

# Homology modeling and docking

---

Επιβλέπων καθηγητής: Γλυκός Νικόλαος  
Φοιτητής: Μπαλλχύσα Εουγκέν

25/11/2018

## Εισαγωγή

Σύμφωνα με τις οδηγίες που παραδόθηκαν, το πρώτο βήμα ήταν η λήψη και εξοικείωση στην χρήση του προγράμματος Modeller. Ύστερα από την επιτυχή λήψη του λογισμικού και την ανάγνωση των βοηθημάτων χρήσης του, επιχειρήθηκε η απομίμηση των βημάτων που ακολουθούνται στα παραδείγματα σε μία πρωτεΐνη, της οποίας η δομή απουσιάζει από την βάση δεδομένων της Protein Data Bank. Η πρωτεΐνη που επιλέχθηκε είναι η Rho-σχετιζόμενη GTP-binding protein «RhoG» του οργανισμού *Mus musculus*. Η πρωτεΐνη αποτελείται από 191 αμινοξέα και έχει μέγιστη ομοιότητα 73% με πρωτεΐνες των οποίων η δομή είναι γνωστή και υπάρχει στην βάση PDB. Σύμφωνα με την βάση UniProt, υπάρχουν τρία σημεία πρόσδεσης GTP (10-17, 57-61 και 115-118). Έτσι η πρωτεΐνη αυτή πληροί τα τρία κριτήρια που χρειάζονται για να πραγματοποιηθεί homology modeling για την δημιουργία μοντέλων της δομής της, καθώς και docking για την μοντελοποίηση της σύνδεσης με GTP και των δυνάμεων που σταθεροποιούν την σύνδεση αυτή.

## Μέθοδοι

### Homology modeling

Η πρωτεϊνική αλληλουχία της RhoG λήφθηκε από το NCBI και με την χρήση BLAST για πρωτεΐνες παρόμοιας αλληλουχίας βρέθηκε εάν υπάρχουν ομόλογες στην βάση PDB. Ύστερα, δημιουργήθηκε το PIR αρχείο με την αλληλουχία του:

```
>P1;test  
sequence:test:::0.00: 0.00  
MQSIKCVVVGDAVGKTCLLICYTTNAFPKEYIPTVFDNYSQAQSAVDGRTVNLNLWDTAGQEEYDRLRTL  
SYPQTNVFCFSIASPPSYENVVRHKWHPEVCHHCPDVPILLVGTGKDLRAQPDTLRRRLKEQGQAPITPQ  
QGQALAKQIHAVRYLECSALQQDGVKEVFAEAVRAVLNPTPIKRGRSCILL *
```

Μετά την λήψη του προγράμματος MODELLER και αφού τροποποιήθηκαν ώστε να αναζητούν πληροφορίες για την RhoG, χρησιμοποιήθηκαν και τα 5 scripts που υπάρχουν στο βασικό παράδειγμα της ιστοσελίδας του MODELLER. Η δομή του μοντέλου που προέκυψε φαίνεται στην Εικόνα 1. Λόγω της ύπαρξης μία αφύσικης «ουράς» στο C-αμινοτελικό άκρο, επιχειρήθηκε η μοντελοποίηση της δομής της με την βοήθεια του on-line προγράμματος “Swiss-Model”. Παρατηρήθηκε η ουρά αυτή αποτελείται από τα τελευταία 14 κατάλοιπα. Σύμφωνα με την υπόθεση πως υπάρχουν μεταμεταφραστικές τροποποιήσεις που πιθανώς αφαιρούν αυτήν την ουρά που προκαλεί την ασταθή δομή του μορίου, αφαιρέθηκαν αυτά τα κατάλοιπα από το PIR αρχείο της πρωτεΐνης και ξανα-χρησιμοποιήθηκαν τα scripts του MODELLER από την αρχή (Εικόνα 2). Πραγματοποιώντας

μία εκτίμηση του μοντέλου αυτού, υπολογίσθηκαν και συγκρίθηκαν οι ενέργειες ανά κατάλοιπο σε τιμές Discrete optimized protein energy (DOPE) μεταξύ της Rhog και του template (Διάγραμμα 1).

