IN SILICO PHOSPHORYLATION-INDEPENDENT ACTIVATION of ARRESTIN-3 PROTEIN by means of SMALL MOLECULES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF

ENGINEERING AND NATURAL SCIENCES

OF ISTANBUL MEDIPOL UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

BIOMEDICAL ENGINEERING AND BIOINFORMATICS

by

Mehmet Hanifi KURT

June, 2019

IN SILICO PHOSPHORYLATION-INDEPENDENT ACTIVATION of
ARRESTIN-3 PROTEIN by means of SMALL MOLECULES
By Mehmet Hanifi Kurt
June, 2019

We certify that we have read this thesis and that in our opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

_____
Assist. Prof. Dr. Özge Şensoy (Advisor)

_____
Assist. Prof. Dr. Mehmet Hikmet Üçışık

_____
Assist. Prof. Dr. Sefer Baday

Approved by the Graduate School of Engineering and Natural Sciences:

_____
Assoc. Prof. Dr. Yasemin Yüksel Durmaz
Director of the Graduate School of Engineering and Natural Sciences

ii

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Mehmet Hanifi KURT

Signature   :

ABSTRACT

IN SILICO PHOSPHORYLATION-INDEPENDENT ACTIVATION of ARRESTIN-3 PROTEIN by means of SMALL MOLECULES

Mehmet Hanifi Kurt

M.S. in Biomedical Engineering and Bioinformatics

Advisor: Assist. Prof. Dr. Özge Şensoy

June, 2019

G protein-coupled receptors (GPCRs) are responsible for communication of the cell with its surroundings. Upon ligand binding, a set of conformational changes occurs at the GPCR, which triggers activation and dissociation of G protein from the receptor. Subsequently, the receptor is phosphorylated, and activated/phosphorylated receptor causes recruitment of Arrestin to terminate signaling. Arrestin family is composed of four proteins, namely, Arrestin(Arr)-1, 2, 3 and 4. In spite of sharing a high structural similarity and conserved structural fold, the members display remarkable differences in their preference for receptor phosphorylation. Specifically, Arr-1/Arr-4 can exclusively bind to activated/phosphorylated Rhodopsin, whereas Arr-2/Arr-3 can bind to various types of GPCRs. Moreover, Arr-3 can also bind to non-phosphorylated receptors depending on the type of the receptor. The phosphorylation-independent activation mechanism remains elusive; but, might be utilized for the treatment of crucial diseases such as *congestive heart failure*. Until now, phosphorylation-independent Arrs have been attempted to be created but ended up with problems like instability of the protein.

In this thesis project, we aim to activate Arr-3 *in silico* by means of small molecules. To do so, we target key regions, which are involved in the activation mechanism, on Arr-3 using small molecules that are retrieved from the *ZINC* database. Our results show that small-molecule/Arrestin complexes are stable and the rotation angle, which is required for activation, is achieved. Therefore, this study provides a framework for the development of phosphorylation-independent Arr-3 and also an insight into the molecular mechanism of non-classical phosphorylation-independent activation mechanism.

ÖZET

ARRESTİN-3 PROTEİNİNİN *İN SİLİKO'da* KÜÇÜK MOLEKÜLLER YARDIMIYLA FOSFORİLASYONDAN BAĞIMSIZ OLARAK AKTİVASYONU

Mehmet Hanifi Kurt

Biyomedikal Mühendisliği ve Biyoinformatik, Yüksek Lisans

Tez Danışmanı: Dr. Öğr. Üyesi Özge Şensoy

Haziran, 2019

G protein-kenetli reseptörler (GPKR) hücre ve hücre çevresi iletişiminden sorumludur. Ligand bağlanmasını takiben GPKR'de, G proteinin aktivasyonuna ve reseptörden ayrılmasına neden olan bir dizi konformasyonel değişiklik meydana gelir. Daha sonra reseptör fosforile edilir ve aktif/fosforile edilmiş reseptör Arrestin alımına, sinyalin sonlandırılmasına neden olur. Arrestin protein ailesi, Arrestin(Arr) -1, 2, 3 ve 4'ten, yani dört proteinden oluşur. Bu ailenin üyeleri, yüksek yapısal benzerliklerlerine ve korunmuş yapısal katlanmalarına rağmen, reseptör forforilasyon tercihlerinde dikkate değer farklılıklar göstermektedir. Özellikle, Arr-1/Arr-4 sadece aktive edilmiş/fosforlanmış Rhodopsin'e bağlanabilirken, Arr-2/Arr-3 çeşitli GPKR tiplerine bağlanabilir. Ayrıca, Arr-3 reseptörün tipine bağlı olarak fosforile edilmemiş reseptörlere de bağlanabilir. Fosforilasyondan bağımsız aktivasyon mekanizması henüz tam olarak anlaşılmasa da; *konjestif kalp yetmezliği* gibi önemli hastalıkların tedavisinde kullanılabilir. Şimdiye kadar fosforilasyondan bağımsız Arrestinler yaratılmaya çalışılmıştır, ancak bu yöntem proteinin kararsızlığı gibi sorunlara neden olmuştur.

Bu tez projesinde, Arr-3'ü *in siliko'da* küçük moleküller ile aktif hale getirmeyi amaçlıyoruz. Bunu yapmak için, *ZINC* veritabanından alınan küçük molekülleri kullanarak, Arr-3'ün aktivasyon mekanizmasına dahil olan kilit bölgelerini hedefliyoruz. Sonuçlarımız, küçük moleküllü Arrestin komplekslerinin stabil olduğunu ve aktivasyon için gerekli olan dönme açısının elde edildiğini göstermektedir. Dolayısıyla bu çalışma, fosforilasyondan bağımsız Arr-3'ü geliştirmek ve klasik olmayan aktivasyon mekanizmasının moleküler mekanizmasını anlamak için kullanılabilecek bir bakış açısı sağlamaktadır.

Anahtar sözcükler: G-protein kenetli reseptörler, fosforilasyondan bağımsız Arrestin-3, konjestif kalp yetmezliği, hücre sinyalizasyonu, ilaç yeniden konumlandırma

Acknowledgment

Foremost, I would like to express my sincere gratitude to my advisor Dr. Özge Şensoy for the continuous support of my MSc study and researches, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my MSc study.

Besides my advisor, I would like to thank my colleagues, *Sensoy's Computational Biophysics Research Group* members. Especially Metehan who helped me with his technical knowledge and Hanife Pekel who helped me about selecting of drug molecules with her pharmacological background.

My sincere thanks also go to Mehmet Ören who is the administrator of the HPC of Medipol University. He has always helped us in any technical problem for the cluster system. Thanks to Medipol University for supplying computers, HPC and software licensing. Also, thanks to TÜBİTAK for supporting our project.

Last but not least, I would like to thank my wife; Zeynep for always supporting me and for being with me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

G protein-coupled receptors (GPCRs) constitute the largest family of cell surface receptors which are involved in cell signaling. Therefore, any problem occurring in this protein family results with the onset of many crucial diseases. As such, they have been targeted by ca. 40% of currently prescribed drugs in the drug market [1]. GPCRs act as conduits by transmitting the extracellular signal to the cytoplasm to fulfill cellular needs. Mechanistically, ligand binding causes a set of conformational changes at the GPCR, which are transmitted across the membrane, from the extracellular part to the intracellular site of the receptor. These conformational changes are recognized by the primary effector, namely heterotrimeric G protein, and then it is activated. Subsequently, G protein dissociates from the receptor to initiate the target signaling cascade in the cytoplasm. Consequently, the receptor is phosphorylated by corresponding GPCR kinases which leads to the recruitment of another cytoplasmic protein, namely Arrestin (Arr), to the activated and phosphorylated receptor. Arr and G protein occupy similar regions at the intracellular site of the receptor. Therefore, Arr binding prevents further G protein coupling, hence signaling is terminated, which is known as *desensitization* [2]. Finally, the receptor together with bound Arr is internalized and the receptor is either recycled back to the membrane or it is directed to the lysosome for *degradation* [3], which causes *down-regulation* [4] as depicted in Figure 1.1. In addition, it has been recently discovered that Arrs can also initiate alternative signaling pathways independently of G proteins [5], which is beyond the scope of this study.

1
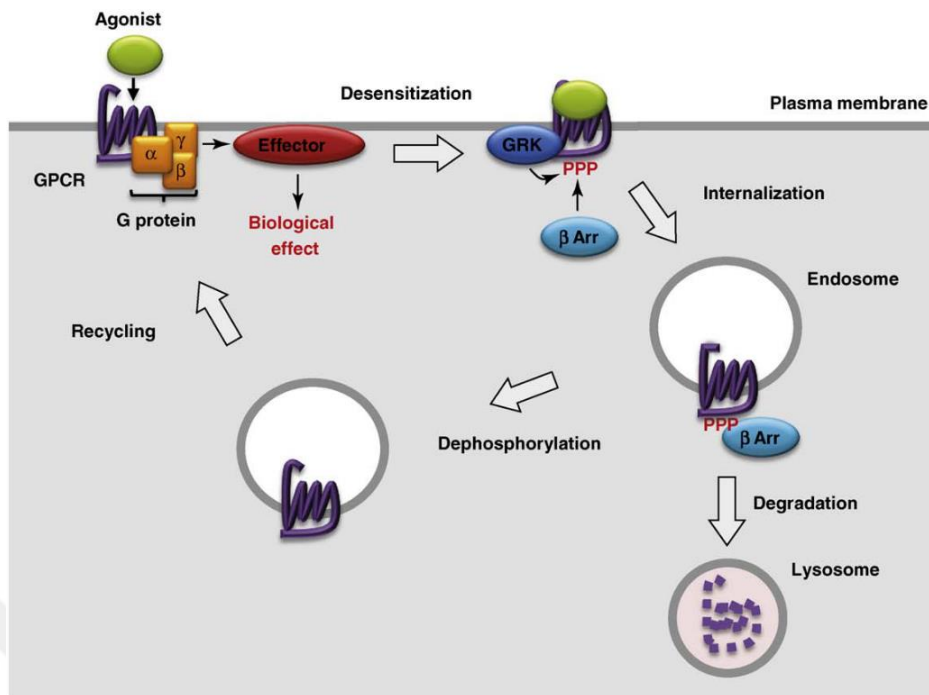
Figure 1.1: *Schematic representation of GPCR-mediated signaling and trafficking. GPCRs transmit information from the extracellular site to the inside of the cell by means of heterotrimeric G protein, which is depicted in orange squares. Activation and subsequent detachment of G protein cause receptor phosphorylation which is done by GPCR kinases (blue oval). The signal is terminated (desensitization) by binding of the activated and phosphorylated receptor to Arrestin, which is represented by a light blue oval. Consequently, the receptor is internalized. Depending on the physiological conditions the receptor is either recycled back to the membrane or directed to the lysosome for degradation (down-regulation). The figure is taken from [6]*

The Arr protein family consists of four members, namely, Arr-1, 2, 3 & 4. Arr-1 and 4 are exclusively expressed in photoreceptors, so they are also known as visual Arrs, whereas Arr-2 (a.k.a β-Arr1) and Arr-3 (a.k.a β-Arr2) are ubiquitous and participate in various physiological pathways [5]. The members in this protein family share a high sequence identity (over 75%) and conserved structural fold. The protein is composed of two β-sheet sandwich domains (N- and C-domains) which are connected by a hinge region (See Figure 1.2). Here, the key regions which are involved in activation of Arrestin such as the "*gate loop*" (residues 291-296), the "*short helix*"(residues 313-317), the "*aromatic core*" (residues 76-245), the "*neighbor of the gate loop*" (residues 276-287) and the "*polar core*" (residues 27, 170, 291, 298, 393) region are also shown.

