Οι GPCRs αποτελούν, όπως έχει ήδη αναφερθεί, τη μεγαλύτερη υπεροικογένεια διαμεμβρανικών υποδοχέων στους ευκαρυωτικούς οργανισμούς, συνεπώς, ο εντοπισμός πιθανών GPCRs σε νεοπροσδιορισμένα πρωτεώματα είναι μείζονος σημασίας. Για τον εντοπισμό και το λεπτομερές σχολιασμό των GPCRs σχεδιάστηκε και κατασκευάστηκε η μέθοδος GPCRripe. Η μέθοδος αποτελείται από δύο επίπεδα (Εικόνα 2.5). Το πρώτο επίπεδο αφορά τον εντοπισμό των υποδοχέων και αποτελεί το κύριο χαρακτηριστικό της μεθόδου. Ο εντοπισμός των GPCRs στηρίζεται σε: (α) ένα HMM ειδικά σχεδιασμένο για αυτόν το σκοπό (Εικόνα 2.6), το οποίο είναι μία τροποποιημένη έκδοση του μοντέλου που χρησιμοποιεί ο αλγόριθμος HMM-TM (Bagos, et al., 2006), και, (β) μία βιβλιοθήκη 39 ρHMMs της Pfam (Πίνακας 2.1) (Punta, et al., 2012), τα οποία είναι ειδικά για τις διάφορες οικογένειες των GPCRs.

Το δεύτερο επίπεδο παρέχει λεπτομερή σχολιασμό του πιθανού υποδοχέα με τη χρήση προϋπαρχόντων εργαλείων. Πιο συγκεκριμένα, οι πληροφορίες που παρέχονται αφορούν:

- την ειδικότητα “σύζευξης” του υποδοχέα με συγκεκριμένες οικογένειες G-πρωτεϊνών με χρήση του αλγορίθμου PRED-COUPLE2 (Sgourakis, et al., 2005),

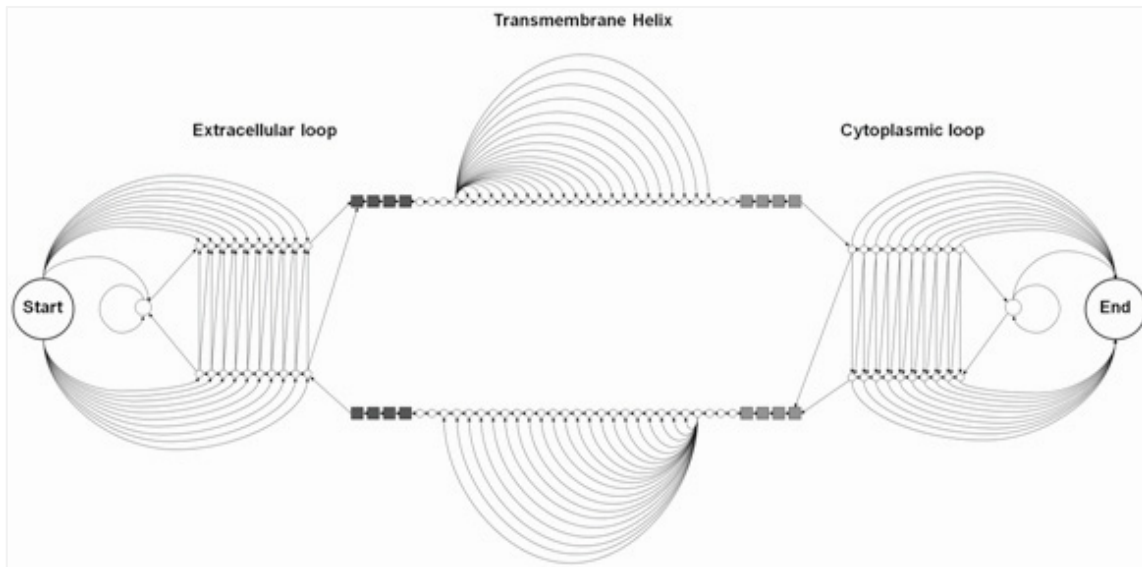
- την τοπολογία μέσω του HMM, της εμφάνισης συγκεκριμένων ρHMMs της Pfam που εμφανίζονται ειδικά ενδοκυτταρικά ή εξωκυτταρικά, και της θέσης του πεπτιδίου οδηγητή μέσω του αλγορίθμου SignalP 4.0 (Petersen, et al., 2011) και
- την οικογένεια στην οποία ανήκει ο υπό μελέτη υποδοχέας με βάση το χαρακτηριστικό ρHMM της Pfam από τη βιβλιοθήκη των ειδικών για GPCRs profiles.



Εικόνα 2.5: Διάγραμμα ροής της μεθόδου GPCRipe. Διακρίνονται τα δύο επίπεδα της μεθόδου.

2.1.2.1 Η αρχιτεκτονική του μοντέλου και η διαδικασία εκπαίδευσης

Όπως αναφέρθηκε ήδη, το HMM που χρησιμοποιείται στο επίπεδο του εντοπισμού είναι παρεμφερές με αυτό του αλγορίθμου HMM-TM (Bagos, et al., 2006). Το μοντέλο είναι κυκλικό και έχει τροποποιηθεί ειδικά για να αναγνωρίζει GPCRs, καθώς το αμινοτελικό (κατάσταση έναρξης) και το καρβοξυτελικό (κατάσταση τερματισμού) άκρο συνδέονται αποκλειστικά με εξωκυττάρους και κυττατοπλασματικούς βρόχους αντίστοιχα (Εικόνα 2.6). Το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας τον αλγόριθμο Baum-Welch για σημασμένες αλληλουχίες (Krogh, 1994) και η αποκωδικοποίηση πραγματοποιήθηκε με χρήση του Optimal Accuracy Posterior Decoder (Kall, et al., 2005).



Εικόνα 2.6. Η αρχιτεκτονική του τροποποιημένου HMM για την πρόγνωση της τοπολογίας των GPCRs. Το μοντέλο αποτελείται από 3 υπομοντέλα που αντιστοιχούν στις διαμεμβανικές α -έλικες, τους εξωκυττάριους και ενδοκυττάριους βρόχους αντίστοιχα. Διακρίνονται οι καταστάσεις έναρξης (Start) και τερματισμού (End) οι οποίες συνδέονται αποκλειστικά με τους εξωκυττάριους και κυτταροπλασματικούς βρόχους αντίστοιχα. Η τροποποίηση αυτή προτυποποιεί την κοινή τοπολογία των GPCRs με το αμινοτελικό άκρο να είναι πάντα εξωκυττάριο.

2.1.2.2 Η βιβλιοθήκη των ειδικών για GPCRs pHMMs της Pfam

Έπειτα από εκτενή έρευνα στη βάση Pfam, κατασκευάστηκε μία βιβλιοθήκη με 39 pHMMs ειδικά για GPCRs (Πίνακας 2.1). Πιο συγκεκριμένα, η βιβλιοθήκη αυτή αποτελείται από 32 pHMMs της υπερομάδας GPCR_A (συνολικά 35 pHMMs) της Pfam, έχοντας αποκλείσει τα υπόλοιπα τρία, καθώς τα δύο από αυτά αφορούν βακτηριακούς υποδοχείς με 7 διαμεμβρανικά (7TM-7TMR_HD και Bac_rhodopsin), οι οποίοι δεν αποτελούν μέλη της υπεροικογένειας των GPCRs, ενώ το Git3_C pHMM περιγράφει το καρβοξυτελικό domain ενός GPCR. Τα 30 από τα 32 pHMMs αφορούν υποδοχείς της Κατηγορίας A, ενώ το 7tm_2 περιγράφει υποδοχείς της Κατηγορίας B και το Ocular_alb ένα πολύ μικρό κομμάτι υποδοχέων της Κατηγορίας C. Επιπλέον στη βιβλιοθήκη ενσωματώθηκαν τρία pHMMs που δεν ανήκουν στην υπερομάδα GPCR_A της Pfam, τα 7tm3 (αντιπροσωπεύει την Κατηγορία C), STE2 και STE3 (υποδοχείς φερομονών μυκήτων, Κατηγορία D). Τέλος, συμπεριλήφθηκαν τέσσερα επιπλέον pHMMs που ανήκουν στην υπερομάδα Chemosens_resp. Αυτά αφορούν χημειοαισθητικούς υποδοχείς εντόμων (γεύσης και όσφρησης), και, παρά τη διαμάχη που υπάρχει για το κατά πόσο μπορούν να θεωρηθούν GPCRs ή όχι, συμπεριλήφθηκαν, μιας και

αποτελούν σημαντικό ποσοστό των πρωτεωμάτων των εντόμων και θα μπορούσαν να αποτελέσουν στόχους εντομοκτόνων.

Πίνακας 2.1 Τα 35 ρHMMs της Pfam (Punta, et al., 2012) που χρησιμοποιούνται στο GPCRpipe και η ταξινόμησή τους με βάση το σύστημα του Kolakowski (Kolakowski, 1994).

Ομάδα	Όνομα ρHMM	Κωδικός της Pfam
Class A	7tm_1	PF00001
	7tm_4	PF13853
	Lung_7-TM_R	PF06814
	GpcrRhopsn4	PF10192
	TAS2R	PF05296
	V1R	PF03402
	Git3	PF11710
Class A Nematode specific chemo-receptors	7TM_GPCR_Srab	PF10292
	7TM_GPCR_Sra	PF02117
	7TM_GPCR_Srbc	PF10316
	7TM_GPCR_Srb	PF02175
	7TM_GPCR_Srd	PF10317
	7TM_GPCR_Srh	PF10318
	7TM_GPCR_Sri	PF10327
	7TM_GPCR_Srj	PF10319
	7TM_GPCR_Srsx	PF10320
	7TM_GPCR_Srt	PF10321
	7TM_GPCR_Sru	PF10322
	7TM_GPCR_Srv	PF10323
	7TM_GPCR_Srw	PF10324
	7TM_GPCR_Srx	PF10328
	7TM_GPCR_Srz	PF10325
	7TM_GPCR_Str	PF10326
Serpentine_r_xa	PF03383	
Sre	PF03125	
Srg	PF02118	

	DUF1182	PF06681
	DUF621	PF04789
Class B	7tm_2	PF00002
Class C	7tm_3	PF00003
	Ocular_alb	PF02101
Class D	STE2	PF02116
	STE3	PF02076
Class E	Dicty_CAR	PF05462
Class F	Frizzled	PF01534
Insect	7tm_6	PF02949
Chemosensory	7tm_7	PF08395
receptors	DUF267	PF03268
	Trehalose_recip	PF06151

2.1.2.3 Σύνολο εκπαίδευσης

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε ένα σύνολο 432 μη ομόλογων GPCRs. Χρησιμοποιήθηκαν τα βασικά σύνολα (seed sets) των 35 pHMMs της βάσης Pfam, στα οποία έγινε μείωση ομολογίας σε επίπεδο 50% με τη χρήση του BLAST(Altschul, et al., 1990), και έτσι προέκυψε το σύνολο εκπαίδευσης. Δεν χρησιμοποιήθηκαν κατά την εκπαίδευση ακολουθίες από χημειοαισθητήριους υποδοχείς εντόμων λόγω της ασάφειας σχετικά με την τοπολογία τους. Η τοπολογία των υποδοχέων προσδιορίστηκε χρησιμοποιώντας πληροφορίες από τη βάση UniProt (UniProt, 2013) και τη διεθνή βιβλιογραφία.

2.1.3 Η μέθοδος GprotPRED

Οι G-πρωτεΐνες αποτελούν τους μοριακούς διακόπτες της μεταγωγής σήματος. Σε αντίθεση με την τεράστια ποικιλομορφία των GPCRs σε επίπεδο ακολουθίας και το μεγάλο αριθμό τους, οι G-πρωτεΐνες είναι πολύ συντηρημένες μεταξύ τους και βρίσκονται στα πρωτεώματα σε πολύ μικρούς αριθμούς. Για τον εντοπισμό και την κατάταξη των G-πρωτεϊνών σε οικογένειες σχεδιάστηκε και υλοποιήθηκε η μέθοδος GprotPRED. Η μέθοδος βασίζεται σε pHMMs καθένα από τα οποία είναι ειδικά κατασκευασμένο για καθεμία από τις τέσσερις

οικογένειες Gα των θηλαστικών, ενώ, επιπλέον, κατασκευάστηκαν και άλλα 2 pHMMs, ένα για την Gβ υπομονάδα και ένα για την Gγ υπομονάδα. Στόχος της μεθόδου είναι να εντοπίζει G-πρωτεΐνες σε νεοπροσδιοριζόμενα πρωτεώματα και να τις κατατάσσει στις διαφορετικές οικογένειες.

2.1.3.1 Κατασκευή των pHMMs

Αρχικά, όλες οι αλληλουχίες των Gα, Gβ και Gγ υπομονάδων ανασύρθηκαν από την UniProt/Swiss-Prot (UniProt Release 2010_09) (UniProt, 2010). Συλλέχθηκαν 190 Gα υπομονάδες, εκ των οποίων οι 112 κατατάσσονταν σε μία από τις τέσσερις γνωστές οικογένειες ετεροτριμερών G-πρωτεϊνών (23 G_s, 55 G_{i/o}, 27 G_{q/11}, 7 G_{12/13}), ενώ οι υπόλοιπες 78 δεν ήταν δυνατόν να καταταχθούν σε κάποια από τις οικογένειες. Επιπλέον, ανασύρθηκαν 77 Gβ και 57 Gγ υπομονάδες. Αλληλουχίες από τις 6 κατηγορίες χρησιμοποιήθηκαν για την κατασκευή pHMMs. Πιο συγκεκριμένα, οι ακολουθίες που επιλέχθηκαν ως βασικό σύνολο (seed set) για καθένα από τα υπό κατασκευή pHMMs στοιχήθηκαν με τη βοήθεια του ClustalW2 (Larkin, et al., 2007). Οι στοιχίσεις χρησιμοποιήθηκαν ως είσοδοι στο πρόγραμμα hmmbuild του πακέτου HMMER v2.3.2 (Eddy, 2006). Εξαιτίας της μεγάλης συγγένειας μεταξύ των τεσσάρων οικογενειών Gα, χρησιμοποιήθηκε η τροποποίηση HMM-ModE (Srivastava, et al., 2007), η οποία επιτρέπει τη δημιουργία pHMMs με τη χρήση τόσο θετικού όσο και αρνητικού συνόλου εκπαίδευσης, με σκοπό τη δημιουργία πιο εξειδικευμένων μοντέλων. Λόγω της μη συμβατότητας του HMM-ModE με την καινούρια έκδοση του πακέτου HMMER v3, τα pHMMs κατασκευάστηκαν με την προηγούμενη έκδοση, όπως αναφέραμε, όμως στη συνέχεια μετατράπηκαν σε μορφή συμβατή με το HMMER v3, ώστε οι αναζητήσεις (hmmsearch, hmmscan) να εκμεταλλεύονται τις δυνατότητες της νέας έκδοσης του HMMER. Αξίζει να σημειωθεί ότι έγινε προσπάθεια δημιουργίας όλων των pHMMs με χρήση αποκλειστικά του HMMER v3, όμως φαίνεται ότι η χρήση και αρνητικού συνόλου εκπαίδευσης είναι σημαντική. Αναλυτικά για την κατασκευή κάθε pHMM (Όλα τα σύνολα δίνονται στον Πίνακα 1.3 του Παραρτήματος και στον ιστότοπο <http://aias.biol.uoa.gr/~mtheo/datasets/>):

- ❖ **pHMM για την G_s οικογένεια:** 23 αλληλουχίες χρησιμοποιήθηκαν ως θετικό σύνολο εκπαίδευσης, ενώ 89 αλληλουχίες που ανήκουν στις υπόλοιπες 3 οικογένειες Gα (G_{i/o}, G_{q/11}, G_{12/13}) ως αρνητικό.

- ❖ **pHMM για την G_{i/o} οικογένεια:** 41 αλληλουχίες ως θετικό σύνολο εκπαίδευσης, ενώ 57 αλληλουχίες από τις υπόλοιπες 3 οικογένειες G_α ως αρνητικό.
- ❖ **pHMM για την G_{q/11} οικογένεια:** 27 αλληλουχίες αποτέλεσαν το θετικό σύνολο εκπαίδευσης, ενώ 85 αλληλουχίες από τις υπόλοιπες 3 οικογένειες G_α το αρνητικό.
- ❖ **pHMM για την G_{12/13} οικογένεια:** Το μοντέλο κατασκευάστηκε από την πολλαπλή στοίχιση 7 αλληλουχιών που ανήκουν στην οικογένεια αυτή, ενώ 105 αλληλουχίες από τις υπόλοιπες οικογένειες αποτέλεσαν το αρνητικό σύνολο.
- ❖ **pHMM για την G β υπομονάδα:** Το συγκεκριμένο μοντέλο κατασκευάστηκε με στόχο να αντιπροσωπεύει ολόκληρη τη β υπομονάδα, σε αντίθεση με το διαθέσιμο στη βάση Pfam μοντέλο (WD40, PF00400) το οποίο αντιπροσωπεύει μόνο την WD40 δομική αυτοτελή περιοχή. Ως θετικό σύνολο εκπαίδευσης χρησιμοποιήθηκαν 50 αλληλουχίες, στις οποίες περιλαμβάνονταν ένας τουλάχιστον αντιπρόσωπος από κάθε οργανισμό και από κάθε διαφορετικό τύπο G β υπομονάδων. Στο αρνητικό σύνολο περιλαμβάνονται 89 αλληλουχίες, οι οποίες προέκυψαν από την ακόλουθη διαδικασία:
 1. Χρησιμοποιήθηκε το WD40 μοντέλο για να πραγματοποιηθεί αναζήτηση έναντι της UniProt/Swiss-Prot και βρέθηκαν 2195 αλληλουχίες
 2. Στη συνέχεια απομονώθηκαν οι αλληλουχίες που περιλάμβαναν το WD40 μοτίβο 7 φορές (τόσες επαναλήψεις περιλαμβάνουν οι β υπομονάδες). Συγκεντρώθηκαν 582 αλληλουχίες.
 3. Από αυτές εξαιρέθηκαν οι αλληλουχίες των β υπομονάδων και απέμειναν 501 αλληλουχίες.
 4. Τέλος, πραγματοποιήθηκε μείωση ομολογίας με χρήση του προγράμματος CD-HIT (Huang, et al., 2010), οδηγώντας σε ένα μη ομόλογο σύνολο 89 αμινοξικών ακολουθιών πρωτεϊνών που έχουν 7 επαναλήψεις του WD40 μοτίβου, αλλά δεν είναι G β υπομονάδες.
- ❖ **pHMM για την G γ υπομονάδα:** Αντίστοιχη διαδικασία ακολουθήθηκε και για την G γ υπομονάδα. Το υπάρχον στην Pfam μοντέλο (G-gamma, PF00631) περιγράφει το GGL domain το οποίο εκτός από τις G γ υπομονάδες, εμφανίζεται και σε RGS πρωτεΐνες. Για την κατασκευή του ειδικού για την G γ υπομονάδα pHMM, χρησιμοποιήθηκαν 24 αλληλουχίες ως θετικό σύνολο εκπαίδευσης, στις οποίες περιλαμβάνονταν ένας τουλάχιστον αντιπρόσωπος από κάθε οργανισμό και από

κάθε διαφορετικό τύπο. Το αρνητικό σύνολο εκπαίδευσης προέκυψε από την αναζήτηση με το μοντέλο G-gamma της Pfam έναντι της UniProt/Swiss-Prot (71 αλληλουχίες) αφού εξαιρέθηκαν όλες οι Gγ υπομονάδες. Τελικά, το αρνητικό σύνολο περιλαμβάνει 14 αμινοξικές ακολουθίες.

2.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών

2.2.1 Η βάση δεδομένων Human-grDB

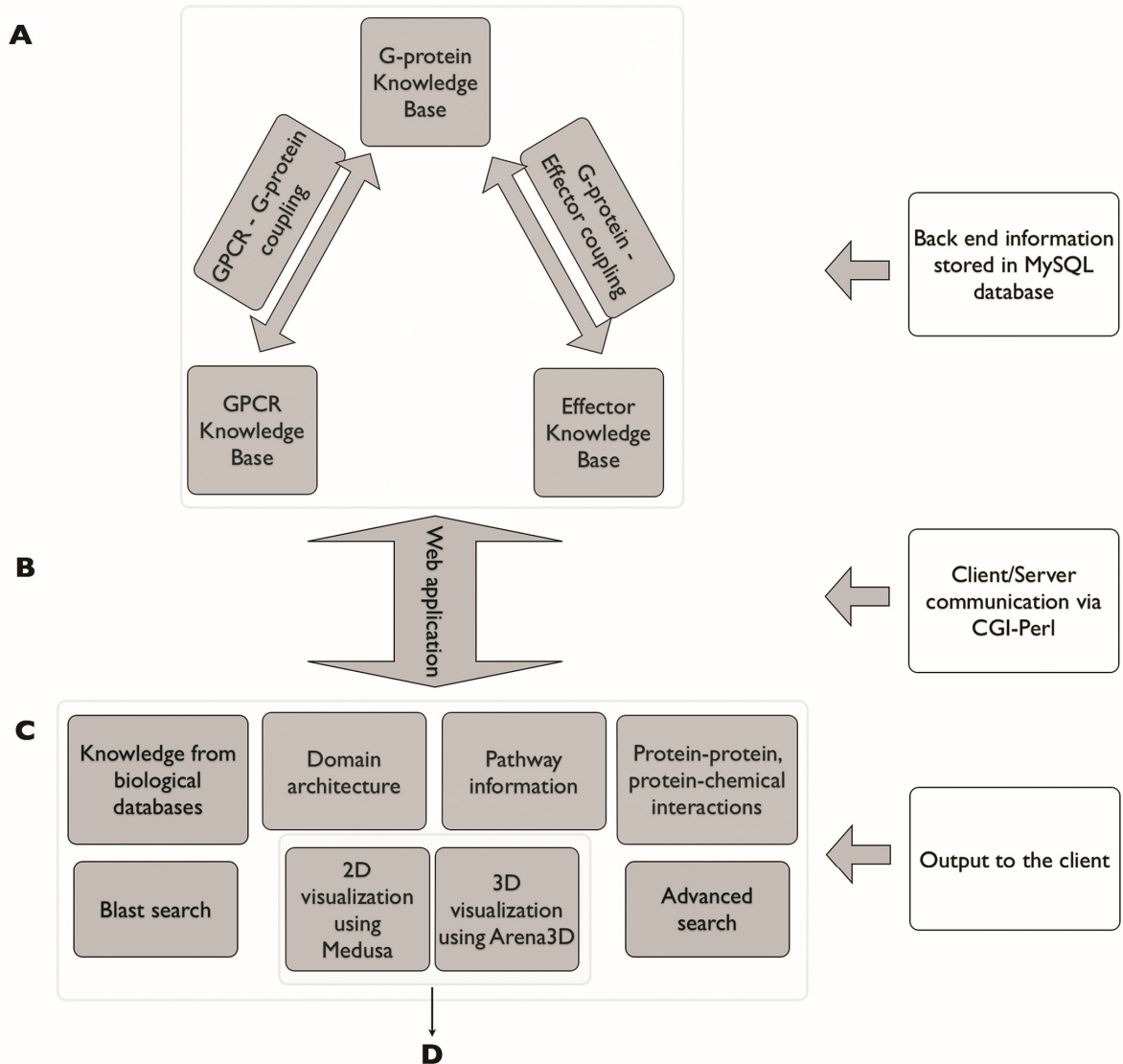
Οι GPCRs, οι G-πρωτεΐνες, οι εκτελεστές και οι αλληλεπιδράσεις μεταξύ τους αποτελούν κύριο μηχανισμό για τη μεταγωγή σήματος και την ενεργοποίηση και απενεργοποίηση μονοπατιών μέσα στο κύτταρο. Η βάση δεδομένων Human-grDB σχεδιάστηκε και κατασκευάστηκε για να συμπεριλάβει το σύνολο των πληροφοριών που αφορούν GPCRs, G-πρωτεΐνες, εκτελεστές και τις αλληλεπιδράσεις μεταξύ τους στον άνθρωπο. Η Human-grDB δεν αποτελεί απλά ένα υποσύνολο της υπάρχουσας βάσης δεδομένων του εργαστηρίου μας grDB (Elefsinioti, et al., 2004; Theodoropoulou, et al., 2008), καθώς όχι μόνο παρουσιάζει πως τα μόρια αυτά αλληλεπιδρούν μεταξύ τους, αλλά αποκαλύπτει πληροφορίες και για τα μονοπάτια στα οποία αυτά εμπλέκονται (Satagoram, et al., 2010). Επιπλέον, η Human-grDB κατασκευάστηκε με στόχο να γίνει ένα χρήσιμο εργαλείο στη φαρμακευτική έρευνα και μία πλατφόρμα που θα αποκαλύπτει νέα πρότυπα για θεραπείες.

2.2.1.1 Συλλογή δεδομένων και κατασκευή της βάσης

Το πρώτο στάδιο ήταν η ανάκτηση των αλληλουχιών ξεχωριστά για τους ανθρώπινους GPCRs, G-πρωτεΐνες και εκτελεστές από τη βάση UniProt/Swiss-Prot (UniProt Release 2012_12) (UniProt, 2012).

Τα δεδομένα αυτά συλλέχθηκαν με τη χρήση κατάλληλων προγραμμάτων γραμμένων σε γλώσσα Perl, τα οποία εντόπιζαν τα πεδία DE (description), GN (gene) ή DR (database cross reference) για κάθε μία εγγραφή. Στη συνέχεια, αποκλείστηκαν από τα δεδομένα οι διπλοεγγραφές. Για τη συλλογή πληροφοριών που αφορούν τις αλληλεπιδράσεις μεταξύ των εν λόγω μορίων, διεξήχθη εκτενής βιβλιογραφική έρευνα. Με αυτό τον τρόπο, δημιουργήθηκε ένας κατάλογος με συνδέσεις μεταξύ των μορίων, ο οποίος συνοδεύεται από τις αντίστοιχες βιβλιογραφικές αναφορές στην PubMed. Τα δεδομένα και για τους GPCRs και τις G-πρωτεΐνες ομαδοποιήθηκαν με βάση τα καθιερωμένα συστήματα

ταξινόμησης, ενώ για τους εκτελεστές η κατηγοριοποίηση έγινε με βάση τη λειτουργία τους, όπως αυτή προτείνεται στην ανανεωμένη έκδοση της βάσης grDB (Theodoropoulou, et al., 2008).



Εικόνα 2.7: Επισκόπηση της διαδικτυακής εφαρμογής της Human-grDB. **A.** Το παρασκήνιο της εφαρμογής αποτελείται από πληροφορίες που συλλέχθηκαν με το χέρι για τις τρεις κατηγορίες μορίων και τις αλληλεπιδράσεις τους, καθώς και πληθώρα δημόσια διαθέσιμων πληροφοριών για κάθε πρωτεΐνη. Όλα τα δεδομένα είναι αποθηκευμένα σε μία βάση δεδομένων MySQL. **B.** Ένα CGI-Perl πρόγραμμα διαχειρίζεται την επικοινωνία μεταξύ διακομιστή και πελάτη και **C.** παρέχει πληθώρα πληροφοριών ως έξοδο στο χρήστη. Αυτό το σύστημα μπορεί να συνδυαστεί με **D.** ένα εργαλείο οπτικοποίησης στις 2 διαστάσεις (Medusa, (Hooper and Borg, 2005) και ένα εργαλείο οπτικοποίησης στις 3 διαστάσεις (Arena3D, (Pavliopoulos, et al., 2008).

Στη συνέχεια, όλα τα δεδομένα εμπλουτίζονται περαιτέρω με πληροφορίες από διάφορες βάσεις δεδομένων που περιέχουν πληροφορίες για πρωτεΐνες, γονίδια, συντηρημένα μοτίβα, τρισδιάστατες δομές, ασθένειες, πρωτεϊνική λειτουργία, φάρμακα και μεταβολικά μονοπάτια. Πηγαίνοντας ένα βήμα παραπέρα, η Human-grDB δεν συνδέεται μόνο με άλλες πηγές, αλλά έρχεται επίσης με κάποιες δυνατότητες ανάλυσης για να κάνει την ενοποίηση δεδομένων πιο χρήσιμη (Satagoram, et al., 2010).

Όλα τα παραπάνω δεδομένα μεταφέρονται στο χρήστη μέσω μιας διαδικτυακής εφαρμογής (Εικόνα 2.7). Η βάση δεδομένων υλοποιήθηκε σε MySQL, ενώ για τη διεπαφή (GUI) χρησιμοποιήθηκε HTML. Τα δυναμικά κομμάτια της διεπαφής υποστηρίζονται από JavaScript. Η επικοινωνία μεταξύ της βάσης δεδομένων και της διεπαφής γίνεται μέσω προγραμμάτων CGI, και, τέλος, όλοι οι απαραίτητοι υπολογισμοί υλοποιούνται με προγράμματα σε Perl.

2.2.2 Η βάση δεδομένων PLHG-DB

Οι υποκαταστάτες που προσδέονται στους GPCRs αποτελούν σημαντικό κομμάτι των επαγόμενων μονοπατιών μεταγωγής σήματος, καθώς είναι ουσιαστικά το εναρκτήριο σήμα της διαδικασίας. Όπως αναλύθηκε στην Εισαγωγή, μεγάλος αριθμός υποκαταστατών είναι πεπτιδικής φύσεως. Η PLHG-DB είναι μια δημόσια διαθέσιμη, σχεσιακή βάση δεδομένων των πεπτιδικών υποκαταστατών των GPCRs του ανθρώπου.

2.2.2.1 Συλλογή δεδομένων και κατασκευή της βάσης

Το πρώτο βήμα για την εύρεση των πεπτιδικών υποκαταστατών των GPCRs του ανθρώπου ήταν η εκτενής αναζήτηση στη διεθνή βιβλιογραφία, με στόχο να βρεθούν οι διάφοροι υποκαταστάτες (Ligands) που προσδέονται σε κάθε GPCR. Προέκυψε ένας κατάλογος με τους πεπτιδικούς υποκαταστάτες (peptide ligands) αντιστοιχισμένους στις διάφορες GPCR υπο-οικογένειες, που υπάρχουν. Οι υποκαταστάτες αυτοί αλληλεπιδρούν με 128 GPCR υπο-οικογένειες, οι οποίες είναι οργανωμένες σε 41 οικογένειες και ανήκουν σε 3 κατηγορίες, την Κατηγορία A (Rhodopsin-like), την Κατηγορία B (Secretin-like) και την Κατηγορία F (Frizzled/Smoothened).

Παρατηρήθηκε ότι, πολλοί πεπτιδικοί υποκαταστάτες μπορούν να αλληλεπιδρούν με πολλούς GPCRs, λόγω της ομοιότητας μεταξύ των υποκαταστατών, αλλά η PLHG-DB

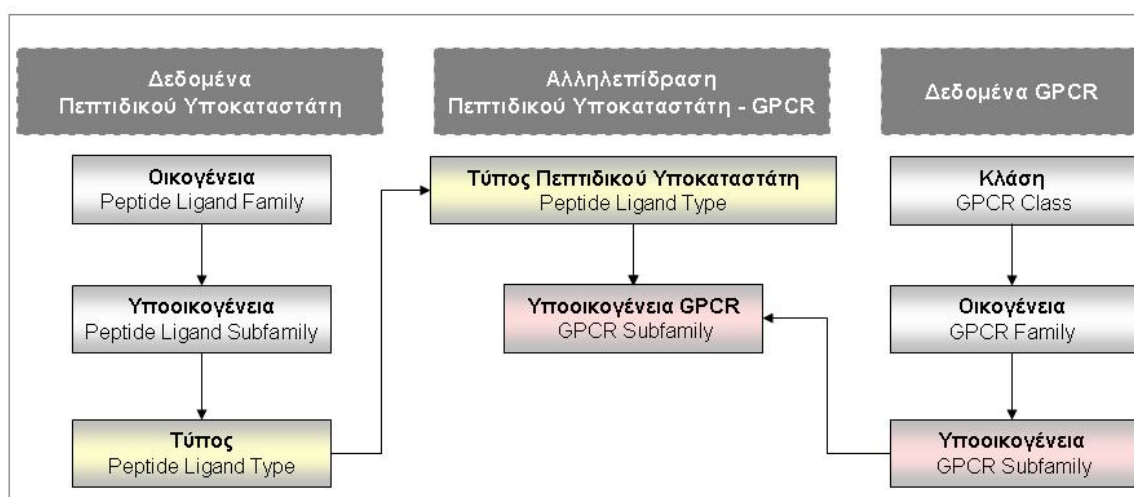
διατήρησε τις βασικές αλληλεπιδράσεις μεταξύ GPCR και πεπτιδικών υποκαταστατών. Όπου παρατηρήθηκε ότι δεν υπάρχει ένας αποκλειστικός πεπτιδικός υποκαταστάτης, έχουν καταγραφεί όλες οι αλληλεπιδράσεις.

Αφού καταγράφηκαν οι αντίστοιχοι πεπτιδικοί υποκαταστάτες του κάθε GPCR, έγινε αναζήτηση στην UniProt (UniProt Release 2012_12) (UniProt, 2012) για να συλλεχθούν οι βασικές αρχικές πληροφορίες που αφορούν κάθε πεπτιδικό υποκαταστάτη. Η συγκεκριμένη αναζήτηση έγινε με την εισαγωγή κατάλληλων λέξεων-κλειδιών (keywords) στα πεδία αναζήτησης DE (Description) και GN (Gene), και εστίασαν στον OS (Organism) Homo sapiens με μία εξαίρεση που αφορούσε τους Formyl Peptide Receptors, των οποίων η φύση τους είναι τέτοια που προσδένουν και υποκαταστάτες από άλλους οργανισμούς.

Από μία UniProt εγγραφή μπορούν να προκύψουν περισσότεροι πεπτιδικοί υποκαταστάτες, οι οποίοι προκύπτουν από το ίδιο προπεπτίδιο. Συνεπώς, αν και το σύνολο των εγγραφών που συλλέχθηκαν ήταν 163, ο αριθμός των πεπτιδικών υποκαταστατών ήταν μεγαλύτερος κι ίσος με 207 πεπτιδικούς υποκαταστάτες, που κατηγοριοποιήθηκαν σε 128 υπο-οικογένειες και 48 οικογένειες με βάση τη λειτουργία τους, αλλά και την αντίστοιχη βιβλιογραφία.

Για κάθε πεπτιδικό υποκαταστάτη, αλλά και το προπεπτίδιο από το οποίο προκύπτει, συλλέχθηκαν διάφορες χρήσιμες πληροφορίες από την εγγραφή της UniProt. Είναι σημαντικό να σημειωθεί ότι οι αλληλουχίες των πεπτιδικών υποκαταστατών που συλλέχθηκαν, διασταυρώθηκαν με τη διαθέσιμη βιβλιογραφία, και, όπου προέκυψαν διαφορές μεταξύ της UniProt και της βιβλιογραφίας, διατηρήθηκαν τα δεδομένα που είναι σύμφωνα με τη βιβλιογραφία. Τα δεδομένα αυτά εμπλουτίστηκαν με πληροφορίες για την αλληλεπίδραση συγκεκριμένων πεπτιδικών υποκαταστατών με συγκεκριμένους υποδοχείς, όπως αυτά προέκυψαν από τη βιβλιογραφική αναζήτηση. Κάθε αλληλεπίδραση συνοδεύεται από ένα σχόλιο, το οποίο έχει προκύψει από τη μελέτη της αντίστοιχης βιβλιογραφίας, καθώς και από συνδέσμους για τα ερευνητικά άρθρα από όπου προέκυψαν οι πληροφορίες αυτές.

Όλα αυτά τα δεδομένα οργανώθηκαν σε μία βάση δεδομένων μέσω της MySQL, της οποίας το συνοπτικό σχεσιακό μοντέλο φαίνεται στην Εικόνα 2.8. Για το σχεδιασμό της διαδικτυακής διεπαφής χρησιμοποιήθηκαν HTML, JavaScript, CSS, καθώς και PHP.



Εικόνα 2.8: Το σχεσιακό μοντέλο της βάσης δεδομένων PLHG-DB.

2.3 Μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες

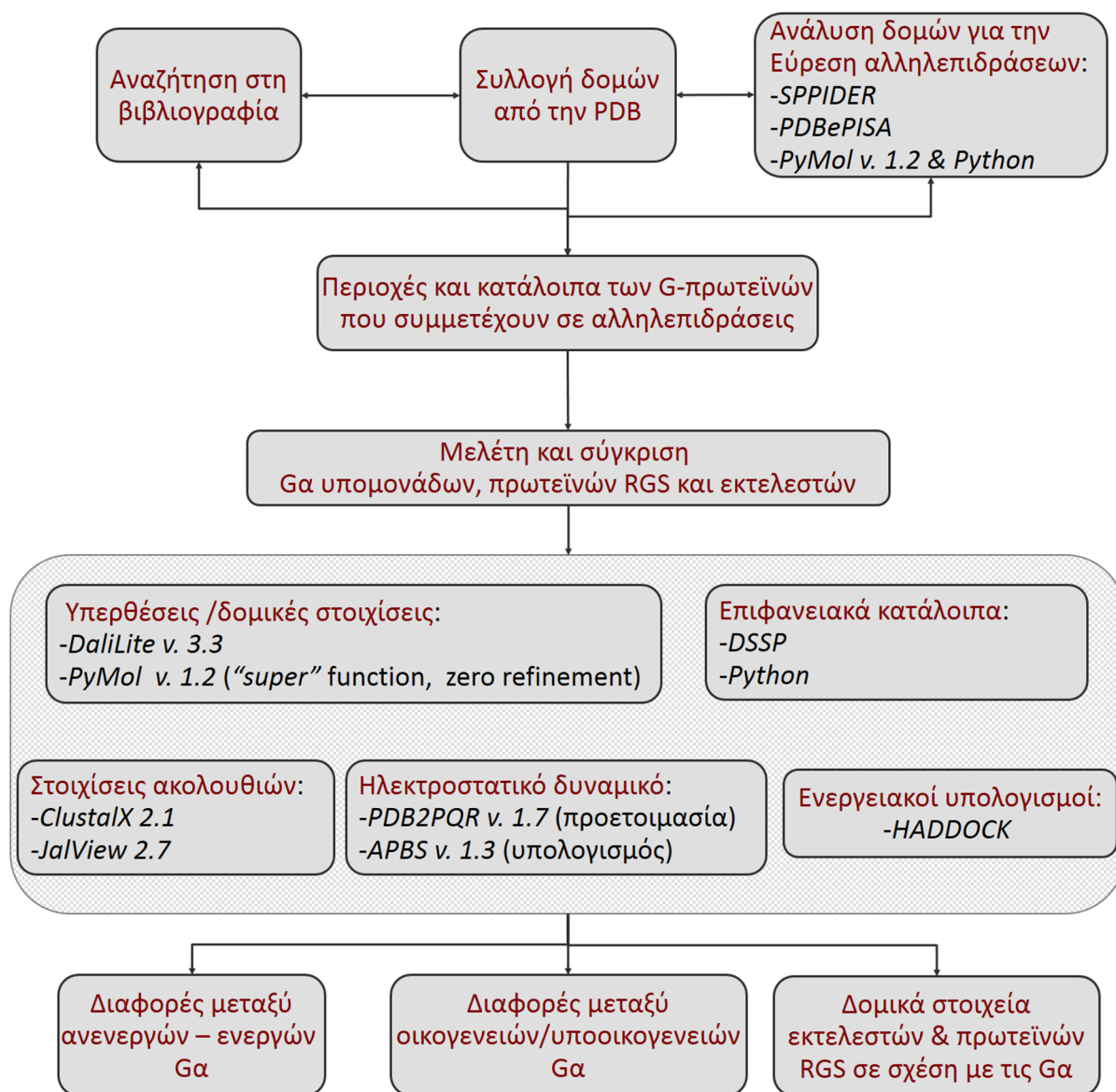
Στην παρούσα ενότητα, μελετήθηκαν οι αλληλεπιδράσεις των G-πρωτεϊνών με τους GPCRs και τα μόρια εκτελεστές, καθώς, ταυτόχρονα και η σχέση των δομικών χαρακτηριστικών των G-πρωτεϊνών με τις αλληλεπιδράσεις αυτές. Για να επιτευχθεί αυτό, ακολουθήθηκε η πορεία εργασίας που απεικονίζεται συνοπτικά στην Εικόνα 2.9.

Αρχικά, μέσω εκτενούς βιβλιογραφικής αναζήτησης για τις αλληλεπιδράσεις των G-πρωτεϊνών, συγκεντρώθηκαν δεδομένα προσδιορισμένων πειραματικά δομών, πειραμάτων μεταλλαξογένεσης και υπολογιστικών μελετών. Όλες οι κρυσταλλογραφικά προσδιορισμένες δομές G-πρωτεϊνών συγκεντρώθηκαν και παρατίθενται στον Πίνακα 1.2 του Παραρτήματος Ι. Επιπλέον συγκεντρώθηκαν και αρκετές δομές RGS πρωτεϊνών. Με τη βοήθεια του προγράμματος SPPIDER (Porollo and Meller, 2007) εντοπίστηκαν τα αμινοξικά κατάλοιπα που συμμετέχουν στις αλληλεπιδράσεις των Gα με τα υπόλοιπα μόρια στις δομές των συμπλόκων. Οι αλληλουχίες για όλες τις Gα υπομονάδες και τις RGS πρωτεΐνες με προσδιορισμένες δομές ανακτήθηκαν από τη βάση UniProt (UniProt Release 2013_2) (UniProt, 2013) και δημιουργήθηκαν πολλαπλές στοιχίσεις με το Clustal X 2.1 (Larkin, et al., 2007). Το JalView 2.7 (Clamp, et al., 2004; Waterhouse, et al., 2009) χρησιμοποιήθηκε για την οπτικοποίηση των πολλαπλών στοιχίσεων.

Αφού εντοπίστηκαν οι περιοχές και τα συγκεκριμένα κατάλοιπα των Gα υπομονάδων σε επίπεδο δομής και αλληλουχίας, στη συνέχεια διερευνήθηκε η συμπεριφορά τους κατά την ενεργοποίηση της G-πρωτεΐνης, συγκρίνοντας τις δομές ανενεργών και ενεργών ή

ενδιάμεσων καταστάσεων των Gα μέσω δομικών στοιχίσεων. Επιπλέον, συγκρίθηκαν τα μέλη των διαφόρων οικογενειών και υπο-οικογενειών μέσω υπέρθεσης των δομών τους. Για την επιλογή των κατάλληλων δομών χρησιμοποιήθηκαν ως κριτήρια τόσο η διακριτικότητα στην οποία έχει προσδιοριστεί η εκάστοτε δομή, όσο και η ομοιότητα που παρουσιάζει η μελετώμενη αλληλουχία και δομή με τη φυσικού τύπου (wild type) πρωτεΐνη. Όπου αυτό ήταν εφικτό, οι δομικές στοιχίσεις κατασκευάστηκαν χρησιμοποιώντας δομές αντίστοιχης διακριτικότητας (2.2 – 2.9 Å) και με το μικρότερο δυνατό αριθμό μεταλλαγών και με τις λιγότερες δυνατές περιοχές χωρίς δομή ή με μη κανονική δομή.

Στη συνέχεια, υπολογίστηκε το ηλεκτροστατικό δυναμικό για τις τέσσερις οικογένειες Gα πρωτεϊνών, καθώς και για τα μόρια εκτελεστές, αλλά και τις RGS πρωτεΐνες, οι οποίες βρέθηκαν σε δομές συμπλόκων με τις Gα πρωτεΐνες. Οι δομικές στοιχίσεις και οι υπολογισμοί του RMSD (τιμή της ρίζας των μέσων τετραγώνων των αποκλίσεων, Root Mean Square Deviation) πραγματοποιήθηκαν μέσω του Dali Server (Holm and Rosenstrom, 2010). Το εργαλείο PDB2PQR (Dolinsky, et al., 2007; Dolinsky, et al., 2004; Unni, et al., 2011) χρησιμοποιήθηκε για την προετοιμασία των δομών για υπολογισμούς, ενώ οι χάρτες ηλεκτροστατικού δυναμικού υπολογίστηκαν με το εργαλείο APBS (Baker, et al., 2001; Unni, et al., 2011). Το πρόγραμμα μοριακών γραφικών PyMol (DeLano, 2002) χρησιμοποιήθηκε για τις μετρήσεις αποστάσεων και την κατασκευή των μοντέλων. Συμπληρωματικά προσδιορίστηκαν τα επιφανειακά κατάλοιπα των Gα υπομονάδων μέσω του DSSP (Define Secondary Structure of Proteins) (Kabsch and Sander, 1983). Τέλος, μια σειρά από κρυσταλλικές δομές συμπλόκων Gα – RGS και Gα – εκτελεστών μελετήθηκαν ενεργειακά, και προσδιορίστηκαν οι τιμές των διαμοριακών ενεργειών των αλληλεπιδράσεων μεταξύ τους με τη χρήση του διαδικτυακού server του HADDOCK (<http://haddock.chem.uu.nl/>) (de Vries, et al., 2010) και συγκεκριμένα του “Refinement Interface”.



Εικόνα 2.9: Πορεία εργασίας για τη μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με τους GPCRs και τα μόρια εκτελεστές.

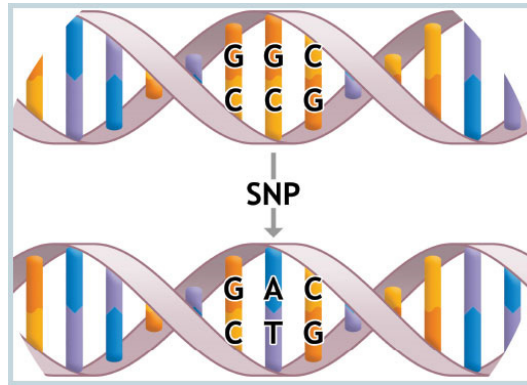
2.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες

2.4.1 Μονονουκλεοτιδικοί πολυμορφισμοί (SNPs)

Ένας μονονουκλεοτιδικός πολυμορφισμός, SNP (Single Nucleotide Polymorphism), είναι μια ενδογενής διαφοροποίηση του γονιδιώματος (Εικόνα 2.10). Τα SNPs είναι η πιο απλή και πιο κοινή «πηγή» γενετικού πολυμορφισμού στο ανθρώπινο γονιδίωμα (90% όλων των ανθρώπινων DNA πολυμορφισμών).

Τα SNPs μπορεί να προκύψουν από δύο είδη αντικαταστάσεων νουκλεοτιδικών βάσεων:

1. Μία **μετάβαση (transition)** συμβαίνει όταν πραγματοποιείται αντικατάσταση πουρίνης με πουρίνη (A,G) ή πυριμιδίνης με πυριμιδίνη (C,T). Από αυτό το είδος της αντικατάστασης δημιουργούνται τα δύο τρίτα των SNPs.
2. Μία **μεταστροφή (transversion)** συμβαίνει όταν πραγματοποιείται αντικατάσταση πουρίνης με πυριμιδίνη ή το αντίστροφο.



Εικόνα 2.10: Σχηματική αναπαράσταση ενός μονονουκλεοτιδικού πολυμορφισμού.

Η ετερογένεια σε επίπεδο ακολουθίας που προκαλείται από τα SNPs μπορεί να μετρηθεί με βάση την ποικιλία νουκλεοτιδίων και από το λόγο του αριθμού των διαφορετικών βάσεων ανάμεσα σε δύο γονιδιώματα σε σχέση με τον αριθμό των βάσεων που συγκρίνονται. Αυτός ο λόγος κυμαίνεται από 1/1000-1/1350 ζεύγη βάσεων μεταξύ δύο ισοδύναμων χρωμοσωμάτων.

Τα SNPs δεν κατανέμονται ομοιόμορφα στο ανθρώπινο γονιδίωμα, ούτε στα χρωμοσώματα, ούτε κατά μήκος ενός μεμονωμένου χρωμοσώματος. Το ένα τρίτο των SNPs εντοπίζονται σε κωδικοποιούσες περιοχές σε σχέση με τα SNPs που εντοπίζονται σε μη κωδικοποιούσες περιοχές. Έχει δειχθεί επίσης ότι η ετερογένεια σε επίπεδο ακολουθίας είναι πολύ μικρότερη στα φυλετικά χρωμοσώματα. Σε ένα μεμονωμένο χρωμόσωμα, τα SNPs μπορεί να συγκεντρώνονται γύρω από μία συγκεκριμένη περιοχή, υποδεικνύοντας συνήθως περιοχές με φαρμακευτικό ή ερευνητικό ενδιαφέρον.

Ένα SNP που βρίσκεται σε κωδικοποιούσα περιοχή μπορεί να έχει δύο πιθανά αποτελέσματα στην τελική πρωτεΐνη:

- ❖ **Συνώνυμη (synonymous)** καλείται η αντικατάσταση που δεν προκαλεί αλλαγή ενός αμινοξικού καταλοίπου στην πρωτεΐνη που παράγεται. Καλείται και σιωπηλή μεταλλαγή (silent mutation).
- ❖ **Μη συνώνυμη (non-synonymous)** είναι η αντικατάσταση που οδηγεί σε αλλαγή του αμινοξικού καταλοίπου που κωδικοποιείται. Μία *παρερμηνεύσιμη* μεταλλαγή

(missense mutation) αλλάζει την πρωτεΐνη προκαλώντας αλλαγή σε ένα κωδικόνιο. Μία *ανερμηνεύσιμη* μεταλλαγή (nonsense mutation) οδηγεί σε αλλαγή θέσης εμφάνισης κωδικονίου λήξης. Το 50% των SNPs που εμφανίζονται σε κωδικοποιούσες περιοχές υπολογίζεται ότι οδηγεί σε μη συνώνυμες αλλαγές κωδικονίων.

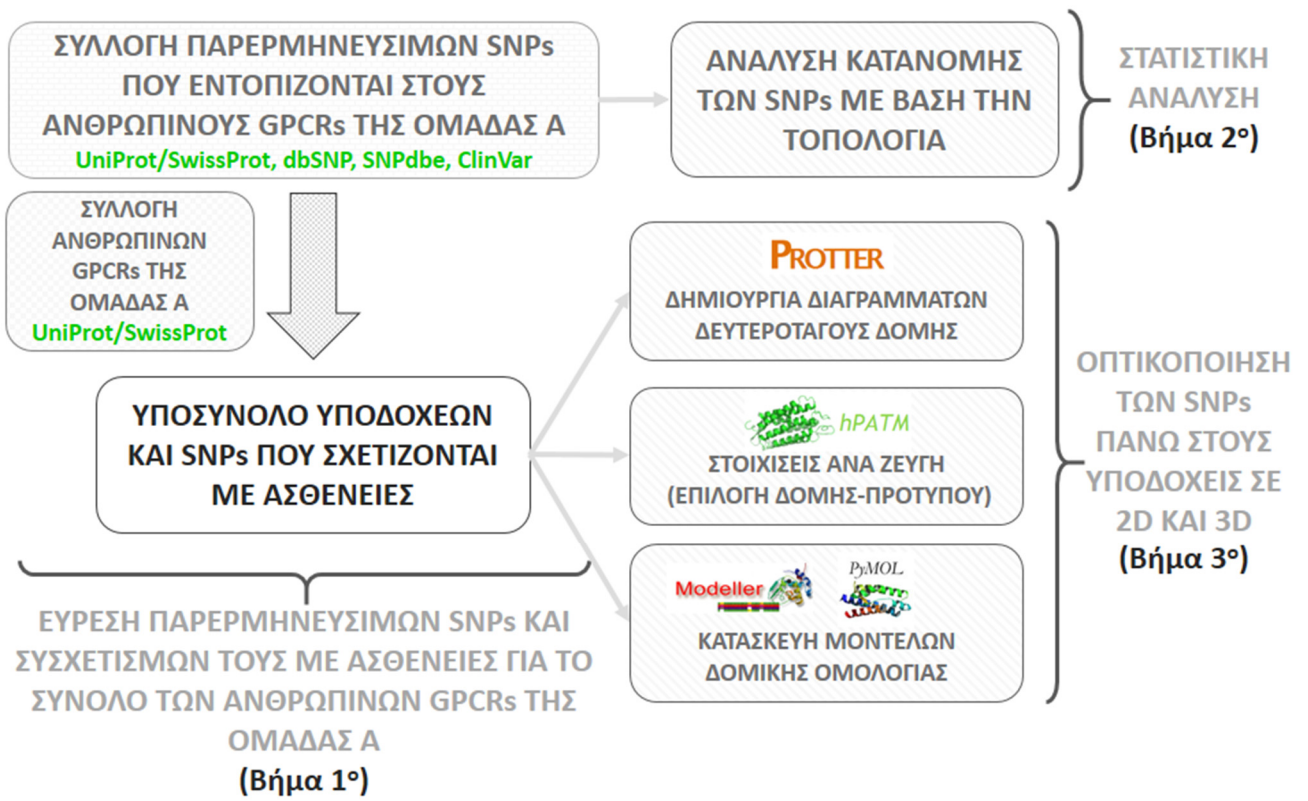
Επίσης SNPs μπορεί να συμβαίνουν και σε ρυθμιστικές περιοχές γονιδίων. Αυτά τα SNPs είναι ικανά να αλλάζουν την **ποσότητα** ή το **χρονικό πλαίσιο** στο οποίο εκφράζεται μία πρωτεΐνη. Αυτά τα SNPs είναι πιο δύσκολο να εντοπιστούν καθώς η γονιδιακή ρύθμιση δεν έχει κατανοηθεί πλήρως μέχρι σήμερα.