Ακολούθησε περαιτέρω βελτιστοποίηση του μοντέλου χρησιμοποιώντας πολλαπλά templates από την οικογένεια fm02480 της 1mh1 (1G4UR, 3BWDD, 5A0FA). Με την χρήση των Scripts των παραδειγμάτων της ιστοσελίδας, δημιουργήθηκαν 5 νέα μοντέλα και το καλύτερο επιλέχθηκε ως αυτό με την χαμηλότερη τιμή molpdf (5455.49854). Ύστερα προστέθηκε το νέο αυτό μοντέλο στο διάγραμμα σύγκρισης της ενέργειας DOPE ανά κατάλοιπο (Διάγραμμα 2).

Παρατηρώντας το διάγραμμα, βλέπουμε μία μεγάλη ασυμφωνία των γραφημάτων στο σημείο όπου υπάρχει ένα loop στην πρωτεΐνη (κατάλοιπα 28-35). Στο σημείο αυτό εφαρμόστηκε η τεχνική loop refining και δημιουργήθηκαν 10 νέα μοντέλα, των οποίων τα DOPE scores υπολογίσθηκαν και συγκρίθηκαν (Πίνακας 1). Παρατηρούμε από το διάγραμμα 3, πως ύστερα από το loop refinement, το DOPE score στην περιοχή του loop έπεσε σημαντικά. Ως αποτέλεσμα, αυτό το νέο μοντέλο θα θεωρηθεί το βέλτιστο έως τώρα και θα προχωρήσουμε στο docking με αυτό (Εικόνα 3). Εδώ πρέπει να σημειωθεί πως μία πολύ ακριβής προσέγγιση με loop refinement απαιτεί την μοντελοποίηση εκατοντάδων ανεξάρτητων δομών και το clustering τους για να επιλεγθεί το πιο αντιπροσωπευτικό loop.

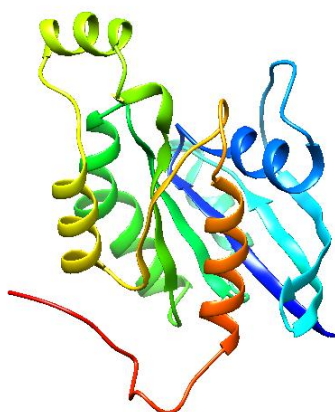
## Docking

Σύμφωνα με την βάση UniProt, η πρωτεΐνη που μελετάται περιέχει θέση πρόσδεσης για GTP. Επομένως η τρισδιάστατη δομή της GTP λήφθηκε από την βάση PDB. Επόμενο βήμα ήταν η εύρεση των περιοχών στις οποίες το Autodock Vina θα προσπαθήσει να βρει το σημείο πρόσδεσης (Εικόνα 4). Για να βρεθεί η περιοχή αυτή, αναζητήθηκε στο Swiss-Model μία πρωτεΐνη με παρόμοια αλληλουχία με την υπό μελέτη. Έτσι, βρέθηκε ένα RhoG domain της πρωτεΐνης RacC του οργανισμού *Entamoeba histolytica* στο μοντέλο του οποίου ήταν προσδεσμένη η GTP. Επόμενο βήμα ήταν η δημιουργία του pdbqt αρχείου του ligand, στον οποίο έχουμε καθορίσει τις μοριακές συνδέσεις στις οποίες επιτρέπουμε ή όχι την περιστροφή (Εικόνα 5). Σύμφωνα με την εικόνα 5, υπάρχουν 15/32 δεσμούς οι οποίοι μπορούν να περιστραφούν, κάτι που προσδίδει μεγάλη ελευθερία στην εύρεση διάφορων μοντέλων από το AutoDock Vina. Τέλος, δημιουργήθηκε το configuration αρχείο που απαιτεί το AutoDock Vina για να τρέξει. Αφού έτρεξε το πρόγραμμα, λήφθηκαν 9 διαφορετικά μοντέλα της σύνδεσης (Εικόνα 6). Αφού έτρεξε το AutoDockTools και λήφθηκαν τα μοντέλα, πραγματοποιήθηκε στο τέλος μία σύγκριση της δομής της GTP κατά την πρόσδεση στο μοντέλο που δημιουργήθηκε και στο μοντέλο που υπάρχει στην PDB (Εικόνα 7).

