Molecular modeling and molecular dynamics simulation study of the human Rab9 and RhoBTB3 C-terminus complex

Muhammad Junaid[1], Ziyad Tariq Muhseen[1], Ata Ullah[3], Abdul Wadood[3], Junjun Liu[2] & Houjin Zhang[1]*

[1]Department of Biotechnology, Huazhong University of Science and Technology, China; [2]Tongji School of Pharmacy, Huazhong University of Science and Technology, China; [3]Department of Biochemistry, Abdul Wali Khan University, Mardan, Pakistan; Houjin Zhang - Email: hjzhang@hust.edu.cn; *Corresponding author

Abstract:
Rab9 is required for the transport of mannose 6-phosphate receptors to the trans-Golgi network from late endosomes through the interaction with its effector: RhoBTB3. Earlier research indicates the C-terminus of RhoBTB3 (Rho_Cterm) is used for the interaction with Rab9. We used the homology modeling along with the molecular dynamics (MD) simulation to study the binding pattern of Rho_Cterm and Rab9 at atomic level. Both modeled structures, Rab9 and Rho_Cterm, are of high quality as suggested by the Ramachandran plot and ProCheck. The complex of Rab9-Rho_Cterm was generated by unrestrained pairwise docking using ZDOCK server. The interface of complex is consistent with the previous experimental data. The results of MD simulation indicate that the binding interface is stable along the simulation process.

Background:

Rab proteins, the largest subfamily of small GTPases, consist of more than 70 members [1,2]. Rab proteins function with their effectors in distinct organelle membranes and regulate the docking, tethering, motility and fusion of intracellular membranes in the eukaryotic organisms [3]. One member of this family, Rab9, cycles between active (GTP-bound) and inactive (GDP-bound) conformations [4-6]. It is localized in the late endosome membranes and required to transport the mannose 6-phosphate receptors to the trans-Golgi network from late endosomes [7-9]. Moreover, Rab9 is also involved in the late endosome morphology and lysosome biogenesis [9, 10]. Rab9 can also act as a cellular target for some pathogens such as HIV and *Salmonella* [11, 12]. Some effectors of Rab9 have been discovered that include RhoBTB3, Tip47, GCC185, all of which are involved in recycling the mannose 6-phosphate receptors [13-17]. RhoBTB3 and GCC186 are localized to the trans-Golgi network. They facilitate the fusion of the Rab9-bound vesicles

and the TGN [15,17]. The two conserved effectors of Rab9 in eukaryotic organisms are RUTBC1 and RUTBC2 [18, 19].

Rho GTPases, the key regulator of actin cytoskeleton, regulate the most aspects of cell dynamics, including morphogenesis migration and division [20, 21]. Some members are classified as atypical GTPases because their sequences are not like classical GTPases. RhoBTB3, an atypical Rho GTPase, is much larger than other Rho GTPases. It consists of a Rho GTPase-related domain near its N-terminus and a BTB domain near its C-terminus [15]. These Rho-GTPases have an *active* GTP-bound and an *inactive* GDP-bound state. The GTP-GDP binding cycle is tightly regulated. The inactive form is activated at the cell membrane by GDP–GTP exchange with the help of a large family (~80 proteins) of guanine nucleotide exchange factors [22]. The BTB domain is found to have a key role in protein-protein interaction that participate in many functions such as cytoskeleton regulation [23, 24], transcription repression [25, 26] and protein ubiquitination/degradation [27].