2.4.2 Συλλογή δεδομένων

Για τη συλλογή των **παρερμηνεύσιμων** SNPs στους ανθρώπινους GPCRs της **Κατηγορίας A** και τη συσχέτισή τους με ασθένειες ακολουθήθηκε η πορεία εργασίας, όπως αυτή περιγράφεται στην Εικόνα 2.11.

1. Συγκέντρωση συνόλου των υπό μελέτη υποδοχέων από τις βάσεις δεδομένων πρωτεϊνικών ακολουθιών UniProtKB/SwissProt (UniProt Release 2012_12) (UniProt, 2012) και RefSeq (Pruitt, et al., 2012) του NCBI. Εύρεση παρερμηνεύσιμων SNPs στο σύνολο των υποδοχέων μέσω των βάσεων dbSNP του NCBI (Sherry, et al., 2001), UniProtKB/SwissProt, SNPdbe (Schaefer, et al., 2012) και ClinVar (Riggs, et al., 2013). Τέλος, συσχέτιση συγκεκριμένων SNPs με ασθένειες. Δυστυχώς δεν υπάρχει βάση δεδομένων η οποία να επικεντρώνεται σε τέτοιου είδους πληροφορία. Εξαίρεση αποτελεί η ClinVar, η οποία παρόλα αυτά δημιουργήθηκε πρόσφατα και δεν θεωρούνταν ακόμη αρκετά αξιόπιστη καθώς τα δεδομένα που παρέχονταν δεν ήταν ακόμη πλήρως επιβεβαιωμένα. Δεδομένα παθογένειας ή μη συγκεκριμένων μονονουκλεοτιδικών πολυμορφισμών παρέχονταν και από τις dbSNP, UniProtKB/SwissProt, SNPdbe χωρίς να είναι μεγάλης αξιοπιστίας. Έτσι προέκυψε ένα κενό πληροφορίας σχετικά με τη συμμετοχή σε παθογένεια ή μη των πολυμορφισμών.
2. Στατιστική ανάλυση κατανομής των SNPs με βάση την τοπολογία των GPCRs στους οποίους εντοπίζονται.
3. Η οπτικοποίηση των μονονουκλεοτιδικών πολυμορφισμών σε επίπεδο δευτεροταγούς και τριτοταγούς δομής. Για το σκοπό αυτό χρησιμοποιήθηκαν τα

εργαλεία Protter (Omasits, et al., 2013) και Modeller (Sali and Blundell, 1993) αντίστοιχα.



Εικόνα 2.11: Η πορεία εργασίας για τη μελέτη των παρερμηνεύσιμων SNPs στους ανθρώπινους GPCRs της Κατηγορίας Α και της συσχέτισής τους με ασθένειες.

2.5 Μέτρηση της επιτυχίας των μεθόδων

Για να μπορέσουμε να μετρήσουμε την επιτυχία και την αξιοπιστία των προγνώσεων που προέρχονται από μια μέθοδο, έχουν προταθεί διάφορα μέτρα. Για την αξιολόγηση των προγνώσεων όπου μία ακολουθία ανήκει ή όχι σε μία κατηγορία (π.χ. είναι ή όχι πιθανός GPCR), χρησιμοποιείται ο γνωστός συντελεστής συσχέτισης του Matthews (C) (Baldi, et al., 2000):

$$C = \frac{TP \times TN - FP \times FN}{\sqrt{(TP \times FN)(TP \times FP)(TN \times FP)(TN \times FN)}} \quad (2.1)$$

Με TP (True Positives) συμβολίζουμε τον αριθμό ορθώς θετικά προσδιορισμένων αλληλουχιών, TN (True Negatives) τον αριθμό των ορθώς αρνητικά προσδιορισμένων αλληλουχιών, FN (False Negatives) τον αριθμό των εσφαλμένως αρνητικά προσδιορισμένων

καταλοίπων, και FP (False Positives) τον αριθμό των εσφαλμένως θετικά προσδιορισμένων καταλοίπων.

Η επιτυχία των μεθόδων αξιολογείται επίσης με τα μέτρα της ευαισθησίας (Sensitivity, S_n) και της ειδικότητας (Specificity, S_p) τα οποία δίνονται από τους παρακάτω τύπους:

$$S_n = \frac{TP}{TP + FN} \quad (2.2)$$

$$S_p = \frac{TN}{TN + FP} \quad (2.3)$$

Με τα TP, TN, FP και FN όπως περιεγράφηκαν προηγούμενα.

Έτσι μια αξιόπιστη μέθοδος θα πρέπει να μπορεί να εντοπίζει σωστά όσο περισσότερες πρωτεΐνες είναι εφικτό από το θετικό σύνολο ελέγχου (μεγάλη ευαισθησία), και, ταυτόχρονα να χαρακτηρίζει λανθασμένα ως πρωτεΐνες της κατηγορίας αυτής αλληλουχίες όσο λιγότερες αλληλουχίες από το αρνητικό σύνολο ελέγχου είναι εφικτό (μεγάλη ειδικότητα). Τέλος σε πολλές περιπτώσεις χρησιμοποιείται και ως περισσότερο αξιόπιστο μέτρο ή σταθμισμένη ακρίβεια (balanced accuracy):

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (2.4)$$

3. Αποτελέσματα και Συζήτηση

3.1 Αλγόριθμοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών

3.1.1 Η μέθοδος GPCRipe

Η μέθοδος GPCRipe εντοπίζει αποτελεσματικά και προσφέρει λεπτομερή σχολιασμό δομής και λειτουργίας για συζευγμένους με G-πρωτεΐνες υποδοχείς. Στο πρώτο επίπεδο της μεθόδου γίνεται ο εντοπισμός των πιθανών υποδοχέων, ενώ στο δεύτερο επίπεδο οι υποδοχείς που έχουν βρεθεί σχολιάζονται περαιτέρω με τη χρήση προϋπαρχόντων αλγορίθμων. Δύο κριτήρια χρησιμοποιούνται για την ανίχνευση των πιθανών GPCRs. Σύμφωνα με το πρώτο κριτήριο, μία πρωτεΐνη είναι GPCR εάν προβλέπεται από το ειδικά σχεδιασμένο HMM να έχει 7 διαμεμβρανικά τμήματα. Λόγω της καθορισμένης θέσης τόσο του αμινοτελικού όσο και του καρβοξυτελικού άκρου στο μοντέλο, είναι σημαντικό να σημειωθεί ότι εντοπίζει πάντα περιττό αριθμό διαμεμβρανικών τμημάτων. Σύμφωνα με το δεύτερο κριτήριο μία πρωτεΐνη είναι GPCR αρκεί να “ταιριάζει” με ένα από τα 39 ειδικά pHMMs. Για κάθε pHMM ως κατώφλι χρησιμοποιείται η τιμή κατωφλιού εμπιστοσύνης (trusted cut-off) που δίνεται από τη βάση Pfam. Η χρήση των δύο αυτών κριτηρίων μπορεί να γίνει είτε εναλλακτικά (OR method) είτε συνδυαστικά (AND method).

3.1.1.1 Σύνολα ελέγχου και αξιολόγηση της μεθόδου

Για την αξιολόγηση της μεθόδου, κατασκευάστηκαν ένα θετικό και ένα αρνητικό σύνολο ελέγχου. Ως θετικό σύνολο, χρησιμοποιήθηκαν 1311 ανθρώπινοι συζευγμένοι με G-πρωτεΐνες υποδοχείς, ενώ ως αρνητικό σύνολο χρησιμοποιήθηκαν 2850 πρωτεΐνες εκ των οποίων οι 169 ήταν α-ελικοειδείς διαμεμβρανικές πρωτεΐνες επιλεγμένες από την PDBTM (Tusnady, et al., 2004; Tusnádý, et al., 2005), οι οποίες δεν είναι GPCRs, και οι υπόλοιπες 2681 είναι σφαιρικές πρωτεΐνες από τη βάση PDB (Berman, et al., 2000). Τα σύνολα αυτά δίνονται στον ιστότοπο <http://aias.biol.uoa.gr/~mtheo/datasets/>.

Η απόδοση της μεθόδου συγκρίθηκε με την απόδοση του αλγορίθμου πρόγνωσης GPCRHMM (Wistrand, et al., 2006) ο οποίος είναι ειδικά σχεδιασμένος για την ανίχνευση

των GPCRs. Επιπλέον χρησιμοποιήθηκαν οι αλγόριθμοι πρόγνωσης τοπολογίας HMM-TM (Bagos, et al., 2006), HMMTOP (Tusnady and Simon, 2001), MEMSAT1 (Jones, et al., 1994), MEMSAT3 (Jones, 2007), MEMSAT-SVM (Nugent and Jones, 2009), OCTOPUS (Viklund and Elofsson, 2008), Philius (Reynolds, et al., 2008), PHOBIUS (Kall, et al., 2004), PolyPhobius (Kall, et al., 2005), PRO (Viklund and Elofsson, 2004), PRODIV (Viklund and Elofsson, 2004), SCAMPI (Bernsel, et al., 2008), SCAMPI-single, TMHMM (Krogh, et al., 2001), TOPCONS (Bernsel, et al., 2009), TOPCONS-single (Hennerdal and Elofsson, 2011) και TopPred2 (von Heijne, 1992), και σε αυτή την περίπτωση ως σωστές προβλέψεις θεωρήθηκαν οι περιπτώσεις όπου μια πρωτεΐνη προβλεπόταν με 7 διαμεμβρανικά τμήματα και με το αμινοτελικό άκρο στον εξωκυττάριο χώρο.

Στο πρώτο στάδιο της αξιολόγησης, εκτιμήθηκε η απόδοση του τροποποιημένου HMM μέσω διαδικασίας cross-validation κατά την οποία το σύνολο εκπαίδευσης χωρίστηκε σε 10 ισομεγέθη υποσύνολα. Το ~97% (419 από 432) των υποδοχέων εντοπίζεται σωστά, ενώ το αρχικό HMM που χρησιμοποιείται στο HMM-TM εντοπίζει μόνο το 65% (282 από 432) των υποδοχέων. Αντίστοιχη βελτίωση της απόδοσης παρατηρείται και στο θετικό σύνολο ελέγχου των 1311 ανθρώπινων GPCRs (97% σωστές προβλέψεις). Παρατηρήθηκε ότι η μέγιστη απόδοση που επιτυγχάνουν οι αλγόριθμοι πρόγνωσης τοπολογίας είναι της τάξης του 82% (Πίνακας 3.1) (Tsirigos, et al., 2012), και, συνεπώς το GPCRpipe είναι σαφώς πιο αποτελεσματικό στον εντοπισμό πιθανών GPCRs. Ο Πίνακας 3.2 συνοψίζει τα αποτελέσματα της σύγκρισης τόσο της εναλλακτικής (OR method) όσο και της συνδυαστικής (AND method) μεθόδου του GPCRpipe με το GPCRHMM αλλά και τους αλγορίθμους πρόγνωσης τοπολογίας.

Περαιτέρω αξιολόγηση της αποτελεσματικότητας της μεθόδου GPCRpipe (τόσο της εναλλακτικής όσο και της συνδυαστικής) πραγματοποιήθηκε στο θετικό και αρνητικό σύνολο δεδομένων (Πίνακας 3.2). Με τη χρήση της συνδυαστικής μεθόδου παρουσιάζεται μεγαλύτερη ειδικότητα στην πρόγνωση. Παρόλα αυτά, η χρήση της εναλλακτικής μεθόδου δίνει τη δυνατότητα εντοπισμού νέων πιθανών GPCRs που δεν ανήκουν στις γνωστές οικογένειες που εκφράζουν τα rHMMs. Συγκεκριμένα, η αρχιτεκτονική του HMM προσφέρει τη δυνατότητα να προγνώσει πιθανούς GPCRs που δεν παρουσιάζουν τα ήδη γνωστά μοτίβα των διαφόρων κατηγοριών και οικογενειών.

Πίνακας 3.1 Σύγκριση της ακρίβειας εντοπισμού (accuracy), στο σύνολο των 1311 ανθρωπίνων GPCRs

Method	Ποσοστό των GPCRs που έχουν προβλεφθεί σωστά
GPCRpipe (AND method)	97%
GPCRpipe (OR method)	100%
GPCRHMM	99%
HMM-TM	38%
HMMTOP	82%
MEMSAT1	56%
MEMSAT3	64%
MEMSAT-SVM	36%
OCTOPUS	68%
Philius	55%
Phobius	68%
PolyPhobius	82%
PRO	52%
PRODIV	76%
SCAMPI	65%
SCAMPI-single	23%
SPOCTOPUS	52%
TMHMM	69%
TOPCONS	74%
TOPCONS-single	53%
TopPred2	38%

Πίνακας 3.2 Τα αποτελέσματα της αξιολόγησης του GPCRpipe στα σύνολα ελέγχου.

Μέθοδος	Sn*	Sp*	Acc*	FPR*	C*
GPCRpipe (AND method)	91.15%	100.00%	97.21%	0.00%	0.936
GPCRpipe (OR method)	100.00%	99.82%	99.88%	0.18%	0.997
GPCRHMM	95.27%	100.00%	98.51%	0.00%	0.966

*Sn: Ευαισθησία, Sp: Ειδικότητα, Acc: Ακρίβεια, FPR: το ποσοστό των αρνητικών παραδειγμάτων που κατηγοριοποιούνται λάθος και C: ο συντελεστής συσχέτισης του Mathew's.

Σε σύγκριση με τον αλγόριθμο GPCRHMM, η μέθοδος εμφανίζει παρόμοια αξιοπιστία αποτελεσμάτων, αλλά προσφέρει επιπλέον πληροφορίες λειτουργίας και δομής, όπως ήδη αναφέρθηκε.

3.1.1.2 Εντοπισμός GPCRs σε πρωτεώματα

Για να αναδειχθεί η λειτουργικότητα του GPCRipe, εντοπίστηκαν οι GPCRs σε πέντε ολοκληρωμένα πρωτεώματα οργανισμών, τα οποία ανακτήθηκαν από τη βάση δεδομένων UniProt (UniProt, 2013). Τα αποτελέσματα συνοψίζονται στον Πίνακα 3.3. Παρατηρείται ότι, το ποσοστό εμφάνισης GPCRs στα πρωτεώματα ποικίλει από ~2% μέχρι ~8%. Οι πιθανοί GPCRs κατηγοριοποιούνται με βάση την οικογένεια στην οποία ανήκουν (βλέπε Πίνακα 2.1).

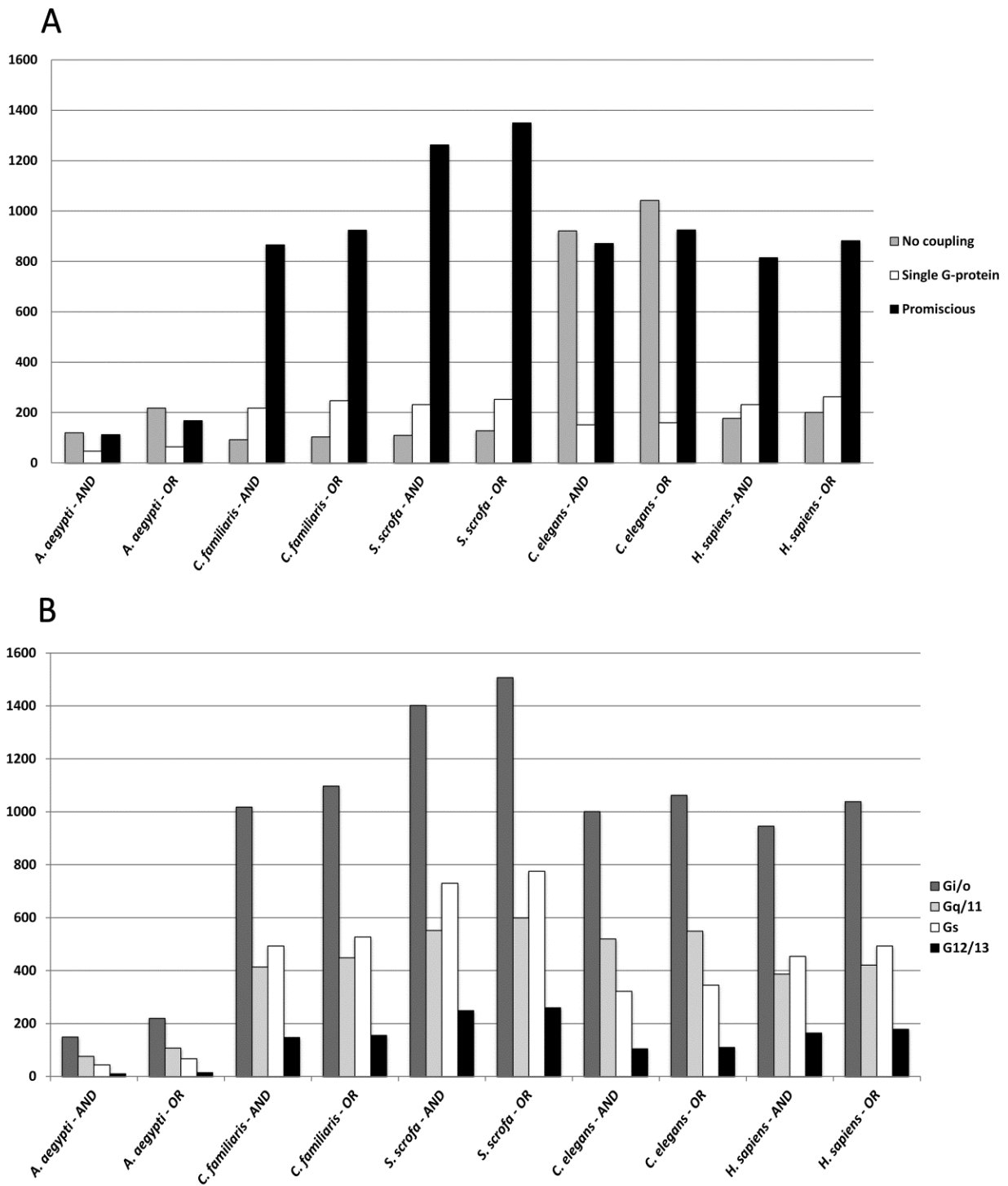
Πίνακας 3.3 Εντοπισμός των πιθανών GPCRs σε πρωτεώματα πέντε οργανισμών με τη μέθοδο GPCRipe.

Οργανισμός	Μέγεθος πρωτεώματος	Πιθανοί GPCRs	Class A	Class B	Class C	Class F	Insect Chemosensory
<i>Aedes aegypti</i>	16663	278 (1.67%) / 450 (2.70%)	93 / 150	6 / 11	12 / 16	7 / 13	160 / 260
<i>Canis familiaris</i>	26423	1176 (4.45%) / 1274 (4.82%)	1061 / 1126	72 / 97	30 / 35	13 / 15	-
<i>Sus scrofa</i>	26070	1602 (6.14%) / 1730 (6.64%)	1530 / 1623	41 / 66	22 / 32	9 / 9	-
<i>Caenorhabditis elegans</i>	25179	1943 (7.72%) / 2126 (8.44%)	1910 / 2076	4 / 6	10 / 15	10 / 10	9 / 18
<i>Homo sapiens</i>	70548	1223 (1.73%) / 1344 (1.91%)	1051 / 1120	108 / 154	49 / 54	15 / 16	-

Οι στήλες 3-8 δείχνουν τον αριθμό των πιθανών GPCRs τόσο με τη συνδυαστική (AND) όσο και με την εναλλακτική (OR) μέθοδο του GPCRipe χωρισμένους με “/”.

Πιο συγκεκριμένα, οι GPCRs φαίνεται να αποτελούν περίπου το 2% του πρωτεώματος του ανθρώπου, με τα μέλη της Κατηγορίας A να είναι τα περισσότερα, και την Κατηγορία F να έχει τους λιγότερους εκπροσώπους. Στο *Aedes aegypti*, πάνω από 2% του πρωτεώματος αντιπροσωπεύει GPCRs, όμως σε αυτή την περίπτωση οι χημειοαισθητήριοι υποδοχείς

των εντόμων είναι οι πιο συχνά εμφανιζόμενοι. Στον *Ceanorhabditis elegans*, περίπου το 8% του πρωτεώματος του αποτελείται από GPCRs, με πολύ μεγάλη αντιπροσώπευση των ειδικών για νηματώδεις χημειοϋποδοχέων την Κατηγορίας A (class A – Nematode specific chemoreceptors) και ελάχιστους αντιπροσώπους από τις υπόλοιπες κατηγορίες. Η παρουσία μικρού αριθμού χημειοαισθητήριων υποδοχέων εντόμων στον *C. elegans* εξηγείται από την ομοιότητα σε επίπεδο αμινοξικής ακολουθίας που εμφανίζουν οι, ειδικοί για νηματώδεις, χημειοϋποδοχείς με τα μέλη της 7tm_7 οικογένειας της Pfam, και, κατά συνέπεια, με τους γευστικούς και οσφρητικούς υποδοχείς εντόμων. Το γεγονός αυτό υποδηλώνει εξελικτική συγγένεια μεταξύ αυτών των υποδοχέων. Στους οργανισμούς *Canis familiaris* και *Sus scrofa* περίπου 5% και 6% του πρωτεώματος αντίστοιχα αφορά GPCRs. Και στους δύο αυτούς οργανισμούς η Κατηγορία A είναι η πολυπληθέστερη, ενώ ειδικά στο *S. scrofa* η διαφορά μεταξύ της A και των υπολοίπων κατηγοριών είναι σαφώς μεγαλύτερη. Η υπερ-εκπροσώπηση της Κατηγορίας A σε αυτούς τους δύο οργανισμούς ίσως μπορεί να εξηγηθεί από την παρουσία μεγάλου αριθμού οσφρητικών υποδοχέων (και τα δύο αυτά είδη φημίζονται για την αίσθηση της όσφρησης). Το γεγονός ότι μεγάλο ποσοστό των υποδοχέων φαίνεται να αλληλεπιδρά με G_s πρωτεΐνες (Εικόνα 3.1) ενισχύει αυτή μας την υπόθεση, καθώς οι G_{olf} , οι οποίες εμπλέκονται στην όσφρηση, είναι μέλη της G_s οικογένειας. Πολλοί από τους πιθανούς GPCRs του *C. elegans* δεν φαίνεται να “συζεύγνυται” με κάποια G-πρωτεΐνη, γεγονός που μπορεί να εξηγηθεί ίσως από τη μεγαλύτερη ετερογένεια που είναι γνωστό ότι εμφανίζουν οι G-πρωτεΐνες του ώστε να συνεργάζονται επιτυχώς με τη μεγάλη ποικιλία χημειοϋποδοχέων, και, επιπλέον, από το γεγονός ότι μόνο τα 4 από τα 21 γονίδια που κωδικοποιούν για G-πρωτεΐνες στον *C. elegans* είναι ομόλογα με τις 4 οικογένειες των θηλαστικών (Bastiani and Mendel, 2006). Μικρός αριθμός των υποδοχέων του *A. aegypti* “συζεύγνυται” με την $G_{12/13}$ οικογένεια, καθώς δεν έχει εντοπιστεί με βεβαιότητα ομόλογου της $G_{12/13}$ στα έντομα. Γενικά, η ικανότητα “πολλαπλής σύζευξης” (promiscuous coupling), δηλαδή η ικανότητα ενός υποδοχέα να «συζεύγνυται με περισσότερες από μία G-πρωτεΐνες», εμφανίζεται πολύ συχνά, στους περισσότερους οργανισμούς, και, ίσως, το γεγονός αυτό να εξηγεί την ποικιλία και την πολυπλοκότητα των μονοπατιών μεταγωγής σημάτων που προκύπτουν από τους GPCRs (Εικόνα 3.1).



Εικόνα 3.1: Η ειδικότητα “σύζευξης” των GPCRs σε πέντε πρωτεύματα. Ο εντοπισμός των GPCRs έγινε τόσο με την εναλλακτική όσο και με τη συνδυαστική μέθοδο του GPCRpipe. Α. Οι πιθανοί GPCRs κατηγοριοποιούνται με τη χρήση των αποτελεσμάτων του επιπέδου του σχολιασμού (Feature Annotation layer), σε τρεις κατηγορίες: χωρίς “σύζευξη”, “σύζευξη” αποκλειστικά με μία οικογένεια G-πρωτεϊνών, “ασύδοτο σύζευξη”. Β. Η προβλεφθείσα ικανότητα “σύζευξης” των GPCRs με κάθε οικογένεια G-πρωτεϊνών αντίστοιχα.

3.1.1.3 Η διαδικτυακή διεπαφή για τη μέθοδο GPCRpipe

Η μέθοδος είναι διαθέσιμη στο διαδίκτυο μέσω της σελίδας του εργαστηρίου στη διεύθυνση: <http://bioinformatics.biol.uoa.gr/GPCRpipe>. Ο χρήστης μπορεί να εισάγει τις προς μελέτη αμινοξικές ακολουθίες σε μορφή FASTA, στη διαθέσιμη φόρμα, ή να τις υποβάλει μέσω ενός αρχείου που να τις περιέχει στην ίδια μορφή. Ο χρήστης οφείλει να επιλέξει την εναλλακτική ή τη συνδυαστική μέθοδο (OR ή AND method) προτού υποβάλλει την επερώτηση (Εικόνα 3.2).

GPCR pipe
A pipeline for the detection of G-protein coupled receptors in proteomes

Home Run About

Paste your sequence in **FASTA** format in the field below

```
>sp|P17693|HLA_HUMAN HLA class I histocompatibility antigen, alpha chain G OS=Homo sapiens GN=HLA-G PE=1 SV=1
MVVMAAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPFIAMGYVDDTQFVRF
DSDSACPMEPRAPWVEQEGPEYEEETRNKHAHQTRMNLQTLRGYYNQSEASSHTLQ
WMIGCDLGSDDGRLLRGYEQAYDGDYLDALNEDLRSWTAADTAAQISKRKCEANVAEQR
RAYLEGTCVWLHRYLENGKEMLRADPPKTHVTHHFVFDYEAIRCWALGFYPAEIIILT
WQRDGEDQIQDVELVETRPAGDGTQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ
SSLPTIPIMGIVAGLVVLAAVVTGAAVAALWRKSSD
>sp|P35412|GPR12_MOUSE G-protein coupled receptor 12 OS=Mus musculus GN=Gpr12 PE=2 SV=1
MNEDEPKVNL SGLPRDCIDAGAPENISAAPVPSQGSVAESEPELVNPNWDIVLCSGTLICC
ENAVVLLII FHSFSLRAPMFLIGSLALADLAGLGLIINFVFAYLLQSEATKLVITIGLI
VASFSASVCSLLAITVDRLSILYALTYHSERTVFTFYVLMVLMGTSICLGLLPVMGWN
CLRDESTCSVVRPLTKNNAIILSISFLFMFALMLQLYIQICKIVMRHAHQIALQHHFLAT
SHVVTTRKGVSTLALILGTFAACWMPFTLYSLADYTFYSIYTYATLLPATYNSIINFVI
YAFRNQEIQKALCLICCGCFSSLSQRARSPSDV
```

• Alternatively, upload a file with **FASTA** formatted sequences: Δεν έχει επιλεγεί κανένα αρχείο

AND - Query sequences are considered as propable GPCRs only if BOTH the HMM AND the Pfam library predict them as such.

OR - Query sequences are considered as propable GPCRs if either the HMM OR the Pfam library predict them as such.

UP TO 100 SEQUENCES PER RUN!!!

Note : To perform a large scale analysis contact mtheo@biol.uoa.gr

This service is freely available to academic users only!!!!

Non-academic users should contact Prof. S.J. Hamodrakas (shamodr@biol.uoa.gr).

University of Athens
Faculty of Biology
Biophysics & Bioinformatics Laboratory

powered by

Εικόνα 3.2: Η ιστοσελίδα μέσω της οποίας εκτελείται διαδικτυακά η μέθοδος GPCRpipe.

Τα αποτελέσματα παρουσιάζονται με μορφή πίνακα για κάθε μία από τις αλληλουχίες που υποβλήθηκαν, περιλαμβανομένης και της τοπολογίας, όταν αυτή είναι διαθέσιμη από τα αποτελέσματα του HMM (Εικόνα 3.3). Στον πίνακα παρατίθενται οι ακόλουθες πληροφορίες:


```

Query:  sp|P35412|GPR12_MOUSE
GPCR:   YES
HMM Reliability Score:  0.947963982524515
Topology:      O: 1-48 M: 49-69      I: 70-80      M: 81-103      O: 104-113      M: 114-
136      I: 137-155      M: 156-177      O: 178-198      M: 199-221      I: 222-252      M: 253-
273      O: 274-280      M: 281-303      I: 304-334
Pfam_GPCR:    7tm_1      LM: 61-301
Length: 334
SignalP:      NO
PRED-COUPLE2:      Gs(0.97)      Gi/o(0.47)
Pfam_IN:      NO HITS
Pfam_OUT:     NO HITS
Sequence:
MNEDPKVNL SGLPRDCIDAGAPENISA AVPSQGSVAESEPELVVNPWDIVLCSSGTLICCENAVVVLII FHSPLRAPMFL LIGSLALADLLAGLG
LIINFVFAYLLQSEATKLVITIGLIVASF SASVCSLLAITVD RYLSLYYALTYHSERTVFTTYVMLVMLWGT SICLGLLPVMGWNCLRDESTCSVVR
PLTKNNAAIL SISFLFMFALMLQLYIQICKIVMRHAHQIALQH HFLATSHYVTRKGVSTLALILGTFAACWMPFTLYSLIADYTPSIYTYATLL
PATYNSIINPVIYAFRNQEIQKALCLICCGCIPSSLSQRARSPSDV
//
Query:  sp|P17693|HLAG_HUMAN
GPCR:   NO
Topology:      -
Pfam_GPCR:    -
SignalP:      -
PRED-COUPLE2: -
Pfam_IN:      -
Pfam_OUT:     -
Sequence:
MVMVAPRTLFLLLSGALTLTETWAGSHSMRYFSA AVSRPGRGEPRIAMGYVDDTQFVRFDSDSACPRMEFRAPWVEQEGPEYWEETRNTKAHAQ
TDRMNLQTLRGYYNQSEASSHTLQWMI GCDLGS DGRLLRGYEQYAYDGKDY LALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWL
HRYLENGKEMLQRADPPKTHVTHHPVFDYEATLRCWALGFYPAE ILLTWQRDGEDQTDVELVETRPAGDGT FQKWA AVVVP SGEEQRYTCHVQHE
GLPEPLMLRWKQSSLPITIPIMGIVAGLVVLA AVVTGA AVA AVLWRKSSD
//

```

Εικόνα 3.4: Το αρχείο κειμένου των αποτελεσμάτων της μεθόδου GPCRpipe το οποίο διατίθεται.

Η μέθοδος GPCRpipe εντοπίζει με αξιοπιστία πιθανούς συζευγμένους με G-πρωτεΐνες υποδοχείς σε πρωτεώματα, παρέχοντας ταυτόχρονα σημαντικές πληροφορίες για τη δομή και τη λειτουργία τους. Το HMM που χρησιμοποιείται προβλέπει με μεγάλη αξιοπιστία τη σωστή τοπολογία των GPCRs ενώ η ύπαρξη κυτταροπλασματικών και εξωκυττάρων rHMMs μπορεί να βελτιώσει περαιτέρω την πρόγνωση αυτή (Bernsel and Von Heijne, 2005). Η κατάταξη των GPCRs που προβλέπονται σε ήδη γνωστές οικογένειες και η πρόγνωση για την ειδικότητα «σύζευξης» τους με G-πρωτεΐνες παρέχουν σημαντικές πληροφορίες για τη λειτουργία τους στο κύτταρο.

3.1.2 Η μέθοδος GprotPRED

Η μέθοδος GprotPRED σχεδιάστηκε για να εντοπίζει G-πρωτεΐνες και να τις κατατάσσει στις 4 οικογένειες, αλλά και για τις Gβ και Gγ υπομονάδες. Για το σκοπό αυτό, κατασκευάστηκαν έξι εξειδικευμένα rHMMs, τα οποία και αποτελούν τον πυρήνα του διαδικτυακού εργαλείου GprotPRED, το οποίο και θα περιγραφεί στη συνέχεια.

3.1.2.1 Αξιολόγηση των pHMMs

Τα έξι μοντέλα ελέγχθηκαν για την αξιοπιστία τους μέσω αναζήτησης έναντι της βάσης UniProt/Swiss-Prot (UniProt Release 2012_01, 534.242 εγγραφές). Για το σκοπό αυτό εγκαταστήσαμε τοπικά τη βάση και πραγματοποιήσαμε τον έλεγχο με χρήση της εντολής *hmmsearch* στο εργαλείο HMMER v3.0 (Eddy, 2011). Μέσω αυτής της διαδικασίας, έγινε προσπάθεια να καθοριστεί μία τιμή κατωφλιού (*cut-off*) για κάθε pHMM, η οποία θα διαχωρίζει τις αλληλουχίες που ανήκουν στην οικογένεια, από αυτές που δεν ανήκουν. Για τον ορισμό των κατωφλιών (*cut-offs*), συγκεντρώθηκαν για κάθε pHMM, οι βαθμολογίες (*scores*) όλων των ακολουθιών που είναι μέλη της και όλων των υπολοίπων αποτελεσμάτων. Σύμφωνα με τις βαθμολογίες (*scores*), ορίσαμε για το κάθε μοντέλο τις ακόλουθες τιμές:

- ❖ Κατώφλι εμπιστοσύνης (*trusted cutoff*, TC): Ως TC ορίζεται η βαθμολογία (*score*) του χαμηλότερου θετικού hit
- ❖ Κατώφλι θορύβου (*noise cutoff*, NC): Ως NC ορίζεται η βαθμολογία (*score*) του υψηλότερου αρνητικού hit.

Στη συνέχεια, έγινε υπολογισμός για κάθε profile, για ένα σύνολο ενδεικτικών τιμών βαθμολογιών μεταξύ των TC και NC, της ευαισθησίας και της ειδικότητας και κατασκευάστηκαν διαγράμματα, ώστε να βρεθεί το ιδανικό κατώφλι. Στην Εικόνα 3.5 δίνονται τα διαγράμματα που προέκυψαν από την παραπάνω διαδικασία καθώς και το κατώφλι που προέκυψε.

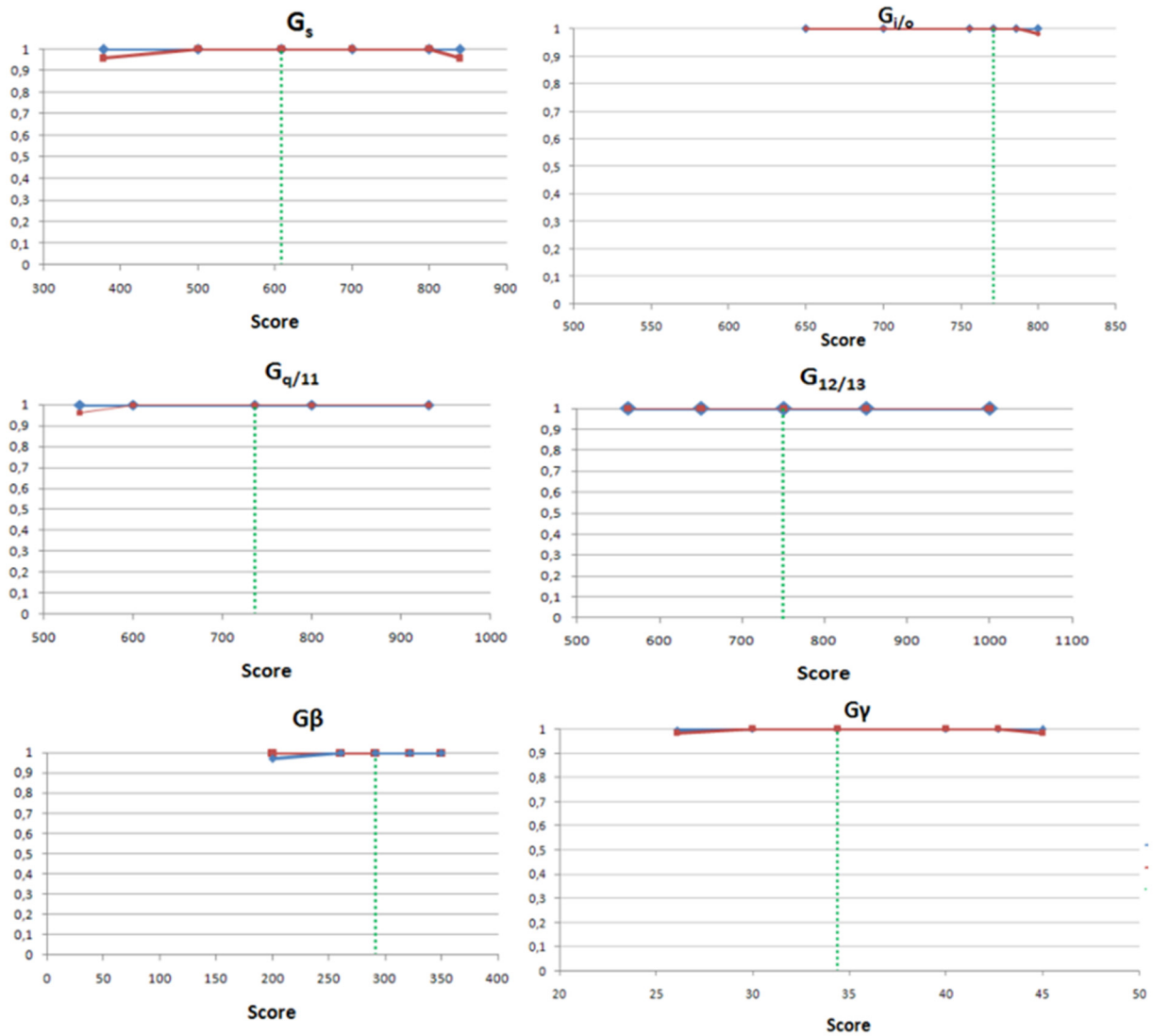
Μέσω της διαδικασίας αυτής, καταλήξαμε ότι η τιμή κατωφλιού ορίζεται ως εξής:

$$cut - off = \frac{TC+NC}{2} \quad (3.1)$$

Ο Πίνακας 3.4 συνοψίζει τις τιμές κατωφλιού που προέκυψαν, καθώς και του γενικού profile G-alpha της Pfam.

Πίνακας 3.4 Οι τιμές κατωφλιού για τα έξι pHMMs που κατασκευάστηκαν και το ήδη υπάρχον γενικό για τις Gα της Pfam.

	Gα (PF00503)	G _s	G _{l/o}	G _{q/11}	G _{12/13}	Gβ	Gγ
cut-off	48.8	608.3	770.7	735.3	639	291.5	34.4



Εικόνα 3.5: Γραφική παράσταση τιμών ευαισθησίας και ειδικότητας για τα έξι rHMMs που κατασκευάστηκαν. Η μπλε γραμμή αναπαριστά την ειδικότητα (specificity), η κόκκινη την ευαισθησία (sensitivity) και η διακεκομμένη πράσινη το κατώφλι.

Παρατηρούμε ότι για όλα τα μοντέλα με εξαίρεση το G_β , στην τιμή κατωφλιού που ορίστηκε τόσο η ευαισθησία όσο και η ειδικότητα ισούνται με 1, ενώ και στην περίπτωση του G_β έχουμε $S_n=1$ και $S_p=0.99993$, γεγονός που αποδεικνύει την αξιοπιστία τους.

3.1.2.2 Εντοπισμός G-πρωτεϊνών σε πρωτεώματα

Στη συνέχεια έγινε προσπάθεια εντοπισμού G-πρωτεϊνών με τη χρήση της νέας μεθόδου στα πρωτεώματα των οργανισμών *Escherichia coli*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae*, *Drosophila melanogaster* και *Homo sapiens*.

Οι οργανισμοί αυτοί επιλέχθηκαν από εξελικτική σκοπιά, έτσι ώστε να καλυφθεί όλο το εύρος, από κατώτερα είδη μέχρι το ανώτερο, που είναι ο άνθρωπος, αλλά στην επιλογή

τους συντέλεσε και το γεγονός ότι είναι πρότυποι οργανισμοί για βιολογικές μελέτες. Τα πρωτεώματα των οργανισμών *E. coli*, *S. cerevisiae* και *H. sapiens* ανακτήθηκαν από τη βάση δεδομένων UniProt, ενώ για το *A. thaliana* από την ειδική για τον οργανισμό αυτό βάση TAIR (The Arabidopsis Information Resource - <http://www.arabidopsis.org/>) (Swarbreck, et al., 2008) και για τη *D. melanogaster* από την ειδική για είδη του γένους *Drosophila* βάση FlyBase (<http://flybase.org/>) (Marygold, et al., 2013). Ο Πίνακας 3.5 συνοψίζει τα αποτελέσματα.

Πίνακας 3.5 Εντοπισμός G-πρωτεϊνών σε πέντε οργανισμούς με τη μέθοδο GprotPRED.

	Μέγεθος πρωτέωμα τος	G _α (PF00503)	G _ς	G _{ι/ο}	G _{α/11}	G _{12/13}	G _β	G _γ
<i>Drosophila melanogaster</i>	23361	29 0.12%	4 0.02%	10 0.04%	11 0.05%	-	9 0.04%	12 0.05%
<i>Homo sapiens</i>	58732	40 0.07%	6 0.01%	13 0.02%	4 0.006%	6 0.01%	28 0.05%	17 0.03%
<i>Arabidopsis thaliana</i>	32350	6 0.02%	-	-	-	-	7 0.02%	-
<i>Saccharomyces cerevisiae</i>	6627	4 0.06%	-	-	-	-	2 0.03%	1 0.02%
<i>Escherichia coli</i>	4302	0	0	0	0	0	0	0

Παρατηρείται ότι το ποσοστό εμφάνισης των G-πρωτεϊνών στα πρωτεώματα είναι πολύ μικρό, όπως ήταν αναμενόμενο. Στην *E. coli* δεν εντοπίστηκαν G-πρωτεΐνες, ενώ στους οργανισμούς *S. cerevisiae* και *A. thaliana* εντοπίστηκαν G_α πρωτεΐνες, οι οποίες όμως δεν μπορούσαν να αντιστοιχιστούν στις 4 οικογένειες που αφορούν κυρίως τα θηλαστικά, αλλά και γενικά ανώτερους οργανισμούς. Επιπλέον, στο *S. cerevisiae* εντοπίστηκαν τόσο G_β όσο και G_γ υπομονάδες, ενώ στο *A. thaliana* μόνο G_β. Τόσο στην *D. melanogaster* όσο και στον *H. sapiens* εντοπίστηκαν G_β και G_γ υπομονάδες, καθώς και μέλη και των τεσσάρων οικογενειών G_α στον άνθρωπο, ενώ στη *Drosophila* δεν εντοπίστηκε G_{12/13}, η ύπαρξη της οποίας δεν είναι επιβεβαιωμένη. Επιπλέον, έγινε σύγκριση των αποτελεσμάτων που προέκυψαν με το σχολιασμό των αντίστοιχων πρωτεϊνών που προσφέρει το κάθε πρωτέωμα, και παρατηρήθηκε συμφωνία του σχολιασμού με τα αποτελέσματα των έξι ειδικών profiles. Αντίθετα, το rHMM της Pfam για την G_α υπομονάδα έχει την τάση να προβλέπει περισσότερες του αναμενόμενου πρωτεΐνες, οδηγώντας σε αρκετές εσφαλμένως θετικές προβλέψεις.

3.1.2.3 Η διαδικτυακή διεπαφή για τη μέθοδο GprotPRED

Η μέθοδος είναι διαθέσιμη στο διαδίκτυο μέσω της σελίδας του εργαστηρίου στη διεύθυνση: <http://bioinformatics.biol.uoa.gr/GprotPRED>. Ο χρήστης μπορεί να εισάγει τις προς μελέτη αμινοξικές ακολουθίες σε μορφή FASTA, στη διαθέσιμη φόρμα, ή να τις υποβάλει μέσω ενός αρχείου που να τις περιέχει στην ίδια μορφή (Εικόνα 3.6).

GprotPRED is an online tool that uses profile Hidden Markov Models (pHMMs) for the four known heterotrimeric Galpha protein families, the Gbeta and the Ggamma subunit in order to classify a set of protein sequences into the appropriate G-protein family. For more information click [here](#)

If you are using this tool, please cite the following reference: Kostiou, V.D., Theodoropoulou, M.C., Hamodrakas, S.J, Detection of the four distinct Galpha families, the Gbeta and the Ggamma subunits using profile Hidden Markov Models (pHMMs) [to be submitted]

Submission Form
Insert your sequence(s) in FASTA format below:

or upload a file (40 MB MAX) : No file chosen

If you have already submitted a query you can retrieve your results by inserting your job id here:

Εικόνα 3.6: Η ιστοσελίδα μέσω της οποίας εκτελείται διαδικτυακά η μέθοδος GprotPRED.

Τα έξι pHMMs που κατασκευάστηκαν, καθώς και το profile της Pfam G-alpha (PF00503), χρησιμοποιούνται για την αναζήτηση πιθανών G-πρωτεϊνών, στις προς μελέτη ακολουθίες. Τα αποτελέσματα επιστρέφονται στο χρήστη με μορφή πίνακα (Εικόνα 3.7), ενώ, επίσης, τα δεδομένα είναι διαθέσιμα είτε με τη μορφή καταλόγου αποτελεσμάτων, όπως αυτή προκύπτει από την αναζήτηση με το πρόγραμμα hmsearch του πακέτου HMMER v3.0 (Εικόνα 3.8), είτε με τη μορφή αρχείων fasta ανά κατηγορία, τα οποία περιλαμβάνουν τις αλληλουχίες των πιθανών G-πρωτεϊνών της αντίστοιχης κατηγορίας. Η μέθοδος GprotPRED εντοπίζει και κατατάσσει σε οικογένειες τις πιθανές G-πρωτεΐνες με μεγάλη ακρίβεια. Είναι γρήγορη και εύχρηστη μέθοδος και ευελπιστούμε να γίνει σημαντικό εργαλείο στο σχολιασμό νέων πρωτεωμάτων. Μελλοντικά, η μέθοδος μπορεί να συμπεριλάβει επιπλέον ειδικά pHMMs για ομάδες οργανισμών των οποίων οι Gα υπομονάδες δεν μπορούν να καταταχθούν στις τέσσερις οικογένειες, όπως για παράδειγμα οι μύκητες ή τα φυτά.

Process completed successfully

Your Job ID is: **58519926101755** (You can use this ID to access your results for 24 hours)[Click here](#) to download your results in text format

Number of predicted G-proteins per family						
Galpha	Gs	Gi/o	Gq/11	G12/13	Gbeta	Ggamma
2	3	3	1	7	9	

Predicted sequences belonging to Galpha family (by Pfam database model)

Download FASTA

Score	e-value	Sequence
-	-	-

Predicted sequences belonging to Gs family

Download FASTA

Score	e-value	Sequence
951.0	3.4e-289	spIP630921GNAS2_HUMAN Guanine nucleotide-binding protein Gs
949.5	9.6e-289	spIP630941GNAS2_MOUSE Guanine nucleotide-binding protein Gs

Predicted sequences belonging to Gio family

Download FASTA

Score	e-value	Sequence
959.2	1.4e-291	spIP630971IGNAI1_BOVIN Guanine nucleotide-binding protein Gi
923.8	7.5e-281	spIP384041IGNAO_LOCFI Guanine nucleotide-binding protein Gi
866.8	1.2e-263	spIP417761IGNAL_HOMAM Guanine nucleotide-binding protein Gi

Predicted sequences belonging to Gq11 family

Download FASTA

Score	e-value	Sequence
1060.9	5.1e-322	spIQ2PKF41IGNAQ_PIG Guanine nucleotide-binding protein Gq
1053.1	1.2e-319	spIP299921IGNAI1_HUMAN Guanine nucleotide-binding protein sub
1012.1	2.7e-307	spIP236251IGNAQ_DROME Guanine nucleotide-binding protein Gq

Predicted sequences belonging to G1213 family

Download FASTA

Score	e-value	Sequence
998.6	9.4e-303	spIQ63210IGNAI2_RAT Guanine nucleotide-binding protein subun

Predicted sequences belonging to Gbeta family

Download FASTA

Score	e-value	Sequence
821.2	1.5e-249	trIQ803H51Q803H5_DANRE Gnb1 protein OS=Danio rerio GN=gnb1b
817.9	1.6e-248	spIP62871IGBB1_BOVIN Guanine nucleotide-binding protein Gf
805.3	1e-244	spIP54313IGBB2_RAT Guanine nucleotide-binding protein Gf
805.3	1e-244	spIP628791IGBB2_HUMAN Guanine nucleotide-binding protein Gf
800.9	2.1e-243	spIP293871IGBB4_MOUSE Guanine nucleotide-binding protein su
790.9	2.2e-240	trIQ770Y91Q770Y9_XENLA Gnb1 protein OS=Xenopus laevis GN=gnb
785.8	7.6e-239	trIQ1RMY81Q1RMY8_HUMAN GNB1 protein (Fragment) OS=Homo sapie

Predicted sequences belonging to Ggamma family

Download FASTA

Score	e-value	Sequence
116.9	2e-32	spIP597681GBG2_HUMAN Guanine nucleotide-binding protein Gi
111.4	8.8e-31	spIQ5RBQ01GBG12_PONAB Guanine nucleotide-binding protein Gi
111.4	8.8e-31	spIQ9UBI61GBG12_HUMAN Guanine nucleotide-binding protein Gi
111.0	1.2e-30	spIQ9DAS91GBG12_MOUSE Guanine nucleotide-binding protein Gi
97.7	1.3e-26	spIP501511GBG10_HUMAN Guanine nucleotide-binding protein Gi
91.5	1e-24	spIP619531GBG11_MOUSE Guanine nucleotide-binding protein Gi
89.8	3.3e-24	spIP026981GBG1_BOVIN Guanine nucleotide-binding protein GT
88.4	8.7e-24	spIQ9P2W31GBG13_HUMAN Guanine nucleotide-binding protein Gi
83.1	3.6e-22	spIQ9JMF31GBG13_MOUSE Guanine nucleotide-binding protein Gi

[Submit another job](#)

Εικόνα 3.7: Μέρος της ιστοσελίδας των αποτελεσμάτων της μεθόδου GprotPRED.

```

List of predicted Gs proteins:

  E-value  Score  Sequence
  -----  -----  -----
3.4e-289  951.0  sp|P63092|GNAS2_HUMAN Guanine nucleotide-binding protein G(s Gs
9.6e-289  949.5  sp|P63094|GNAS2_MOUSE Guanine nucleotide-binding protein G(s Gs

2 hit(s).

*****
*****
List of predicted Gi/o proteins:

  E-value  Score  Sequence
  -----  -----  -----
1.4e-291  959.2  sp|P63097|GNAI1_BOVIN Guanine nucleotide-binding protein G(i Gio
7.5e-281  923.8  sp|P38404|GNAO_LOCOMI Guanine nucleotide-binding protein G(o Gio
1.2e-263  866.8  sp|P41776|GNAI_HOMAM  Guanine nucleotide-binding protein G(i Gio

3 hit(s).

*****
*****
List of predicted Gq/11 proteins:

  E-value  Score  Sequence
  -----  -----  -----
5.1e-322  1060.9 sp|Q2PKF4|GNAQ_PIG    Guanine nucleotide-binding protein G(q Gq11
1.2e-319  1053.1 sp|P29992|GNA11_HUMAN Guanine nucleotide-binding protein sub Gq11
2.7e-307  1012.1 sp|P23625|GNAQ_DROME  Guanine nucleotide-binding protein G(q Gq11

3 hit(s).

*****
*****
List of predicted G12/13 proteins:

  E-value  Score  Sequence
  -----  -----  -----
9.4e-303  998.6  sp|Q63210|GNA12_RAT  Guanine nucleotide-binding protein subun G1213

1 hit(s).

*****
*****
List of predicted Gbeta proteins:

  E-value  Score  Sequence
  -----  -----  -----
1.5e-249  821.2  tr|Q803H5|Q803H5_DANRE Gnb11 protein OS=Danio rerio GN=gnb1b Gbeta
1.6e-248  817.9  sp|P62871|GBB1_BOVIN  Guanine nucleotide-binding protein G( Gbeta
1e-244    805.3 sp|P54313|GBB2_RAT    Guanine nucleotide-binding protein G( Gbeta
1e-244    805.3 sp|P62879|GBB2_HUMAN  Guanine nucleotide-binding protein G( Gbeta
2.1e-243  800.9  sp|P29387|GBB4_MOUSE  Guanine nucleotide-binding protein su Gbeta
2.2e-240  790.9  tr|Q7T0Y9|Q7T0Y9_XENLA Gnb1 protein OS=Xenopus laevis GN=gmb Gbeta
7.6e-239  785.8  tr|Q1RMY8|Q1RMY8_HUMAN GNB1 protein (Fragment) OS=Homo sapie Gbeta

7 hit(s).

*****
*****
List of predicted Ggamma proteins:

  E-value  Score  Sequence
  -----  -----  -----
2e-32     116.9  sp|P59768|GBG2_HUMAN  Guanine nucleotide-binding protein G(I Ggamma
8.8e-31   111.4  sp|Q5RBQ0|GBG12_PONAB Guanine nucleotide-binding protein G(I Ggamma
8.8e-31   111.4  sp|Q9UBI6|GBG12_HUMAN Guanine nucleotide-binding protein G(I Ggamma
1.2e-30    111.0  sp|Q9DAS9|GBG12_MOUSE Guanine nucleotide-binding protein G(I Ggamma
1.3e-26    97.7   sp|P50151|GBG10_HUMAN Guanine nucleotide-binding protein G(I Ggamma
1e-24     91.5   sp|P61953|GBG11_MOUSE Guanine nucleotide-binding protein G(I Ggamma
3.3e-24    89.8   sp|P02698|GBG1_BOVIN  Guanine nucleotide-binding protein G(T Ggamma
8.7e-24    88.4   sp|Q9P2W3|GBG13_HUMAN Guanine nucleotide-binding protein G(I Ggamma
3.6e-22    83.1   sp|Q9JMF3|GBG13_MOUSE Guanine nucleotide-binding protein G(I Ggamma

9 hit(s).

*****
*****

```

Εικόνα 3.8: Ο κατάλογος των αποτελεσμάτων της μεθόδου GprotPRED, η οποία είναι διαθέσιμη σε μορφή αρχείου κειμένου.