## Αποτελέσματα

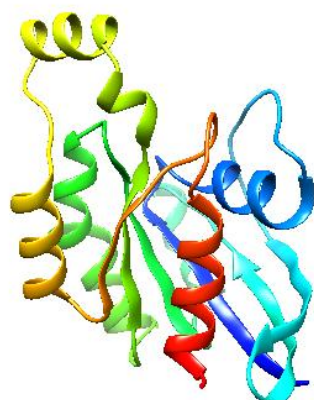
### Homology modeling

Από το πρώτο script2 προέκυψε πως η 1mh1\_A είναι η βέλτιστη πρωτεΐνη για το homology modeling, καθώς έχει την υψηλότερη ομοιότητα (72%), την χαμηλότερη e-value (0), και την καλύτερη ανάλυση (1.38 Å). Μετά το alignment της πρωτεΐνης υπό εξέταση και του template (1mh1) προέκυψαν 5 μοντέλα, από τα οποία το καλύτερο κρίθηκε ως αυτό με την χαμηλότερη τιμή DOPE (-21573.29297) και το υψηλότερο GA341 score (1) (Εικόνα 1).

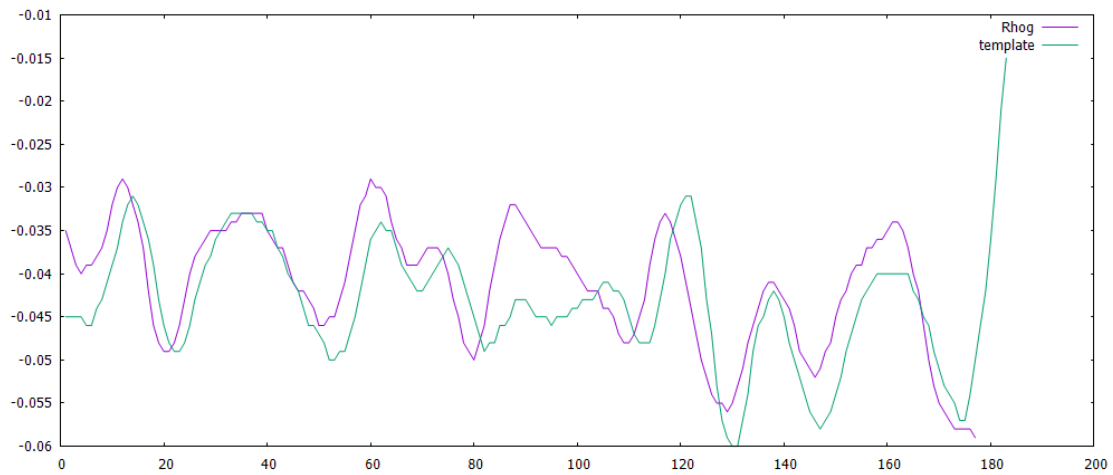


**Εικόνα 1.** Δομή της πρωτεΐνης RhoG όπως προέκυψε από τα 5 αρχικά scripts που χρησιμοποιήθηκαν στο MODELLER. Η εικόνα εμφανίζεται από το πρόγραμμα Chimera.