N-Domain   *C-Domain*
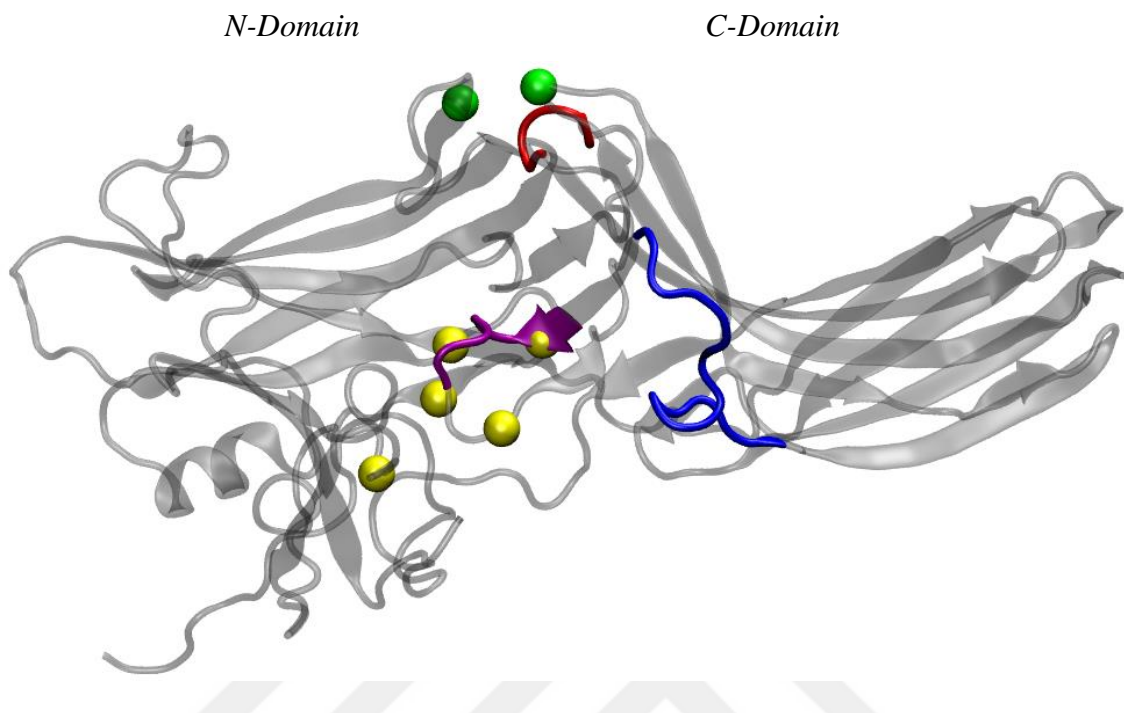
Figure 1.2: *Structural representation of Arr3 protein on the crystal structure of inactive Arr3 (PDB ID: 3P2D). The key regions that are involved in the activation mechanism are indicated by colors: "gate loop", the "short helix", and the "neighbor of the gate loop" are shown in purple, red and blue, respectively. The Cα atoms of the residues that make up the "aromatic-core" and the "polar-core" are represented by green and yellow dots, respectively.*

In spite of having a conserved structural fold, the members display remarkable differences in their requirement of receptor phosphorylation which is needed for activation of Arr as well as receptor binding. Specifically, Arr-1 and Arr-4 can exclusively bind to activated and phosphorylated Rhodopsin to form high-affinity complex [7], [8], whereas Arr-2 and Arr-3 can bind to various types of GPCRs. More interestingly, phosphorylation requirement of Arr-3 depends on the type of the receptor [9]. For instance, Arr-3 cannot bind to non-phosphorylated $\beta_2 AR$ (beta-2-adrenergic receptor) whereas it can bind to M_2R (Muscarinic 2 Receptor) independent of the phosphorylation status of the receptor [9]. In general, binding of Arr to activated and phosphorylated GPCR is thought to proceed via the *"two-step binding"* mechanism [4]. According to that, in the first step,

the phosphorylated C-terminus of the receptor interacts with the phosphate sensor residue, which is located on the N-domain of Arr, that contributes to the stability of the "*polar core*" region. There is an intricate charge balance within that region, therefore any perturbation, for instance, interaction with receptor-attached phosphorus atoms, leads to disruption of the stability of the region. Subsequently, this leads to the release of the C-terminus of Arr (See Figure 1.3.B), which is known as the "*hallmark of Arrestin activation*" [10], [11], [5], [12]. In the second step, structural constraints that stabilize the inactive conformation of Arr are released and the key regions which are required for receptor binding are exposed upon displacement of the C-terminus of Arr. The most remarkable conformational change that occurs upon activation of Arr is the rotation of the C-domain with respect to N-domain by an angle of 17^0 as shown in Figure 1.3.C. Consequently, activated Arr binds to the receptor forming a high-affinity complex [2], [4], [13] as shown in Figure 1.3.A.

A

B

C

C-Domain   N-Domain

17.7^0
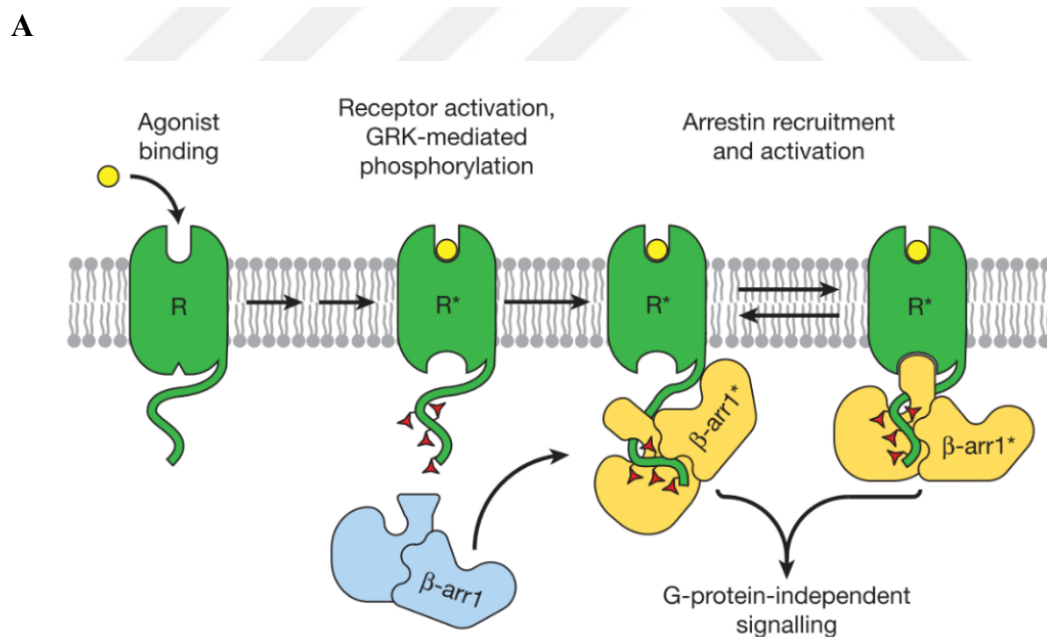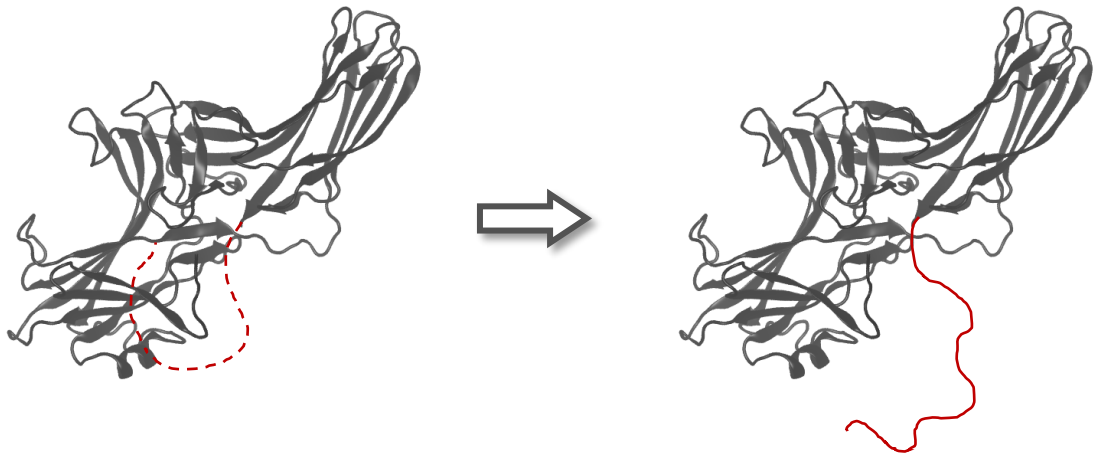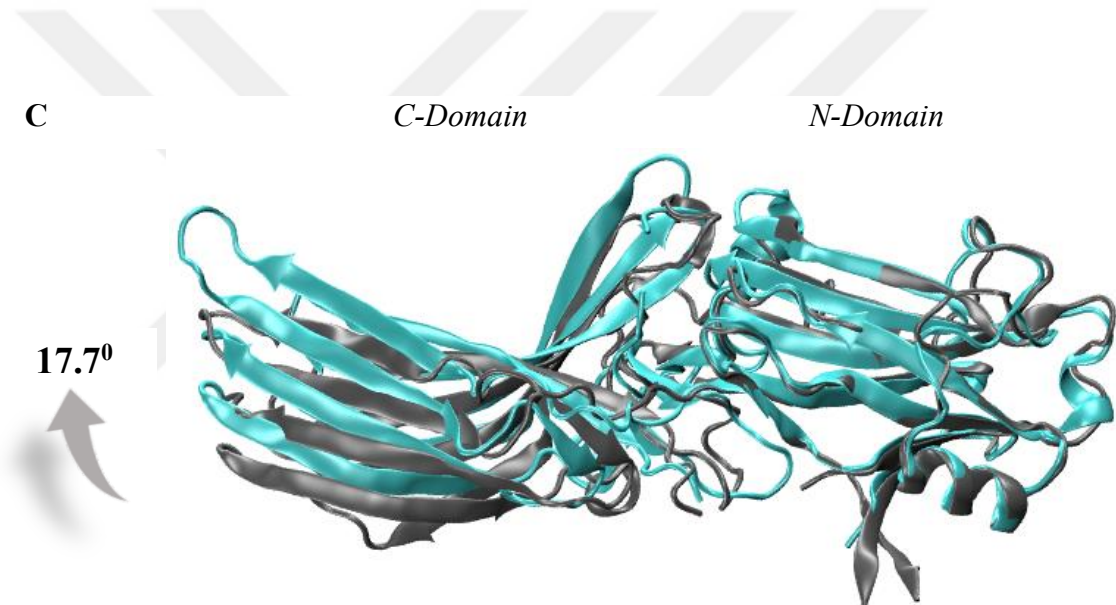
Figure 1.3: *A. The "two-step binding mechanism" which is proposed for receptor binding. The agonist is represented by a yellow circle, whereas the receptor is shown in green color. The receptor-attached phosphorus atoms are represented by red triangles. The inactive Arrestin is shown in blue color, whereas active Arrestin is shown in yellow color. In the first step, the phosphorylated C-terminus of the receptor binds to the N-domain of Arrestin which leads to release of the structural restraints that stabilize the inactive conformation of the protein. In the second step, Arrestin binds to the receptor by means of these key regions and a high-affinity complex is formed. The figure is taken from [13]. B. Depiction of the "hallmark of Arrestin activation". The C-terminus of Arrestin is shown in continuous and dashed red lines in inactive and active Arr, respectively. C. Representation of the rotation of the C-domain with respect to the N-domain upon Arr activation. The crystal structure of active (PDB ID:5TV1, cyan) and inactive Arr-3 (PDB ID:3P2D, gray) are aligned with respect to their N-domains*

In a recent experimental study, where receptor-induced conformational changes were investigated in non-visual Arrs, it was shown that the C-terminus of Arr-3 can sample short distances from the body of the protein upon receptor binding suggesting that Arr-3 can bind to the GPCR independent of C-terminal displacement which might be provided by an alternative activation mechanism that is different from the "*two-step binding mechanism*" [4], [14], [15], [16]. In accordance with that, it has been shown in a computational study that Arr-3 can adopt activation-related structural rearrangements independent of the perturbation of the "*polar core*", but rather using its intrinsic flexibility [17]. Here, it is important to point out that the rotation angle is measured as 7.5^0 in the crystal structure of the "*polar core mutant*", which can bind to non-phosphorylated and activated receptor Rhodopsin [18]. This shows that pre-activated Arr needs smaller domain rotation angle to bind to the non-phosphorylated receptor.