BIOINFORMATION
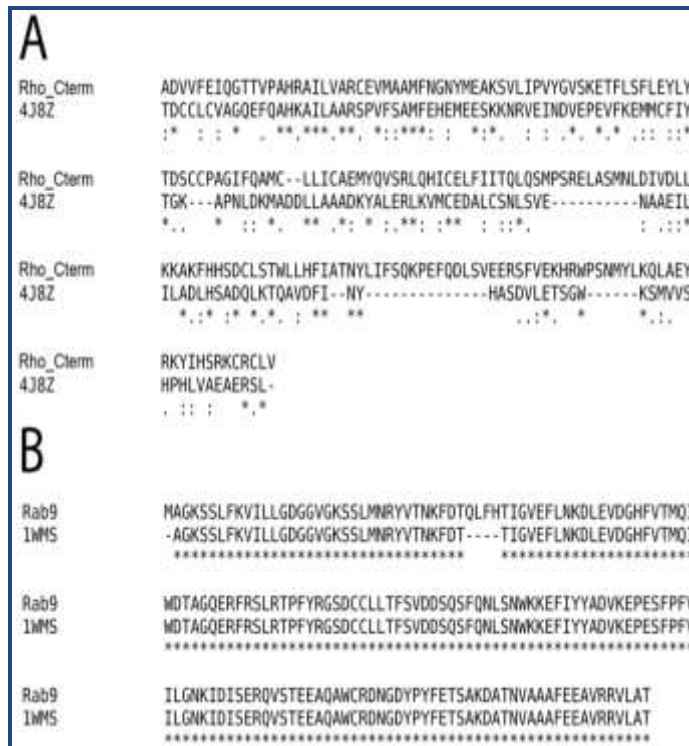
Figure 1: A) Partial Alignment of Rho_Cterm with its template: Human SPOP BTB Domain (PDB ID 4J8Z); **B)** Partial Alignment of Rab9 with its template: Human Rab9GTPase mutant (PDB ID 1WMS).
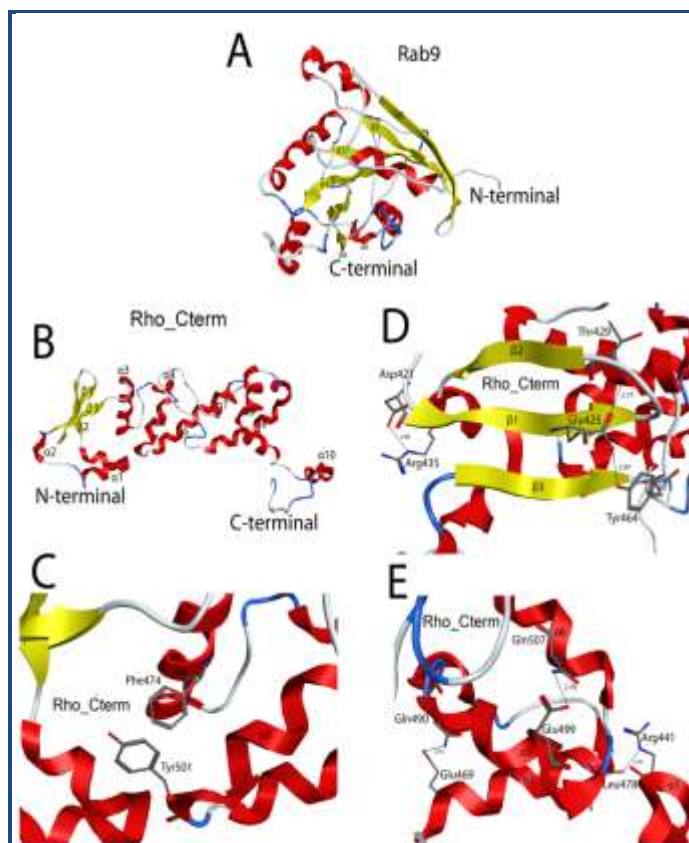
Figure 2: **A)** The modeled structure of Rab9; **B)** The modeled structure of Rho_Cterm; **C)** The 3D representation of aromatic interactions between helices α3 and α5 of Rho_Cterm; **D)** The 3D representation of hydrogen bonding network between

Asp421 and Arg435, Thr429 and Glu425, Glu425 and Tyr464 of Rho_Cterm; **E)** The 3D representation of hydrogen bonding network between Gln490 and Glu469, Gln507 and Glu499, Arg441 and Leu478 of Rho_Cterm.

In the present study we report the prediction of 3D structures of human Rab9 GTPase and RhoBTB3 C-terminus (designated as Rho_Cterm thereafter) based on the crystal structures of human Rab9 GTPase mutant and human SPOP BTB domain respectively. The modeled structures of Rab9 and Rho_Cterm are of high quality as suggested by the Ramachandran plot. The ZDOCK online server **[28]** was used to predict the complex structure. The binding interface of the resultant complex matches the experimental data. And the MD simulation performed with the complex shows that the complex is stable.

