NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATION

INTERDISCIPLINARY POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"

MASTER THESIS

Investigation of the S427F oncogenic mutation effects on the RXRα structure and dynamics

Georgia S. Stergiou

Supervisor: **Dr. Zoe Cournia**, Researcher - Assistant Professor Level, Biomedical Research Foundation of the Academy of Athens (BRFAA)

ATHENS

JULY 2019

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Διερεύνηση των επιδράσεων της ογκογενούς μετάλλαξης S427F στη δομή και τη δυναμική της πρωτεΐνης RXRα

Γεωργία Σ. Στεργίου

Επιβλέπουσα: **Δρ. Ζωή Κούρνια,** Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)

ΑΘΗΝΑ

ΙΟΥΛΙΟΣ 2019

MASTER THESIS

Investigation of the S427F oncogenic mutation effects on the RXRα structure and dynamics

Georgia S. Stergiou
S.N.: PIV0187

SUPERVISOR: **Dr. Zoe Cournia**, Researcher - Assistant Professor Level, Biomedical Research Foundation of the Academy of Athens (BRFAA)

EXAMINATION COMMITTEE:

Dr. Zoe Cournia, Researcher - Assistant Professor Level, Biomedical Research Foundation of the Academy of Athens (BRFAA)
Dr. Stavros Perantonis, Research Director, Institute of Informatics and Telecommunications National Center for Scientific Research “Demokritos”
Dr. Georgia Xanthou, Researcher - Assistant Professor Level, Biomedical Research Foundation of the Academy of Athens (BRFAA)

JULY 2019

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Διερεύνηση των επιδράσεων της ογκογενούς μετάλλαξης S427F στη δομή και τη δυναμική της πρωτεΐνης RXRα

Γεωργία Σ. Στεργίου

Α.Μ.: ΠΙΒ0187

ΕΠΙΒΛΕΠΟΥΣΑ: **Δρ. Ζωή Κούρνια,** Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Δρ. Ζωή Κούρνια, Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)
Δρ. Σταύρος Περαντώνης, Διευθυντής Έρευνας, Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών, Εθνικό Κέντρο Επιστημονικής Έρευνας «Δημόκριτος»
Δρ. Γεωργία Ξάνθου, Ερευνήτρια Γ', Ίδρυμα Ιατροβιολογικών Ερευνών Ακαδημίας Αθηνών (ΙΙΒΕΑΑ)

ΙΟΥΛΙΟΣ 2019

ABSTRACT

Retinoid X Receptor alpha (RXRα) is a nuclear hormone receptor of the vitamin A metabolite, 9-cis Retinoic Acid (RA). When bound to 9-cis RA, RXRα can activate the transcription of specific genes by serving either as an obligate heterodimer partner for many nuclear receptors (NRs) or as a homodimer partner. These dimers play a significant role in many physiological conditions such as regulation of metabolic rate, cell growth and differentiation. However, some of them are involved in pathological conditions such as premature birth, skin deceases and cancer development.

Studies have shown that 5-8% of bladder cancer patients bear the S427F single point mutation on RXRα, where a serine is replaced by phenylalanine at amino acid position 427. Biological experiments demonstrated that this mutation increases the expression of genes regulated by the Peroxisome Proliferator-Activated Receptor (PPAR). PPAR, in turn, switches on genes that aid cancer cells growth. Although valuable, biochemical studies aimed at determining the exact molecular mechanism of how the S427F mutation exerts its oncogenic action on RXRα in relation to its partners lack sufficient resolution for investigating the detailed underlying molecular interactions that bring about the observed result. Computational experiments such as Molecular Dynamics simulations (MD) together with atomic-level detailed experiments such as X-ray crystallography supplement biological experiments as they provide atomic resolution of the conformational changes that biomolecules undergo during a biological process. With the use of MD simulations, details of conformational changes are sampled to clarify mechanistic characteristics of biological processes that may otherwise be elusive.

In the present study, following up on a series of experiments that showed that the mutant RXRα inhibits the transcription of genes regulated by the RXRα homodimer, we investigated the effects of the S427F mutation on the structure and dynamics of the RXRα homodimer in the presence and absence of the ligands 9-cis RA. MD simulations were performed in order to elucidate the structural and dynamical consequences of this oncogenic mutation on the RXRα protein and its interactions within both Wild Type (WT) and mutant RXRα homodimers. Moreover, control simulations were performed to compare with previously reported data from the WT and mutant RXRα-PPARγ heterodimer. Results were also compared with previously performed RAR-RXRα heterodimer simulations. Furthermore, Dynamical Network Analysis was used to illustrate differences in the inter-residue pathways of the three dimers, in order to examine if there is a disruption in the communication.

Our results show that the mutation S427F does not affect the dimerization ability of the RXRα homodimer neither in the presence nor the absence of the endogenous ligand 9-cis RA. However, further analysis of the apo mutant demonstrates an increase of the solvent accessible surface area of the binding site in both monomers of the dimer in the absence of the ligands compared to the WT protein. This fact could prohibit the ligand 9-cis RA from binding resulting to a loss of transcriptional activity for RXRα. Finally, we validated our results analyzing the findings of the simulated RXRα-PPARγ heterodimer and compared to previous RAR-RXRα simulations, and finding that the two calculations are in agreement.

SUBJECT AREA: Computational Molecular Biophysics

KEYWORDS: RXRα, PPARγ, bladder cancer, molecular dynamics, dynamical network analysis

ΠΕΡΙΛΗΨΗ

Ο υποδοχέας του ρετινοειδούς Χ άλφα (RXRα) είναι ένας υποδοχέας πυρηνικής ορμόνης του μεταβολίτη βιταμίνης Α, 9-cis ρετινοϊκό οξύ (RA). Όταν συνδέεται με το 9-cis RA, ο RXRα μπορεί να ενεργοποιήσει τη μεταγραφή συγκεκριμένων γονιδίων εξυπηρετώντας είτε ως υποχρεωτικός συνεργάτης ετεροδιμερούς για πολλούς πυρηνικούς υποδοχείς (NRs) είτε ως συνεργάτης ομοδιμερούς. Αυτά τα διμερή παίζουν σημαντικό ρόλο σε πολλές φυσιολογικές καταστάσεις όπως η ρύθμιση του μεταβολικού ρυθμού, η κυτταρική ανάπτυξη και η διαφοροποίηση. Εντούτοις, μερικοί από αυτούς εμπλέκονται σε παθολογικές καταστάσεις όπως η πρόωρη γέννηση, οι δερματικές ασθένειες και η ανάπτυξη του καρκίνου.

Μελέτες έχουν δείξει ότι το 5-8% των ασθενών με καρκίνο της ουροδόχου κύστης φέρουν τη σημειακή μετάλλαξη S427F στον RXRα, όπου μια σερίνη αντικαθίσταται από φαινυλαλανίνη στη θέση αμινοξέος 427. Βιολογικά πειράματα κατέδειξαν ότι αυτή η μετάλλαξη αυξάνει την έκφραση γονιδίων που ρυθμίζονται από τον πολλαπλασιαστή υπεροξυσώματος - ενεργοποιημένος υποδοχέας (PPAR). Ο PPAR, με τη σειρά του, ενεργοποιεί τα γονίδια που βοηθούν την ανάπτυξη των καρκινικών κυττάρων. Αν και πολύτιμες, οι βιοχημικές μελέτες που στοχεύουν στον προσδιορισμό του ακριβούς μοριακού μηχανισμού με τον οποίο η μετάλλαξη S427F ασκεί την ογκογόνο δράση της στον RXRα σε σχέση με τους συνεργάτες της, δεν έχουν επαρκή ανάλυση για να διερευνήσουν τις λεπτομερείς υποκείμενες μοριακές αλληλεπιδράσεις που επιφέρουν τον παρατηρούμενο φαινότυπο. Υπολογιστικά πειράματα, όπως οι προσομοιώσεις Μοριακής Δυναμικής (ΜΔ) μαζί με πειράματα σε επίπεδο ατομικής ανάλυσης, όπως η κρυσταλλογραφία ακτίνων Χ, συμπληρώνουν τα βιολογικά πειράματα καθώς παρέχουν ατομική ανάλυση των διαμορφωτικών αλλαγών που υφίστανται τα βιομόρια κατά τη διάρκεια μιας βιολογικής διαδικασίας. Με τη χρήση προσομοιώσεων ΜΔ, εξετάζονται λεπτομερώς οι μεταβολές της διαμόρφωσης για να αποσαφηνιστούν τα μηχανιστικά χαρακτηριστικά των βιολογικών διεργασιών που διαφορετικά θα παρέμεναν ασαφή.

Στην παρούσα μελέτη, ακολουθώντας μια σειρά πειραμάτων που έδειξαν ότι ο μεταλλαγμένος RXRα αναστέλλει τη μεταγραφή των γονιδίων που ρυθμίζονται από το ομοδιμερές RXRα, ερευνήσαμε την επίδραση της μετάλλαξης S427F στη δομή και τη δυναμική του ομοδιμερούς RXRα με την παρουσία και την απουσία των προσδετών 9-cis RA. Διεξήχθησαν προσομοιώσεις ΜΔ προκειμένου να διασαφηνισθούν οι δομικές και δυναμικές συνέπειες αυτής της ογκογενούς μετάλλαξης στην πρωτεΐνη RXRα και η αλληλεπιδράσεις της στο ομοδιμερές φυσικού τύπου (WT) καθώς και στο μεταλλαγμένο RXRα ομοδιμερές. Επιπλέον, πραγματοποιήθηκαν προσομοιώσεις ελέγχου για σύγκριση με προηγούμενα αναφερόμενα δεδομένα από το WT και μεταλλαγμένο RXRα-PPARγ ετεροδιμερές. Τα αποτελέσματα συγκρίθηκαν επίσης με προηγούμενες προσομοιώσεις του ετεροδιμερούς RAR-RXRα. Επιπροσθέτως, χρησιμοποιήθηκε Δυναμική Ανάλυση Δικτύων για την εξέταση της επικοινωνίας μεταξύ των αμινοξέων της πρωτεΐνης στα τρία διμερή, προκειμένου να εξεταστεί εάν υπάρχει διαταραχή στην επικοινωνία λόγω της μετάλλαξης.

Τα αποτελέσματά μας δείχνουν ότι η μετάλλαξη S427F δεν επηρεάζει την ικανότητα διμερισμού του ομοδιμερούς RXRα ούτε παρουσία ούτε απουσία των ενδογενών προσδετών 9-cis RA. Εντούτοις, περαιτέρω ανάλυση του μεταλλαγμένου ομοδιμερούς καταδεικνύει αύξηση της επιφάνειας προσβάσιμης από το διαλύτη στην περιοχή πρόσδεσης του 9-cis RA και στα δύο μονομερή του διμερούς απουσία των προσδετών σε σύγκριση με την WT πρωτεΐνη. Το γεγονός αυτό θα μπορούσε να απαγορεύσει τη δέσμευση του προσδέτη 9-cis RA με αποτέλεσμα να εμποδίζεται η μεταγραφική

λειτουργία του RXRα. Τέλος, επικυρώσαμε τα αποτελέσματά μας αναλύοντας τα ευρήματα της προσομοίωσης του ετεροδιμερούς RXRα-PPARγ και συγκρίνοντάς τα με προηγούμενες προσομοιώσεις του RAR-RXRα, τα οποία βρίσκονται σε συμφωνία.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Μοριακή Βιοφυσική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: RXRα, PPARγ, καρκίνος της ουροδόχου κύστης, μοριακή δυναμική, δυναμική ανάλυση δικτύων

To my family

.

ACKNOWLEDGMENTS

I would like to take this opportunity to acknowledge those who have helped me complete this thesis. First and foremost, I would like to express my gratitude to my supervisor, Dr. Zoe Cournia, Researcher - Assistant Professor Level, at the Biomedical Research Foundation of the Academy of Athens (BRFAA) - her encouragement, support, and thoughtful advice have been immensely valuable for me and my academic steps. Her guidance helped me in my research and in writing this thesis.

I would like to thank Giannis Galdadas, PhD student at the University College London, UK, and previous member of the Cournia's lab, whose research was the groundwork of this diploma thesis for his useful advice and guidance in setting up, running and analyzing Molecular Dynamic simulations as well as rationalizing the results.

For their precious advice, I would also like to thank the two members of my MSc diploma thesis examination committee, Dr. Stavros Perantonis, Research Director, Institute of Informatics and Telecommunications National Center for Scientific Research "Demokritos" and Dr. Georgia Xanthou, Researcher - Assistant Professor Level, Biomedical Research Foundation of the Academy of Athens (BRFAA).

I also would like to thank my fellow lab mates Alexios Chatzigoulas, Stefan Doerr, Dimitris Ntekoumes, Anastasia Theodoropoulou, Theodora Argyropoulou, Vasilis Georgilas for all the valuable discussions and suggestions throughout my thesis and presentation.

Last but not least, I would like to thank my family: my parents, my sister, my brother in law and all my friends for supporting me spiritually throughout this thesis and in my life in general.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The master thesis “Investigation of the S427F oncogenic mutation effects on the RXRα structure and dynamics” has been conducted at the Biomedical Research Foundation Academy of Athens for the completion of the Postgraduate Program "Information Technologies in Medicine and Biology" (I.T.M.B.), Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece.

The first Chapter presents an overview of a specific category of nuclear receptors, the Retinoid X Receptors alpha. Their distinctive characteristics and function in an organism are illustrated, along with their implication in bladder cancer, as a result of their hot-spot mutation at position 427. In the end of this chapter, the motivation of the study is presented.

In the second Chapter, the theoretical foundations of the present work are described. First, the Molecular Dynamics (MD) theory is introduced, followed by the basic principles of Statistical Mechanics and the approximations that an MD simulation includes. Finally, the theory behind the Dynamical Network Analysis, which aims to describe the allosteric pathway communication in proteins, is outlined.

The results of the present thesis are presented in Chapter three. First, the procedure that was followed for the MD simulations of the biological systems is detailed, followed by the results from the trajectory and Dynamical Network Analysis for the studied systems and the main findings of this study are discussed.

Finally, the conclusions constitute the epilogue of this thesis, along with the possible future perspectives of this study in Chapters four and five.

1. INTRODUCTION

In this Chapter, an overview of the nuclear receptor, Retinoid X Receptor (RXR), is presented. The structure and mechanism of the specific type of RXRs, Retinoid X Receptor alpha (RXRα) is demonstrated along with its biological role. Moreover, the involvement of RXRα in bladder cancer is illustrated. Finally, the objective of this study is explained.

1.1 Transcription Factors in biology

Transcription factors are proteins that control the rate of transcription of genetic information (from DNA to messenger RNA), by binding to specific DNA sequences [1]. They regulate genes in order to ensure that they are expressed correctly in each cell and in the right amount. Also, transcription factors directly manage cell division and growth as well as cell migration and organization during the embryonic development. Finally, they respond to signals from outside the cell, such as hormones.

