Istanbul Medipol University

School of Engineering and Natural Sciences

Graduation Project

2021-2022

PROJECT TITLE
A new approach to the activation mechanism of ADP-ribosylation factor (ARF): Investigation of the effect of the membrane and posttranslational modification on the activation mechanism with the help of computational methods
FACULTY ADVISOR
Asst. Prof. Ozge Sensoy
TEAM MEMBERS
Rumeysa Hanife Kars

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

Project Code : BME4110775
Project Title: A new approach to the activation mechanism of ADP-ribosylation factor (ARF): Investigation of the effect of the membrane and posttranslational modification on the activation mechanism with the help of computational methods
Faculty Advisor: Asst. Prof. Ozge Sensoy
Project Team Members: Rumeysa Hanife Kars
Sponsor Company (if any) : TUBITAK

BUDGET (TL)	PROPOSED	CONSENTED
IMU FUNDING		
SPONSOR COMPANY FUNDING	4000 TL	3952 TL
TOTAL		3952 TL

PROJECT PLAN	PROPOSED	CONSENTED
PROJECT PLAN	**14 Weeks**	**14 Weeks**
STARTING DATE	**04.10.2021**	**04.10.2021**

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

Project Code : BME4110775	
ADVISOR	**DEPARTMENT CHAIR**
Name: Asst. Prof. Ozge Sensoy	**Name: Prof. Yasemin Yuksel Durmaz**
Contact Information: Tel 0216 681 5621 E-mail : osensoy@medipol.edu.tr	**Contact Information:** Tel : E-mail : ydurmaz@medipol.edu.tr
Signature:	**Signature:**

TEAM MEMBER	**TEAM MEMBER**
Name: Rumeysa Hanife Kars	**Name:**
Contact Information: Tel : E-mail : rumeysa.kars@std.medipol.edu.tr	**Contact Information:** Tel : E-mail :
Signature:	**Signature:**

Istanbul Medipol University
School of Engineering and Natural Sciences
Graduation Project

Project Title: A new approach to the activation mechanism of ADP-ribosylation factor (ARF): Investigation of the effect of the membrane and posttranslational modification on the activation mechanism with the help of computational methods

Faculty Advisor: Asst. Prof. Ozge Sensoy

Team Members: Rumeysa Hanife Kars

PROJECT OVERVIEW/SUMMARY/ABSTRACT

In eukaryotic cells, the maintenance of the intracellular traffic is required for ensuring precise communication among intracellular compartments. In particular, ADP-ribosylation factors (ARF), which are encoded by CDKN2a locus, play an important role in regulation of membran traffic, dynamics and intracellular transport. Besides, ARFs are also involved in regulation of cell proliferation, senescence, and apoptosis. Structurally, ARF protein bears an amphiphilic helix at its N-terminus, unlike other GTPases, which can undergo posttranslational modificatio such as myristoylation, which has been shown to assist protein in its interaction with the membrane. Importantly, stabilization of the inactive conformation of the protein is provided by this amphipathic helix in the GDP-bound state. [1]

Existing crystal structures of ARF has revealed that major conformational changes occur especially in Switch 1 (SWI), Switch 2 (SWII) and the region that connects SWI and SWII, namely inter-switch region (ISR) during activation. In the GTP-bound protein, the N-terminal helix is displaced which triggers shifting of the ISR upwards by two amino acids (2-residue shi which has been thought to be catalyzed by guanine-exchange factors, thus making the nucleoti binding pocket available for GTP binding. Recently, it has been shown by our collaborator that ARF could adopt two-residue shift even in GDP-bound protein in the presence of the membran which has led to the questioning of the activation mechanism accepted so far. With this motivation, we set out to investigate the impact of membrane and its composition on the activation mechanism of ARF by using computational tools such as molecular docking and molecular dynamics simulations. Moreover, we will calculate the energy required for adopting active state of the protein in the presence of membrane with different compositions. The output of the project will provide a mechanistic insight to the activation mechanism of ARF, hence provide a platform for effective modulation of the protein under disease conditions [2].

Keywords: ADP-ribosylation factor (ARF), molecular dynamics, cancer, activation mechanism, membrane

School of Engineering and Natural Sciences
Graduation Project

Project Code BME4110775

BUDGET PROPOSED– (TL)

	ITEMS				
	PEOPLE	MACHINE-INSTRUMENT	MATERIALS	SERVICE	TRAVEL
IMU FUND	0	0	0	0	0
SPONSOR COMPANY FUND	0	4000 TL	0	0	0
TOTAL	0	4000 TL	0	0	0

BUDGET APPROVED– (TL)

	ITEMS				
	PEOPLE	MACHINE-INSTRUMENT*	MATERIALS*	SERVICE	TRAVEL
IMU FUND	0	0	0	0	0
SPONSOR COMPANY FUND	0	3952 TL	0	0	0
TOTAL	0	3952 TL	0	0	0

1. OBJECTIVE OF THE PROJECT: ADP-ribosylation factors (ARF) are involved in the regulation of membrane traffic, dynamics, intracellular transport, cell proliferation, senescence, and apoptosis. Many studies have revealed that cancer cells suppress proteins such as ARF, thus causing cell migration and invasion. Therefore, ARF has been considered as a hot target in cancer-focused therapies in an effort to prevent tumor progression. In this respect, having a holistic understanding of the activation mechanism of ARF is crucial for effective modulation of the protein. [3]

2. LITERATURE REVIEW:

The ability of ARF to function depends on the level of phosphorylation of the nucleotide attached to the protein (Figure 1). Accordingly, the protein is in inactive form when guanine nucleotide di-phosphate (GDP) is bound, and in active form when guanine nucleotide tri-phosphate (GTP) is bound. The transitions between the inactive and active forms of the protein proceed very slowly. For this reason, in order to speed up the processes, ARF is accompanied by two proteins, GEF (guanine nucleotide exchange factor) and GAP (GTPase activating protein). GEF accelerates the exchange of GDP with GTP and provides activation of the protein. Activated ARF, on the other hand, interacts with relevant cytosolic effectors, allowing the activation of signaling pathways to be controlled. On the other hand, GAP accelerates the hydrolysis of GTP and inactivates the protein. GDP-bound ARF was found in the cytoplasm, while GTP-bound protein was localized in the membrane. [4]