Methodology:
Homology modeling
The Rab9 GTPase and RhoBTB3 with accession No. P51151 and O94955 were used as target sequences respectively. The crystal structure of human Rab9 GTPase fragment (residues 1-177) is available at Protein Data Bank (PDB). So, we have made an attempt to model the complete sequence (residues 1-201) of human Rab9 GTPase. To find suitable templates for the modeling of these proteins, BLASTP server and the other server (http://toolkit.tuebingen.mpg.de/psi_blast/) were used against PDB using default parameters **[29, 30].** Both servers give the same results .The high-resolution crystal structure of human Rab9 GTPase (PDB ID: 1WMS) and crystal structure of the human SPOP BTB Domain (PDB ID: 4J8Z) were selected as templates for Rab9 GTPase and Rho_Cterm (residues 420-607) respectively. The models were built using the Molecular Operating Environment (MOE). A series of 10 independent models for each target protein were built using the Boltzmann-weighted randomized procedure **[31]** combined with specialized logic for the handling of sequence insertions and deletions **[32]**. Out of 10 models, the models with best MOE packing scores were selected for each target. Both models were superimposed over its templates using protein superpose module implemented in MOE. The structural evaluation and stereo chemical analysis were done using Ramachandran plot, ProSA **[33]** and Procheck **[34].**

Protein-protein docking
For protein-protein docking, we used unrestrained pair wise rigid body docking for Rab9 GTPase and Rho_Cterm. The coordinates of Rab9 and Rho_Cterm models were used for energy minimization prior to docking procedure. The energy was minimized using AMBER99 force field implemented in MOE software. ZDOCK server **[28]** was utilized for protein-protein docking to predict and evaluate the interactions in Rab9-Rho_Cterm complex. Docking was carried out without specifying the binding residues so that the docking results will reflect the most possible interaction patterns without any arbitrary restrain.

This ZDOCK server ranks the 100 most probable predictions on the basis of electrostatic complementarity, hydrophobicity and geometry of the molecular surface out of thousands of candidates. By manual analysis of the complexes, the important residues of RhoBTB3 mentioned by Espinosa EJ *et al.* **[15],** were

found in the complex that is ranked second on the basis of docking Z-Score.
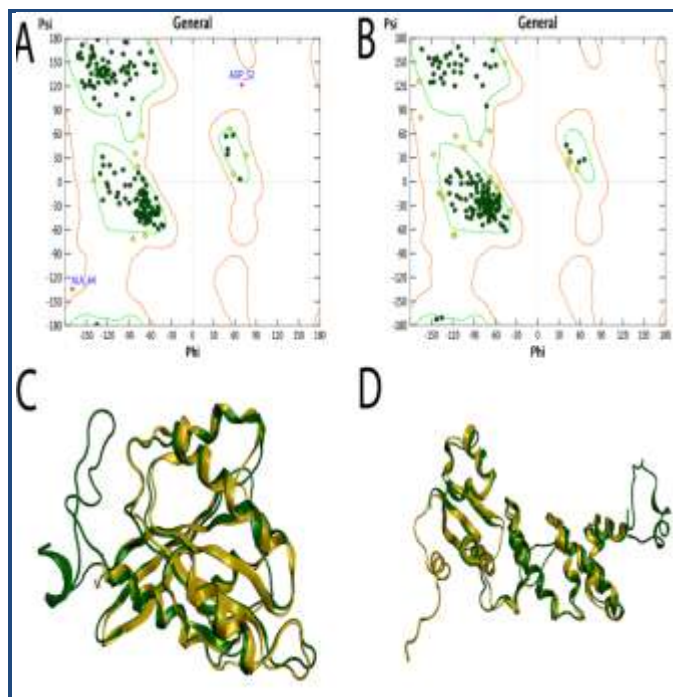
Figure 3: A) Ramachandran Plot of Rab9 after MD simulation. The residues shown in blue color are in outlier region; **B)** Ramachandran Plot of Rho_Cterm. After MD simulation there is no residue in outlier region; **C)** The superposition of Rab9 homology model over its template 1WMS, The Rab9 is shown in green while the template is shown in golden color; **D)** The superposition of Rho_Cterm homology model over its template 4J8Z, The Rho_Cterm is shown in green while the template is shown in golden color.
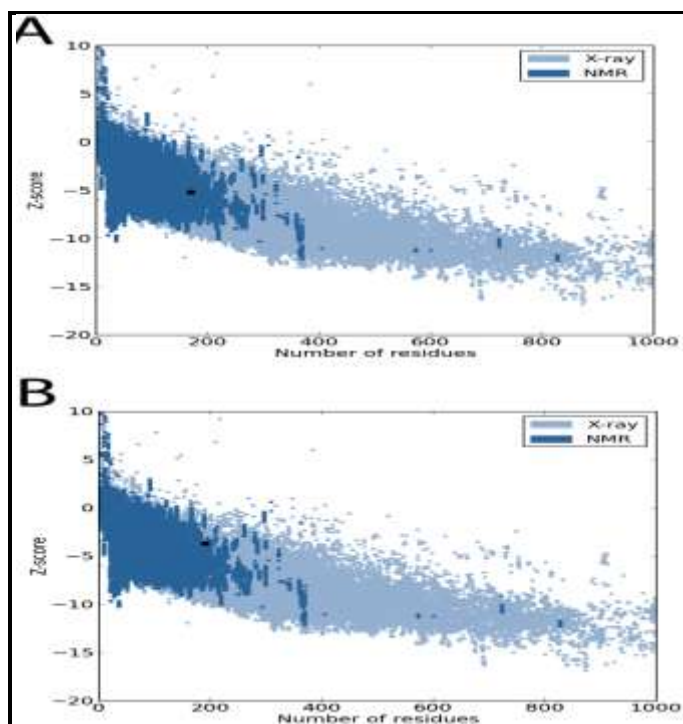
Figure 4: **A)** Protein model quality scores of Rab9 and **B)** Rho_Cterm; The Z-scores of Rab9 and Rho_Cterm are