Another function of transcription factors is that they work either alone or with other proteins in a complex, in order to promote, functioning as an activator, or to repress, functioning as a repressor, the recruitment of RNA polymerase which is the enzyme that performs the transcription of genes [2].

One distinctive characteristic of transcription factors is that they contain at least one DNA-binding domain (DBD), which binds to a specific sequence of DNA adjacent to the genes that they regulate [3].

1.2 Nuclear Receptors

One class of transcription factors is known as nuclear receptors (NRs). NRs are responsible to sense thyroid and steroid hormones. They are classified as transcription factors because they have the unique ability to directly bind to DNA and control the expression of the adjacent genes. Also, as transcription factors, they can interact with other proteins in order to regulate the development, homeostasis and metabolism of an organism [4][5].

The control of gene expression by nuclear receptors usually only occurs when a ligand (a small molecule that affects the receptor's behavior) is binding. Because these small molecules were unknown at first, the receptors took the name ‘‘orphan’’ receptors [6]. More specifically, when the ligand binds to a nuclear receptor promotes a conformational change, which, consequently, leads to the activation of the receptor which in turn regulates transcription of its target genes. Generally, nuclear receptors could be classified according to either mechanism [7][8] or homology [9][10].

1.3 The Retinoid X Receptor alpha

The first uncovering of nuclear receptors occurred from the field of endocrinology and through the identification of lipophilic hormones that function as their ligands [4]. Steroid and thyroid hormones, along with vitamins A and D, were decoded based on their endocrine origin and the physiologic processes that they manage. Each of these molecules is chemically distinct, however, biochemical experiments uncovered the

presence of an intracellular receptor that, upon ligand binding, activated transcription of target genes [7][8].

For many decades, the basic nature of the nuclear receptors, their mechanism for regulating the gene expression and their ability of identification of specific molecules were unclear. Eventually, in the mid-1980s, the first biomolecular and genetic characterization of the genes, encoding the first steroid receptors, was established [9].

One essential member of nuclear receptors that was identified is the Retinoid X Receptor, that was first characterized as an orphan receptor. RXRs belong to the steroid/thyroid hormone superfamily of NRs. There are three isotypes of RXR: RXR-alpha (RXRα), RXR-beta (RXRβ), and RXR-gamma (RXRγ), encoded by the RXRA, RXRB, RXRG genes, respectively.

It has been found that RXRs are activated by a ligand called 9–cis retinoic acid (9-cis-RA), which has reported to be an endogenous ligand [10]. Thus, RXRs can be classified in the category of “adopted” orphan receptors [6]. 9-cis RA or alitretinoin belong to a class of chemical compounds that are vitamers of vitamin A. Retinoids, in general, are used in medicine where they control epithelial cell growth. Also, they have many significant functions throughout the body involving vision, control of cell proliferation and differentiation, immune function and activation of tumor suppressor genes. Research is also being done into their ability to treat skin cancers [11]. Nowadays, 9-cis-RA can be used topically to help treat skin wounds from Kaposi's sarcoma [12].

Subsequently, we focus on the structure, distinctive characteristics and functions of RXRα.

1.3.1 Structure of RXRα

The structure and domains of RXRα (Figure 1), just like all nuclear receptors, are described below [13] [14].

They consist of a N-terminal domain, which includes the activation function 1 (AF-1); it acts independently from the presence of ligand [15]. The AF-1 has not very strong transcriptional activity, but it collaborates with AF-2 in the E-domain in order to induce a dominant upregulation of gene expression.

Next, there is the DNA-binding domain (DBD), which contains two zinc fingers that bind to specific sequence of DNA known as Hormone Response Elements (HRE). It is a highly preserved domain.

Then, the Hinge region connects the DBD with the Ligand Binding Domain (LBD) and is assumed to be responsible for the flexibility of the protein, intracellular connection and distribution.

LBD is generally preserved in sequence and highly preserved in structure among the different nuclear receptors. The structure of the LBD is known as “alpha helical sandwich fold” in which three anti parallel alpha helices (“sandwich filling”) are surrounded by two alpha helices on one side and three on the other (“bread”). The ligand binding region is enclosed by the LBD and three anti parallel alpha helical “sandwiches filling”. LBD, also, contributes to the dimerization interface of the receptor and the binding of the coactivators and corepressors. In addition, LBD includes the activation function 2 (AF-2), which acts based on the presence of the ligand [15].

Finally, the C-terminal domain is highly changeable in sequence among the nuclear receptors.

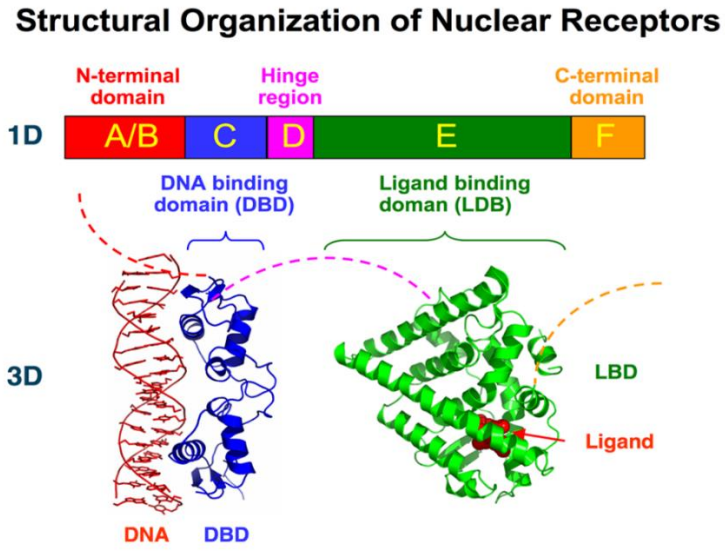

Figure 1: Structural Organization of Nuclear Receptors. (Top) Representation 1D amino acid sequence of a nuclear receptor. (Bottom) 3D structures of the DBD and LBD regions of the nuclear receptor. The structures shown are of the estrogen receptor. Experimental structures of N-terminal domain (A/B), hinge region (D), and C-terminal domain (F) have not been determined therefore are illustrated by red, purple, and orange dashed lines, respectively [16]**.**

The domain that is of great interest is the Ligand Binding Domain (Figure 2). It consists of 12 a-helices, a single β-sheet and short connecting loops.

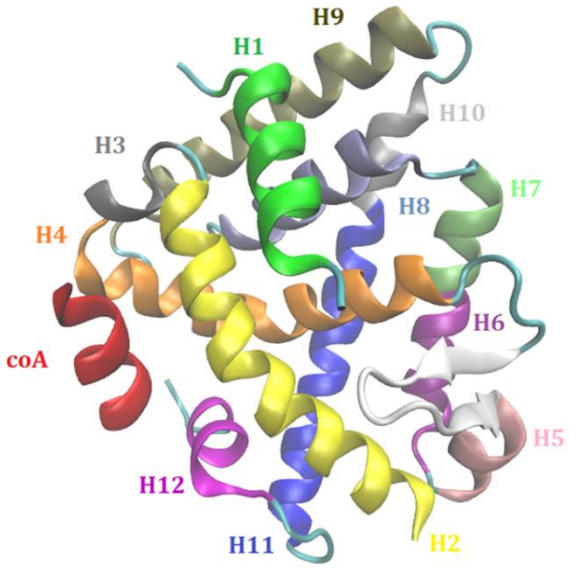

Figure 2: Representation of the Ligand Binding Domain (LBD) of the RXRα Homodimer. The helices are shown with different colors, β-sheet in white and coactivator with red (pdb code: 1MVC).

1.3.2 Heterodimeric Partners of RXRα

RXRα belongs to subfamily 1 of the nuclear receptors (Figure 3). It functions either as homodimer or as obligate heterodimeric partner for other nuclear receptors from subfamily 1 such as Farnesoid X Receptor (FXR), Liver X Receptor (LXR), Peroxisome Proliferator-Activated Receptor (PPAR), Retinoic Acid Receptor (RAR), Thyroid Receptor (TR), and Vitamin D Receptor (VDR). All the heterodimeric partners of RXRα play important physiological role in regulation of metabolic rate, cell growth, differentiation etc. [13]. However, they are involved in some pathological conditions, such as premature birth, skin disease and cancer development [17].

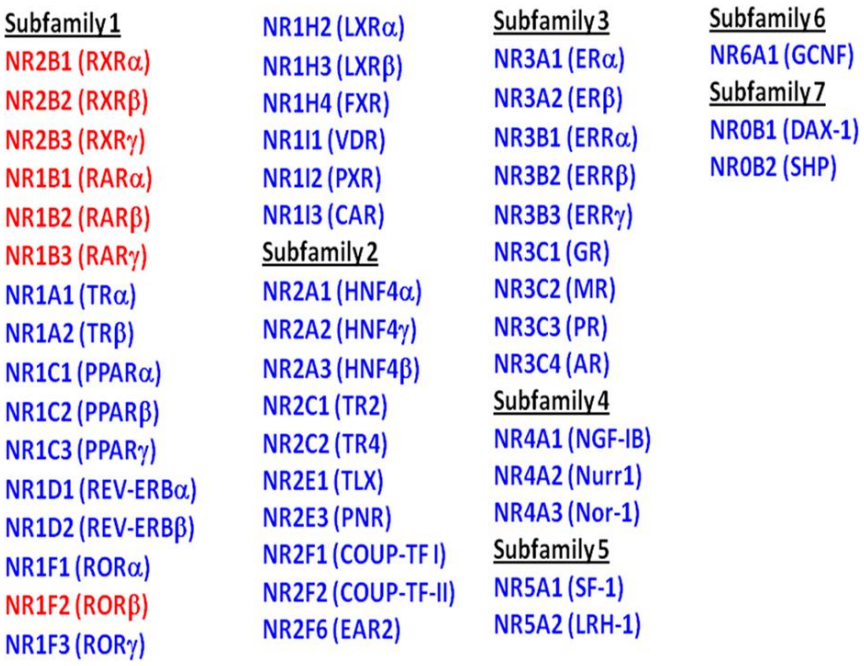

Figure 3: The NR family. Shown are the human NRs divided into subfamilies based on sequence alignments. Receptors in red are the high-affinity retinoic acid binding receptors [13]**.**

RXRα, also, forms homodimers. The main difference between homodimers and heterodimers is that they activate the expression of different genes by binding to different response elements across gene promoters [18]. RXRα dimers can recognize and bind to specific response elements, called direct repeats (DRs). DRs are formed by two conserved hexanucleotides, spaced by 1-5 bases which are called DR1 to DR5 respectively (Figure 4). The RXRα-RXRα homodimers and RXR-PPAR heterodimers bind preferentially to DR1 response elements, whereas, for example, the RXR-RAR heterodimers bind preferentially to DR5 and DR1 response elements [19]. However, it is difficult to find target genes that are exclusive targets of RXR-RXR homodimers or RXR-PPAR heterodimers.

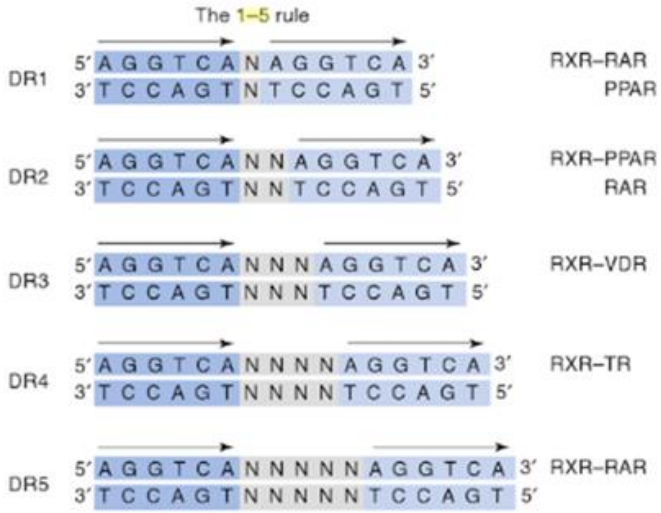

Figure 4: The 1-5 rule of DNA direct repeat binding by RXR and its nuclear receptors partners. The base pair size of the spacing between AGGTCA sequences can vary from one to five [20]**.**

1.3.3 Mechanism of RXRα activation

Depending on the receptor and the chemical structure of the ligand involved, all nuclear receptor ligands may display entirely diverse effects, such as agonism or antagonism [21]. Agonist ligands are the chemical compounds that upon binding to the receptor induce an upregulation of gene expression. This phenomenon is known as agonist response. On the contrary, antagonists are the ligands that the down-regulate the gene expression and inhibit the binding of the agonists.

As a result, the transcriptional activities of RXRα are initiated by the ligand interactions in the LBD region. Specifically, in the absence of an agonist ligand, the dimers associate with corepressors. The binding of corepressors leads to recruitment of some complexes, HDACs, that down-regulate the gene expression. The HDACs alter the histone tails in order to remove acetylation marks that are associated with active transcription. This process leads to histone condensation and gene silencing [22].

On the other hand, when an agonist binds to the LBD, it provokes the dissociation of the corepressors and the binding of coactivators. The coactivators, in turn, recruit some complexes, among them the RNA polymerase, which lead to chromatin decondensation and activation of transcription [22]. The interesting point is how the coactivators are able to bind to the dimer. The answer is that the binding of the agonist ligand induces a change in the conformation of helix 12 that allows the coactivators to bind [13].

The mechanism is described below (Figure 5). Without an agonist ligand bound, the helix 12 of the RXRα subunit is away from the LBD. The moment an agonist binds to the ligand binding pocket, helix 12 moves in position to enclose the ligand. Helix 2 unwinds and helices 3, 4 and 12 develop a surface where the coactivator binds; in a region right above of helix 12, being arranged in a so-called “gamma position” [13].

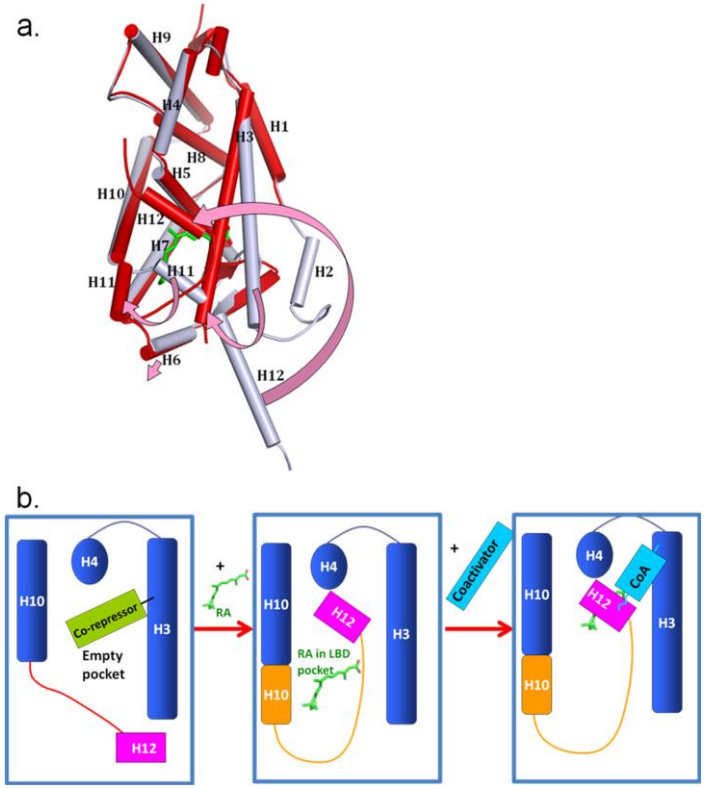

Figure 5: Representation of the ligand-induced positioning of helix-12 into the "active conformation" (a) Superposition of RXRα LBD in state without a ligand (gray) and the same receptor in state with the ligand (red) 9-cis RA (green) in the pocket. The four arrows (pink) show the helical movements induced by ligand binding. (b) Representation of how corepressors, ligand, and coactivators interact with a NR LBD. The blue box represents the LBD portion of NRs [13]**.**

The resulting conformation is known as "active conformation", whereas in the absence of a ligand as "inactive conformation". Some examples are shown in Figure 6.