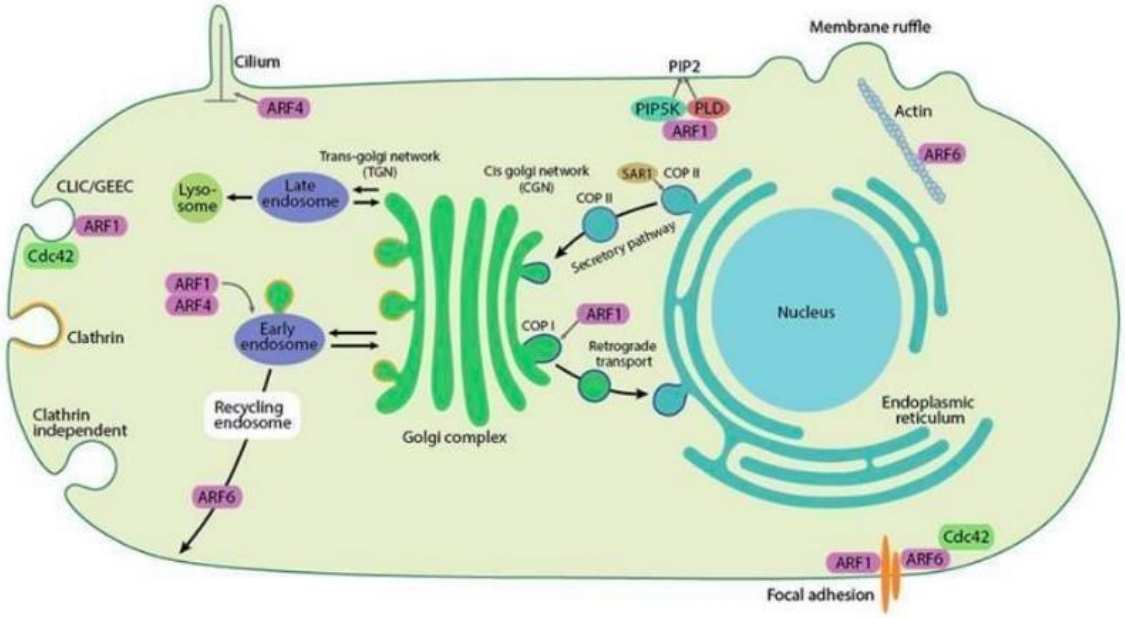

Figure 1. Schematic representation of the processes in the cell in which ARF is involved [5]

It has been shown that membrane localization of ARF is mediated by the amphiphilic helix located in the N-terminal region of the protein and attached to the myristoyl posttranslational group. The important point here is that the problems that occur in posttranslational modifications related to ARF play a role in the formation of cancer. Therefore, it is of great importance to examine the impact of myristoylation on the ARF function. [6]

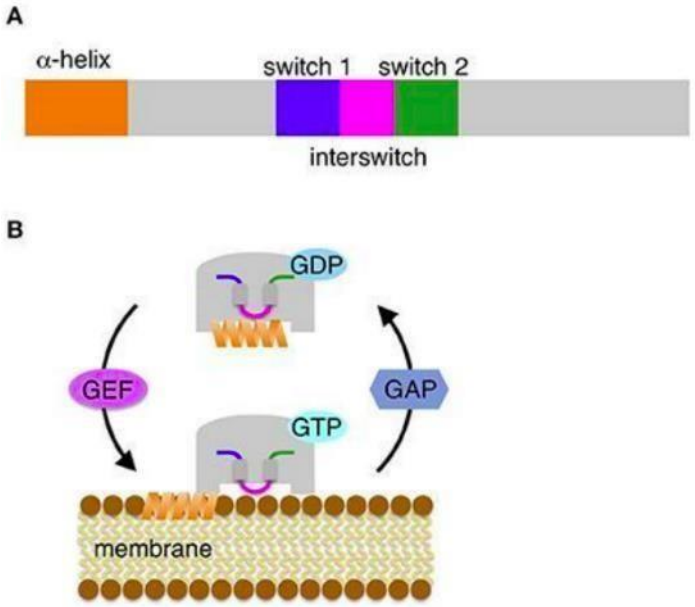

Figure 2. The role of GAP (GTPase activating protein) and GEF (guanine nucleotide exchange factor) in ARF activation/inactivation [7]

When the structures in which all amino acids belonging to ARF are present are examined, it has been observed that major conformational changes occur especially in the Switch 1 (SWI) and Switch 2 (SWII) regions of the protein during activation via GEF.[8] In addition, conformational changes occur in the inter-switch region (ISR), which connects SWI and SWII. According to the "ISR toggle" model, in the GDP-bound protein, the ISR is pulled to the center of the protein.[9] Aspartate amino acid, which is found in the ISR and is protected in the ARF protein family, closes the nucleotide binding pocket and prevents GTP and GEF binding to the protein in the cytoplasm (Figure 3).

In this conformation, the hydrophobic part of the amphiphilic helix at the N-terminus of the protein is packed into the nucleotide binding pocket, preventing the ISR from moving. In this way, the inactive conformation of ARF is stabilized in the cytoplasm. On the other hand, ARF is attached to the membrane by myristylation of the N- terminal helix. In this way, the N-terminal helix, which acts as a cap on the ISR, moves away from the ISR.[8] As a result of these conformational rearrangements, GEF binds to the binding pocket of the ARF, triggering an upward pull of the ISR by two amino acids (2-residue-shift). This conformational change also helps to pull up the conserved aspartate amino acid (indicated by D) that holds the nucleotide binding pocket, making the nucleotide binding pocket available for GTP to bind. The ISR, which shifts two amino acids up, prevents the N-terminal helix from moving back towards the nucleotide binding pocket and closing this region.

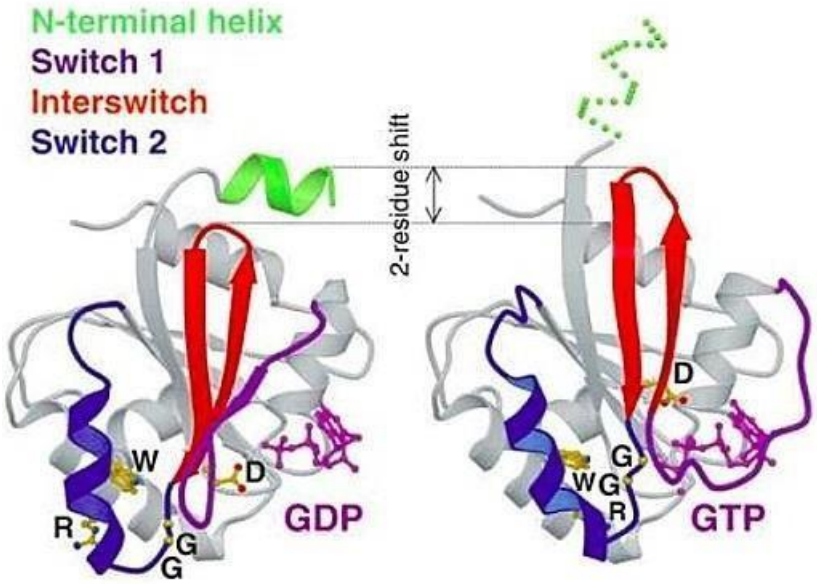

Figure 3. Schematic representation of the two-residue shift along with regions involved in the process. [9]

With the project proposal, the effect of membrane and myristoylation on the ARF activation mechanism will be examined with the help of computational methods. Our hypothesis that we want to investigate with the project proposal is that "conformational change (2-residue-shift) in the ISR region, which is of great importance in ARF activation, is made more energetically possible in the GDP-bound protein through membrane and myristoylation". In this way, it will be possible to test that, contrary to what has been thought up to now, GEF plays a role in nucleotide change rather than triggering the conformational change observed in the ISR. The project is very important as it has the potential to provide a new approach to the activation mechanism of ARF. As mentioned above, ARF has recently been included in studies as an alternative target protein in cancer therapies. In this respect, revealing the activation mechanism of ARF at the atomistic level will contribute to the effective control of the protein's function and thus to the treatment of diseases caused by ARF. [10]

3. ORIGINALITY: With the project proposal, the impact of the membrane and membrane composition on the ARF activation mechanism will be examined by utilizing computational methods. The working hypothesis that is aimed to test is that conformational rearrangement pertaining to the 2-residue-shift con occurring at the ISR region, which is of contribution to ARF activation, is made energetically more favorable in the GDP-bound protein through membrane. In this way, it will be possible to test that, contrary to what has been thought up to now, GEF plays a role in nucleotide change rather than triggering the conformational change observed in the ISR. The project is very important as it has the potential to provide a new approach to the activation mechanism of ARF. As mentioned above, ARF has recently been included in studies as an alternative target protein in cancer therapies. In this respect, revealing the activation mechanism of ARF at the atomistic level will contribute to the effective control of the protein's function, thus proposing effective strategies for treatment of diseases that are caused by ARF dysfunction.