represented in the plot by large black dots. The Z-scores of all proteins in PDB, determined by X-ray crystallography, are shown by light blue color. The Z-scores of all proteins in PDB, determined by NMR spectroscopy, are shown by dark blue color.
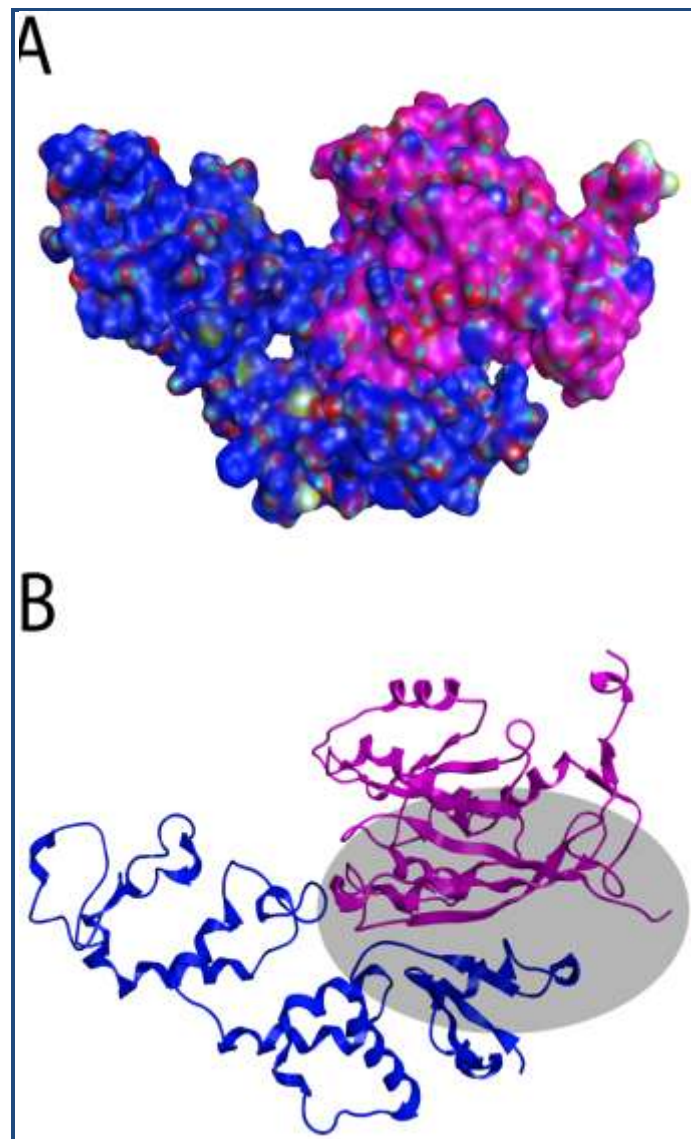

Figure 5: A) The 3D representation of shape complementarity of Rab9-Rho_Cterm complex; **B)** The overall 3D representation of Rab9-Rho_Cterm complex. The shaded portion shows the binding interface area. The Rho Cterm is colored in blue and the Rab9 is colored in magenta.