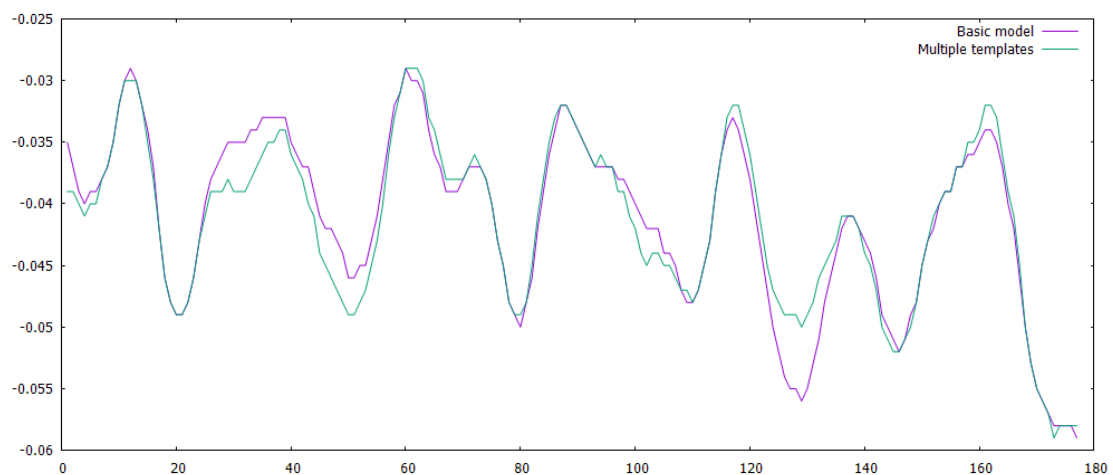
Παρόμοια, μετά την αφαίρεση των 14 καταλοίπων, προέκυψε το εξής καλύτερο μοντέλο της εικόνας 2, με DOPE (-20894.24805) και GA341 score (1).



**Εικόνα 2.** Δομή της πρωτεΐνης RhoG όπως προέκυψε από τα 5 αρχικά scripts που χρησιμοποιήθηκαν στο MODELLER ύστερα από την αφαίρεση των 14 καταλοίπων του C-αμινοτελικού άκρου. Η εικόνα εμφανίζεται από το πρόγραμμα Chimera.



**Διάγραμμα 1.** Απεικόνιση του DOPE score ανά κατάλοιπο (άξονας γ) προς τα σημεία του alignment μεταξύ των πρωτεϊνών (άξονας χ) ύστερα από μοντελοποίηση με ένα template. Η εικόνα λήφθηκε από το πρόγραμμα GNUPLOT.

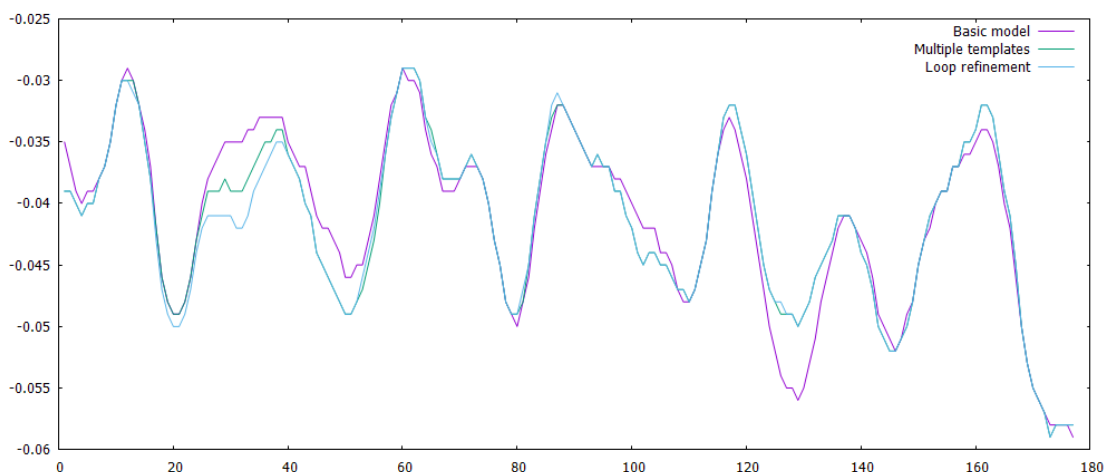


**Διάγραμμα 2.** Απεικόνιση του DOPE score ανά κατάλοιπο (άξονας γ) προς τα σημεία του alignment μεταξύ των πρωτεϊνών (άξονας χ) ύστερα από μοντελοποίηση με πολλαπλά templates.