4. SCOPE OF THE PROJECT AND EXPERIMENTS/METHODS:

The project will be performed by an undergraduate student, Rumeysa Hanife Kars, under the

supervision of Asst. Prof. Ozge Sensoy. Within the scope of this project, the impact of the membrane and membrane composition on the activation mechanism of ARF will be analyzed in by using molecular dynamics simulations and energy calculations. This project will be accomplished between October, 2021 and June, 2022, in fourteen weeks. Within the context of the project, homology modeling, molecular docking, molecular dynamics simulations and energy calculations will be performed.
In order to examine the effect of the myristoyl group on the activation mechanism of the protein, the crystal structure representing the cytoplasmic conformation of the protein will be simulated in the presence and absence of myristoyl. On the other hand, there is no myristoyl group in the new crystal structure representing the cytoplasmic conformation. There is no structure of ARF5 myristoylated at the N terminus in PDB. However, there is an NMR structure of myristoylated and GDP-bound ARF1 protein of S.cerevisiae with PDB ID of 2K5U.[11] There is more than one conformation of ARF1 in the NMR structure in question. First of all, the new crystal structure representing the cytoplasmic conformation of ARF5 will be aligned with the NMR structures and the structure with the crystal structure of ARF5 and the root mean square deviation (RMSD) value and root mean square fluctuation (RMSF) value will be determined.

In this way, it will be possible to transfer the myristoyl group in the NMR structure to ARF5. In the absence of such a structure in the NMR community, the myristoylated ARF5 protein will be modeled with reference to the myristoylated ARF1 protein. The amino acid sequence similarity between these two structures is around 70%. In the new crystal structure representing the membrane system, there is no N- terminal amphiphilic helix as mentioned before. Therefore, to model the N-terminal amphiphilic helix of the protein, the myristoylated and bicelle-bound ARF1 protein from S.cerevisiae with PDB ID 2KSQ will be used.[11] The membrane composition will be prepared to be 30% cholesterol and 70% sphingomyelin. This is because it can represent the membrane region, called the lipid raft, where signaling takes place in the membrane. The systems to be worked on within the scope of the project will be created using the CHARMM-GUI server. [12]

Molecular dynamics simulations to be performed in the membrane and cytoplasm are planned as 1 microsecond. At the end of this period, the simulation time will be extended if the system does not reach equilibrium (high RMSD). Two simulations with different initial velocity distribution will be made for each system. In this way, the reliability of the results obtained from the simulation will be tested. The simulations will be done using the Gromacs simulation package. [13] Charmm36m force field will be used to model the protein and TIP3P will be used to model water molecules. [12] 0.15 M KCl will be used to notify the electrical charge of the systems. Trajectory obtained from molecular dynamics simulations will be examined.

using the principal component analysis method. [14] According to this method, the trajectories obtained at the end of the molecular dynamics simulation will be aligned with a reference structure. As the reference structure, the Cα atoms of the crystal structures used as the starting

structure in the simulations will be used. In this way, the covariance matrices of the systems will be obtained. Then, these matrices will be diagonalized and the eigenvector and the eigenvalues of these vectors will be calculated using the g_covar and g_anaeig modules of the GROMACS simulation package. [13]

The eigenvalues obtained will be sorted in descending order and the eigenvectors corresponding to these eigenvalues will be examined and the basic dynamic modes of the system will be

determined. The eigenvector with the largest eigenvalue is thought to be responsible for the biological function of the protein. Structural and dynamical similarities and differences of the systems will be revealed by comparing the modes of trajectories of ARF5 with myristoylated and non-myristoylated N-terminal helix in the cytoplasm and membrane-bound ARF5 with myristoylated N-terminal helix. Then, PSN-Ensemble software will be used to determine the amino acids that are important for the dynamics of the protein in these systems. [15] Finally, dynamic correlation matrices will be used to examine the correlation between amino acids in these systems. Dynamic cross-correlation maps will be obtained to measure the correlation between amino acids. Accordingly, the normalized covariance values of amino acids will be calculated with calculating Cα atoms and seeing the below equation (Equation 1):

$$DCC(,j)= \frac{<\Delta r_i(t).\Delta r_j(t)>_t}{\sqrt{||\Delta r_i(t)||^2>_t}} \frac{1}{\sqrt{||\Delta r_j(t)||^2>_t}} \quad (1)$$

In this equation, ri(t) and rj(t) represent the positions of the i and jth atoms as a number of time t. < > shows the time ensemble average. Δri(t) = ri(t) − (< ri(t) >)t and Δrj(t) = rj(t) − (<rj(t) >)t. The range of relation of residues is from -1.0 to 1.0, in which the positively relative alteration is between 0-1.0 and the anti-correlation displacement is between - 1.0-0. GDP nucleotide is attached to the protein in the newly obtained crystal structure and the ISR has made the aforementioned 2-amino acid shift. This conformational change was not observed in the existing crystal and NMR structures of ARF5, which represent GDP-bound and membrane. Within the scope of the project proposal, the effect of the membrane on this conformational change will be investigated. In this context, energy calculations will be started with a GDP-bound structure in which 2 amino acid shifts are not observed in the ISR. For energy calculations, metadynamics, one of the enhanced conformational sampling techniques available in the Plumed program, will be used. [16] The free energy is measured as a number of n predetermined order parameters, the collective variables (CVs) utilized to generate the historical bias potential can be calculated through Equation 2 [17] :

$$V(,t)= \sum_{t\prime<t} w_{t^i} \prod_{i=1}^{n} \exp\left(\frac{[s_i(R(t)) - s_i(R(t\prime))^2\}}{\sigma_i^2}\right) \quad (2)$$

In this equation, si(R), 1 ≤ i ≤ n, represents CVs, t ′ is a multiple of the deposition duration τ and wt ′ and σi represent the width and height of the accumulated Gaussian potentials (peaks), respectively. The following equation (Equation 3) is a kind of well-ordered metadynamics, the original metadynamics algorithm in which the Gaussian height wt' is spontaneously rearranged. In this equation, in which ΔT is a value with the temperature dimension, kB is the Boltzmann constant and w is a value energy showing the highest height of the Gaussian potentials. [18]

$$w_{t'} = w exp\left(\frac{-V(R,t')}{k_B \Delta T}\right) \quad (3)$$

Accordingly, the distance from the center of mass of ARF5 to the membrane will be used as the collective variable. In this way, the conformational state of the ISR can be examined according to the change in the distance of ARF5 from the membrane. Our prediction is that when there is an energetically optimal orientation of ARF5 relative to the membrane, the expected conformational change in the ISR will occur. While making energy calculations, the correlation between amino acids and the amino acids that play a role in the activation mechanism will also

be determined. Finally, since the N-terminal amphiphilic helix interacts with the membrane, the composition of the membrane will also have an effect on the energy values to be obtained. For this reason, these energy calculations will be repeated in the presence of negatively charged lipids (phosphatidylinositol 4,5-bisphosphate (PIP2)) and energy values will be compared.