3.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών

3.2.1 Η βάση δεδομένων Human-gpDB

Η Human-gpDB είναι μια δημόσια διαθέσιμη, βάση δεδομένων για τους ανθρώπινους GPCRs, G-πρωτεΐνες, εκτελεστές και τις αλληλεπιδράσεις μεταξύ τους. Η διαδικτυακή διεπαφή της είναι διαθέσιμη στις ηλεκτρονικές διευθύνσεις:

- http://bioinformatics.biol.uoa.gr/human_gpdb/ και
- http://schneider.embl.de/human_gpdb.

Περιλαμβάνει δεδομένα για 713 υποδοχείς, 36 G-πρωτεΐνες και 99 πρωτεΐνες-εκτελεστές, καθώς επίσης και πληροφορίες για 1663 αλληλεπιδράσεις μεταξύ GPCRs και G-πρωτεϊνών και 1618 αλληλεπιδράσεις μεταξύ G-πρωτεϊνών και εκτελεστών. Η αρχική σελίδα της διαδικτυακής διεπαφής φαίνεται στην Εικόνα 3.9.

Human-gpDB
A database of human GPCRs, G-proteins, Effectors and their interactions

home browse blast search advanced search help Quick search: search...

Introduction
Human-gpDB is a publicly accessible, relational database of human G-Proteins and their interactions with human GPCRs and Effectors. The database currently holds information about 713 human GPCRs, 36 human G-Proteins and 99 human Effectors. The collection of the information about the interactions between these molecules was done manually and the current status of Human-gpDB reveals information about 1663 connections between GPCRs and G-Proteins and 1618 connections between G-Proteins and Effectors. The sequences of G-Proteins and GPCRs are classified according to a hierarchy of different classes, families and sub-families, whereas the Effectors sequences are classified in families, subfamilies and types, based on extensive literature search. The classification of GPCRs follows the IUPHAR classification, while the Effectors classification is a unique feature and is based on their function. Advanced data integration techniques make Human-gpDB very rich in context since all of the bioentities are linked to a rich variety of external data sources. High quality visualization methods make the networks more informative and the extraction of information easier. Human-gpDB is currently a very useful tool for drug targeting investigation.

Statistics

- GPCRs : 713
- G-proteins : 36
- Effectors : 99
- GPCR-G-protein interactions : 1663
- Effectors-G-protein interactions : 1618

Send comments to Venkata P. Satagopam

Εικόνα 3.9: Η αρχική σελίδα της Human-gpDB στην οποία αναφέρονται τα στατιστικά στοιχεία της βάσης.

Ο χρήστης μπορεί να περιηγηθεί στη βάση μέσω της σελίδας browse (Εικόνα 3.10). Πληροφορίες για τις τρεις κατηγορίες μορίων της βάσης παρουσιάζονται, ακολουθώντας ιεραρχικά την ομαδοποίηση των πρωτεϊνών σε κατηγορίες, οικογένειες και υπο-οικογένειες. Ουσιαστικά, ο χρήστης μπορεί να ακολουθήσει το ιεραρχικό δέντρο χρησιμοποιώντας τα '+' και '-' και σε κάθε επίπεδο να αντλεί πληροφορίες. Στο κατώτερο επίπεδο της ιεραρχίας που περιέχει τις πρωτεΐνες, ο χρήστης βλέπει συνοπτικές πληροφορίες για το εν λόγω μόριο (όνομα, περιγραφή, όνομα γονιδίου) καθώς και συνδέσμους στην UniProt αλλά και στη σχετική σελίδα της εγγραφής (entry page), το περιληπτικό φύλλο (summary sheet) που προσφέρει όλες τις διαθέσιμες διασυνδέσεις με άλλες βάσεις δεδομένων και την ενότητα οπτικοποίησης (visualization module). Επίσης, σε οποιοδήποτε επίπεδο της ιεραρχίας, ο χρήστης μπορεί να ανακαλύψει περαιτέρω πληροφορίες μέσω ενός αναδυόμενου παραθύρου (popup window). Από οποιαδήποτε σελίδα της διεπαφής, προσφέρεται η δυνατότητα γρήγορης αναζήτησης μέσω της φόρμας που υπάρχει πάντα στο δεξί άνω άκρο των σελίδων (Quick search).

The figure consists of three screenshots of the Human-gpDB website interface, illustrating navigation and search capabilities.

Top Screenshot: Shows the main header with the Human-gpDB logo and title "Human-gpDB A database of human GPCRs, G-proteins, Effectors and their interactions". Navigation tabs include "home", "browse", "blast search", "advanced search", and "help". A "Quick search" field is present.

Middle Screenshot: Shows a "Description" section with the text "Click the following categories for further exploration". Below this, a "Categories" section lists "GPCRs", "G-proteins", and "Effectors", with a blue bracket indicating "3 categories of molecules contained in Human-gpDB".

Bottom Screenshot: Shows a "Legend" section with instructions: "E: Hyperlink to Entry-Page" and "S: Hyperlink to Summary-Sheet". It includes "Expand All" and "Collapse All" buttons. A tree structure of "Human-gpDB" is shown, with a blue bracket indicating the "Tree structure of Human-gpDB". A table lists entries under "GPCRs" and "GPRC6".

Name	Description	Gene Names	UniProt Acc	URLs
CALCIUM-SENSING RECEPTORS_CASR_Homo sapiens	RecName: Full=Extracellular calcium-sensing receptor; Short=CaSR; AltName: Full=Parathyroid Cell calcium-sensing receptor; AltName: Full=PcAr1; Flags: Precursor	CASR; GPRC2A; PCAR1	P41180	E S

A red circle highlights the "Quick search" field in the bottom screenshot. A text box next to it states: "Quick search using auto completing forms based on the name, the IDs, the gene names and the cross-references provided by our database".

A pop-up window for "Galpha-11" is shown, listing options: "Knowledge from biological databases", "Domain architecture", "Pathway information", "P-P, P-C interactions", "Visualization of GPCR, G-protein, Effector interaction in Medusa", and "Arena3D visualization". A text box next to it states: "Pop-up window with all the available options. The pop-up window appears by left clicking the respective level."

Εικόνα 3.10: Η σελίδα περιήγησης στην Human-gpDB και οι διάφορες διαθέσιμες προς το χρήστη δυνατότητες.

Επιπλέον, η Human-gpDB προσφέρει τη δυνατότητα αναζήτησης στη βάση, με τη χρήση του BLAST (Lopez, et al., 2003) (Εικόνα 3.11). Ο χρήστης παρέχει μία ή περισσότερες αλληλουχίες σε Fasta μορφή και, με τη χρήση του εργαλείου BLAST, παίρνει ως αποτέλεσμα ομόλογες αλληλουχίες που περιέχονται στη βάση μας.

Human-gpDB
A database of human GPCRs, G-proteins, Effectors and their interactions

home browse blast search **advanced search** help Quick search: search...

Blast Search

Enter sequence(s) (fasta format):

Or upload fasta sequence(s) file:

Search target: G-protein : GPCR : Effector :

Εικόνα 3.11: Η σελίδα αναζήτησης μέσω BLAST της Human-gpDB.

Τέλος, στη σελίδα προηγμένης αναζήτησης (advanced search) προσφέρεται η δυνατότητα αναζήτησης έναντι της βάσης, με τη χρήση συγκεκριμένων πεδίων αυτής (Εικόνα 3.12).

Human-gpDB
A database of human GPCRs, G-proteins, Effectors and their interactions

home browse blast search **advanced search** help Quick search: search...

Advanced Search

Gene/Protein:

Class:

Family:

Sub-family:

Type:

Description:

Function:

Search target: G-protein : GPCR : Effector :

The fields by which the user may perform a query against Human-gpDB

Selection of the type of molecule against which the query will be performed

Εικόνα 3.12: Η σελίδα προηγμένης αναζήτησης της Human-gpDB.

3.2.2 Η βάση δεδομένων PLHG-DB

Η PLHG-DB είναι μία δημόσια διαθέσιμη βάση δεδομένων για τους πεπτιδικούς υποκαταστάτες των ανθρώπινων GPCRs. Η διαδικτυακή διεπαφή της βάσης βρίσκεται στη σελίδα του εργαστηρίου μας: http://bioinformatics.biol.uoa.gr/plhg_db/. Η βάση περιλαμβάνει δεδομένα για 207 πεπτιδικούς υποκαταστάτες, για 128 ανθρώπινους GPCRs. Οι πεπτιδικοί υποκαταστάτες κατηγοριοποιούνται με βάση τη λειτουργία τους, όπως αυτή προέκυψε από την εκτενή βιβλιογραφική αναζήτηση, σε οικογένειες υπο-οικογένειες και τύπους. Συγκεκριμένα, οι πεπτιδικοί υποκαταστάτες κατηγοριοποιούνται σε 48 οικογένειες, οι οποίες χωρίζονται περαιτέρω σε 128 υπο-οικογένειες, που περιλαμβάνουν τους 207 τύπους, οι οποίοι ουσιαστικά αποτελούν αυτούς καθαυτούς τους πεπτιδικούς υποκαταστάτες που περιλαμβάνονται στη βάση.

Οι υποκαταστάτες αυτοί αλληλεπιδρούν με 128 διαφορετικές υπο-οικογένειες GPCRs, οι οποίες είναι μέλη 41 οικογενειών των τριών από τις τέσσερις κατηγορίες υποδοχέων του ανθρώπου. Στη βάση περιλαμβάνονται συνολικά 504 αλληλεπιδράσεις GPCR-πεπτιδικών υποκαταστατών (Εικόνα 3.13).

PLHG-DB:
a DataBase of Peptide Ligands of human G-protein coupled receptors

Introduction Navigation Text Search Blast Search Manual

PLHG-DB is a relational database of the Peptide Ligands that interact with human GPCRs. The sequences of the Peptide Ligands are classified according to a hierarchy of different families, subfamilies and types, based on extensive literature search.

The application provides full sequence information of the Peptide Ligands with cross-references to the publicly available databases, Uniprot and Pubmed. There are also displayed the interactions between the Peptide Ligands and the GPCRs with cross-references to Pubmed, that support the specific interactions.

The user also has the possibility to use extra tools. These are the Text Search tool and the Blast Search tool. Through the Text Search tool the user is able to submit advanced queries for text search and by using the Blast Search tool the BLAST program is run against the database. In addition, an analytical manual is provided that offers the user full guidance about the use of the application and important information about the current database.

Peptide Ligand Families: 48
Peptide Ligand Subfamilies: 128
Peptide Ligand Types: 207

GPCR Classes: 3
GPCR Families: 41
GPCR Subfamilies: 128

Peptide Ligand(s) - GPCR(s) Interactions: 504

Εικόνα 3.13: Η αρχική σελίδα της PLHG-DB στην οποία αναφέρονται και τα στατιστικά στοιχεία της βάσης.

Η διαδικτυακή διεπαφή της PLHG-DB δίνει τη δυνατότητα στο χρήστη να περιηγηθεί στα δεδομένα της βάσης για τους πεπτιδικούς υποκαταστάτες, είτε μέσω της ιεραρχίας των πεπτιδικών υποκαταστατών, είτε μέσω της ιεραρχίας των υποδοχέων με τους οποίους αυτοί αλληλεπιδρούν (Εικόνα 3.14). Αξίζει να σημειωθεί ότι στη βάση δεν περιέχονται δεδομένα για τους υποδοχείς, όμως δίνεται για κάθε υποδοχέα σύνδεσμος στη βάση του

εργαστηρίου μας Human-grpDB, στην οποία περιλαμβάνονται πληθώρα πληροφοριών, όπως περιγράφηκε προηγούμενα.

The screenshot displays the PLHG-DB website interface. At the top, the logo 'PLHG-DB' is shown next to the text 'PLHG-DB: a DataBase of Peptide Ligands of human G-protein coupled receptors'. Below this is a navigation bar with buttons for 'Introduction', 'Navigation', 'Text Search', 'Blast Search', and 'Manual'. The main content area is divided into two sections: 'GPCRs Classification' and 'Peptide Ligands Classification'. The 'GPCRs Classification' section lists 'CLASS A', 'CLASS B', and 'FRIZZLED / SMOOTHENED', with a red arrow pointing to 'GPCR Classes'. The 'Peptide Ligands Classification' section lists numerous peptide families, including 'ADIPOKINES', 'AGOUTI', 'AMYLOIDOGENIC PEPTIDES', 'ANAPHYLATOXINS', 'ANGIOTENSINS', 'ANNEXINS', 'APELIN', 'BOMBESIN-LIKE PEPTIDES', 'CALCITONIN', 'CATHELICIDINS', 'CATHESPINS', 'CHEMOKINES', 'CHOLECYSTOKININ / GASTRIN', 'CORTICOTROPIN-RELEASING FACTOR', 'ENDOTHELINS', 'F2L PEPTIDE', 'GALANIN', 'GLUCAGON / SECRETIN', 'GLYCOPROTEIN HORMONES', 'GONADOTROPIN-RELEASING HORMONE', 'HEDGEHOG', 'HUMANIN', 'KININS', 'MELANIN-CONCENTRATING HORMONE', 'MELANOCORTINS', 'MOTILIN / GHRELIN', 'NEUROMEDIN S (NMS)', 'NEUROMEDIN U (NMU)', 'NEUROPEPTIDE S (NPS)', 'NEUROPEPTIDE Y PEPTIDES (NPY PEPTIDES)', 'NEUROPEPTIDES B/W', 'NEUROTENSIN (NT)', 'NORRIN', 'OPIOIDS', 'OREXINS', 'PARATHYROID HORMONES', 'PROTEINASE-ACTIVATED RECEPTORS PEPTIDES (PAR_PEPTIDES)', 'R-SPONDINS', 'RELAXINS', 'RFamide PEPTIDES', 'SOMATOSTATIN', 'TACHYKININS', 'TEMPORINS', 'THYRO TROPIN-RELEASING HORMONE (TRH)', 'UROKINASE PLASMINOGEN ACTIVATOR SURFACE RECEPTOR (U-PAR)', 'UROTENSIN II', 'VASOPRESSIN / OXYTOCIN', and 'WNTS'. A red arrow points from this list to 'Peptide Ligand Families'. At the bottom, there is contact information for the University of Athens, Faculty of Biology, Dept. of Cell Biology and Biophysics, and the Biophysics & Bioinformatics Laboratory.

Εικόνα 3.14: Η σελίδα περιήγησης της PLHG-DB. Η περιήγηση μπορεί να γίνει ακολουθώντας είτε την ιεραρχία των υποδοχέων είτε αυτή των υποκαταστατών.

Ακολουθώντας οποιαδήποτε από τις δύο διαδρομές, ο χρήστης οδηγείται στην εγγραφή για κάθε πεπτιδικό υποκαταστάτη (Εικόνα 3.15). Εδώ παρέχονται στο χρήστη μια πληθώρα πληροφοριών, από τις οποίες, άλλες προέκυψαν με αυτοματοποιημένο τρόπο από τις εγγραφές τις UniProt/Swiss-Prot, και άλλες προέρχονται από την εκτενή αναζήτηση στη διεθνή βιβλιογραφία.

PLHG-DB:
a DataBase of Peptide Ligands of human G-protein coupled receptors

Introduction Navigation Text Search Blast Search Manual

Entry 1 of 1

PEPTIDE LIGAND CLASSIFICATION	
FAMILY:	MELANOCORTINS
SUBFAMILY:	MELANOCYTE-STIMULATING HORMONES
TYPE:	Alpha-MELANOCYTE-STIMULATING HORMONE (Alpha-MSH)

PEPTIDE LIGAND INFORMATION			
NAME:	Alpha-MELANOCYTE-STIMULATING HORMONE (Alpha-MSH)		
DESCRIPTION:	Melanotropin alpha Alpha-MSH		
SPECIES: Homo sapiens	COMMON NAME:	Human	NCBI TAXONOMY: 9606
SEQUENCE:	SYSMHFRRGKPV		LENGTH: 13
BEGIN: 138	END: 150		
COMMENT:	Alpha-Melanocyte-Stimulating Hormone (Alpha-MSH) belongs to the family of Melanocortins (MC) that also consists of Beta-Melanocyte-Stimulating Hormone (Beta-MSH), Gamma-MSH, Adrenocorticotrophic Hormone (ACTH), Corticotropin-like Intermediate Lobe Peptide (CLIP), Beta-lipotropin and Beta-endorphin. All Melanocortins are generated by post-translational modification of the 267 amino acid POMC prohormone by proteolytic cleavage. Alpha-MSH is a 13 amino acid peptide cleaved from ACTH by Proconvertase 2 (PC2). Alpha-MSH is further translationally modified by amidation of the C-terminus and acetylation of the N-terminus. Moreover all Melanocortins have a conserved sequence His-Phe-Arg-Trp. Melanocortins are expressed throughout the brain as well in peripheral tissues including testis, ovary, placenta, duodenum, liver, kidney, thymus, lymphocytes, macrophages, and skin. They are best known for their ability to stimulate eumelanin synthesis in mammalian melanocytes and steroidogenesis in adrenocortical cells. Other effects include regulation of food intake and energy metabolism, and neuronal regeneration. Melanocortins act as endogenous antipyretic agents, and have systemic as well as peripheral antiinflammatory effects. Melanocortins have sebotropic effects, induce lipolysis, regulate exocrine glands, cardiac and testicular functions and natriuresis.		
REFERENCES:	14535656		

PROTEIN INFORMATION	
UNIPROT:	P01169 P78442 Q9UD39 Q9UD40
PROTEIN DESCRIPTION:	Pro-opiomelanocortin (POMC) Corticotropin-lipotropin
GENE:	POMC

PEPTIDE LIGAND(s) - GPCR(s) INTERACTIONS		
GPCR(s) THAT BIND:		
Nº of Interaction	GPCR Subfamily:	GPCR(s) References:
1.	MC1	14535656 11315190 17057195
2.	MC3	14535656 11315190 17057195
3.	MC4	14535656 11315190 17057195
4.	MC5	14535656 11315190 17057195

PEPTIDE LIGAND(s) - GPCR(s) BINDING INFORMATION	
1.	Melanocortin (MC) Receptors are receptors for peptides of the family of Melanocortins. These peptides are the Adrenocorticotrophic Hormone (ACTH), Alpha-Melanocyte-Stimulating Hormone (Alpha-MSH), Beta-MSH and Gamma-MSH. They are also receptors for Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP). Melanocortin Receptors comprise 5 types of receptors: MC1, MC2, MC3, MC4 and MC5. Melanocortins, Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP) show different affinities for each receptor type. Based on these observations, each receptor is shown to interact with specific ligands. Therefore, MC1 interacts with Alpha-MSH, ACTH, Beta-MSH, Gamma-MSH, and ASP. MC2 interacts with ACTH, MC3 interacts with Alpha-MSH, Beta-MSH, Gamma-MSH and AGRP. MC4 interacts with Alpha-MSH, ACTH, Beta-MSH, ASP and AGRP, and finally MC5 interacts with Alpha-MSH, ACTH and Beta-MSH.
2.	Melanocortin (MC) Receptors are receptors for peptides of the family of Melanocortins. These peptides are the Adrenocorticotrophic Hormone (ACTH), Alpha-Melanocyte-Stimulating Hormone (Alpha-MSH), Beta-MSH and Gamma-MSH. They are also receptors for Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP). Melanocortin Receptors comprise 5 types of receptors: MC1, MC2, MC3, MC4 and MC5. Melanocortins, Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP) show different affinities for each receptor type. Based on these observations, each receptor is shown to interact with specific ligands. Therefore, MC1 interacts with Alpha-MSH, ACTH, Beta-MSH, Gamma-MSH, and ASP. MC2 interacts with ACTH, MC3 interacts with Alpha-MSH, Beta-MSH, Gamma-MSH and AGRP. MC4 interacts with Alpha-MSH, ACTH, Beta-MSH, ASP and AGRP, and finally MC5 interacts with Alpha-MSH, ACTH and Beta-MSH.
3.	Melanocortin (MC) Receptors are receptors for peptides of the family of Melanocortins. These peptides are the Adrenocorticotrophic Hormone (ACTH), Alpha-Melanocyte-Stimulating Hormone (Alpha-MSH), Beta-MSH and Gamma-MSH. They are also receptors for Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP). Melanocortin Receptors comprise 5 types of receptors: MC1, MC2, MC3, MC4 and MC5. Melanocortins, Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP) show different affinities for each receptor type. Based on these observations, each receptor is shown to interact with specific ligands. Therefore, MC1 interacts with Alpha-MSH, ACTH, Beta-MSH, Gamma-MSH, and ASP. MC2 interacts with ACTH, MC3 interacts with Alpha-MSH, Beta-MSH, Gamma-MSH and AGRP. MC4 interacts with Alpha-MSH, ACTH, Beta-MSH, ASP and AGRP, and finally MC5 interacts with Alpha-MSH, ACTH and Beta-MSH.
4.	Melanocortin (MC) Receptors are receptors for peptides of the family of Melanocortins. These peptides are the Adrenocorticotrophic Hormone (ACTH), Alpha-Melanocyte-Stimulating Hormone (Alpha-MSH), Beta-MSH and Gamma-MSH. They are also receptors for Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP). Melanocortin Receptors comprise 5 types of receptors: MC1, MC2, MC3, MC4 and MC5. Melanocortins, Agouti-Signaling Protein (ASP) and Agouti-Related Protein (AGRP) show different affinities for each receptor type. Based on these observations, each receptor is shown to interact with specific ligands. Therefore, MC1 interacts with Alpha-MSH, ACTH, Beta-MSH, Gamma-MSH, and ASP. MC2 interacts with ACTH, MC3 interacts with Alpha-MSH, Beta-MSH, Gamma-MSH and AGRP. MC4 interacts with Alpha-MSH, ACTH, Beta-MSH, ASP and AGRP, and finally MC5 interacts with Alpha-MSH, ACTH and Beta-MSH.

Contact @ bioinformatics@ua.gr for comments, corrections and data (sequence) submission.

University of Athens
Faculty of Biology
Dept. of Cell Biology and Biophysics

Biophysics
& Bioinformatics
Laboratory

Εικόνα 3.15: Παράδειγμα εγγραφής για έναν πεπτιδικό υποκαταστάτη στην PLHG-DB.

Οι πληροφορίες που περιλαμβάνονται στην εγγραφή κάθε πεπτιδικού υποκαταστάτη είναι:

- Η κατάταξη του πεπτιδικού υποκαταστάτη με βάση τη λειτουργία του

- Πληροφορίες που αφορούν τον πεπτιδικό υποκαταστάτη, όπως το όνομα του, η αλληλουχία (ελεγμένη προσεκτικά και σύμφωνη με τα βιβλιογραφικά δεδομένα), το μήκος του πεπτιδίου, οι θέσεις της αλληλουχίας του προπεπτιδίου από όπου προέρχεται και, τέλος, ένα σχόλιο συνοδευόμενο από αντίστοιχη βιβλιογραφική αναφορά. Το σχόλιο αυτό έχει προκύψει από τη μελέτη των επιστημονικών άρθρων που συγκεντρώθηκαν.
- Συνοπτικές πληροφορίες για το προπεπτίδιο
- Οι πληροφορίες που αφορούν την αλληλεπίδραση με συγκεκριμένους υποδοχείς συνοδευόμενες από την αντίστοιχη βιβλιογραφία και από ένα σχόλιο που έχει προκύψει από τη μελέτη της.

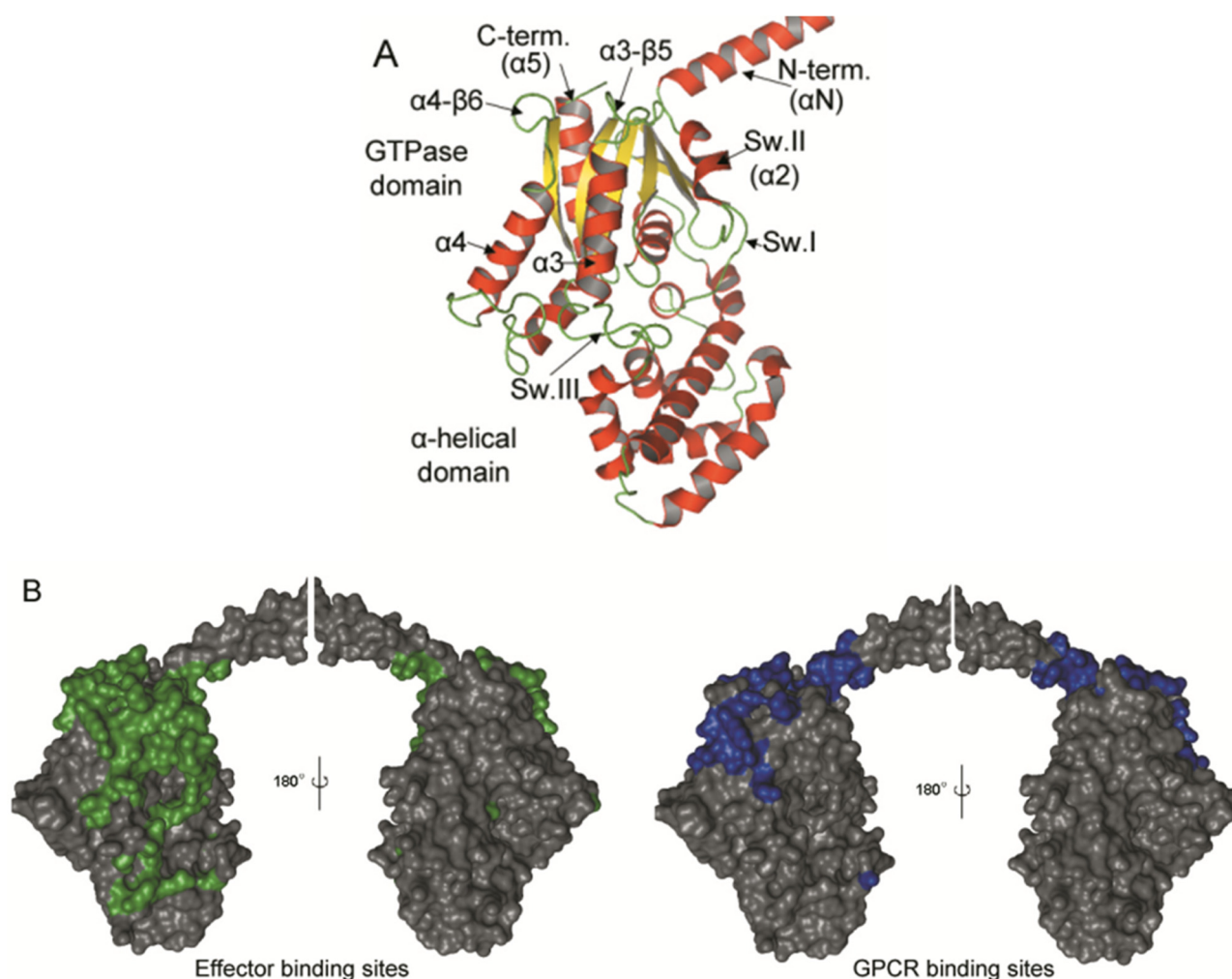
Τέλος, προσφέρεται στο χρήστη η δυνατότητα αναζήτησης στην PLHG-DB, είτε μέσω BLAST είτε μέσω προηγμένης αναζήτησης, χρησιμοποιώντας συγκεκριμένα πεδία της βάσης.

3.3 Μελέτη των αλληλεπιδράσεων των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες

3.3.1 Αλληλεπικαλυπτόμενες περιοχές αλληλεπιδράσεων στην επιφάνεια των Gα

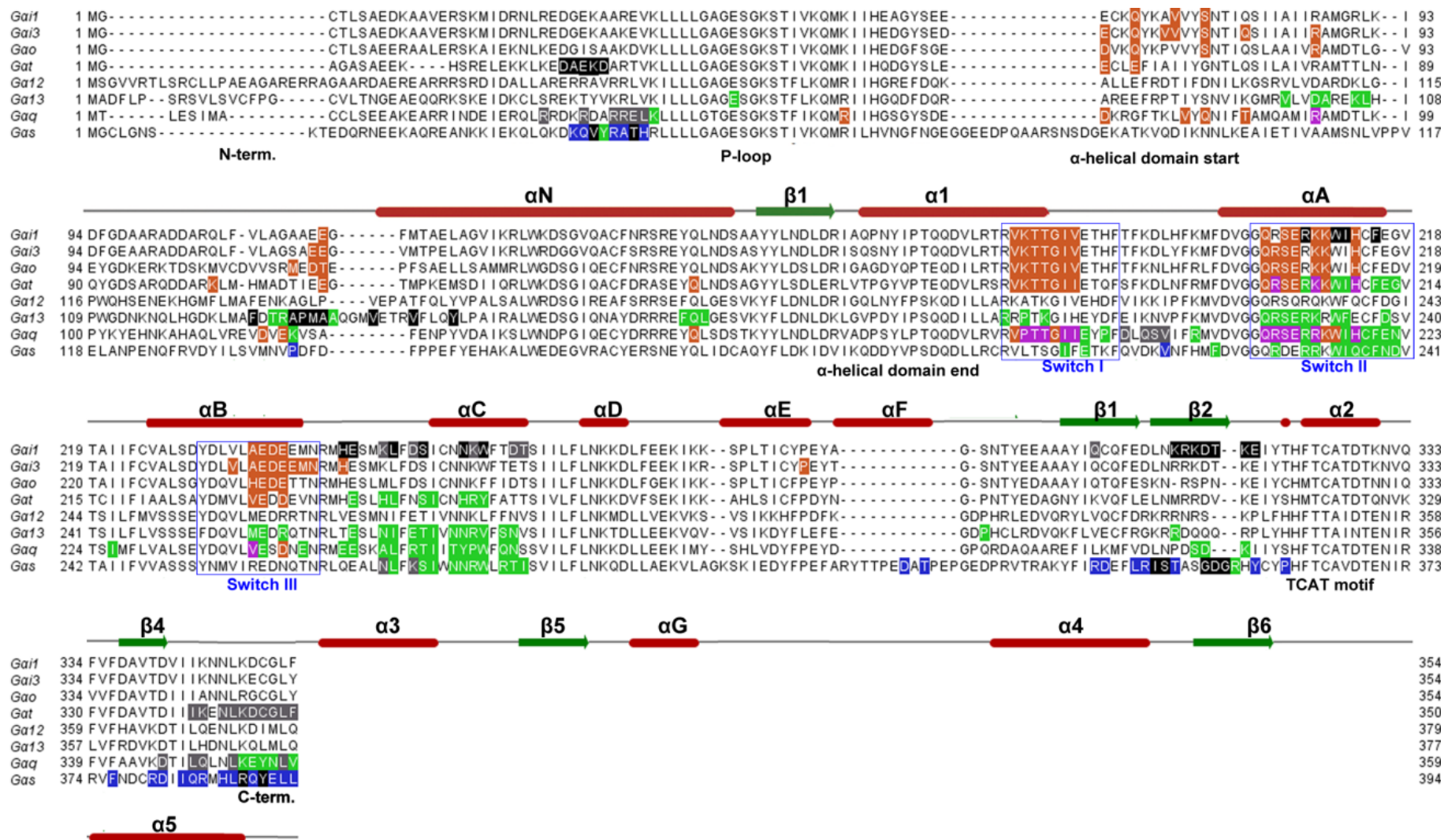
Από τη μελέτη των κρυσταλλικών δομών των συμπλόκων των G-πρωτεϊνών, σε συνδυασμό με βιβλιογραφικά στοιχεία προερχόμενα από πειράματα μεταλλαξογένεσης και υπολογιστικές μελέτες, αναγνωρίσαμε στην επιφάνεια των Gα υπομονάδων μια σειρά από περιοχές και κατάλοιπα που μπορούν σε πολλές περιπτώσεις να αλληλεπιδράσουν τόσο με GPCRs όσο και με εκτελεστές (Εικόνα 3.16, Εικόνα 3.17, Πίνακας 3.6).

Σε αυτές τις περιοχές περιλαμβάνονται τόσο το αμινοτελικό όσο και το καρβοξυτελικό άκρο, καθώς και ο βρόχος α4-β6. Ο βρόχος α3-β5, που αποτελεί κοινή επιφάνεια αλληλεπίδρασης με τα μόρια εκτελεστές, φαίνεται να εμπλέκεται στη ρύθμιση της ενεργοποίησης της Gα υπομονάδας, αλλά και στην πρόσδεση με τον υποδοχέα, από διάφορες βιοχημικές μελέτες και μελέτες μεταλλαξογένεσης, και για αυτό το λόγο μπορεί να θεωρηθεί περιοχή αλληλεπίδρασης με υποδοχείς, παρόλο που δεν υπάρχουν επαρκή δομικά δεδομένα που να υποστηρίζουν την υπόθεση αυτή.



Εικόνα 3.16. A. Δομικά χαρακτηριστικά των $G\alpha$ υπομονάδων. **B.** Οι θέσεις αλληλεπίδρασης εκτελεστών και GPCRs πάνω στην επιφάνεια της $G\alpha$, σύμφωνα με δεδομένα από κρυσταλλικές δομές, πειράματα μεταλλαξογένεσης και υπολογιστικές μελέτες. Οι επιφάνειες επαφής χρωματίζονται πράσινες για τους εκτελεστές και μπλε για τους GPCRs. Η δομή που χρησιμοποιείται είναι η $G\alpha_{i1}$ (PDB ID: 1GP2).

Πρέπει να σημειωθεί ότι συγκεκριμένα αμινοξικά κατάλοιπα του καρβοξυτελικού άκρου και του βρόχου $\alpha 4\text{-}\beta 6$, και, ίσως, του αμινοτελικού άκρου και του βρόχου $\alpha 3\text{-}\beta 5$, εμπλέκονται σε αλληλεπιδράσεις τόσο με μόρια εκτελεστές όσο και με GPCRs. Τέλος, συγκεκριμένες περιοχές της αυτοτελούς περιοχής α -έλικας, συγκεκριμένα η αA έλικα και ο $\alpha B\text{-}\alpha C$ βρόχος, έχει δειχθεί ότι αλληλεπιδρούν με αρκετά μόρια εκτελεστές και RGS πρωτεΐνες, και στο σύμπλοκο $\beta_2AR - G\alpha_s G\beta\gamma$ επίσης φαίνεται να παίζουν ρόλο στην αλληλεπίδραση υποδοχέα- $G\alpha$ υπομονάδας.



Εικόνα 3.17. Στοιχισμός αλληλουχιών των Ga υπομονάδων με προσδιορισμένες κρυσταλλογραφικά δομές στην PDB. Η δευτεροταγής δομή βασίζεται στη δομή της $G\alpha_1$ (PDBID: 1GP2) και απεικονίζεται με κόκκινους κυλίνδρους για τις α -έλικες και πράσινα βέλη για τους β -κλώνους. Οι τρεις περιοχές «διακόπτες» (switches) σημαίνονται με κυανά πλαίσια. Τα κατάλοπα των Ga που συμμετέχουν σε αλληλεπιδράσεις, σύμφωνα με τις δομές των συμπλόκων, χρωματίζονται πράσινα για εκτελεστές, πορτοκαλί για RGS, μωβ και για τους δύο και μπλε για GPCRs. Επιπλέον αλληλεπιδράσεις που προτείνονται από δομές συμπλόκων υποδοχέων ή G-πρωτεϊνών με πεπτιδία χρωματίζονται μαύρα και γκριζα για εκτελεστές και GPCRs, αντίστοιχα.

Πίνακας 3.6 Οι θέσεις αλληλεπίδρασης του β_2 αδρενεργικού υποδοχέα και των εκτελεστών πάνω στην επιφάνεια των $G\alpha$, με βάση τις προσδιορισμένες δομές της PDB. Τα σύμβολα (+) και (-) υποδεικνύουν ύπαρξη ή απουσία αλληλεπίδρασης, αντίστοιχα. Παρατίθενται σημαντικά τμήματα της αυτοτελούς περιοχής GTPάσης των $G\alpha$. Άλλες θέσεις αλληλεπίδρασης τοποθετούνται στην κατηγορία «Άλλες». Στην περίπτωση του p115RhoGEF παρατίθενται δεδομένα και από τις δύο δομές, καθώς παρουσιάζονται διαφορές στις αλληλεπιδράσεις του εκτελεστή με τη χιμαιρική $G\alpha_{13/i}$ και με τη φυσικού τύπου $G\alpha_{13}$.

Δομές	Θέσεις Αλληλεπίδρασης στις $G\alpha$											
	Σύμπλοκο GPCR	$G\alpha$	PDB	N-άκρο.	Sw. I	Sw. II	Sw. III	$\alpha 3$	$\alpha 3$ - $\beta 5$	$\alpha 4$ - $\beta 6$	C-άκρο.	Άλλες
												+
Gαβγ-β ₂ AR	$G\alpha_s$	3SN6	+	-	-	-	-	-	+	+		(β2-β3, αG-α4, περιοχή α-έλικας)
Σύμπλοκα εκτελεστών												
												+
$G\alpha_s$ -Αδ. Κυκλ.	$G\alpha_s$	1AZT κλπ.	+	+	+	-	+	+	-	-		(περιοχή α-έλικας)
PDEγ-Gα _{t/i} -RGS9	$G\alpha_{t/i}$	1FQJ	-	-	+	-	+	+	-	-		-
$G\alpha_{13/i}$ -p115RhoGEF rgRGS	$G\alpha_{13/i}$	1SHZ	-	+	+	+	+	+	-	-		(περιοχή α-έλικας)
$G\alpha_q$ -GRK2-Gβγ	$G\alpha_q$	2BCJ	-	-	+	-	+	+	-	-		-
$G\alpha_q$ -p63RhoGEF - RhoA	$G\alpha_q$	2RGN	-	-	+	-	+	+	+	+		-
$G\alpha_{13}$ -PDZRhoGEF rgRGS	$G\alpha_{13}$	3CX8 κλπ.	-	+	+	+	+	+	+	-		(περιοχή α-έλικας)
$G\alpha_q$ -PLCβ3	$G\alpha_q$	3OHM	+	+	+	+	+	+	+	-		(περιοχή α-έλικας)
$G\alpha_{13}$ -p115RhoGEF rgRGS	$G\alpha_{13}$	1SHZ	-	+	+	+	+	+	+	-		(περιοχή α-έλικας)
$G\alpha_{i1}$ -KB-1753	$G\alpha_{i1}$	2G83	-	-	+	-	+	+	-	-		-

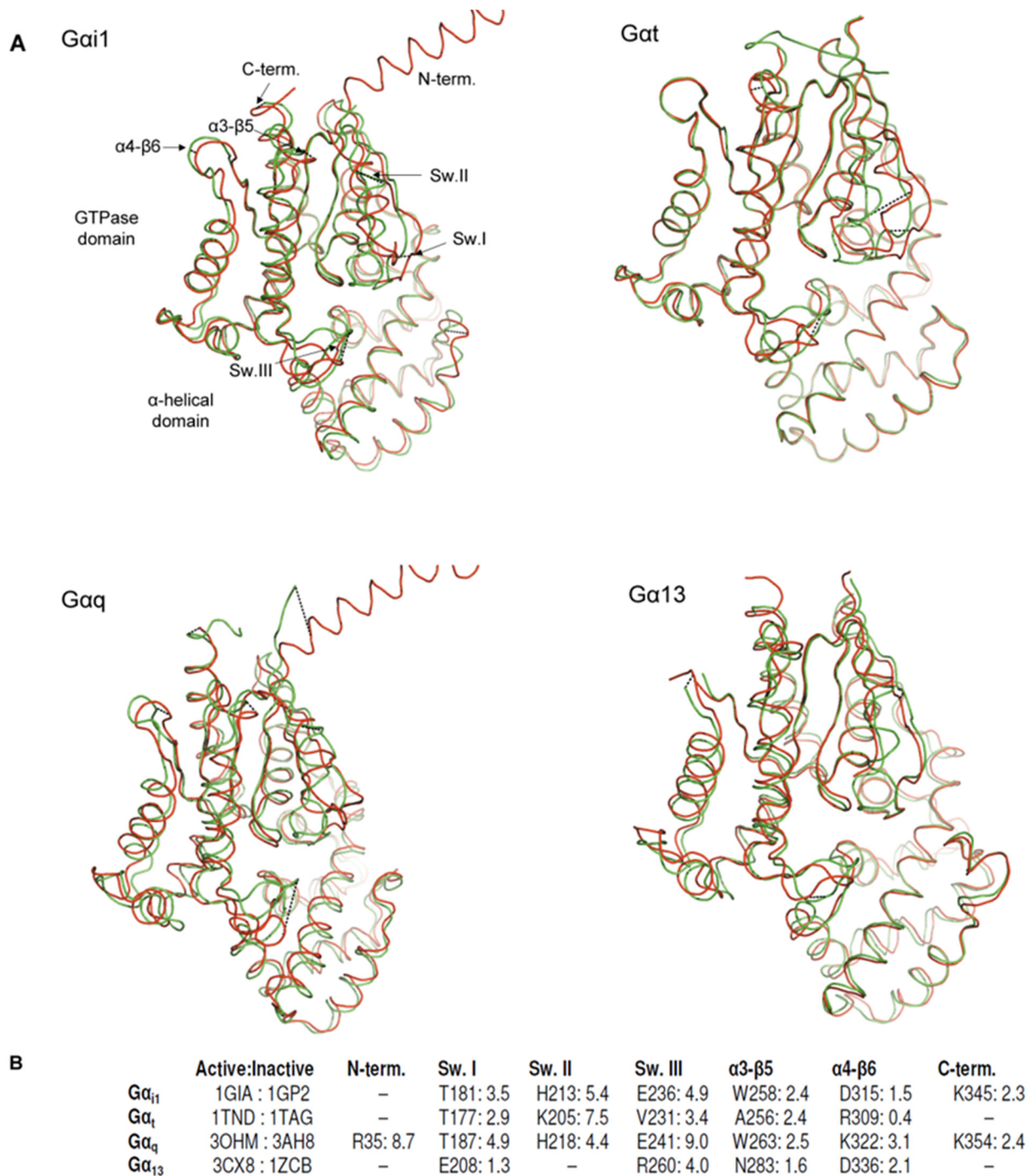
3.3.2 Μεταβολές στη δομή κατά την ενεργοποίηση των Gα

Η σύγκριση ενεργών και ανενεργών Gα υπομονάδων με τις μεθόδους που αναφέρθηκαν δείχνει πως οι συγκεκριμένες περιοχές διαφοροποιούνται κατά την ενεργοποίηση των G-πρωτεϊνών, γεγονός που θα μπορούσε να επηρεάζει το ρόλο τους στις αλληλεπιδράσεις των Gα. Κατά την ενεργοποίησή τους παρατηρούνται μια σειρά από αλλαγές στη στερεοδιάταξη των Gα υπομονάδων. Η πρόσδεση του ετεροτριμερούς των G-πρωτεϊνών στην κυτταροπλασματική «πλευρά» του ενεργοποιημένου υποδοχέα οδηγεί στο άνοιγμα της κοιλότητας που περιέχει το νουκλεοτίδιο για την ανταλλαγή του GDP σε GTP, μέσω της εκτεταμένης μετακίνησης της αυτοτελούς περιοχής α-έλικας (Rasmussen, et al., 2011). Η ενεργοποίηση της Gα όμως οδηγεί και σε ανακατατάξεις συγκεκριμένων τμημάτων της περιοχής GTPάσης, και ιδίως των περιοχών που συμμετέχουν στις αλληλεπιδράσεις των Gα. Οι συγκεκριμένες ανακατατάξεις γίνονται ορατές με τη σύγκριση ενεργών και ανενεργών υπομονάδων από τις τέσσερις οικογένειες των Gα (Εικόνα 3.18, Πίνακας 3.7).

Πίνακας 3.7 Τιμές RMSD από τη σύγκριση ενεργών και ανενεργών Gα υπομονάδων. Δίνονται οι κωδικοί των δομών στην PDB, καθώς και οι κωδικοί των αλυσίδων που συγκρίθηκαν σε παρένθεση.

Gα	PDB	RMSD (Å)
Gα _{i1}	1GP2 (A) – 1GIA (A)	1.5
Gα _t	1TAG (A) – 1TND (A)	1.1
Gα _q	3AH8 (A) – 3OHM (A)	1.9
Gα ₁₃	1ZCB (A) – 3CX8 (A)	1.4
Gα _s GTPase	3SN6 (A) – 1AZT (A) GTPase regions	2.1

Πρέπει να σημειωθεί πως οι περισσότερες κρυσταλλικές δομές των Gα δεν διαθέτουν συντεταγμένες για το μεγαλύτερο μέρος του αμινοτελικού άκρου, κάτι που εμποδίζει την πλήρη παρατήρηση της συμπεριφοράς του κατά την ενεργοποίηση της υπομονάδας αλλά και την αλληλεπίδραση με υποδοχείς ή εκτελεστές. Το συγκεκριμένο γεγονός οδηγεί πιθανώς και σε σχετικά χαμηλές τιμές RMSD κατά τη σύγκριση δομών Gα μέσω υπερθέσεων.



Εικόνα 3.18: **A.** Σύγκριση ενεργών και ανενεργών Gα υπομονάδων μέσω δομικών στοιχίσεων. Οι ενεργές υπομονάδες χρωματίζονται πράσινες και οι ανενεργές κόκκινες. Σημαντικά δομικά χαρακτηριστικά σημειώνονται στην Gα_{i1}, και οι υπόλοιπες υπομονάδες ακολουθούν τον ίδιο προσανατολισμό με αυτήν. Οι τιμές RMSD δίνονται στον Πίνακα 3.7. **B.** Πίνακας αποστάσεων (σε Å) ανάμεσα σε CA άνθρακες καταλοίπων που συμμετέχουν σε αλληλεπιδράσεις με εκτελεστές, πρωτεΐνες RGS και υποδοχείς. Οι δομές δίνονται στη στήλη «Active:Inactive». Με «-» συμβολίζονται οι περιπτώσεις όπου είτε απουσιάζουν τμήματα των δομών είτε δεν υπάρχουν γνωστά αλληλεπιδρώντα κατάλοιπα.

3.3.3 Δομική ετερογένεια μεταξύ των διαφορετικών οικογενειών των Gα

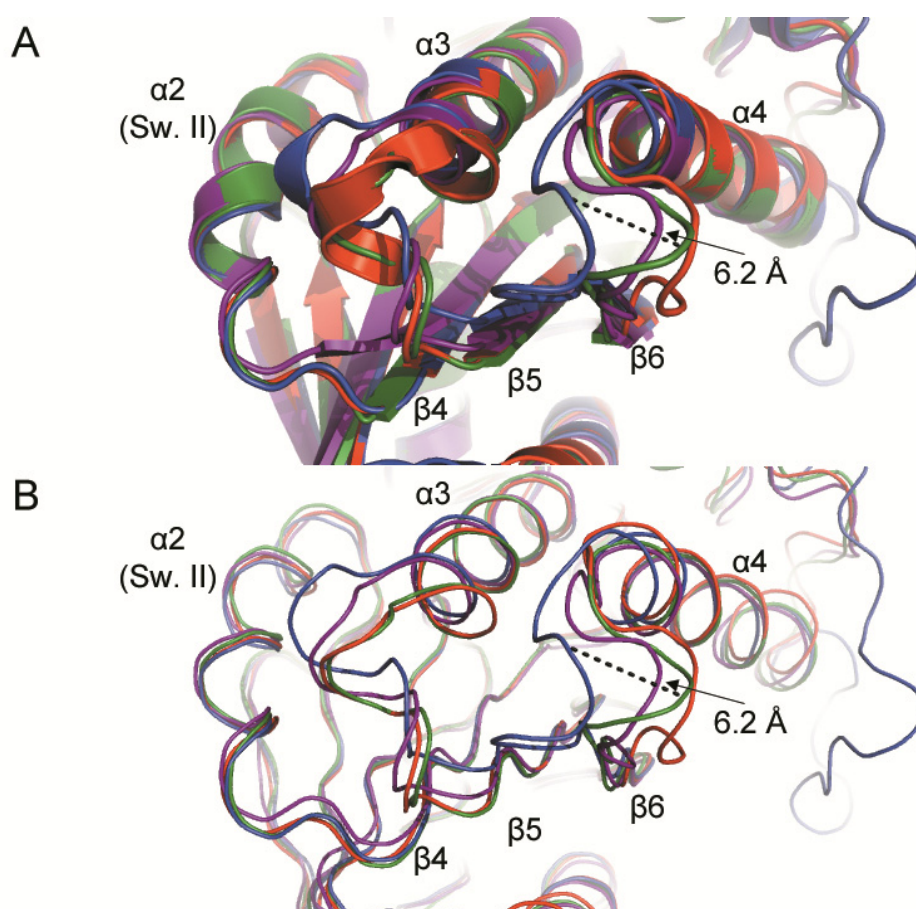
Η υπέρθεση δομών διαφορετικών Gα δείχνει πως το μεγαλύτερο μέρος της αυτοτελούς περιοχής GTPάσης είναι υψηλά συντηρημένο, ωστόσο παρατηρούνται κάποιες σημαντικές διαφορές. Η μεγαλύτερη ετερογένεια στη δομή παρουσιάζεται από την Gα_s, η οποία διαθέτει μια εκτεταμένη ακολουθία αμινοξικών καταλοίπων μετά τη μικρή έλικα αG, που δεν παρατηρείται στις υπόλοιπες υπομονάδες. Οι βρόχοι α3-β5 και α4-β6 διαφοροποιούνται τόσο στην ακολουθία όσο και στη δομή μεταξύ των τεσσάρων οικογενειών (Εικόνα 3.19). Ο βρόχος α3-β5 των Gα_s και Gα₁₂ υιοθετεί μια δομή εκτεταμένης έλικας, σε αντίθεση με την Gα_q και τα μέλη των Gα_{i/o}. Ο συγκεκριμένος βρόχος αποτελεί τμήμα της κοινής επιφάνειας πρόσδεσης εκτελεστών πάνω στις Gα, η διαφοροποίηση που παρατηρείται, ίσως αποτελεί έναν παράγοντα επιλεκτικότητας για την πρόσδεση εκτελεστών. Η θέση του βρόχου α4-β6 διαφοροποιείται επίσης ανάμεσα στην Gα_s και τις υπόλοιπες υπομονάδες. Η διαφορά αυτής της θέσης μάλιστα, ανάμεσα στις Gα_s και Gα_{i1} (~ 6 Å) έχει εμπλακεί στη διαφορετική φύση των αλληλεπιδράσεων τους με την Αδενυλική Κυκλάση (Dessauer, et al., 1998), ενώ είναι πιθανό να επηρεάζει τη φύση των αλληλεπιδράσεων των Gα με επιπλέον εκτελεστές, ίσως και με GPCRs.