Considering the fact that receptor phosphorylation -in general- is required for Arr binding, any problem occurs in this step causes various crucial diseases. For instance, any mutation that occurs in one of the phosphorylation sites of Rhodopsin prevents binding of Arr-1 to the receptor. Consequently, signaling cascade cannot be terminated and this causes the death of rod photoreceptor cells as seen in *retinitis pigmentosa* (night blindness) [19], [20], [21], [22]. In another example, a mutation (R137H) that occurs in the transmembrane region of Vasopressin receptor causes activation of the receptor without agonist binding which leads to excessive phosphorylation and down-regulation of the receptor as seen in *nephrogenic diabetes insipidus* [23]. Similarly, high concentration of an endogenous ligand like catecholamine increases the expression of G protein kinase-2 in the cell which leads to excessive phosphorylation and down-regulation of beta 2 adrenergic receptor as seen in "*congestive heart failure (CHF)*" [24], [25], [26], [27]. It has been shown that phosphorylation-independent Arr can be used to prevent down-regulation of the receptor since the interaction between the receptor and Arr will be weaker in the absence of phosphorus atoms, thus shortening the time required for recycling of the receptor back to the membrane [28], [29], [30], [31], [32].

In this thesis project, we aim to activate Arr-3 *in silico* by means of small molecules, independent of phosphorylation. To do so, as a first step, we performed molecular dynamics simulations of Arr-3 in the absence of ligand to pick up representative structures to which small molecules will be docked. Subsequently, possible binding pockets on Arr-

3 are determined and pharmacophore groups are created accordingly. Afterwards, small molecule candidates are searched in the *ZINC* compound database and docked to Arr-3. Successful candidates that have relatively more negative binding energy and acceptable molecular descriptor properties are tested for their stability and impact on dynamics of Arr-3 by using molecular dynamics simulations. Our results show that the active-like rotation angle can be achieved with small molecules that stably bind to the "*short helix*" region of Arr-3. Moreover, it is also observed that ligand binding increases correlation within both N- and C-domain which might trigger domain rotation.

Chapter 2

Methods

In this project, to test our working hypothesis, which is *"Arr-3 can be activated by means of small molecules independent of phosphorylation",* a variety of computational tools such as molecular docking, classical and accelerated molecular dynamics simulations were used. The workflow which is shown in Figure 2.1 was followed in the study. Specifically, we started with performing molecular dynamics simulations on Arr-3 in the absence of the ligand to investigate dynamics of the protein alone which will be used as a reference throughout the study. In the second step, we picked up representative structures from Arr-3 trajectory and identified possible binding pockets on these structures by using *SiteMap* module of Schrodinger software [33]. In the third step, we determined corresponding pharmacophore groups that fit the binding pockets in terms of both chemistry and geometry. In the fourth step, we searched for possible candidate molecules in *ZINC* compound database and docked the candidate molecules to the binding pockets. Afterwards, we calculated binding free energies and investigated molecular descriptors to check if the candidates are physiologically relevant. Finally, the stability of the successful candidates and their impact on the dynamics of Arr-3 were tested using both classical and accelerated molecular dynamics simulations.
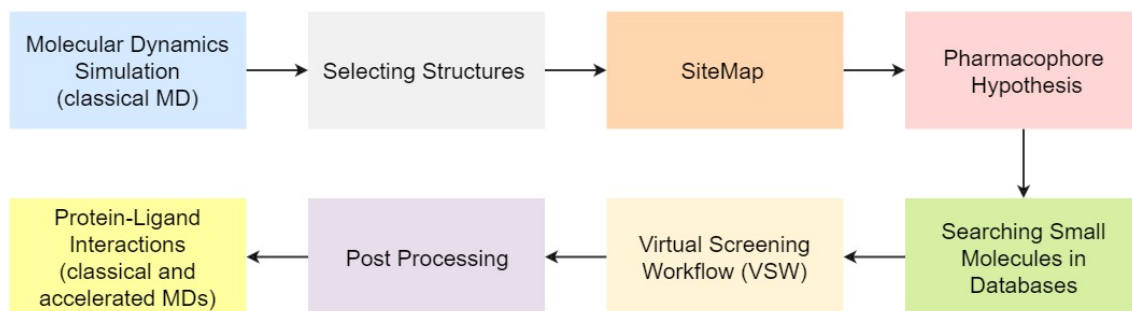
Figure 2.1: *Workflow of the method*

2.1 Fully Atomistic Molecular Dynamics Simulations

Molecular dynamics (MD) simulations allow us to investigate time-dependent structural and dynamical properties of systems at the atomistic level, thus providing a mechanistic insight into the macroscopic behavior of the system studied. MD simulations act as a bridge between the theory and experiments providing that the *"ergodicity"* is satisfied, which states that the *"time average"* properties resemble the *"ensemble average"* properties of the system. In this project, we performed MD simulations on both Arr-3 and Arr-3/ligand complexes to investigate the impact of the ligands on the dynamics of the protein in a time-dependent manner. We used the crystal structure of inactive Arr-3 (PDB ID: 3P2D) [34] as the starting conformation in systems that do not include any ligand. Systems were prepared using the CHARMM-GUI server [35]. The protein was modeled using the CHARMM36 force field (Chemistry at Harvard Macromolecular Mechanics force field) [36], whereas water molecules were modeled using the TIP3P model [37]. Simulations were performed using the *NPT* ensemble, where the pressure and the temperature were set to 1 atm and 310 K, respectively. The cut off value used to calculate non-bonded interactions was set to 12 Å and long-range electrostatic interactions were computed using particle mesh Ewald method [38]. Protonation states of ionizable amino acids were calculated using PropKa server [39] at pH 7.4 and systems were neutralized with NaCI. Simulations were run for 1µs using GROMACS package [40] on the high-performance computing cluster system of Istanbul Medipol University. We used two replicas for Arr-3 system, each of which started with a different velocity. Finally, trajectories were analyzed in terms of both local and global properties of the system. Specifically, we calculated probability distributions for certain atom pair distances, root-

mean-square deviation of a set of residues and made principal component (PCA) and dynamic cross correlation analysis (DCCM) whose details are given below.

2.1.1 Probability Distributions of Atom-Pair (Cα-Cα) Distance

We calculated probability distributions of distances between certain residue-pairs by using their corresponding Cα atoms. Among them are, the *"aromatic core"* (residues 76 & 245), the *"short helix" (*residues 313 & 317) and the *"gate loop"* (residues 291 & 296) which have been shown to be involved in the activation mechanism of the protein. To do so, we first calculated timeline distance plots by using *'gmx distance'* module of GROMACS Software [40] and then constructed corresponding probability distributions. In addition, we constructed similar probability distributions for the crystal structures of inactive (PDB ID:3P2D) [34] and active Arr-3 (PDB ID:5TV1) [41] which were used as references to group Arr-3 trajectories into different conformational states like *inactive, intermediate* and *active*.

2.1.2 Rotation Analysis

As mentioned above the most remarkable conformational change occurs upon activation is the rotation of the C-domain with respect to the N-domain. Therefore, we calculated rotation angles for trajectories of both Arr-3 and Arr-3/ligand complexes to investigate the activation state of the systems. We used available crystal structures of inactive and active Arr-3 to build up the pseudo-trajectory and calculated the first principal component which was defined as the rotation axis. The rotation angle which is required to bind to the activated and phosphorylated receptor is measured as 18^0 [5], [7], [42], [41]. However, smaller rotation angle (minimum of 7.5^0) is required to bind to the activated but non-phosphorylated receptor, as seen in the crystal structure of the *"polar core mutant"* and which is targeted to achieve in this study [18].

2.1.3 RMSD (root-mean-square deviation) of Atomic Position

RMSD is defined as the measure of the average distance between the atoms of superimposed structures and given by the following equation:

$$RMSD_\alpha(t_j) = \sqrt{\frac{\sum_{\alpha=1}^{N_\alpha}\left(\vec{r}_\alpha(t_j) - \langle\vec{r}_\alpha\rangle\right)^2}{N_\alpha}} \qquad (2.1.3a)$$

$$\langle\vec{r}_\alpha\rangle = \frac{1}{N_t}\sum_{j=1}^{N_t}\vec{r}_\alpha(t_j) \qquad (2.1.3b)$$

where N_α is the number of atoms, $\vec{r}_\alpha(t_j)$ is the position of the atom $\propto$ at a time t_j, N_t is the number of time steps and $\langle\vec{r}_\alpha\rangle$ is the average value of the position of atom α with respect to the position of $\vec{r}_\alpha(t_j)$. RMSD can be used to calculate either the difference between two structures or time evolution of the structural changes of a given molecule during the trajectory. In this study, we did both types of RMSD calculation using the *"RMSD Trajectory Tool"* of VMD [43].

2.1.4 RMSF (root-mean-square-fluctuation)

RMSF is the deviation between the position of particle i and some reference position, which corresponds to -in general- time-average position of the same particle, and given by the following equation:

$$RMSF = \sqrt{\left(\left(\frac{1}{N}\right)\sum_{n=1}^{N}(X_i(n) - \bar{X}_i)\right)} \qquad (2.1.4)$$

where N is the duration of the simulation. $X_i(n)$ is the coordinate of the backbone atom X_i at time n. RMSF was calculated by using *'gmx rmsf'* module of GROMACS [44] using backbone atoms of the protein.

2.2   Principal Component Analysis (PCA)

Principal component analysis is a method to reduce multidimensional set of variables to a small set that represents the dominant behavior in the system. In our study, we applied PCA to investigate overall dominant collective motions of the protein in the presence of the ligands. To do so, we first calculated the first and the second eigenvectors of Arr-3 system using Cα atoms of the protein. Subsequently, Arr-3/ligand trajectories were projected onto these eigenvectors to comparatively investigate the effect of the ligand on the dynamics of Arr-3. PCA was made by using *'gmx covar'* and *'gmx anaeig'* modules of Gromacs software [44] using the following equations:

$$C_{ij} = \langle M_{ij}\Delta r_i \Delta r_j \rangle \qquad (2.7.2a)$$

$$C_v = \delta^2 v \qquad (2.7.2b)$$

where C_{ij} corresponds to the covariance matrix. Changing of position in time average along with all coordinates equal to C_{ij}. Respectively δ^2 and v represents the diagonalization of covariance matrices which are obtained by eigenvalues and eigenvectors.

2.3 DCCMs (dynamic cross correlation maps)

Dynamic Cross Correlation Maps are used to determine the cross-correlations of atomic displacements. To assess the extent of the correlation of atomic fluctuations within a system the magnitude of all pairwise cross-correlation coefficients, C_{ij}, can be examined [45]. When C_{ij} equals to 1, -1 and 0, the fluctuations of atoms i and j are completely correlated, completely anticorrelated, and are not correlated, respectively. In our study, to investigate the impact of the ligands on the communication profile of amino acid residue pairs we created DCCM maps of both Arr-3 and Arr-3/ligand systems using the "*bio3d*" library of R package [46].

2.4 Enhanced Sampling Method

2.4.1 Accelerated Molecular Dynamics (aMD)

Accelerated molecular dynamics (aMD) [47] is one of the enhanced-sampling methods that is used to increase conformational sampling in the system by decreasing energy barriers that separate local minima found on the potential energy surface. According to the method, energy wells that fall below a certain threshold level are raised by adding external potential to the system whereas those that are above the threshold level are not changed (See equation below).

$$E_{dihed} = V_{dihed\_avg} + \left(\Lambda \times V_{dihed\_avg} \right) \qquad (2.8.1a)$$

$$\alpha_{dihed} = \Lambda \times V_{dihed\_avg} \div 5 \qquad (2.8.1b)$$

$$E_{total} = V_{total\_avg} + \left(0.2 \times N_{atoms} \right) \qquad (2.8.1c)$$

$$\alpha_{total} = 0.2 \times N_{atoms} \qquad (2.8.1d)$$

where V_{total_avg} and V_{dihed_avg} are the average of the total and dihedral energies, N_{atoms} is the total number of atoms, Λ is an adjustable acceleration parameter [48]. In this way, barriers that separate adjacent conformational states are reduced thus allowing the system to sample regions on the conformational space, which are otherwise not accessible by classical MD simulation. In our study, to further test the stability of the ligands we performed aMD simulations on Arr-3/ligand systems. V_{dihed_avg} and V_{total_avg} values were obtained from classical MD trajectories of the systems and corresponding parameters were included in NAMD [49] configuration file. Here, it is important to point out that acceleration of the system dynamics must not distort the structure of the target system and can be controlled by properly adjusting the Λ parameter.