The RMSD graph, is the calculation of the medium distance along the atoms of some proteins that are superimposed, of each system has been delivered and to indicate the differences between systems and what are the distinctions in the systems has been observed respectively, calculated using Equation 4 where reference structure, r_{ref} and where M = Σi mi and ri (t) is the conformation of atom i at duration t after smallest square orientation the structure to the reference molecule.

$$\mathrm{RMSD}(t) = \left[\frac{1}{M} \sum_{i=1}^{N} m_i |\mathbf{r}_i(t) - \mathbf{r}_i^{\mathrm{ref}}|^2 \right]^{1/2} \qquad (4)$$

The RMSF graph, is a measure of the deviation between the position of particle i and some reference position, of each system has been delivered and to indicate the differences between systems and what are the distinctions in the systems has been observed respectively, shown in Equation 5, in which mT is the time over that one needs to mean and r_i^{ref} is the reference conformation of particle i. This reference place will be the time-averaged conformation of the same molecule i.

$$\mathrm{RMSF}_i = \left[\frac{1}{T} \sum_{t_j=1}^{T} |\mathbf{r}_i(t_j) - \mathbf{r}_i^{\mathrm{ref}}|^2 \right]^{1/2} \qquad (5)$$

5. PROJECT TARGETS AND SUCCESS CRITERIA:

It has been thought that, the conformational change seen in the ISR is catalyzed by guanine exchange factor during activation; however it has been shown by our collaborator that the conformational rearrangement pertaining to two-amino acid shift rea has been observed in the crystal structure of the GDP-bound ARF in the presence of the membrane. Therefore, this observation has caused to the questioning of the activation mechanism accepted so far: Does the membrane also contribute to ARF activation? With this motivation, in our project, for the first time, the impact of membrane and membrane composition will be examined with the help of computational methods. The aim of the project proposal is to examine the impact of membrane and membrane composition on the activation mechanism of ARF by means of computational methods. In order to achieve this aim:

1. The structure and dynamics of ARF will be examined in the cytoplasm and at the membrane having different lipid compositions.
2. The energy required to adopt two-residue conformational shift will be calculated at different membrane compositions.
3. The allosteric mechanism that provides GDP-bound ARF to undergo two-residue-shift mechanism will be elucidated.

I. Modeling and preparation of the systems studied for using in molecular dynamics simulations (20%)
II. Performing molecular dynamics simulations for at least 1 microsecond and with 2 replicas for each system (45%)
III. Identifying differences or similarities between structural and dynamic properties of systems studied in the cytoplasm and at the membrane (35%)

6. RISKS AND B PLANS:

WP	Risks	B plan
1	Getting errors in molecular dynamics simulations related to modeled systems	Using alternative structures with high sequence similarity instead of reference structures used in modeling, or starting the main simulation with a structurally more stable system by making the minimization steps longer before starting the main molecular dynamics simulations.
2	a) The results obtained from the second replica planned for each system are different from those of the first replica b) There is no difference between the systems planned to work within the scope of the project.	a) In such cases, a 3rd replica will be made and it will be examined which system the 3rd replica behaves more similar to. b) Instead of classical molecular dynamics simulation, techniques such as accelerated molecular dynamics that can increase conformational sampling will be used.
3	Failure to achieve balance in energy calculations and therefore a high margin of error	In such cases, different collective variables can be used or the parameters to be used in metadynamics simulations can be changed.

7. WORK TIME PLAN OF THE PROJECT:

WP No	Detailed Definition of Work and Activity
1	Preparation of systems to be examined within the scope of the project
2	Performing and Analysis of Molecular Dynamics Simulations
3	Energy Calculations

Work package	Target	Measurable outcome	Contribution to overall success(%)
1	Ability to use modeled systems in molecular dynamics simulations without any errors	Finishing 1microsecond molecular dynamics simulations without any error	20
2	Performing molecular dynamics simulations on the systems specified in the project proposal for at least 1 microsecond and with 2 replicas for each system and Identifying differences or similarities between structural and dynamic properties of systems	Understanding the properties of systems properly	45
3	Calculation of the energy calculation with the help of the metadynamic algorithm implemented in the Plumed program and repeat for different membrane compositions	Elucidation of the allosteric mechanism which is responsible for activation of the protein and understanding the impact of membrane composition on activation mechanism	35
			Total:100

Project Member	WP1	WP2	WP3
R.Hanife Kars	100	100	100
Total	100	100	100

8. FINANCIAL EVALUATION:

With the budget requested within the scope of the project, it is planned to purchase 2 5 TB portable external disks. The reason for this is that the data is not backed up in the parallel computer system within the university. The data to be obtained within the scope of the project can be stored with the external disk. Therefore, in cases where the computer system is intensive, the executive will provide himself an alternative source by using the remaining budget after the external disk is procured to support the purchase of a tablet.

9. RESULTS:

In this project, it is also aimed to see the effects of lipids to the membrane like charged lipids and also neutral lipids. In order to see the differences of distinct systems, four systems have been prepared. The systems are prepared for molecular dynamics (MD) simulations using the parameters given in the methods section. Water environment system and its replica, the system that has neutral lipids and the system that has charged lipids. The neutral lipid systems, it includes DPPC as a lipid type. For the charged lipid system, it has PIP2 and DPPS as a charged lipid types. After finishing the molecular dynamics simulations, we will able to observe the impact of different types of lipid as well as the environment. These MD simulations have been performed as 1 microseconds. The systems are generated using CHARMM-GUI server.
Firstly, using ARF5 Human protein has been used to prepare systems. Several alignments, minimizations and homology modeling methods have been performed. GLYM lipid tail has been added to first residues and adjusted to interact with the membrane.
The water box system can be seen in Figure 4.

Figure 4. Difference between first frame and last frame in the water box system

(yellow protein represents 1st frame's protein and blue protein represents last frame's protein; brown licorice represents GLYM in the 1st frame and green licorice represents GLYM in the last frame; purple licorice represents GDP in the 1st frame, pink licorice represents GDP in the last frame)

Similarly, the neutral charged lipid system that has DPPC in the structure, has been prepared as 50% of DPPC in the upper leaflet and 50% of DPPC in the lower leaflet. The neutral lipid system can be seen in Figure 5.

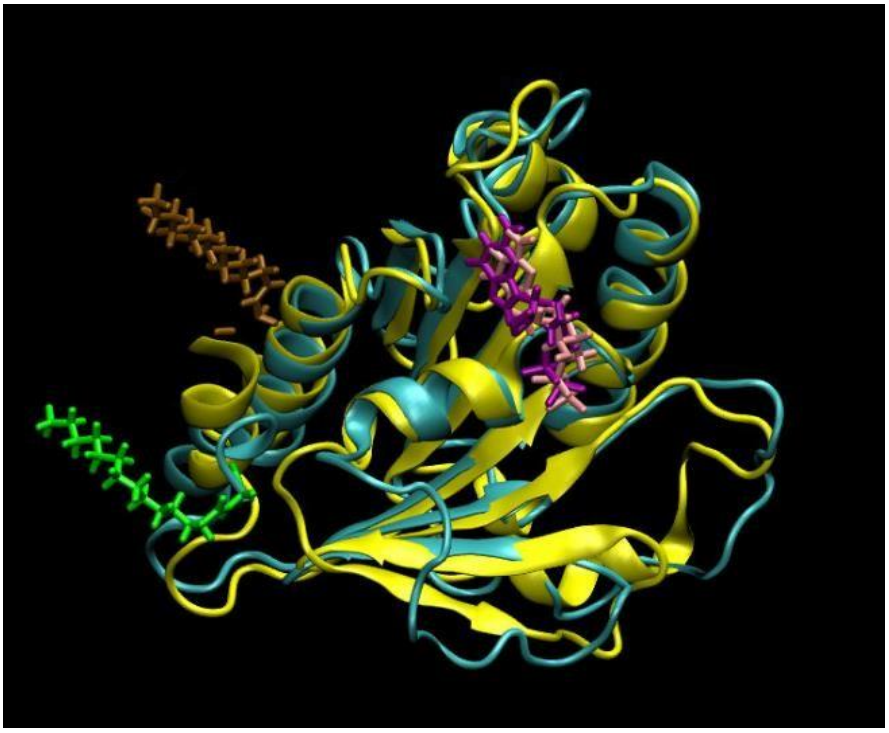

Figure 5. Difference between first frame and last frame in the neutral lipid system