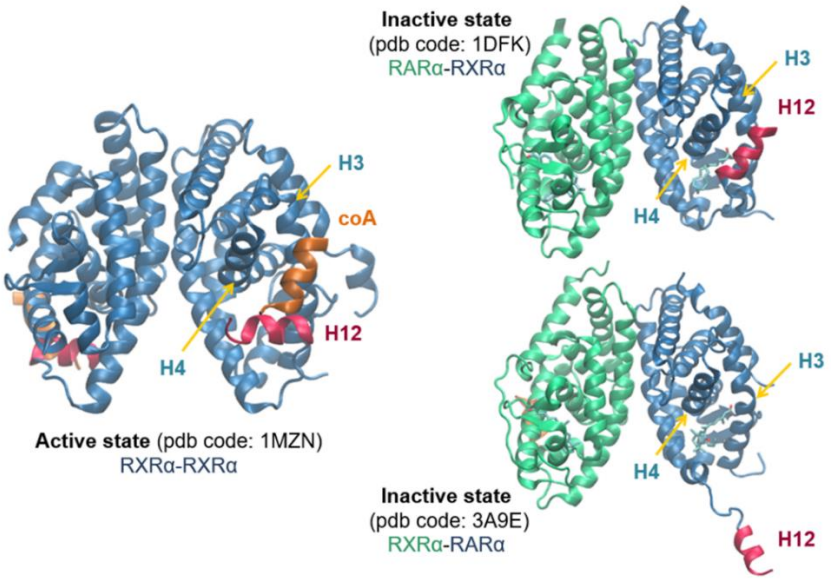

Figure 6: Examples of the active and inactive states (conformations) of the RXRα-RARα heterodimer. For the active state the RXRα-RXRα homodimer (PDB code:1MZN) is shown in blue,

helix 12 in red and coactivator in orange. Helix 12 is in a position that the coactivator can bind. For the inactive states the structures of RARα-RXRα heterodimer is shown. On top RARα is illustrated in green color, RXRα in blue and helix 12 in red, whereas at the bottom the RXRα is shown in green, RARα in blue and helix 12 in red. In both structures the helix 12 is in a position that the coactivator cannot bind [23]**.**

There are two mechanisms, with which RXRα can be activated resulting to the active conformation; these are described in the next section.

1.3.3 RXRα as active or silent transcription partner

RXRα has the potential to be activated by either the RXRα ligand or its partner receptor ligand. As a result, in some cases, in the absence of the RXRα ligand there is no transcriptional activity. Due to this fact, RXRα dimers are classified as non-permissive or permissive groups [6] [13].

Non-permissive RXRα dimers are activated only when the partner's ligand is present while the RXRα is silent [13]. Such examples are the complexes RXR-RAR and RXR-VDR. This phenomenon may be caused by the fact that corepressors are unable to be released from the RXR subunit of the heterodimer. It has been suggested that corepressors obstruct the binding of coactivators to the partner receptor [24]. When agonist ligands bind to both subunits of the complex lead to supportive coactivator binding [24]. Remarkably, in the case of RXR-RAR heterodimer, in the presence of its ligands, RAR "becomes" permissive. RAR and RXR agonists together have better effect than the RAR agonist alone, which means that the binding of RAR agonist boosts the transcriptional activity of the heterodimer by 30%. Therefore RAR-RXR is called a conditional permissive heterodimer [24].

On the other hand, permissive RXRα dimers can be activated by the ligand either of RXRα or its partner [13] [24]. Such complexes are the RXR-PPAR, RXR-LXR and RXR-FXR. A reasonable explanation for this ability may be due to that the RXR subunit in these complexes is gets a conformation that does not allow the corepressors to bind [24].

Besides the function of RXRα as heterodimeric partner, it can also form homodimers as well as tetramers. Their mechanism and functions are presented in the next section.

1.3.4 Tetramer and dimer forms of RXRα and their biological role

In vitro studies have indicated that the RXRα LBD can self-associate into tetramers with high affinity, and could exist predominately in the tetrameric form in the absence of a ligand [25] [26]. At this conformation, the tetramer (Figure 7) has been suggested to represent an inactive form and that does not exhibit any transcriptional activity [26]. This may be caused by two reasons. The first one, is due to an autorepression mechanism of the tetramer [26]. Specifically, helix 12 of each monomer extend outward from each LBD into the coactivator binding site of each monomer thus causing autorepression [13].

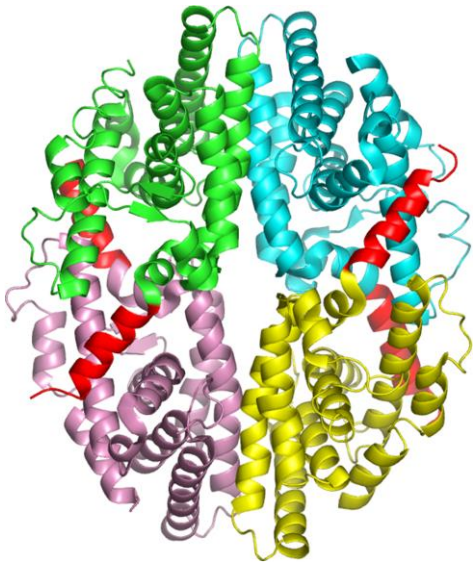

Figure 7: Structure of the RXRα LBD tetramer in the absence of ligand. The four subunits are shown in different colors (helix-12 is in red) [13]**.**

The second reason is that, when 9-cis RA binds to the RXR tetramer, it induces a rapid dissociation into homodimers and monomers [27]. Thus, it was proposed that the dissociation of the tetramer, caused by ligand binding, might be the first step toward activating RXR [28].

As a result, homodimerization is triggered by the 9-cis RA binding to the tetramer (Figure 8) [28]. RXR is the only member of the nuclear receptor family that can form homodimers and numerous studies suggest that RXRα homodimers play distinctive role in various biological processes [29] [30]. Specifically, two studies described, below, support the existence of an RXRα homodimer signaling pathway [31] [32].

First, *in vivo* studies reported that a specific RXRα homodimer agonist effectively lowers blood glucose in an animal model of insulin-resistant diabetes [30]. The authors found that Rexinoids (agonist ligands of RXRs) function as insulin sensitizers and can decrease hyperglycemia through an RXRα homodimer-mediated mechanism. Another, *in vivo*, finding is that RXRα homodimers bind to PPAR response elements (PPREs), and in combination with the recruitment of a specific coactivator (SRC1), lead to activate PPAR target genes [29]. In the presence of an RXR agonist, coactivator SRC1 preserve the binding of RXRα homodimers to PPREs.

Furthermore, it has been suggested that agonist ligands bound to RXRα homodimers upregulate the p21 cyclin-dependent kinase inhibitor (CKI) [31]. p21 contains DR-1 response elements, where RXRα homodimers bind selectively connect. Thus, RXRs are able to bind directly to p21 promoters. Ligands that activate other partners, such as the RARs or PPARs, failed to induce p21. Finally, it has been found that the RXRα homodimers are involved in innate immune response to inflammatory stimuli [32]. Agonist ligands, such as 9-cis RA and LG100268, of RXRα homodimers have been proved to induce the expression of CCL6 and CCL9 chemokines in mice macrophages. Once again, the promoters of chemokines contain DR-1 response elements. However, excessive activation of these chemokines may cause harmful consequences, such as sepsis.

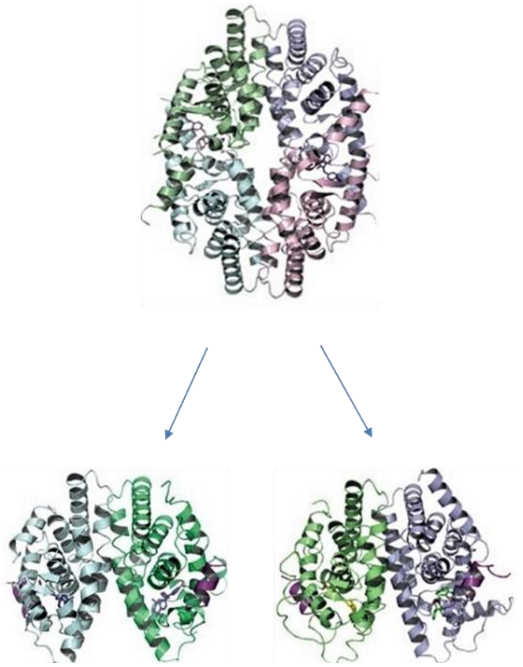

Figure 8: Representation of the dissociation of the tetramer (on top) into homodimer (bottom left) and heterodimer (bottom right) [33]**.**

Although, RXRα is involved in many physiological conditions, various mutations in its structure can affect the heterodimerization or homodimerization ability. These RXRα mutations are explained below.

1.3.5 RXRα mutations

Studies have shown that different mutations in RXRα lead to various consequences according to the heterodimeric partner. For instance, in Ref. [34] the authors investigated the effect of several mutations on the ability of RXRα to form RXRα homodimers as well as RXRα-RARα heterodimers (Figure 9).

Location	Mutant	Amino acid(s) mutated	Activity with dimerization partner	
			RXR	RAR
H7	m1	E357A	135 ± 15	102 ± 1
	m2	K361A	<5	115 ± 13
Loop L7–8	m3	R363A	90 ± 10	127 ± 10
	m4	D364A	70 ± 5	105 ± 7
Loop L8–9	m5	N382A	<5	119 ± 7
	m6	D384A	<5	98 ± 8
H9	m7	E395A	<5	110 ± 5
	m8	D384A/E395A	<5	102 ± 6
	m9	R398A	160 ± 40	106 ± 25
	m10	E399A	24 ± 4	136 ± 15
	m11	Y402A	228 ± 15	23 ± 7
	m12	E406A	30 ± 10	120 ± 15
	m13	E399A/Y402A/E406A	230 ± 30	30 ± 5
	m14	R363A/D384A/R426A	<5	115 ± 13
	m15	R363A/E395A/R426A	<5	90 ± 8
H10	m16	F420A	>300	81 ± 21
	m17	K422A	<5	105 ± 7
	m18	L424A	<5	125 ± 16
	m19	R426A	<5	70 ± 8
	m20	K422A/R426A	<5	68 ± 5
	m21	L427A	8 ± 1	122 ± 8
	m22	R431A	145 ± 15	117 ± 13
	m23	L435A	15 ± 10	108 ± 15
Loop L10–11	m24	E439A	68 ± 8	72 ± 5
	m25	D364A/D384A/E399A/Y402A/E406A/R426A/E439A	98 ± 3	43 ± 7

[a] After autoradiography, the EMSAs were quantified by densitometry and the results were expressed as the percentages of RXR mutant dimerization activities compared with the wtRXR activity, which was arbitrarily defined as 100%. The results are the means ± standard deviations of two to three independent experiments.

Figure 9: Homo- and heterodimerization abilities of the RXR mutants as assayed by EMSA (Electrophoretic Mobility Shift Assa). The results were expressed as the percentages of RXR mutant dimerization activities compared with the RXR activity, which was arbitrarily defined as 100%. The results are the means ± standard deviations of two to three independent experiments. [34].

Briefly, they suggest that mutations such as E357A and R398A reinforce the RXRα-RXRα homodimerization, while they have no effect on the RARα-RXRα heterodimerization. Moreover, mutations such as Y402A or E406A enhance the homodimerization, but abolish heterodimerization. On the contrary, L424A and L427A mutations inhibit the RXRα-RXRα homodimerization and enhance the RARα-RXRα heterodimerization, whereas E439A mutation abolishes both.

Moreover, many studies report that a single mutation in RXRα plays a crucial role in bladder cancer [35] [36]. The next section explains in detail this claim.

1.3.6 The mutation S427F of RXRα in bladder cancer: implications of heterodimeric partners

Analysis from the Cancer Genome Atlas (TCGA) [37] [38] [39] has identified a number of gene mutations in samples from muscle invasive, urothelial tumors and peripheral blood,

that had not previously reported as altered, in a significant portion in bladder cancer. Statistically significant levels of these mutations (8-9% of the samples) showed, among others, that the mutated gene corresponds to RXRα. Seven of twelve mutations in RXRα appeared at the same amino acid 427, five of them being S427F and two of them S427Y, in the LBD of the receptor.

In order to explore how the single mutation S427F in RXRα affects the different heterodimers, computational and biological experiments have been performed. To avoid any misunderstanding, those experiments were conducted in order to investigate the transcriptional output (activity) of dimers bearing this mutation and not whether this event is tumor suppressing or oncogenic. The oncogenic profile of the mutation is dependent to a large extend to the specific tissue cell type, as it happens in the case of the RXRα-PPARγ heterodimer. Thus, our objective was to observe whether the mutation would have any putative effect on only the transcriptional activity of the dimers bearing this mutation.

Consequently, according to previously-performed microsecond simulations in the Cournia lab for the heterodimer RXRα-RARα [23], in the absence of the ligand, the mutation S427F affects the communication between the helices 3, 10 and 11 (Figure 10). Helix 12 rotates back to a position that would allow the coactivator binding. The helix 11 almost immediately rejoins with helix 10, and helix 3 collapses into the void that this motion leaves. This movement of helix 3 causes a dramatic decrease of the binding site volume, leaving no space for 9-cis RA to bind.