3.3.4 Η πιθανή επιρροή του ηλεκτροστατικού δυναμικού στις αλληλεπιδράσεις των G-πρωτεϊνών

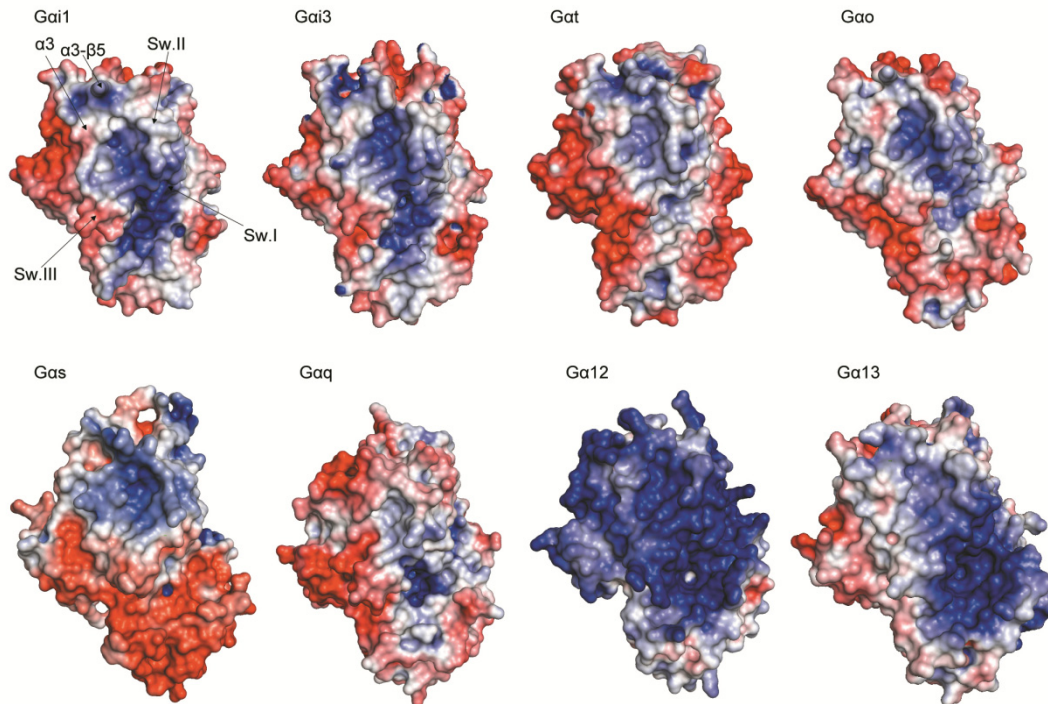
Η σύγκριση των ηλεκτροστατικών ιδιοτήτων των τεσσάρων οικογενειών καθώς και των υπο-οικογενειών των Gα υποδεικνύει έναν πιθανό επιπρόσθετο παράγοντα στη ρύθμιση των αλληλεπιδράσεων των G-πρωτεϊνών. Το ηλεκτροστατικό δυναμικό των Gα υπομονάδων με προσδιορισμένες κρυσταλλογραφικά δομές παρουσιάζεται στις Εικόνες 3.20 και 3.21. Από τη σύγκριση φαίνεται καθαρά πως η κατά τα άλλα υψηλά συντηρημένη κοινή επιφάνεια πρόσδεσης των εκτελεστών διαφοροποιείται έντονα ως προς τις ηλεκτροστατικές ιδιότητες της μεταξύ των τεσσάρων οικογενειών και, σε συγκεκριμένες περιπτώσεις, και ανάμεσα στα μέλη της ίδιας οικογένειας (Εικόνα 3.20, Εικόνα 3.21).

Μελετήθηκαν αναλυτικά και οι χάρτες ηλεκτροστατικού δυναμικού, τόσο των RGS πρωτεϊνών, όσο και των εκτελεστών, σε σύμπλοκα με Gα πρωτεΐνες. Συνολικά, παρατηρήθηκε πως οι περισσότερες αλληλεπιδράσεις επιφάνειας των Gα υπομονάδων είναι σε πολλές περιπτώσεις συμπληρωματικές στις πρωτεϊνικές επιφάνειες με τις οποίες

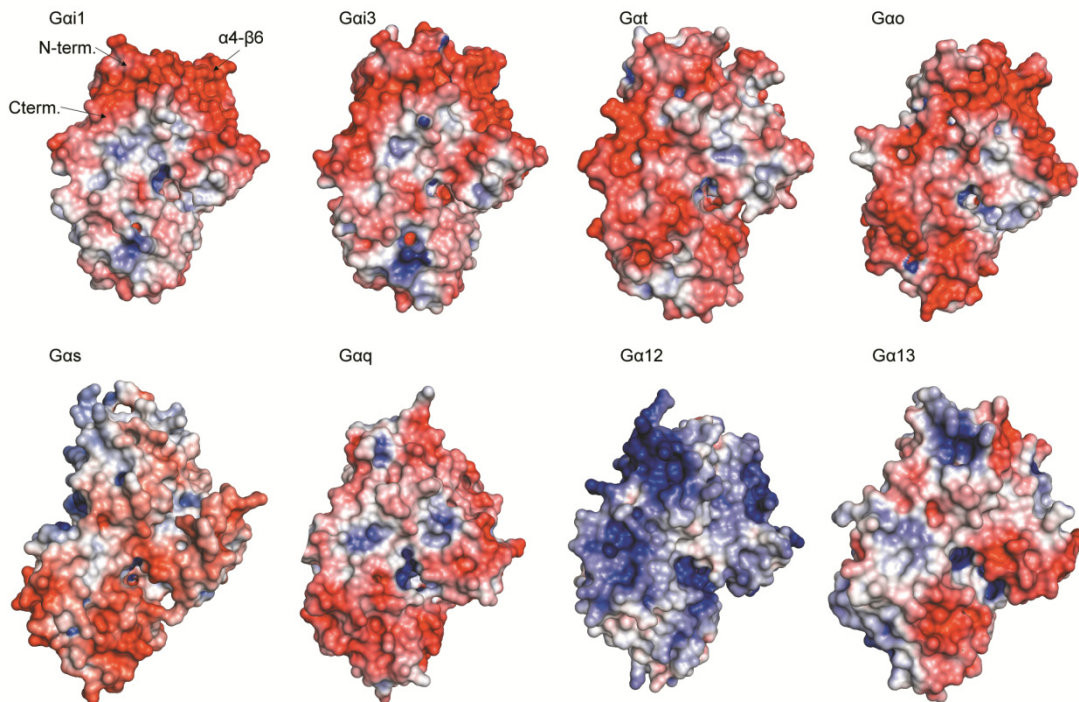
έρχονται σε επαφή, όσον αφορά τις ηλεκτροστατικές ιδιότητές τους. Η ετερογένεια που παρατηρείται, επιπλέον, όχι μόνο στις ήδη γνωστές αλληλεπιδρώσεις επιφάνειες αλλά και στην αυτοτελή περιοχή α -έλικας των $G\alpha$, σε συνδυασμό με δεδομένα που υποδεικνύουν αλληλεπιδράσεις της τελευταίας με υποδοχείς, εκτελεστές και πρωτεΐνες RGS δείχνει πως η ο ρόλος της ίσως είναι πιο σημαντικός από ότι πιστευόταν ως τώρα. Είναι, συνεπώς, πιθανό το ηλεκτροστατικό δυναμικό να διαδραματίζει καθοριστικό ρόλο στις αλληλεπιδράσεις των $G\alpha$ με εκτελεστές, πρωτεΐνες RGS και, πιθανότατα, ετεροδιμερή $G\beta\gamma$ ή GPCRs.



Εικόνα 3.19: Σύγκριση των περιοχών που έρχονται σε επαφή με εκτελεστές και GPCRs των $G\alpha_{11}$, $G\alpha_s$, $G\alpha_q$ και $G\alpha_{12}$ σε απεικόνιση “cartoon” (A) και “ribbon” (B). Συγκρίνονται ο «διακόπτης» (switch) II, η έλικα α_3 και ο βρόχος α_3 - β_5 , που αποτελούν την κοινή επιφάνεια αλληλεπίδρασης εκτελεστών, καθώς και ο βρόχος α_4 - β_6 , για τον οποίο μετράται απόσταση σε Å ανάμεσα στις $G\alpha_s$ και $G\alpha_{11}$. Οι χρωματισμοί είναι μπλε για την $G\alpha_s$ (1AZT.A), πράσινο για την $G\alpha_{11}$ (1GIA.A), μωβ για την $G\alpha_q$ (2RGN.A) και κόκκινο για την $G\alpha_{12}$ (1ZCA.A). Οι έλικες α_2 (Switch II) και α_3 παρουσιάζουν ελάχιστες αλλαγές. Αντίθετα, οι βρόχοι α_3 - β_5 και α_4 - β_6 διαφοροποιούνται ανάμεσα στις τέσσερις οικογένειες. Ο βρόχος α_4 - β_6 απέχει 4-6 Å σε σχέση με τους αντίστοιχους βρόχους των άλλων τριών υπομονάδων.



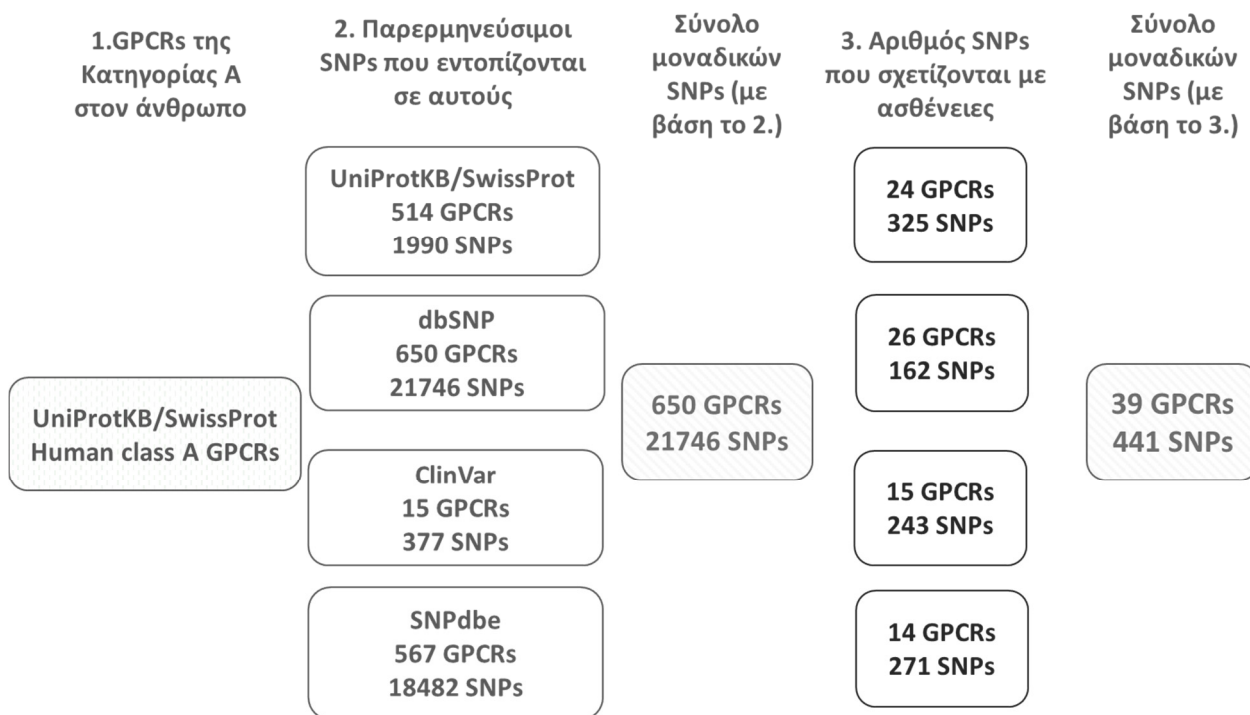
Εικόνα 3.20: Το ηλεκτροστατικό δυναμικό των μοριακών επιφανειών των Ga υπομονάδων. Όλες οι υπομονάδες ακολουθούν τον ίδιο προσανατολισμό, παρουσιάζοντας την όψη που φέρει την κοινή επιφάνεια αλληλεπίδρασης εκτελεστών. Σημαντικά δομικά χαρακτηριστικά σημειώνονται στη δομή της Ga_{i1} . Το εύρος της απεικόνισης κυμαίνεται από -5 (κόκκινο) ως $+5$ Kt/e⁻ με βάση το ηλεκτροστατικό δυναμικό της προσβάσιμης στο διαλύτη επιφάνειας Lee-Richards (Solvent Accessible Surface, ASA). Χρησιμοποιούνται οι ενεργές υπομονάδες των Ga_{i1} (1GIA), Ga_t (1TND), Ga_s (1AZT), Ga_q (2RGN), Ga_{12} (1ZCA) και Ga_{13} (3CX8) και οι προσδεδμεμένες σε RGS υπομονάδες των Ga_{i3} (2V4Z) και Ga_o (3C7K).



Εικόνα 3.21: Το ηλεκτροστατικό δυναμικό των μοριακών επιφανειών των Ga υπομονάδων. Οι υπομονάδες απεικονίζονται στραμμένες κατά 180° σε σχέση με τον προσανατολισμό της Εικόνας 3.20. Σημαντικά δομικά στοιχεία σημειώνονται στη δομή της Ga_{i1} . Η απεικόνιση ακολουθεί τους κανόνες της Εικόνας 3.20.

3.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες

Μέσα από τη διαδικασία που αναλύθηκε στην παράγραφο 2.4, προέκυψαν τα αποτελέσματα που συνοψίζονται στην Εικόνα 3.22.



Εικόνα 3.22: Συνοπτικό σχεδιάγραμμα στο οποίο φαίνονται οι πηγές των δεδομένων και τα αριθμητικά δεδομένα συνολικά για τους SNPs της Κατηγορίας A των GPCRs στον άνθρωπο.

3.4.1 Στατιστική επεξεργασία των αποτελεσμάτων

Για το μοναδικό σύνολο των 650 ανθρώπινων κατηγορίας A (class A) GPCRs και των 21746 παρερμηνεύσιμων SNPs που εντοπίστηκαν σε αυτούς, αφού εξαιρέθηκαν οι υποδοχείς όσφρησης (olfactory receptors) (λόγω μεγάλης ομοιογένειας θα αποτελούσαν «θόρυβο» σε μία στατιστική ανάλυση), προέκυψε ένα σύνολο 283 GPCRs και 9662 SNPs, οι οποίοι κατηγοριοποιήθηκαν με βάση τη θέση τους στην πρωτεϊνική ακολουθία και με βάση την τοπολογία της εκάστοτε πρωτεΐνης.

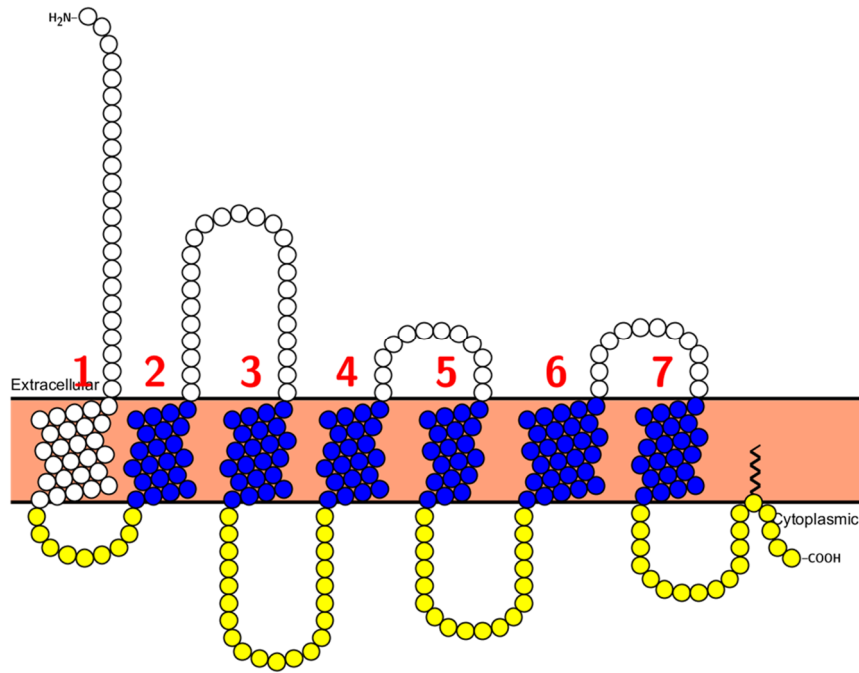
Για κάθε πρωτεΐνη του εκάστοτε συνόλου, πραγματοποιήθηκε ανάλυση του ποσοστού των παρερμηνεύσιμων μονονουκλεοτιδικών πολυμορφισμών χρησιμοποιώντας ως κριτήριο την τοπολογία. Έτσι για την πρωτεΐνη X, έστω ότι στο αμινοτελικό άκρο εντοπίστηκαν 2 SNPs, ενώ στη χ πρωτεΐνη εντοπίστηκαν συνολικά 10 SNPs. Επίσης, έστω ότι το αμινοτελικό άκρο αριθμούσε 20 κατάλοιπα, ενώ το σύνολο της πρωτεΐνης 200. Τότε, αναμένεται στο αμινοτελικό άκρο να βρούμε $(10/200) \cdot 20 = 1$ SNP αναμενόμενη

τιμή (expected value). Όμως αυτό διαφέρει κατά ένα από το παρατηρηθέν (observed value). Η διαδικασία αυτή έγινε για όλες τις τοπολογικές περιοχές, όλων των πρωτεϊνών του συνόλου και στη συνέχεια συγκρίθηκε η κατανομή των παρατηρηθέντων με αυτή των αναμενόμενων τιμών, με χρήση του t-test ανα ζεύγη. Από τα παραπάνω προέκυψαν οι παρακάτω πίνακες (Πίνακας 3.8, Πίνακας 3.9) καθώς και οι σχηματικές αναπαραστάσεις (Εικόνα 3.23, Εικόνα 3.24).

Τα παρόντα αποτελέσματα δείχνουν μία «προτίμηση» που παρουσιάζουν τα SNPs να εμφανίζονται σε συγκεκριμένες τοπολογικές περιοχές, ενώ σε κάποιες άλλες υπάρχει στατιστικά μειωμένη εμφάνιση τους. Η εμφάνιση ή όχι SNPs σε συγκεκριμένες τοπολογικές περιοχές, πιθανώς να σχετίζεται, στην περίπτωση της μειωμένης εμφάνισης, με περιοχές όπου έχουμε μεγάλη συντηρητικότητα, ή το αντίθετο, σε περιπτώσεις αυξημένης εμφάνισης. Θα είχε ενδιαφέρον η σύγκριση της υπερεμφάνισης ή της υποεμφάνισης παθολογικών SNPs σε συγκεκριμένες τοπολογικές περιοχές, όμως, προς το παρόν, το σύνολο των δεδομένων δεν επιτρέπει την εξαγωγή στατιστικά σημαντικών αποτελεσμάτων.

Πίνακας 3.8 Αποτελέσματα των t-test κατά ζεύγη για το σύνολο των ανθρώπινων GPCRs της κατηγορίας A (εκτός των οσφρητικών υποδοχέων) και για κάθε τοπολογική περιοχή.

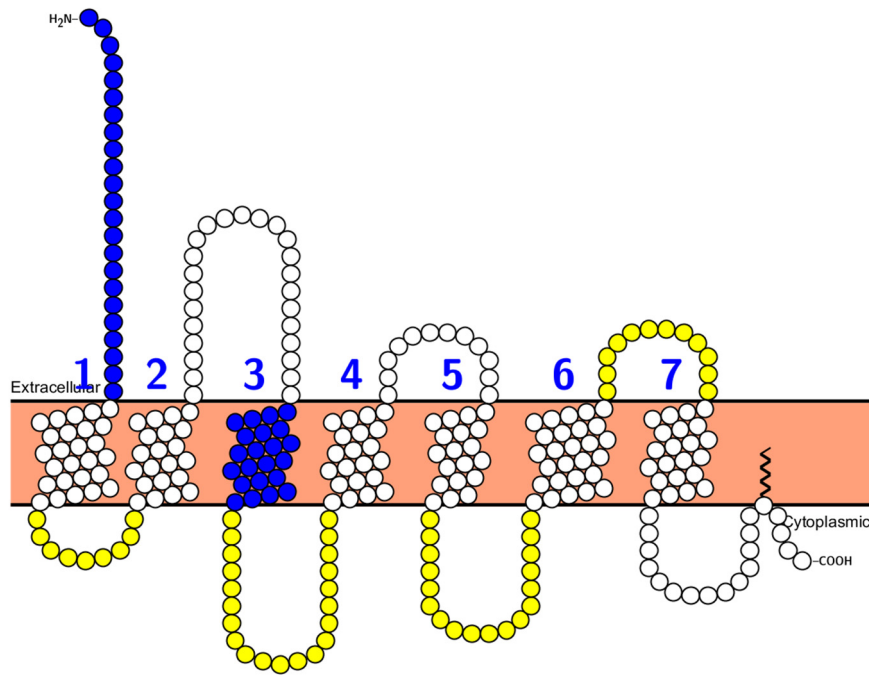
	T-TEST	P-VALUE	CI	CI	TSTAT	DF	SD
N- TERMINAL	0	0,2943	-0,1413	0,4648	1,0507	282	2,5897
TM1	0	0,3127	0,3087	0,0991	-1,0115	282	1,7425
ICL1	1	0,0038	0,0609	0,312	2,9224	282	1,0732
TM2	1	1,4957e-08	-0,631	-0,3125	-5,8323	282	1,3607
ECL1	0	0,8870	-0,149	0,1722	0,1422	282	1,3727
TM3	1	5,2560e-06	-0,4983	-0,2016	-4,6436	282	1,2678
ICL2	1	1,8570e-04	0,166	0,5251	3,7879	282	1,5346
TM4	1	0,0305	0,3575	-0,0178	-2,1748	282	1,4516
ECL2	0	0,5784	0,2458	0,1374	-0,5564	282	1,6375
TM5	1	4,3915e-06	-0,5002	-0,2041	-4,6834	282	1,2649
ICL3	1	2,8118e-10	0,6705	12,474	6,5442	282	2,4651
TM6	1	3,0724e-04	0,4792	-0,1437	-3,6545	282	1,4336
ECL3	0	0,4542	0,0899	0,2004	0,7495	282	1,2405
TM7	1	3,4222e-07	-0,5986	-0,2709	-5,2232	282	1,4003
C- TERMINAL	1	0,0075	0,1472	0,9469	2,6933	282	3,4169



Εικόνα 3.23: Διάγραμμα δευτεροταγούς δομής των GPCRs. Χρωματίζονται οι τοπολογικές περιοχές που παρουσίασαν στατιστικά αυξημένη εμφάνιση πολυμορφισμών (κίτρινο) και εκείνες με στατιστικά μειωμένη εμφάνιση (μπλε) στο σύνολο των ανθρώπινων GPCRs της κατηγορίας A (εκτός των οσφρητικών υποδοχέων).

Πίνακας 3.9 Αποτελέσματα των t-test κατά ζεύγη για το σύνολο των ανθρώπινων οσφρητικών GPCRs και για κάθε τοπολογική περιοχή.

	T-TEST	P-VALUE	CI	CI	TSTAT	DF	SD
N- TERMINAL	1	3,1335e-04	-0,4789	-0,1429	-3,639	363	1,63
TM1	0	0,0565	-0,3053	0,0042	-1,9129	363	1,5016
ICL1	1	0,0101	0,0342	0,2518	2,5856	363	1,0556
TM2	0	0,1313	0,2374	0,031	-1,5126	363	1,3017
ECL1	0	0,7822	-0,1254	0,1665	0,2767	363	1,4159
TM3	1	3,6664e-04	0,4076	-0,1194	-3,5969	363	1,3977
ICL2	1	1,8646e-12	0,4325	0,7515	7,2973	363	1,5478
TM4	0	0,0977	-0,254	0,0214	-1,6604	363	1,3361
ECL2	0	0,2058	-0,3236	0,07	-1,2675	363	1,9092
TM5	0	0,3456	-0,2326	0,0817	-0,9443	363	1,5245
ICL3	1	2,7703e-08	0,2765	0,5693	5,6796	363	1,4205
TM6	0	0,5947	-0,1078	0,1879	0,5325	363	1,4348
ECL3	1	0,0474	0,0015	0,2476	1,9899	363	1,1939
TM7	0	0,1364	-0,2587	0,0354	-1,4928	363	1,427
C- TERMINAL	0	0,2753	-0,237	0,0677	-1,0925	363	1,4781



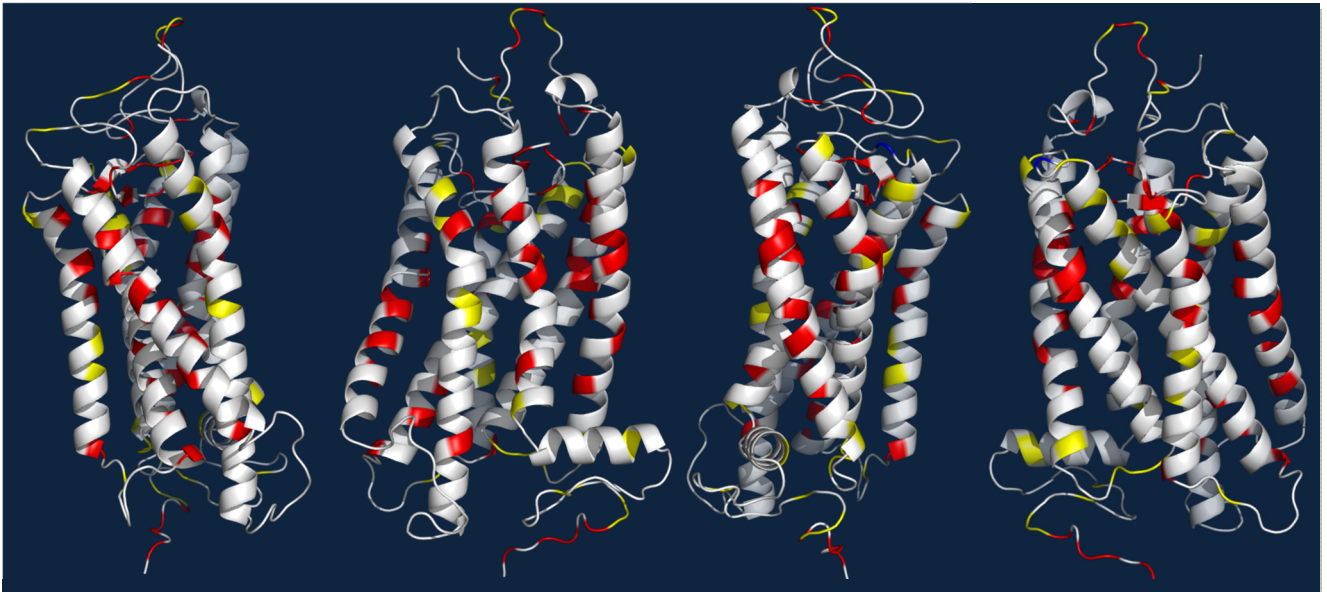
Εικόνα 3.24: Διάγραμμα δευτεροταγούς δομής των GPCRs. Χρωματίζονται οι τοπολογικές περιοχές που παρουσίασαν στατιστικά αυξημένη εμφάνιση πολυμορφισμών (κίτρινο) και εκείνες με στατιστικά μειωμένη εμφάνιση (μπλε) στο σύνολο των ανθρώπινων οσφρητικών GPCRs.

3.4.2 Οπτικοποίηση SNPs σε επίπεδο δευτεροταγούς και τριτοταγούς δομής

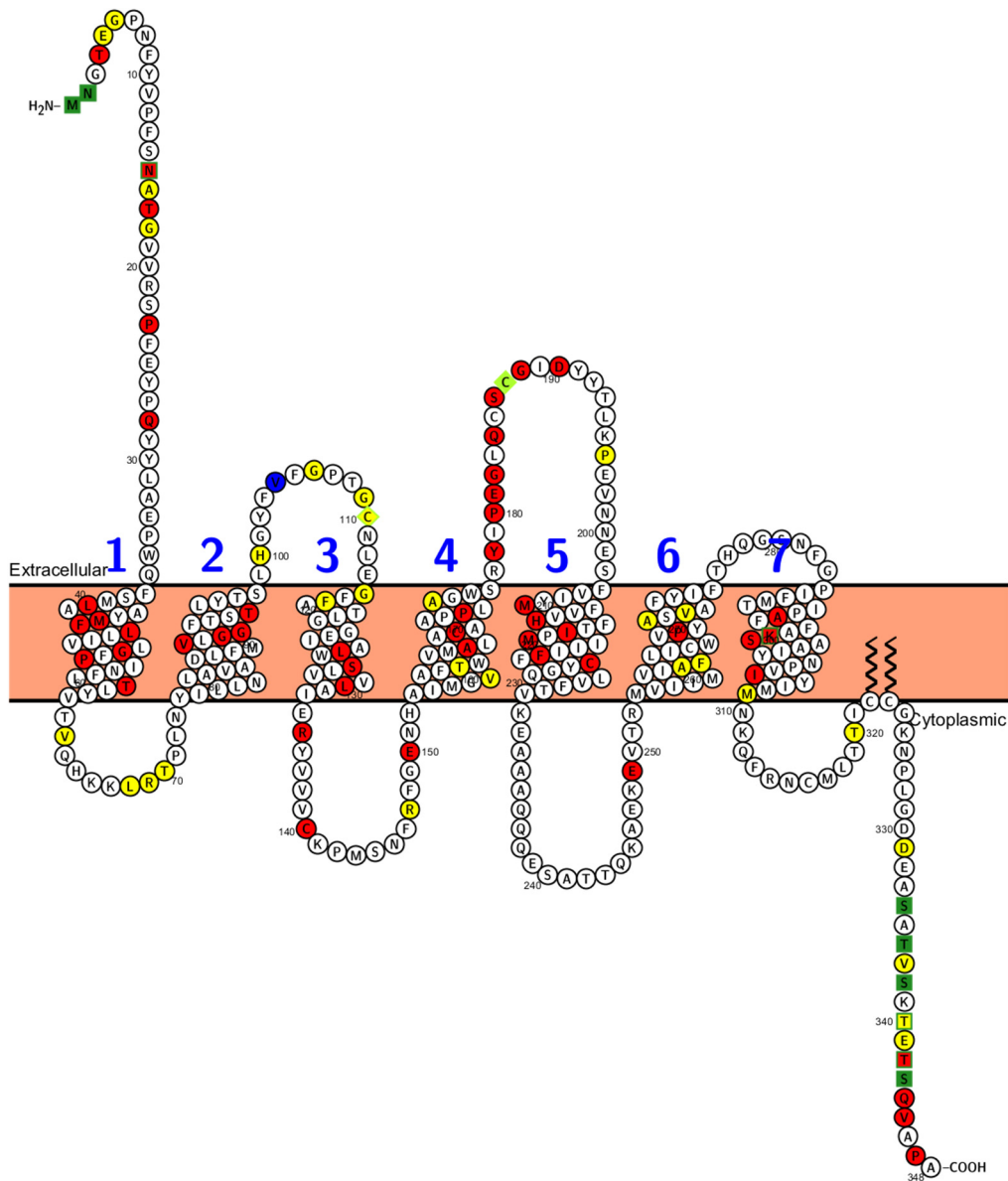
Όπως ήδη αναφέρθηκε, συγκεντρώθηκαν 39 υποδοχείς, οι οποίοι εμφανίζουν τουλάχιστον έναν SNP που σχετίζεται με κάποια ασθένεια, σύμφωνα με τα δεδομένα που συλλέξαμε. Για όσους από τους 39 υποδοχείς δεν είχαν προσδιορισμένη τρισδιάστατη δομή, κατασκευάστηκαν δομικά μοντέλα ομολογίας με χρήση του Modeller (Sali and Blundell, 1993). Σε αυτά τα τρισδιάστατα μοντέλα, αλλά και στα μοντέλα δευτεροταγούς δομής μέσω του Protter, επισημάνθηκαν όλοι οι μονονουκλεοτιδικοί πολυμορφισμοί που εμφανίζει ο κάθε υποδοχέας, σημαίνοντας διαφορετικά τους σχετιζόμενους με κάποια ασθένεια πολυμορφισμούς από τους ουδέτερους. Ενδεικτικά παρουσιάζονται στις Εικόνες 3.25 και 3.26 οι μονονουκλεοτιδικοί πολυμορφισμοί που εντοπίστηκαν στη ροδοψίνη. Με κόκκινο χρώμα σημαίνονται οι πολυμορφισμοί που σχετίζονται με ασθένειες και με κίτρινο οι ουδέτεροι πολυμορφισμοί.

Συνολικά υπήρχε μεγάλη ετερογένεια στα δεδομένα, γεγονός που οφείλεται στα ελλιπή δεδομένα συσχέτισης με ασθένειες, αλλά και στο γεγονός ότι συχνά κάποια θέματα είναι μεγαλύτερου ενδιαφέροντος και έτσι είναι πιο εκτενώς μελετημένα. Παρόλα τα προβλήματα, σε κάποιους από τους υποδοχείς παρατηρήθηκε μια «συγκέντρωση» των

πολυμορφισμών σε ορισμένες περιοχές στο χώρο. Η παρατήρηση αυτή πρέπει να αξιολογηθεί και να ποσοτικοποιηθεί, γεγονός που αποτελεί αντικείμενο μελέτης.



Εικόνα 3.25: Τρισδιάστατη αναπαράσταση της δομής της Ροδοψίνης όπου επισημαίνονται οι παρερμηνεύσιμοι SNPs. Με κόκκινο σημαίνονται οι πολυμορφισμοί που σχετίζονται με ασθένειες, ενώ με κίτρινο αυτοί για τους οποίους δεν είναι γνωστό αν σχετίζονται ή όχι με ασθένειες.



Εικόνα 3.26: Διάγραμμα δευτεροταγούς δομής της Ροδοψίνης όπου επισημαίνονται οι παρερμηνεύσιμοι SNPs. Με κόκκινο σημαίνονται οι πολυμορφισμοί που σχετίζονται με ασθένειες, ενώ με κίτρινο αυτοί για τους οποίους δεν είναι γνωστό αν σχετίζονται ή όχι με ασθένειες. Με πράσινο σημαίνονται οι θέσεις μεταμεταφραστικών τροποποιήσεων.

4. Συμπεράσματα

Κλείνοντας την παρούσα διδακτορική διατριβή, θα επιχειρηθεί μία σύνοψη των σημαντικότερων συμπερασμάτων που προέκυψαν, σε όλα τα στάδια της μελέτης. Θα γίνει μια προσπάθεια ανάδειξης της σημασίας των μεθόδων που αναπτύχθηκαν και των αναλύσεων που διεξήχθησαν και, τέλος, θα τεθούν οι μελλοντικοί στόχοι και οι προεκτάσεις που θα μπορούσαν να έχουν όλες οι πτυχές της παρούσας μελέτης.

4.1 Μέθοδοι εντοπισμού και ταξινόμησης GPCRs και G-πρωτεϊνών

Στα πλαίσια της παρούσας διδακτορικής διατριβής αναπτύχθηκαν δύο νέες μέθοδοι. Η μία αφορά τον εντοπισμό από την αλληλουχία και μόνο και το λεπτομερή σχολιασμό πιθανών συζευγμένων με G-πρωτεΐνες υποδοχέων (GPCRripe), ενώ η μέθοδος GprotPRED έρχεται να συμπληρώσει την προηγούμενη μέθοδο, καθώς εντοπίζει και ταξινομεί σε οικογένειες πιθανές G-πρωτεΐνες.

Η μέθοδος GPCRripe είναι μία αξιόπιστη μέθοδος για τον εντοπισμό πιθανών συζευγμένων με G-πρωτεΐνες υποδοχέων, προχωρώντας όμως και ένα βήμα παραπέρα. Για κάθε πιθανό GPCR, μέσω του δεύτερου επιπέδου της μεθόδου που αφορά το λειτουργικό σχολιασμό, προσφέρει στο χρήστη σημαντικές πληροφορίες για τη λειτουργία και την τοπολογία της πρωτεΐνης. Συγκεκριμένα, η μέθοδος GPCRripe παρέχει πληροφορίες για τη συγκεκριμένη οικογένεια στην οποία ανήκει ο υποδοχέας. Επιπλέον, παρέχεται και η τοπολογία του υποδοχέα στη μεμβράνη, καθώς και το πεπτίδιο οδηγητή, όπου αυτό υπάρχει. Περαιτέρω τοπολογικές αλλά και λειτουργικές πληροφορίες προκύπτουν από την εμφάνιση συγκεκριμένων εξωκυττάρων ή/και κυτταροπλασματικών συντηρημένων περιοχών (εμφάνιση εξειδικευμένων rHMMs της Pfam με τις συγκεκριμένες ιδιότητες). Τέλος, μέσω της μεθόδου PRED-COUPLE2 (Sgourakis, et al., 2005), παρέχεται η εξειδίκευση “σύζευξης” με συγκεκριμένες οικογένειες G-πρωτεϊνών, πληροφορία χρήσιμη για τη διαλεύκανση της συγκεκριμένης λειτουργίας του υποδοχέα στο κύτταρο. Η συγκεκριμένη μέθοδος θα αποτελέσει πολύτιμο εργαλείο στα χέρια επιστημόνων που μελετούν τη σημαντική αυτή υπεροικογένεια ευκαρυωτικών υποδοχέων, αλλά και στον αυτοματοποιημένο σχολιασμό νεοπροσδιοριζόμενων πρωτεωμάτων.

Η μέθοδος GprotPRED σχεδιάστηκε για να εντοπίζει με ταχύτητα και αξιοπιστία G-πρωτεΐνες με μόνο δεδομένο την αλληλουχία τους. Ειδικότερα, με τα ειδικά κατασκευασμένα rHMMs

που περιλαμβάνει, μπορεί να εντοπίσει Gα υπομονάδες και μάλιστα να της ταξινομήσει στις τέσσερις βασικές οικογένειες που συναντώνται στα θηλαστικά (G_s , $G_{i/o}$, $G_{q/11}$, $G_{12/13}$), χωρίς όμως να αποκλείονται Gα υπομονάδες που δεν ανήκουν σε κάποια από αυτές τις οικογένειες (αυτό επιτυγχάνεται με τη χρήση του γενικού ρHMM της Pfam για την Gα υπομονάδα με κωδικό PF00503). Επιπλέον, με τα δύο ειδικά κατασκευασμένα ρHMMs για τη β και τη γ υπομονάδα, ο εντοπισμός G-πρωτεϊνών είναι ολοκληρωμένος. Η μέθοδος GprotPRED θα αποτελέσει ένα χρήσιμο εργαλείο για τον αυτοματοποιημένο σχολιασμό πρωτεωμάτων.

4.2 Βάσεις δεδομένων GPCRs, G-πρωτεϊνών, εκτελεστών και υποκαταστατών

Δύο νέες βάσεις δεδομένων που σχετίζονται με τους GPCRs ήταν το αποτέλεσμα της παρούσας διδακτορικής διατριβής.

Η Human-grDB, συγκρινόμενη με την προϋπάρχουσα βάση δεδομένων του εργαστηρίου μας, grDB (Elefsinioti, et al., 2004; Theodoropoulou, et al., 2008), περιλαμβάνει περισσότερο ολοκληρωμένα και, σύμφωνα με τις τελευταίες εξελίξεις στο πεδίο, δεδομένα. Εστιάζει στους ανθρώπινους GPCRs, G-πρωτεΐνες και εκτελεστές, χωρίς όμως να είναι απλά ένα υποσύνολο της grDB.

Συνδυάζει πλήρως πληροφορίες από εξωτερικές πηγές δεδομένων, προσφέροντας με αυτόν τον τρόπο πληροφορίες που δεν υπάρχουν διαθέσιμες σε άλλες βάσεις, ενώ παρέχει μία φιλική προς το χρήστη διεπαφή, που υποστηρίζεται από προηγμένες τεχνικές οπτικοποίησης.

Η Human-grDB είναι ήδη χρήσιμη σύντροφος των ερευνητών στις βιοεπιστήμες, καθώς συνδυάζει τις επίκαιρες και προσεκτικά συγκεντρωμένες πληροφορίες για τους ανθρώπινους GPCRs, G-πρωτεΐνες, εκτελεστές και τις αλληλεπιδράσεις μεταξύ τους, με προηγμένες τεχνικές οπτικοποίησης, με στόχο να γίνει ο μεγάλος αυτός όγκος δεδομένων πιο χρήσιμος. Ευελπιστούμε ότι, η Human-grDB θα γίνει οδηγός αναφοράς στη φαρμακευτική έρευνα, και, ειδικά, στην ανακάλυψη χημικών και φαρμάκων για ανθρώπινες ασθένειες.

Η PLHG-DB περιλαμβάνει μόνο τους πεπτιδικούς υποκαταστάτες των ανθρώπινων GPCRs, ομαδοποιημένους σε οικογένειες, υπο-οικογένειες και τύπους σύμφωνα με λειτουργικά χαρακτηριστικά. Το χαρακτηριστικό αυτό δεν παρατηρείται σε κάποια άλλη δημόσια διαθέσιμη πηγή. Το σχεσιακό της μοντέλο είναι ειδικά σχεδιασμένο για να παρουσιάζει τις

επιλεκτικές αλληλεπιδράσεις των πεπτιδικών υποκαταστατών με τους GPCRs, σύμφωνα με τα επιστημονικά δεδομένα της διεθνούς βιβλιογραφίας.

Για κάθε πεπτιδικό υποκαταστάτη, η PLHG-DB, παρουσιάζει την αλληλεπίδραση με τον ή τους υποδοχείς, την αμινοξική ακολουθία, την έναρξη και τη λήξη της ακολουθίας στο προπεπτίδιο. Επίσης περιλαμβάνεται πληροφορία για την ιδιαιτερότητα του πεπτιδικού υποκαταστάτη, την οικογένεια στην οποία ανήκει καθώς και τη λειτουργία, που επιτελεί στον οργανισμό ο συγκεκριμένος πεπτιδικός υποκαταστάτης, μέσω της αλληλεπίδρασης με τον εκάστοτε υποδοχέα. Ταυτόχρονα, προβάλλεται πληροφορία σε μορφή σχολίου για την αλληλεπίδραση του πεπτιδικού υποκαταστάτη με τον υποδοχέα και οι ιδιαιτερότητές τους. Η παραπάνω πληροφορία, δεν παρουσιάζεται συγκεντρωμένη με αυτό τον τρόπο σε καμία άλλη διαθέσιμη δημόσια βάση δεδομένων.

Παρόλο που η PLHG-DB δεν αποτελεί τη μοναδική διαθέσιμη βάση δεδομένων που περιλαμβάνει πεπτιδικούς υποκαταστάτες των GPCRs, είναι η μόνη που προσφέρει στο χρήστη γνώση που έχει εξορυχθεί με μη αυτόματο τρόπο από τη διεθνή βιβλιογραφία.

4.3 Οι αλληλεπιδράσεις των G-πρωτεϊνών με GPCRs, εκτελεστές και RGS πρωτεΐνες

Από τη μελέτη των κρυσταλλικών δομών των συμπλόκων των G-πρωτεϊνών, σε συνδυασμό με βιβλιογραφικά στοιχεία προερχόμενα από πειράματα μεταλλαξογένεσης και υπολογιστικές μελέτες αναγνωρίσαμε στην επιφάνεια των Gα υπομονάδων μια σειρά από περιοχές και κατάλοιπα που μπορούν, σε πολλές περιπτώσεις, να αλληλεπιδράσουν τόσο με GPCRs, όσο και με εκτελεστές. Αυτές οι περιοχές περιλαμβάνουν το αμινοτελικό και καρβοξυτελικό άκρο, και τους βρόχους α3-β5 και α4-β6.

Η σύγκριση ενεργών και ανενεργών Gα υπομονάδων δείχνει πως οι συγκεκριμένες περιοχές διαφοροποιούνται κατά την ενεργοποίηση των G-πρωτεϊνών, γεγονός που θα μπορούσε να επηρεάζει το ρόλο τους στις αλληλεπιδράσεις των Gα. Η σύγκριση της ακολουθίας και της δομής αυτών των περιοχών ανάμεσα σε διαφορετικές Gα υπομονάδες υποδεικνύει πως η ετερογένειά τους μπορεί να επηρεάσει τις αλληλεπιδράσεις τους με εκτελεστές και ίσως με υποδοχείς.

Τέλος, η αξιοσημείωτη ετερογένεια των ηλεκτροστατικών ιδιοτήτων που παρατηρείται όχι μόνο ανάμεσα στις διαφορετικές οικογένειες των Gα υπομονάδων αλλά ακόμη και ανάμεσα σε μέλη της ίδιας οικογένειας, καθώς και η συμπληρωματικότητά τους με τις ηλεκτροστατικές ιδιότητες των αντίστοιχων αλληλεπιδρώντων με αυτές εκτελεστών και

πρωτεϊνών RGS, προτείνει πως το ηλεκτροστατικό δυναμικό μπορεί να δράσει καθοριστικά στις αλληλεπιδράσεις των G-πρωτεϊνών. Οι μετρήσεις των διαμοριακών ενεργειών των συμπλόκων $G\alpha - RGS$ και $G\alpha - GPCRs$ έδειξαν πως, σε περιπτώσεις με γνωστή και καταγεγραμμένη διαφορά στη συγγένεια των αλληλεπιδράσεων, σημειώνεται σημαντική διαφοροποίηση στις τιμές της ηλεκτροστατικής ενέργειας, γεγονός που ενισχύει το ρόλο του ηλεκτροστατικού δυναμικού, σε αυτές τις αλληλεπιδράσεις.

Ένα ακόμη σημαντικό συμπέρασμα που προκύπτει από τη μελέτη των αλληλεπιδράσεων σε συνδυασμό με τις συγκρίσεις του ηλεκτροστατικού δυναμικού είναι πως η σηματοδοτική και ρυθμιστική ικανότητα των $G\alpha$ υπομονάδων ίσως να μην περιορίζεται στην αυτοτελή περιοχή GTPάσης τους, αλλά να περιλαμβάνει και στοιχεία στην αυτοτελή περιοχή α -έλικας. Τα αποτελέσματά μας ενισχύουν την υπόθεση πως η συγκεκριμένη αυτοτελής περιοχή, που μέχρι πρόσφατα είχε απασχολήσει ελάχιστα την έρευνα στο πεδίο των G-πρωτεϊνών, ίσως διαδραματίζει σημαντικό ρόλο στη ρύθμιση των αλληλεπιδράσεων των $G\alpha$.

Η πληροφορία που παρέχεται από τα αποτελέσματα της παρούσας διδακτορικής διατριβής θα μπορούσε να φανεί χρήσιμη στην περαιτέρω μελέτη της δομικής φύσης των αλληλεπιδράσεων των G-πρωτεϊνών με GPCRs και περισσότερα μόρια εκτελεστές, αλλά και στη μελέτη των αλληλεπιδράσεων πρωτεϊνών - πρωτεϊνών γενικότερα.

4.4 Μελέτη των SNPs στους ανθρώπινους GPCRs και συσχέτιση με ασθένειες

Ύστερα από τη μελέτη που πραγματοποιήθηκε, κατέστη σαφές ότι οι βάσεις δεν παρέχουν ακόμη επαρκείς πληροφορίες για τη συσχέτιση SNPs-ασθενειών. Καθώς το ενδιαφέρον γύρω από την παθογένεια των SNPs αυξάνεται, αναμένεται και αύξηση αυτής της πληροφορίας μελλοντικά. Ένα επιπλέον πρόβλημα που αντιμετωπίσαμε ήταν η ετερογένεια των δεδομένων από βάση σε βάση και η ανάγκη για ενοποίηση ώστε να προκύψει ένα μοναδικό πλέον σύνολο. Η διαδικασία αυτή ήταν αρκετά επώδυνη, καθώς συχνά ως σημείο αναφοράς χρησιμοποιούνται πρωτεϊνικές ακολουθίες από διαφορετικές βάσεις, με αποτέλεσμα να προκύπτει ασυμβατότητα.

Η στατιστική ανάλυση κατανομής των πολυμορφισμών με βάση την τοπολογία κατέστη δυνατή μόνο για το σύνολο των πολυμορφισμών. Αρχικός στόχος ήταν να γίνει και για το σύνολο των παθογόνων και των μη παθογόνων πολυμορφισμών, αλλά τα αντίστοιχα σύνολα ήταν αρκετά μικρά και κατά συνέπεια δεν προέκυψε κάποιο στατιστικά σημαντικό αποτέλεσμα. Η κατανομή του συνόλου των πολυμορφισμών παρέχει σημαντική

πληροφορία, καθώς υποδεικνύει αυξημένη εμφάνιση πολυμορφισμών στην κυτταροπλασματική πλευρά, γεγονός που έρχεται σε αντίθεση με τη δομική συντηρητικότητα της περιοχής αυτής. Ενδιαφέρον αποτελεί το γεγονός ότι, η κυτταροπλασματική πλευρά του υποδοχέα είναι υπεύθυνη για την αλληλεπίδραση με τις G-πρωτεΐνες, και, ίσως η αυξημένη εμφάνιση πολυμορφισμών να σχετίζεται με την εξειδίκευση της «σύζευξης» αυτής. Επόμενο βήμα είναι η αξιολόγηση της «φύσης» των πολυμορφισμών αυτών, δηλαδή του είδους των αλλαγών αμινοξικών καταλοίπων που γίνονται και εάν το είδος των αλλαγών δικαιολογεί τη δομική συντηρητικότητα.

Μέσω της οπτικοποίησης των SNPs σε διαγράμματα δευτεροταγούς δομής και σε τρισδιάστατα μοντέλα, παρατηρήθηκε σε συγκεκριμένους υποδοχείς μια «συγκέντρωση» των πολυμορφισμών σε ορισμένες περιοχές στο χώρο. Η παρατήρηση αυτή πρέπει να αξιολογηθεί και να ποσοτικοποιηθεί, γεγονός που αποτελεί αντικείμενο μελέτης.

Η σημασία των μονονουκλεοτιδικών πολυμορφισμών και ο συσχετισμός τους με ασθένειες αποτελεί ένα σχετικά καινούριο πεδίο έρευνας, και η μελέτη τους μπορεί να αποκαλύψει σημαντικές πληροφορίες όχι μόνο γύρω από την παθογένεια και τη θεραπεία ασθενειών αλλά και να αποκαλύψει πληροφορίες σχετικά με τη λειτουργικότητα των πρωτεϊνών και πως αυτή καθορίζεται από τη σύσταση αλλά και τη δομή της πρωτεΐνης.

4.5 Μελλοντικές κατευθύνσεις

Ολοκληρώνοντας την παρούσα διδακτορική διατριβή και έχοντας μελετήσει εκτενώς από υπολογιστικής σκοπιάς τους συζευγμένους με G-πρωτεΐνες υποδοχείς, υπάρχουν ορισμένες προτάσεις τις οποίες θα ήθελα εδώ να αναφέρω. Τα δεδομένα που αφορούν το συγκεκριμένο θέμα, όπως και όλα τα βιολογικά δεδομένα, συσσωρεύονται και ανανεώνονται με ξέφρενους ρυθμούς. Για αυτό το λόγο είναι σκόπιμη η συνεχής ανανέωση τόσο των βάσεων δεδομένων που δημιουργήθηκαν, όσο και η περαιτέρω εξέλιξη των μεθόδων και των αναλύσεων που προτείνονται.

Σε ότι αφορά τις μεθόδους εντοπισμού και ταξινόμησης GPCRripe και GprotPRED, ενδιαφέρον θα ήταν ο συνδυασμός τους σε μία κοινή πλατφόρμα, η οποία θα έχει ως στόχο τον αυτοματοποιημένο σχολιασμό πρωτεωμάτων, για αυτούς τους τύπους πρωτεϊνών. Επιπλέον, στη μέθοδο GprotPRED είναι σκόπιμη η ενσωμάτωση, όταν αυτό καταστεί δυνατό, ειδικών και για άλλες οικογένειες Gα υπομονάδων pHMMs.

Η συνεχής ανανέωση και βελτίωση των δύο βάσεων δεδομένων είναι εξέχουσας σημασίας. Παρά την επίπονη διαδικασία εξόρυξης των πληροφοριών που προέρχονται από τη βιβλιογραφία, είναι σημαντικό οι συλλογές δεδομένων να είναι επίκαιρες.

Τα δομικά δεδομένα που αφορούν τόσο τους GPCRs, όσο και τις G-πρωτεΐνες, αλλά και σύμπλοκα, κυρίως των δεύτερων με άλλα μόρια, συνεχώς αυξάνονται. Η μελέτη τους μπορεί να προσφέρει καλύτερη κατανόηση για τους μηχανισμούς μεταγωγής σήματος και την ακριβή λειτουργία των μορίων αυτών στο κύτταρο. Επιπλέον, λόγω της μεγάλης εμπλοκής των συγκεκριμένων μορίων σε ασθένειες, ίσως καταστεί δυνατή η αποσαφήνιση της αιτιολογίας διαφόρων ασθενειών.