(yellow protein represents 1st frame's protein and blue protein represents last frame's protein; brown licorice represents GLYM in the 1st frame and green licorice represents GLYM in the last frame; purple licorice represents GDP in the 1st frame, pink licorice represents GDP in the last frame)

Additionally, the charged lipid system that has DPPC, DPPS, PIP2 in the structure, has been prepared as 42% of DPPC, 8% of PIP2 and 50% DPPS has been used to prepare system. 50% DPPS in the upper leaflet and 42% of DPPC, 8% of PIP2 (that interacts with the protein) in the lower leaflet.

The charged lipid system can be seen in Figure 6.

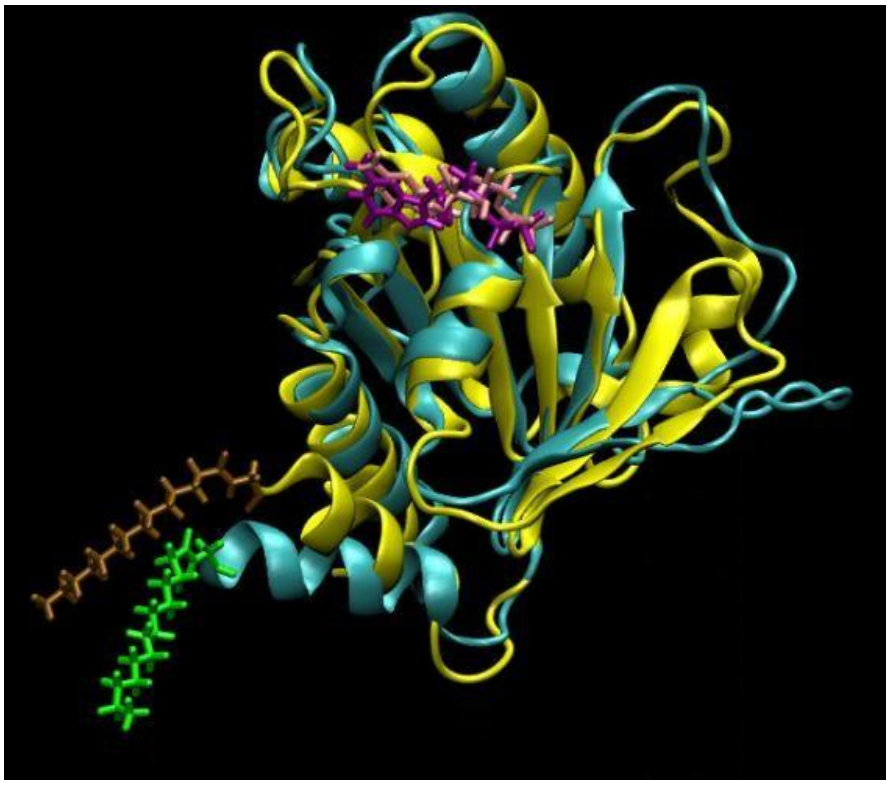

Figure 6. Difference between first frame and last frame in the charged lipid system

(yellow protein represents 1st frame's protein and blue protein represents last frame's protein; brown licorice represents GLYM in the 1st frame and green licorice represents GLYM in the last frame; purple licorice represents GDP in the 1st frame, pink licorice represents GDP in the last frame)

Several analyzes have been performed, RMSF (root- mean-square fluctuation), RMSD (root-mean-square deviation) and DCCM (dynamic cross correlation maps) representations have been shown accordingly.

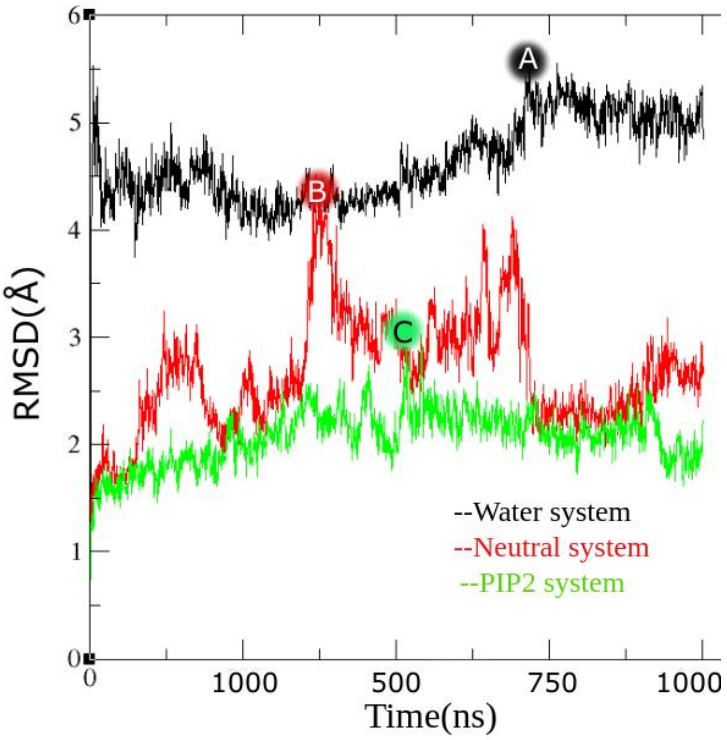

Figure 7. RMSD graphs of each system representation.

It has been observed that membrane composition has an effect on the system's fate. The black line represents the charged lipid system which has more nm value compared to neutral lipid system. At 150th frame of charged lipid system has great value of nm whereas neutral lipid does not have that much high number. At 500th frame of neutral system, the highest nm value has been observed obviously.

The charged lipid system has decreased sharply around 700th frame since the structure is kind of going to left-hand side and far away with each other.

In the water box environment, at 10th frame of system, it has a sharp increase in the first replica. It has been obviously seen that the behavior of water box environments does not change too much, it has a same trend line almost.

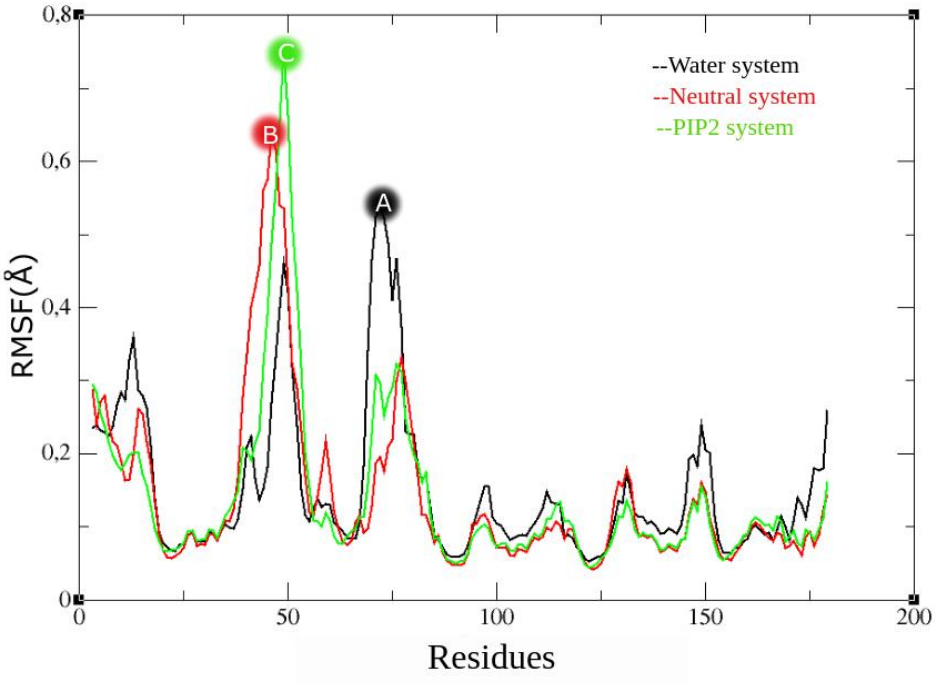

Figure 8. RMSF graphs of each system representation.