2.5   Binding Site Identification using Sitemap

Once Arr-3 trajectories are classified with respect to their conformational states we picked up 9 representative structures from the trajectories and identified possible binding sites, to which small molecules will be docked, on these structures using the *SiteMap* tool of Schrodinger software [50], which evaluates the pocket candidates using the following parameters [33] such as draggability score (Dscore), number of site points, exposure, enclosure, hydrophobic/hydrophilic and donor/acceptor character of the pocket and volume, all of which are included in the *SiteScore*. In general, it is suggested that *SiteScore* values, which are composed of high draggability score, enclosure, volume, and smaller exposure values, should be considered. It is important to emphasize that binding pockets with relatively higher *SiteScore* values were shown to cluster around the *"short helix"* which is one of the key regions that is involved in the activation mechanism of Arr-3.

2.6   Pharmacophore Hypothesis

The pharmacophore model is a geometrical description of the chemical functionalities such as capability of acting as hydrogen bond donors/acceptors, hydrophobicity,

aromaticity of the target region, etc. [51], [52]. In general, the pharmacophore model can be created in different ways depending on the availability of the structural information of both the target and the ligand. In our case, we built up the pharmacophore model using "receptor cavity" and E-pharmacophore options implemented under the *"Phase"* module of Schrödinger software [53] since the structure of the target protein is known. We used the "centroid of residue option" to explicitly specify the residues that will be considered in the model. For instance, to build up corresponding pharmacophore model for the "*short helix*" region we picked up the coordinates of the residues that are located in that region.

2.7   Searching Molecules in Databases

Using the pharmacophore models created in the previous step we did candidate molecule search in *ZINC* compound Database via ZincPharmer [54]. Here, we used the "number of rotatable bond option" as the filter to filter out molecules that have eight or more rotatable bonds.

2.8   Docking and Virtual Screening

Molecular docking is a method which is used to predict the preferred orientation of a given molecule within the target site [55]. We used this method to eliminate candidate molecules that cannot favorably bind to the target site which was assessed by comparing the relative binding energies of the molecules. To do so, we utilized the *Glide* docking tool of Schrodinger software [56]. Before the docking step, the ligands were prepared using the *LipPrep* tool of Schrodinger software [57] in accordance with their chemical properties. The ionization states of the residues were determined at pH 7.4 using PropKa [58], and alternative tautomeric forms -if any- were also prepared. In the second step, candidate molecules were docked to the target site using SP docking algorithm [56]. During molecular docking, the ligand together with the side chains of the residues, which were in the grid box, were kept flexible. We used the same coordinates, which were used

in the pharmacophore generation step, to create the grid box for ensuring consistency. The size of the grid box was set to 20 Å and 40 Å for the inner and the outer box, respectively as shown in Figure 2.2. In addition, we also used another docking algorithm, namely Autodock-Vina [59], to check the reliability and reproducibility of the best docking poses obtained from Glide.
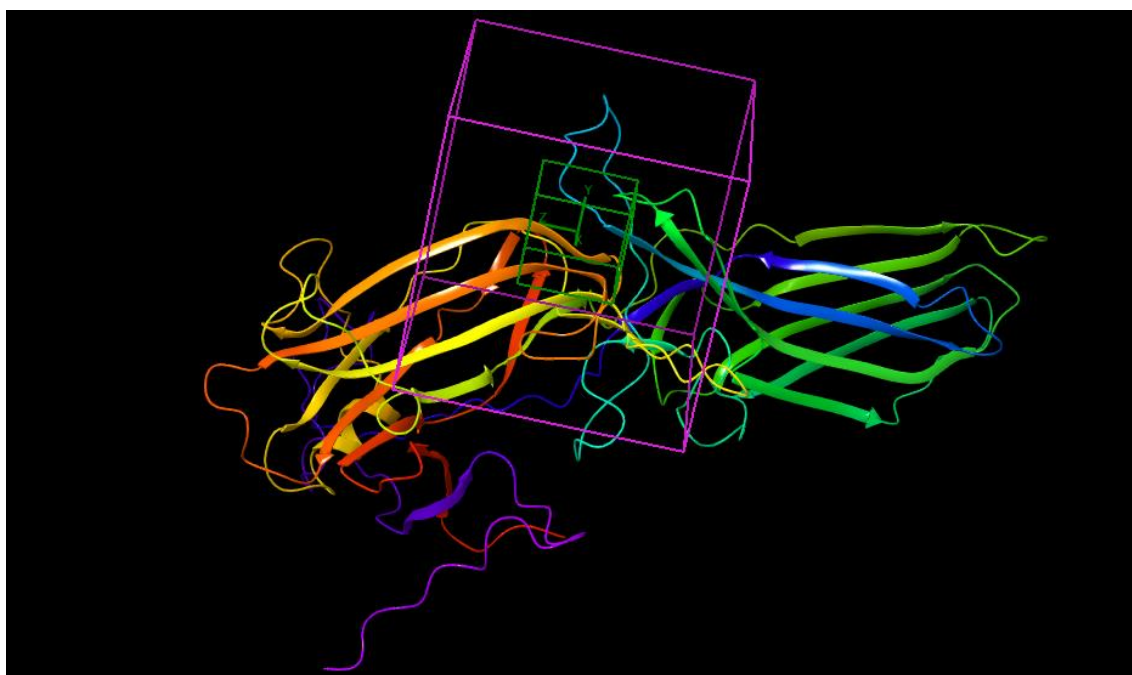

Figure 2.2: *The grid box, which is used to dock candidate molecules to the "short-helix" region, is presented. The inner box, which is used to center the ligand, is shown in green, whereas the outer box, which is used to set the maximum volume to which the candidate molecules will be docked, is shown in magenta.*

2.9 Assessment of the Biological Relevance of the Candidate Molecules having Favorable Binding Energy

In this step, we further assessed the biological relevance of the candidate molecules, which have relatively more negative binding energy scores (*glide_gscore*). For this purpose, we calculated the following parameters using the *Molecular Descriptor* module of Schrodinger software [60]: predisposition to metabolite *(#metab)*, oral absorption value

(HumanOralAbsorption), Lipinski rule of five *(RuleOfFive)*, molecular weight *(mol_MW)*, affinity to HERG channel *(QPlogHERG)*, and prediction of binding to human serum albumin *(QplogKhsa)*. In addition to *glide_score* we also calculated *glide_ligand_efficiency*, which is given by *glide_score/#of heavy atoms*. Here, this value gives more reliable results, in particular, for bigger molecules since the binding score is normalized by the number of heavy atoms. Finally, we also checked the similarity of the candidates by using *'Canvas Similarity and Clustering'* module using *Tanimoto* as the similarity metric, and obtained 29 candidate molecules to be used in the following step.

2.10 Testing Stability of Candidates and Investigation of Their Impact on Dynamics of Arrestin-3

The stability and the impact of 29 selected candidate molecules were tested by using molecular dynamics simulations. We used the same parameters that were given in detail for Arr-3 system *(see: 2.1 Fully Atomistic Molecular Dynamics Simulations)*. Similar to the Arr-3 system, we performed at least two replicas for each ligand, each of which started with a different velocity. The systems, in which the ligand stably bound to the target region during the course of 1 µs simulation, were further analyzed to investigate the impact of the ligand on the dynamics of Arr-3. We made similar analyses, which were given in detail above, for also Arr-3/ligand complexes.

Chapter 3

Results

It has been shown that Arr-3 displays more structural flexibility compared to other members, and it can transiently adopt active-like conformations even in the absence of any perturbation like the disruption of the *"polar core"* by phosphorylated C-terminus of the receptor [34], [17]. In this study, we aim to target and stabilize these transient active-like structures by means of small molecules, thus shifting the conformational ensemble towards the active-like state. Twenty-nine candidate molecules that had favorable binding scores and acceptable molecular descriptor values were tested for their stability and impact on dynamics of Arr-3. 9 out of 29 molecules were stably bound the the target site and could shift the ensemble towards the active-like state. Here, we presented 4 of these successful candidates in this thesis project.

3.1   Grouping of Trajectories into Specific Conformational States

As the first step in this study, we grouped Arr-3 trajectories into different conformational states like *inactive*, *intermediate* and *active* using reference values of Cα-Cα atom distances of a certain set of residue pairs which are involved in the activation mechanism of Arr-3, as shown in Table 3.1.1.

Table 3.1.1: *Distance values (Å) measured between Cα-Cα atoms of a certain set of residue pairs, which are involved in the activation mechanism, using inactive (PDB ID: 3P2D) and active crystal (PDB ID: 5TV1) structures of Arr-3. Green color represents the active whereas red color represents the inactive crystal structure.*

	Inactive (Å)	Active (Å)
Aromatic Core	7.2	9.7
Short Helix	7.8	9.5
Gate Loop	10.1	13.9

As shown in Table 3.1.2, Arr-3 more resembles the *intermediate* state for the *"aromatic core"* and the *"gate loop"* whereas it resembles the inactive state for the *"short helix"*. We did not consider the *"active state"* both for the *"short helix"* and the *"gate loop"* regions as this state was barely sampled as given in Table 3.1.1. Some of the combinations found in the selected structures were as follows: the *"aromatic core"* (inactive)/the *"short helix"* (inactive)/the *"gate loop"* (inactive), the "aromatic core" (intermediate)/the *"short helix"* (inactive)/the *"gate loop"* (inactive), the *"aromatic core"* (intermediate)/the *"short helix"* (intermediate)/the *"gate loop"* (intermediate), etc.

Table 3.1.2: *The percentage values of the inactive, intermediate and active conformational states calculated for Arr-3 trajectory using reference distance values given in Table 3.1.1*

	Inactive	Intermediate	Active
Aromatic Core	**10.3%** < 7.2	7.2 < **78.8%** < 9.7	**10.9%** > 9.7
Short Helix	**78.3%** < 7.8	7.8 < **18.9%** < 9.5	**2.8%** > 9.5
Gate Loop	**27.7%** < 10.1	10.1 < **72.3%** < 13.9	**0.01%** > 13.9

Using these representative structures, we determined possible binding pocket candidates on Arr-3 and picked up those having relatively higher *site_score* values. Below, we

provided the top ten binding pocket candidates obtained for a conformation in which the *"aromatic core"*, the *"short helix"* and the *"gate loop"* all adopted the intermediate state. Here, it is important to point out that the binding pocket with the highest *site_score* in Table 3.1.3 (site_1) corresponds to the *"short-helix"* region, which is involved in the activation mechanism of Arr-3 as shown in Figure 1.2. Similar results were obtained for other selected structures as well. We also showed a corresponding pharmacophore model built up for the *"short helix"* region along with the docked ligand in Figure 3.1.1, Figure 3.1.2 and Figure 3.1.3

Table 3.1.3: *The top ten binding pocket candidates which are obtained by SITEMAP tool. The individual terms, which are used to calculate the final Site_Score values, are also shown.*