Τέλος, όπως προαναφέραμε, η σημασία των μονονουκλεοτιδικών πολυμορφισμών και ο συσχετισμός τους με ασθένειες αποτελεί ένα σχετικά καινούριο πεδίο έρευνας και η μελέτη αυτών μπορεί να αποκαλύψει σημαντικές πληροφορίες όχι μόνο γύρω από την παθογένεια και τη θεραπεία ασθενειών αλλά και να αποκαλύψει πληροφορία σχετικά με τη λειτουργικότητα των πρωτεϊνών και πως αυτή καθορίζεται από τη σύσταση αλλά και τη δομή της πρωτεΐνης. Όσο τα δεδομένα σε αυτό το πεδίο αυξάνουν, τόσο πιο επιτακτική γίνεται η ανάγκη εκτενούς ανάλυσης και μελέτης τους.

Βιβλιογραφία

- Alberts, B., *et al.* (2007) Mechanisms of Cell Communication. In, *Molecular Biology of the Cell*.
- Almen, M.S., *et al.* (2009) Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin, *BMC biology*, **7**, 50.
- Altschul, S.F., *et al.* (1990) Basic local alignment search tool, *J Mol Biol*, **215**, 403-410.
- Altschul, S.F., *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res*, **25**, 3389-3402.
- Anantharaman, V., *et al.* (2011) Comparative genomics uncovers novel structural and functional features of the heterotrimeric GTPase signaling system, *Gene*, **475**, 63-78.
- Antunes, L.C. and Ferreira, R.B. (2009) Intercellular communication in bacteria, *Crit Rev Microbiol*, **35**, 69-80.
- Aris, L., *et al.* (2001) Structural requirements for the stabilization of metarhodopsin II by the C terminus of the alpha subunit of transducin, *J Biol Chem*, **276**, 2333-2339.
- Attwood, T.K. and Findlay, J.B. (1994) Fingerprinting G-protein-coupled receptors, *Protein Eng*, **7**, 195-203.
- Aubry, L. and Firtel, R.A. (1998) Spalten, a protein containing Galpha-protein-like and PP2C domains, is essential for cell-type differentiation in Dictyostelium, *Genes & development*, **12**, 1525-1538.
- Azpiazu, I., *et al.* (1999) A G protein gamma subunit-specific peptide inhibits muscarinic receptor signaling, *J Biol Chem*, **274**, 35305-35308.
- Bürglin, T.R. (2008) The Hedgehog protein family, *Genome biology*, **9**, 241.
- Bae, H., *et al.* (1997) Molecular determinants of selectivity in 5-hydroxytryptamine1B receptor-G protein interactions, *J Biol Chem*, **272**, 32071-32077.

- Bae, H., *et al.* (1999) Two amino acids within the alpha4 helix of Galphai1 mediate coupling with 5-hydroxytryptamine1B receptors, *J Biol Chem*, **274**, 14963-14971.
- Bagos, P.G., Liakopoulos, T.D. and Hamodrakas, S.J. (2006) Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins, *BMC bioinformatics*, **7**, 189.
- Baker, N.A., *et al.* (2001) Electrostatics of nanosystems: application to microtubules and the ribosome, *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 10037-10041.
- Baldi, P., *et al.* (2000) Assessing the accuracy of prediction algorithms for classification: an overview, *Bioinformatics*, **16**, 412-424.
- Barton, G.J. (1990) Protein multiple sequence alignment and flexible pattern matching, *Methods Enzymol*, **183**, 403-428.
- Bashford, D., Chothia, C. and Lesk, A.M. (1987) Determinants of a protein fold. Unique features of the globin amino acid sequences, *J Mol Biol*, **196**, 199-216.
- Bastiani, C. and Mendel, J. (2006) Heterotrimeric G proteins in *C. elegans*, *WormBook*, 1-25.
- Benton, M.J. and Ayala, F.J. (2003) Dating the tree of life, *Science*, **300**, 1698-1700.
- Berg JM, T.J., Stryer L (2002) *Biochemistry*. New York: W H Freeman. 2002
- Berman, H.M., *et al.* (2000) The Protein Data Bank, *Nucleic Acids Res*, **28**, 235-242.
- Bernsel, A., *et al.* (2008) Prediction of membrane-protein topology from first principles, *Proc Natl Acad Sci U S A*, **105**, 7177-7181.
- Bernsel, A., *et al.* (2009) TOPCONS: consensus prediction of membrane protein topology, *Nucleic Acids Res*, **37**, W465-468.
- Bernsel, A. and Von Heijne, G. (2005) Improved membrane protein topology prediction by domain assignments, *Protein Science*, **14**, 1723-1728.

- Bjarnadottir, T.K., *et al.* (2006) Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse, *Genomics*, **88**, 263-273.
- Boeckmann, B., *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res*, **31**, 365-370.
- Boto, T., Gomez-Diaz, C. and Alcorta, E. (2010) Expression analysis of the 3 G-protein subunits, Galpha, Gbeta, and Ggamma, in the olfactory receptor organs of adult *Drosophila melanogaster*, *Chem Senses*, **35**, 183-193.
- Cabrera-Vera, T.M., *et al.* (2003) Insights into G protein structure, function, and regulation, *Endocr. Rev.*, **24**, 765-781.
- Cabrera-Vera, T.M., *et al.* (2003) Insights into G protein structure, function, and regulation, *Endocr Rev*, **24**, 765-781.
- Cai, K., Itoh, Y. and Khorana, H.G. (2001) Mapping of contact sites in complex formation between transducin and light-activated rhodopsin by covalent crosslinking: use of a photoactivatable reagent, *Proc Natl Acad Sci U S A*, **98**, 4877-4882.
- Chen, C.A. and Manning, D.R. (2001) Regulation of G proteins by covalent modification, *Oncogene*, **20**, 1643-1652.
- Chen, Z., *et al.* (2005) Structure of the p115RhoGEF rgRGS domain-Galpha13/i1 chimera complex suggests convergent evolution of a GTPase activator, *Nat. Struct. Mol. Biol.*, **12**, 191-197.
- Choe, H.W., *et al.* (2011) Crystal structure of metarhodopsin II, *Nature*, **471**, 651-655.
- Churchill, G.A. (1989) Stochastic models for heterogeneous DNA sequences, *Bull Math Biol*, **51**, 79-94.
- Clamp, M., *et al.* (2004) The Jalview Java alignment editor, *Bioinformatics*, **20**, 426-427.
- Coleman, D.E., *et al.* (1994) Structures of active conformations of Gi alpha 1 and the mechanism of GTP hydrolysis, *Science*, **265**, 1405-1412.

Conant, G.C. and Wolfe, K.H. (2008) Turning a hobby into a job: how duplicated genes find new functions, *Nature reviews. Genetics*, **9**, 938-950.

Cooper GM, H.R. (2007) *The Cell: A Molecular Approach*. ASM Press – Sinauer Associates, Inc.

Davies, M.N., *et al.* (2007) Proteomic applications of automated GPCR classification, *Proteomics*, **7**, 2800-2814.

de Vries, S.J., van Dijk, M. and Bonvin, A.M. (2010) The HADDOCK web server for data-driven biomolecular docking, *Nature protocols*, **5**, 883-897.

DeLano, W.L. (2002) *The Pymol Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.

Dessauer, C.W., *et al.* (1998) Identification of a G α binding site on type V adenylyl cyclase, *The Journal of biological chemistry*, **273**, 25831-25839.

Dolinsky, T.J., *et al.* (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic. Acids. Res.*, **35**, W522-525.

Dolinsky, T.J., *et al.* (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations, *Nucleic. Acids. Res.*, **32**, W665-667.

Dorsam, R.T. and Gutkind, J.S. (2007) G-protein-coupled receptors and cancer, *Nature reviews. Cancer*, **7**, 79-94.

Downes, G.B. and Gautam, N. (1999) The G protein subunit gene families, *Genomics*, **62**, 544-552.

Dratz, E.A., *et al.* (1993) NMR structure of a receptor-bound G-protein peptide, *Nature*, **363**, 276-281.

Durbin, R., *et al.* (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids* Cambridge University Press, Cambridge, UK.

Eddy, S. (2006) HMMER User's Guide. Biological sequence analysis using profile hidden Markov models.

- Eddy, S.R. (1995) Multiple alignment using hidden Markov models, *Proc Int Conf Intell Syst Mol Biol*, **3**, 114-120.
- Eddy, S.R. (1998) Profile hidden Markov models, *Bioinformatics*, **14**, 755-763.
- Eddy, S.R. (2004) What is a hidden Markov model?, *Nat Biotechnol*, **22**, 1315-1316.
- Eddy, S.R. (2011) Accelerated Profile HMM Searches, *PLoS Comput Biol*, **7**, e1002195.
- Elefsinioti, A.L., *et al.* (2004) A database for G proteins and their interaction with GPCRs, *BMC bioinformatics*, **5**, 208.
- Elion, E.A. (2000) Pheromone response, mating and cell biology, *Current opinion in microbiology*, **3**, 573-581.
- Feng, D.F., Cho, G. and Doolittle, R.F. (1997) Determining divergence times with a protein clock: update and reevaluation, *Proc Natl Acad Sci U S A*, **94**, 13028-13033.
- Filipek, S., *et al.* (2003) G protein-coupled receptor rhodopsin: a prospectus, *Annual review of physiology*, **65**, 851-879.
- Ford, C.E., *et al.* (1998) Molecular basis for interactions of G protein betagamma subunits with effectors, *Science*, **280**, 1271-1274.
- Fredriksson, R., *et al.* (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints, *Mol Pharmacol*, **63**, 1256-1272.
- Fredriksson, R. and Schioth, H.B. (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes, *Mol Pharmacol*, **67**, 1414-1425.
- Fujisawa, Y., Kato, H. and Iwasaki, Y. (2001) Structure and function of heterotrimeric G proteins in plants, *Plant & cell physiology*, **42**, 789-794.
- Gether, U. (2000) Uncovering molecular mechanisms involved in activation of G protein-coupled receptors, *Endocrine reviews*, **21**, 90-113.

Gribskov, M., Luthy, R. and Eisenberg, D. (1990) Profile analysis, *Methods Enzymol*, **183**, 146-159.

Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) Profile analysis: detection of distantly related proteins, *Proc Natl Acad Sci U S A*, **84**, 4355-4358.

Grishina, G. and Berlot, C.H. (2000) A surface-exposed region of G(salpa) in which substitutions decrease receptor-mediated activation and increase receptor affinity, *Mol Pharmacol*, **57**, 1081-1092.

Hamm, H.E., *et al.* (1988) Site of G protein binding to rhodopsin mapped with synthetic peptides from the alpha subunit, *Science*, **241**, 832-835.

Henikoff, S. (1996) Scores for sequence searches and alignments, *Curr Opin Struct Biol*, **6**, 353-360.

Henikoff, S., *et al.* (1997) Gene families: the taxonomy of protein paralogs and chimeras, *Science*, **278**, 609-614.

Hennerdal, A. and Elofsson, A. (2011) Rapid membrane protein topology prediction, *Bioinformatics*, **27**, 1322-1323.

Ho, M.K. and Wong, Y.H. (2000) The amino terminus of Galpha(z) is required for receptor recognition, whereas its alpha4/beta6 loop is essential for inhibition of adenylyl cyclase, *Mol Pharmacol*, **58**, 993-1000.

Hoffmann, K.F., *et al.* (2001) The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni*, *Molecular and biochemical parasitology*, **112**, 113-123.

Hollmann, M.W., *et al.* (2005) Receptors, G proteins, and their interactions, *Anesthesiology*, **103**, 1066-1078.

Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic. Acids. Res.*, **38**, W545-549.

- Hooper, S.D. and Bork, P. (2005) Medusa: a simple tool for interaction graph analysis, *Bioinformatics*, **21**, 4432-4433.
- Horn, F., *et al.* (2003) GPCRDB information system for G protein-coupled receptors, *Nucleic acids research*, **31**, 294-297.
- Huang, C.C. and Tesmer, J.J. (2011) Recognition in the face of diversity: interactions of heterotrimeric G proteins and G protein-coupled receptor (GPCR) kinases with activated GPCRs, *The Journal of biological chemistry*, **286**, 7715-7721.
- Huang, Y., *et al.* (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences, *Bioinformatics*, **26**, 680-682.
- Ishii, I., *et al.* (1997) Alanine exchanges of polar amino acids in the transmembrane domains of a platelet-activating factor receptor generate both constitutively active and inactive mutants, *J Biol Chem*, **272**, 7846-7854.
- Itoh, Y., Cai, K. and Khorana, H.G. (2001) Mapping of contact sites in complex formation between light-activated rhodopsin and transducin by covalent crosslinking: use of a chemically preactivated reagent, *Proc Natl Acad Sci U S A*, **98**, 4883-4887.
- Janetopoulos, C., Jin, T. and Devreotes, P. (2001) Receptor-mediated activation of heterotrimeric G-proteins in living cells, *Science*, **291**, 2408-2411.
- Janz, J.M. and Farrens, D.L. (2004) Rhodopsin activation exposes a key hydrophobic binding site for the transducin alpha-subunit C terminus, *J Biol Chem*, **279**, 29767-29773.
- Jaynes, E.T. (1995) *Probability Theory: The Logic of Science*.
- Jiang, Q., *et al.* (1997) A mutational analysis of residues essential for ligand recognition at the human P2Y1 receptor, *Mol Pharmacol*, **52**, 499-507.
- Jiang, Q., *et al.* (1997) Mutagenesis reveals structure-activity parallels between human A2A adenosine receptors and biogenic amine G protein-coupled receptors, *Journal of medicinal chemistry*, **40**, 2588-2595.

- Johnston, C.A., *et al.* (2008) A sweet cycle for Arabidopsis G-proteins: Recent discoveries and controversies in plant G-protein signal transduction, *Plant signaling & behavior*, **3**, 1067-1076.
- Jones, D.T. (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics*, **23**, 538-544.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology, *Biochemistry*, **33**, 3038-3049.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, **22**, 2577-2637.
- Kall, L., Krogh, A. and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal peptide prediction method, *Journal of molecular biology*, **338**, 1027-1036.
- Kall, L., Krogh, A. and Sonnhammer, E.L. (2005) An HMM posterior decoder for sequence feature prediction that includes homology information, *Bioinformatics*, **21 Suppl 1**, i251-257.
- Kamesh, N., Aradhyam, G.K. and Manoj, N. (2008) The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*, *BMC evolutionary biology*, **8**, 129.
- Kenakin, T. (2004) Principles: receptor theory in pharmacology, *Trends Pharmacol Sci*, **25**, 186-192.
- Khelashvili, G., *et al.* (2010) GPCR-OKB: the G Protein Coupled Receptor Oligomer Knowledge Base, *Bioinformatics*, **26**, 1804-1805.
- Kim, J.Y., Haastert, P.V. and Devreotes, P.N. (1996) Social senses: G-protein-coupled receptor signaling pathways in *Dictyostelium discoideum*, *Chemistry & biology*, **3**, 239-243.
- Kisselev, O., *et al.* (1995) Receptor-G protein coupling is established by a potential conformational switch in the beta gamma complex, *Proc Natl Acad Sci U S A*, **92**, 9102-9106.
- Kisselev, O.G., *et al.* (1998) Light-activated rhodopsin induces structural binding motif in G protein alpha subunit, *Proc Natl Acad Sci U S A*, **95**, 4270-4275.

Koenig, B.W., *et al.* (2002) Structure and orientation of a G protein fragment in the receptor bound state from residual dipolar couplings, *J Mol Biol*, **322**, 441-461.

Kokkola, T., *et al.* (2003) Important amino acids for the function of the human MT1 melatonin receptor, *Biochem Pharmacol*, **65**, 1463-1471.

Kolakowski, L.F., Jr. (1994) GCRDb: a G-protein-coupled receptor database, *Receptors Channels*, **2**, 1-7.

Kristiansen, K. (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function, *Pharmacol. Ther.*, **103**, 21-80.

Krogh, A. (1994) Hidden Markov models for labelled sequences. *12th IAPR International Conference on Pattern Recognition*. pp. 140-144.

Krogh, A., *et al.* (1994) Hidden Markov models in computational biology. Applications to protein modeling, *J Mol Biol*, **235**, 1501-1531.

Krogh, A., *et al.* (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, *J Mol Biol*, **305**, 567-580.

Lagerstrom, M.C. and Schioth, H.B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery, *Nat Rev Drug Discov*, **7**, 339-357.

Lambright, D.G., *et al.* (1994) Structural determinants for activation of the alpha-subunit of a heterotrimeric G protein, *Nature*, **369**, 621-628.

Lambright, D.G., *et al.* (1996) The 2.0 Å crystal structure of a heterotrimeric G protein, *Nature*, **379**, 311-319.

Langer, I., *et al.* (2003) Lysine 195 and aspartate 196 in the first extracellular loop of the VPAC1 receptor are essential for high affinity binding of agonists but not of antagonists, *Neuropharmacology*, **44**, 125-131.

Larkin, M.A., *et al.* (2007) Clustal W and Clustal X version 2.0, *Bioinformatics*, **23**, 2947-2948.

Liapakis, G., *et al.* (2000) The forgotten serine. A critical role for Ser-2035.42 in ligand binding to and activation of the beta 2-adrenergic receptor, *J Biol Chem*, **275**, 37779-37788.

Lichtarge, O., Bourne, H.R. and Cohen, F.E. (1996) Evolutionarily conserved Galphabeta gamma binding surfaces support a model of the G protein-receptor complex, *Proc Natl Acad Sci U S A*, **93**, 7507-7511.

Lin, S.H. and Civelli, O. (2004) Orphan G protein-coupled receptors: targets for new therapeutic interventions, *Ann Med*, **36**, 204-214.

Lopez, R., *et al.* (2003) WU-Blast2 server at the European Bioinformatics Institute, *Nucleic Acids Res*, **31**, 3795-3798.

Luttrell, L.M. (2006) Transmembrane signaling by G protein-coupled receptors, *Methods Mol Biol*, **332**, 3-49.

Luttrell, L.M. (2008) Reviews in molecular biology and biotechnology: transmembrane signaling by G protein-coupled receptors, *Mol Biotechnol*, **39**, 239-264.

Macfarlane, S.R., *et al.* (2001) Proteinase-activated receptors, *Pharmacol Rev*, **53**, 245-282.

Malbon, C.C. (2005) G proteins in development, *Nat Rev Mol Cell Biol*, **6**, 689-701.

Man, O., Gilad, Y. and Lancet, D. (2004) Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons, *Protein science : a publication of the Protein Society*, **13**, 240-254.

Marygold, S.J., *et al.* (2013) FlyBase: improvements to the bibliography, *Nucleic Acids Res*, **41**, D751-757.

Mazzoni, M.R. and Hamm, H.E. (1996) Interaction of transducin with light-activated rhodopsin protects it from proteolytic digestion by trypsin, *J Biol Chem*, **271**, 30034-30040.

McCudden, C.R., *et al.* (2005) G-protein signaling: back to the future, *Cell. Mol. Life Sci.*, **62**, 551-577.

Menon, S.T., Han, M. and Sakmar, T.P. (2001) Rhodopsin: structural basis of molecular physiology, *Physiological reviews*, **81**, 1659-1688.

- Miettinen, H.M., *et al.* (1997) The ligand binding site of the formyl peptide receptor maps in the transmembrane region, *Journal of immunology*, **159**, 4045-4054.
- Mixon, M.B., *et al.* (1995) Tertiary and quaternary structural changes in Gi alpha 1 induced by GTP hydrolysis, *Science*, **270**, 954-960.
- Muller, G. (2000) Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach, *Curr Med Chem*, **7**, 861-888.
- Nemoto, W., Fukui, K. and Toh, H. (2009) GRIP: a server for predicting interfaces for GPCR oligomerization, *Journal of receptor and signal transduction research*, **29**, 312-317.
- Nishi, S., *et al.* (2002) The ectodomain of the luteinizing hormone receptor interacts with exoloop 2 to constrain the transmembrane region: studies using chimeric human and fly receptors, *J Biol Chem*, **277**, 3958-3964.
- Nishimura, A., *et al.* (2010) Structural basis for the specific inhibition of heterotrimeric Gq protein by a small molecule, *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 13666-13671.
- Noel, J.P., Hamm, H.E. and Sigler, P.B. (1993) The 2.2 Å crystal structure of transducin-alpha complexed with GTP gamma S, *Nature*, **366**, 654-663.
- Nordstrom, K.J., Fredriksson, R. and Schiöth, H.B. (2008) The amphioxus (*Branchiostoma floridae*) genome contains a highly diversified set of G protein-coupled receptors, *BMC evolutionary biology*, **8**, 9.
- Nordstrom, K.J., *et al.* (2011) Independent HHsearch, Needleman--Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families, *Molecular biology and evolution*, **28**, 2471-2480.
- Nugent, T. and Jones, D.T. (2009) Transmembrane protein topology prediction using support vector machines, *BMC bioinformatics*, **10**, 159.
- Oka, Y., *et al.* (2009) The fifth class of Galpha proteins, *Proc Natl Acad Sci U S A*, **106**, 1484-1489.

Oldham, W.M. and Hamm, H.E. (2006) Structural basis of function in heterotrimeric G proteins, *Q Rev Biophys*, **39**, 117-166.

Oldham, W.M. and Hamm, H.E. (2008) Heterotrimeric G protein activation by G-protein-coupled receptors, *Nat. Rev. Mol. Cell. Biol.*, **9**, 60-71.

Omasits, U., *et al.* (2013) Protter: interactive protein feature visualization and integration with experimental proteomic data, *Bioinformatics*.

Ono, Y., Fujibuchi, W. and Suwa, M. (2005) Automatic gene collection system for genome-scale overview of G-protein coupled receptors in eukaryotes, *Gene*, **364**, 63-73.

Onrust, R., *et al.* (1997) Receptor and betagamma binding sites in the alpha subunit of the retinal G protein transducin, *Science*, **275**, 381-384.

Pagani, I., *et al.* (2012) The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata, *Nucleic Acids Res*, **40**, D571-579.

Palczewski, K., *et al.* (2000) Crystal structure of rhodopsin: A G protein-coupled receptor, *Science*, **289**, 739-745.

Papasaikas, P.K., *et al.* (2003) A novel method for GPCR recognition and family classification from sequence alone using signatures derived from profile hidden Markov models, *SAR and QSAR in environmental research*, **14**, 413-420.

Papasaikas, P.K., *et al.* (2004) PRED-GPCR: GPCR recognition and family classification server, *Nucleic acids research*, **32**, W380-382.

Pavlopoulos, G.A., *et al.* (2008) Arena3D: visualization of biological networks in 3D, *BMC systems biology*, **2**, 104.

Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison, *Proc Natl Acad Sci U S A*, **85**, 2444-2448.

Perlman, J.H., *et al.* (1996) A refined model of the thyrotropin-releasing hormone (TRH) receptor binding pocket. Experimental analysis and energy minimization of the complex between TRH and TRH receptor, *Biochemistry*, **35**, 7643-7650.

- Petersen, T.N., *et al.* (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature Methods*, **8**, 785-786.
- Peterson, K.J. and Butterfield, N.J. (2005) Origin of the Eumetazoa: testing ecological predictions of molecular clocks against the Proterozoic fossil record, *Proc Natl Acad Sci U S A*, **102**, 9547-9552.
- Pierce, K.L., Premont, R.T. and Lefkowitz, R.J. (2002) Seven-transmembrane receptors, *Nat Rev Mol Cell Biol*, **3**, 639-650.
- Pin, J.P., Galvez, T. and Prezeau, L. (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors, *Pharmacology & therapeutics*, **98**, 325-354.
- Pin, J.P., *et al.* (2007) International Union of Basic and Clinical Pharmacology. LXVII. Recommendations for the recognition and nomenclature of G protein-coupled receptor heteromultimers, *Pharmacological reviews*, **59**, 5-13.
- Porollo, A. and Meller, J. (2007) Prediction-based fingerprints of protein-protein interactions, *Proteins*, **66**, 630-645.
- Pruitt, K.D., *et al.* (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy, *Nucleic Acids Res*, **40**, D130-135.
- Punta, M., *et al.* (2012) The Pfam protein families database, *Nucleic acids research*, **40**, D290-301.
- Quan, F., Wolfgang, W.J. and Forte, M. (1993) A Drosophila G-protein alpha subunit, Gf alpha, expressed in a spatially and temporally restricted pattern during Drosophila development, *Proc Natl Acad Sci U S A*, **90**, 4236-4240.
- Rasmussen, S.G., *et al.* (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex, *Nature*, **477**, 549-555.
- Rasmussen, S.G.F., *et al.* (2011) Structure of a nanobody-stabilized active state of the [bgr]2 adrenoceptor, *Nature*, **469**, 175-180.

Reynolds, S.M., *et al.* (2008) Transmembrane topology and signal peptide prediction using dynamic bayesian networks, *PLoS Comput Biol*, **4**, e1000213.

Riggs, E.R., *et al.* (2013) Towards a Universal Clinical Genomics Database: the 2012 International Standards for Cytogenomic Arrays Consortium Meeting, *Human mutation*, **34**, 915-919.

Rompler, H., *et al.* (2007) G protein-coupled time travel: evolutionary aspects of GPCR research, *Molecular interventions*, **7**, 17-25.

Rosenbaum, D.M., Rasmussen, S.G. and Kobilka, B.K. (2009) The structure and function of G-protein-coupled receptors, *Nature*, **459**, 356-363.

Saitoh, O., *et al.* (1997) Molecular identification of a G protein-coupled receptor family which is expressed in planarians, *Gene*, **195**, 55-61.

Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J Mol Biol*, **234**, 779-815.

Satagopam, V.P., *et al.* (2010) GPCRs, G-proteins, effectors and their interactions: human-gpDB, a database employing visualization tools and data integration techniques, *Database : the journal of biological databases and curation*, **2010**, baq019.

Schaefer, C., *et al.* (2012) SNPdbe: constructing an nsSNP functional impacts database, *Bioinformatics*, **28**, 601-602.

Scheerer, P., *et al.* (2008) Crystal structure of opsin in its G-protein-interacting conformation, *Nature*, **455**, 497-502.

Schioth, H.B. and Fredriksson, R. (2005) The GRAFS classification system of G-protein coupled receptors in comparative perspective, *Gen Comp Endocrinol*, **142**, 94-101.

Schmidt, C.J., *et al.* (1992) Specificity of G protein beta and gamma subunit interactions, *J Biol Chem*, **267**, 13807-13810.

- Sgourakis, N.G., Bagos, P.G. and Hamodrakas, S.J. (2005) Prediction of the coupling specificity of GPCRs to four families of G-proteins using hidden Markov models and artificial neural networks, *Bioinformatics*, **21**, 4101-4106.
- Sgourakis, N.G., *et al.* (2005) A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile Hidden Markov Models, *BMC bioinformatics*, **6**, 104.
- Sharman, J.L., *et al.* (2012) IUPHAR-DB: updated database content and new features, *Nucleic acids research*.
- Sherry, S.T., *et al.* (2001) dbSNP: the NCBI database of genetic variation, *Nucleic Acids Res*, **29**, 308-311.
- Siehl S, M.G. (2011) *G Protein-Coupled Receptors: Structure, Signalling and Physiology*. Cambridge University Press.
- Skrabaneck, L., *et al.* (2007) Requirements and ontology for a G protein-coupled receptor oligomerization knowledge base, *BMC bioinformatics*, **8**, 177.
- Slep, K.C., *et al.* (2001) Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 Å, *Nature*, **409**, 1071-1077.
- Smith, T.F. and Waterman, M.S. (1981) Identification of common molecular subsequences, *J Mol Biol*, **147**, 195-197.
- Smotrys, J.E. and Linder, M.E. (2004) Palmitoylation of intracellular signaling proteins: regulation and function, *Annu Rev Biochem*, **73**, 559-587.
- Sondek, J., *et al.* (1996) Crystal structure of a G-protein beta gamma dimer at 2.1Å resolution, *Nature*, **379**, 369-374.
- Song, Z.H. and Bonner, T.I. (1996) A lysine residue of the cannabinoid receptor is critical for receptor recognition by several agonists but not WIN55212-2, *Mol Pharmacol*, **49**, 891-896.
- Soppa, J. (1994) Two hypotheses--one answer. Sequence comparison does not support an evolutionary link between halobacterial retinal proteins including bacteriorhodopsin and eukaryotic G-protein-coupled receptors, *FEBS Lett*, **342**, 7-11.

- Soundararajan, M., *et al.* (2008) Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits, *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 6457-6462.
- Sprang, S.R. (1997) G protein mechanisms: insights from structural analysis, *Annu Rev Biochem*, **66**, 639-678.
- Srivastava, P.K., *et al.* (2007) HMM-ModE--improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences, *BMC bioinformatics*, **8**, 104.
- Stevens, R.C., *et al.* (2013) The GPCR Network: a large-scale collaboration to determine human GPCR structure and function, *Nat Rev Drug Discov*, **12**, 25-34.
- Strader, C.D., *et al.* (1989) Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor, *J Biol Chem*, **264**, 13572-13578.
- Strader, C.D., *et al.* (1987) Identification of residues required for ligand binding to the beta-adrenergic receptor, *Proc Natl Acad Sci U S A*, **84**, 4384-4388.
- Strotmann, R., *et al.* (2011) Evolution of GPCR: change and continuity, *Molecular and cellular endocrinology*, **331**, 170-178.
- Suwa, M. and Ono, Y. (2009) Computational overview of GPCR gene universe to support reverse chemical genomics study, *Methods in Molecular Biology*, **577**, 41-54.
- Swarbreck, D., *et al.* (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation, *Nucleic Acids Res*, **36**, D1009-1014.
- Tamai, K., *et al.* (2000) LDL-receptor-related proteins in Wnt signal transduction, *Nature*, **407**, 530-535.
- Taylor, J.M., *et al.* (1994) Binding of an alpha 2 adrenergic receptor third intracellular loop peptide to G beta and the amino terminus of G alpha, *J Biol Chem*, **269**, 27618-27624.
- Taylor, J.M., *et al.* (1996) Receptor and membrane interaction sites on Gbeta. A receptor-derived peptide binds to the carboxyl terminus, *J Biol Chem*, **271**, 3336-3339.

- Taylor, W.R. (1986) Identification of protein sequence homology by consensus template alignment, *J Mol Biol*, **188**, 233-258.
- Temple, B.R., Jones, C.D. and Jones, A.M. (2010) Evolution of a signaling nexus constrained by protein interfaces and conformational States, *PLoS Comput Biol*, **6**, e1000962.
- Theodoropoulou, M.C., *et al.* (2008) gpDB: a database of GPCRs, G-proteins, effectors and their interactions, *Bioinformatics*, **24**, 1471-1472.
- Tierney, A.J. (2001) Structure and function of invertebrate 5-HT receptors: a review, *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, **128**, 791-804.
- Tota, M.R. and Strader, C.D. (1990) Characterization of the binding domain of the beta-adrenergic receptor with the fluorescent antagonist carazolol. Evidence for a buried ligand binding site, *J Biol Chem*, **265**, 16891-16897.
- Tsirigos, K.D., *et al.* (2012) A guideline to proteome-wide alpha-helical membrane protein topology predictions, *Proteomics*, **12**, 2282-2294.
- Tusnady, G.E., Dosztanyi, Z. and Simon, I. (2004) Transmembrane proteins in the Protein Data Bank: identification and classification, *Bioinformatics*, **20**, 2964-2972.
- Tusnady, G.E., Dosztanyi, Z. and Simon, I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank, *Nucleic acids research*, **33**, D275-278.
- Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server, *Bioinformatics*, **17**, 849-850.
- UniProt, C. (2010) The Universal Protein Resource (UniProt) in 2010, *Nucleic Acids Res*, **38**, D142-148.
- UniProt, C. (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt), *Nucleic Acids Res*, **40**, D71-75.

UniProt, C. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013, *Nucleic Acids Res*, **41**, D43-47.

UniProt_Consortium (2010) The Universal Protein Resource (UniProt) in 2010, *Nucleic Acids Res*, **38**, D142-148.

Unni, S., *et al.* (2011) Web servers and services for electrostatics calculations with APBS and PDB2PQR, *Journal of computational chemistry*, **32**, 1488-1491.

Unson, C.G. (2002) Molecular determinants of glucagon receptor signaling, *Biopolymers*, **66**, 218-235.

van der Horst, E., *et al.* (2010) A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: a potential strategy for receptor de-orphanization, *BMC bioinformatics*, **11**, 316.

Vander AJ, S.J., Luciano DS (2001) *Human Physiology: The Mechanisms of Body Function*. McGraw-Hill.

Viklund, H. and Elofsson, A. (2004) Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information, *Protein science : a publication of the Protein Society*, **13**, 1908-1917.

Viklund, H. and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar, *Bioinformatics*, **24**, 1662-1668.

von Heijne, G. (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule, *J Mol Biol*, **225**, 487-494.

von Heijne, G. (1999) Recent advances in the understanding of membrane protein assembly and structure, *Q Rev Biophys*, **32**, 285-307.

Vu, T.K., *et al.* (1991) Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation, *Cell*, **64**, 1057-1068.

- Waldo, G.L., *et al.* (2010) Kinetic scaffolding mediated by a phospholipase C-beta and Gq signaling complex, *Science*, **330**, 974-980.
- Wall, M.A., *et al.* (1995) The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2, *Cell*, **83**, 1047-1058.
- Waterhouse, A.M., *et al.* (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics*, **25**, 1189-1191.
- Wauson, E.M., Lorente-Rodriguez, A. and Cobb, M.H. (2013) Minireview: nutrient sensing by g protein-coupled receptors, *Molecular endocrinology*, **27**, 1188-1197.
- Wilkie, T.M., *et al.* (1992) Evolution of the mammalian G protein alpha subunit multigene family, *Nat Genet*, **1**, 85-91.
- Wistrand, M., Kall, L. and Sonnhammer, E.L. (2006) A general model of G protein-coupled receptor sequences and its application to detect remote homologs, *Protein science : a publication of the Protein Society*, **15**, 509-521.
- Woehler, A. and Ponimaskin, E.G. (2009) G protein--mediated signaling: same receptor, multiple effectors, *Curr. Mol. Pharmacol.*, **2**, 237-248.
- Wonerow, P., *et al.* (2001) Thyrotropin receptor mutations as a tool to understand thyrotropin receptor action, *Journal of molecular medicine*, **79**, 707-721.
- Xu, F., *et al.* (2011) Structure of an agonist-bound human A2A adenosine receptor, *Science*, **332**, 322-327.
- Yabuki, Y., *et al.* (2005) GRIFFIN: a system for predicting GPCR-G-protein coupling selectivity using a support vector machine and a hidden Markov model, *Nucleic Acids Res*, **33**, W148-153.
- Yohannan, S., *et al.* (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors, *Proc Natl Acad Sci U S A*, **101**, 959-963.
- Zhang, F.L. and Casey, P.J. (1996) Protein prenylation: molecular mechanisms and functional consequences, *Annu Rev Biochem*, **65**, 241-269.

Zhang, J. and Zhang, Y. (2010) GPCRRD: G protein-coupled receptor spatial restraint database for 3D structure modeling and function annotation, *Bioinformatics*, **26**, 3004-3005.

Zhu, H., *et al.* (2009) Arabidopsis extra large G-protein 2 (XLG2) interacts with the Gbeta subunit of heterotrimeric G protein and functions in disease resistance, *Molecular plant*, **2**, 513-525.

Παράρτημα I

Πίνακας I.1 Οι προσδιορισμένες δομές GPCRs στην PDB

Όνομα δομής	PDB	PubMed	Αλυσ.	Res. (Å)	R-value	Σχόλια
Ροδοψίνη Βοδιού (Bovine Rhodopsin)	Κατηγορία: A					UniProt: P02699
CRYSTAL STRUCTURE OF BOVINE RHODOPSIN	1F88	10926528	A, B	2.80	0.186	1 ^η δομή GPCR
CRYSTAL STRUCTURE OF BOVINE RHODOPSIN	1HZX	11425302	A, B	2.80	0.175	
Structure of bovine rhodopsin (dark adapted)	1JFP	11570894	A	NMR	NMR	
Crystal structure of bovine rhodopsin at 2.6 angstroms RESOLUTION	1L9H	11972040	A, B	2.60	0.188	
STRUCTURE OF BOVINE RHODOPSIN (Metarhodopsin II)	1LN6	12044163	A	NMR	NMR	Πρώτη δομή Μεταροδοψίνης II
STRUCTURE OF BOVINE RHODOPSIN IN A TRIGONAL CRYSTAL FORM	1GZM	15491621	A, B	2.65	0.201	Δισουλφιδικός δεσμός μεταξύ καταλοίπων 110-187
Crystal Structure of Bovine Rhodopsin at 2.2 Angstroms Resolution	1U19	15327956	A, B	2.20	0.200	
Crystallographic model of bathorhodopsin	2G87	16586416	A, B	2.60	0.178	all-trans μορφή στα 100 K
Crystallographic model of lumirhodopsin	2HPY	16908857	A, B	2.80	0.218	Δομή φωτοενεργού ενδιάμεσου
Crystal structure of rhombohedral crystal form of ground-state rhodopsin	2I35	17060607	A	3.80	0.375	
Crystal structure of trigonal crystal form of ground-state rhodopsin	2I36	17060607	A, B, C	4.10	0.374	Δομή λειτουργικού ολιγομερούς
Crystal structure of a photoactivated rhodopsin	2I37	17060607	A, B, C	4.15	0.378	Δομή ενεργοποιημένου λειτουργικού ολιγομερούς
CRYSTAL STRUCTURE OF A RHODOPSIN STABILIZING MUTANT EXPRESSED IN MAMMALIAN CELLS	2J4Y	17825322	A, B	3.40	0.290	Μεταλλαγές: N2C, D282C
Crystallographic model of 9-cis-rhodopsin	2PED	17449675	A, B	2.95	0.243	
Structure of ground-state bovine rhodopsin in a hexagonal crystal form	3C9L	18645239	A	2.65	0.190	
Structure of a mutant bovine rhodopsin in hexagonal crystal form	3C9M	18645239	A	3.40	0.176	Μεταλλαγές: N2C, D282C

Crystal Structure of Native Opsin: the G Protein-Coupled Receptor Rhodopsin in its Ligand-free State	3CAP	18563085	A, B	2.90	0.231	Δομη Οψίνης
Crystal structure of the active G-protein-coupled receptor opsin in complex with a C-terminal peptide derived from the Galpha subunit of transducin	3DQB	18818650	A	3.20	0.215	Δομή Οψίνης Σύμπλοκο με C-άκρο Gat
Crystal structure of bovine rhodopsin with beta-ionone	3OAX	20923672	A, B	2.60	0.224	
Crystal structure of the constitutively active E113Q,N2C, D282C Rhodopsin mutant with bount Galpha C-t Peptide	2X72	21389983	A	3.00	0.212	Δομή ενεργής Ροδοψίνης Σύμπλοκο με C-άκρο Gat Μεταλλαγές: E113Q,N2C, D282C
Crystal structure of Metarhodopsin II in complex with a C-terminal peptide derived from the Galpha subunit of transducin	3PQR	21389988	A	2.85	0.217	Δομή Μεταροδοψίνης II Σύμπλοκο με C-άκρο Gat
Crystal structure of Metarhodopsin II	3PXO	21389988	A	3.00	0.217	Δομή Μεταροδοψίνης II
Crystal structure of the light-activated constitutively active N2C, M257Y,D282C rhodopsin mutant in complex with a peptide resembling the C-terminus of the Galpha-protein subunit (GaCT)	4A4M	22198838	A	3.30	0.218	Δομή ενεργής Ροδοψίνης Σύμπλοκο με C-άκρο Gat/ανάλογο Μεταλλαγές: N2C, M257Y,D282C
Ροδοψίνη καλαμαριού (Squid Rhodopsin)						
			Κατηγορία: A		UniProt: P31356	
Crystal structure of squid rhodopsin	2Z1Y	18463093	A	3.70	0.302	Πρώτη δομή GPCR σε ασπόνδυλο
Crystal structure of squid rhodopsin	2Z73	18480818	A	2.50	0.188	
Crystal structure of the batho intermediate of squid rhodopsin	3AYM	21906602	A	2.80	0.295	
Crystal structure of squid isorhodopsin	3AYN	21906602	A	2.70	0.180	
Ανθρώπινος β2 αδρενεργικός υποδοχέας (Human β2 adrenergic Receptor)						
			Κατηγορία: A		UniProt: P07550	
High resolution crystal structure of human B2-adrenergic G protein-coupled receptor.	2RH1	17962520	A	2.40	0.198	Μεταλλαγές: N187E, C54T, C97A
Crystal structure of the human beta2	2R4R	17952055	A	3.40	0.226	

adrenoceptor						
Crystal structure of the human beta2 adrenoceptor	2R4S	17952055	A	3.40	0.217	
Cholesterol bound form of human beta2 adrenergic receptor	3D4S	18547522	A	2.80	0.230	Μεταλλαγές: E122W, N187E, C1054T, C1097A
Crystal structure of a Methylated beta2 Adrenergic Receptor-Fab complex	3KJ6	20054398	A	3.40	0.242	
Crystal structure of the human beta2 adrenergic receptor in complex with the inverse agonist ICI 118,551	3NY8	20669948	A	2.84	0.242	Μεταλλαγές: E122W, N187E, C1054T, C1097A
Crystal structure of the human beta2 adrenergic receptor in complex with a novel inverse agonist	3NY9	20669948	A	2.84	0.229	Μεταλλαγές: E122W, N187E, C1054T, C1097A
Crystal structure of the human beta2 adrenergic receptor in complex with the neutral antagonist alprenolol	3NYA	20669948	A	3.16	0.222	Μεταλλαγές: E122W, N187E, C1054T, C1097A
Irreversible Agonist-Beta2 Adrenoceptor Complex	3PDS	21228876	A	3.5	0.241	Δομή ενεργού υποδοχέα Μεταλλαγές: H93C, N187E, C265A
Structure of a nanobody-stabilized active state of the beta2 adrenoceptor	3POG	21228869	A	3.5	0.243	Δομή ενεργού υποδοχέα Μεταλλαγές: N187E
Crystal structure of the beta2 adrenergic receptor-Gs protein complex	3SN6	21772288	A	3.2	0.228	Δομή ενεργού υποδοχέα Μεταλλαγές: C54T, C97A, M96T, M98T, N187E Συμπλοκο με Gα-Gβγ
N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor	4GBR	23056231	A	3.99	0.268	Μεταλλαγές: M96T, M98T Παρουσία 3 ^{ου} ενδοκ. βρόχου
β1 αδρενεργικός υποδοχέας γαλοπούλας						
(Turkey β1 adrenergic Receptor)		Κατηγορία: A			UniProt: P07700	
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND CYANOPINDOLOL	2VT4	18594507	A, B, C, D	2.70	0.215	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND	2Y00	21228877	A, B	2.50	0.228	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C

BOUND PARTIAL AGONIST DOBUTAMINE (CRYSTAL DOB92)						
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND PARTIAL AGONIST DOBUTAMINE (CRYSTAL DOB102)	2Y01	21228877	A, B	2.60	0.228	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND AGONIST CARMOTEROL	2Y02	21228877	A, B	2.60	0.231	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND AGONIST ISOPRENALINE	2Y03	21228877	A, B	2.85	0.222	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND PARTIAL AGONIST SALBUTAMOL	2Y04	21228877	A, B	3.05	0.225	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND ANTAGONIST CARAZOLOL	2YCW	21540331	A, B	3.00	0.250	Μεταλλαγές: S68R, L116C, A227Y, A272K, S273R, R275T, K276S, A327F, M338F, A358C Παρουσία 3 ^{ου} ενδοκ. βρόχου
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND ANTAGONIST CYANOPINDOLOL	2YCX	21540331	A, B	3.25	0.274	Μεταλλαγές: S68R, L116C, A227Y, A272K, S273R, R275T, K276S, A327F, M338F, A358C Παρουσία 3 ^{ου} ενδοκ. βρόχου
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND ANTAGONIST CYANOPINDOLOL	2YCY	21540331	A, B	3.15	0.255	Μεταλλαγές: S68R, L116C, A227Y, A272K, S273R, R275T, K276S, A327F, M338F, A358C Παρουσία 3 ^{ου} ενδοκ. βρόχου
TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND ANTAGONIST IODOCYANOPINDOLOL	2Y CZ	21540331	A, B	3.65	0.250	Μεταλλαγές: S68R, L116C, A227Y, A272K, S273R, R275T, K276S, A327F, M338F, A358C Παρουσία 3 ^{ου} ενδοκ. βρόχου
Turkey beta1 adrenergic receptor with stabilising mutations and bound biased agonist bucindolol	4AMI	22579251	A, B,	3.20	0.244	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C

Turkey beta1 adrenergic receptor with stabilising mutations and bound biased agonist carvedilol	4AMJ	22579251	A, B	2.30	0.204	Μεταλλαγές: S68R, L116C, A227Y, A327F, M338F, A358C
Ανθρώπινος Α2α αδενোসινικός Υποδοχέας						
(Human A2α adenosine Receptor)		Κατηγορία: A			UniProt: P29274	
The 2.6 Å Crystal Structure of a Human A2A Adenosine Receptor bound to ZM241385.	3EML	18832607	A	2.60	0.198	
Agonist bound structure of the human adenosine A2a receptor	3QAK	21393508	A	2.71	0.220	Δομή ενεργού υποδοχέα
Thermostabilised Human A2A receptor with adenosine bound	2YDO	21593763	A	3.00	0.246	Δομή ενεργού υποδοχέα Πρόσδεση φυσικού υποκαταστάτη (αδενοσίνη) Μεταλλαγές: A89Q, A154N
Thermostabilised Human A2A receptor with NECA bound	2YDV	21593763	A	2.60	0.234	Δομή ενεργού υποδοχέα Μεταλλαγές: A89Q, A154N
Thermostabilised Adenosine A2A Receptor	3PWH	21885291	A	3.30	0.278	Μεταλλαγές: A54L, T88A, R107A, K122A, L202A, L235A, V239A, S277A
Thermostabilised adenosine A2A receptor in complex with XAC	3REY	21885291	A	3.31	0.300	Μεταλλαγές: A54L, T88A, R107A, K122A, L202A, L235A, V239A, S277A
Thermostabilised adenosine A2A receptor in complex with caffeine	3RFM	21885291	A	3.60	0.298	Μεταλλαγές: A54L, T88A, R107A, K122A, L202A, L235A, V239A, S277A
Crystal structure of human adenosine A2A receptor with an allosteric inverse-agonist antibody at 2.7 Å resolution	3VG9	22286059	A	2.70	0.208	Μεταλλαγές: N154Q
Crystal structure of human adenosine A2A receptor with an allosteric inverse-agonist antibody at 3.1 Å resolution	3VGA	22286059	A	3.10	0.199	Μεταλλαγές: N154Q
Thermostabilised Adenosine A2A receptor in complex with 6-(2,6-Dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine	3UZA	22220592	A	3.27	0.295	Μεταλλαγές: A54L T88A R107A K122A L202A L235A V239A S277A
Thermostabilised Adenosine A2A receptor in complex with 4-(3-	3UZC	22220592	A	3.34	0.282	Μεταλλαγές: A54L T88A R107A K122A L202A L235A V239A S277A

amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol						
Crystal structure of the chimeric protein of A2aAR-BRIL in complex with ZM241385 at 1.8A resolution	4E1Y	22798613	A	1.80	0.176	Χιμαιρική πρωτεΐνη Μεταλλαγές: M215W, H310I, R314L
Ανθρώπινος CXCR4 χημειοκινικός Υποδοχέας (Human CXCR4 Chemokine Receptor) Κατηγορία: A UniProt: P61073						
The 2.5 A structure of the CXCR4 chemokine receptor in complex with small molecule antagonist IT1t	3ODU	20929726	A, B	2.50	0.240	
Crystal structure of the CXCR4 chemokine receptor in complex with a cyclic peptide antagonist CVX15	3OE0	20929726	A	2.90	0.217	
Crystal structure of the CXCR4 chemokine receptor in complex with a small molecule antagonist IT1t in I222 spacegroup	3OE6	20929726	A	3.20	0.235	
Crystal structure of the CXCR4 chemokine receptor in complex with a small molecule antagonist IT1t in P1 spacegroup	3OE8	20929726	A, B, C	3.10	0.252	
Crystal structure of the chemokine CXCR4 receptor in complex with a small molecule antagonist IT1t in P1 spacegroup	3OE9	20929726	A, B	3.10	0.254	
Ανθρώπινος D3 ντοπαμινικός Υποδοχέας (Human D3 Dopamine Receptor) Κατηγορία: A UniProt: P35462						
Structure of the human dopamine D3 receptor in complex with eticlopride	3PBL	21097933	A, B	2.89	0.245	Μεταλλαγές: L119W, C1054T, C1097A
Ανθρώπινος H1 ισταμινικός Υποδοχέας (Human H1 Histamine Receptor) Κατηγορία: A UniProt: P35367						
Structure of the human histamine H1 receptor in complex with doxepin	3RZE	21697825	A	3.10	0.217	
Ανθρώπινος M2 μουσκαρινικός Υποδοχέας (Human M2 Muscarinic Receptor) Κατηγορία: A UniProt: P08172						
Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist	3UON	22278061	A	3.00	0.227	Μεταλλαγές: N2D, N3D, N6D, N9D, C54T, C97A
Ανθρώπινος M3 μουσκαρινικός Υποδοχέας						

(Human M3 Muscarinic Receptor) Κατηγορία: A UniProt: P08483						
Structure of the M3 Muscarinic Acetylcholine Receptor	4DAJ	22358844	A	3.40	0.254	Μεταλλαγές: C1054T, C1097A
Ανθρώπινος λιπιδικός S1P1 υποδοχέας (Human Lipid S1P1 Receptor) Κατηγορία: A UniProt: P21453						
Crystal Structure of a Lipid G protein-Coupled Receptor at 3.35A	3V2W	22344443	A	3.35	0.228	
Crystal Structure of a Lipid G protein-Coupled Receptor at 2.80A	3V2Y	22344443	A	2.80	0.231	
Ανθρώπινος κ οπιοειδής Υποδοχέας (Human κ opioid receptor) Κατηγορία: A UniProt: P41145						
Structure of the human kappa opioid receptor in complex with JD1c	4JDH	22437504	A, B	2.90	0.228	Μεταλλαγές: I135L, C54T, C97A
Ανθρώπινος μ οπιοειδής Υποδοχέας (Human μ opioid receptor) Κατηγορία: A UniProt: P42866						
Crystal structure of the μ-opioid receptor bound to a morphinan antagonist.	4DKL	22437502	A	2.80	0.235	
Ανθρώπινος δ οπιοειδής Υποδοχέας (Human δ opioid receptor) Κατηγορία: A UniProt: P32300						
Structure of the delta opioid receptor bound to naltrindole	4EJ4	22596164	A	3.40	0.255	
Ανθρώπινος N/OFQ οπιοειδής Υποδοχέας (Human N/OFQ opioid Receptor) Κατηγορία: A UniProt: P41146						
Structure of the N/OFQ Opioid Receptor in Complex with a Peptide Mimetic	4EA3	22596163	A, B	3.01	0.250	Μεταλλαγές: M29W, H124I, K128L Παρουσία 3 ^{ου} ενδοκ. βρόχου
Ανθρώπινος CXCR1 χημειοκινικός Υποδοχέας (Human CXCR1 Chemokine Receptor) Κατηγορία: A UniProt: P25024						
Structure of human CXCR1 in phospholipid bilayers	2LNL	23086146	A	NMR	NMR	Παρουσία 3 ^{ου} ενδοκ. βρόχου
NTS1 Νευροτενσινικός υποδοχέας Του αρουραίου (Rat NTS1 Neurotensin Receptor) Κατηγορία: A UniProt: P20789						
The crystal structure of the neurotensin receptor NTS1 in complex with neurotensin (8-13)	4GRV	23051748	A	2.80	0.228	Δομή ενεργοποιημένου υποδοχέα
PAR1 Ενεργοποιούμενος από πρωτεάση υποδοχέας 1 Του ανθρώπου (Human PAR1 Protease Activated Receptor) Κατηγορία: A UniProt: P25116						
Crystal structure of human protease-activated receptor 1 (PAR1) bound with antagonist vorapaxar at 2.2 angstrom	3VW7	23222541	A	2.20	0.219	Μεταλλαγές: N250G, N259S, D1020N, C1054T, C1097A
Ανθρώπινος 5-υδροξυτρυπταμινικός υποδοχέας 1B (Human 5-hydroxytryptamine 1B receptor (5-HT1B)) Κατηγορία: A UniProt: P28222						

Crystal structure of the chimeric protein of 5-HT1B-BRIL in complex with ergotamine	4IAR	23519210	A	2.70	0.225	Μεταλλαγές: L138W, M29W, H124I, R128L
Crystal structure of the chimeric protein of 5-HT1B-BRIL in complex with dihydroergotamine	4IAQ	23519210	A	2.80	0.237	Μεταλλαγές: L138W, M29W, H124I, R128L
Ανθρώπινος 5-υδροxyτρυπταμινικός υποδοχέας 2B (Human 5-hydroxytryptamine 2B receptor (5-HT2B)) Κατηγορία: A UniProt: P41595						
Crystal structure of the chimeric protein of 5-HT2B-BRIL in complex with ergotamine	4IB4	23519215	A	2.70	0.227	Μεταλλαγές: M144W, M29W, H124I, R128L
Ανθρώπινος CCR5 χημειοκινικός Υποδοχέας (Human CCR5 Chemokine Receptor) Κατηγορία: A UniProt: P51681						
Crystal Structure of the CCR5 Chemokine Receptor	4MBS	24030490	A, B	2.71	0.218	Μεταλλαγές: C58Y, G163N, A233D, K303E
Ανθρώπινος Υποδοχέας του παράγοντα απελευθέρωσης κορτικοτροπίνης Τύπου 1 (Human Corticotropin-releasing factor receptor 1) Κατηγορία: B UniProt: P34998						
Structure of class B GPCR corticotropin releasing factor receptor 1	4K5Y	23863939	A, B, C	2.98	0.243	Μεταλλαγές: V120A, L144A, W156A, S160A, N40S, A41V, C54S, C97S, T151A, K228A, F260A, I277A, Y309A, F330A, S349A, Y363A
Ανθρώπινος Υποδοχέας του Γλυκογόνου (Human Glucagon Receptor) Κατηγορία: B UniProt: P47871						
Structure of the class B human glucagon G protein coupled receptor	4L6R	23863937	A	3.30	0.286	Μεταλλαγές: M7W, H102I, R106L
Ανθρώπινος Υποδοχέας Smoothened (Human Smoothened Receptor) Κατηγορία: F UniProt: Q99835						
Structure of the human smoothened receptor bound to an antitumour agent.	4JKV	23636324	A, B	2.45	0.201	Μεταλλαγές: M7W, H101I, R107L

Πίνακας Ι.2 Οι προσδιορισμένες δομές Gα υπομονάδων στην PDB (Ιούνιος 2012).