In the Figure 8, it has been understood that around 10th residue (SER), we have more fluctuation as well as the 50th (GLY) residue. Especially, PIP2 system that has charged lipids inside it, has a great raise in the 50th residue. Additionally, in around 70th (GLY) residue, 110th (GLY) residue and 150th (SER)residue, there is a big value of RMSF as nm. The highest RMSF value belongs to PIP2 system whereas the lowest RMSF value belongs to the water system. High RMSF reports on more fluctuation in a given system. One can say that around residues between 40 to 50, there is more fluctuation in systems.

When the first and the last frames of membrane-based systems were compared, there was a indeed 2-2.5 Å displacement in the interswitch region(ISR) as can be seen in Figure 9. However, as we predicted, this was not observed in the water box system. This situation proves that we have a shifting in the presence of the membrane.

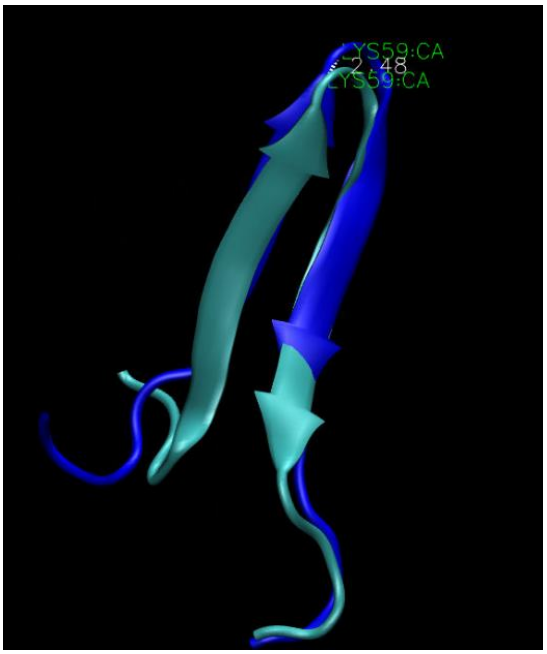

Figure 9. Shifting in the interested area.

To understand the distribution of the distances between systems, we have plotted the overall data in a single chart. By doing so, we are able to see the differences in the systems to be compared. Figure 10 shows this distribution among systems using histograms,which is a graphical representation of data that uses bars of varying heights to group numbers into ranges since it can aid in the detection of any odd findings or data gaps.

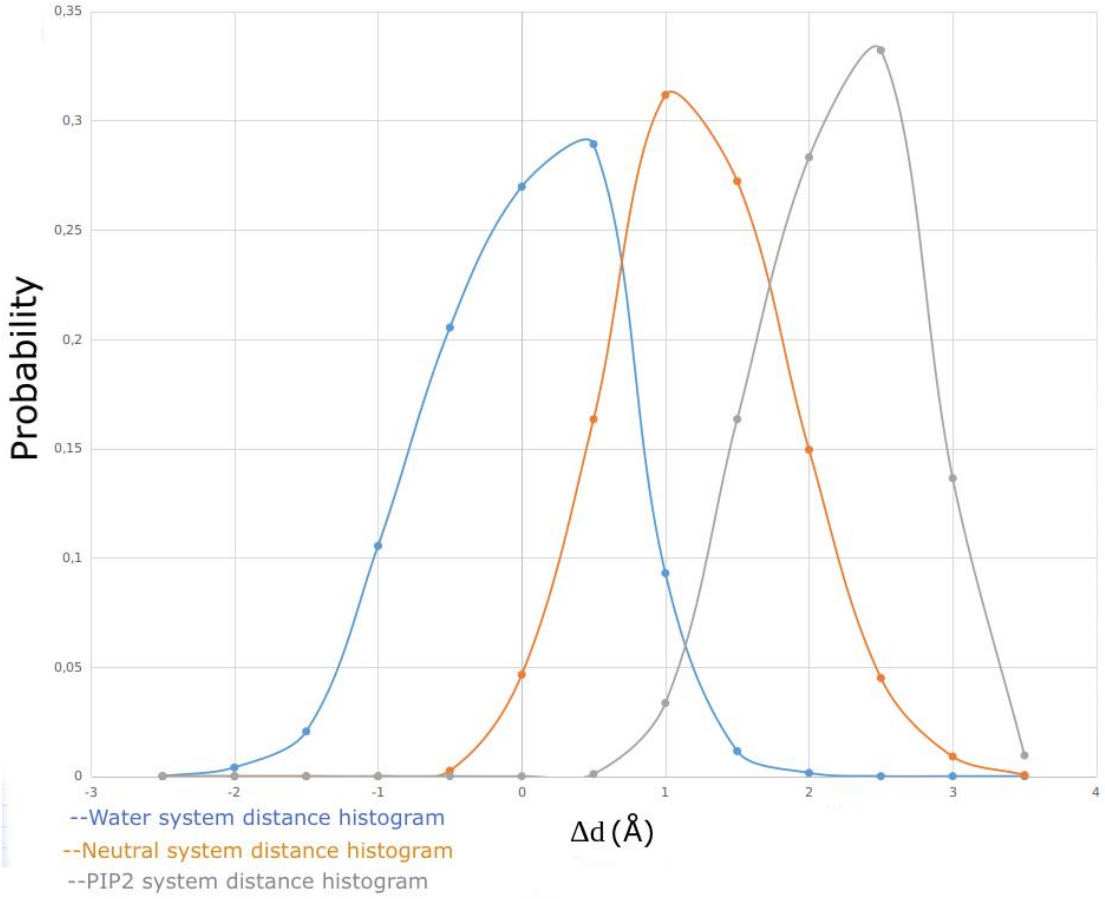

Figure 10. Histograms that show the distance change in the interswitch region of the systems studied.

This figure shows the probability of the changes in the ISR distances for all systems studied. It is seen that, when we add charged lipids, we have more shifting in residues around as much as 2.5Å. On the other hand, in water system and neutral systems, there are less shifting. The water system , neutral system and PIP2 system displayed changes between -2.5Å to 3Å, -0.5Å to 3.5Å, 0.5Å to 3.5Å, respectively. This chart shows the data by having probability on the y-axis and Δd (alterations in distances within the system) on the x-axis.

As it is stated in the methods, lastly energy calculations were calculated for each system. To predict energy values of the systems, we load the data to MATLAB and calculated the energy required to shift the ISR in the systems as shown in Figure 11.

In the neutral system, there is the least energy needed for the ISR shift whereas in water system required more energy to change its position.