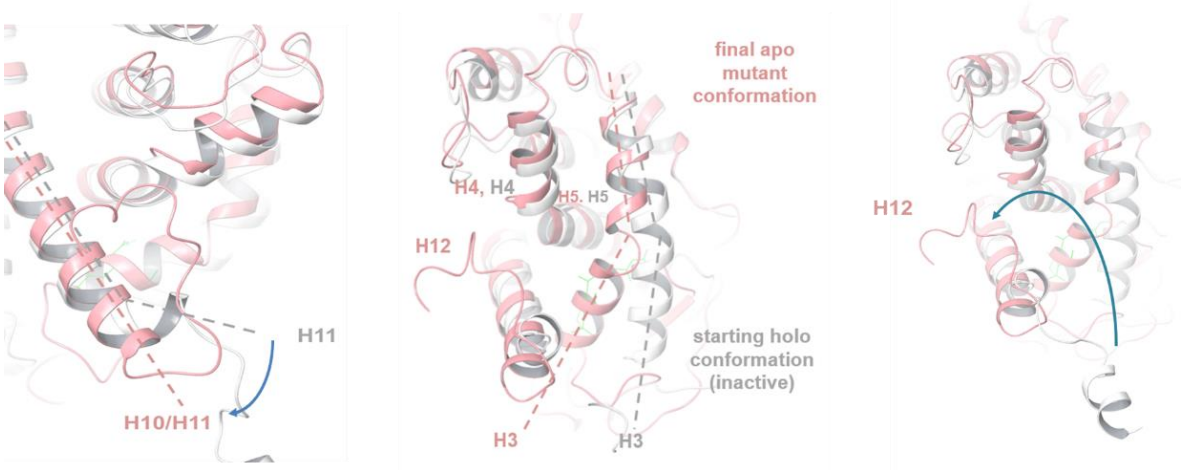

Figure 10: The conformational change of helices 3,10,11 and 12 in the mutant RXR-RAR heterodimer without a ligand [23]**.**

Biological experiments investigated the transcriptional activity of the RXRα-RAR heterodimer [40]. It was found that, in the presence of the RAR ligand, the transcriptional output of the WT dimer is increased by 30%, and when an RXR agonist binds to the dimer the transcriptional activity is increased again. On the other hand, although the mutant RXRα-RAR, in the presence of the RAR ligand, is still activated, the addition of 9-cis RA (RXRα agonist) does not increase the transcriptional output compared to WT.

Additional experiments explored how the mutation affects the RXRα-RXRα homodimer activation and transcriptional output [40]. Luciferase assays experiments for both WT and mutant homodimers showed that in the absence of the agonist ligands the luciferase expression of the mutant RXR-RXR was approximately increased 3-fold compared to luciferase expression that was observed in WT. However, when agonists 9-cis RA bind

to the mutant RXRα-RXRα dimer the transcriptional activity remains the same compared to WT, which was increased 10-fold.

Moreover, reported analysis based on TCGA data-sets connected the mutant RXRα with its heterodimeric partner, the PPARs [36]. It has been found that RXRα mutant is implicated in a hyperactive PPAR pathway signaling. These findings suggest that the RXRα hot-spot mutation (S427F/Y) may affect 20-25% of human bladder cancer [35], as the mutant RXRα causes abnormally high activity in PPARs. PPARs, in turn, switch on genes that help cancer cells to grow and multiply.

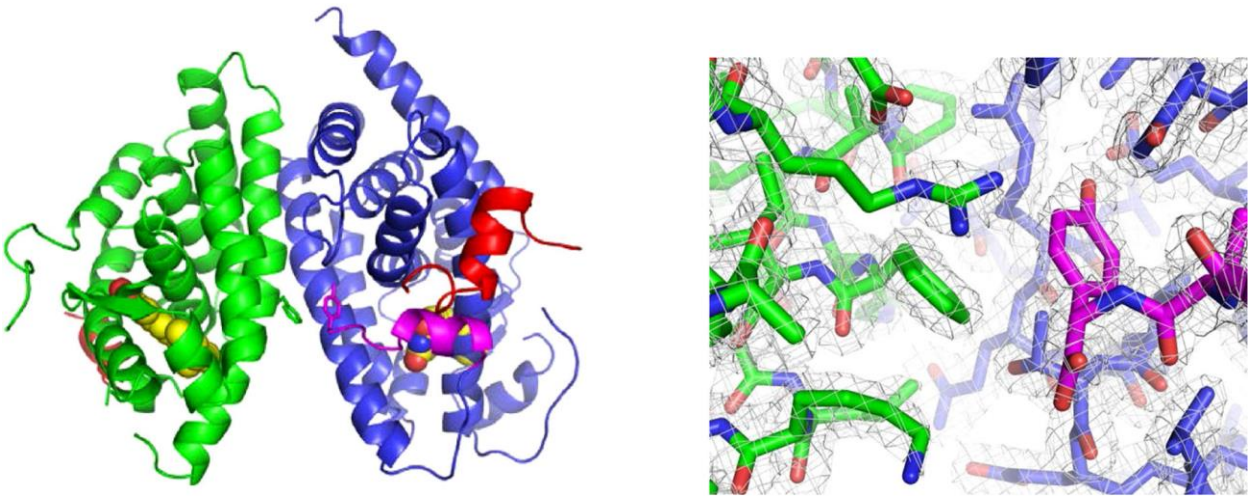

Figure 11: (left) Overall crystal structure of the heterodimer complex of mutant RXRα (green) and PPARγ (blue) with the co-activator peptide (red). The AF-2 helix (Helix H12) of PPARγ is shown in magenta. (right) Zoom in of the heterodimer interface shows the S427F mutation of RXRα (green) introduces a π-stacking interaction with Y477 of PPARγ (blue) at the C-terminal (magenta) [36]**.**

In order to support these findings, biomolecular experiments, in combination with computational techniques, such as Molecular Dynamics (MD) simulations explored how the single mutation from serine to phenylalanine at position 427 in RXRα resulted in structural and dynamical changes, which may be connected to altering the transcriptional activity of the RXRα-PPARγ dimer and consequently be implicated in cancer [35] [36]. They showed that the mutation S427F in RXRα allosterically regulates the AF-2 region (helix 12) of PPARγ via a π-π interaction between the F427 in RXRα and the terminal tyrosine of the C-terminal in PPARγ [36]. This means that the mutation may activate the RXRα-PPARγ dimer without the need of agonist ligand binding.

Figure 11 (left) shows the overall structure of the heterodimer complex of the mutant RXRα^{S427F} and PPARγ with the coactivator peptide, whereas Figure 10 (right) illustrates the π-π interaction between the mutant residue PHE427 of RXRα and TYR477 of PPARγ.

Considering all the above, the S427F mutation in RXRα appears to affect differently the structure of the two heterodimers RXRα-PPARγ and RXRα-RARα. For the RXRα-PPARγ heterodimer, the mutation controls helix 12 allosterically by forming a π-π interaction with the terminal tyrosine of PPARγ (TYR477) resulting to an upregulation of the PPAR target genes. On the other hand, for the RXRα-RARα heterodimer, mutant residue PHE427 causes a decrease of the binding site in RXR subunit by regulating the motion of helix 3, thus the ligand 9-cis RA is not able to bind.

In order to study the mechanistic details of the S427F mutation on the structure and dynamics of RXRα in complex with its different partners, the inter and intra-molecular interactions that occur between the biological molecules need to be studied in detail. These interactions are introduced in the next section.

1.4 Intra and Inter-molecular Interactions

Forces that are developed between molecules and other types of particles (atoms or ions), including attraction or repulsion, are called intermolecular interactions, while forces that occur within the same molecule such as covalent bonds are called intramolecular interactions.

The major types of intermolecular interactions are dipole-dipole interactions, van der Waals forces and hydrogen bonds [41] [42]. Dipole – dipole interactions occur from the electrostatic interactions of the positive and negative ends of molecules. The magnitude of the dipole corresponds to the strength of the interaction. Van der Waals forces emerge from the generation of immediate dipole moment in polar or non-polar molecules as a result of short fluctuations of the electron charge distribution. Moreover, hydrogen bonds are strong dipole-dipole interactions between molecules that are hydrogen bonded with an electronegative atom, such as O, N or F. The positively charged H atom (bond donor) interact strongly with a lone pair of electrons of a negatively charged O, N, or F atom (bond acceptor) (Figure 12).

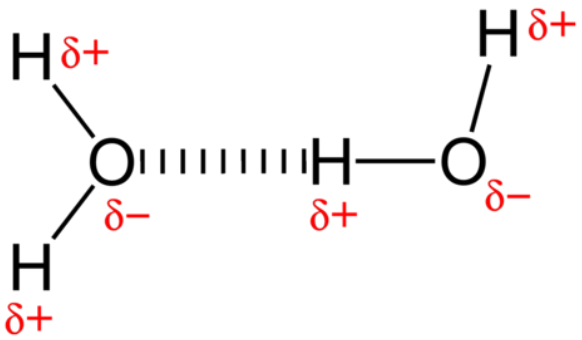

Figure 12: A hydrogen bond in water between the hydrogen atom of one water molecule and the lone pair of electrons on an oxygen atom of the neighboring water molecule [43]**.**

Another important type of intermolecular interactions is the π-π stacking which is an attractive interaction between aromatic rings as they contain pi bonds. For instance, two adjacent benzene atoms can form π-π interactions by being arranged in three possible configurations (Figure 13).

Figure 13: Three conformations of π-π interactions [44]**.**

The first two conformations, which are energy-friendly, are called parallel displaced and edge-to-face, while the third one is called sandwich and it is not so stable. This instability is due to the high electrostatic excitation of the electrons in the two-ring pendulums.

Within the same molecule, the interactions that are developed are called covalent, ionic and metallic bonds. Covalent bonds are formed between atoms with similar electronegativities [45]. Because both atoms have similar need for electrons there is little or no charge separation. A nonpolar covalent bond is formed between atoms with very small difference in electronegativity (less than 0.5) [42]. Whereas, a polar covalent bond is formed between atoms of slightly different electronegativities (between 0.5 and 1.9). For example, the hydrogen (O-H) bond in water H_2O [42]. Non-covalent bonds are formed by a complete transmission of valence electrons between atoms. Through this transfer two oppositely charged ions are generated [42]. Finally, metallic bonds occur between atoms of metals, in which the electrons are moving freely. These bonds are formed through the attraction of the mobile electrons and the positively charged metal ions [42].

1.5 Study Objectives

As previously mentioned, the single mutation on the human RXRα, S427F, which is found in 5-8% of patients with bladder cancer [37] [39] [38] affects differently the RXRα-RAR and RXRα-PPARγ dimers. Functional studies have shown that the S427F mutation activates, without the need of an agonist ligand, the heterodimer RXRα-PPARγ controlling the helix 12 allosterically by forming a π-π interaction with the terminal tyrosine of PPARγ (TYR477) [35] [36].

Computational and biological experiments observed how the mutation functions in RXRα-RAR heterodimer [23] and in RXRα-RXRα homodimer respectively indicating that mutation causes a decrease in of transcriptional activity in both cases. However, for the RXRα homodimer the mechanism of the mutation is still unknown.

To address the question of the effect of the S427F mutation on the RXRα homodimer and how its mechanism regulates the transcriptional activity of the dimer, we perform Molecular Dynamics (MD) simulations. Also, in order to validate our results with previous ones from the literature [35] [36], we perform MD simulations for the heterodimer RXRα-PPARγ. Finally, we use the Dynamical Network Analysis for both systems in order to explore the communications between their residues.

2. METHODS

In this Chapter, the theoretical background of the present study is presented. Firstly, an introduction to Molecular Dynamics (MD) simulations is outlined. Next, the algorithms employed for energy minimization and integration of the equations of motion are described. The principles of statistical mechanics, the basic theory behind MD simulations, are presented. Finally, the Dynamical Network Analysis, a method that explores the allosteric communication pathways within a biomolecule, is introduced.

2.1 Molecular Dynamics Simulations

MD simulations is a computational method that describes the time dependent performance of a molecular system. MD simulations can provide atomic-level information on the structure, dynamics and thermodynamics of biomolecules such as proteins and nucleic acids.

Alder and Wainwright were the first that introduced MD simulations in the late 1950's with a study on hard spheres [46]. Next, in 1974, Rahman and Stillinger conducted the first MD simulation of a realistic molecular system, whereas the first protein simulation was performed in 1977 by Karplus and his team [47]. Karplus, Warshel and Levitt received the Nobel Prize in Chemistry in 2013 in recognition of their work "for the development of multiscale models for complex chemical systems" [48] essentially acknowledging their contributions in MD simulations for the study of biological systems. As the Nobel Prize Committee mentioned in their announcement: "*Martin Karplus, Michael Levitt and Arieh Warshel laid the foundation for the powerful programs that are used to understand and predict chemical processes. Computer models mirroring real life have become crucial for most advances made in chemistry today. Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments*".

Nowadays, MD simulations are extensively used to study the biomolecular processes of proteins, protein-DNA complexes, and lipid systems, as well as for structure determination in experimental techniques such as X-ray crystallography or NMR. Below, we outline the basic principles behind the MD simulation technique.

2.1.1 Force Field: The Empirical Potential Energy Function

The energy of a system is represented by the Hamiltonian:

$$H = T + V \quad (2.1)$$

where T is the Kinetic Energy and V is the Dynamical (potential) Energy. In classical mechanics, the kinetic energy, T, of the system can be described by $\frac{1}{2}mv^2$; however, the description of the potential energy term, V, is not obvious. In order to describe the potential energy of a biomolecular system, intermolecular interactions need to be taken

into account. Thus, the potential energy, $V(r)$, can be defined as the non-bonded and bonded energy terms of the system [49] (see also section 1.4):

$$V(r) = V_{bonded} + V_{non-bonded} \quad (2.2)$$

In the empirical potential energy function, the non-bonded interactions are presented by the intermolecular van der Waals and Coulombic interactions, whereas the bonded terms involve the simple covalent binding in addition to the complex hybridization and π-orbital effects. In order to simulate the bonded and non-bonded energy terms of a molecular system in a classical mechanics framework, an empirical potential describing these interactions needs to be introduced with a set of equations. Such equations need to be simple, but at the same time describing the physicochemical nature of the system as well as computationally efficient. Thus, in MD simulations an empirical potential energy function or force field is used for the simulation of the system [50] [51]. Below, we outline the different equations for bonded and non-bonded interactions that comprise the potential energy function.