Όνομα δομής	PDB	PubMed	Gα τύπος	Αλυσ.	Res. (Å)	R-value	Σχόλια
G-protein Heterotrimer Giα1 β1 γ2 with GDP bound	1GP2	8521505	Gαi1	A	2.30	0.226	→GDP bound →Gα – Gβγ complex
G-protein Heterotrimer Giα1 (G203A) β1 γ2 with GDP bound	1GG2	8521505	Gαi1	A	2.30	0.205	→GDP bound →Mutation: G203A →Gα – Gβγ complex
Structure of Active Conformations of Giα1 and the mechanism of GTP hydrolysis	1GIA	8073283	Gαi1	A	2.00	0.175	→GTPγS bound →Truncated N-terminus
Structure of Active Conformations of Giα1 and the mechanism of GTP hydrolysis	1GFI	8073283	Gαi1	A	2.20	0.214	→GDP-AIF4- bound →Truncated N-terminus
Structure of Active Conformations of Giα1 and the mechanism of GTP hydrolysis	1GIL	8073283	Gαi1	A	2.30	0.222	→GTPγS bound →Truncated N-terminus →Mutation: Q204L
Structure of GTP-binding Protein	1GIT	8939752	Gαi1	A	2.60	0.186	→GDP-Pi bound →Truncated N-terminus →Mutation: G203A
GTPγS bound G42V Giα1	1AS0	9398294	Gαi1	A	2.00	0.206	→GTPγS bound →Truncated N-terminus →Mutation: G42V
GDP+Pi bound G42V Giα1	1AS2	9398294	Gαi1	A	2.80	0.189	→GDP-Pi bound →Truncated N-terminus →Mutation: G42V
GDP bound G42V Giα1	1AS3	9398294	Gαi1	A	2.40	0.212	→GDP bound →Truncated N-terminus →Mutation: G42V
Complex of AIF4-activated Giα1 with RGS4	1AGR	9108480	Gαi1	A, D	2.80	0.216	→GDP-AIF4- bound →Gα – RGS complex
Giα1 bound to GDP and Magnesium	1BOF	9772163	Gαi1	A	2.20	0.227	→ GDP bound → Switch II, Switch III disordered
A326S mutant of an inhibitory α subunit	1BH2	9705312	Gαi1	A	2.10	0.190	→GTPγS bound →Truncated N-terminus →Mutation: A326S

Gi α 1 subunit of Guanine nucleotide-binding protein complexed with a GTP analogue	1CIP	10358003	Gai1	A	1.50	0.213	→GppNHp bound →Truncated N-terminus
Crystal Structure of Human Gai1 Bound to the GoLoco Motif of RGS14	1KJY	11976690	Gai1	A, C	2.70	0.238	→GDP bound →Truncated N-terminus →G α – RGS complex
Structure of the K180P mutant of Gi α subunit bound to AIF4 and GDP	1SVK	15128951	Gai1	A	2.00	0.197	→GDP-AIF4- bound →Truncated N-terminus →Mutation: K180P
Structure of the K180P mutant of Gi α subunit bound to GppNHp	1SVS	15128951	Gai1	A	1.50	0.200	→GppNHp bound →Truncated N-terminus →Mutation: K180P
Structure of Gai1 bound to a GDP-selective peptide provides insight into guanine nucleotide exchange	1Y3A	16004878	Gai1	A, B, C, D	2.50	0.255	→GDP bound →Truncated N-terminus →G α – peptide complex → Switch III disordered
Structure of activated Gai1 bound to a nucleotide-state-selective peptide: Minimal determinants for recognizing the active form of a G protein α subunit	2G83	16981699	Gai1	A, B	2.80	0.301	→GDP-AIF4- bound →Truncated N-terminus →G α – peptide complex
Crystal Structure Of Human Gai1 Bound To The Goloco Motif Of Rgs14	2OM2	17603074	Gai1	A, C	2.20	0.227	→GDP bound →Truncated N-terminus →G α – RGS complex
Crystal structure of the heterodimeric complex of human RGS1 and activated Gi α 1	2GTP	18434541	Gai1	A, B	2.55	0.228	→GDP-AIF4- bound →Truncated N-terminus →G α – RGS complex
Crystal structure of the heterodimeric complex of human RGS16 and activated Gi α 1	2IK8	18434541	Gai1	A, B	2.71	0.235	→GDP-AIF4- bound →Truncated N-terminus →G α – RGS complex
Mechanism underlying the critical contribution of a switch II residue in a heterotrimeric G-protein α subunit during <i>C. elegans</i> asymmetric cell division	2EBC	18519563	Gai1	A	2.24	0.229	→GDP bound → Switch II disordered → Mutation: G202D
Crystal Structure of a fast activating G protein	3FFA	19703466	Gai1	A	2.30	0.188	→GTP γ S bound

mutant							→Truncated N-terminus →Mutation: T329A
G α 1 mutant in GDP bound form	3FFB	19703466	G α i1	A	2.57	0.181	→GDP bound →Truncated N-terminus →Switch II, Switch III disordered →Mutation: T329A
Crystal Structure of the G Protein Fast-Exchange Double Mutant I56C/Q333C	3D7M	19222191	G α i1	A	2.90	0.249	→GDP-ALF4- bound →Truncated N-terminus →Mutation: I56C, Q333C
Structure of the K349P mutant of G α 1 subunit bound to ALF4 and GDP	2ZJY	-	G α i1	A	2.80	0.181	→GDP-ALF4- bound →Truncated N-terminus →Mutation: K349P
Structure of the K349P mutant of G α 1 subunit bound to GDP	2ZJZ	-	G α i1	A, B	2.60	0.221	→GDP bound →Truncated N-terminus →Mutation: K349P
Structure of a G α i1 mutant with enhanced affinity for the RGS14 GoLoco motif.	3ONW	21115486	G α i1	A, B	2.38	0.230	→GDP bound →Truncated N-terminus →Mutation: Q147L
Crystal structure of human G α i1 bound to a designed helical peptide derived from the GoLoco motif of RGS14	2XNS	21388199	G α i1	A, B	3.41	0.223	→GDP bound →Truncated N-terminus
Crystal structure of the G202A mutant of human G α i1	3UMS	20351284	G α i1	A	2.34	0.183	→GDP bound →Truncated N-terminus →Switch II disordered →Mutation: G202A
A G α i1 P-loop mutation prevents transition to the activated state	3QE0	22383884	G α i1	A, B, C	3.00	0.249	→GDP bound →Truncated N-terminus →Switch II disordered →Mutation: G42R
A G α P-loop mutation prevents transition to the activated state: G42R bound to RGS14 GoLoco	3QJ2	22383884	G α i1	A, B	2.80	0.200	→GDP bound →Truncated N-terminus →Mutation: G42R →G α – RGS complex
Structure of LGN GL4/G α 1 complex	4G5Q	-	G α i1	A, B, C, D	2.90	0.209	→GDP bound →Truncated N-terminus →G α – G-protein signaling modulator complex

Crystal structure of the heterodimeric complex of human RGS8 and activated G α 3	2ODE	18434541	Gai3	A, C	1.90	0.181	→GDP-AIF4- bound →Truncated N-terminus →G α – RGS complex
Crystal structure of the heterodimeric complex of human RGS10 and activated G α 3	2IHB	18434541	Gai3	A	2.71	0.209	→GDP-AIF4- bound →Truncated N-terminus →G α – RGS complex
The crystal structure of the human G-protein subunit α (GNAI3) in complex with an engineered Regulator of G-Protein Signaling Type 2 domain (RGS2)	2V4Z	19478087	Gai3	A	2.80	0.210	→GDP-AIF4- bound →Truncated N-terminus →G α – RGS complex
Structure of LGN GL4/Gai3(Q147L) complex	4G5O	-	Gai3	A, B, C, D	2.90	0.210	→GDP bound →Truncated N-terminus →Mutation: Q147L →G α – G-protein signaling modulator complex
Structure of LGN GL4/Gai3 complex	4G5R	-	Gai3	A, B, C, D	3.48	0.211	→GDP bound →Truncated N-terminus →G α – G-protein signaling modulator complex
Structure of LGN GL4/Gai3 complex	4G5S	-	Gai3	A, B, C, D	3.62	0.210	→GDP bound →Truncated N-terminus →G α – G-protein signaling modulator complex
The 2.2 Angstroms crystal structure of Transducin- α complexed with GTP γ S	1TND	8259210	Gat	A, B, C	2.20	0.190	→GTP γ S bound →Truncated N-terminus
Structural determinants for activation of the α -subunit of a heterotrimeric G protein.	1TAG	8208289	Gat	A	1.80	0.187	→GDP bound →Truncated N-terminus
GTPase mechanism of Gproteins from the 1.7-A crystal structure of transducin α -GDP-AIF-4	1TAD	7969474	Gat	A, B, C	1.70	0.209	→GDP-AIF4- bound →Truncated N-terminus
Heterotrimeric complex of a G α /G α chimera and the G β γ subunits	1GOT	8552184	Gat	A	2.00	0.207	→GDP bound →Chimera: 216-294 of Gat have been replaced with 220-298 of Gai1 →G α – G β γ complex

Crystal structure of the heterotrimeric complex of the RGS domain of RGS9, the γ subunit of Phosphodiesterase and the Gt/i1 chimera α subunit [(RGS9)-(PDE γ)-(Gt/i1 α)-(GDP)-(AlF4 $^-$)-(Mg $^{2+}$)]	1FQJ	11234020	Gat	A, D	2.02	0.233	→GDP-AlF4 $^-$ bound →Truncated N-terminus →Chimera: 216-294 of Gat have been replaced with 220-298 of Gai1 →G α – effector – RGS complex
Crystal structure of the heterodimeric complex of the RGS domain of RGS9, and the Gt/i1 chimera α subunit [(RGS9)-(Gt/i1 α)-(GDP)-(AlF4 $^-$)-(Mg $^{2+}$)]	1FQK	11234020	Gat	A, C	2.30	0.231	→GDP-AlF4 $^-$ bound →Truncated N-terminus →Chimera: 216-294 of Gat have been replaced with 220-298 of Gai1 →G α – RGS complex
Studies of a constitutively active G-alpha subunit provide insights into the mechanism of G protein activation.	3V00	22448927	Gat	A, B, C	2.90	0.221	→Chimera: 216-294 of Gat have been replaced with 220-298 of Gai1 →Mutation: G56P, K244H, D247N
Molecular architecture of Galphao and the structural basis for RGS16-mediated deactivation	3C7K	18434540	Gao	A, C	2.90	0.250	→GDP-AlF4 $^-$ bound →Truncated N-terminus →G α – RGS complex
G α complexed with GTP γ S	1AZT	9395396	Gas	A, B	2.30	0.219	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase	1AZS	9417641	Gas	C	2.30	0.219	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with Adenosine 5'-(α thio) triphosphate (RP), Mg and Mn	1CJK	10427002	Gas	C	3.00	0.220	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with β -L-2'-3'-dideoxyATP, Mn and Mg	1CJT	10427002	Gas	C	2.80	0.206	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex

Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with β -L-2'-3'-dideoxyATP and Mg	1CJU	10427002	Gas	C	2.80	0.222	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with β -L-2'-3'-dideoxyATP, Mg and Zn	1CJV	10427002	Gas	C	3.00	0.203	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with 2'-deoxy-Adenosine 3'-monophosphate, pyrophosphate and Mg	1CS4	11087399	Gas	C	2.50	0.221	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of G α with the Catalytic domains of mammalian Adenylyl cyclase: Complex with 2'-deoxy-Adenosine 3'-triphosphate and Mg	1CUL	11087399	Gas	C	2.40	0.221	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex Of G α - With The Catalytic Domains Of Mammalian Adenylyl Cyclase: Complex With 2'(3')-O-(N-methylanthraniloyl)-guanosine 5'-triphosphate and Mn	1TL7	15591060	Gas	C	2.80	0.254	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Structural basis for the inhibition of mammalian Adenylyl Cyclase by MANT-GTP	1UOH	15591060	Gas	C	2.90	0.245	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex Of G α - With The Catalytic Domains Of Mammalian Adenylyl Cyclase: Complex With TNP-ATP and Mn	2GVD	16766715	Gas	C	2.90	0.245	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Crystal Structure of Complex of G α - with The Catalytic Domains of Mammalian Adenylyl Cyclase: Complex with MANT-ATP and Mn	2GVZ	16766715	Gas	C	3.27	0.275	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex

Complex of GS- α with the Catalytic Domains of Mammalian Adenylyl Cyclase: Complex with Pyrophosphate and Ca	3C14	19243146	Gas	C	2.68	0.248	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of GS- α with the Catalytic Domains of Mammalian Adenylyl Cyclase: Complex with Pyrophosphate and Mg	3C15	19243146	Gas	C	2.78	0.240	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of GS- α with the Catalytic Domains of Mammalian Adenylyl Cyclase: Complex with Adenosine-5'-Triphosphate and Ca	3C16	19243146	Gas	C	2.87	0.252	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of GS- α with the Catalytic Domains of Mammalian Adenylyl Cyclase: Complex with Adenosine 5-O-(l-Thiophosphate) and Low Ca Concentration	3MAA	19243146	Gas	C	3.00	0.242	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Complex of GS- α with the catalytic domains of mammalian adenylyl cyclase: complex with MANT-ITP and Mn	3G82	-	Gas	C	3.11	0.240	→GTP γ S bound →Truncated N-terminus → A part of α -helical domain disordered →G α – effector complex
Crystal structure of the β 2 adrenergic receptor-Gs protein complex	3SN6	21772288	Gas	A	3.20	0.228	→No nucleotide bound → Helical domain moves to open nucleotide cleft →Mutation: G72S →G α – G $\beta\gamma$ – receptor complex
Crystal Structure of G Protein-Coupled Receptor Kinase 2 in Complex with G α -q and G $\beta\gamma$ Subunits	2BCJ	16339447	G α q	Q	3.06	0.236	→GDP-AlF ₄ - bound →Truncated N-terminus →Chimera: N-terminus from G α i1 →G α – effector complex
Crystal Structure of p63RhoGEF complex with G α -q and RhoA	2RGN	18096806	G α q	A, D	3.50	0.243	→GDP-AlF ₄ - bound →Truncated N-terminus →Chimera: N-terminus from G α i1 →G α – effector complex
Structure of heterotrimeric G protein G α -q beta gamma in	3AH8	20639466	G α q	A	2.90	0.262	→GDP bound →Chimera: N-terminus

complex with an inhibitor YM-254890							from G α i1 →G α – G $\beta\gamma$ – inhibitor peptide complex
Crystal structure of activated G α Q bound to its effector phospholipase C β 3	3OHM	20966218	G α q	A	2.70	0.207	→GDP-AIF4- bound →Truncated N-terminus →G α – effector complex
Crystal structure of G α 12 in complex with GDP, Mg $^{2+}$ and AIF4-	1ZCA	16388592	G α 12	A, B	2.90	0.239	→GDP-AIF4- bound →Truncated N-terminus →Chimera: N-terminus from G α i1
Crystal structure of G α 13 in complex with GDP	1ZCB	16388592	G α 13	A	2.00	0.209	→GDP bound →Truncated N-terminus →Chimera: N-terminus from G α i1 →Switch II, α 4- β 6 disordered
Crystal Structure of the p115RhoGEF rgRGS Domain in A Complex with G α (13):G α (i1) Chimera	1SHZ	15665872	G α 13	A, D	2.85	0.229	→GDP-AIF4- bound →Truncated N-terminus →Chimera: residues 21-47, 185-210, 213-230, 240-353 of G α i1 and residues 64-207, 234-235, 254-262 of G α 13 →G α – effector complex
Crystal Structure of PDZRhoGEF rgRGS Domain in a Complex with Galpha-13 Bound to GDP	3CX6	18940608	G α 13	A	2.50	0.266	→GDP bound →forced active conformation →Truncated N-terminus → α 4- β 6 disordered →G α – effector complex
Crystal Structure of PDZRhoGEF rgRGS Domain in a Complex with Galpha-13 Bound to GDP-AIF4	3CX7	18940608	G α 13	A	2.25	0.264	→GDP-AIF4 bound →Truncated N-terminus → α 4- β 6 disordered →G α – effector complex
Crystal Structure of PDZRhoGEF rgRGS Domain in a Complex with Galpha-13 Bound to GTP γ S	3CX8	18940608	G α 13	A	2.00	0.243	→GTP γ S bound →Truncated N-terminus → α 4- β 6 disordered →G α – effector complex
Crystal structure of p115RhoGEF RGS domain in complex with G α 13	3AB3	21507947	G α 13	A, C	2.40	0.205	→GDP-AIF4 bound →Truncated N-terminus →Chimera: N-terminus from G α i3 → α 4- β 6 disordered →G α – effector complex

Crystal structure of the $G\alpha$ protein ATGPA 1 from <i>Arabidopsis thaliana</i>	2XTZ	21304159	GPA1	A, B, C	2.34	0.212	→GTPγS bound →Truncated N-terminus →Plant G-protein
---	------	----------	------	---------	------	-------	---

Πίνακας I.3 Τα σύνολα εκπαίδευσης που χρησιμοποιήθηκαν για την κατασκευή των ειδικών pHMMs των G-πρωτεϊνών. Για τις $G\alpha$ χρησιμοποιήθηκαν ως αρνητικό σύνολο τα μέλη των υπολοίπων τριών οικογενειών. Τα αρνητικά σύνολα για τις $G\beta$ και $G\gamma$ δίνονται αναλυτικά στον ιστότοπο <http://aias.biol.uoa.gr/~mtheo/datasets/>.

A.A	UniProt AC	Τύπος	Οργανισμός	A.A	UniProt AC	Τύπος	Οργανισμός
Οικογένεια G_s				Οικογένεια $G_{12/13}$			
1	Q5JWF2	$G\alpha_s$	Homo sapiens	1	Q03113	$G\alpha_{12}$	Homo sapiens
2	Q6R0H7	$G\alpha_s$	Mus musculus	2	P27600	$G\alpha_{12}$	Mus musculus
3	Q63803	$G\alpha_s$	Rattus norvegicus	3	Q63210	$G\alpha_{12}$	Rattus norvegicus
4	P04896	$G\alpha_s$	Bos taurus	4	Q14344	$G\alpha_{13}$	Homo sapiens
5	P63092	$G\alpha_s$	Homo sapiens	5	P27601	$G\alpha_{13}$	Mus musculus
6	P63094	$G\alpha_s$	Mus musculus	6	Q19572	$G\alpha_{12}$	Caenorhabditis elegans
7	P63095	$G\alpha_s$	Rattus norvegicus	7	Q613V4	$G\alpha_{12}$	Caenorhabditis briggsae
8	Q7PD79	$G\alpha_s$	Anopheles gambiae	$G\beta$			
9	P63091	$G\alpha_s$	Canis familiaris	1	P62871	$G\beta$	Bos taurus
10	Q8R4A8	$G\alpha_s$	Cricetulus griseus	2	P17343	$G\beta$	Caenorhabditis elegans
11	P16052	$G\alpha_s$	Cricetulus longicaudatus	3	P62872	$G\beta$	Canis familiaris
12	P20354	$G\alpha_s$	Drosophila melanogaster	4	Q6TMK6	$G\beta$	Cricetulus griseus
13	Q292P9	$G\alpha_s$	Drosophila pseudoobscura	5	Q6PH57	$G\beta$	Danio rerio
14	Q9XZV5	$G\alpha_s$	Geodia cydonium	6	P26308	$G\beta$	Drosophila melanogaster
15	O16118	$G\alpha_s$	Homarus americanus	7	O45040	$G\beta$	Homarus americanus
16	P30684	$G\alpha_s$	Lymnaea stagnalis	8	P62873	$G\beta$	Homo sapiens
17	P63093	$G\alpha_s$	Mesocricetus auratus	9	P62874	$G\beta$	Mus musculus
18	P29797	$G\alpha_s$	Sus scrofa	10	Q5R5W8	$G\beta$	Pongo abelii
19	P30669	$G\alpha_s$	Schistosoma mansoni	11	P54311	$G\beta$	Rattus norvegicus
20	P24799	$G\alpha_s$	Xenopus laevis	12	P93397	$G\beta$	Nicotiana tabacum
21	P38405	$G\alpha_{olf}$	Homo sapiens	13	P79959	$G\beta$	Xenopus laevis
22	Q8CGK7	$G\alpha_{olf}$	Mus musculus	14	P11017	$G\beta$	Bos taurus
23	P38406	$G\alpha_{olf}$	Rattus norvegicus	15	Q20636	$G\beta$	Caenorhabditis elegans

Οικογένεια $G_{i/o}$				16	P29829	$G\beta$	<i>Drosophila melanogaster</i>
1	P19086	$G\alpha_z$	<i>Homo sapiens</i>	17	P62879	$G\beta$	<i>Homo sapiens</i>
2	O70443	$G\alpha_z$	<i>Mus musculus</i>	18	P54313	$G\beta$	<i>Rattus norvegicus</i>
3	P19627	$G\alpha_z$	<i>Rattus norvegicus</i>	19	P93398	$G\beta$	<i>Nicotiana tabacum</i>
4	P27045	$G\alpha_k$	<i>Xenopus laevis</i>	20	P79147	$G\beta$	<i>Canis familiaris</i>
5	P38403	$G\alpha_k$	<i>Cavia porcellus</i>	21	P16520	$G\beta$	<i>Homo sapiens</i>
6	P08754	$G\alpha_k$	<i>Homo sapiens</i>	22	Q40507	$G\beta$	<i>Nicotiana tabacum</i>
7	Q9DC51	$G\alpha_k$	<i>Mus musculus</i>	23	Q9HAV0	$G\beta$	<i>Homo sapiens</i>
8	P08753	$G\alpha_k$	<i>Rattus norvegicus</i>				24
9	P04695	$G\alpha_{t1}$	<i>Bos taurus</i>	25	O14775	$G\beta$	<i>Homo sapiens</i>
10	Q28300	$G\alpha_{t1}$	<i>Canis familiaris</i>	26	P62881	$G\beta$	<i>Mus musculus</i>
11	P11488	$G\alpha_{t1}$	<i>Homo sapiens</i>	27	P49177	$G\beta$	<i>Arabidopsis thaliana</i>
12	P20612	$G\alpha_{t1}$	<i>Mus musculus</i>	28	O14435	$G\beta$	<i>Cryphonectria parasitica</i>
13	P04696	$G\alpha_{t2}$	<i>Bos taurus</i>	29	P36408	$G\beta$	<i>Dictyostelium discoideum</i>
14	P19087	$G\alpha_{t2}$	<i>Homo sapiens</i>	30	P23232	$G\beta$	<i>Loligo forbesi</i>
15	P50149	$G\alpha_{t2}$	<i>Mus musculus</i>	31	Q08706	$G\beta$	<i>Lymnaea stagnalis</i>
16	POC7Q4	$G\alpha_{t3}$	<i>Bos taurus</i>	32	P49178	$G\beta$	<i>Zea mays</i>
17	A8MTJ3	$G\alpha_{t3}$	<i>Homo sapiens</i>	33	Q40687	$G\beta$	<i>Oryza sativa</i>
18	Q3V3I2	$G\alpha_{t3}$	<i>Mus musculus</i>	34	Q10282	$G\beta$	<i>Schizosaccharomyces pombe</i>
19	P29348	$G\alpha_{t3}$	<i>Rattus norvegicus</i>	35	P18851	$G\beta$	<i>Saccharomyces cerevisiae</i>
20	P38407	$G\alpha_t$	<i>Xenopus laevis</i>	36	O24456	$G\beta$	<i>Arabidopsis thaliana</i>
21	P63097	$G\alpha_{i1}$	<i>Bos taurus</i>	37	Q93134	$G\beta$	<i>Biomphalaria glabrata</i>
22	P38401	$G\alpha_{i1}$	<i>Cavia porcellus</i>	38	P63243	$G\beta$	<i>Bos taurus</i>
23	P50146	$G\alpha_{i1}$	<i>Gallus</i>	39	O42248	$G\beta$	<i>Danio rerio</i>
24	P63096	$G\alpha_{i1}$	<i>Homo sapiens</i>	40	P46800	$G\beta$	<i>Dictyostelium discoideum</i>
25	P87383	$G\alpha_{i1}$	<i>Oryzias latipes</i>	41	O18640	$G\beta$	<i>Drosophila melanogaster</i>
26	Q5RAD4	$G\alpha_{i1}$	<i>Pongo abelii</i>	42	P63244	$G\beta$	<i>Homo sapiens</i>
27	P10824	$G\alpha_{i1}$	<i>Rattus norvegicus</i>	43	Q25189	$G\beta$	<i>Hydra vulgaris</i>
28	P27044	$G\alpha_{i1}$	<i>Xenopus laevis</i>	44	Q4R7Y4	$G\beta$	<i>Macaca fascicularis</i>
29	P38400	$G\alpha_{i2}$	<i>Canis familiaris</i>	45	Q01369	$G\beta$	<i>Neurospora crassa</i>
30	P38402	$G\alpha_{i2}$	<i>Cavia porcellus</i>	46	P49027	$G\beta$	<i>Oryza sativa</i>
31	P08239	$G\alpha_o$	<i>Bos taurus</i>	47	P63246	$G\beta$	<i>Sus scrofa</i>
32	P51875	$G\alpha_o$	<i>Caenorhabditis elegans</i>	48	Q10281	$G\beta$	<i>Schizosaccharomyces pombe</i>
33	P59216	$G\alpha_o$	<i>Cricetulus longicaudatus</i>	49	Q39836	$G\beta$	<i>Glycine max</i>
34	P16378	$G\alpha_o$	<i>Drosophila melanogaster</i>	50	P38011	$G\beta$	<i>Saccharomyces cerevisiae</i>
35	Q9XZV3	$G\alpha_o$	<i>Geodia cydonium</i>	Gγ			

36	P51877	Gα ₀	Helisoma trivolvis	1	P50151	Gγ	Homo sapiens
37	P09471	Gα ₀	Homo sapiens	2	P61954	Gγ	Rattus norvegicus
38	P38404	Gα ₀	Locusta migratoria	3	Q5RBQ0	Gγ	Pongo abelii
39	P30683	Gα ₀	Lymnaea stagnalis	4	P63210	Gγ	Canis familiaris
40	P53359	Gα ₀	Manduca sexta	5	P38040	Gγ	Drosophila melanogaster
41	P18872	Gα ₀	Mus musculus	6	Q9NFZ2	Gγ	Calliphora vicina
Οικογένεια G_{q/11}				7	Q75BU5	Gγ	Ashbya gossypii
1	P38409	Gα ₁₁	Bos taurus	8	Q4VT26	Gγ	Caenorhabditis briggsae
2	P29992	Gα ₁₁	Homo sapiens	9	P54406	Gγ	Caenorhabditis elegans
3	P45645	Gα ₁₁	Meleagris gallopavo	10	Q6FJ50	Gγ	Candida glabrata
4	P21278	Gα ₁₁	Mus musculus	11	Q95ZG8	Gγ	Dictyostelium discoideum
5	Q2XSV9	Gα ₁₁	Sus scrofa	12	Q6CPB4	Gγ	Kluyveromyces lactis
6	Q9JID2	Gα ₁₁	Rattus norvegicus	13	Q870G5	Gγ	Lentinula edodes
7	P43444	Gα ₁₁	Xenopus laevis	14	Q01821	Gγ	Loligo forbesi
8	P38408	Gα ₁₄	Bos taurus	15	Q7RWT0	Gγ	Neurospora crassa
9	O95837	Gα ₁₄	Homo sapiens	16	O94309	Gγ	Schizosaccharomyces pombe
10	P30677	Gα ₁₄	Mus musculus	17	Q6CHP5	Gγ	Yarrowia lipolytica
11	O73819	Gα ₁₄	Xenopus laevis	18	P18852	Gγ	Saccharomyces cerevisiae
12	P30679	Gα ₁₅	Homo sapiens	19	Q9JMF3	Gγ	Mus musculus
13	P30678	Gα ₁₅	Mus musculus	20	P63216	Gγ	Mus musculus
14	Q9TU29	Gα ₁₅	Oryctolagus cuniculus	21	P63212	Gγ	Bos taurus
15	O88302	Gα ₁₅	Rattus norvegicus	22	Q5R639	Gγ	Pongo abelii
16	Q28294	Gα _q	Canis familiaris	23	P63219	Gγ	Rattus norvegicus
17	P23625	Gα _q	Drosophila melanogaster	24	O60262	Gγ	Homo sapiens
18	Q9XZV4	Gα _q	Geodia cydonium				
19	P91950	Gα _q	Homarus americanus				
20	P50148	Gα _q	Homo sapiens				
21	P38412	Gα _q	Loligo forbesi				
22	P38411	Gα _q	Lymnaea stagnalis				
23	P21279	Gα _q	Mus musculus				
24	O15975	Gα _q	Patinopecten yessoensis				
25	Q2PKF4	Gα _q	Sus scrofa				
26	P82471	Gα _q	Rattus norvegicus				
27	P38410	Gα _q	Xenopus laevis				

Δημοσιεύσεις

Original article

GPCRs, G-proteins, effectors and their interactions: human-gpDB, a database employing visualization tools and data integration techniques

Venkata P. Satagopam¹, Margarita C. Theodoropoulou², Christos K. Stampoulakis², Georgios A. Pavlopoulos^{1,3}, Nikolaos C. Papandreou², Pantelis G. Bagos³, Reinhard Schneider¹ and Stavros J. Hamodrakas^{2,*}

¹Structural and Computational Biology Unit, EMBL, Meyerhofstrasse 1, Heidelberg D69117, Germany, ²Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01 and ³Department of Computer Science and Biomedical Informatics, University of Central Greece, Lamia 35 100, Greece

*Corresponding author: Tel: +30 210 727 4931; Fax: +30 210 727 4254. Email: shamodr@biol.uoa.gr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Submitted 1 June 2010; Revised 21 July 2010; Accepted 22 July 2010

G-protein coupled receptors (GPCRs) are a major family of membrane receptors in eukaryotic cells. They play a crucial role in the communication of a cell with the environment. Ligands bind to GPCRs on the outside of the cell, activating them by causing a conformational change, and allowing them to bind to G-proteins. Through their interaction with G-proteins, several effector molecules are activated leading to many kinds of cellular and physiological responses. The great importance of GPCRs and their corresponding signal transduction pathways is indicated by the fact that they take part in many diverse disease processes and that a large part of efforts towards drug development today is focused on them. We present Human-gpDB, a database which currently holds information about 713 human GPCRs, 36 human G-proteins and 99 human effectors. The collection of information about the interactions between these molecules was done manually and the current version of Human-gpDB holds information for about 1663 connections between GPCRs and G-proteins and 1618 connections between G-proteins and effectors. Major advantages of Human-gpDB are the integration of several external data sources and the support of advanced visualization techniques. Human-gpDB is a simple, yet a powerful tool for researchers in the life sciences field as it integrates an up-to-date, carefully curated collection of human GPCRs, G-proteins, effectors and their interactions. The database may be a reference guide for medical and pharmaceutical research, especially in the areas of understanding human diseases and chemical and drug discovery.

Database URLs: http://schneider.embl.de/human_gpdb; http://bioinformatics.biol.uoa.gr/human_gpdb/

Background

Signal transduction refers to these cellular processes by which stimuli, either physical or chemical, induce specific cellular responses, through chosen molecular mechanisms. The specificity of a cellular response to a signal depends on the receptor expressed on the target cell.

G-protein coupled receptors (GPCRs) are a very important superfamily of cell membrane receptors in eukaryotic cells. They may interact with both the environment outside and inside the cell and they play a crucial role in receiving stimuli signals from the environment. In response they induce certain cellular responses. GPCRs have a characteristic

structure comprised of seven transmembrane-spanning α -helices, an extracellular N terminus, an intracellular C terminus and three interhelical loops on each side of the membrane (1). Several classification systems have been used for this superfamily categorization. The most frequently system used (2,3) classifies GPCRs in six classes, based on their sequence homology and their functional similarity. These are: Class A or 1 Rhodopsin-like, Class B or 2 Secretin receptor family, Class C or three Metabotropic glutamate/pheromone, Class D or four Fungal mating pheromone receptors, Class E or five Cyclic AMP receptors and Class F or six Frizzled/Smoothed like, first presented by (4). GPCRs that are not yet characterized or classified are called orphan GPCRs. Furthermore, a number of putative classes of some newly discovered GPCRs exist, whose nomenclature has not yet been accepted by the scientific community (5,6). Ligands bind to GPCRs on the outside of the cell, activating the GPCRs by causing a conformational change, and allowing them to bind to G-proteins (7).

G-proteins form heterotrimers composed of $G\alpha$, $G\beta$ and $G\gamma$ subunits, which possess a binding site for a GTP or a GDP molecule. They are characterized by their α -subunits, which are further grouped into the $G\alpha_s$, $G\alpha_i/o$, $G\alpha_q$ and $G\alpha_{12}$ families (8). The stimulation of GPCRs leads to the activation of G-proteins, which dissociate into their $G\alpha$ and $G\beta\gamma$ subunits. The subunits then activate several effector molecules that lead to many kinds of cellular and physiological responses (1). Effectors form a diverse group of proteins through their interaction with G-proteins that act either as secondary messengers, or lead directly to a cellular and physiological response. Many proteins such as tubulins, adenylate cyclases, ion channels and others act as effectors (5). GPCRs, G-proteins, effectors and their interactions compose one of the main mechanisms for signal transduction and activation or deactivation of pathways within the cell. A large part of efforts towards drug development today is focused on finding chemicals that affect the ability of ligands to bind to GPCRs (9) either to inhibit or accelerate certain cellular processes. GPCRs play a crucial role in a wide range of human diseases.

Human-gpDB was developed as a tool for integrating together human GPCRs, G-proteins and effectors. It does not only present how they interact with each other but it also reveals information about the pathways they are involved in. Human-gpDB was built as a useful tool for drug research and as a platform that reveals new patterns for therapeutic paths.

Construction and content

Data integration

Our initial step was to collect sequence information individually about human GPCRs, G-proteins and effectors from

the UniProt/Swissprot database (10). The entries were acquired using suitable scripts written in Perl in order to parse the DE (description), the GN (gene) or the DR (database cross reference) field of a respective database entry. The data sets were then checked manually in order to eliminate duplicates. Our main goal was to include database unique entries from UniProt/SwissProt. Perl scripts were used for data manipulation.

For each of the three sets, the next step was to isolate and keep these proteins that have at least one connection with another protein, a GPCR with a G-protein and a G-protein with an effector and vice versa. For the extraction of information concerning the connections between human GPCRs, G-proteins and effectors, an extensive literature search was performed attempting to detect terms that co-occur in the same abstract and are biologically related. None of the available text mining engines was used in order to avoid false negative results and to increase the reliability of the results that are presented. Currently the database holds 1663 connections between GPCRs and G-proteins and 1618 connections between G-proteins and effectors. In addition, PubMed reference articles that provide the literature support for each recorded connection are included.

Efficient ways were used to show how G-proteins and effectors might be categorized into families, subfamilies and types. G-proteins' classification is the most commonly used and is based on their α subunits sequence homology, while effectors' classification is based on their function. Many different classifications exist regarding GPCRs; the classification used here was according to the IUPHAR classification (2). All classifications were done manually.

UniProt (10) identifiers were used as starting points to integrate the Human-gpDB with various external data sources. The systems that were used to help us with this integration were ENSEMBL (11), BioMart (12) and SRS (13). For each of the proteins, information about the name, the sequence, the description, the family and the subfamily it belongs to, together with the full record coming from the Dasty2 DAS client (14) was collected. Furthermore, a collection of a vast variety of linked identifiers was obtained, to enrich the information for each protein. Thus, information comes from various databases like for example Uniprot (10), RefSeq Proteins (15), Entrez Proteins (16) and Ensembl Proteins databases (11). Information about the gene location and its properties are provided from Ensembl (11), EMBL (17), EntrezGene (18), RefSeq DNA (15) and UniGene databases. Domain links are provided for Smart (19), InterPro (20) and Pfam (21,22) databases. Structures are linked to PDB (23), HSSP (24) and PSSH (25) databases. Information about diseases comes from the OMIM (26) database and information about protein function from the Gene Ontology (27) database. Chemical information is provided by HMDB (28) and

DrugBank (29) and pathway information comes from KEGG (30), Panther (31) and Reactome (32) databases. Drugs related to three categories of molecules of Human-gpDB (GPCRs, G-proteins, effectors) were collected from DrugBank (29), Madator (33) and AKS2 (34). All available interactions between drugs and the three categories of molecules of the database are presented to the user using visualization tools (see 'Visualization' section for more details). All of the aforementioned information was collected for each protein using ENSEMBL, BioMart and SRS and the results were stored in a MySQL database.

Going one step further, Human-gpDB is not only linked to other sources but it also comes with some analysis features to make the data integration part more useful. Therefore, Human-gpDB comes with domain architecture analysis, protein-protein/protein-chemical interactions and pathway enrichment that will be explained in detail in the 'Utility and Discussion' section below.

Implementation details

All the results are delivered to the user through a web application. The database was implemented in MySQL whereas for the graphical user interface (GUI), the HTML language was used. The dynamic parts of the interface, as for example the auto-complete forms and the advanced search capabilities, are supported by Javascript. The communication between the GUI and the database to extract information from Human-gpDB was achieved with the use of CGI scripts and all the calculations were performed using Perl. The entire application is set up behind an apache web server. The pop-up window that provides links to external or internal data sources was implemented with the help of the Overlib library. For the visualization, the Arena3D standalone Java application (35) was used to support the projections of large-scale networks whereas for smaller networks the Medusa Java applet (36) visualization tool was used. Figure 1 shows an overview of Human-gpDB web application.

Visualization

The visualization module in Human-gpDB was designed in such a way, in order to give maximum flexibility to the user to visualize the interactions between different knowledge domains (GPCRs, G-proteins, effectors, drugs related to GPCRs, drugs related to G-proteins and drugs related to effectors) at different levels (depths) by taking advantage of the hierarchical categories of GPCRs, G-proteins and effectors.

Medusa, a 2D visualization tool (36), was used to graphically visualize interaction partner bioentities such as proteins or drugs as they come from the DrugBank, MATADOR and AKS2 for each of the GPCR, G-protein and effector proteins. The newer version of Medusa tool that supports the current version of Human-gpDB is now more interactive

and many layout algorithms are implemented that make the networks much more informative and the extraction of the biological knowledge easier. Like Arena3D, Medusa comes with a set of layout algorithms that are able to minimize the crossovers between the nodes and make the network visually simpler. Medusa is currently provided as a java applet.

To show either the whole network consisting of drugs, GPCRs, G-proteins, effector proteins and their interactions or some large scale sub-networks, the Arena3D standalone visualization tool (35) was used. According to Arena3D, drugs, GPCRs, G-protein and effectors were separated onto four different layers following a multi-layer graph concept, a stack of 2D networks. Arena3D among others currently comes with a clustering layout algorithm that is able to visualize very efficiently predefined distinct clusters. The separation of the clusters is done by placing the nodes that belong to the same cluster together either in 2D or 3D groups. The rich color scheme helps the user to immediately recognize which node belongs to which cluster since nodes that belong to the same cluster are colored similarly. Taking advantage of this functionality, GPCRs, G-proteins and effectors were clustered individually onto their different layers into families and subfamilies according to their properties as mentioned in the previous section, whereas drugs were not clustered. Arena3D is highly interactive and gives the opportunity to the user to isolate either individual, or a set of, connections that reaches his interests for a more focused research. Users can visually highlight and observe patterns that can be easier processed by the user. Such an example could be the answer to a question like whether a protein targets a specific protein family or not. In this version of Human-gpDB, pre-generated input files for the interaction networks are available for download and can be used as input files for the Arena3D application, since it is currently available as a standalone application only and not as a web based tool.

To visually show the domain structure, a static HTML view was implemented. For the pathway visualization, the KEGG schemas are given as they come from the KEGG database. The parts of the pathways that the proteins of the Human-gpDB are involved in are then highlighted to allow the users to easier mark and distinguish the signal transition paths that each protein gets involved in. This functionality is also a strong point of Human-gpDB since researchers can immediately see the influence of a specific protein on a pathway.

Utility and discussion

Browse section

In this section, information about the three distinct categories according to the protein type is presented. These are

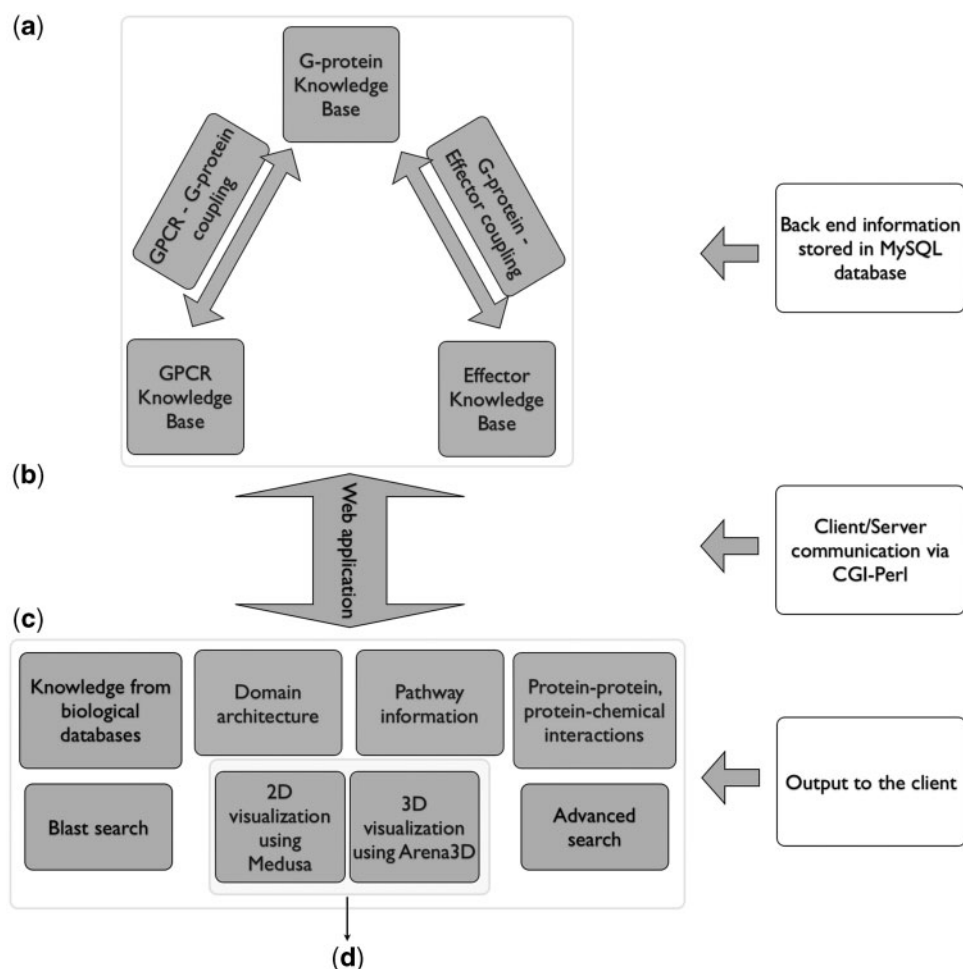


Figure 1. Over view of Human-gpDB web application. (a) Backend of the application consisting of manually collected information regarding GPCRs, G-proteins, effectors and their interactions as well as a wide range of publicly available information for each of these proteins stored in a MySQL database ('Data Integration' section for more details). (b) A CGI-Perl script handles the communication between the client and the server and (c) provides a wide range of information as output to the user (see 'Utility' section for more details). The system can be coupled with (d) a 2D visualization tool (Medusa) and a 3D visualization tool (Arena3D), which allows the easy visualization of the relationships between GPCRs, G-Proteins, effectors and the drugs (see 'Visualization' sections for more details).

the 'GPCRs', 'G-proteins' and 'effectors' respectively. Each category of proteins is further sub-divided into classes, families, sub-families hierarchically. One can easily navigate the results by choosing the respective category and then by following the '+' (expand) and the '-' (collapse) button to navigate through the different levels of the hierarchy. Once the user fully expands these categories and goes to the lowest levels that contain the proteins, s/he can see some information about these proteins like the name, the description, the gene names, hyper links to UniProt and links to the relevant entry page, a summary sheet and visualization module (node level).

Entry page. The entry page of each protein contains information about the class, the family, the subfamily

that each protein belongs to according to the IUPHAR (2) classification, the sequence and a variety of cross references to various available databases like the UniProt, EMBL, InterPro, PRODOM, GENEID, MIM, PRINTS and PFAM. In the entry page, links about proteins that interact with a selected protein in the form of a sorted list and link to node level visualization module. Redirections to the literature (PubMed) that reveal the evidence of the recorded interaction are also given, so that users can observe the biological relevance of each connection and see where the interaction comes from.

Summary sheet. It provides information about each individual protein coming from a vast variety of publicly available databases. The collected information is further

split into different sections; these are labeled as genes, proteins, domains, structural features, diseases, gene ontology (GO), pathways, chemicals and orthology. The information of each section can be further explored by giving to researchers the opportunity to go deeper into the volume of knowledge in a more efficient way. All the information is presented under the same web page, which not only makes the exploration easier but it also reduces dramatically the loss of time that someone needs to find the information by browsing and querying the available databases. For example, if one checks the information in the section 'Proteins', he can see that it comes from four resources; these are 'Ensembl Proteins', 'Entrez Proteins', 'RefSeq Proteins' and UniPort. The information in these four resources related to the protein of interest is either complementary to each other or it provides additional evidence. Furthermore, Dasty2 DAS client links are given, to present many of the protein features. Simultaneously, a link to visualization module is provided to visualize the interactions between GPCRs, G-proteins and effectors graphically. In contrast to the entry page, the summary sheet provides to the user all the relevant information about how each protein is linked to other bioentities like genes, structures, diseases, pathways, domains, chemicals, etc. The links that are provided were explained in a previous paragraph of this section.

Visualization module. Human-gpDB provides a very flexible visualization module, where a user can select knowledge domains of interest, level or depth and also the tool in order to visualize the particular network of user's choice. The user may choose to include all types of molecules (GPCRs, G-proteins, effectors, drugs related to GPCRs, drugs related to G-proteins and drugs related to effectors) in the network or some of them. Also the user can decide the level/depth (Class, Family, Subfamily, Type and Node) in which the visualization will be made. This feature is available for GPCRs, G-proteins, effectors, but not for drugs. Two visualization tools are offered as options: Medusa (36), for 2D representations, and Arena 3D (35), for 3D visualization. Drugs are not classified and thus they appear as individual nodes connected to the respective type of molecule (GPCRs, G-proteins and effectors). The main reason for giving the possibility to visualize the interactions at these levels is because otherwise the information gets overcrowded and it becomes difficult to clearly see the information. Arena3D software was chosen to overcome the problem of the 2D space limitations for visualizing larger scale networks. These networks consist of hundreds of nodes and hundreds of connections. The main feature of Arena3D is that it utilizes 3D space to project the data. Like Medusa, it also comes with efficient algorithms to minimize the crossovers between the connections so that the network becomes more informative. The Arena3D tool was

used to visualize the interactions between Drugs, GPCRs, G-proteins and effectors for any selected category and any protein level. Simultaneously, the four different molecule categories were separated onto four different 2D layers by following a multi-layered graph representation. These are the Drug, the GPCR, the G-protein and the effector layers. Arena3D does not only visualize the nodes and the edges of the network but it can also very efficiently visualize precalculated clusters, which in the case of Human-gpDB, represent the subfamilies of the proteins (see 'Visualization' section for more details).

Pop-up window. While a user navigates through the data in the browse section, he/she may further answer questions that refer to a set of proteins that belong to a specific level of the hierarchy and not only to individual ones. A researcher for example, might want to see some information about the whole B or C class of the GPCRs or the 'Gamma-aminobutyric' acid receptor subfamily. This way, a user may explore the biological knowledge related to each category and sub category by following the hyperlinks provided by a pop-up window after clicking on highlighted names. The information that a particular user may retrieve about a set of proteins is explained below.

Knowledge from biological databases. Biological information from different databases for all of the members of the selected category and sub category are displayed in a table view. A variety of sorting choices is provided so that researchers can sort the target proteins according to their names, their description, the family, the class or the subfamily that they belong to if any.

Domain architecture. A static comparative domain architecture view in order to detect patterns and investigate if the selected category possesses specific structural features is provided. Domain information of each protein was collected from the SMART database and a HTML based visualization tool was developed, to display this knowledge. In this view, each protein is hyperlinked to the SMART and ENSEMBL databases. Mouse actions over a specific domain allow the user to interact with the GUI interactively and get further information about the selected domain.

Pathway information. Each protein was mapped to KEGG pathways in order to find which pathways are enriched for the selected category. The results are provided in a tabular view accompanied by a KEGG pathway identifier, a pathway name and a list of proteins involved in that pathway. Each pathway is hyperlinked to a red flag, which ultimately displays the pathway, with the proteins from the selected category highlighted in red color.

Protein–protein and protein–chemical interactions. Information about protein–protein and protein–chemical interactions are provided by the STITCH database (37). The collection of this information was done through the available API. The generated networks do not only include proteins stored in Human-gpDB database but also proteins that are recorded in the STITCH database. These proteins do not necessarily belong to one of the three categories of proteins that Human-gpDB holds. The STITCH database (37) goes one step further by providing interacting proteins that were found experimentally besides proteins that co-occur in the literature. Together with the variety of information that STITCH database holds, this is the main reason why this tool was selected to provide to the user the relevant protein–protein and protein–chemical interaction networks.

Visualization. This feature of the pop-up window redirects the user to the Visualization module referring to the particular level of the tree [see Visualization module and ‘Visualization’ section (main manuscript) for more details].

Blast search section

This feature of Human-gpDB database gives the opportunity to the users to search for homologies by providing one or more protein sequences in Fasta format. Wu-BLAST (38) was used to align a given set of sequences against the selected protein categories of Human-gpDB like for example against the GPCR category. The user may provide the advanced blast options in order to narrow down the search results. The results are then grouped according to their category and sorted by significance. Each result that is found is then hyperlinked to an entry page, summary sheet and node level visualization module. Here, the user can see the alignments as well.

Advanced search section

This feature gives users the option to search the given fields in the database. The user can enter any word in one or more of the available boxes under the name: Gene/Protein, Class, Family, Subfamily, Type, Description and Function. Expressions in separate search fields are combined with the AND operator, so every entry of the result set will satisfy the expressions of all the search fields the user has chosen. The user has the option to choose whether the query will be performed against the GPCRs, the G-proteins or the effectors included in the database.

Results

The database currently holds information about 713 human GPCRs, 36 human G-proteins and 99 human effectors.

The collection of information about the interactions between these molecules was done manually and the current status of Human-gpDB reveals information about 1663 connections between GPCRs and G-proteins and 1618 connections between G-proteins and effectors. GPCRs are categorized in four classes. Table 1 shows the number of families and subfamilies in each GPCR class, while Table 2 shows the distribution of GPCRs’ subfamilies based on the number of $G\alpha$ families with which they interact. G-proteins are categorized in $G\alpha$, $G\beta$ and $G\gamma$ groups. $G\alpha$ consists of four respective families, as described initially in the ‘Background’ section. From the 36 human G-proteins, 17 are characterized as $G\alpha$, 7 as $G\beta$ and 12 as $G\gamma$. Effectors are categorized in 20 families, 29 subfamilies and 63 types based on their biological function (Theodoropoulou, M.C.,

Table 1. Number of families and subfamilies in each GPCR class

GPCRs’ Class	No. of GPCR Families	No. of GPCR Subfamilies
Class A	55	640 (422 subfamilies of olfactory receptor)
Class B	6	16
Class C	4	41 (29 subfamilies of taste receptors)
Frizzled/Smoothened	2	11
Total	67	708

Class A is the largest and consists of 55 families and 640 subfamilies (422 subfamilies of olfactory receptors). Class B consists of 6 families and 16 subfamilies. Class C consists of 4 families and 41 subfamilies (29 subfamilies of taste receptors). Frizzled/Smoothened class consists of 2 families and 11 subfamilies.