Molecular Dynamics Simulation
Molecular dynamics simulation was performed utilizing AMBER14 software package. The atomic coordinates of the Rho_Cterm-Rab9 complex were obtained from the ZDOCK online server. The system was made neutral by adding counter-ions. The resultant system was then solvated in a rectangular box of TIP3P water with a buffer of 8 Å **[35].** Long-range electrostatic interactions were computed by employing Particle Mesh Ewald (PME) with the default setting in AMBER14. Force field ff14SB was used for the MD simulation. The system was energy-minimized by steepest descent and conjugate gradient methods. Subsequently, the system was heated from 0 to 300 K

for 100 ps and then equilibrated for 400 ps. After equilibration, a MD simulation was executed for 40ns.
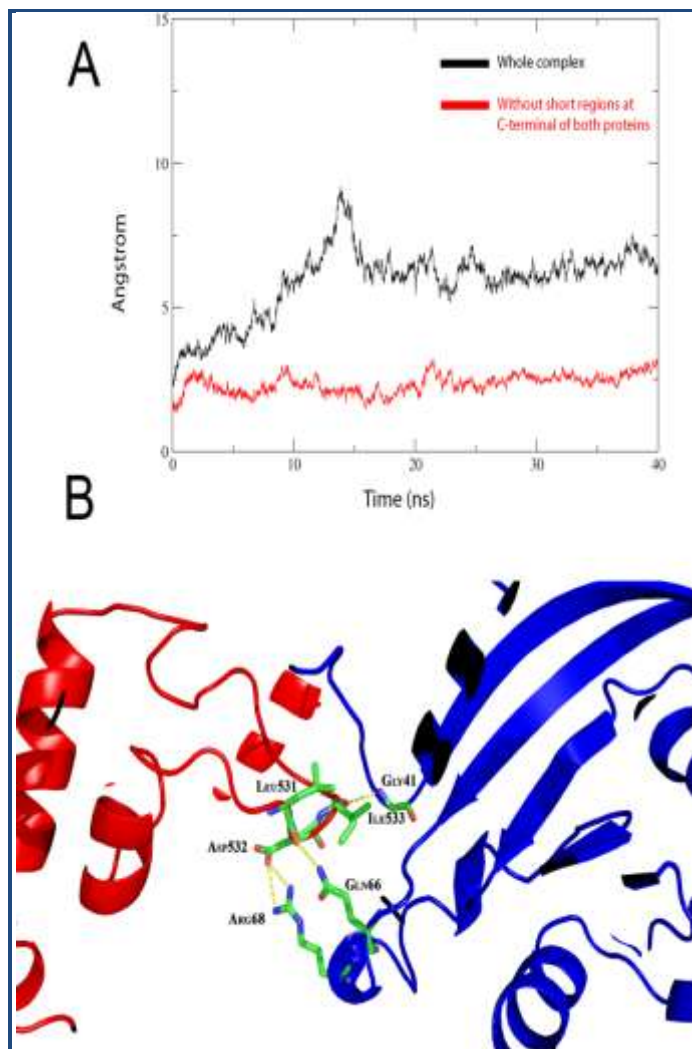
Figure 6: A) RMSD versus time plot of Rab9-Rho_Cterm complex during 40ns of simulation. The black line represents the RMSD of whole complex. The RMSD converges at about 6 Å after 16 ns and it remains stable afterwards. The red line represents the RMSD of truncated complex (without short regions at C-terminal of both proteins). The RMSD converges at about 2.5 Å after 7 ns; **B)** The 3D representation of hydrogen bonding network between Rab9 and Rho_Cterm after MD simulation of 40ns. The Rho_Cterm is colored in red and Rab9 is colored in blue.