|	Site Score	size	Dscore	volume	exposure	enclosure	phobic	philic
site_1	1.032	146	0.924	282.970	0.484	0.746	0.156	1.420
site_2	0.976	169	0.986	473.680	0.654	0.662	0.395	1.076
site_3	1.004	139	0.920	449.670	0.621	0.704	0.200	1.356
site_4	0.961	107	0.946	227.070	0.587	0.639	0.198	1.157
site_5	0.898	84	0.899	263.770	0.729	0.637	0.332	1.056
site_6	0.858	63	0.850	150.920	0.611	0.682	0.573	0.938
site_7	0.724	59	0.693	137.540	0.728	0.539	0.094	1.080
site_8	0.712	38	0.624	71.690	0.591	0.688	0.351	1.130
site_9	0.908	43	0.541	83.350	0.482	0.937	0.020	1.956
site_10	0.656	41	0.528	99.130	0.721	0.578	0.000	1.290
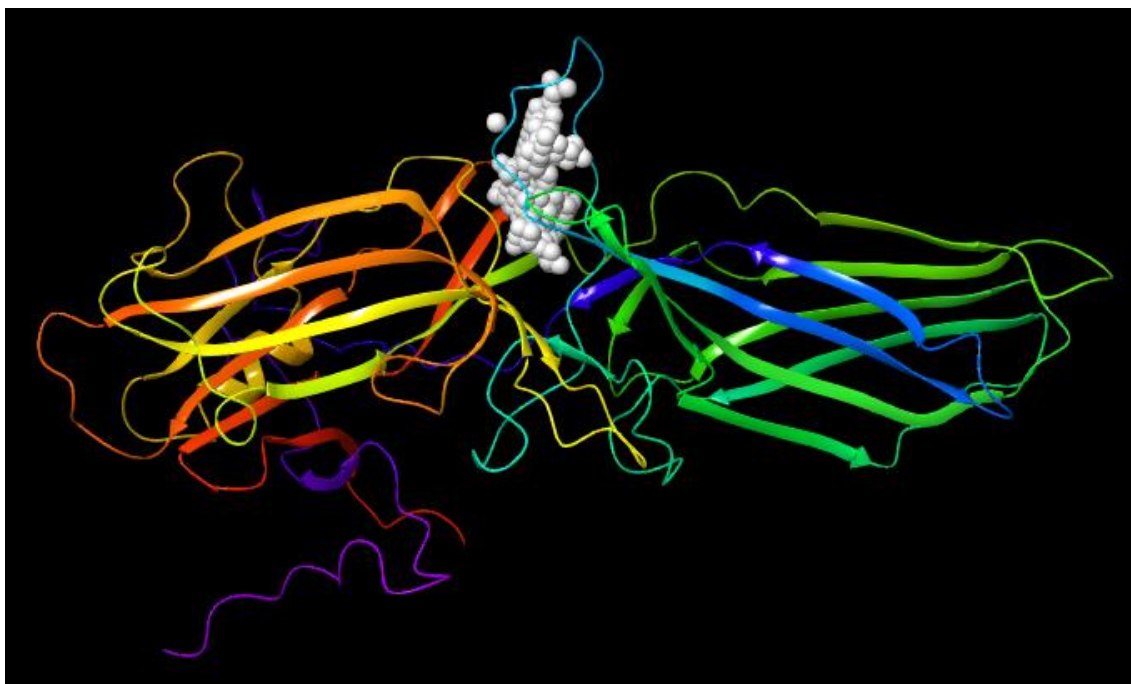
Figure 3.1.1: *Shows an example to one of the results given by the SiteMap. Sites (white beads) collected near the Short Helix.*

Figure 3.1.2: *The corresponding pharmacophores model that fits the region given in Figure 3.1.1.*

Figure 3.1.3: *The depiction of one of the Arr3-ligand complexes with high binding energy scores*

As mentioned in the Methods section we performed docking by using two different programs to check the reliability and reproducibility of the docking poses and obtained similar results as shown in Figure 3.1.4.
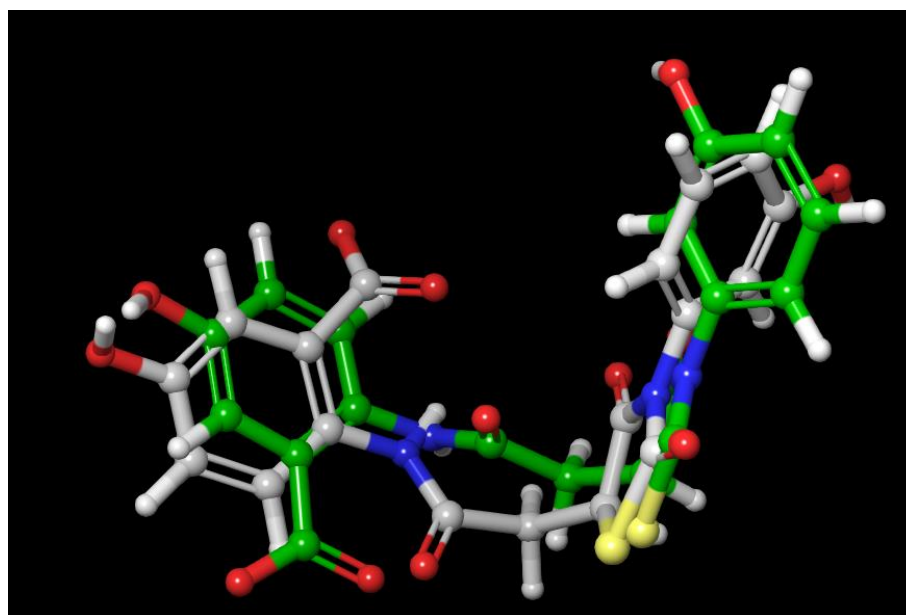

Figure 3.1.4: *The comparison of the best binding poses obtained from Autodock Vina (green) and Glide-SP (gray)*

Table 3.1.4: *Molecular descriptor values which were used to assess the biological relevance of the candidates that have favorable binding score are shown along with their limit values. Scores were obtained using Molecular Descriptor – QikProp module of Schrodinger.*

|	Ref. Values	Ligand-1	Ligand-2	Ligand-3	Ligand-4
HumanOral Absorption	(low to high) 1-2-3	3	2	3	2
RuleOfFive	max. 4	0	0	0	0
mol_MW	130.0 – 725.0	365.345	367.378	360.857	402.378
QplogHERG	concern below −5	-4.873	-3.680	-6.19	-3.822
QplogBB	−3.0 – 1.2	-1.638	-1.748	-0.425	-2.576
#rtvFG	0 – 2	1	0	1	0
#metab	1 – 8	1	4	2	4
glide_ligand _efficiency	n/a	-0.301	-0.316	-0.357	-0.351
glide_gscore	n/a	-8.914	-8.224	-8.575	-9.838

Below, we provide the 2D ligand interaction maps that show key molecular interactions between the target region and successful ligands (See Figure 3.1.5). The snapshots were taken from the beginning of the trajectories, but these interactions were stable throughout molecular dynamics simulations. As seen from Figure 3.1.5, Ligand-1 and Ligand-4 interact dominantly via electrostatic interactions with the target *"short helix"*, whereas Ligand-2 and Ligand-3 interact via hydrophobic interactions. We also determined the degree of similarity between successful ligands to check if we had diversity among them. The molecules were structurally distant from each other as evident from the dendrogram shown in Figure 3.1.6.
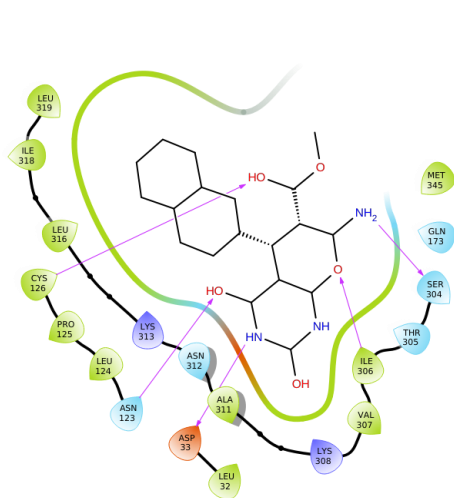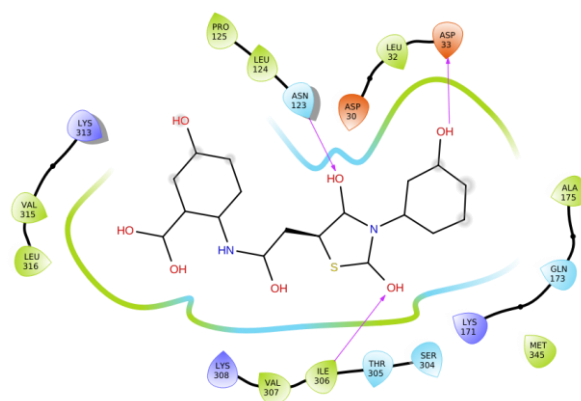
Ligand-1
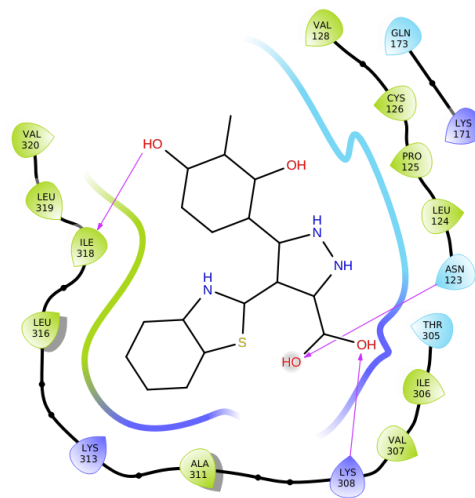
Ligand-2
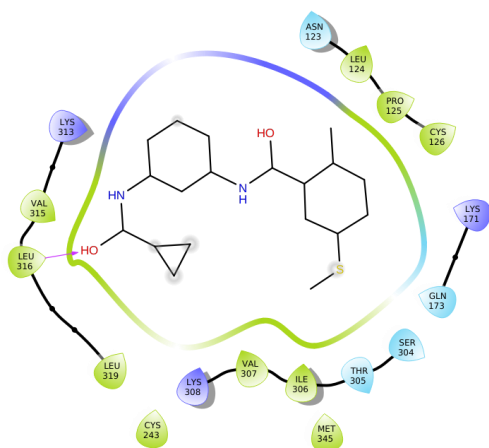
Ligand-3

Ligand-4

Figure 3.1.5: *2D ligand interaction maps that show key interactions formed between the ligands and Arr-3.*
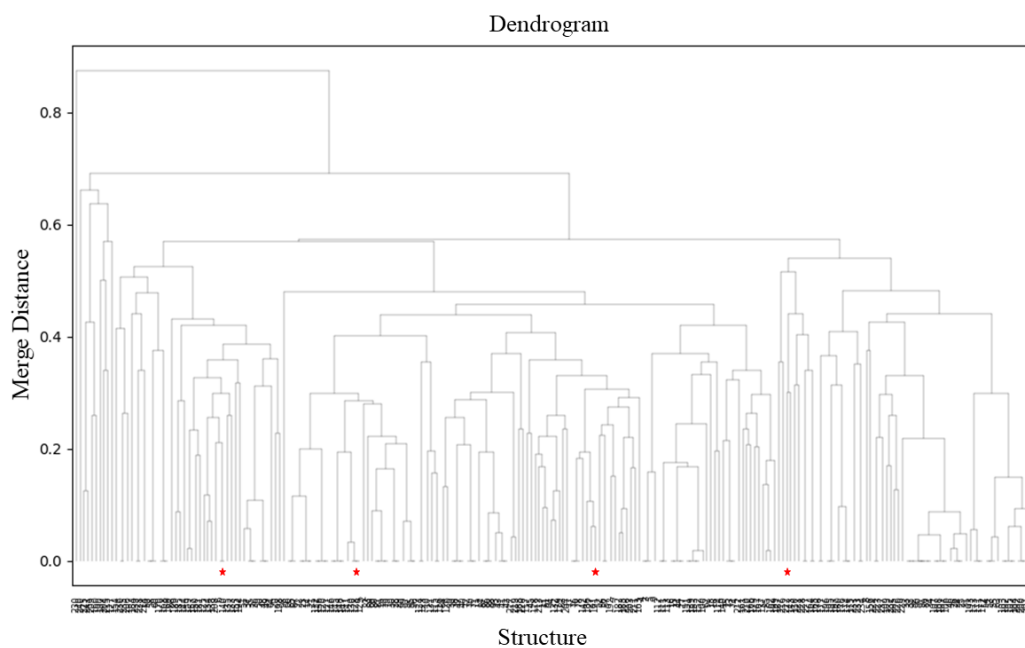
Figure 3.1.6: *The dendrogram shows the structural similarity between ligands. The successful ligands are indicated by red stars at the bottom of the plot.*

3.2   Testing the Stability and the Impact of Candidate Molecules on Dynamics of Arr-3

The candidate molecules that have favorable binding energy and acceptable molecular descriptor values were tested for their stability in long molecular dynamics runs. The results showed that 9 out of 29 molecules were stably bound to the target *"short helix"* region during the 1µ simulation. 4 of them are shown here as shown in Figure 3.2.1.