Table 2. Distribution of GPCR subfamilies based on the number of $G\alpha$ families with which they interact

Couples with	No. of GPCRs’ Subfamilies
1 $G\alpha$ family	623
2 $G\alpha$ families	48
3 $G\alpha$ families	15
All 4 $G\alpha$ families	1
Unknown coupling	21
Total	708

One subfamily of GPCRs, the TSHR family, couples with members of all four $G\alpha$ families. Most of the GPCR subfamilies couple with members of one $G\alpha$ family (623 out of the 708 subfamilies of GPCRs). Fifteen GPCR subfamilies couple with members from 3 $G\alpha$ families, whereas 48 couple with members from 2 $G\alpha$ families. Twenty-one GPCR subfamilies do not have known coupling.

Bagos, P.G. and Hamdrakas, S.J., manuscript in preparation). The two most highly populated effectors' families are Ion Channels and Tubulins.

Visualization of the interactions between GPCRs, G-proteins, effectors and drugs together with the rich data integration part is one of the main features of Human-gpDB. As described in the 'Visualization' section Medusa application was used for 2D representation of the networks of interactions. Arena3D was chosen for 3D and more efficient representation of either the whole network of interactions or dense subparts of it. In order for the user to evaluate the visualization tools offered, five chosen different examples of visualization are given in Figures 2–6. Medusa, which is a Java applet, offers the user a first glance of the respective network. However Medusa still has disadvantages compared to Arena3D mainly due to the fact that the visualization it offers is in 2D so the space might be a limiting factor for larger or dense networks.

Case study of human Prostanoid TP receptor's network of interactions

In order to demonstrate the utility of Human-gpDB, a case study of human Prostanoid TP receptor's network of interactions (Figure 5) is presented. Human Prostanoid TP receptor may couple with all four subfamilies (G_q , G_{11} , G_{14} and $G_{15/16}$) of the $G_{q/11}$ family of G-proteins. These subfamilies

of G-proteins interact with 21 different types of effectors belonging to 8 different families (Tubulins, PI3/PI4 kinases, Phospholipases C, Ser/Thr protein kinases, TPR repeat proteins, Tyr protein kinases, Guanine nucleotide exchange factors and Ezrin-radixin-moesin-binding phosphoproteins). For this specific receptor 23 different drugs exist. Using information from STITCH (37) about protein–protein and protein–chemical interactions, information regarding mostly known natural or synthetic ligands is presented, complementing the functional role of the receptor. Prostanoid TP receptor is related with a bleeding disorder in cases of defection of the receptor [information retrieved from 188070 entry of OMIM (26) database]. Regarding the receptor's participation in particular KEGG (30) pathways, the receptor is involved in a calcium-signaling pathway (hsa04020). Using the known coupling preferences of the receptor, a researcher may relate the receptor with other specific KEGG pathways (for example gap junction), in which $G_{q/11}$ G-proteins are known to participate. After launching BLAST against Human-gpDB using human Prostanoid TP receptor as query sequence, Prostanoid FR receptor is the most similar entry. Prostanoid FR receptor interacts with the same subfamilies of G-proteins, shares four mutual ligands and agonists [comparison of the networks presented by STITCH (37)] and two mutual drugs [comparison of the networks presented by Medusa (36)] with Prostanoid TP receptor, and, both receptors

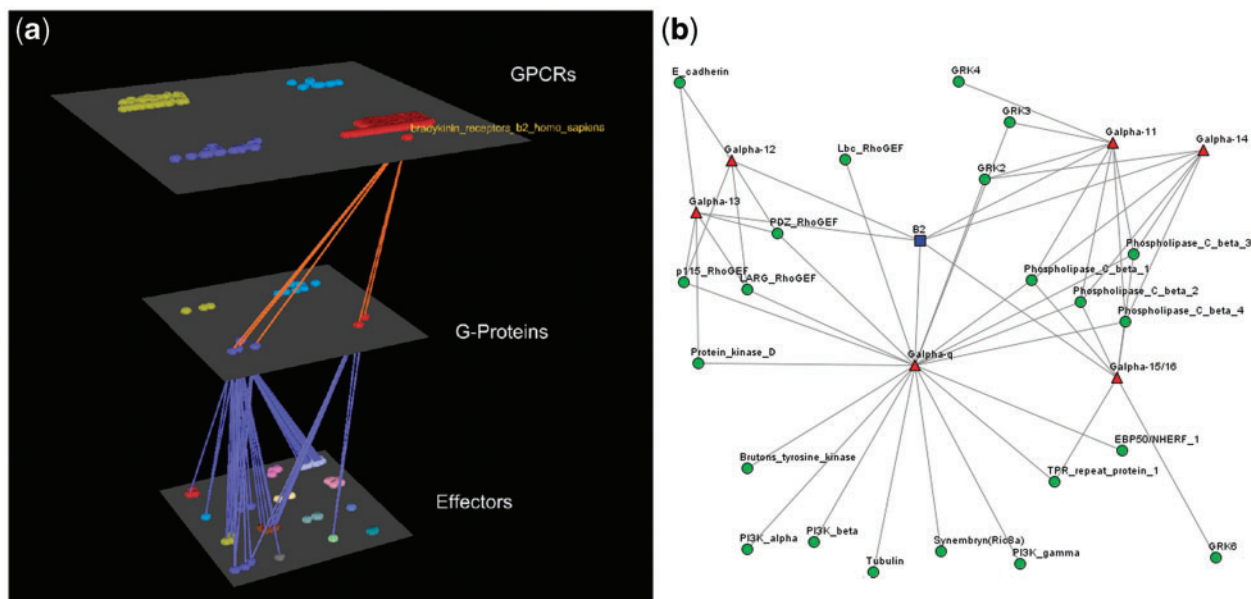


Figure 2. Visualization of human Bradykinin B2 receptor's interactions (a) Arena3D Visualization: human Bradykinin B2 receptor targets six different subfamilies of G_{α} G-Proteins belonging to $G_{\alpha-q/11}$ and $G_{\alpha-12/13}$ families. The G-Proteins are connected to 22 different types of effectors belonging to nine families. (b) Medusa 2D Visualization: human Bradykinin B2 receptor targets G_{α} G-Proteins belonging to six distinct G_{α} subfamilies ($G_{\alpha-q}$, $G_{\alpha-11}$, $G_{\alpha-14}$, $G_{\alpha-15/16}$, $G_{\alpha-12}$ and $G_{\alpha-13}$). These G_{α} G-Proteins interact with 22 types of effectors.

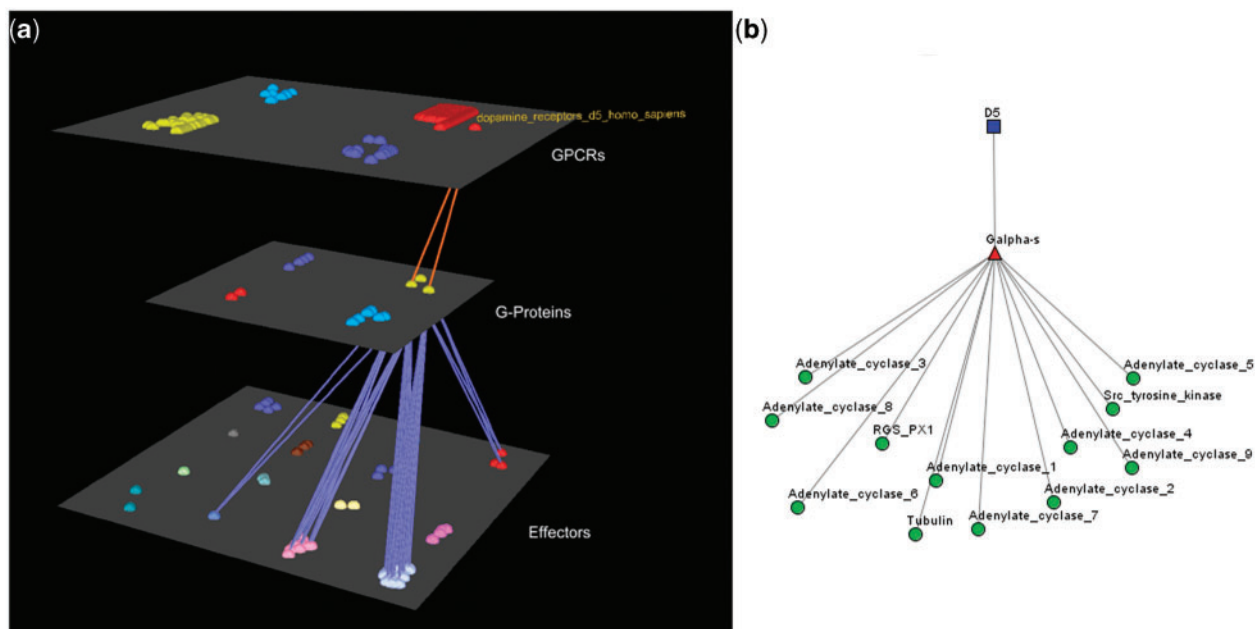


Figure 3. Visualization of human Dopamine D5 receptor's interactions. (a) Arena3D Visualization: Human Dopamine D5 receptor targets two $G\alpha$ -s G-Proteins. The G-Proteins are connected to 12 types of effectors belonging to four specific families. These are: Adenylate cyclases, Regulators of G-Protein signaling, Tyr protein kinases and tubulins. (b) Medusa 2D Visualization: human dopamine D5 receptor targets $G\alpha$ -s G-Proteins, which interact with the following four families of effectors: Adenylate cyclases, Regulators of G-Protein signaling, Tyr protein kinases and tubulins.

participate in the calcium-signaling pathway. In general, the two receptors seem to share similar functions. Based on that, a researcher may assume that chemicals that interact with one of the receptors may also interact with the other, leading to similar results (if a chemical is an agonist for one receptor, probably it will also be for the other one too). Therefore drugs known to affect the function of Prostanoid TP receptor (for this receptor more drugs are known) may also affect Prostanoid FR receptor too. Prostanoid FR receptor is not yet related with any disease (according to OMIM [26]), nevertheless there are 12 drugs related with this receptor. One of these drugs is Latanoprost, which is used for controlling the progression of glaucoma or ocular hypertension by reducing intraocular pressure and is a prostaglandin analogue. Based on the known usage of this drug and also the fact that Prostanoid FR receptor participates in the calcium-signaling pathway, there are indications that this specific receptor may be related with hypertension. As already shown, the combination of different information retrieved from Human-gpDB may help the researchers to design specific experiments by which they will clarify the pathways in which the receptors participate, propose a mechanism for the specific disease in the case of Prostanoid TP receptor, propose a relation between Prostanoid FR receptor and hypertension and/or comprehend the side effects of drugs.

Conclusions

Human-gpDB compared to the previous gpDB databases (39,40) is now richer and focuses only on human GPCRs, G-proteins and effectors. Human-gpDB is not simply a gpDB subset, since it contains more recent data (last update of gpDB was done in March 2008, whereas all data of Human-gpDB were retrieved until December 2010), but it also contains new information concerning the classification of GPCRs (11 new subfamilies were added and all existing subfamilies are classified based on the IUPHAR classification) and also contains interactions between all molecules. It is fully integrated with external data sources by bridging information that did not exist in the previous versions (e.g. drugs and chemicals) and it now comes with a new user-friendly environment supported by advanced visualization techniques. The interface makes the navigation friendlier, the exploration of information more efficient and the extraction of new knowledge easier. Human-gpDB database was built to provide a simple but yet a powerful tool for researchers in the life sciences field as it integrates a current, careful collection of human GPCRs, G-proteins, effectors and their interactions. Human-gpDB uses advanced visualization techniques to make the volume of data more informative and the advanced data integration techniques make Human-gpDB a unique tool, a reference

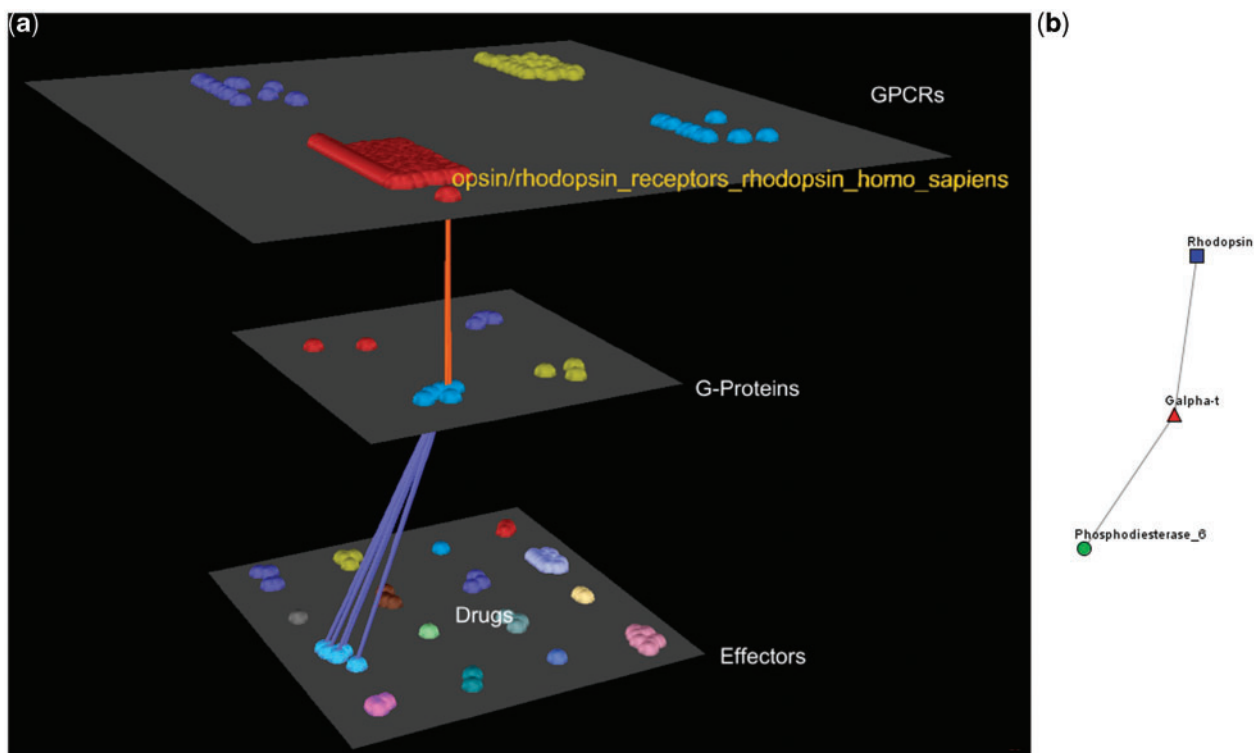


Figure 4. Visualization of human Rhodopsin receptors' interactions. (a) Arena3D Visualization: Human Rhodopsin receptor targets three $G\alpha$ G-Proteins that belong to the $G\alpha-t$ subfamily. The three G-Proteins interact with five effectors belonging to the Rhodopsin-sensitive cGMP-specific PDEases subfamily and more specifically to Phosphodiesterase 6 type of effectors. (b) Medusa 2D Visualization: the Rhodopsin subfamily of the Opsin/Rhodopsin family of the Class A of the GPCRs interacts with the $G\alpha-t$ subfamily of the G-Proteins which interact with the Rhodopsin-sensitive cGMP-specific PDEases effectors' subfamily.

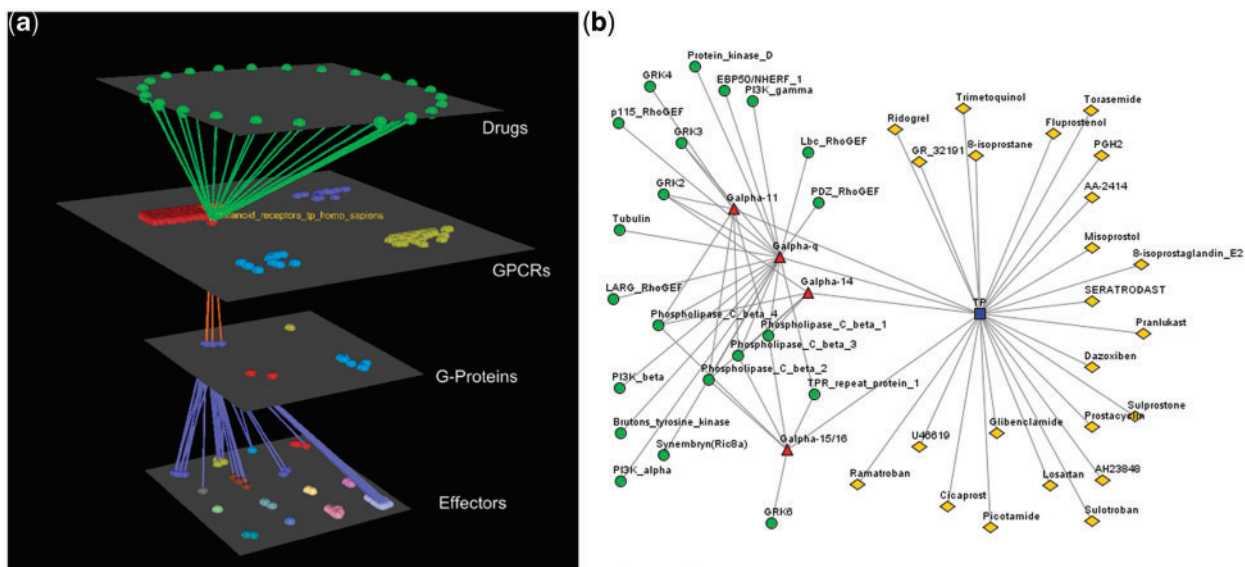


Figure 5. Visualization of human Prostanoid TP receptor's interactions (drugs included). (a) Arena3D Visualization: Human Prostanoid TP receptor protein of Class A GPCR family targets four $G\alpha$ G-Proteins that belong to $G\alpha_{q/11}$ family. The G-Proteins are connected to effectors proteins belonging to eight specific families. For this specific receptor 23 different drugs exist. (b) Medusa 2D Visualization: Human Prostanoid TP receptor protein targets $G\alpha$ G-Proteins that belong to $G\alpha_{q/11}$ family. These G-Proteins interact with 11 different subfamilies of effectors. For this specific receptor 23 different drugs exist.

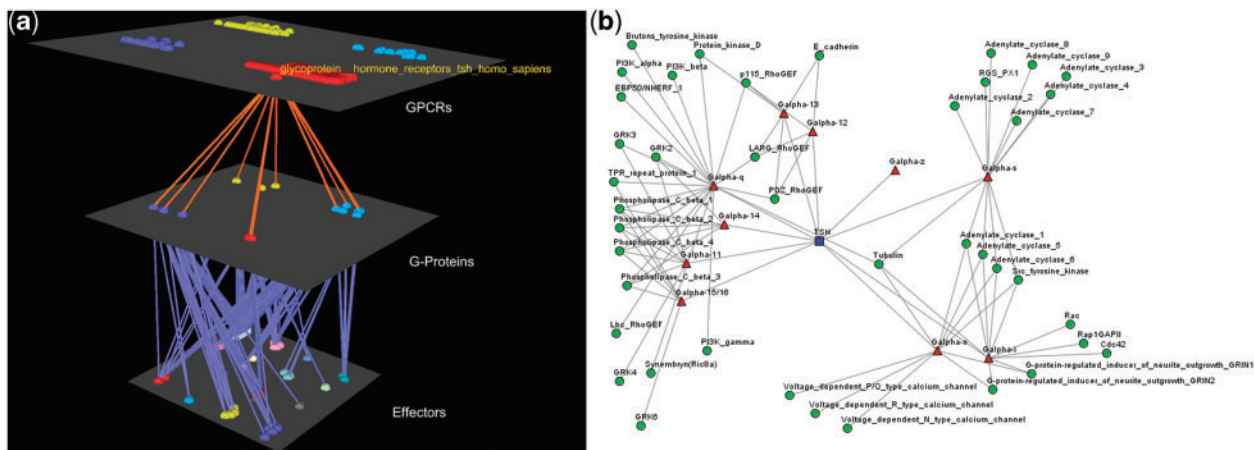


Figure 6. Visualization of human Glucoprotein Hormone TSH receptors' interactions. (a) Arena3D Visualization: Human Glucoprotein Hormone TSH receptor targets 13 G-Proteins of 10 different subfamilies, which belong to all four $G\alpha$ families. These G-proteins target 16 different families of effectors. (b) Medusa 2D Visualization: Human Glucoprotein Hormone TSH receptor targets 13 $G\alpha$ G-Proteins from all four $G\alpha$ families. More accurately, these $G\alpha$ G-Proteins belong to 10 respective $G\alpha$ subfamilies and interact with 19 subfamilies of effectors.

guide in pharmaceutical research and especially in the areas of chemical and drug discovery for human diseases. In the future, the expansion of the current version of the database for other organisms starting from the ones that are evolutionarily closer to Humans is essential.

Availability and requirements

Currently, two Human-gpDB servers are set up, one running at EMBL (http://schneider.embl.de/human_gpdb) and the other running at the Department of Cell Biology and Biophysics of the University of Athens (http://bioinformatics.biol.uoa.gr/human_gpdb/). Both servers hold the same copy of the Human-gpDB database. Concerning the linking to Human-gpDB from external sources, other databases can link to our database by using, for example, the following URLs: http://schneider.embl.de/cgi-bin/human_gpdb.cgi?search=P21918 or http://schneider.embl.de/cgi-bin/human_gpdb.cgi?search=DRD5_HUMAN, based on Uniprot ID or Accession Number.

Acknowledgements

The authors thank the reviewers of this article for their useful criticism. M.C.T. collected the data, V.P.S. developed the web application and data integration, G.A.P. worked on visualization tools, N.C.P., C.K.S., P.G.B. helped in the development of the application, R.S. and S.J.H. supervised the project. All authors are involved in the writing of the article.

Funding

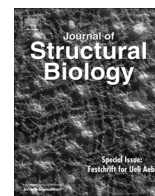
University of Athens, the University of Central Greece and EMBL, Heidelberg. Funding for open access charge: EMBL, Heidelberg.

Conflict of interest. None declared.

References

- Oldham, W.M. and Hamm, H.E. (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat. Rev. Mol. Cell. Biol.*, **9**, 60–71.
- Harmar, A.J., Hills, R.A., Rosser, E.M. *et al.* (2009) IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic Acids Res.*, **37**, D680–D685.
- Horn, F., Bettler, E., Oliveira, L. *et al.* (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.
- Kolakowski, L.F. Jr (1994) GCRDB: a G-protein-coupled receptor database. *Receptors Channels*, **2**, 1–7.
- Kristiansen, K. (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.*, **103**, 21–80.
- Pierce, K.L., Premont, R.T. and Lefkowitz, R.J. (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell. Biol.*, **3**, 639–650.
- McCudden, C.R., Hains, M.D., Kimple, R.J. *et al.* (2005) G-protein signaling: back to the future. *Cell. Mol. Life Sci.*, **62**, 551–577.
- Cabrera-Vera, T.M., Vanhauwe, J., Thomas, T.O. *et al.* (2003) Insights into G protein structure, function, and regulation. *Endocrine Rev.*, **24**, 765–781.
- Attwood, T.K. (2001) A compendium of specific motifs for diagnosing GPCR subtypes. *Trends Pharmacol. Sci.*, **22**, 162–165.

10. UniProt Consortium. (2009) The universal protein resource (UniProt) 2009. *Nucleic Acids Res.*, **37**(Database issue), D169–D174.
11. Hubbard, T.J., Aken, B.L., Ayling, S. et al. (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**(Database issue), D690–D697.
12. Smedley, D., Haider, S., Ballester, B. et al. (2009) BioMart—biological queries made easy. *BMC Genomics*, **10**, 22.
13. Etzold, T. and Verde, G. (1997) Using views for retrieving data from extremely heterogeneous databanks. *Pac. Symp. Biocomput.* 134–141
14. Jimenez, R.C., Quinn, A.F., Garcia, A. et al. (2008) Dasty2, an Ajax protein DAS client. *Bioinformatics*, **24**, 2119–2121.
15. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**(Database issue), D61–D65.
16. Schuler, G.D., Epstein, J.A., Ohkawa, H. et al. (1996) Entrez: molecular biology database and retrieval system. *Methods Enzymol.*, **266**, 141–162.
17. Cochrane, G., Aldebert, P., Althorpe, N. et al. (2006) EMBL nucleotide sequence database: developments in 2005. *Nucleic Acids Res.*, **34**(Database issue), D10–D15.
18. Maglott, D., Ostell, J., Pruitt, K.D. et al. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**(Database issue), D54–D58.
19. Letunic, I., Doerks, T. and Bork, P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.*, **37**(Database issue), D229–D232.
20. Mulder, N.J., Apweiler, R., Attwood, T.K. et al. (2002) InterPro: an integrated documentation resource for protein families, domains and functional sites. *Brief. Bioinform.*, **3**, 225–235.
21. Sammut, S.J., Finn, R.D. and Bateman, A. (2008) Pfam 10 years on: 10,000 families and still growing. *Brief. Bioinform.*, **9**, 210–219.
22. Finn, R.D., Tate, J., Mistry, J. et al. (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**(Database issue), D281–D288.
23. Berman, H.M., Battistuz, T., Bhat, T.N. et al. (2002) The protein data bank. *Acta Crystallogr. D Biol. Crystallogr.*, **58**(Pt 6 No 1), 899–907.
24. Sander, C. and Schneider, R. (1993) The HSP data base of protein structure-sequence alignments. *Nucleic Acids Res.*, **21**, 3105–3109.
25. Schafferhans, A., Meyer, J.E. and O'Donoghue, S.I. (2003) The PSSH database of alignments between protein sequences and tertiary structures. *Nucleic Acids Res.*, **31**, 494–498.
26. Hamosh, A., Scott, A.F., Amberger, J. et al. (2002) Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.
27. Harris, M.A., Clark, J., Ireland, A. et al. (2004) The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**(Database issue), D258–D261.
28. Wishart, D.S., Knox, C., Guo, A.C. et al. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.*, **37**(Database issue), D603–D610.
29. Wishart, D.S. (2008) DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics*, **9**, 1155–1162.
30. Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
31. Mi, H., Guo, N., Kejariwal, A. et al. (2007) PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.*, **35**(Database issue), D247–D252.
32. Matthews, L., Gopinath, G., Gillespie, M. et al. (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.*, **37**(Database issue), D619–D622.
33. Gunther, S., Kuhn, M., Dunkel, M. et al. (2008) SuperTarget and matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, **36**(Database issue), D919–D922.
34. Welcome to AKS2. [http://www.bioalma.com/aks2/index.php] (30 July 2010, date last accessed).
35. Pavlopoulos, G.A., O'Donoghue, S.I., Satagopam, V.P. et al. (2008) Arena3D: visualization of biological networks in 3D. *BMC Systems Biol.*, **2**, 104.
36. Hooper, S.D. and Bork, P. (2005) Medusa: a simple tool for interaction graph analysis. *Bioinformatics*, **21**, 4432–4433.
37. Kuhn, M., von Mering, C., Campillos, M. et al. (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, **36**(Database issue), D684–D688.
38. Lopez, R., Silventoinen, V., Robinson, S. et al. (2003) WU-Blast2 server at the European bioinformatics institute. *Nucleic Acids Res.*, **31**, 3795–3798.
39. Elefantioti, A.L., Bagos, P.G., Spyropoulos, I.C. et al. (2004) A database for G proteins and their interaction with GPCRs. *BMC bioinformatics*, **5**, 208.
40. Theodoropoulou, M.C., Bagos, P.G., Spyropoulos, I.C. et al. (2008) gpDB: a database of GPCRs, G-proteins, effectors and their interactions. *Bioinformatics*, **24**, 1471–1472.



Interactions of the α -subunits of heterotrimeric G-proteins with GPCRs, effectors and RGS proteins: A critical review and analysis of interacting surfaces, conformational shifts, structural diversity and electrostatic potentials

Fotis A. Baltoumas, Margarita C. Theodoropoulou, Stavros J. Hamodrakas *

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01, Greece

ARTICLE INFO

Article history:

Received 4 December 2012
Received in revised form 6 March 2013
Accepted 11 March 2013
Available online 21 March 2013

Keywords:

G-proteins
G-protein-coupled receptors (GPCRs)
RGS
Effectors
Signal transduction

ABSTRACT

G-protein coupled receptors (GPCRs) are one of the largest families of membrane receptors in eukaryotes. Heterotrimeric G-proteins, composed of α , β and γ subunits, are important molecular switches in the mediation of GPCR signaling. Receptor stimulation after the binding of a suitable ligand leads to G-protein heterotrimer activation and dissociation into the $G\alpha$ subunit and $G\beta\gamma$ heterodimer. These subunits then interact with a large number of effectors, leading to several cell responses. We studied the interactions between $G\alpha$ subunits and their binding partners, using information from structural, mutagenesis and Bioinformatics studies, and conducted a series of comparisons of sequence, structure, electrostatic properties and intermolecular energies among different $G\alpha$ families and subfamilies. We identified a number of $G\alpha$ surfaces that may, in several occasions, participate in interactions with receptors as well as effectors. The study of $G\alpha$ interacting surfaces in terms of sequence, structure and electrostatic potential reveals features that may account for the $G\alpha$ subunit's behavior towards its interacting partners. The electrostatic properties of the $G\alpha$ subunits, which in some cases differ greatly not only between families but also between subfamilies, as well as the G-protein interacting surfaces of effectors and regulators of G-protein signaling (RGS) suggest that electrostatic complementarity may be an important factor in G-protein interactions. Energy calculations also support this notion. This information may be useful in future studies of G-protein interactions with GPCRs and effectors.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

G-protein coupled receptors (GPCRs) are one of the largest and most diverse groups of cell membrane receptors in eukaryotic organisms. Nearly 800 human genes encode GPCRs that mediate most cellular responses to hormones and neurotransmitters, as well as sensory stimuli (Bjarnadottir et al., 2006). Furthermore, these receptors have been identified as key elements in a number of diseases, including various metabolism and nervous system disorders, some types of cancer and HIV infection. As a result, today, GPCRs are the targets for ~30% of pharmaceuticals on the market

(Oldham and Hamm, 2008). Several classification systems have been proposed for this superfamily categorization, each focusing on different GPCR aspects. The most popular system (Kolakowski, 1994) classifies GPCRs into six families based on their sequence homology and functional similarity.

All GPCRs are characterized by the presence of seven transmembrane α -helical segments, an extracellular N-terminus, an intracellular C-terminus and three interhelical loop regions on each side of the membrane (Kristiansen, 2004; Rosenbaum et al., 2009). This widely accepted common GPCR topology has been confirmed by crystal structures which include Rhodopsin, β_2 and β_1 adrenergic receptors, $\alpha_2\alpha$ adenosine, CXCR₄ chemokine, D₃ dopamine and H₁ histamine, M₂ and M₃ muscarinic, S1P₁ sphingosin, and the recently solved κ , μ , δ , and nocicepin/orphanin FQ opioid receptors (Chien et al., 2010; Granier et al., 2012; Haga et al., 2012; Hanson et al., 2012; Jaakola et al., 2008; Kruse et al., 2012; Manglik et al., 2012; Palczewski et al., 2000; Rasmussen et al., 2007; Shimamura et al., 2011; Thompson et al., 2012; Warne et al., 2008; Wu et al., 2010, 2012). Furthermore, a recent NMR study revealed the three-dimensional structure of another human chemokine receptor, CXCR₁ (Park et al., 2012).

Abbreviations: GPCR, G-protein-coupled receptor; $G\alpha$, α subunit of heterotrimeric G-proteins; $G\beta$, β subunit of heterotrimeric G-proteins; $G\gamma$, γ subunit of heterotrimeric G-proteins; β_2 AR, β_2 adrenergic receptor; RGS, regulator of G-protein signaling; GAP, GTPase accelerating protein; GRK2, G-protein coupled receptor kinase 2; PDE γ , phosphodiesterase γ ; PLC β_3 , phospholipase C β_3 ; RhoGEF, rho guanine nucleotide exchange factor; LARG, leukemia associated rho guanine nucleotide exchange protein.

* Corresponding author. Fax: +30 210 7274254.

E-mail addresses: fotis.baltoumas@gmail.com (F.A. Baltoumas), mtheodo@bio.uoa.gr (M.C. Theodoropoulou), shamodr@biol.uoa.gr (S.J. Hamodrakas).

Table 1
The $G\alpha$ subunit families and subfamilies.

Families	Subfamilies	Subunits with available structures
$G\alpha_{i/o}$	$G\alpha_i$	$G\alpha_{i1}$: 1GP2, 1GIA etc. $G\alpha_{i3}$: 2V4Z, 2IHB etc.
	$G\alpha_t$	$G\alpha_t$: 1TAG, 2FQJ etc.
	$G\alpha_o$	$G\alpha_o$: 3C7 K
	$G\alpha_z$	
$G\alpha_s$	$G\alpha_s$	$G\alpha_s$: 1AZT, 1AZS etc.
	$G\alpha_{of}$	
$G\alpha_{q/11}$	$G\alpha_q$	$G\alpha_q$: 3AH8, 2RGN etc.
	$G\alpha_{11}$	
	$G\alpha_{15/16}$	
$G\alpha_{12/13}$	$G\alpha_{12}$	$G\alpha_{12}$: 1ZCA
	$G\alpha_{13}$	$G\alpha_{13}$: 1ZCB, 3CX8 etc.

Heterotrimeric G-proteins are molecular switches that regulate intracellular signaling cascades in response to GPCR activation. They are composed by α , β and γ subunits, and possess a binding site for GTP (active conformation) or GDP (inactive), located in the $G\alpha$ subunit (McCudden et al., 2005). In mammals there are 21 $G\alpha$ subunits encoded by 16 genes, 6 $G\beta$ subunits encoded by 5 genes and 12 $G\gamma$ subunits (Downes and Gautam, 1999). G-proteins are typically grouped depending on $G\alpha$ similarity into four main classes, $G\alpha_s$, $G\alpha_{i/o}$, $G\alpha_{q/11}$ and $G\alpha_{12/13}$ (Cabrera-Vera et al., 2003) (Table 1). Stimulation of GPCRs by agonists leads to the activation of G-proteins, which dissociate to $G\alpha$ and $G\beta\gamma$. These subunits then interact with several effector molecules, leading to a wide range of cellular and physiological responses. The signal is terminated when the intrinsic GTPase activity of $G\alpha$ hydrolyzes GTP to GDP, thus shifting $G\alpha$ to the inactive state.

This step is often catalyzed by proteins known as regulators of G-protein signaling (RGS) that bind $G\alpha$ and accelerate GTP hydrolysis (Oldham and Hamm, 2008). The RGS domain, a conserved fold composed of nine α -helices, regulates this activity. RGS-containing proteins are grouped into eight subfamilies, four of which (R4, R7, RZ, R12) contain members that regulate the GTPase activity of $G\alpha$ subunits (Soundararajan et al., 2008).

A number of crystal structures of different $G\alpha$ subunits from all four families, as well as $G\beta\gamma$ heterodimers, are available and have provided the framework for understanding the basis of G-protein signaling (Table 1). $G\alpha$ subunits adopt a conserved fold composed of an α -helical N-terminus, a helical domain of six α -helices and a GTPase domain formed by five α -helices surrounding a β -sheet of six antiparallel strands (Lambright et al., 1994) (Fig. 1A). The GTPase domain hydrolyzes GTP and provides most of $G\alpha$'s binding surfaces for $G\beta\gamma$, receptors, effectors and G-protein regulators such as RGS proteins. Three flexible sites in this domain, called Switches I, II and III may adopt different conformations during $G\alpha$ activation. The $G\beta$ subunit adopts a seven bladed propeller structure composed by seven WD40 sequence repeats and an α -helical N-terminus. $G\gamma$ is a small subunit composed by two α -helices that bind to the N-terminus and the 5th and 6th blade of $G\beta$. $G\beta$ binds to $G\alpha$ by contacting a hydrophobic pocket made by Switches I and II, as well as a part of the N-terminus. There is no structural evidence for direct interactions between $G\alpha$ and $G\gamma$ (Wall et al., 1995).

Effectors form a diverse group of proteins that, through their interactions with $G\alpha$ and $G\beta\gamma$, either act as second messengers or lead to direct physiological responses (Kristiansen, 2004). Many proteins can act as effectors, including enzymes such as adenylyl cyclases and phospholipases, ion channels, adhesion proteins and tubulins, and various studies over the years have identified a large number of novel G-protein effectors (Hewavitharana and Wedegartner, 2012; Woehler and Ponimaskin, 2009). Each $G\alpha$ family and $G\beta\gamma$ dimer can bind to a number of different effectors, and many effectors can be regulated by more than one G-proteins.

In this work, our goal was to study the interactions between $G\alpha$ subunits, GPCRs and effectors, and to arrive at structural implications for these interactions, using information both from crystal structures and from experimental and computational studies.

2. Methods

2.1. Data collection and interface identification

Initially, we performed an extensive literature search on the interactions of G-proteins with GPCRs, effectors and RGS proteins, gathering information concerning solved structures, mutagenesis experiments and computational studies. We then compiled a dataset of solved structures of $G\alpha$ subunits, as well as several structures of various RGS proteins from the Protein Data Bank (Berman et al., 2000). All G-protein structures were examined in terms of resolution, number of $G\alpha$ chains, Guanine nucleotide binding state and the presence of mutations, disordered or truncated regions and interacting partners (Supplementary Table S1). Our initial dataset was updated each time a new $G\alpha$ structure was published.

In several occasions, the literature accompanying solved structures did not provide enough information regarding specific residues participating in interactions. Therefore, we conducted our own interface identification of $G\alpha$ complexes. Residues participating in interactions of $G\alpha$ subunits with their binding partners were identified by analyzing the structures of $G\alpha$ complexes with SPPIDER (Porollo and Meller, 2007). SPPIDER calculates the difference in Relative Solvent Accessible (RSA) surface values between the unbound and bound protein chains of a complex for each residue, and applies a user defined cut-off for the identification of interacting residues. In our analysis we applied the default cut-off value of 4% RSA. A sum of interactions between $G\alpha$ subunits and their partners is presented in Fig. 1C and Table 2. Sequences for all $G\alpha$ subunits and RGS domains with solved structures were obtained from UniProt (UniProt Consortium, 2012) and aligned with Clustal X 2.1 (Larkin et al., 2007). Further editing of alignment results was performed with JalView 2.7 (Clamp et al., 2004; Waterhouse et al., 2009).

2.2. Structural comparison and electrostatics

Having located interacting sites and residues on $G\alpha$ sequences and structures, the next step was to examine how these areas behave during G-protein activation, by comparing structures of inactive and active or transition state $G\alpha$ subunits through structural alignments for $G\alpha_{i1}$, $G\alpha_t$, $G\alpha_s$, $G\alpha_q$ and $G\alpha_{13}$. We also compared $G\alpha$ families and subfamilies, superimposing structures of $G\alpha_{i1}$, $G\alpha_{i3}$, $G\alpha_t$, $G\alpha_o$, $G\alpha_s$, $G\alpha_q$, $G\alpha_{12}$ and $G\alpha_{13}$ subunits. Criteria applied in the selection of $G\alpha$ structures was resolution and resemblance of sequence and structure to wild type G-proteins. When available, the alignments were performed using structures of similar resolution (2.2–2.9 Å) and with the minimum number of mutations and truncated or disordered regions possible. Structural alignments and RMSD calculations were performed with DaliLite v.3 on the Dali Server (Holm and Rosenstrom, 2010), as well as PyMol v.1.2 (DeLano, 2002). Distance measurements and modeling were prepared with PyMol. Calculated RMSD values are presented in Table 3 and Supplementary Tables S2 and S3.

Finally, we calculated the electrostatic potential of members from the four $G\alpha$ families using the Poisson–Boltzmann equation. The electrostatic potential was also calculated for effectors appearing in structures of complexes with $G\alpha$ subunits, as well as specific RGS proteins. Calculations for atomic radii, charges and hydrogen atoms were prepared with PDB2PQR 1.7, (Dolinsky et al., 2004, 2007; Unni et al., 2011), using the PARSE force field. Potential maps

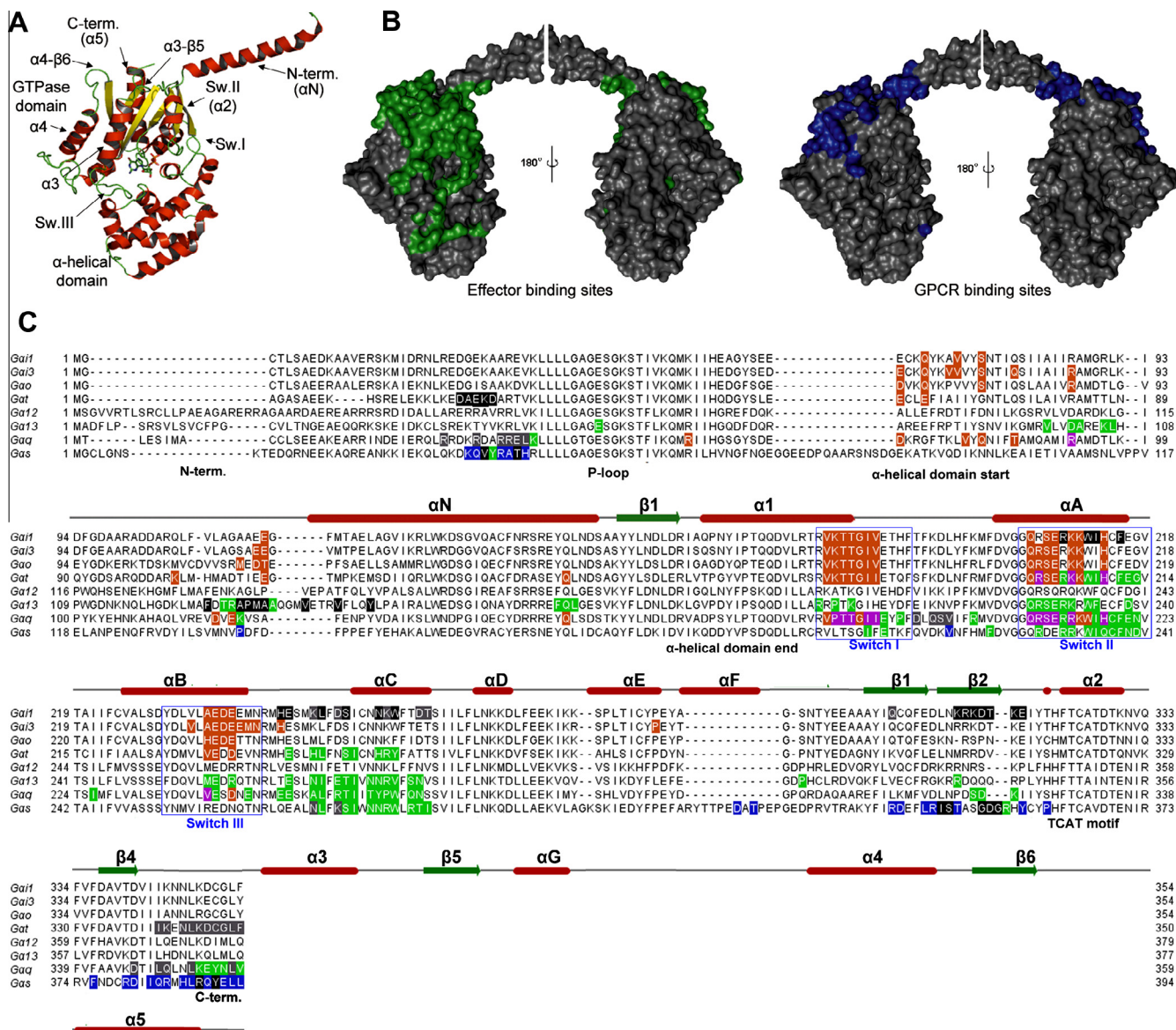


Fig. 1. A. Structural elements of $G\alpha$ subunits. B. Effector and GPCR interacting regions on the surface of $G\alpha$, according to crystal structures, mutagenesis and computational studies. Structures are colored gray, while contact sites of effectors and GPCRs are colored green and blue, respectively. The structure of $G\alpha$ in A and B is inactive $G\beta\gamma$ -bound $G\alpha_1$ (PDB: 1GP2). C. Sequence alignment of $G\alpha$ subunits with solved crystal structures. Secondary structure is represented by red cylinders for helices and green arrows for strands. The Switch regions are identified with cyan boxes. $G\alpha$ residues participating in interactions, as are identified in the crystal structures of complexes, are orange for RGS proteins, green for effectors, purple for both and blue for GPCRs. Additional interactions suggested by complexes of receptors or G-proteins with peptides, mutagenesis and computational studies are colored grey for GPCRs and black for effectors. The sequences used are rat $G\alpha_{11}$ (UniProt: P10824), human $G\alpha_{13}$ (UniProt: P08754), bovine $G\alpha_4$ (UniProt: P04695) and $G\alpha_5$ (UniProt: P04896), and mouse $G\alpha_6$ (UniProt: P18872), $G\alpha_{12}$ (UniProt: P27600), $G\alpha_{13}$ (UniProt: P27601) and $G\alpha_q$ (UniProt: P21279).

were calculated with APBS 1.3 (Baker et al., 2001; Unni et al., 2011). Temperature was set to 298.15 K, and biomolecular and solvent dielectric constants were set to 2 and 78.54, respectively. Modeling and presentation of results were prepared with PyMol.

2.3. Calculation of intermolecular energies

In order to evaluate the importance of electrostatics in $G\alpha$ interactions, we performed a series of energy analyses in known $G\alpha$ -RGS and $G\alpha$ -effector complexes. Structures of the complexes were subjected to calculation of intermolecular energies using HADDOCK v. 2.1 (de Vries et al., 2007) through the HADDOCK web server (de Vries et al., 2010). HADDOCK is a data-driven protein-protein docking approach that incorporates structural knowledge (ambiguous interaction restraints, AIRs), derived from various

experimental and/or computational methods, to drive the procedure (Dominguez et al., 2003). The HADDOCK docking protocol consists of three stages, namely, a rigid body energy minimization, a semi flexible refinement in torsion angle space and a final refinement in explicit solvent (de Vries et al., 2010).

For the calculation of intermolecular energies the last stage was utilized. The structures of the complexes were subjected to a gentle refinement in explicit water (van Dijk and Bonvin, 2006), using the "Refinement Interface" of the web server (Kastritis et al., 2012). The coordinates of ions such as Mg^{2+} and Ca^{2+} were retained, since they often play a vital part in $G\alpha$ functions. The results generated by HADDOCK include van der Waals (E_{VDW}), electrostatic (E_{ELE}), and desolvation energies (E_{DES}), buried surface area (BSA) and the HADDOCK score, a weighted sum of the above in combination with AIRs restraint energy, which can be used as criteria in the selection

Table 2
GPCR and effector interaction sites on the $G\alpha$ subunit, identified from crystal structures.

Structures	$G\alpha$	PDB ID	N-term.	Sw. I	Sw. II	Sw. III	$\alpha 3$	$\alpha 3$ – $\beta 5$	$\alpha 4$ – $\beta 6$	C-term.	Other
<i>GPCR Complex</i>											
$G\alpha\beta\gamma$ – β_2 AR	$G\alpha_s$	3SN6	+ ^b	– ^c	–	–	–	–	+	+	+($\beta 2$ – $\beta 3$, αG – $\alpha 4$, helical domain)
<i>Effector complexes</i>											
$G\alpha_s$ –Adenylyl Cyclase	$G\alpha_s$	1AZT etc.	+	+	+	–	+	+	–	–	+(helical domain)
PDE γ – $G\alpha_t$ –RGS9	$G\alpha_{t/i}$ ^a	1FQJ	–	–	+	–	+	+	–	–	–
$G\alpha_{13/i}$ –p115RhoGEF rgRGS	$G\alpha_{13/i}$ ^a	1SHZ	–	+	+	+	+	+	–	–	+(helical domain)
$G\alpha_q$ –GRK2– $G\beta\gamma$	$G\alpha_q$	2BCJ	–	–	+	–	+	+	–	–	–
$G\alpha_q$ –p63RhoGEF–RhoA	$G\alpha_q$	2RGN	–	–	+	–	+	+	+	+	–
$G\alpha_{13}$ –PDZRhoGEF rgRGS	$G\alpha_{13}$	3CX8 etc.	–	+	+	+	+	+	+	–	+(helical domain)
$G\alpha_q$ –PLC $\beta 3$	$G\alpha_q$	3OHM	+	+	+	+	+	+	+	–	+(helical domain)
$G\alpha_{13}$ –p115RhoGEF rgRGS	$G\alpha_{13}$	1SHZ	–	+	+	+	+	+	+	–	+(helical domain)
$G\alpha_{11}$ –KB-1753 phage display peptide	$G\alpha_{11}$	2G83	–	–	+	–	+	+	–	–	–

^a Chimeric subunits with Switch regions replaced with $G\alpha_{11}$.

^b The plus (+) sign indicates the sites on $G\alpha$, which interact in each respective complex.

^c The minus (–) sign indicates the sites on $G\alpha$ that do not interact.

Table 3
RMSD (Å) values of aligned active and inactive $G\alpha$ subunits.

$G\alpha$	PDB ^a	RMSD (Å)
$G\alpha_{11}$	1GP2 (A)–1GIA (A)	1.5
$G\alpha_t$	1TAG (A)–1TND (A)	1.1
$G\alpha_q$ ^b	3AH8 (A)–3OHM (A)	1.9
$G\alpha_{13}$	1ZCB (A)–3CX8 (A)	1.4
$G\alpha_s$ GTPase domain ^c	3SN6 (A)–1AZT (A) GTPase domains	2.1

^a The PDB IDs and chains (in parentheses) of structures used.

^b In the case of $G\alpha_q$, the GDP–AlF₄[–] bound subunit is used as an active state.

^c RMSD (Å) value of superimposed GTPase domains of active and empty-state $G\alpha_s$. Alignment of complete subunits gives a higher RMSD value (18.1 Å) due to the vast movement of the helical domain.

of the best docking solution. Since our goal was energy calculation of known structures of complexes, the score is omitted from presentation. CNS 2.1 is utilized for performing structural calculations (Brunger et al., 1998). Non bonded interactions are calculated with the OPLS force field with a cut-off of 8.5 Å (Jorgensen and Tirado-Rives, 1988). The electrostatic potential (E_{ELE}) is calculated by using a shift function, while a switching function (between 6.5 and 8.5 Å) is used to define the van der Waals potential (E_{VDW}) (de Vries et al., 2010; Dominguez et al., 2003). The results of the calculations are presented in Supplementary Tables S4 and S5.

3. Results and discussion

3.1. The GPCR– $G\alpha$ complex

Until recently little was known concerning GPCR–G-protein interactions. Mutagenesis studies and trials with G-protein chimeras suggested that $G\alpha$ subunits use primarily their C-terminus and $\alpha 4$ – $\beta 6$ loop to interact with the cytoplasmic pocket opened by receptor activation (Aris et al., 2001; Bae et al., 1999; Cai et al., 2001; Hamm et al., 1988). Crystal structures of complexes between Rhodopsin active intermediates and the C-terminus of $G\alpha_t$ supported this assumption (Choe et al., 2011; Scheerer et al., 2008). Additional studies included residues in the N-terminus (Ho and Wong, 2000; Taylor et al., 1994), as well as more sites on the surface of the GTPase domain. Mutagenesis studies with members of the $G\alpha_s$ family showed that specific mutations in the sequence of the $\alpha 3$ – $\beta 5$ loop altered the $G\alpha$ subunit's binding affinity towards adrenergic receptors (Grishina and Berlot, 2000). Subsequent biochemical and computational studies also advocated the role of $\alpha 3$ – $\beta 5$ as a GPCR interacting site on $G\alpha$ (Yu et al., 2008). Finally,

a recent computational study proposed interactions between the acetylcholine receptor M3R and residues in the N- and C-termini, as well as the $\beta 2$ – $\beta 3$ loop of inactive $G\alpha_q$ (Hu et al., 2010).

The first breakthrough in unveiling the nature of GPCR– $G\alpha$ interactions was the solved structure of the β_2 AR– $G\alpha_s$ $G\beta\gamma$ complex (Rasmussen et al., 2011a). In the deposited structure, β_2 AR was crystallized in its activated form, adapting a conformation similar to that found in previous activated GPCR structures (Rasmussen et al., 2011b; Scheerer et al., 2008). $G\alpha_s$ appears in an intermediate, nucleotide – empty state, the most striking feature of which is the extensive movement of the α -helical domain. The complex is stabilized by the insertion of the Nb35 nanobody between the open $G\alpha$ subunit and $G\beta\gamma$, and T4 Lysozyme, which binds at the extracellular area of β_2 AR.