BONDED INTERACTIONS

Bond Stretching

In molecules, atoms are connected with covalent bonds which vibrate (Figure 14). In order to describe bond vibrations, a harmonic potential according to Hooke's law can be employed:

$$F = -kx \quad (2.3)$$

Taking into account that the force and the potential energy of the system are connected through the following formula:

$$F = -kx = -\nabla V \quad (2.4)$$

the potential of bond stretching becomes:

$$V = -kx^2 \quad (2.5)$$

which is expressed to describe the force constant k_b of the vibration of the spring and the displacement, b, from the equilibrium position, b_0, of the two bonded atoms:

$$V_{bonds} = k_b(b - b_0)^2 \quad (2.6)$$

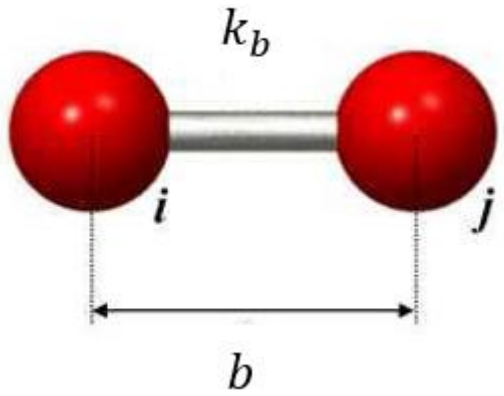

Figure 14: Definition of bond stretching between two atoms (i,j) [52]**.**

Bond Angle Bending

Bond Angles in molecules are designed to reproduce the bond geometry in molecules, which in reality is controlled by hybridization of electronic orbitals. When three covalently bonded atoms are present, a conformational change of the valence angles may ensue (Figure 15). This conformational change can be described again using Hooke's law (eq. 2.3) turning into an angle bending term, which as the bond stretching potential, is expressed as:

$$V_{angles} = k_\theta(\theta - \theta_0)^2 \qquad (2.7)$$

where θ is the angle between three atoms. The parameters that are used to describe each angle in the system are: the reference angle θ_0 and a force constant k_θ.

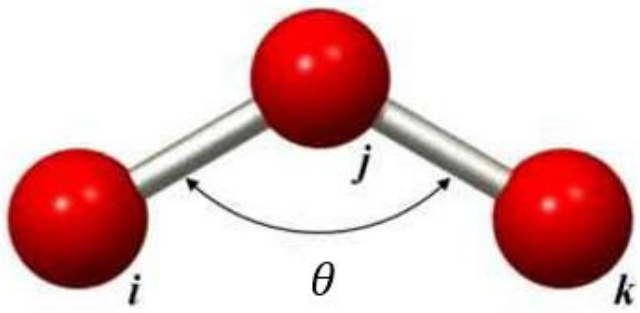

Figure 15: Definition of bond angle bending between three atoms (i,j,k) [52]**.**

Dihedral Torsion Term

Torsional terms represent rotations occurring between four covalently bonded atoms. The need for the proper torsional potential arises because bond-stretching or angle-bending potentials cannot describe the energetics of rotatable bonds, as they appear in four covalently bonded atoms (Figure 16).

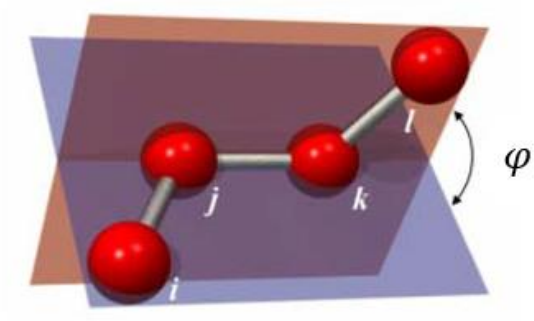

Figure 16: Definition of dihedral torsion between four atoms (i,j,k,l) [52]**.**

Torsional potentials can be expressed by a cosine function that represents the periodicity of the rotation along a covalent bond:

$$V_{dihedrals} = k_{\varphi}\left(1 - cos(n\varphi - \delta)\right) \qquad (2.8)$$

where φ is the angle between the planes formed by the first and the last three of the four atoms. The parameters that define this interaction are: δ, the minimum energy angle, $k\varphi$, a force constant and n the periodicity.

Improper Dihedral Term

The improper dihedral angles are used to impose the correct geometry or chirality of the atoms. If we have four atoms (i, j, k, l), and j is covalently bonded with i, l, k. The improper angle is defined as the angle between the $j - l$ and the plane $i - j - k$ (Figure 17).

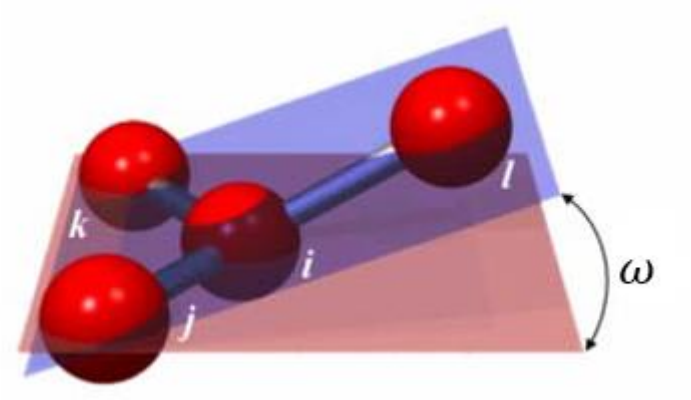

Figure 17: Definition of improper dihedral for four atoms (i,j,k,l) [52]**.**

The improper potential is expressed by a harmonic function:

$$V_{impropers} = k_{\omega}(\omega - \omega_0)^2 \qquad (2.9)$$

where ω is the angle between the plane formed by the central atom and two peripheral atoms and the plane formed by the peripheral atoms.

Finally, the V_{bonded} term is the summation of the terms that correspond to the types of the atom movement (Figure 18):

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} \qquad (2.10)$$

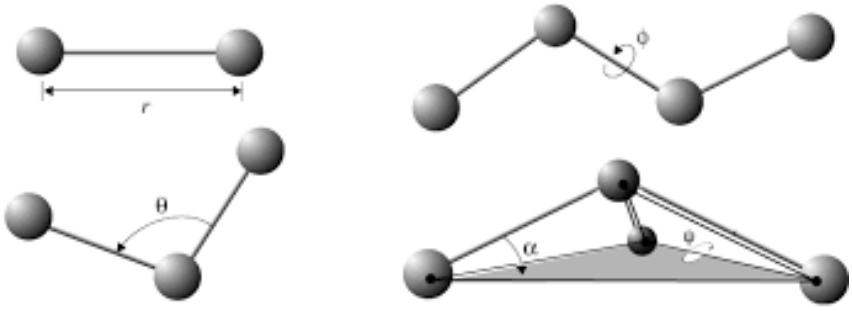

Figure 18: Symbolic representation of the bonded interactions: bond stretching r (upper left), bond angle bending θ (bottom left), proper dihedral φ (upper right) and improper dihedral ψ (bottom right) and the small out-of-plane angle α [53]**.**

NON-BONDED INTERACTIONS

Non-bonded interactions involve two types of interactions: the van der Waals interaction energy and the electrostatic interaction energy:

$$V_{non-bonded} = V_{vdw} + V_{electrostatic} \qquad (2.11)$$

These interactions are the most computationally demanding, as they include long-range interactions between the atoms in the system.

Van Der Waals Interactions

The van der Waals interaction appears from the balance between opposite (attractive) and same (repulsive) forces. Repulsion appears at short distances where the electron interaction is strong. Attraction, also known as dispersion force, appears when there are vacillations in the charge dispersion in the electron clouds. These vacillations on a single molecule results in a creation of a direct dipole which, consecutively, generates a dipole in a second molecule, causing an attractive interaction [54].

Van der Waals interaction has strongly repulsive effects at short lengths, whereas at medium lengths has weakly attractive effects. It is described by the Lennard-Jones potential:

$$V_{vdw} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (2.12)$$

where r is the distance between two atoms. It is characterized by the parameters σ, which is the collision parameter (the separation for which the energy is zero) and ε, which is the depth of the potential well (Figure 19).

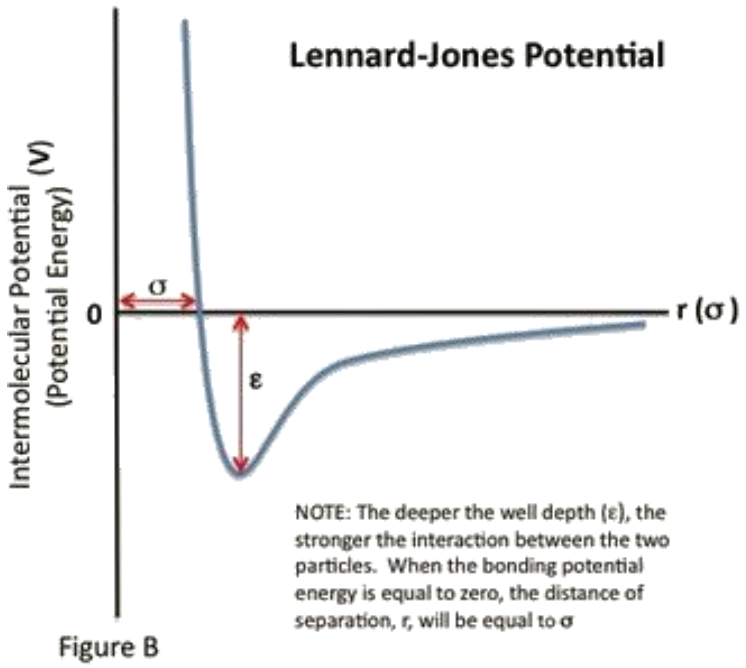

Figure 19: The Lennard-Jones potential: σ is the collision parameter and ε the well depth [55]**.**

Electrostatic Interactions

Electronegative elements are more attracted to electrons rather than the less electronegative elements, which consecutively create an unequal charge dispersion in a molecule. The most common way to illustrate this charge dispersion is by congregating fractional point charges all through the molecule. The charges are designed to replicate the electrostatic properties of the molecule, and the ones enclosed to the nuclear centers are known as partial atomic charges [56].

Coulomb's law describes the long-distance electrostatic interaction between two atoms:

$$V_{electrostatic} = \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}} \qquad (2.13)$$

Equation of the potential energy function

Finally, the equation of the potential energy function describing the force field is:

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{proper\ dihedrals} k_\varphi\big(1 - cos(n\varphi - \delta)\big) + \sum_{improper\ dihedrals} k_\omega(\omega - \omega_0)^2 + \sum_{vdw} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{electrostatic} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \qquad (2.14)$$

2.1.2 Energy Minimization

Prior to beginning a MD simulation, the system should undergo an energy minimization; through this process steric clashes between atoms, distorted bond angles or lengths, as well as any non-physical van der Waals contacts, will be relieved. Otherwise, these interactions may lead to an unbalanced simulation. At the energy minimum, the net force on each atom disappears [57]. Energy minimization aims to locate a local energy minimum in order to ensure that there is a realistic starting structure (Figure 20).

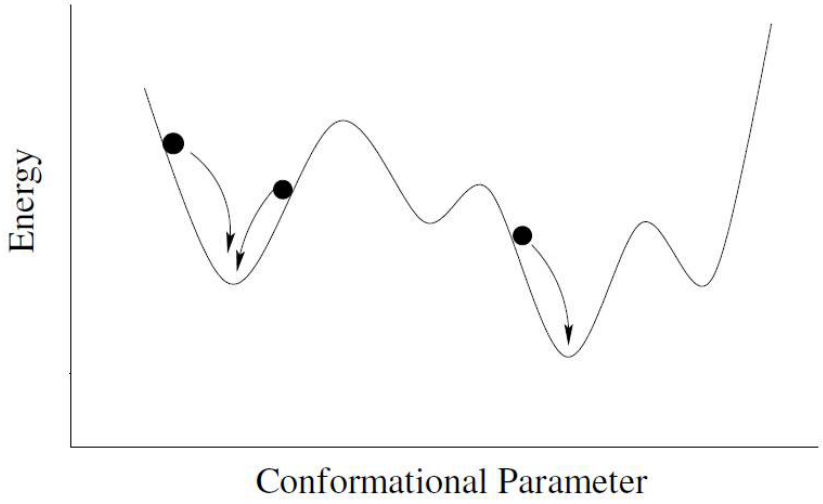

Figure 20: Representation of one-dimensional potential energy surface. Minimization methods move downhill to the nearest minimum [58]**.**

Common minimization algorithms use derivatives of the energy with respect to the initial coordinates, in order to locate the closest energy minimum. The direction is determined by the magnitude of the first derivative and in order to reach a minimum, a change in the coordinates is needed. The magnitude of the first derivative is a precise approach to indicate convergence. When derivatives are close to zero the system has reached a minimum [57]. To achieve that, the configuration must be constantly updated by modifying the coordinates (taking a step) and examining the convergence [57].

The three main algorithms that are used for the energy minimization are: Steepest Descent, Conjugate Gradient and Newton-Raphson algorithms. The appropriate selection of the method is based on the quantity of iterations essential to converge, but also on the quantity of function assessments necessary per iteration [59].

Steepest Descent

The Steepest Descent method makes use of the first derivative to form the direction towards the minimum.

$$F(r) = -\frac{dV(r)}{dr} \qquad (2.15)$$

It must be used in combination with a line search in order to predict the step size. The line search uses the direction vector obtained from the first derivative of the potential function, and in this way, locates the optimal step size close to this direction vector [59].

The moment this restricted minimum close to the direction of the derivative is settled, the step can be acquired. To find the best possible step size, the line search needs a huge number of function assessments. This strategy is strong, but it is not very efficient [59].

Conjugate gradient and Newton-Raphson

Conjugate Gradient and Newton-Raphson algorithms are more efficient methods for energy minimization. The Conjugate Gradient method takes advantage of information from earlier first derivatives to assume the most optimal direction for a line search, whereas the Newton-Raphson method utilizes the second derivatives along with the first ones to accomplish the same result, but with greater accuracy. The Newton-Raphson, also, makes use of the curvature to forecast the point where the gradient of the function will be changed into another direction [59]. For large systems the second derivative matrix is not recommended.

$$V(r) = V(r_0) + V'(r_0)(r - r_0) + \frac{V"(r_0)(r - r_0)^2}{2} + \dots$$

$$\frac{dV}{dr} = 0; \frac{d^2V}{dr^2} > 0 \qquad (2.16)$$

2.1.3 MD Formalism

After having defined the potential energy function and minimizing the energy of a system, the system is ready to be simulated with MD simulations. Here, we revise the principles behind MD simulations.

MD simulations are based on Newton's second law or the equation of motion:

$$F_i = m_i a_i = m_i \frac{\partial^2 r_i}{\partial t^2}, \; i = 1, \dots, N \qquad (2.17)$$

where m_i is the mass of each atom i and a_i is the acceleration on atom i.

Forces are the negative derivatives of the potential energy function $V(r_1, r_2, r_3, \dots, r_N)$ and can be written as:

$$F_i = -\frac{\partial V}{\partial r_i} \qquad (2.18)$$

Combining the two equations:

$$\frac{\partial V}{\partial r_i} = - m_i \frac{\partial^2 r_i}{\partial t^2} \qquad (2.19)$$

and solving this differential equation the trajectory of the system can be obtained.

The equations are solved synchronously in short time steps. The system is continued for some time, making sure the parameters such as the temperature and pressure maintain constant at the desired values. The coordinates, related to time, represent the trajectory of the system. Generally, the system reaches an equilibrium state and averaging the measures of the equilibrium trajectories, most of the macroscopic properties can be obtained.

To begin a MD simulation, an initial structure of the system is needed. Most often, in biomolecular simulations, the initial configuration is an X-ray crystallography or NMR structure, that can be easily obtained by the Protein Data Bank.