Figure 3.2.1: *The timeline plot that shows the distance change between the center of mass of the ligand and the "short helix" obtained from the trajectories of Arr-3/ligand.*
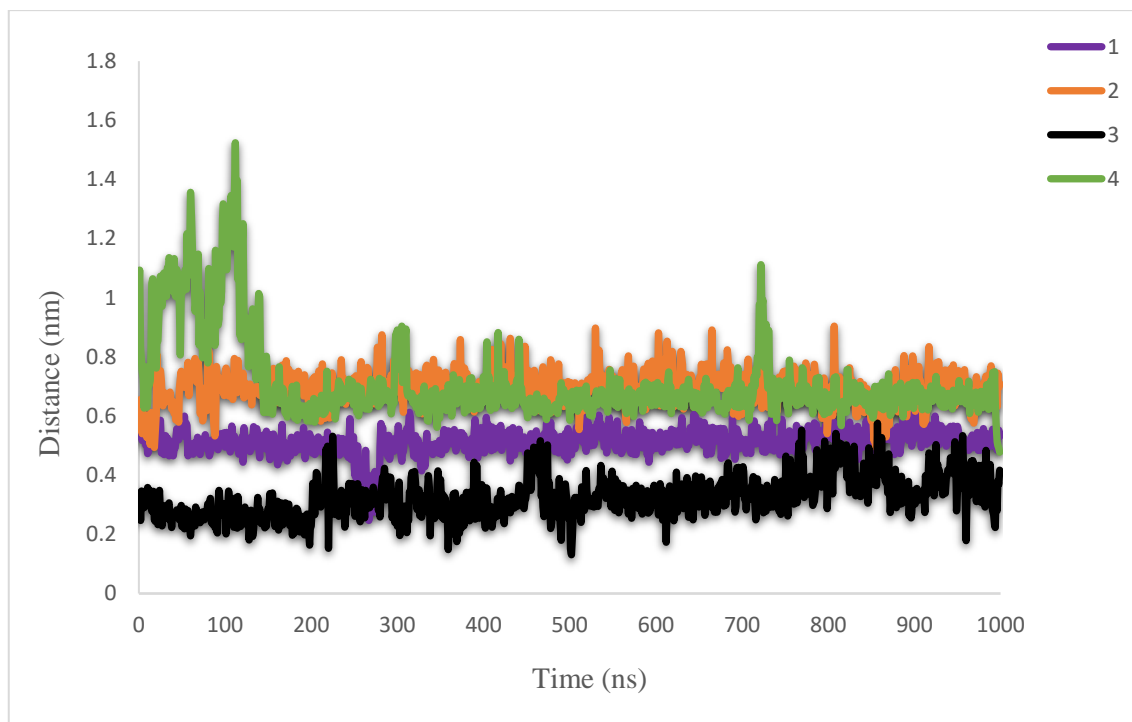
We further analyzed trajectories of the successful ligands, to investigate the impact of the molecules on the dynamics of Arr-3. As mentioned above, the C-domain of wild type Arr needs to rotate around 17^0 with respect to the N-domain to bind to the activated-phosphorylated receptor. On the other, the *"polar core"* Arr mutant, which does not require receptor phosphorylation to get activated, can bind to the activated and non-phosphorylated receptor with a rotation angle of 7.5^0. As shown in Figure 3.2.2, the ligands could stabilize the active-like conformation of Arr-3 as evident by higher rotation angles adapted compared to that in the absence of any ligand (Compare blue and the other colors, Figure 3.2.2).
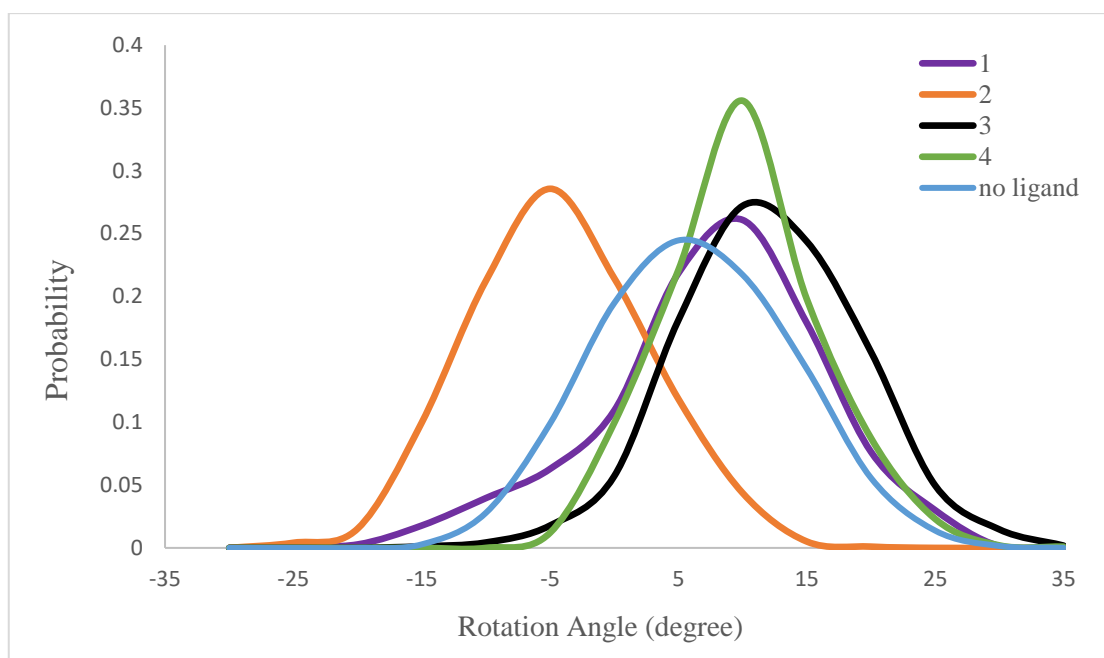
Figure 3.2.2: *The probability distribution of the rotation angle between N- and C-domains. Numbers correspond to different ligands.*

To further investigate the impact of the successful ligands on dynamics of key regions that are involved in the activation mechanism of the protein, such the "*aromatic core*", the "*short helix*", and the "*gate loop*" we created 2D plots that show the relationship between these regions. We observed that the flexibility of the "*short helix*" region decreases for all of the ligands tested, whereas the change in the flexibility of the "*gate loop*" depends on the ligand as shown in Figure 3.2.3.A. Specifically, the flexibility of the "*gate loop*" region decreases in the presence of both Ligand-1 and Ligand-4, whereas it was the same for the Ligand-3, but higher for the Ligand-2 as shown in Figure 3.2.3.A. Moreover, we also looked at the relation between the "*aromatic core*" and the "*short helix*" both of which are shown to be involved in the activation mechanism of the protein. We observed that the state (*exposed vs. shielded*) of the "*aromatic core*" was similar for all of the ligands -except Ligand-1, -as shown in Figure 3.2.3.B. It was in the shielded state in the presence of the Ligand-2, 3 and 4 which was evident by shorter distances measured between the Cα atoms of the two residues that make up the "*aromatic core*".

A

B
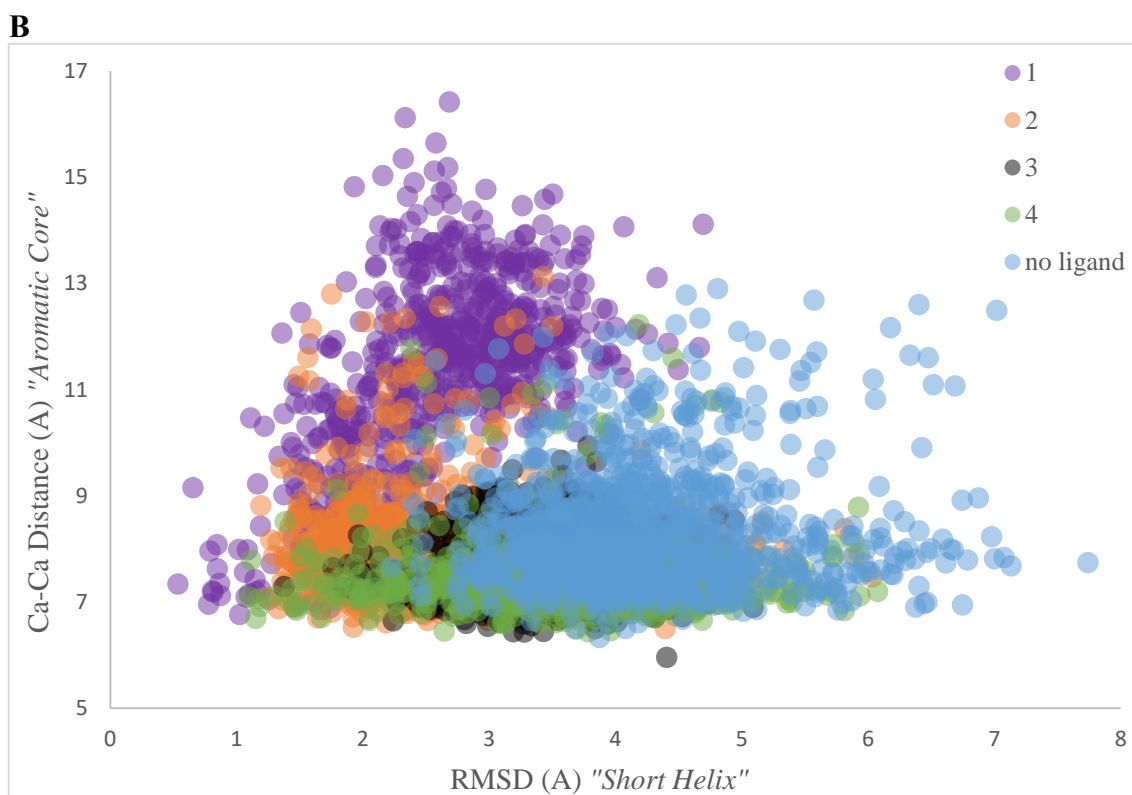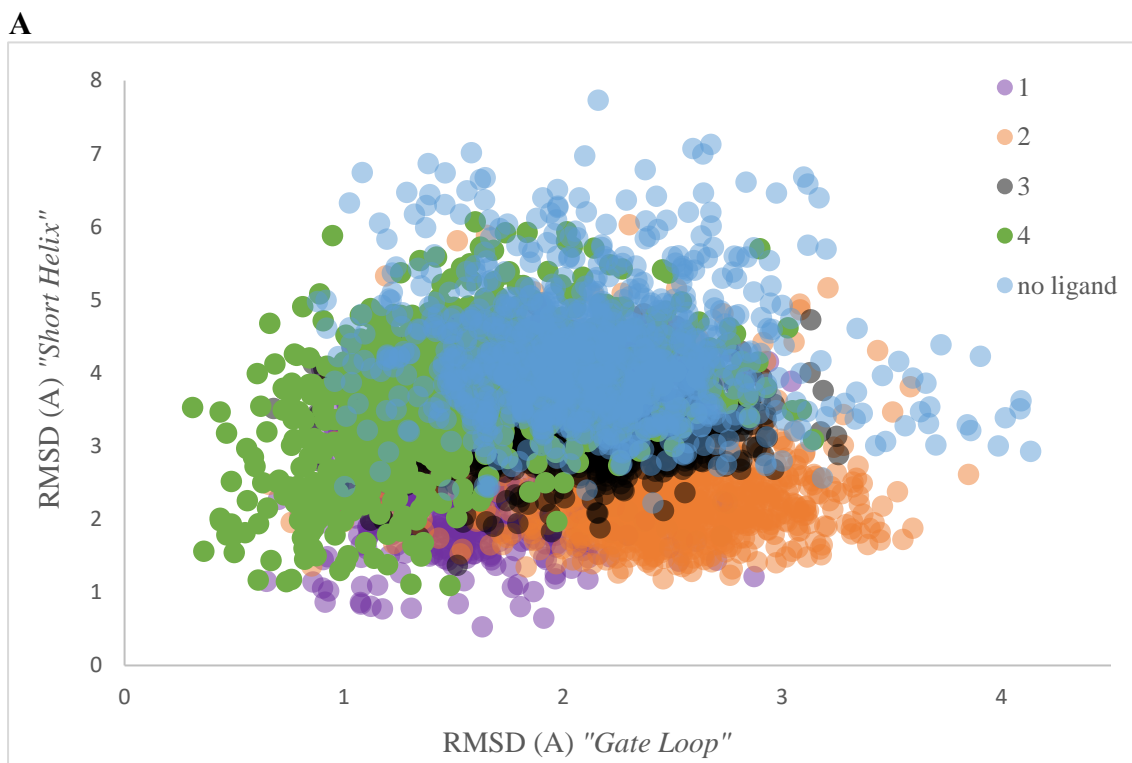
Figure 3.2.3: *A. RMSD values of the "gate loop" versus the "short helix". B. RMSD value of the "short helix" versus Cα atoms distance of the residue pair that make up the "aromatic core"*

We also created dynamics cross-correlation maps to investigate the impact of the ligands on the correlation patterns of residue pairs. We observed that the residue pairs, which are located in the N-domain and C-domain, become positively correlated in the presence of the ligand -except Ligand-1, as shown in Figure 3.2.4 (See dashed black rectangle) which might trigger the rigid-body motion, hence the rotation of the domain with respect to the N-domain.