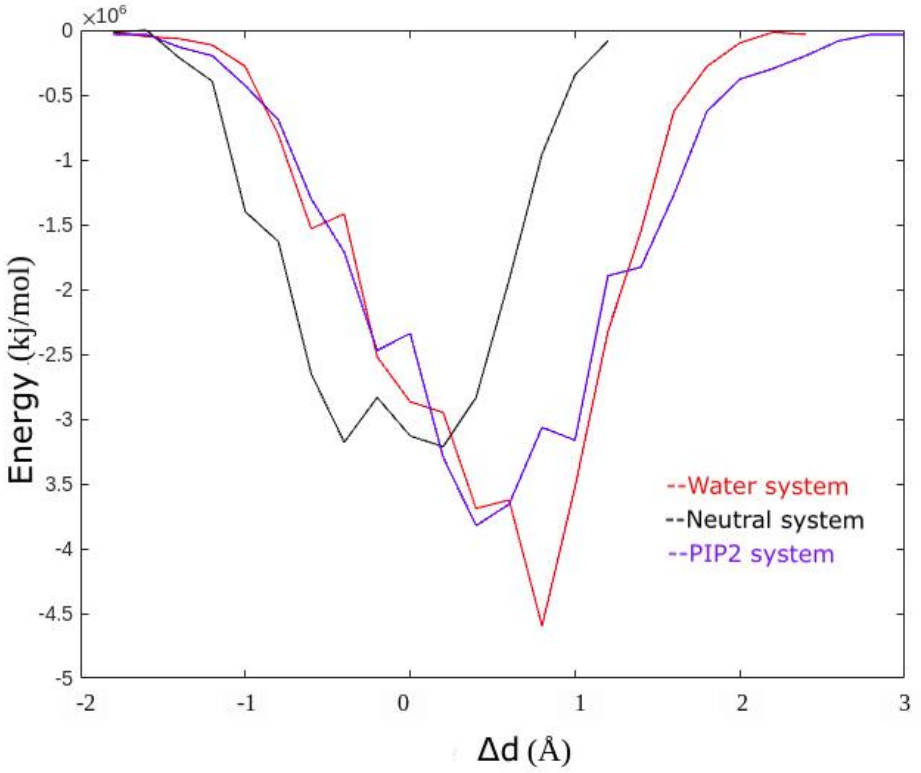

Figure 11: Comparison of energy values required for the ISR shift of the systems studied.

10. DISCUSSION:

In the water box environment, when we align the first and the last residues, the change in the GDP and GLYM residues can be seen in the Figure 4. Since in the water system, there is more fluctuation compared to other systems, the residues GLYM, GDP are more likely to replace more. These residues want to change their places more frequently than the other two systems, charged lipid and neutral lipid. However, the place of GDP did not alter too much and change in GDP residues were the same for all the systems. GLYM residues changed their positions more compared to charged lipid one.

In the neutral lipid system, when we align the first and the last residues, the change in the GDP and GLYM residues can be seen in the Figure 5. The rotation of GLYM has changed quite but the angle between GLYM and protein did not change too much, it seems nearly 90 degree. Also, the place of GDP did not change quite and alteration in GDP residues were the same for all the systems as well as the neutral lipid one.

In the charged lipid system, when we align the first and the last residues, the change in the GDP and GLYM residues can be seen in the Figure 6. The placement of GLYM has altered but the angle between GLYM and the protein has changed, it wanted to twist more by the finishing of the simulation since it is curled a lot in the last frame.

After analyzing RMSD graphs, it has been understood that water box systems (1st and 2nd replicas) have almost the same chart. At the 10th residue in the water box environments, there was a huge distinction. It may not get used to its place for a while around 10th residue, then it satisfies with the place where it can orient around the environment by the effects of water molecules. Once it recognizes its place, the water box environment is more likely to become stable. This can be understood by the changing of RMSD value. If RMSD value is a small number, it means represents good reproduction of the correct pose and it shows the better the model is in comparison to the target structure. In this manner, water box environment became more correct place while the running of simulation. It can be said that the average value of water box environments is around 3.5 nm for 1 microsecond simulation. In order to comprehend the fluctuation, RMSF chart should be considered. When one sees the RMSF of water box environments, the values are quite similar to each other, they follow the same line as well. Since water environment is more flexible compared to other systems, its RMSF value is higher, which means that it fluctuated more in the simulation. This outcome was expectable since water molecules are more mobile than other lipid compositions. The average value of RMSF values in the water box 0.2 nm which is quite higher than others.

For the difference in the charged lipid and the neutral lipid system in RMSD chart, they are not following the same trend. Neutral system is more stable than charged lipid system. In the charged one, there is more fluctuation as more increasings and decreasings. However, in the neutral lipid system, there are two increasings for the whole simulation. It can be said that charged lipid membranes are mostly dynamic molecules included in several cell biological events like morphogenesis. Especially, phosphatidylinositol 4,5- bisphosphate (PI(4,5)P2), that is used in the simulations, has localized conformation and high density of charge and it is not distributed uniformly and may be localized in raft domains. Another reason charged lipid systems are more mobile is that PIP2 reorganizes surrounding lipids and stimulates the creation of a PIP2 lipid microdomain, allowing it to behave as a protein-attaching molecule. The other reason may be coming out of the DPPS. It is evident that the presence of charged DPPS lipids in a leaflet alters the structural and dynamic characteristics of the lipids on the opposite leaflet of the membrane.

In contrast, the order parameter of the DPPC in the leaflet incorporating DPPS increased considerably as compared to values measured from a DPPC bilayer. The discrepancy might be attributed to the order and flexibility of lipids. For example, PIP2 was observed to induce local perturbations in the bilayer. All these reasons can show us why we have more replacements in RMSD graph.

Furthermore, these reasons can also explain the RMSF graph. In RMSF graph, actually there is no big change in the charged lipid system and neutral lipid system. One point around 50th residue in the charged lipid system has high value of RMSF. The reason can be explained because of the complicated interplay of several physical interactions between the peptides and the lipid membrane. 50th residue corresponds to GLY residue and between 47th and 52nd residue, we have a beta sheet that we want it to transform into a coil structure in the future studies. It has been shown in previous analyzes that this beta sheet structure from 47th residue to 52nd residue

is so mobile, and it has moving structure. Since the GLY is belong to that fragment.

Eventually, the system behaviors and reactions are almost like each other although in some places has some differences. In the end, they are following the same trend line in the RMSF graph. However, in the RMSD graph, the systems are not quite same, there were several differences.

When it comes to distance changes in the interested area, the systems have several distinctions. If these changes are ranked, PIP2 system has the highest displacement, then neutral lipid system, lastly water box system,shown in Figure 10. This proves that membrane and lipids have an effect on the movement in the interswitch region. When we compare PIP2 system with neutral lipid system, there were more changes in PIP2 system and this shows that charged lipids in the membrane is effective for movement. In same way, if we compare membrane system with water box system, there were more movement in the membrane systems and this situation indicates that membrane effect is available and this effect was observed. This movements were shown in the probability chart in order to see the distribution of values.

Another analysis was energy calculations for whole systems to see how much energy is required for the displacements.In order to calculate this, several commands were implemented in MATLAB and result were shown in Figure 11. In neutral system, there is the least energy required for the movement in a system whereas in water system and PIP2 system require more energy to alter its position. This can show that when we have neutral lipid in the system, we need less energy to make an alteration compared to charged lipid system, PIP2 system. As we have more charged lipids, we have more energy to make these kinds of changes. On the other hand, in water box environment, we require more energy compared to membrane systems. This may indicate that as we have membrane component, we give less effort to displace in a system, it comes easy to change its position.