Results & Discussion:
Homology modeling
In order to build the 3D models for Rab9 GTPase and Rho_Cterm, the BLASTP searches were carried out against the PDB. We identified the entry 1WMS with 95% identity to Rab9 GTPase and entry 4J8Z with 33% identity to Rho_Cterm (Residues 420-607), which were used as templates for the modeling. The target and template protein sequences were aligned as shown in **Figure 1**. The modeled structure of Rab9 consists of six β-sheets surrounded by six α-helices. Rab9 has a classical nucleotide-binding fold that is present in all Rab GTPase family members **(Figure 2A)**. The modeled structure of Rho_Cterm consists of three β-sheets and ten α-helices **(Figure**

2B). The aromatic interactions and hydrogen bonds between α-helices and β-sheets of Rho_Cterm model contribute toward the stability of the structure. The aromatic interaction in Rho_Cterm is found between the Phe474 of α3 and Tyr501 of α5 **(Figure 2C).** The β1 and β2 sheets of Rho_Cterm are antiparallel to each other. This arrangement is supported on one side by hydrogen bonds between Asp421 of β1 and Arg435 of α1 and on the opposite site by the interaction of Thr429, Glu425 and Tyr464 through hydrogen bonding network **(Figure 2D).** Along with this, all other α-helices are connected to each other through a hydrogen-bonding network. The hydrogen bond between the Gln507 and Glu499 supports the parallel arrangement of α5 and α6. Moreover, the α3 is held perpendicular to the α1 by the hydrogen bond between Arg441 of α1 and Leu478 of α3 **(Figure 2E)**. The residues Ala498, Asp532 and Ile533, which are proved to be important for the interaction with Rab9, are present in the α5 and in the loop between α7/ α8 respectively **[15]**.
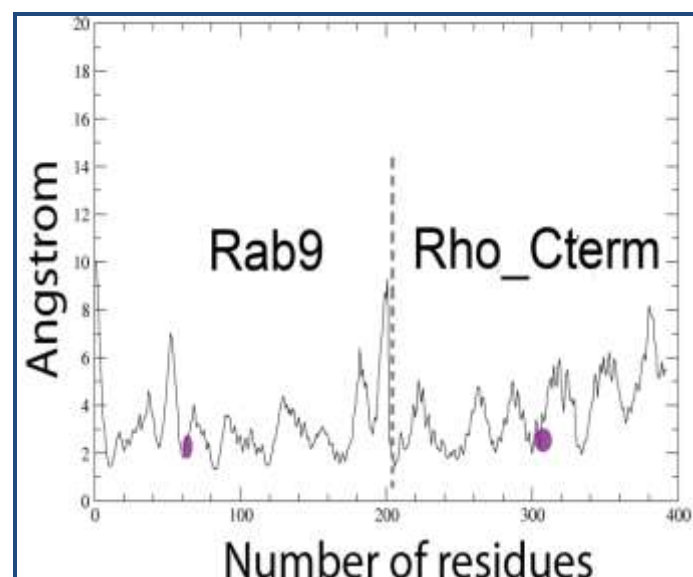
Figure 7: RMSF graph of Rab9-Rho_Cterm complex. The shaded areas show the residues involved in binding. The residues at the left of dotted line are for Rab9 and residues at the right of dotted line are for Rho_Cterm.

The quality of modeled structures was evaluated through Ramachandran plot. The evaluation of backbone Psi and Phi dihedral angles for Rab9 model revealed that 85.9 % residues lie in favored regions, 11 % residues lie in allowed regions and only 3 % residues lie in the outlier regions. In addition, the Ramachandran plot tools implemented in MOE show that Ala2, Gln35 and Leu36 fall in the outlier regions but after MD simulation there are only two residues Asp52 and Ala64 in the outlier regions **(Figure 3A).** The stereo chemical evaluation of backbone Psi and Phi dihedral angles for Rho_Cterm model revealed that 82.6 % residues lie in favored regions, 12.6 % residues lie in allowed regions and only 4.7 % residues lie in outlier regions. The residues in the outlier region are Ser520, Asp535, Ser545 and Met589 but after the MD simulation no residues were found in the outlier regions **(Figure 3B).** The structural superposition of alpha carbon (Cα) of Rab9 and Rho_Cterm models over their templates has root mean square deviation (RMSD) of 0.94 Å and 1.008 Å respectively **(Figure 3C & D).** The qualities of the models were further evaluated by