Sample-1   Sample-2

Sample-3   Sample-4

Residue Cross Correlation
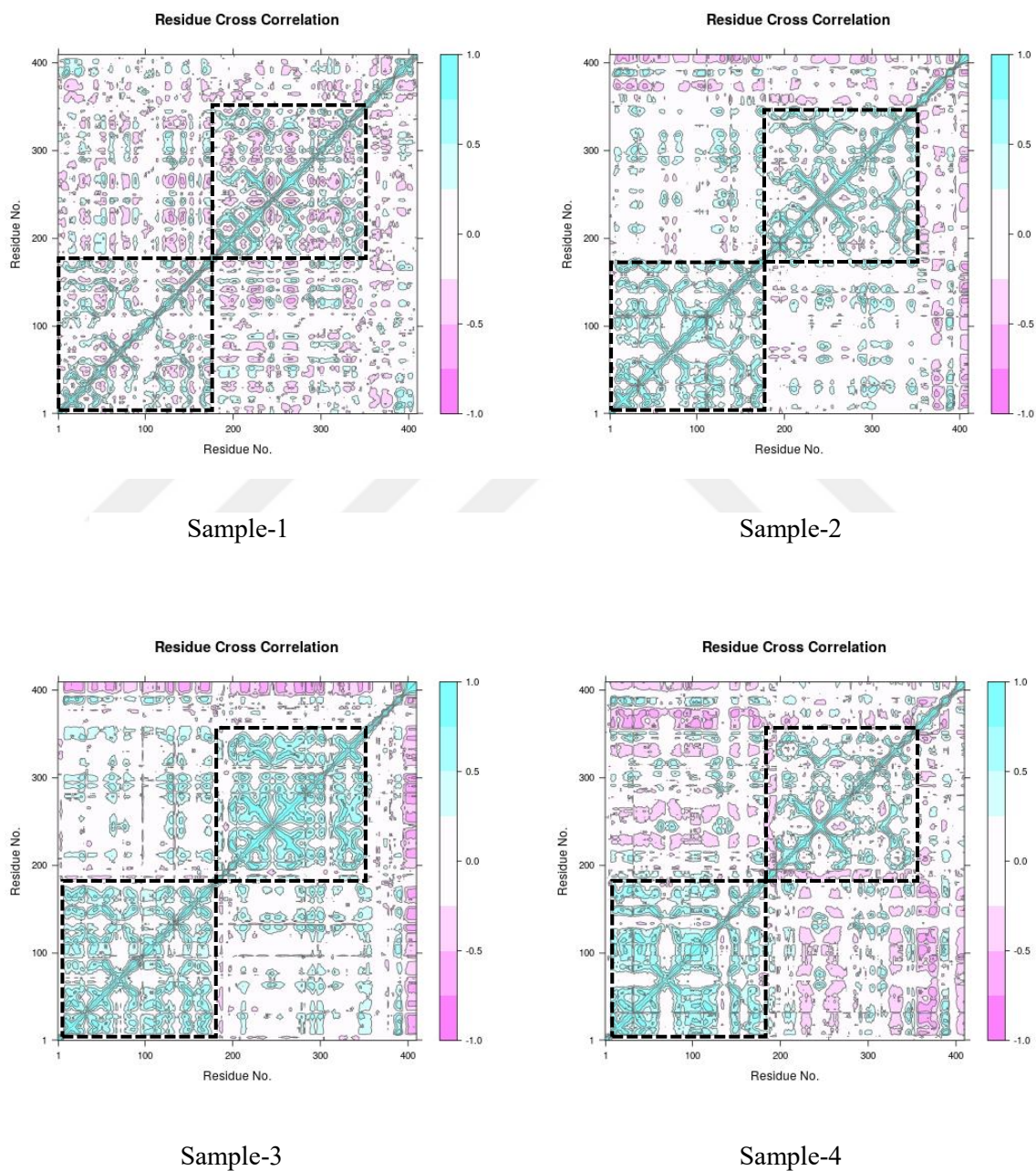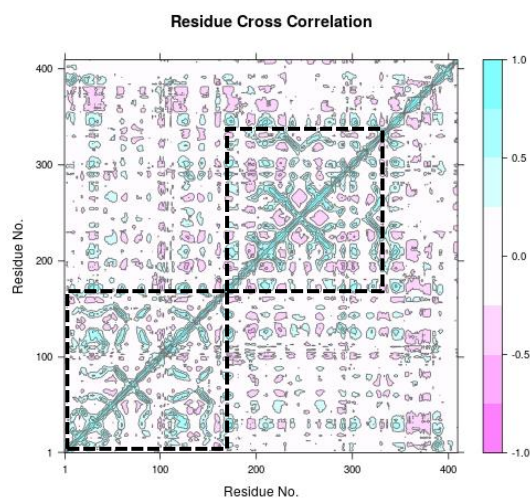
Sample-no ligand

Figure 3.2.4: *DCCMs created for Arr-3 and Arr-3/ligands systems. Cyan color corresponds to the positive correlation between residue pairs, whereas the pink color corresponds to negative correlation.*

Lastly, we made principal component analysis to investigate the impact of the ligands on the global collective motions of Arr-3. As shown in Figure 3.2.5, all of the Arr3-ligand systems sampled similar region on the conformational space that is spanned by the first and the second eigenvector. On the other hand, ligand-free Arr-3 sampled additional region which was not visited by any of the ligands.
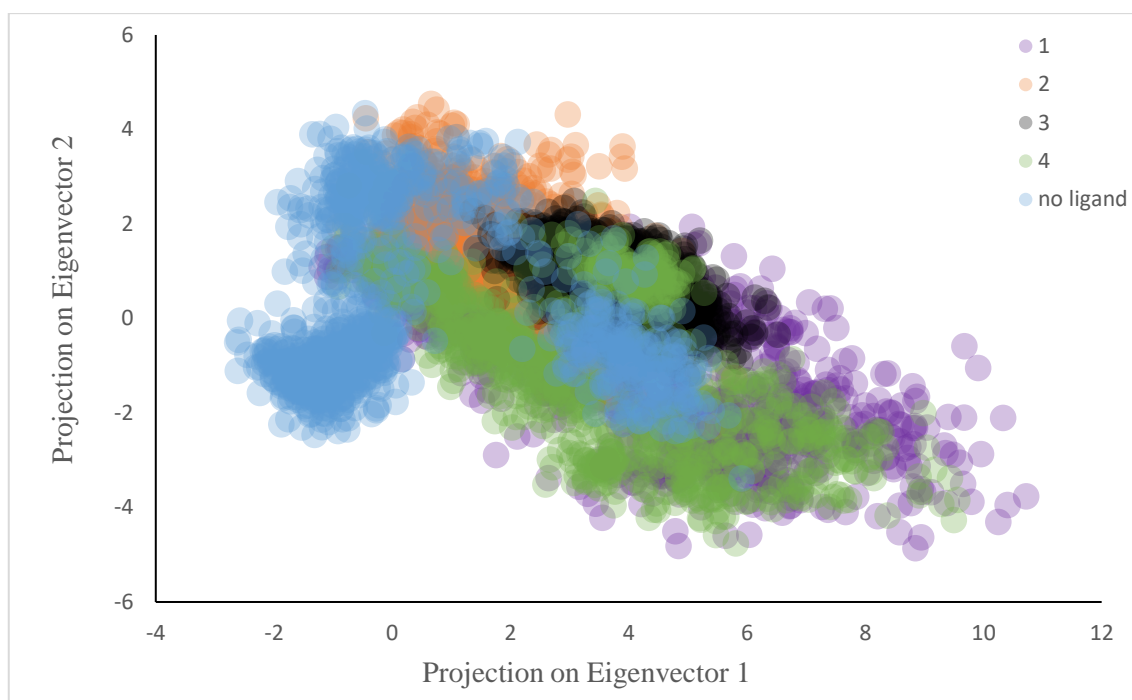
Figure 3.2.5: *2D PCA projection of trajectories. All ligand-Arr3 systems visited similar regions, whereas the ligand-free Arr3 scans a more diffuse surface.*

3.3   Enhanced Sampling Method

3.3.1   Accelerated Molecular Dynamics (aMD)

As mentioned above to further test the stability of the ligand we performed accelerated MD (aMD) simulations on Arr-3/ligand systems. Interestingly, the ligands, which stably bound to the target region in classical molecular dynamics simulations, immediately left the binding site in aMD simulations. Consequently, to test if the dynamics of the protein increased -in particular- around the "*short helix*" region we performed aMD simulations on Arr-3 alone. We observed that aMD caused an increase in the percentage of the *intermediate state* of the "*short helix*" as shown in Table 3.1.2 which explains why the ligands could not stably bind to the region. Interestingly, however, the conformational distribution of the "*gate loop*" did not change which might be due to the location of the region on the protein, which is found in a more shielded region, whereas the "*short helix*" is located in a more exposed region, thus making acceleration of the dynamics of the latter easier.

31

Table 3.3.1.1: *Comparison of the percentage of the distribution of conformational states between classical and accelerated MD trajectories.*

classical MD	Inactive	Intermediate	Active
Aromatic Core	**10.3%** < 7.2	7.2 < **78.8%** < 9.7	**10.9%** > 9.7
Short Helix	**78.3%** < 7.8	7.8 < **18.9%** < 9.5	**2.8%** > 9.5
Gate Loop	**27.7%** < 10.1	10.1 < **72.3%** < 13.9	**0.01%** > 13.9

aMD	Inactive	Intermediate	Active
Aromatic Core	**52.2%** < 7.2	7.2 < **45.8%** < 9.7	**1.9%** > 9.7
Short Helix	**61.2%** < 7.8	7.8 < **33.1%** < 9.5	**5.7%** > 9.5
Gate Loop	**22%** < 10.1	10.1 < **78%** < 13.9	0

Chapter 4

4.1 Discussion

In spite of having a high sequence similarity and conserved structural fold, the members of this small protein family display remarkable differences in their preferences towards the phosphorylation state of the receptor. Arr-1 and Arr-4 can exclusively bind to the activated and phosphorylated Rhodopsin, whereas Arr-2 and Arr-3 can bind to various types of GPCRs. Interestingly, Arr-3 can also bind to the activated and non-phosphorylated GPCR depending on the type of the receptor. Considering the fact that phosphorylation is required to release the structural constraints that stabilize the inactive conformation of Arr, the capability of binding to non-phosphorylated receptor suggests that Arr-3 is intrinsically more flexible compared to the other members. The possible source of this structural flexibility was discussed in the study that focused on the crystal structure of inactive Arr-3 [34]. Consequently, this also brings up the fact that Arr-3 uses an alternative mechanism, as opposed to the well-established "*two-step* binding mechanism" when it binds to the non-phosphorylated receptor. Indeed, in an experimental study, it was shown that Arr-3 can bind to the receptor without displacement of the C-terminus of the protein [61]. This is also verified in a recent computational study where Arr-3 was shown to adopt active-like conformational changes in the absence of perturbation of the "*polar core*" [17]. In the same study, it was also shown that unraveling of the "*short-helix*" increased the domain rotation angle [17].

In this study, we are motivated to utilize the intrinsic structural flexibility of Arr-3 and stabilize active-like conformers found in the conformational ensemble by means of small molecules. In this way, one would be able to pre-activate Arr-3 independent of phosphorus atoms attached to the C-terminus of the receptor, and also avoid problems that were encountered when attempted to create phosphorylation-independent Arr-3. Our results showed that the "*short helix"* emerges as a favorable binding pocket candidate for the ligands as it got the highest score nearly for all of the conformations used. Indeed, most of the ligands that bound to this region could stabilize the active-like conformation of Arr-3. Importantly, the analyses showed that the flexibility of the "*short helix*" region decreased independent of the type of the ligand bound; however, the state of the "*aromatic core*" and the "*gate loop*", which are also involved in the activation mechanism, seemed to be dependent on the type of the ligand used. Moreover, the ligands seemed to impact the type of correlation between residue pairs. In general, the correlation within the N- and the C-domain increased in the presence of the ligand which might trigger domain rotation hence the activation of the protein. We also tested the capability of the ligands that were docked the "*gate loop*" in the stabilization of the required angle (data not shown); however, they were not successful and maintained the rotation angle near 0. Therefore, the "*short helix*" turns out to be a key region which helps Arr-3 to stabilize the rotation angle that is needed to bind to the activated and non-phosphorylated receptor.