Interactions between the receptor and $G\alpha$ include the cytoplasmic ends of the 5th and 6th transmembrane helices and the 2nd loop of β_2 AR, and residues at the C-terminus, $\alpha 4$ – $\beta 6$ loop and N-terminal αN – $\beta 1$ loop of $G\alpha$. Additional interactions include the C-terminus of β_2 AR and residues in the $\beta 2$ – $\beta 3$ loop and α -helical region of $G\alpha_s$ (Table 2). There are no direct contacts between β_2 AR and $G\beta\gamma$, although it is possible that the latter could interact with a second receptor in cases of GPCR oligomerization. However, the lack of coordinates for the 3rd intracellular loop of β_2 AR, a region that has been indicated to be pivotal in the formation of the GPCR– $G\alpha$ complex could mean the presence of more, unobserved interactions between the receptor and $G\alpha$, perhaps with elements like the $\alpha 3$ – $\beta 5$ loop. Furthermore, this structure is a snapshot of the complex of β_2 AR with the intermediate state of $G\alpha_s$, but reveals little information regarding the early interactions of the receptor with the GDP-bound subunit.

3.2. RGS proteins use conserved residues to interact with $G\alpha$ subunits

A number of structures of complexes between $G\alpha$ subunits and various RGS proteins or RGS fragments have been deposited on the PDB, shedding light on the nature of $G\alpha$ –RGS interactions. Most subunits in these structures are from the $G\alpha_{i/o}$ family, including $G\alpha_{11}$, $G\alpha_{13}$, $G\alpha_t$ and $G\alpha_o$ (Kimple et al., 2002, 2009; Sammond et al., 2007; Slep et al., 2001, 2008; Soundararajan et al., 2008; Tesmer et al., 1997b). Furthermore, a recently solved structure of the $G\alpha_q$ –RGS2 complex has revealed the nature of interactions between RGS proteins and other $G\alpha$ families (Nance et al., 2013). The binding of RGS proteins to $G\alpha$ subunits allows the stabilization and study of the transition state for GTP hydrolysis (Kimple et al., 2009).

Many RGS proteins can regulate the activity of more than one $G\alpha$ subunits. For example, RGS4 interacts with $G\alpha_q$ as well as $G\alpha_i$, and RGS16 can regulate both $G\alpha_i$ and $G\alpha_o$ subfamilies. On the other hand, some RGS domains display specificity towards their interacting partners. For example, RGS2 is normally $G\alpha_q$ exclusive, and is the only known member of the R4 subfamily to express such selectivity. However, mutations in three specific residues enable interactions with $G\alpha_{i3}$ without affecting its GAP activity towards its original partner (Kimple et al., 2009). Other RGS domains contact $G\alpha_{i/o}$ members exclusively, and some display an even greater selectivity towards their partners. RGS6 is selective towards $G\alpha_{i1}$. RGS12 and RGS14 cannot interact with $G\alpha_{i/o}$ subunits other than $G\alpha_{i1}$ and $G\alpha_{i3}$, and RGS9 shows higher affinity towards $G\alpha_t$ rather than $G\alpha_{i1}$ and $G\alpha_{i3}$ (Slep et al., 2001; Soundararajan et al., 2008), even though most of the interface residues of these three subunits are identical or similar (Soundararajan et al., 2008). As far as $G\alpha_i$ only selective RGS domains are concerned, RGS12 and RGS14 specificity towards $G\alpha_{i1}$ has been attributed to the GoLoco motif located in their C-terminus. However, other RGS domains, such as RGS6 and RGS10, also display similar specificity, without having this motif in their sequence. In fact, the structure of the RGS10– $G\alpha_{i3}$ complex shows that this selective towards $G\alpha_i$ GAP uses identical or similar residues with other RGS proteins that regulate both $G\alpha_{i/o}$ and $G\alpha_q$ (Supplementary Fig. S1). It is worth noting that $G\alpha_i$ only selective RGS domains interact with residues in the helical domain of $G\alpha$ as well as the Switch regions of $G\alpha_{i1}$ and $G\alpha_{i3}$, and cannot regulate $G\alpha_t$.

Structural data demonstrate that RGS proteins contact specific residues in all three Switch regions, thus regulating GTPase activity, and in certain cases they also contact parts of the helical domain of $G\alpha_{i/o}$ proteins, specifically residues in the αA helix. Interacting residues in Switch I and Switch II of RGS interacting $G\alpha$ subunits are highly conserved, with $G\alpha_q$ being the most diverse. Switch III displays most differences between $G\alpha$ subunits (Fig. 1C). Several RGS proteins also contact the helical domain. Most interactions include the αA helix. Furthermore, our analysis with SPPIDER showed additional interactions with residues in the αB – αC loop in certain cases. The recently solved $G\alpha_q$ –RGS2 structure also displays interactions of this kind, which are believed to contribute to the selectivity of RGS2. Different RGS domains, even though they adopt a similar fold, show distinct diversity in their sequences. However, most interacting residues are identical or similar (Supplementary Fig. S1).

3.3. The nature of $G\alpha$ –effector interactions

Our current understanding of the structural basis of interactions between G-proteins and their effectors depends heavily on the crystal structures of $G\alpha$ –effector complexes deposited in PDB. These structures include $G\alpha_s$ bound to the catalytic C1/C2 domains of adenylyl cyclase, $G\alpha_t$ binding the phosphodiesterase γ (PDE γ) subunit in the presence of RGS9, $G\alpha_q$ in complex with GRK2 and G $\beta\gamma$, p63RhoGEF and phospholipase C β_3 , and $G\alpha_{i3}$ in complex with the N-terminal rgRGS domains of p115RhoGEF and PDZRhoGEF (Chen et al., 2005, 2008; Hajicek et al., 2011; Lutz et al., 2007; Slep et al., 2001; Tesmer et al., 1997a, 2005; Waldo et al., 2010). An additional structure of $G\alpha_{i1}$ and an active state selective peptide demonstrated similar interfaces (Johnston et al., 2006). The structural data indicate that all effectors bind to a common recognition surface of $G\alpha$, comprised of Switch II, the $\alpha 3$ helix and the $\alpha 3$ – $\beta 5$ loop (Sprang et al., 2007). Switch II and $\alpha 3$ interacting residues are highly conserved among different $G\alpha$ subunits, with $G\alpha_s$ and $G\alpha_{12/13}$ being the most diverse. However, the $\alpha 3$ – $\beta 5$ loop differs both in sequence and in structure between families, and has been suggested to be a key in the selectivity of G-proteins towards their effectors (Sprang et al., 2007).

Apart from the effector binding site of Switch II, $\alpha 3$ and $\alpha 3$ – $\beta 5$, certain effectors demonstrate additional interactions with other $G\alpha$ surfaces (Table 2). The p63RhoGEF– $G\alpha_q$ complex displays interactions of the effector with the C-terminus and the $\alpha 4$ – $\beta 6$ loop. PLC β_3 , and p115RhoGEF, which have been shown to act as GTPase accelerating proteins (GAPs) towards $G\alpha_q$ and $G\alpha_{i3}$, respectively, contact residues in Switch I, III and the N-terminal part of Switch II, which form the binding surface of RGS proteins (Fig. 1). The p115RhoGEF rgRGS domain and N-terminus additionally contact certain parts of the α -helical domain of $G\alpha_{i3}$, though not the same as most RGS proteins. PDZRhoGEF, highly homologous to p115RhoGEF, displays similar interactions, although it lacks GAP activity towards $G\alpha_{i3}$ (Chen et al., 2008).

Several interactions were also identified by mutagenesis, as well as *in silico* studies. $G\alpha_i$ subunits have been suggested to use the $\alpha 4$ – $\beta 6$ loop to bind to a site of adenylyl cyclase distinct from the one of $G\alpha_s$ (Dessauer et al., 1998; Grishina and Berlot, 1997). $G\alpha_t$ has been shown to use residues in the N-terminus to bind PDE γ , a fact not observed in the structure of the complex because $G\alpha$ had a truncated N-terminus (Grant et al., 2006). Studies using chimeric proteins identified $G\alpha_s$ and $G\alpha_{i1}$ binding to β -tubulin using residues from the GTPase domain, and a protein–protein docking trial between $G\alpha_s$ and β -tubulin suggested interactions with the common effector site, as well as the N- and C-terminus and $\alpha 4$ – $\beta 6$ (Chen et al., 2003; Dave et al., 2009; Layden et al., 2008). A recent study, combining NMR, mutagenesis and protein–protein docking trials, indicates a new interaction site on the helical domain of $G\alpha_{i3}$ for p115RhoGEF (Chen et al., 2012) formed by the DH/PH domains of the effector and the αB – αC loop of $G\alpha_{i3}$. This particular loop also contacts RGS domains in some $G\alpha$ –RGS complexes. Interface residues of $G\alpha$ that contact various binding partners are shown in Fig. 1C, displaying information both from crystal structures and from mutagenesis or computational studies.

Our analysis of several different structures of the $G\alpha_s$ –adenylyl cyclase complex shows additional interactions with some residues in Switch I and the N-terminus of $G\alpha$. Similarly, the $G\alpha_q$ –PLC β_3 complex displays interfaces with the N-terminal $\beta 1$ strand. Additional contacts are observed between p63RhoGEF and residues in the $\alpha 4$ – $\beta 6$ loop and C-terminus of $G\alpha_q$. Interactions with $\alpha 4$ – $\beta 6$ are also observed in a more recent structure of the complex between $G\alpha_{i3}$ –p115RhoGEF (Hajicek et al., 2011), as well as the complexes with PDZRhoGEF (Chen et al., 2008). In these structures $G\alpha_{i3}$ is mostly native type, while the chimeric subunit in the original $G\alpha_{i3/i1}$ –p115RhoGEF complex (Chen et al., 2005) has a number of effector binding residues replaced with the ones from $G\alpha_{i1}$.

Unlike $G\alpha$ subunits, which follow a conserved structural motif, G-protein effectors are a large and diverse group of various enzymes, ion channels, regulators and cytoskeleton components, with very different structural and functional characteristics, thus making it more difficult for researchers to establish a set of common features. $G\alpha_s$ activates adenylyl cyclase by interacting with elements in both of its cytoplasmic domains, C1 and C2 (Tesmer et al., 1997a), while mutagenesis experiments indicate that the adenylyl cyclase binding site for inhibitory $G\alpha_i$ proteins is located in its C1 domain only (Dessauer et al., 1998). Effectors such as p63RhoGEF and Phospholipase C β_3 use helix-loop-helix domains to contact the $G\alpha_q$ Switch II and $\alpha 3$ regions with residues both in the two helices and in the loop, and a similar structural feature is used by GRK2 in its complex with $G\alpha_q$ (Lutz et al., 2007; Tesmer et al., 2005; Waldo et al., 2010), however that is not the case with adenylyl cyclase, which uses residues both from multiple α -helices from the C2 domain as well as a loop and a β -strand from C1 to contact $G\alpha_s$. Mutagenesis experiments suggest that β -tubulin uses its nucleotide binding domain, formed by a β -strand, a helix and the loop connecting them, to contact $G\alpha_s$, which also differs from

the structural feature used by the three $G\alpha_q$ effectors (Chen et al., 2003; Layden et al., 2008).

The Guanine nucleotide exchange factor (GEF) activity of Rho-GEFs is located in their DH/PH domains, and p63RhoGEF contacts $G\alpha_q$ with residues in these domains (Lutz et al., 2007). On the other hand, RGS containing p115RhoGEF and PDZRhoGEF use their rgRGS or RGS-box domains to interact with $G\alpha_{13}$ in an effector like fashion, but it is not clear how this interface leads to the activation of their GEF ability, since the DH/PH domains are missing from the structures of their complexes with $G\alpha_{13}$ (Chen et al., 2005; Chen et al., 2008; Hajicek et al., 2011). However, a recent combination of biochemical and protein–protein docking studies suggests that the DH/PH domains of p115RhoGEF interact with residues in the α -helical domain of $G\alpha_{13}$ (Chen et al., 2012).

Effectors containing RGS homologous domains, such as the RH domain of GRK2 and the rgRGS of p115RhoGEF and PDZRhoGEF use these domains to interact with the common effector surfaces of $G\alpha_q$ and $G\alpha_{13}$, respectively, but these interactions do not affect the $G\alpha$ subunits' GTPase activity in any way. On the contrary, p115 acts as a GAP towards $G\alpha_{13}$ using residues outside its RGS-box, namely the N-terminal EDED motif (Chen et al., 2005). Mutation of this sequence to its PDZRhoGEF equivalent, EEDY, abolishes GAP activity towards $G\alpha_{13}$ (Chen et al., 2008). Similarly, phospholipase $C\beta_3$, while it does not have any structural similarities with RGS proteins, it displays GAP activity when bound to $G\alpha_q$, by contacting residues in the three Switch regions and the α -helical domain with its third and fourth EF hands, as well as residues in the linker between the TIM barrel and the C2 domain (Waldo et al., 2010).

3.4. Overlapping interacting sites on $G\alpha$ surfaces

The study of the literature, as well as our own observations reveal that certain surfaces of $G\alpha$ subunits can often participate in binding both effectors and GPCRs (Fig. 1, Table 2). These include the N- and C-terminus and the $\alpha 4$ – $\beta 6$ loop. The $\alpha 3$ – $\beta 5$ loop, a part of the common effector surface, has also been implicated in the regulation of $G\alpha$ activation and binding to receptors by many biochemical and mutagenesis studies, and therefore can be considered among the receptor interacting sites of $G\alpha$, even though there are no definitive structural data supporting this interface. It should be noted that certain residues in the C-terminus and $\alpha 4$ – $\beta 6$, and perhaps the N-terminus and $\alpha 3$ – $\beta 5$ sites, have been implicated in forming contacts both with effectors and with GPCRs. For instance, the residues of the LRIST peptide in the $\alpha 4$ – $\beta 6$ loop of $G\alpha_s$ have been shown to contact adenylyl cyclase $\beta_2A.R.$ and, possibly, β -tubulin. Finally, specific sites in the α -helical domain, namely the αA helix and the αB – αC loop, have been shown to contact several effectors and RGS proteins, and the β_2AR – $G\alpha_s G\beta\gamma$ complex also displays an interaction between the receptor and the helical domain of $G\alpha_s$.

3.5. Structural shifts during $G\alpha$ activation

A series of conformational shifts occur during $G\alpha$ activation. Binding of $G\alpha$ – $G\beta\gamma$ to a receptor will result in the opening of the nucleotide cleft for nucleotide exchange, by vast movement of the α -helical domain (Rasmussen et al., 2011a). $G\alpha$ activation also causes rearrangements in distinct sites of the GTPase domain. Comparison of active and inactive structures of $G\alpha_{i1}$, $G\alpha_t$ and $G\alpha_q$ subunits through structure superposition shows Switch II moving about 4–8 Å from its position in the inactive subunit, as well as a 3–6 Å movement of Switch III (Supplementary Fig. S2). In the case of inactive $G\alpha_{13}$ most of the Switch II region is disordered, but it can be assumed to move in a similar fashion. Superposition of the GTPase domains of active and empty $G\alpha_s$ also displays

shifts of the Switch regions (Supplementary Fig. S3). On the other hand, the $\alpha 3$ helix and its adjacent loop show little or no movement. The RMSD values of the alignments between active and inactive $G\alpha$ subunits are shown in Table 3.

Alignments of $G\alpha_q$ and $G\alpha_s$ active forms to inactive and empty – state subunits, respectively, presents displacement of the $\alpha 4$ – $\beta 6$ loop (3–5 Å) and C-terminus (2–10 Å), however these sites show little movement in $G\alpha_{i0}$ and $G\alpha_{13}$ (Supplementary Figs. S2 & S3). Most $G\alpha$ structures are truncated in their N-terminus, preventing full observation of its behavior during activation or binding to effectors. This also affects most structural alignments between $G\alpha$ subunits, resulting in relatively low RMSD values. However, alignment of inactive and RGS4-bound $G\alpha_{i1}$ subunits, as well as subunits expressed in the presence of ions such as SO_4^- , all of which have coordinates for the N-terminus, display vast structural changes of the N-terminal α -helix (Supplementary Fig. S4). This flexibility of the N-terminus could account for its participation in various $G\alpha$ interactions, including binding to effectors or RGS proteins.

3.6. Structural diversity of $G\alpha$ interaction sites between different subunits

Structure superpositions of different $G\alpha$ subunits show that the GTPase domain structure is conserved among the four $G\alpha$ families; still, a few deviations are observed (Supplementary Tables S2 & S3). The $\alpha 3$ – $\beta 5$ and especially the $\alpha 4$ – $\beta 6$ loops of $G\alpha_{i1}$ and $G\alpha_s$ differentiate not only in sequence but also in structural conformation, as shown by superposition of their active subunits; with the $G\alpha_s$ $\alpha 4$ – $\beta 6$ loop located ~5–6 Å away from the $G\alpha_{i1}$ loop. Similar deviations are observed in the alignments of $G\alpha_s$ with $G\alpha_q$ and $G\alpha_{12}$ (Supplementary Fig. S5). It is possible that this feature may differentiate $G\alpha$ binding to GPCRs and certain effectors. On the other hand, members of the same $G\alpha$ family show little difference in most of their sequence and structure features. Comparison of $G\alpha_{12/13}$ structures shows differences in the α -helical region, specifically the αB – αC loop, and the $\alpha 4$ – $\beta 6$ loop of $G\alpha_{13}$ appears to be one residue longer than the one of $G\alpha_{12}$. However most of the α -helical and GTPase domains show no differences between $G\alpha_{12/13}$ subunits, and almost all effector-contacting residues of $G\alpha_{13}$ are present in $G\alpha_{12}$. Similarly, $G\alpha_{i0}$ subfamilies are highly conserved in sequence and structure, with minor changes mostly located in the helical domain.

3.7. Electrostatic diversity of $G\alpha$ surfaces

Many effectors can be contacted by different G-proteins, with their $G\alpha$ interacting surfaces tolerating substantial variation. This allows the use of chimeric $G\alpha$ subunits in structural studies of G-protein–effector complexes (Sprang et al., 2007). Still, a number of binding partners display specificity in their interactions with G-proteins, even in the level of $G\alpha$ subfamily. A component of microtubules, β -tubulin, can interact with $G\alpha_{i1}$, but not $G\alpha_t$ or $G\alpha_o$ (Chen et al., 2003), even though these members of the $G\alpha_{i0}$ family are highly conserved in both sequence and structure. One additional example is p115RhoGEF, which binds and expresses GAP activity to both members of the $G\alpha_{12/13}$ family, but its GEF activity is activated only by $G\alpha_{13}$ (Hart et al., 1998). Another RhoGEF, LARG, can be stimulated by $G\alpha_q$ and $G\alpha_{13}$ but not $G\alpha_{12}$, unless it is tyrosine – phosphorylated (Booden et al., 2002; Suzuki et al., 2003), even though $G\alpha_{12}$ can still contact its unphosphorylated form. Binding specificity is also observed in several $G\alpha$ –RGS interactions, as mentioned in Section 3.2. Many of these cases cannot be fully explained by differences in amino acid sequence or secondary structure.

Comparison of the electrostatic properties of the four different families, as well as their subfamilies, indicates a potential factor in determining $G\alpha$ contacts (Fig. 2, Supplementary Fig. S6). The electrostatic potential of the otherwise highly conserved common effector site is significantly diverse among different $G\alpha$ families and, in specific cases, among members of the same family, as shown in Fig. 2. Striking examples of this diversity are $G\alpha_{12}$ and $G\alpha_{13}$, which differ greatly, in this regard, both from other $G\alpha$ subunits and from each other. The electrostatic surface potential of both subunits is generally more positive compared to other $G\alpha$ families; however the Switch II/ $\alpha 3$ binding pocket differentiates between $G\alpha_{12}$ and $G\alpha_{13}$ as well. Diversity among subfamilies is also observed in subunits from the $G\alpha_{i/o}$ family, though not as radical as in the case of $G\alpha_{12/13}$. $G\alpha_q$, compared to other $G\alpha$ subunits, has a mostly non-polar effector site. Deviations are also observed in the $\alpha 3$ – $\beta 5$ loop as well as the Switch I and III regions. Differences of electrostatic potential in surfaces formed by the α -helical domain, the $\alpha 3$ – $\beta 5$ and $\alpha 4$ – $\beta 6$ loops and C-terminus (Supplementary Fig. S6) are expected, due to the sequence diversity of these sites.

3.8. The electrostatics of RGS domains

In addition to $G\alpha$ subunits, the electrostatic properties of several RGS domains were calculated and compared. Furthermore, calculation and comparison of the intermolecular energies of known $G\alpha$ –RGS structures reveals the participation of different types of interactions in the intermolecular energy between $G\alpha$ and RGS

domains (Supplementary Table S4). In almost all $G\alpha$ –RGS complexes, the values of desolvation (E_{DES}), and van der Waals (E_{VDW}) energies are similar among different structures. Diversity is observed, in specific cases, mainly in the values of electrostatic energy (E_{ELE}). Since almost all structures contain complexes of RGS domains with $G\alpha_{i/o}$ subunits, it is not possible to compare binding energies between complexes of the same RGS with different $G\alpha$ families. However, the present data is sufficient in order to make some observations.

RGS domains with the ability to regulate both $G\alpha_{i/o}$ and $G\alpha_q$ subunits generally show very negatively charged $G\alpha$ interacting surfaces (Supplementary Fig. S7), and interact mainly with residues in the three Switch regions of the GTPase domain of $G\alpha$. These domains belong mostly to the R4 subfamily of RGS proteins. On the other hand, comparison of electrostatic properties between $G\alpha_i$ selective RGS proteins shows that the structures possess slightly less negative surfaces compared to RGS domains with no interacting specificity (Supplementary Fig. S7). It seems that, while almost all interacting residues are identical or conserved among RGS domains, the diversity of the surrounding regions leads to difference in surface potential.

An interesting case is RGS9, which displays a highly positive surface (Supplementary Fig. S7), despite its sequence similarity with other RGS domains (Supplementary Fig. S1). RGS9 interacts with $G\alpha_t$ exclusively, and also contacts the γ subunit of phosphodiesterase, as shown in the $G\alpha_t$ –PDE γ –RGS9 complex. Despite sequence and structure similarity between $G\alpha_{i/o}$ members, RGS9 lacks GAP ability towards other members of the family.

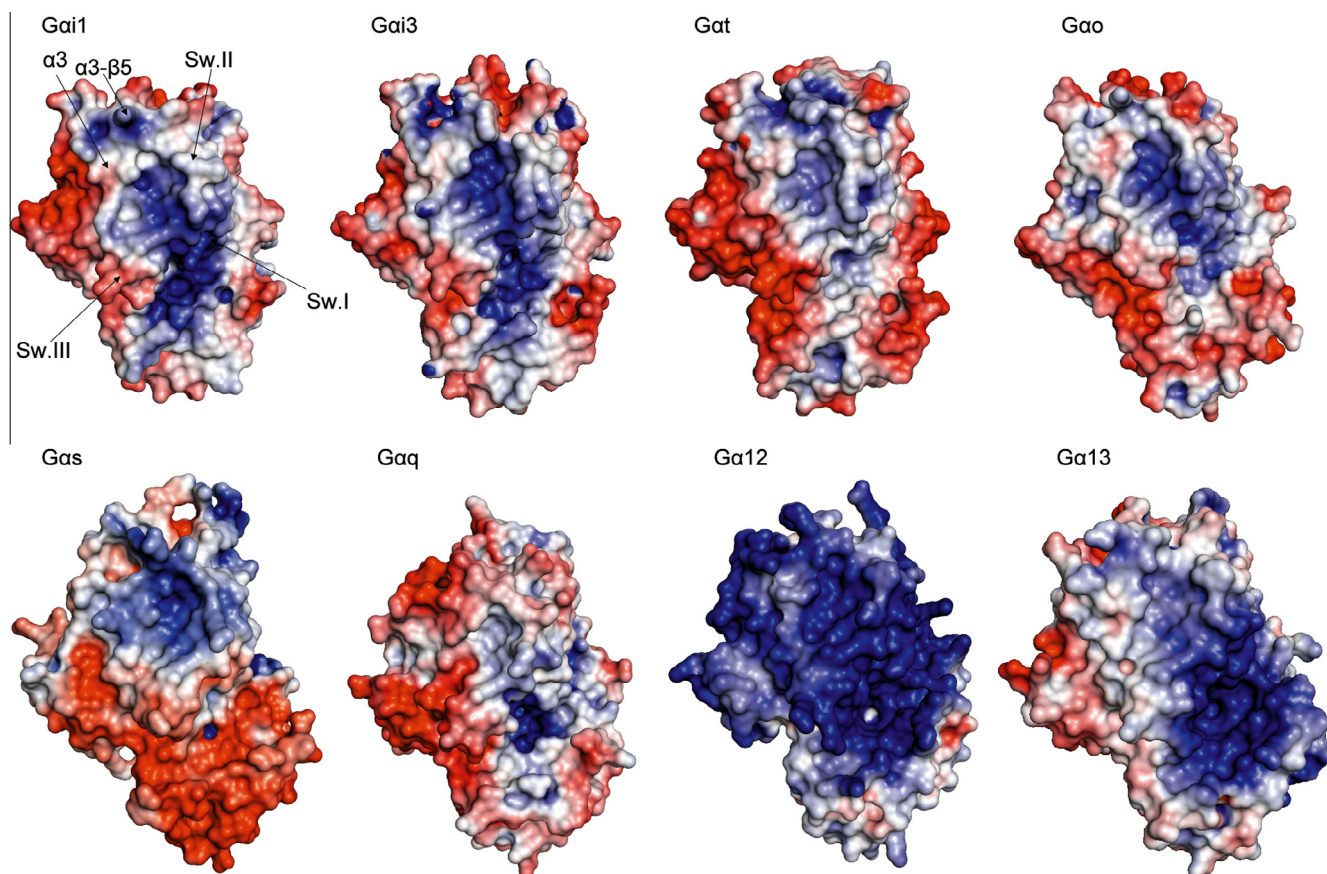


Fig. 2. Electrostatic molecular surfaces of $G\alpha$ subunits. All subunits are oriented in the same way as in Fig. 1. Charged surfaces are colored in shades of blue for positive and red for negative charges, while uncharged surfaces are colored white. Subunit surfaces are contoured from -5 (red) to $+5$ (blue) kT/e^- based on the potential of the solvent accessible surface. All subunits are in their activated state, with the exception of $G\alpha_{13}$ and $G\alpha_o$, which are in their RGS-bound state. There is a conserved, effector-binding pocket between Switch II and $\alpha 3$. A visualization of the potential of the opposite side of the $G\alpha$ subunits is shown in Supplementary Fig. S6. Crystal structures used are $G\alpha_{11}$ (PDB: 1GIA), $G\alpha_{13}$ (PDB: 2V4Z), $G\alpha_t$ (PDB: 1TND), $G\alpha_o$ (PDB: 3C7K), $G\alpha_s$ (PDB: 1AZT), $G\alpha_q$ (PDB: 3AH8), $G\alpha_{12}$ (PDB: 1ZCA) and $G\alpha_{13}$ (PDB: 3CX8).

As mentioned in Section 3.2, wild type RGS2 normally regulates only $G\alpha_q$. However, a triple mutant of the domain (C106S, N184D, E191K) gains the ability to contact and regulate $G\alpha_i$ members. Comparison of the potential between the wild type and the mutant domain reveals that the three mutations alter the potential of the C-terminal surface of RGS2, which contacts the helical domain of $G\alpha_{13}$. Specifically, the substitution of E191 to K results in a surface less charged, compared to that of the wild type (Supplementary Fig. S7). Comparison of energies also reveals a significant diversity in Electrostatic energy values (Supplementary Table S4), with a difference greater than 200 kcal/mol.

It is known from literature that mutant RGS2, while capable of regulating $G\alpha_i$, displays lower affinity towards it ($K_D = 1.25 \mu\text{M}$) compared to that of $G\alpha_q$ and RGS2 ($K_D = 22 \pm 9 \text{ nM}$) (Kimple et al., 2009; Nance et al., 2013). Through our calculations we see that the $G\alpha_q$ -RGS2 W.T. complex displays a more favorable E_{ELE} value ($\sim -500 \text{ kcal/mol}$) as opposed to $G\alpha_{13}$ -RGS2 mutant ($\sim -300 \text{ kcal/mol}$). The difference in E_{ELE} values suggests that electrostatic diversity of RGS2 surfaces, caused by the three mutations, is a factor in RGS2 affinity towards $G\alpha$ subunits.

3.9. Complementarity of effectors and $G\alpha$ interacting surfaces

Since effectors show high diversity in sequence, structure and function, it is difficult to establish a set of common features. However, calculation and comparison of their electrostatic properties show that, at least in some cases, similarities can be observed. On the other hand, the diversity of effector sizes, which range from small chains such as PDE γ to large proteins such as adenylyl cyclase or PLC β_3 , leads to vast differences in Buried Surface Area (BSA) of $G\alpha$ -effector complexes. Therefore, attempts of comparison of intermolecular energies (Supplementary Table S5) can only be made in cases with similar BSA values.

Effector interacting sites can be grouped into two categories: sites that interact with the conserved effector binding pocket, formed by Switch II and α_3 , and sites that interact with other parts of $G\alpha$, namely the N- and C-termini, the α_3 - β_5 and α_4 - β_6 loops and parts of the helical domain. The sites that contact the Switch II/ α_3 pocket generally complement its electrostatic properties. The effector pocket of both $G\alpha_s$ and $G\alpha_i$ subunits is positively charged. The known $G\alpha_s$ site and the proposed $G\alpha_i$ site on the surface of adenylyl cyclase are negatively charged, complementing their G-protein partners (Supplementary Fig. S8A).

The negative surface of phosphodiesterase γ comes in contact with the effector pocket of $G\alpha_t$, as well as the RGS9 domain, which are both relatively positively charged (Supplementary Fig. S8B). The $G\alpha_t$ -PDE γ interface is driven by hydrophobic contacts of W70 from PDE γ . However, the overall structure of PDE γ has a negatively charged surface, while the switch II/ α_3 pocket of $G\alpha_t$ has a relatively positive contour. Therefore, the charge of the surrounding regions could affect the nature of interactions, and could be a factor in the reasons why PDE γ cannot interact with other $G\alpha_{i/o}$ members. Also, the presence of RGS9 in the complex should not be neglected. It is an interesting fact that, according to HADDOCK's calculations, the E_{ELE} value of the RGS9- $G\alpha_t$ interface is more favorable in the presence of PDE γ (Supplementary Table S5).

In the case of $G\alpha_q$ regulated effectors, namely phospholipase $C\beta_3$, GRK2 and p63RhoGEF, the amphipathic sites that contact the effector pocket are mostly uncharged, and complement the generally non polar surface of the $G\alpha_q$ pocket (Supplementary Fig. S9A). Similarly, the rgRGS domains of p115RhoGEF and PDZRhoGEF come in contact with the positively charged pocket of $G\alpha_{13}$ (Supplementary Fig. S9B). This complementarity probably allows p115RhoGEF to bind to and regulate the GTPase activity of $G\alpha_{12}$ as well. Interestingly, the complexes of $G\alpha_{13}$ with p115RhoGEF and PDZRhoGEF display similar values of E_{VDW} but diverse in E_{DES}

and, more importantly, in E_{ELE} (Supplementary Table S5). The displayed diversity in E_{ELE} values could suggest difference in the affinity of the two effectors towards $G\alpha_{13}$. Indeed, measurements with isothermal titration calorimetry show differences in the K_D values for the p115RhoGEF- $G\alpha_{13/i}$ ($K_D = 3-5 \mu\text{M}$) and PDZRhoGEF- $G\alpha_{13}$ ($K_D \sim 300-500 \text{ nM}$) complexes (Chen et al., 2005; Chen et al., 2008).

Sites that contact other surfaces of $G\alpha$ also display interesting electrostatic properties. In the case of p63RhoGEF, all sites that contact the negatively charged C-terminus and α_4 - β_6 loop are highly positive. The EF hands 3 and 4 of PLC β_3 , which regulate the GTPase activity of $G\alpha_q$, are slightly positive as well, complementing the negatively charged surface of Switch III, while the TIM barrel-C2 linker, which contacts the positive surface of Switch I, is negatively charged (Supplementary Fig. S9A). HADDOCK's calculations show that $G\alpha_q$ -p63RhoGEF and $G\alpha_q$ -PLC β_3 complexes also display differences in E_{ELE} values (Supplementary Table S5). This difference can be attributed to the p63RhoGEF's positive potential in DH/PH surfaces that contact the negative surfaces of α_4 - β_6 and the C-terminus in $G\alpha_q$.

The most interesting case, however, is p115RhoGEF. While it is the rgRGS domain that makes effector-like interactions with $G\alpha_{13}$, the GEF activity is believed to be regulated by elements outside this region, specifically the DH and PH domains. As mentioned above, the rgRGS domain complements the effector surfaces of both $G\alpha_{12/13}$ proteins, however the effector can only be activated by $G\alpha_{13}$. A recent combination of biochemical, NMR and computational studies suggests that regulation of GEF ability by G-proteins is performed in distinct sites of p115RhoGEF. One is the MGMT sequence following the RGS box, which interacts with the α_3 - β_5 loop of $G\alpha_{13}$. Mutations in this site prevent GEF activation of the effector from G-proteins, thus showing its importance in p115- $G\alpha$ interactions (Chen et al., 2012). Furthermore, the α_3 - β_5 loop diverges between $G\alpha_{12}$ and $G\alpha_{13}$ in terms of electrostatic properties despite sequence similarity. Another proposed site is the surface of the DH domain, not present in the $G\alpha_{13}$ -p115 rgRGS complexes. This surface is suggested to contact the αB - αC site of the helical domain in $G\alpha_{13}$. This site diverges greatly between the two members of $G\alpha_{12/13}$, however its significance seems to be less important, since mutations of this area in $G\alpha_{13}$ do not abolish GEF activation of p115RhoGEF (Chen et al., 2012).

3.10. Electrostatic diversity as a potential factor in $G\alpha$ interactions

Diversity observed in the Switch regions of $G\alpha_{i/o}$ proteins could account for RGS coupling specificity, while the small but significant changes in the effector binding pocket between $G\alpha_{11}$ and $G\alpha_{13}$, $G\alpha_t$ and $G\alpha_o$ could explain affinity towards β -tubulin. The electrostatic properties of the α -helical domain sites that contact RGS proteins also vary between $G\alpha_{11}/G\alpha_{13}$, $G\alpha_t$ and $G\alpha_q$. This observation suggests that these sites also play a part in the selectivity of $G\alpha$ -RGS interactions. Considering the fact that the α -helical domain of $G\alpha$ subunits also contacts several effectors using residues in this site, it could be assumed that its electrostatic properties can affect interactions with effectors as well.

The potential of the effector pocket is generally positive in $G\alpha$ subunits with the exception of $G\alpha_q$. This could be one of the reasons that, certain effectors can be regulated by $G\alpha_q$ proteins, exclusively. Electrostatic complementarity could also be a fact in the activation of p115RhoGEF by $G\alpha_{13}$ rather than $G\alpha_{12}$. The potential of α_3 - β_5 loop, which contacts the MGMT motif, diverges between the two $G\alpha_{12/13}$ subunits and could, therefore, be a factor in the effector's selectivity towards them. It should be noted that $G\alpha_{12}$, on the whole, displays a positively charged solvent accessible surface, whereas $G\alpha_{13}$ in some aspects resembles other $G\alpha$ subunits. This diversity could also affect $G\alpha_{12}$'s coupling to effectors, since electrostatic properties have been shown to regulate protein-pro-

tein interactions from distances up to 10 Å (van Dijk and Bonvin, 2006).

Furthermore, we have observed that in several cases the interface area extends to the surface of the α -helical domain, specifically in regions that diverge electrostatically between different subunits. Overall, what we see is that most interacting surfaces of $G\alpha$ subunits are in many cases complementary, in terms of electrostatic properties, to the protein surfaces they come in contact with. Moreover, in certain cases, diversity in electrostatic energy values suggests that electrostatic complementarity can influence affinity. It is therefore possible that the electrostatic potential of protein surfaces may play a significant role in $G\alpha$ interacting with various effectors, RGS proteins and perhaps $G\beta\gamma$ heterodimers or GPCRs.

4. Conclusions

Heterotrimeric G-proteins are the mediators in most GPCR – mediated signaling pathways, acting as molecular switches for the regulation of a large number of cell responses, and are consequently a key element in the study of signal transduction, as well as the treatment of various related diseases. We have identified certain surfaces of $G\alpha$ subunits that can, in many cases, participate in binding both receptors and effectors. These surfaces include the N- and C-terminal regions, the $\alpha 4$ – $\beta 6$ loop and the $\alpha 3$ – $\beta 5$ loop. The differences displayed in the sequence and structure of these sites can perhaps account for $G\alpha$ specificity towards their binding partners. Furthermore, the diversity in the electrostatic potential of $G\alpha$ surfaces, combined with observed electrostatic properties of various effectors and RGS structures, suggests that electrostatic complementarity can be an additional factor in the regulation of effectors by G-proteins, as well as $G\alpha$ interactions with RGS proteins. Finally, our observations concerning the α -helical domain reveal features that strengthen the hypothesis of an occasionally important role in interactions for this part of $G\alpha$ subunits, which has been mostly neglected in the past. Information provided by this study could find applications, in future, more detailed studies of the structural basis of G-protein interactions with GPCRs and novel effectors.

Author contributions

F.A.B. and M.C.T. collected the data, performed structural comparisons and evaluated the results. F.A.B. calculated and compared the maps of the electrostatic potential. M.C.T. performed the energy calculations, analysis and comparisons. S.J.H. supervised the whole project and evaluated the results. All authors contributed to the writing of the manuscript. All authors have read and approve the final version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We thank the University of Athens for support. The present work was funded by the SYNERGASIA 2009 PROGRAMME. This Programme is co-funded by the European Regional Development Fund and National resources (Project Code O9SYN-13-999), General Secretariat for Research and Technology of the Greek Ministry of Education and Religious Affairs, Culture and Sports. We should like to thank the handling editor and the reviewers of this manuscript for their very useful and constructive criticism.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2013.03.004>.

References

- Aris, L., Gilchrist, A., Rens-Domiano, S., Meyer, C., Schatz, P.J., et al., 2001. Structural requirements for the stabilization of metarhodopsin II by the C terminus of the alpha subunit of transducin. *J. Biol. Chem.* 276, 2333–2339.
- Bae, H., Cabrera-Vera, T.M., Depree, K.M., Graber, S.G., Hamm, H.E., 1999. Two amino acids within the alpha4 helix of Galpha1 mediate coupling with 5-hydroxytryptamine1B receptors. *J. Biol. Chem.* 274, 14963–14971.
- Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037–10041.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., et al., 2000. The protein data bank. *Nucleic Acids Res.* 28, 235–242.
- Bjarnadottir, T.K., Gloriam, D.E., Hellstrand, S.H., Kristiansson, H., Fredriksson, R., et al., 2006. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* 88, 263–273.
- Booden, M.A., Siderovski, D.P., Der, C.J., 2002. Leukemia-associated Rho guanine nucleotide exchange factor promotes G alpha q-coupled activation of RhoA. *Mol. Cell. Biol.* 22, 4053–4061.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., et al., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.
- Cabrera-Vera, T.M., Vanhauwe, J., Thomas, T.O., Medkova, M., Preinerger, A., et al., 2003. Insights into G protein structure, function, and regulation. *Endocr. Rev.* 24, 765–781.
- Cai, K., Itoh, Y., Khorana, H.G., 2001. Mapping of contact sites in complex formation between transducin and light-activated rhodopsin by covalent crosslinking: use of a photoactivatable reagent. *Proc. Natl. Acad. Sci. USA* 98, 4877–4882.
- Chen, N.F., Yu, J.Z., Skiba, N.P., Hamm, H.E., Rasenick, M.M., 2003. A specific domain of Galpha required for the transactivation of Galpha by tubulin is implicated in the organization of cellular microtubules. *J. Biol. Chem.* 278, 15285–15290.
- Chen, Z., Singer, W.D., Sternweis, P.C., Sprang, S.R., 2005. Structure of the p115RhoGEF rgRGS domain-Galpha13/i1 chimera complex suggests convergent evolution of a GTPase activator. *Nat. Struct. Mol. Biol.* 12, 191–197.
- Chen, Z., Singer, W.D., Danesh, S.M., Sternweis, P.C., Sprang, S.R., 2008. Recognition of the activated states of Galpha13 by the rgRGS domain of PD2RhoGEF. *Structure* 16, 1532–1543.
- Chen, Z., Guo, L., Hadas, J., Gutowski, S., Sprang, S.R., et al., 2012. Activation of p115-RhoGEF requires direct association of Galpha13 and the Dbl homology domain. *J. Biol. Chem.* 287, 25490–25500.
- Chien, E.Y., Liu, W., Zhao, Q., Katritch, V., Han, G.W., et al., 2010. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330, 1091–1095.
- Choe, H.W., Kim, Y.J., Park, J.H., Morizumi, T., Pai, E.F., et al., 2011. Crystal structure of metarhodopsin II. *Nature* 471, 651–655.
- Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., 2004. The Jalview Java alignment editor. *Bioinformatics* 20, 426–427.
- Dave, R.H., Saengsawang, W., Yu, J.Z., Donati, R., Rasenick, M.M., 2009. Heterotrimeric G-proteins interact directly with cytoskeletal components to modify microtubule-dependent cellular processes. *Neurosignals* 17, 100–108.
- de Vries, S.J., van Dijk, M., Bonvin, A.M., 2010. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* 5, 883–897.
- de Vries, S.J., van Dijk, A.D., Krzemiński, M., van Dijk, M., Thureau, A., et al., 2007. HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69, 726–733.
- DeLano, W.L., 2002. The Pymol Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.
- Dessauer, C.W., Tesmer, J.J., Sprang, S.R., Gilman, A.G., 1998. Identification of a Galpha binding site on type V adenylyl cyclase. *J. Biol. Chem.* 273, 25831–25839.
- Dolinsky, T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., 2004. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667.
- Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., et al., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522–525.
- Dominguez, C., Boelens, R., Bonvin, A.M.J., 2003. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737.
- Downes, G.B., Gautam, N., 1999. The G protein subunit gene families. *Genomics* 62, 544–552.
- Granier, S., Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., et al., 2012. Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485, 400–404.
- Grant, J.E., Guo, L.W., Vestling, M.M., Martemyanov, K.A., Arshavsky, V.Y., et al., 2006. The N terminus of GTP gamma S-activated transducin alpha-subunit interacts with the C terminus of the cGMP phosphodiesterase gamma-subunit. *J. Biol. Chem.* 281, 6194–6202.

- Grishina, G., Berlot, C.H., 1997. Identification of common and distinct residues involved in the interaction of α 2 and α 5 with adenylyl cyclase. *J. Biol. Chem.* 272, 20619–20626.
- Grishina, G., Berlot, C.H., 2000. A surface-exposed region of G(α) in which substitutions decrease receptor-mediated activation and increase receptor affinity. *Mol. Pharmacol.* 57, 1081–1092.
- Haga, K., Kruse, A.C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., et al., 2012. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482, 547–551.
- Hajicek, N., Kukimoto-Niino, M., Mishima-Tsumagari, C., Chow, C.R., Shirouzu, M., et al., 2011. Identification of critical residues in G(α)13 for stimulation of p115RhoGEF activity and the structure of the G(α)13–p115RhoGEF regulator of G protein signaling homology (RH) domain complex. *J. Biol. Chem.* 286, 20625–20636.
- Hamm, H.E., Deretic, D., Arendt, A., Hargrave, P.A., Koenig, B., et al., 1988. Site of G protein binding to rhodopsin mapped with synthetic peptides from the alpha subunit. *Science* 241, 832–835.
- Hanson, M.A., Roth, C.B., Jo, E., Griffith, M.T., Scott, F.L., et al., 2012. Crystal structure of a lipid G protein-coupled receptor. *Science* 335, 851–855.
- Hart, M.J., Jiang, X., Kozasa, T., Roscoe, W., Singer, W.D., et al., 1998. Direct stimulation of the guanine nucleotide exchange activity of p115 RhoGEF by Galpha13. *Science* 280, 2112–2114.
- Hewavitharana, T., Wedegaertner, P.B., 2012. Non-canonical signaling and localizations of heterotrimeric G proteins. *Cell Signal.* 24, 25–34.
- Ho, M.K., Wong, Y.H., 2000. The amino terminus of Galpha(z) is required for receptor recognition, whereas its alpha4/beta6 loop is essential for inhibition of adenylyl cyclase. *Mol. Pharmacol.* 58, 993–1000.
- Holm, L., Rosenstrom, P., 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 38, W545–549.
- Hu, J., Wang, Y., Zhang, X., Lloyd, J.R., Li, J.H., et al., 2010. Structural basis of G protein-coupled receptor-G protein interactions. *Nat. Chem. Biol.* 6, 541–548.
- Jaakola, V.P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y., et al., 2008. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322, 1211–1217.
- Johnston, C.A., Lobanova, E.S., Shavkunov, A.S., Low, J., Ramer, J.K., et al., 2006. Minimal determinants for binding activated G alpha from the structure of a G alpha(i1)-peptide dimer. *Biochemistry* 45, 11390–11400.
- Jorgensen, W.L., Tirado-Rives, J., 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110, 1657–1666.
- Kastritis, P.L., van Dijk, A.D., Bonvin, A.M., 2012. Explicit treatment of water molecules in data-driven protein–protein docking: the solvated HADDOCKing approach. *Methods Mol. Biol.* 819, 355–374.
- Kimple, A.J., Soundararajan, M., Hutsell, S.Q., Roos, A.K., Urban, D.J., et al., 2009. Structural determinants of G-protein alpha subunit selectivity by regulator of G-protein signaling 2 (RGS2). *J. Biol. Chem.* 284, 19402–19411.
- Kimple, R.J., Kimple, M.E., Betts, L., Sondek, J., Siderovski, D.P., 2002. Structural determinants for GoLoco-induced inhibition of nucleotide release by Galpha subunits. *Nature* 416, 878–881.
- Kolakowski Jr., L.F., 1994. GCRDB: a G-protein-coupled receptor database. *Recept. Channels* 2, 1–7.
- Kristiansen, K., 2004. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther.* 103, 21–80.
- Kruse, A.C., Hu, J., Pan, A.C., Arlow, D.H., Rosenbaum, D.M., et al., 2012. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552–556.
- Lambright, D.G., Noel, J.P., Hamm, H.E., Sigler, P.B., 1994. Structural determinants for activation of the alpha-subunit of a heterotrimeric G protein. *Nature* 369, 621–628.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Layden, B.T., Saengsawang, W., Donati, R.J., Yang, S., Mulhearn, D.C., et al., 2008. Structural model of a complex between the heterotrimeric G protein, Galpha, and tubulin. *Biochim. Biophys. Acta* 1783, 964–973.
- Lutz, S., Shankaranarayanan, A., Coco, C., Ridilla, M., Nance, M.R., et al., 2007. Structure of Galphaq-p63RhoGEF-RhoA complex reveals a pathway for the activation of RhoA by GPCRs. *Science* 318, 1923–1927.
- Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Mathiesen, J.M., 2012. Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* 485, 321–326.
- McCudden, C.R., Hains, M.D., Kimple, R.J., Siderovski, D.P., Willard, F.S., 2005. G-protein signaling: back to the future. *Cell. Mol. Life Sci.* 62, 551–577.
- Nance, M.R., Kreutz, B., Tesmer, V.M., Sterne-Marr, R., Kozasa, T., et al., 2013. Structural and functional analysis of the regulator of G protein signaling 2-galalphaq complex. *Structure* 21, 438–448.
- Oldham, W.M., Hamm, H.E., 2008. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat. Rev. Mol. Cell. Biol.* 9, 60–71.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., et al., 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289, 739–745.
- Park, S.H., Das, B.B., Casagrande, F., Tian, Y., Nothnagel, H.J., 2012. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491, 779–783.
- Porollo, A., Meller, J., 2007. Prediction-based fingerprints of protein–protein interactions. *Proteins* 66, 630–645.
- Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., et al., 2007. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450, 383–387.
- Rasmussen, S.G., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., et al., 2011a. Crystal structure of the beta2 adrenergic receptor–Gs protein complex. *Nature* 477, 549–555.
- Rasmussen, S.G.F., Choi, H.-J., Fung, J.J., Pardon, E., Casarosa, P., et al., 2011b. Structure of a nanobody-stabilized active state of the [br]2 adrenoceptor. *Nature* 469, 175–180.
- Rosenbaum, D.M., Rasmussen, S.G., Kobilka, B.K., 2009. The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363.
- Sammond, D.W., Eletr, Z.M., Purbeck, C., Kimple, R.J., Siderovski, D.P., et al., 2007. Structure-based protocol for identifying mutations that enhance protein–protein binding affinities. *J. Mol. Biol.* 371, 1392–1404.
- Scheerer, P., Park, J.H., Hildebrand, P.W., Kim, Y.J., Krauss, N., et al., 2008. Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455, 497–502.
- Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., et al., 2011. Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475, 65–70.
- Slep, K.C., Kercher, M.A., He, W., Cowan, C.W., Wensel, T.G., et al., 2001. Structural determinants for regulation of phosphodiesterase by a G protein at 2.0 Å. *Nature* 409, 1071–1077.
- Slep, K.C., Kercher, M.A., Wieland, T., Chen, C.K., Simon, M.I., et al., 2008. Molecular architecture of Galphao and the structural basis for RGS16-mediated deactivation. *Proc. Natl. Acad. Sci. USA* 105, 6243–6248.
- Soundararajan, M., Willard, F.S., Kimple, A.J., Turnbull, A.P., Ball, L.J., et al., 2008. Structural diversity in the RGS domain and its interaction with heterotrimeric G protein alpha-subunits. *Proc. Natl. Acad. Sci. USA* 105, 6457–6462.
- Sprang, S.R., Chen, Z., Du, X., 2007. Structural basis of effector regulation and signal termination in heterotrimeric Galpha proteins. *Adv. Protein Chem.* 74, 1–65.
- Suzuki, N., Nakamura, S., Mano, H., Kozasa, T., 2003. Galpha 12 activates Rho GTPase through tyrosine-phosphorylated leukemia-associated RhoGEF. *Proc. Natl. Acad. Sci. USA* 100, 733–738.
- Taylor, J.M., Jacob-Mosier, G.G., Lawton, R.G., Remmers, A.E., Neubig, R.R., 1994. Binding of an alpha 2 adrenergic receptor third intracellular loop peptide to G beta and the amino terminus of G alpha. *J. Biol. Chem.* 269, 27618–27624.
- Tesmer, J.J., Sunahara, R.K., Gilman, A.G., Sprang, S.R., 1997a. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Galpha.GTPgammaS. *Science* 278, 1907–1916.
- Tesmer, J.J., Berman, D.M., Gilman, A.G., Sprang, S.R., 1997b. Structure of RGS4 bound to AlF4-activated G(i alpha1): stabilization of the transition state for GTP hydrolysis. *Cell* 89, 251–261.
- Tesmer, V.M., Kawano, T., Shankaranarayanan, A., Kozasa, T., Tesmer, J.J., 2005. Snapshot of activated G proteins at the membrane: the Galphaq-GRK2-Gbetagamma complex. *Science* 310, 1686–1690.
- Thompson, A.A., Liu, W., Chun, E., Katritch, V., Wu, H., et al., 2012. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399.
- UniProt Consortium, 2012. Reorganizing the protein space at the universal protein resource (UniProt). *Nucleic Acids Res.* 40, D71–D75.
- Unni, S., Huang, Y., Hanson, R.M., Tobias, M., Krishnan, S., et al., 2011. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* 32, 1488–1491.
- van Dijk, A.D., Bonvin, A.M., 2006. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* 22, 2340–2347.
- Waldo, G.L., Ricks, T.K., Hicks, S.N., Cheever, M.L., Kawano, T., et al., 2010. Kinetic scaffolding mediated by a phospholipase C-beta and Gq signaling complex. *Science* 330, 974–980.
- Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., et al., 1995. The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 83, 1047–1058.
- Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., et al., 2008. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454, 486–491.
- Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191.
- Woehler, A., Ponimaskin, E.G., 2009. G protein-mediated signaling: same receptor, multiple effectors. *Curr. Mol. Pharmacol.* 2, 237–248.
- Wu, B., Chien, E.Y., Mol, C.D., Fenalti, G., Liu, W., et al., 2010. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071.
- Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G.W., 2012. Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 485, 327–332.
- Yu, M.Y., Ho, M.K., Liu, A.M., Wong, Y.H., 2008. Mutations on the Switch III region and the alpha3 helix of Galpha16 differentially affect receptor coupling and regulation of downstream effectors. *J. Mol. Signal.* 3, 17.