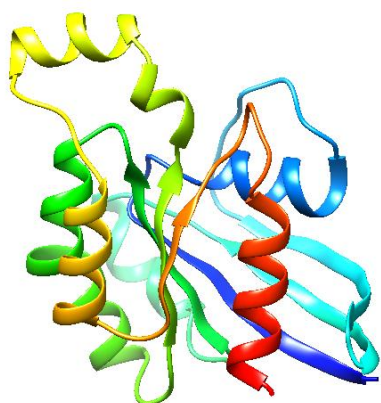
Είναι ενδιαφέρον το γεγονός πως στις περιοχές όπου, σύμφωνα με το UniProt, προσδένονται GTP (10-17, 57-61 και 115-118), έχουμε σχεδόν απόλυτη ταύτιση των ενεργειών στα δύο μοντέλα.

**Πίνακας 1.** Τιμές DOPE των 10 μοντέλων που προέκυψαν από το loop refining. Φαίνεται πως το βέλτιστο μοντέλο είναι το 9.

Αριθμός μοντέλου	DOPE score
1	-20421.304688
2	-20514.527344
3	-20074.203125
4	-20383.332031
5	-20389.105469
6	-20229.382813
7	-20294.496094
8	-20534.859375
9	-20624.242188
10	-20409.332031



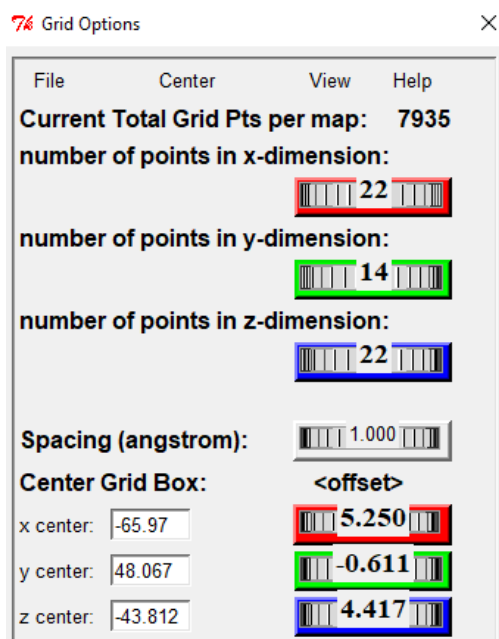
**Διάγραμμα 3.** Απεικόνιση του DOPE score ανά κατάλοιπο (άξονας γ) προς τα σημεία του alignment μεταξύ των πρωτεϊνών (άξονας χ) ύστερα από μοντελοποίηση με πολλαπλά templates και loop refinement, σε σύγκριση με το αρχικό μοντέλο.