4.2 Conclusion

The possible advantages of phosphorylation-independent Arrs have been proposed but no method has been found to be successful yet. In this study, we provide a framework to understand the mechanism of phosphorylation-independent activation which might be used to create such constructs. Our results show that the active-like conformation of Arr3 can be stabilized by means of small molecules which target the "*short-helix*" region, but not the "*gate loop*" that is also involved in the activation mechanism.

Bibliography

[1]   C. Munk *et al.*, "An online resource for GPCR structure determination and analysis," *Nat. Methods*, 2019.

[2]   N. J. Freedman and R. J. Lefkowitz, "Desensitization of G protein-coupled receptors.," *Recent Prog. Horm. Res.*, 1996.

[3]   C. J. van Koppen, "Arrestin-Independent Internalization of G Protein-Coupled Receptors," *Mol. Pharmacol.*, 2004.

[4]   K. L. Pierce, R. T. Premont, and R. J. Lefkowitz, "Seven-transmembrane receptors," *Nature Reviews Molecular Cell Biology*, vol. 3, no. 9. pp. 639–650, Sep-2002.

[5]   A. K. Shukla *et al.*, "Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide," *Nature*, vol. 497, no. 7447, pp. 137–141, 2013.

[6]   D. Calebiro, V. O. Nikolaev, L. Persani, and M. J. Lohse, "Signaling by internalized G-protein-coupled receptors," *Trends Pharmacol. Sci.*, vol. 31, no. 5, pp. 221–228, May 2010.

[7]   N. R. Latorraca *et al.*, "Molecular mechanism of GPCR-mediated arrestin activation," *Nature*, vol. 557, no. 7705, pp. 452–456, May 2018.

[8]   Q. Chen, T. M. Iverson, and V. V Gurevich, "Structural Basis of Arrestin-

Dependent Signal Transduction," *Trends Biochem. Sci.*, vol. 43, no. 6, pp. 412–423, Jun. 2018.

[9]   L. E. Gimenez, S. Kook, S. A. Vishnivetskiy, M. R. Ahmed, E. V Gurevich, and V. V Gurevich, "Role of receptor-attached phosphates in binding of visual and non-visual arrestins to G protein-coupled receptors.," *J. Biol. Chem.*, vol. 287, no. 12, pp. 9028–40, Mar. 2012.

[10]   Q. Chen *et al.*, "Structural basis of arrestin-3 activation and signaling," *Nat. Commun.*, vol. 8, no. 1, p. 1427, Dec. 2017.

[11]   Y. J. Kim, K. P. Hofmann, O. P. Ernst, P. Scheerer, H. W. Choe, and M. E. Sommer, "Crystal structure of pre-activated arrestin p44," *Nature*, 2013.

[12]   Y. Kang *et al.*, "Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser," *Nature*, 2015.

[13]   A. K. Shukla *et al.*, "Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide.," *Nature*, 2013.

[14]   K. N. Nobles, Z. Guan, K. Xiao, T. G. Oas, and R. J. Lefkowitz, "The active conformation of β-arrestin1: Direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of β-arrestins1 and -2," *J. Biol. Chem.*, 2007.

[15]   A. K. Shukla, J. D. Violin, E. J. Whalen, D. Gesty-Palmer, S. K. Shenoy, and R. J. Lefkowitz, "Distinct conformational changes in -arrestin report biased agonism at seven-transmembrane receptors," *Proc. Natl. Acad. Sci.*, 2008.

[16]   A. K. Shukla *et al.*, "Visualization of arrestin recruitment by a G-protein-coupled receptor," *Nature*, vol. 512, no. 7513, pp. 218–222, 2014.

[17]   O. Sensoy, I. S. Moreira, and G. Morra, "Understanding the Differential Selectivity of Arrestins toward the Phosphorylation State of the Receptor," *ACS Chem. Neurosci.*, vol. 7, no. 9, pp. 1212–1224, Sep. 2016.

[18]   J. Granzin, A. Stadler, A. Cousin, R. Schlesinger, and R. Batra-Safferling,

"Structural evidence for the role of polar core residue Arg175 in arrestin activation," *Sci. Rep.*, vol. 5, no. 1, p. 15808, Dec. 2015.

[19] V. V. Gurevich, X. Song, S. A. Vishnivetskiy, and E. V. Gurevich, "Enhanced phosphorylation-independent arrestins and gene therapy," *Handb. Exp. Pharmacol.*, 2014.

[20] G. Restagno *et al.*, "A large deletion at the 3′ end of the rhodopsin gene in an italian family with a diffuse form of autosomal dominant retinitis pigmentosa," *Human Molecular Genetics*. 1993.

[21] R. Y. Kim *et al.*, "Dominant Retinitis Pigmentosa Associated With Two Rhodopsin Gene Mutations: Leu-40-Arg and an Insertion Disrupting the 5′-Splice Junction of Exon 5," *Arch. Ophthalmol.*, 1993.

[22] E. Apfelstedt-Sylla *et al.*, "Ocular findings in a family with autosomal dominant retinitis pigmentosa and a frameshift mutation altering the carboxyl terminal sequence of rhodopsin," *Br. J. Ophthalmol.*, 1993.

[23] L. S. Barak, R. H. Oakley, S. A. Laporte, and M. G. Caron, "Constitutive arrestin-mediated desensitization of a human vasopressin receptor mutant associated with nephrogenic diabetes insipidus," *Proc. Natl. Acad. Sci.*, 2012.

[24] M. R. Bristow *et al.*, "Decreased Catecholamine Sensitivity and β-Adrenergic-Receptor Density in Failing Human Hearts," *N. Engl. J. Med.*, vol. 307, no. 4, pp. 205–211, Jul. 1982.

[25] M. Ungerer, M. Bohm, J. S. Elce, E. Erdmann, and M. J. Lohse, "Altered expression of β-adrenergic receptor kinase and β1-adrenergic receptors in the failing human heart," *Circulation*, 1993.

[26] H. A. Rockman *et al.*, "Expression of a beta-adrenergic receptor kinase 1 inhibitor prevents the development of myocardial failure in gene-targeted mice.," *Proc. Natl. Acad. Sci. U. S. A.*, 1998.

[27] S. A. Akhter, A. D. Eckhart, H. A. Rockman, K. Shotwell, R. J. Lefkowitz, and W. J. Koch, "In vivo inhibition of elevated myocardial β-adrenergic receptor

kinase activity in hybrid transgenic mice restores normal β-adrenergic signaling and function," *Circulation*, 1999.

[28] L. Pan, E. V. Gurevich, and V. V. Gurevich, "The nature of the arrestin·receptor complex determines the ultimate fate of the internalized receptor," *J. Biol. Chem.*, 2003.

[29] V. V. Gurevich and J. L. Benovic, "Visual arrestin interaction with rhodopsin. Sequential multisite binding ensures strict selectivity toward light-activated phosphorylated rhodopsin," *J. Biol. Chem.*, 1993.

[30] V. V. Gurevich and J. L. Benovic, "Visual arrestin binding to rhodopsin: Diverse functional roles of positively charged residues within the phosphorylation-recognition region of arrestin," *J. Biol. Chem.*, 1995.

[31] A. Kovoor, J. Celver, R. I. Abdryashitov, C. Chavkin, and V. V. Gurevich, "Targeted construction of phosphorylation-independent β-arrestin mutants with constitutive activity in cells," *J. Biol. Chem.*, 1999.

[32] J. Celver, S. A. Vishnivetskiy, C. Chavkin, and V. V. Gurevich, "Conservation of the phosphate-sensitive elements in the arrestin family of proteins," *J. Biol. Chem.*, 2002.

[33] T. Halgren, "New method for fast and accurate binding-site identification and analysis," *Chem. Biol. Drug Des.*, vol. 69, no. 2, pp. 146–148, 2007.

[34] X. Zhan, L. E. Gimenez, V. V. Gurevich, and B. W. Spiller, "Crystal Structure of Arrestin-3 Reveals the Basis of the Difference in Receptor Binding Between Two Non-visual Subtypes," *J. Mol. Biol.*, vol. 406, no. 3, pp. 467–478, Feb. 2011.

[35] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *J. Comput. Chem.*, 2008.

[36] K. Vanommeslaeghe *et al.*, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J. Comput. Chem.*, 2010.

[37]   W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, 1983.

[38]   U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, 1995.

[39]   H. Li, A. D. Robertson, and J. H. Jensen, "Very fast empirical prediction and rationalization of protein pK a values," *Proteins Struct. Funct. Genet.*, 2005.

[40]   D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*. 2005.

[41]   Q. Chen *et al.*, "Structural basis of arrestin-3 activation and signaling," *Nat. Commun.*, vol. 8, no. 1, 2017.

[42]   X. E. Zhou *et al.*, "Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors," *Cell*, vol. 170, no. 3, pp. 457-469.e13, 2017.

[43]   J. P. Tim Isgro Marcos Sotomayor, Elizabeth Villa, Hang Yu, David Tanner, Yanxin Liu *et al.*, "NAMD Tutorial: Unix/MacOSX Version," *Univ. Illinois Urbana-Champaign, NIH Resour. Macromol. Model. Bioinformatics, Beckman Institute, Comput. Biophys. Work.*, 2012.

[44]   M. Abraham, B. Hess, D. van der Spoel, and E. Lindahl, "user manual," in *SpringerReference*, Berlin/Heidelberg: Springer-Verlag, 2015.

[45]   J. A. McCammon and S. C. Harvey, "Dynamics of proteins, nucleic acids, and their solvent surroundings," in *Dynamics of proteins and nucleic acids*, Cambridge: Cambridge University Press, 2012, pp. 25–34.

[46]   B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, "Bio3d: an R package for the comparative analysis of protein structures," *Bioinformatics*, vol. 22, no. 21, pp. 2695–2696, Nov. 2006.

[47] D. Hamelberg, J. Mongan, and J. A. McCammon, "Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules," *J. Chem. Phys.*, 2004.

[48] F. Zhang *et al.*, "Using accelerated molecular dynamics simulation to shed light on the mechanism of activation/deactivation upon mutations for CCR5," *RSC Adv.*, vol. 8, no. 66, pp. 37855–37865, 2018.

[49] J. C. Phillips, Gengbin Zheng, S. Kumar, and L. V. Kale, "NAMD: Biomolecular Simulation on Thousands of Processors," 2015.

[50] T. A. Halgren, "Identifying and characterizing binding sites and assessing druggability," *J. Chem. Inf. Model.*, vol. 49, no. 2, pp. 377–389, 2009.

[51] A. Voet *et al.*, "Pharmacophore modeling: advances, limitations, and current utility in drug discovery," *J. Receptor. Ligand Channel Res.*, vol. 7, p. 81, Nov. 2014.

[52] P. Jones, G. and Willet, "Pharmacophore Perception, Development and Use in Drug Design (Güner, O.F., ed.)," *Int. Univ. Line*, 2000.

[53] S. L. Dixon, A. M. Smondyrev, and S. N. Rao, "PHASE: A novel approach to pharmacophore modeling and 3D database searching," *Chemical Biology and Drug Design*. 2006.

[54] D. R. Koes and C. J. Camacho, "ZINCPharmer: Pharmacophore search of the ZINC database," *Nucleic Acids Res.*, 2012.

[55] G. M. Morris and M. Lim-Wilby, "Molecular docking," *Methods Mol. Biol.*, 2008.

[56] R. A. Friesner *et al.*, "Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes," *J. Med. Chem.*, 2006.

[57] Schrödinger, "LigPrep | Schrödinger," *Schrödinger Release 2018-2*. 2018.

[58] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker, "PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations," *Nucleic Acids Res.*, 2004.

[59] Oleg Trott and A. J. Olson, "AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading," *J. Comput. Chem.*, 2009.

[60] W. L. Jorgensen, "QikProp," *Schrödinger LLC, New York*. 2006.

[61] Y. Zhuo, S. A. Vishnivetskiy, X. Zhan, V. V. Gurevich, and C. S. Klug, "Identification of receptor binding-induced conformational changes in non-visual arrestins," *J. Biol. Chem.*, 2014.

IN SILICO PHOSPHORYLATION-INDEPENDENT ACTIVATION of ARRESTIN-3 PROTEIN by means of SMALL MOLECULES