11. CONCLUSION:

In ARF activation, the conformational change seen in the ISR was thought to be catalyzed by GEF, but it was observed that the said change was present in the crystal structure representing GDP and membrane-bound ARF obtained by our collaborator group. For the first time, the observation of the conformational change in the ISR in GDP-dependent ARF has led to the questioning of the activation mechanism accepted so far and suggesting that the membrane also contributes to ARF activation. With this motivation, in our project, for the first time, the effect of membrane and myristilization on the activation mechanism of ARF will be examined with the help of computational methods. To understand this concept, several systems and compositions has been proceeded. Firstly, water box environment system has been created and given to MD simulation. Secondly, neutral lipid systems has been created by CHARMM-GUI server and given to the simulation and lastly, charged lipid system has been prepared and given to simulation. The alterations in the system will help us to get an idea of the effect of the composition in the system and how they are affected. RMSD, RMSF and DCCM analyzes and energy calculations have been done properly. It has been seen that there are a few distinctions are available in several residues and at the several frames. At the end, proposed displacement was observed for membrane based systems as 2-2.5 Armstrong, not in water box system. These displacements were shown in probability distribution chart as well for better understanding. Energy calculations has been made in all systems, showing that water box needs more energy than PIP2 system and PIP2 system needs more energy than neutral system in order to realize a displacement.

12. PLAN FOR FUTURE STUDIES:

The question “why” has been greatly answered and will be answered in the future systems as well. The future systems will include the hybrid systems that has both GDP and GTP structure as well as the beta sheet/coil conformation. These systems have been given to MD simulations, and in the continuity of this project, they will be analyzed and answered as well.

It is planned that the project outputs will be published in high impact journals together with the group that obtained the new crystal structures. As the outputs obtained within the scope of the project proposal will bring a new perspective to the activation mechanism of the ARF5 protein and other ARF/ARL protein family proteins, it has a high impact value and it is planned to apply for a patent after in vitro experiments for amino acids that are thought to play a role in the activation mechanism. After completing her undergraduate education, the project coordinator plans to target regions that play a role in the activation mechanism with small therapeutic molecules during her graduate studies. In this context, the project has a high potential to contribute to the training of researchers and the creation of new projects.

13. REFERENCES:

[1] Ko, A., Han, S. Y., and Song, J. 2018. “Regulatory Network of ARF in Cancer Development” ,Molecules and cells, 41, 381–389.

[2] Jiao, Y., Feng, Y., and Wang, X. 2018. “Regulation of Tumor Suppressor Gene CDKN2A

and Encoded p16-INK4a Protein by Covalent Modifications". Biochemistry (Moscow), 83, 1289-1298.
[3] Casalou, C., Ferreira, A., ve Barral, D. 2020. "The Role of ARF Family Proteins and Their Regulators and Effectors in Cancer Progression: A Therapeutic Perspective", Frontiers In Cell And Developmental Biology, 8, 4.
[4] Hofmann, I. ve Munro, S. 2006. "An N-terminally acetylated Arf-like GTPase is localised to lysosomes and affects their motility". Journal Of Cell Science, 119, 1494-1503.
[5] Jackson, C. ve Bouvet, S. 2014. "Arfs at a Glance", Journal Of Cell Science, 127, 4103-4109.
[6] Franco, M., vd. 1995. "Myristoylation of ADP-ribosylation Factor 1 Facilitates Nucleotide Exchange at Physiological Mg Levels", Journal Of Biological Chemistry, 270, 1337-1341.
[7]Gillingham, A. and Munro, S. 2007. "The Small G Proteins of the Arf Family and Their Regulators", Annual Review Of Cell And Developmental Biology, 23, 579-611.
[8]Goldberg, J. 1998. "Structural Basis for Activation of ARF GTPase", Cell, 95, 237-248.
[9]Pasqualato, S., vd. 2001. "The structural GDP/GTP cycle of human Arf6", EMBO Reports, 2, 234-238.
Cherfils, J., vd. 2000.Structure of Arf6–GDP suggests a basis for guanine nucleotide exchange factors specificity, Nature Structural Biology, 7, 466-469.
[10] Liu, Y., Kahn, R. and Prestegard, J. 2009. "Structure and Membrane Interaction of Myristoylated ARF1", Structure,17, 79-87.
[11] Mackerell, A., Feig, M. and Brooks, C. 2004. "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations", Journal Of Computational Chemistry, 25, 1400-1415.
[12] Abraham, M.J., vd. 2016. "GROMACS User Manual version".
[13] David, C. and Jacobs, D. 2013. "Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins", Protein Dynamics, 1084, 193-226.
[14] Bhattacharyya, M., Bhat, C. and Vishveshwara, S. 2013. "An automated approach to network features of protein structure ensembles", Protein Science, 22, 1399-1416.
[16]Bonomi, M., vd. 2009. "PLUMED: A portable plugin for free-energy calculations with molecular dynamics", Computer Physics Communications, 180, 1961-1972.
[17]Sensoy, O. and Weinstein, H. 2015. "A mechanistic role of Helix 8 in GPCRs: Computational modeling of the dopamine D2 receptor interaction with the GIPC1–PDZ domain", Biochimica Et Biophysica Acta (BBA) - Biomembranes, 1848, 976-983.
[18]Barducci, A., Bonomi, M. ve Parrinello, M. 2011. "Metadynamics", Wires Computational Molecular Science, 15, 826-843.

Istanbul Medipol University

School of Engineering and Natura Sciences

Graduation Project

PROJECT ACTIVITIES AND WORK PLAN

Work and Activity	1. week	2. week	3. week	4. week	5. week	6. week	7. week	8. week	9. week	10. week	11. week	12. week	13. week	14. week
1. Preparing the optimum protein composition for human ARF5 with myristyl group	X	X	X											
2. Given the protein to Charmm-Gui to prepare the system (atomic) in a water box			X	X										
3. Starting the MD simulations for water box environment				X	X									
4. Given the protein to Charmm-Gui to prepare the system (atomic) in a water box					X	X								
5. Given the protein to Charmm-Gui to prepare the system (atomic) in a water box						X	X							
6.Principal component analysis and using GROMACS for eigenvalue and eigenvector							X	X	X					
7. Obtaining dynamic cross-correlation maps									X	X				
8. Energy calculations										X	X	X	X	X

LIST OF WORK PACKAGES

WP No	Detailed Definition of Work and Activity
1	Preparation of systems to be examined within the scope of the project
2	Performing and Analysis of Molecular Dynamics Simulations
3	Energy Calculations

Work package	Target	Measurable outcome	Contribution to overall success(%)
1	Ability to use modeled systems in molecular dynamics simulations without any errors	Finishing 1microsecond molecular dynamics simulations without any error	20
2	Performing molecular dynamics simulations on the systems specified in the project proposal for at least 1 microsecond and with 2 replicas for each system and Identifying differences or similarities between structural and dynamic properties of systems	Understanding the properties of systems properly	45
3	Calculation of the energy calculation with the help of the metadynamic algorithm implemented in the Plumed program and repeat for different membrane compositions	Elucidation of the allosteric mechanism which is responsible for activation of the protein and understanding the impact of membrane composition on activation	35
			Total:100

WORK PACKAGE DISTRUBUTION						
Project Member	WP1	WP2	WP3	WP4	WP5	WP6
R.Hanife Kars	100	100	100	100	100	100
Total	100	100	100	100	100	100