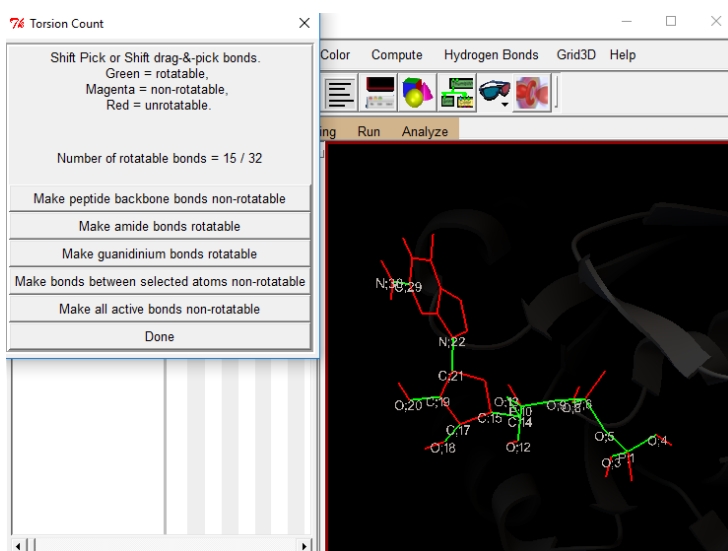
**Εικόνα 3.** Τελικό μοντέλο της δομής της πρωτεΐνης RhoG ύστερα από homology modeling και loop refinement με την χρήση MODELLER. Η εικόνα εμφανίζεται από το πρόγραμμα Chimera.

## Docking

Ακολουθώντας όλες τις συμβουλές του εκπαιδευτικού βίντεο της ιστοσελίδας του AutoDock Vina, λήφθηκαν τα παρακάτω αποτελέσματα.



**Εικόνα 4.** Περιοχή στην οποία θα αναζητηθεί, από το AutoDock Vina, η θέση πρόσδεσης της GFP. Η περιοχή αυτή καθορίστηκε με την βοήθεια της 4mit πρωτεΐνης της βάσης BDP. Το πρόγραμμα που χρησιμοποιήθηκε για τον καθορισμό των συντεταγμένων είναι το AutoDockTools.



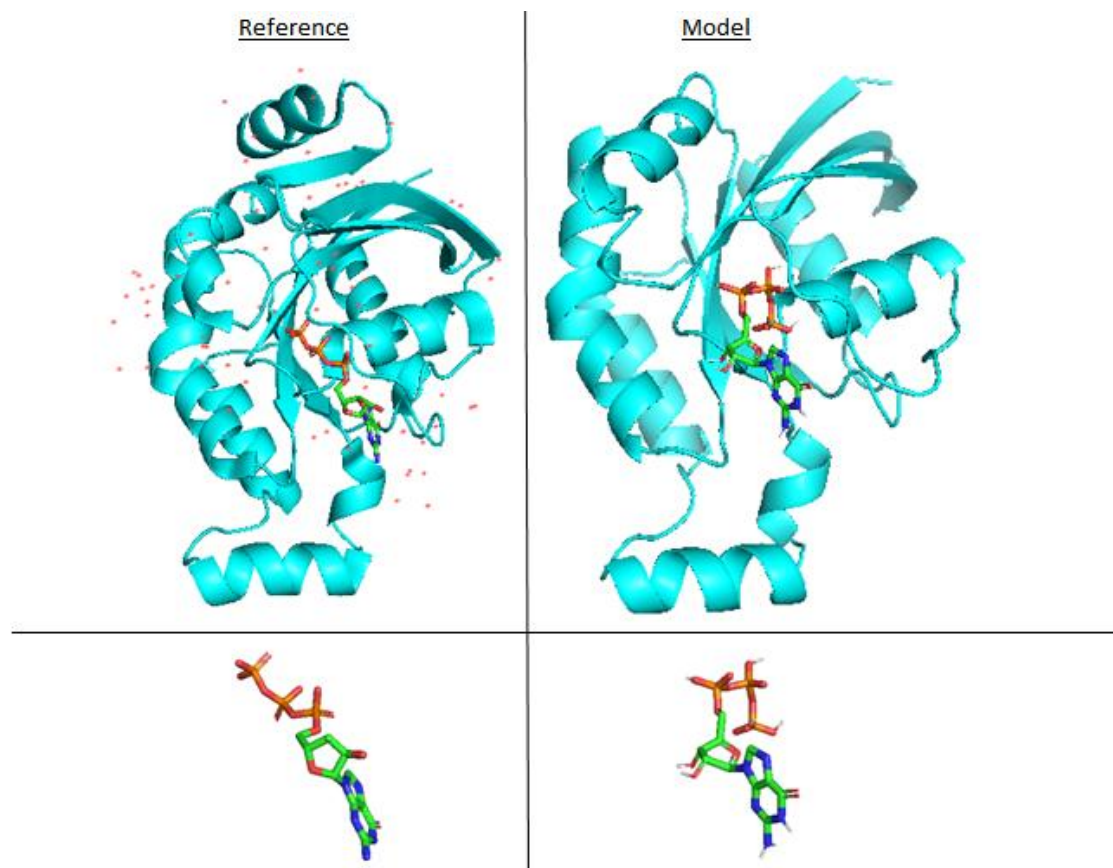
**Εικόνα 5.** Δομή του ligand (GTP) στην οποία οι ενώσεις στις οποίες επιτρέπεται περιστροφή είναι χρωματισμένες πράσινες και σε αυτές που δεν επιτρέπεται, κόκκινες. Ο προσδιορισμός αυτός πραγματοποιήθηκε στο AutoDockTools.