BIOINFORMATION

ProSA that checks the potential errors in the models. This server calculates the Z-score of the input structure that is the measure of quality of the model and the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The value of the Z-score was plotted with the Z-scores of all similar size proteins determined experimentally through NMR and X-ray. Analysis of the Rab9 and Rho_Cterm with ProSA shows a Z-score of - 5.21 and -3.73 respectively, indicating no significant deviation from the scores determined for the proteins of similar size **(Figure 4)**. To confirm the above results, the model was also evaluated by Procheck. The results of Procheck are given in **Table 1 (see in supplementary material).**

Protein-Protein Docking and MD simulation
Protein-protein docking is an important method to understand the structural information on protein-protein interactions **[36-38]**. The structural interface between RhoBTB3 and Rab9 has been described previously through experimental procedure **[15]** but their atomic level interactions have not been available yet. The unavailability of the complex structure is a hindrance to understand that how Rab9 binds with RhoBTB3 to facilitate the protein transport from endosomes to the Trans Golgi network.

The docking of Rab9 and Rho_Cterm was performed using ZDOCK program. The **Figure 5A** shows the shape complementarity of Rab9 and Rho_Cterm. The interface of the complex is formed by the α-helices and β-sheets from both partners **(Figure 5B)**. To study the stability of this complex, the molecular dynamics (MD) simulation of the complex was performed with the program Amber14. The complex energy was minimized; the model was solvated and ionized with the addition of Na+. The MD was performed at 300K for 40ns. The stability of the complex was monitored during MD simulation using RMSD with respect to the initial structure. The RMSD converges around 6 Å after 16ns simulation and has shown no more significant fluctuation afterwards, revealing that the complex of Rho_Cterm and Rab9 is stable. Most of the RMSD is contributed by the residues 175-201 of Rab9 and the residues 584-607 of RhoBTB3, which are not the binding regions. Without these regions, the RMSD converges at about 2.5 Å after 7ns **(Figure 6A).** The most important residues for the binding, Asp532 of RhoBTB3, as mentioned by Eric J. Espinosa et.al **[15]**, makes two hydrogen bonds with Arg68 of Rab9. The OD1 of Asp532 makes two hydrogen bonds with the HH11 and HH21 of Arg68. Another important residue Ile533 of RhoBTB3, as mentioned by Eric J. Espinosa et.al, also makes a hydrogen bond with Gly41 of Rab9. In addition, Leu531 of RhoBTB3 makes a hydrogen bond with Gln66 and further strengthens the binding interface (**Figure 6B).** We observed from the MD simulation that the binding site fluctuates a little bit but the important residues involved in binding do not change before and after the simulation.

The root mean square fluctuation (RMSF) of the system was also calculated. As the RMSF calculation is used to identify the flexible region in protein or complex, the residues having low RMSF value is regard as more stable. As shown in **Figure 7**, the binding residues in the interface, which are highlighted in gray, have significant lower RMSF values (ranging from 2-3 Å) than the rest of the protein. Therefore the binding interface lies in the stable portion of the protein complex.

Conclusions:
The modeled structures of human Rab9 and Rho_Cterm are of good quality as suggested by the Ramachandran plot, Procheck program and ProSA server. The unrestrained pair wise docking performed with ZDOCK indicates that Asp532 and Ile533 of human RhoBTB3 are involved in the binding of Rab9, which is consistent with the published experimental results. As revealed by the complex structure, the binding interface on the Rab9 side includes Gly41, Gln66 and Arg68. The MD simulation performed with the complex shows the complex is stable along the 40ns simulation process. Further PCR mutagenesis is needed to confirm the importance of interface residues in Rab9 that are indicated in our prediction.