The initial coordinates are defined from these experimental structures, whereas the velocities can be derived from a distribution, e.g. Maxwell – Boltzmann distribution, which is expressed as:

$$\frac{dN}{N} = \sqrt{\frac{m}{2\pi k_b T}}\, e^{\frac{-mu^2}{2k_b T}}\, du \qquad (2.20)$$

where dN/N is the fraction of molecules moving at velocity u to $u + du$, m is the mass of the molecule, k_b is the Boltzmann constant and *T* is the temperature [60].

The MD simulation method is deterministic, which means that when the positions and velocities of each atom are known, it can be easily assumed in which state the system is at any time. Although the MD simulations can be time-consuming and have high computational cost, the continuous evolution of technology provides faster and cheaper computers that can decrease the time and the cost. It has been reported that the time scale of solvated protein simulations is up to nanoseconds, although, simulations up to milliseconds have been recorded [61].

2.1.4 Methods for integrating Newton's equations of motion

As previously mentioned, the potential energy is a function that describes the atomic positions of all atoms in the system. This complexity does not allow to use an analytical solution in order to solve the equations of motion. Thus, the classical equations must be treated numerically. For this purpose, there have been developed a lot of algorithms that integrate the equations of motion. The choice of the appropriate algorithm should be made considering the computational efficiency, the permission of a long-time step and the conservation of energy. Some of those algorithms are described below [62].

All the integration algorithms predict the positions, velocities and accelerations using the Taylor series expansion:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots$$
$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \ldots$$
$$a(t+\delta t) = a(t) + b(t)\delta t + \ldots \quad (2.21)$$

where r is the position, v is the velocity (first derivative with respect to time), α is the acceleration (second derivative with respect to time), etc. [62] Also, δt is restricted by the fastest vibration of the system (the C-H bond) and in most cases $\delta t = 1\ fs = 10^{-15}\ s$.

The total force on each particle in the structure at time t is the vector sum of its interactions with other particles. From the force, the acceleration of the particles is determined, and it is combined with the positions and velocities at time t in order to be calculated at time $t + \delta t$. The force remains constant during the time step (Figure 21).

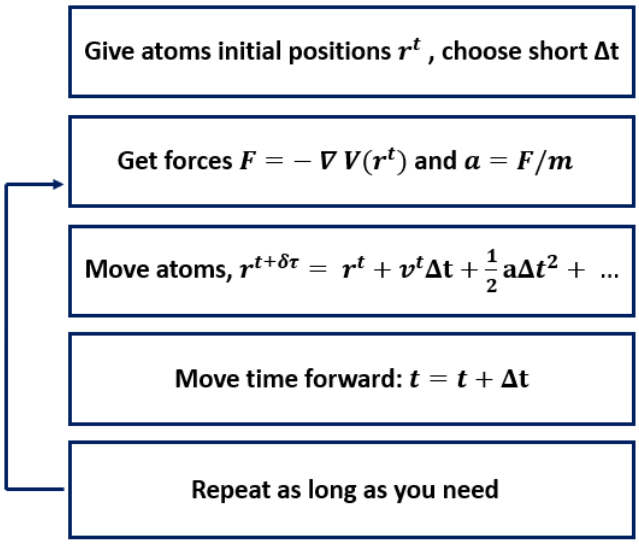

Figure 21: Iterative process in MD simulations

Verlet Algorithm

The most common integration algorithm is the Verlet algorithm, which uses two equations:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots$$
$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots \quad (2.22)$$

and the sum of these two is the basic formalism of this algorithm:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 \qquad (2.23)$$

The Verlet algorithm makes use of the positions (r) and accelerations (a) at time t and the new positions at time t + δt are calculated from the positions at time t – δt. The velocities can be calculated from this equation:

$$v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} \qquad (2.24)$$

This algorithm requires small amount of storage memory, however it generates a lot of errors. [62]

Leap-Frog Algorithm

Another integration algorithm is the Leap–Frog algorithm, which uses these two equations:

$$r(t+\delta t) = r(t) + v\left(t+\frac{1}{2}\delta t\right)$$
$$v\left(t+\frac{1}{2}\delta t\right) = v\left(t-\frac{1}{2}\delta t\right) + a(t)\delta t \qquad (2.25)$$

Specifically, it calculates the velocities (v) at time t + 1/2δt, and these velocities are used to calculate the positions (r) at time t + δt. This algorithm computes the velocities directly, but not at the same time with the positions. It, also, generates less errors than the Verlet algorithm. [62]

2.1.5 Introduction to Statistical Mechanics

MD simulations provide microscopic information of the systems. The microscopic state of a system is defined in a 6N-dimensional space, that involves the momentum, p, and the spatial coordinates r of the N particles. The macroscopic state is described by a small set of parameters, such as the temperature T and pressure P [63].

The basic theory behind MD simulations is Statistical Mechanics; through its framework the macroscopic information of a biomolecular system (internal energy, temperature, etc.) is generated from microscopic properties using statistical ensembles. A statistical ensemble is a set of representative points in the 6N-dimensional phase space. Ensembles present a collection of all possible systems which have different microscopic states corresponding to identical macroscopic (thermodynamic) states [64].

The Boltzmann factor

At the heart of much of physical chemistry lies the Boltzmann's law:

$$\frac{n_i}{n_j} = e^{-(\varepsilon_i - \varepsilon_j)/kT} \quad (2.26)$$

where n_i, n_j are the population in energy states i and j, $\varepsilon_i, \varepsilon_j$ represent the energy in each state, T is the temperature and k is the Boltzmann constant ($k = 1.38\ x\ 10^{-23}\ JK^{-1}$) [65].

For the lowest energy state ($\varepsilon_j = \varepsilon_0$) the equation becomes:

$$n_i = n_0 e^{-(\varepsilon_i)/kT} \quad (2.27)$$

where the $e^{-(\varepsilon_i)/kT}$ is the Boltzmann factor [66].

The molecular and canonical partition functions

The molecular partition function determines how particles distribute themselves over the accessible states [67]. It is known as molecular partition function per particle and is defined as:

$$q = \sum e^{-\varepsilon_i/kT} \quad (2.28)$$

The magnitude of the partition function shows how easily particles spread over the available quantum states. The above treatment applies to quantum statistical mechanics (discrete states) [68].

In classical mechanics, the canonical partition function, as its analogue (the molecular partition function), is defined as:

$$Q_N = \sum_{system\ states} e^{-\varepsilon_i/kT} \quad (2.29)$$

However, the position and momentum variables of each particle can vary continuously, so the set of microstates is actually uncountable. Thus, the best way to describe the partition function is to use an integral rather than a sum [68]. For this purpose, the microcanonical partition function is described as:

$$Q = \frac{1}{h^{3N} N!} \int d^{3N} q d^{3N} p \quad (2.30)$$

Then, the macroscopic thermodynamics will be connected to the microscopic via this equation:

$$S = k_b lnQ \quad (2.31)$$

Some of the different statistical ensembles with various characteristics are described below.

NVE – Microcanonical Ensemble

In NVE ensemble the parameters that are kept constant are the Number of particles, Volume and Energy of the system. This is also known as the “natural” ensemble for the MD simulations, where the equations of motion can be employed to the system without any changes [64]. To convert to other statistical ensembles, the integration of other equations, in place of classical equations, is needed, in order the sampling to be conducted in another statistical ensemble. The equations are the following (see also sections 2.1.1, 2.1.3):

$$H = K + V = constant$$

$$K = \sum_i \frac{1}{2} m_i v_i^2 \ , V = \sum_{i,j} U(r_{ij})$$

$$\frac{dr_i}{dt} = \frac{\partial H}{\partial p_j} = v_j \ , \frac{dp_i}{dt} = \frac{\partial H}{-\partial r_j} = \frac{\partial V}{-\partial r_j} = F_j \quad (2.32)$$

NVT – Canonical Ensemble

In NVT ensemble the constant parameters are the Number of particles, Volume and Temperature of the system. This ensemble is also known as the "Isothermal-isochoric" [64]. Since the temperature should be maintained constant, a thermostat must be applied. The Hamiltonian of the systems becomes:

$$H = K + V + K_s + V_s = constant \quad (2.33)$$

where K is the kinetic energy, V is the potential energy, and K_s and V_s are the kinetic and potential energies coupled to the thermostat:

$$K_s = \frac{1}{2} Q p_s^2$$

$$V_s = (f + 1) k_b T ln(s) \quad (2.34)$$

where p_s is the conjugate thermostat momentum, Q the conjugate mass, s the generalized thermostat coordinate and f the number of degrees of freedom.

The most common algorithm that keeps the temperature constant is the Nose-Hoover Constant Temperature algorithm, in which the equations are the following [69]:

$$\dot{r} = \frac{p_i}{m_i}$$
$$\dot{p_\iota} = F_i - \frac{p_\eta}{Q} p_i$$
$$\dot{\eta} = \frac{p_\eta}{Q}$$
$$\dot{p_\eta} = \sum_i \frac{{p_i}^2}{m_i} - dNkT \quad (2.35)$$

where η is the thermostat position, p_η the thermostat momentum, T the temperature, N the number of atoms, k the Boltzmann constant and d the number of spatial dimensions. $Q = dNkT\tau^2$ and defines the time scale of the thermostat motion based on the parameter τ, which is selected corresponding the vibrational period [69].

NPT Ensemble

In NPT ensemble the pressure, and thus also the density, must be stabilized. The Number of particles, Pressure and Temperature are kept all constant. This ensemble is also referred to as the "Isothermal-isobaric" ensemble, and mimic experimental conditions more carefully [70]. The equations of motion become more complicated because of the additional barostat to the system [71]:

$$\dot{r_\iota} = \frac{p_i}{m_i} + \frac{p_\varepsilon}{W} r_i$$
$$\dot{p_\iota} = F_i - \left(1 + \frac{1}{N}\right) \frac{P_\varepsilon}{W} p_i - \frac{P_\eta}{Q} p_i$$
$$\dot{V} = \frac{dV p_\varepsilon}{W}$$
$$\dot{p_\varepsilon} = dV\,(P_{int} - P_{ext}) + \frac{1}{N} \sum_i \frac{{p_i}^2}{m_i} - \frac{p_\eta}{Q} p_\varepsilon$$
$$\dot{\eta} = \frac{P_\eta}{Q}$$
$$\dot{p_\eta} = \sum_i \frac{{p_i}^2}{m_i} + \frac{{p_\varepsilon}^2}{W} - (dN + 1)kT \quad (2.36)$$

where p_ε is a momentum conjugate to the logarithm of the volume, W is the mass parameter, $\varepsilon = \ln(v/v_o)$, P_{ext} is the external pressure and P_{int} the internal:

$$P_{int} = \frac{1}{dV}\left[\sum_i \frac{p_i^2}{m_i} + \sum_i r_i F_i - (dV)\frac{\partial U}{\partial V}\right] \quad (2.37)$$

In this way, p_ε becomes a barostat and makes the system to reach the equilibrium state of $< P_{int} > = P_{ext}$ [71].

2.1.5 Periodic Boundary Conditions

Biomolecules are usually simulated in water or liquid solutions and not in *vacuum* [72]. This is due to the importance of water on properties of biomolecules. However, typical MD simulations can conduct a limited number of particles in order to achieve a small computational cost due to the need to integrate Newton's equation of motion every 1 fs, which is dictated by the fastest vibration of the system, the C-H bond. As a result, most of the molecules would be divided on or near the perimeters of the system, which, consequently, leads to the inspection of unrealistic surface effects (Figure 22) [72].

To overcome these finite size effects, periodic boundary conditions (PBC) are commonly applied. PBC are used in MD simulations in order to inhibit these surface effects and mimic the properties of mass systems. With this method, the system is enclosed by its own replicas in all directions, to produce an infinite periodic frame of identical cells. The nitrogen atoms, only, located within the central cell are considered exceptionally; however, when one of the atoms moves out of the cell, an image particle enters from the adverse side to take the place of it [72].

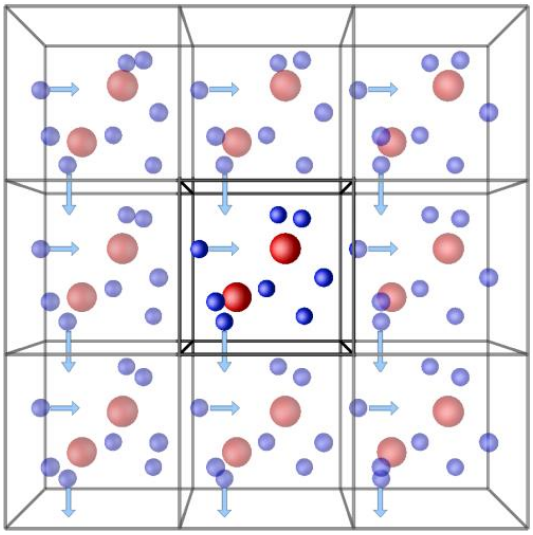

Figure 22: Schematic representation of the idea of periodic boundary conditions [73]**.**

For simulations of solvated macromolecules, the interactions between the molecule and its periodic images could result to unpleasant effects and should be avoided. Thus, a typical procedure is to build a water box around the macromolecule with the length of each box vector exceeding the length of the molecule plus double the cutoff radius.

Finally, these modeling conditions are effective in generating a more realistic representation of the in vivo environment than a water sphere enclosed in vacuum supplies [72].

2.1.6 Approximations in MD simulations

Although MD simulations have proven a valuable tool in studying biomolecular processes in atomic level detail [74] [75] [76], they do not come without approximations. It is important to be aware of these approximations in order to design meaningful simulations and correctly interpret their results.

Limits of the method are related to the parameter sets used, and to the underlying molecular mechanics force fields. As discussed in Section 2.1.1 MD simulations use interatomic potentials (force fields) that describe atomic interactions by employing functions parameterized with experimental or ab initio data (quantum mechanical calculations) to describe properties of the molecular systems. These potentials are thus less accurate and less transferable than ab initio calculations but are computationally less intensive enabling large-scale simulations (up to billions of atoms in current petaFLOP supercomputers). For example, hydrogen bonds are not explicitly represented in modern force fields but evaluated as Coulomb interactions of atomic point charges (Equation 2.13). This is a crude approximation as polarization effects are not represented and thus charge transfer events are neglected. Furthermore, the parameterization of the electrostatic interactions is usually in vacuum, although in reality biomolecules are simulated in aqueous solution bearing a much higher dielectric constant. Finally, van der Waals interactions in MD are described by Lennard-Jones potentials (Equation 2.12) that are static and cannot change based on the atom's environment. However, van der Waals forces are ultimately of electrostatic origin and therefore depend on dielectric properties of the environment [77]. The dependence of van der Waals forces based on their environment is neglected in standard MD simulations but could be included by developing polarizable force fields. Another important limitation of the empirical force field is that neither the bond making nor the bond breaking can be modeled. The force filed does not permit any changes in the electronic structure.