mode	affinity (kcal/mol)	dist from best mode rmsd l.b.   rmsd u.b.
1	-7.4	0.000   0.000
2	-7.2	1.729   2.563
3	-7.2	3.299   5.442
4	-6.5	3.254   5.275
5	-6.5	1.620   2.747
6	-6.3	1.805   2.643
7	-6.0	5.041   6.592
8	-5.9	4.534   7.984
9	-5.8	4.433   8.499

Writing output ... done.

**Εικόνα 6.** Δεδομένα των 9 μοντέλων που δημιουργήθηκαν από το AutoDock Vina. Ως “best mode” προέκυψε το μοντέλο με affinity -7,4.

Το μοντέλο με την χαμηλότερη ενέργεια ανά mol αντιδρόντων, θεωρείται ως αυτό που αντικατοπτρίζει την πιο σταθερή σύνδεση μεταξύ της πρωτεΐνης και του ligand (μεγαλύτερο affinity). Το μοντέλο με την σταθερότερη σύνδεση συγκρίθηκε με το μοντέλο γνωστής δομής σύνδεσης (reference).



**Εικόνα 7.** Σύγκριση των μοντέλων πρόσδεσης της πρωτεΐνης RhoG με την GTP. Ως reference λειτουργεί η γνωστή δομή της επικράτειας RhoG της πρωτεΐνης RacC, κατά την πρόσδεση της GTP σε αυτήν.

## Συμπεράσματα

Παρατηρούμε πως οι δύο διαμορφώσεις της GTP κατά την πρόσδεση είναι πολύ διαφορετικές μεταξύ τους. Αυτό μπορεί να ευθύνεται στο ότι παρόλο που πρόκειται για μία παρόμοια επικράτεια, συγκρίνουμε δύο διαφορετικές εν τέλει πρωτεΐνες οι οποίες προέρχονται η μία από ένα ζώο και η άλλη από μία αμοιβάδα· δύο οργανισμοί εξελικτικά απόμακροι. Όμως, ακόμα και έτσι, χρειαζόμαστε περισσότερα δεδομένα για να μπορέσουμε να αποδεχθούμε τα μοντέλο του docking. Πρέπει να συγκριθούν παρόμοιες επικράτειες από πολλούς διαφορετικούς οργανισμούς και να τρέξουμε το modeling με περισσότερο exhaustiveness (εδώ ήταν ίσο με 10). Παρόλο που αυτό απαιτεί αρκετά μεγάλη υπολογιστική δύναμη, χρειάζεται για ένα καλύτερο μοντέλο. Επομένως είναι προφανές πως, για να λάβουμε μία ασφαλή πρόβλεψη της δομής μίας πρωτεΐνης ή της σύνδεσής της με έναν ligand, απαιτείται εξαντλητική προσπάθεια και από τον ερευνητή και από τον εξοπλισμό του.