References:
[1] Barr F & *Lambright DG, Curr Opin Cell Biol.* 2010 **22:** 461 [PMID: 20466531]
[2] Zerial M & *McBride H, Nat Rev Mol Cell Biol.* 2001 **2:** 107 [PMID: 11252952]
[3] Pfeffer SR, *Trends Cell Biol.* 2001 **11: 487** [PMID: 11719054]
[4] Pfeffer SR, *Curr Opin Cell Biol.* 1994 **6**: 522 [PMID: 7986528]
[5] Bourne HR *et al. Nature.* 1991 **349**: 117 [PMID: 1898771]
[6] Sprang SR, *Annu Rev Biochem.* 1997 **66**: 639 [PMID: 9242920]
[7] Barbero P *et al. J Cell Biol.* 2002 **156**: 511 [PMID: 11827983]
[8] Lombardi D *et al. EMBO J.* 1993 **12:** 677 [PMID: 8440258]
[9] Riederer MA *et al. J Cell Biol.* 1994 **125:** 573 [PMID: 7909812]
[10] Ganley IG *et al. Mol Biol Cell.* 2004 **15**: 5420 [PMID: 15456905]
[11] Murray JL *et al. J Virol.* 2005 **79:** 11742 [PMID: 16140752]
[12] McGourty K *et al. Science.* 2012 **338:** 963 [PMID: 23162002]
[13] Carroll KS *et al. Science.* 2001 **292:** 1373 [PMID: 11359012]
[14] Diaz E *et al. J Cell Biol.* 1997 **138:** 283 [PMID: 9230071]
[15] Espinosa EJ *et al. Cell.* 2009 **137:** 938 [PMID: 19490898]
[16] Hanna J *et al. Proc Natl Acad Sci U S A.* 2002 **99:** 7450 [PMID 12032303]
[17] Reddy JV *et al. Mol Biol Cell.* 2006 **17:** 4353 [PMID: 16885419]
[18] Nottingham RM *et al. J Biol Chem.* 2011 **286:** 33213 [PMID: 21808068]
[19] Nottingham RM *et al. J Biol Chem.* 2012 **287:** 22740 [PMID: 22637480]
[20] Jaffe AB & *Hall A, Annu Rev Cell Dev Biol.* 2005 **21:** 247 [PMID: 16212495]
[21] Ridley AJ, *Trends Cell Biol.* 2006 **16:** 522 [PMID: 16949823]
[22] Aznar S *et al. Cancer Lett.* 2004 **206:** 181 [PMID: 15013523]
[23] Kang MI *et al. Proc Natl Acad Sci U S A.* 2004 **101:** 2046 [PMID: 14764898]
[24] Bomont P *et al. Nat Genet.* 2000 **26**: 370 [PMID: 11062483]
[25] Ahmad KF *et al. Mol Cell.* 2003 **12:** 1551 [PMID: 14690607]
[26] Melnick A *et al. Mol Cell Biol.* 2000 **20:** 6550 [PMID: 10938130]

[27] Kobayashi A *et al. Mol Cell Biol.* 2004 **24:** 7130 [PMID: 15282312]

[28] Chen R *et al. Proteins.* 2003 **52:** 80 [PMID: 12784371]

[29] Altschul SF *et al. J Mol Biol.* 1990 **215:** 403 [PMID: 2231712]

[30] Berman HM *et al. Nucleic Acids Res.* 2000 **28: 235** [PMID: 10592235]

[31] Levitt M, *J Mol Biol.* 1992 **226:** 507 [PMID: 1640463]

[32] Fechteler T *et al. J Mol Biol.* 1995 **253**: 114 [PMID: 7473707]

[33] Wiederstein M & Sippl MJ, *Nucleic Acids Res.* 2007 **35:** W407 [PMID: 17517781]

[34] Laskowski RA *et al. J Biomol NMR.* 1996 **8:** 477 [PMID: 9008363]

[35] Jorgensen WL *et al. The Journal of chemical physics.* 1983 **79**

[36] Grosdidier S *et al. Curr Pharm Des.* 2012 **18:** 4607 [PMID: 22650255]

[37] Zu X *et al. Antiviral Res.* 2014 **104:** 7 [PMID: 24468276]

[38] Kumar K *et al. Adv Virol.* 2013 **2013:** 594319 [PMID: 24288532]

BIOINFORMATION

Supplementary material:

Table 1: Determination of Protein Geometry Parameters by Procheck

	Observed	
	Rab9	Rho_Cterm
Residues in the most favored regions	84.1%	81.7%
Residues in additional allowed regions	14.3%	18.3%
Residues in generously allowed regions	1.1%	0.0%
Residues in disallowed regions	0.5%	0.0%