The Born-Oppenheimer approximation is another major approximation in MD simulations. According to it, motions of the nuclei are separated from the motions of the electrons. The physical basis behind this approximation is that the nuclei are much heavier than the electrons. Consequently, electrons move faster than the nuclei, and thus they may be treated separately [78]. Therefore, the heavy nuclei are modeled as point masses and their motion is described by classical mechanics.

The pair-wise additive approximation calculates the interaction energy between one atom and the rest of the system as a summation of pair-wise interactions. The synchronously interaction between three or more atoms is not computed, so the force field will not particularly involve specific polarization effects. This, in turn, generates some differences between simulated and experimental results. For example, there may be a variation in the computation of the experimental pK changes of ionizable amino acid residues produced by the electrostatic field of the protein [79].

Moreover, a fundamental axiom of statistical mechanics is the ergodic hypothesis [80] which states that the time spent by a system in some phase space of microstates with same energy is equal to the volume of this region. The basic theory is that if a system

evolves in time endlessly, it will eventually go through all possible states. Thus, MD simulations generate enough representative conformations in order to satisfy this equality [62].

Finally, as described in Section 2.1.5, in MD simulations PBC are used in order to approximate a large system. Through this process large systems include an infinite number of unit cells; one of these is the original simulation box and the others are the copies which called images. The PBC are applied in simulations in order to provide a more realistic biomolecular system.

2.2 Dynamical Network Analysis

In biological systems, allosteric communication is a long-distance interaction between sites within a protein. It considered to be very important, as most of the proteins display this behavior [81]. It has been proposed that distinct areas of the protein may communicate through networks of amino acids building up pathways of communication within the protein [82].

Network theory is an efficient strategy to analyze these pathways. A simple network includes a set of nodes and edges that connect pairs of nodes. For instance, in residue - residue interaction networks, nodes represent specific residues, and if two residues interact with each other, an edge is drawn between their nodes (Figure 23).

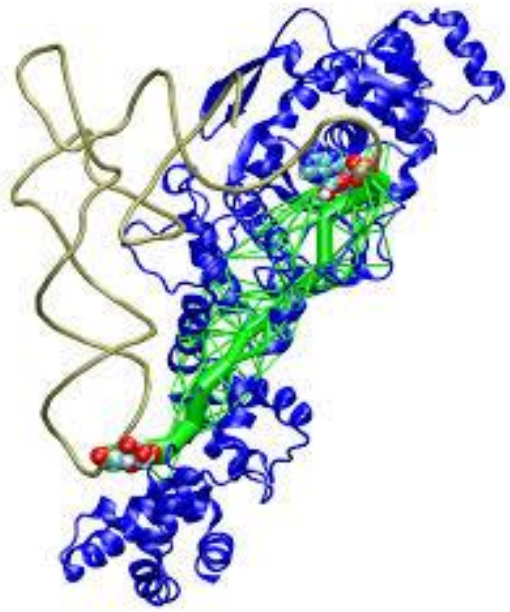

Figure 23: A snapshot of GluRS:tRNA:Glu-AMP complex (from T. thermophilus; pdb code: 1N78) in the active form. The subunit GluRS is shown in blue, tRNA in yellow and the network edges in green [83]**.**

A method for performing this type of analysis is the Dynamical Network Analysis, which explores the communication pathways within the molecule or the complexes based on the protein motions from an MD trajectory. Below, principles of the method along with the procedure for analyzing amino acid networks within a protein is described.

2.2.1 Network generation

First, the protein structure network is modeled in the following way. Residues (or sets of atoms) are defined as the nodes of the network and they are connected to each other by edges (Figure 23). The edges are based on the strength of the pair node interaction, which is the absolute value of their motional correlation, which is calculated throughout an MD trajectory. The motional pairwise correlations, C_{ij}, are expressed as:

$$C_{ij} = \frac{<\Delta r_i(t)\Delta r_j(t)>}{(<\Delta r_i(t)^2><\Delta r_j(t)^2>)^{1/2}}\ ,\ \Delta r_i(t) = r_i(t) - < r_i(t) > \quad (2.36)$$

where $r_i(t)$ is the position of the atom corresponding to the i^{th} node [83]. Motional pairwise correlations define the information transfer between the nodes, and edge weight, w_{ij}, is obtained from the probability of information transfer across a given edge:

$$w_{ij} = -\log(|C_{ij}|) \quad (2.37)$$

2.2.2 Communities generation

The dynamical network of nodes and edges that has been constructed as above can be clustered into subunits or communities of highly correlated regions: the nodes that are highly correlated (e.g. correlation above 0.5), based on their edge weight, belong to the same community (Figure 24). The community analysis (hierarchical clustering) is performed using the Girvan-Newman algorithm [84], in which the edge with the highest betweenness is removed and the betweenness between the remaining edges is recomputed in every step. The betweenness of an edge is expressed as the number of the shortest paths that go through the edge and is used as a measure for the significance of the edge communication. Each node is necessarily a part of a community even if it just a community of its own, but there are edges that lie between communities connecting the nodes of one community to those of another [83].

Figure 24: Network communities (shown in different colors) for GluRS-tRNAGlu [83]**.**

2.2.3 Critical Nodes and optimal/suboptimal paths

The nodes that connect different communities and bear the greatest edge betweenness are called “critical nodes” and “critical edges”, respectively. Because critical nodes and edges carry a large amount of information it is possible that they can transfer the allosteric signaling between communities [82].

Finally, in this analysis, it is assumed that the most biologically significant pathways are those that minimize the distance between two nodes in the network (optimal path). The length of this path, d_{ij}, is defined as the sum of the edge weights between the consecutive nodes (k, l):

$$d_{ij} = \sum_{k,l} w_{k,l} \quad (2.38)$$

where i and j are the distant nodes.

The shortest path, d_{ij}^{0}, is the most dominant form of the communication and it is calculated by the Floyd Warshall algorithm [85]. The suboptimal paths are described as paths that are slightly longer than the optimal path (Figure 25). So, it needs to be clarified how much long they can be. A given suboptimal path will not visit any node more than once [83].

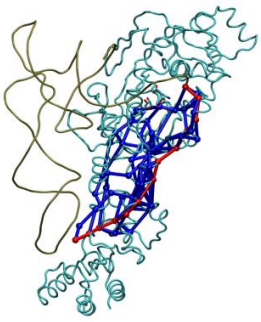

Figure 25: Suboptimal paths are shown (in blue color) between the base of U35 and the sugar of A76 on the tRNA. The optimal path is shown in red [83]**.**

3. RESULTS

In this Chapter, we address the question of the effect of the S427F mutation on the structure and dynamics of RXRα-RXRα and RXRα-PPARγ. Thus, we perform MD simulations to understand how the change of serine to phenylalanine at position 427 is affecting functional domains of the protein such as the dimerization interface, the binding pocket of 9-cis RA, helix 12 responsible for the activation of the complex and residues SER/PHE427 of the two complexes. We compare the results to RXRα-RARα heterodimer findings from previous unpublished work [23].

MD simulations of the heterodimer RXRα-PPARγ are conducted in order to validate our computations with previous work from the literature; [35].

3.1 The protein systems

The protein systems that are used for our study are the heterodimer RXRα-PPARγ and the homodimer RXRα-RXRα. It should be also noted that previous simulations of the RAR-RXRα are referred to in this thesis for comparison of the results; however, no simulations of the RAR-RXRα were performed.

In more detail, in order to investigate the effect of the S427F mutation in the RXRα, we create various systems for every dimer that correspond to the following compositions:

RXRα-PPARγ Heterodimer

i. The RXRα-PPARγ heterodimer in aqueous solution
ii. The RXRα-PPARγ heterodimer in aqueous solution with SER427 of the RXRα mutated into PHE427

RXRα-RXRα Homodimer

i. The RXRα-RXRα homodimer in aqueous solution
ii. The RXRα-RXRα homodimer in aqueous solution with SER427 of the RXRα mutated into PHE427 in one monomer (chain A)
iii. The RXRα-RXRα homodimer in aqueous solution with SER427 of the RXRα mutated into PHE427 in both monomers (chain A and chain C)
iv. The RXRα-RXRα homodimer in aqueous solution with the ligands 9-cis-RA bound to RXRα monomers
v. The RXRα-RXRα homodimer in aqueous solution with SER427 of the RXRα mutated into PHE427 in one monomer (chain A) and with the ligands 9-cis-RA bound to RXRα monomers
vi. The RXRα-RXRα homodimer in aqueous solution with SER427 of the RXRα mutated into PHE427 in both monomers (chain A and chain C) and with the ligands 9-cis-RA bound to RXRα monomers

From now on, systems (i) - (ii) of the RXRα-PPARα and (i) – (vi) of the RXRα-RXRα will be divided in two categories. The first category contains the systems (i) and (ii) of the RXRα-PPARγ dimer and (i) to (iii) of RXRα-RXRα dimer and will be called as “apo” and “apo-mutated” (due to the absence of the ligands). The second category contains the systems (iv) to (vi) and will be called as “holo” and “holo-mutated” (due to the presence

of the ligands). The steps that we followed in order to create those systems are explained below.

3.1.1 Systems Preparation for MD simulations

RXRα-RXRα Homodimer

For the setup of the RXRα-RXRα we use the crystal structure of the RXRα-RXRα homodimer formed by the human RXR alpha LBD bound to an agonist ligand BMS 649 and a coactivator peptide (PDB code 1MVC) [86]. The agonist ligands are replaced with 9-cis RA agonists using the Glide - Ligand Docking tool of the Schrödinger suite [87].

Initially, we generate the dimer structure from the asymmetric unit of the 1MVC crystal structure using the PyMOL software [88]. Also, in the “apo” systems, the ligands 9cis-RA and the co-crystalized water molecules at distance higher than 5 Å away from the ligands are deleted. For the “holo” systems, the ligands and the water molecules are maintained during the whole setup of the systems.

According to the PDB file of the crystal structure of the homodimer [86], several residues that connect functional regions of the receptor were not crystallized (Table 1). Missing residues located at N and C termini of the receptor are not mentioned, as they will not be included in the simulations due to the lack of solid data regarding their secondary structure conformation.

Table 1: Missing residues of the crystal structure [86]**, whose position and conformation were modeled using the Prime software** [89] **of the Schrödinger suite** [90]**.**

RXRα	**RXRα**
LYS 245	LYS 245
THR 246	THR 246
GLU 247	GLU 247
THR 248	THR 248
TYR 249	TYR 249
VAL 250	VAL 250
GLU 251	GLU 251
ALA 252	ALA 252
ASN 253	ASN 253
MET 254	MET 254
GLY 255	GLY 255
LEU 256	LEU 256
ASN 257	ASN 257
PRO 258	PRO 258
SER 259	SER 259
SER 260	SER 260
PRO 261	PRO 261
ASN 262	ASN 262

The prediction of the position and conformation of the missing region LYS245-ASN262 in the RXRα homodimer is performed using the Prime 4.2 tool [89] [91] of the Schrödinger suite [90]. The same protocol is followed in order to model this region in both of the “apo(-mutated)” and the “holo(-mutated)” systems.

RXRα-PPARγ Heterodimer

For the setup of the RXRα-PPARγ systems we use the 2.1 A resolution crystal structure of the heterodimer of the human RXRα and PPARγ ligand binding domains bound with 9-cis retinoic acid and rosiglitazone along with co-activator peptides. It should be noted that in order the simulation to be consistent with the structure that was used in the literature [36] all the co-crystallized ligands and the coactivator peptides are removed. Next, the same procedure as described above is followed.

The output structures of the two dimers are then post processed. Hydrogen atoms that are missing from the crystal structure are added using Maestro (Schrödinger LCC) [90]. The chain termini are left uncapped. The protonation state of the residues is predicted using the PROPKA algorithm [92]. Then, the output structures are energy minimized using the OPLS3 force-field [93].

The introduction of S427F mutation is performed using Maestro 18.4 (Schrödinger LCC) [90]. The orientation of the side - chain of PHE after the mutation is predicted using the “Side-chain Prediction” tool of Prime 4.2 [89]. The prediction is based on a conformational analysis of the side - chain and the outcome corresponds to the first rotamer that does not produce a clash between PHE and the proximal residues. The mutated structures are, then, energy minimized using the OPLS3 force-field.

3.1.2 Systems Setup

In order to simulate the dynamic behavior of the protein parameters within the empirical potential energy function (Eq. 2.14) have to be assigned that are necessary to describe the interactions of the protein, water and ligand atoms. In this study, we use parameters from the Amber ff99SB*-ILDN [94] for the RXRα-RXRα homodimer and the Amber03ff [94] for the RXRα-PPARγ heterodimer. For the assignment of these parameters to all protein atoms, we used a previously created in house python script that allow us to make all the atom types, that were obtained by the Maestro software, compatible with the Amber force field.

Concerning the holo (-mutated) systems of the RXRα-RXRα homodimer, the necessary parameters that would allow us to simulate the 9-cis RA ligand, are obtained from previous work in Cournia’s lab.

Each complex is solvated into a cubic periodic box using water molecules modeled with the TIP3P potential [95]. The minimum distance between the protein atoms and the box walls is set to be 15 Å. Counterions are added to neutralize the net charges of the systems. Then each system is subjected to an energy minimization using the steepest descent method (see Section 2.1.2) until the maximum force is lower than 1000 kJ/(mol·nm).

3.1.3 Systems Equilibration

NVT equilibration

Energy minimization, which was performed in the previous step, ensures a reasonable starting structure, in terms of geometry and solvent orientation. In order to run an NPT MD simulation, we must then equilibrate the solvent and ions around the protein. Because energy minimization is performed at 0 K, first, the system needs to be heated to the temperature we wish to simulate.

Therefore, for the heating of the system and the equilibration of the solvent around the protein and protein-ligand systems we perform MD simulations for 1 ns for each system using the NVT ensemble (see Section 2.1.5). All MD simulations are performed using the parallelized MD program GROMACS 2018.6 [96]. During this step, the position of the heavy atoms of the protein molecules and the ligands are restrained by applying a constant force of 1000 kJ/(mol·nm) in all directions. The length of the covalent bonds is constrained using the LINCS (LInear Constraint Solver) [97], allowing integration step of 2 fs. Non-bonded interactions are cut off at 16.0 Å, and long-range electrostatic interactions are computed using the particle mesh Ewald (PME) method [98], with a B-spline interpolation of order 4. The simulated systems are coupled into an external temperature bath at 310 K (controlled with a Nose-Hoover thermostat – see Section 2.1.5) with a coupling constant of τ_T = 0.1 ps. Initial velocities are assigned to the atoms using the Maxwell distribution at 310 K.

NPT Equilibration

The previous step stabilized the temperature of the system. Before we collect the data, we must also equilibrate (relax) the system at this temperature. Equilibration is conducted under an NPT or “isothermal-isobaric” ensemble (see Section 2.1.5). During this step, the position of the heavy atoms of the protein are restrained by applying a constant force of 1000 kJ/(mol·nm) in all directions. The equilibration lasts 10 ns and is initiated from the final snapshot of the corresponding NVT equilibration. The simulated systems are coupled into an external temperature bath at 310 K with a coupling constant of τ_T = 0.1 ps and isotropic pressure coupling with time constant of τ_P = 2 ps is applied to keep the pressure at 1.013 bar (controlled with isotropic Parrinello-Rahman coupling method [99][100]).

3.1.4 Production Run

After the NPT equilibration, the systems are ready for the position restraints to be released and the production run to begin. This process uses the checkpoint file from the NPT in order to start the simulation from the final step of the equilibration (see Section 3.1.3). Tables 2 and 3, below, show exactly the duration of each simulation (it does not consider the time for the NVT and NPT equilibration).

All production runs are conducted in agreement with the simulations performed earlier [23]. In replica systems, different initial velocities are generated from a Maxwell distribution with a random seed and lasted for 300 ns each, as such timescale has been found to be sufficient to monitor the changes that the mutation provokes. Multiple simulations were performed in order to ensure that the results are reproducible.

Table 2: Simulation time for the production run for all RXRα-RXRα simulated systems

RXRα-RXRα System	Production run	
Apo – Wild Type	Replica 1	370 ns
	Replica 2	300 ns
	Replica 3	300 ns
Apo – Mutant	Replica 1	370 ns
	Replica 2	300 ns
	Replica 3	300 ns
Apo – Double Mutant	Replica 1	370 ns
	Replica 2	300 ns
	Replica 3	300 ns
Holo – Wild Type	Replica 1	370 ns
Holo – Mutant	Replica 1	370 ns
Holo – Double Mutant	Replica 1	370 ns

Table 3: Simulation time for the production run for all the RXRα-PPARγ simulated systems

RXRα-PPARγ System	Production run	
Apo – Wild Type	Replica 1	370 ns
Apo – Mutant	Replica 1	370 ns

After the production run of all systems, periodic boundary conditions were removed (Section 2.1.5) in order to proceed with the analysis. Also, the analysis of the trajectory for both dimers that is described in the next section is conducted using GROMACS tools. The representative structure is calculated using the GROMOS algorithm [101] and cutoff 1.5 A. This method counts the number of neighbors and takes the structure with the largest number of them with all its neighbors as cluster. Then it eliminates this structure from a large number of clusters, and it repeats for the remaining structures. Finally, the Dynamical Network Analysis is performed using the program Visual Molecular Dynamics (VMD) [102].

3.2 Trajectory Analysis for the apo – RXRα/PPARγ Heterodimer (WT and Mutant)

Below, the results from the trajectory analysis of the apo RXRα-PPARγ for WT and mutant systems are presented in detail. The effect of the mutation S427F at heterodimerization interface is tested. Also, our results are compared with the ones from the literature.

3.2.1 Convergence Analysis (density, pressure, temperature)

The time series of the temperature (Figure 26), pressure (Figure 27) and density (Figure 28) for the apo RXRα-PPARγ WT and mutant systems show that these properties have reached stable values. This means that these properties have converged, and the system has probably reached thermodynamic equilibrium; thus, we are ready to proceed with the analysis.

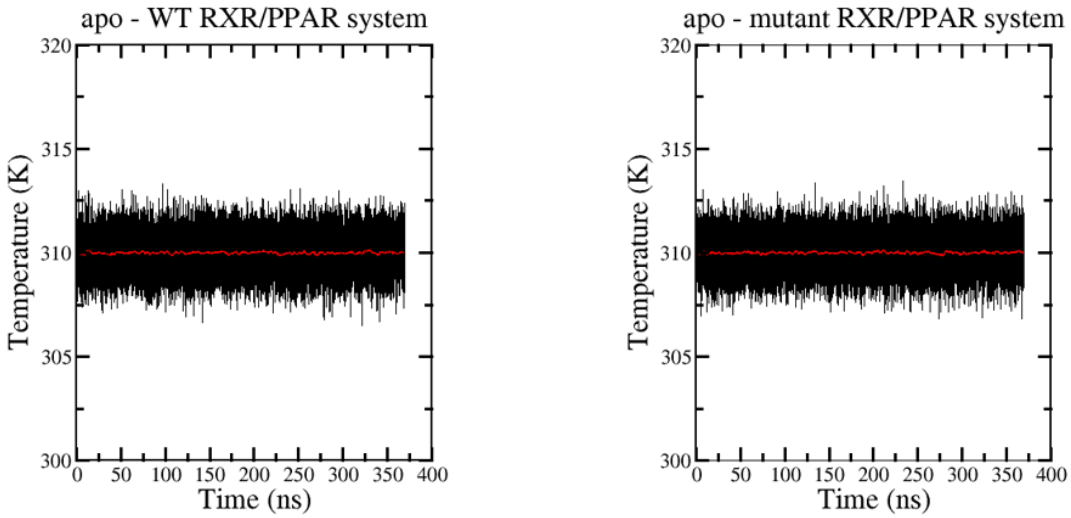

Figure 26: Temperature time series for the apo RXRα-PPARγ. Red lines are running averages over time.

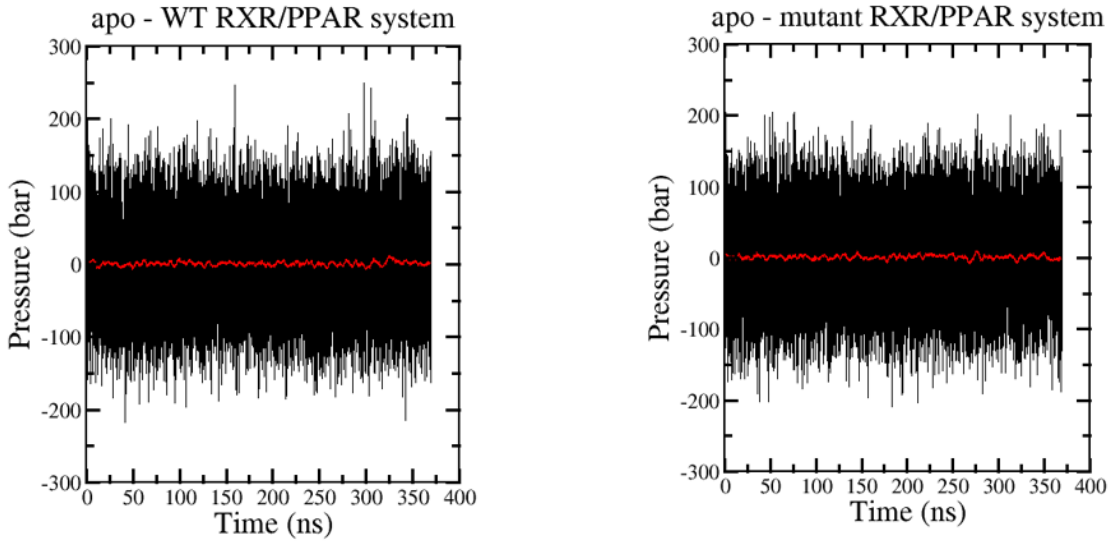

Figure 27: Pressure time series for the apo RXRα-PPARγ. Red lines are running averages over time.

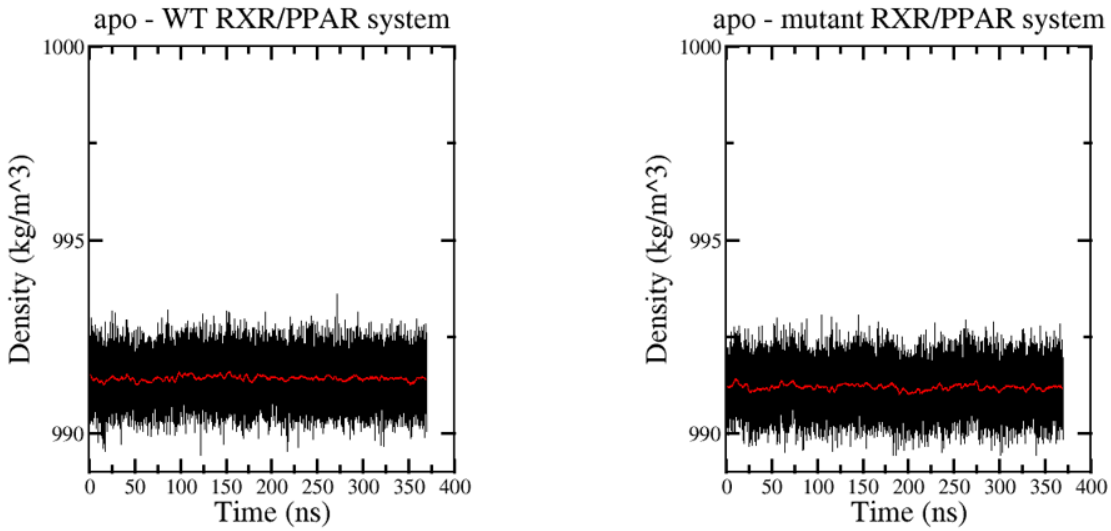

Figure 28: Density time series for the apo RXRα-PPARγ. Red lines are running averages over time.

3.2.2 Root mean square deviation of the monomers

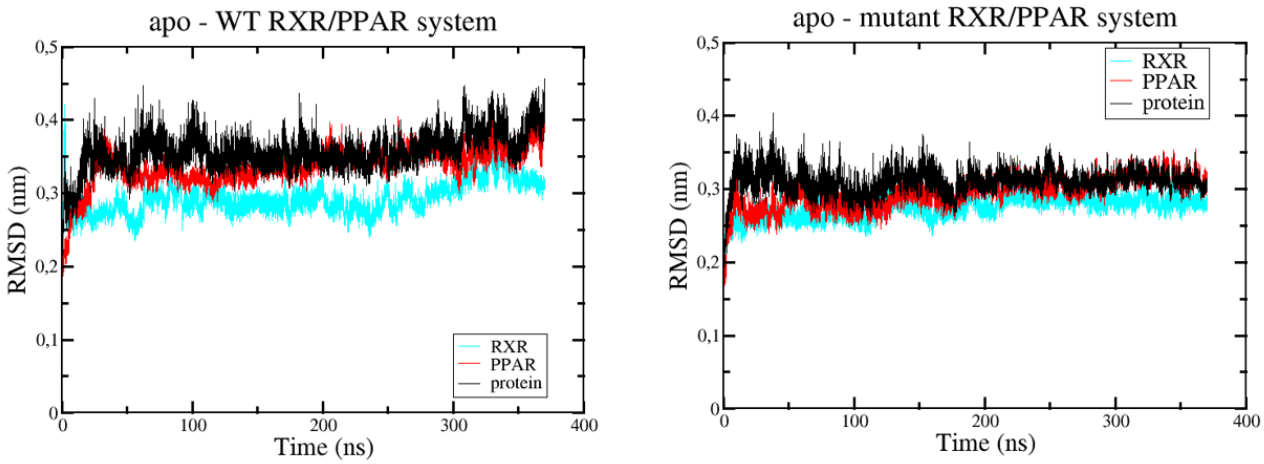

Figure 29: RMSD calculation for the apo RXRα-PPARγ systems. (Left) apo WT RXR-PPAR system. (Right) apo mutant RXR-PPAR system.

Another metric to monitor converge of the system is the root mean square deviation (RMSD) of the protein backbone atoms compared to the relaxed crystal structure (first frame of our calculation). The RMSD is defined as:

$$RMSD = \sqrt{\frac{\sum_{I=1}^{N}(r_i - r_j)^2}{N}} \qquad (3.1)$$

The RMSD time series for the monomers as well as for the whole protein in the apo WT and mutant dimers (Figure 29) illustrate that the systems display a uniform behavior over time. Therefore, the trajectory can be collected and further analyzed.

3.2.3 Distance of the center of mass of the monomers over time

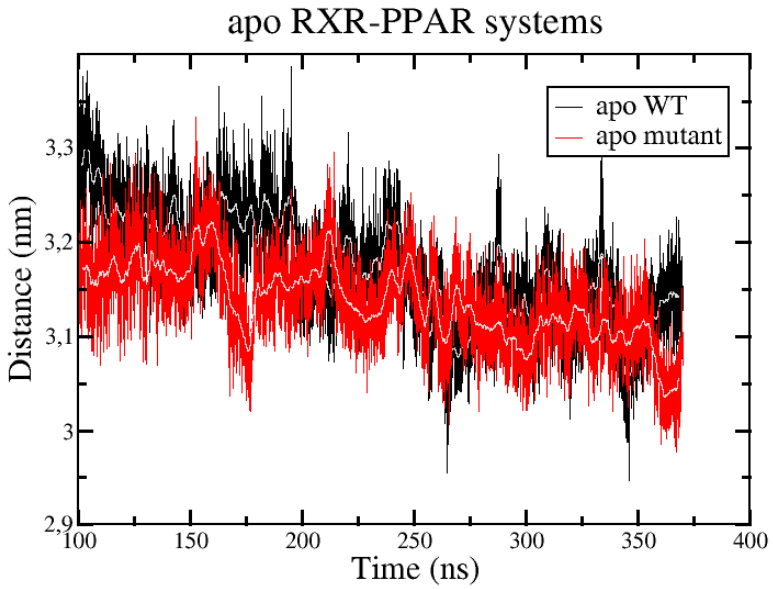

Figure 30: Time series of the distance of the center of mass between the monomers for the apo RXR/PPAR systems. Grey lines are running averages over time.

The average values derived from the calculation of the distance of the center of mass between the monomers (Figure 30) are 3.17 ± 0.06 nm for the apo WT RXRα-PPARγ and 3.13 ± 0.04 nm for the apo mutant. We observe that distance between the monomers is approximately the same. However, this property shows that the system has not yet reached its equilibrium and further simulation time may be needed.

3.2.4 Hydrogen bond analysis for the dimer

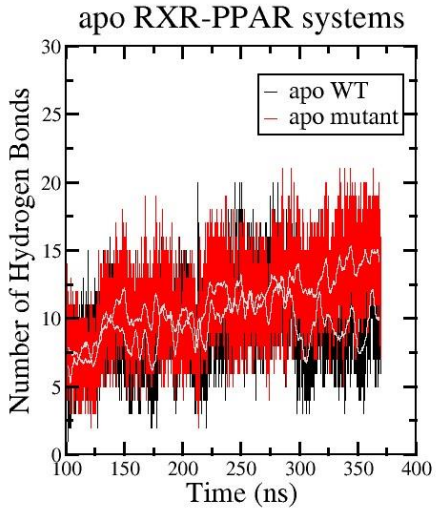

Figure 31: Calculation of the average number of hydrogen bonds between the monomers of the apo RXR/PPAR systems. Grey lines are running averages over time.

The average number of hydrogen bonds between the monomers (Figure 31), is slightly increased (11.1 ± 2.5) in the apo mutant system compared to the apo WT (9.52 ± 2.5) albeit this difference is within the standard deviation. This property also shows that the system has not yet reached its equilibrium and further simulation time may be needed.

3.2.5 Interaction energy between the monomers

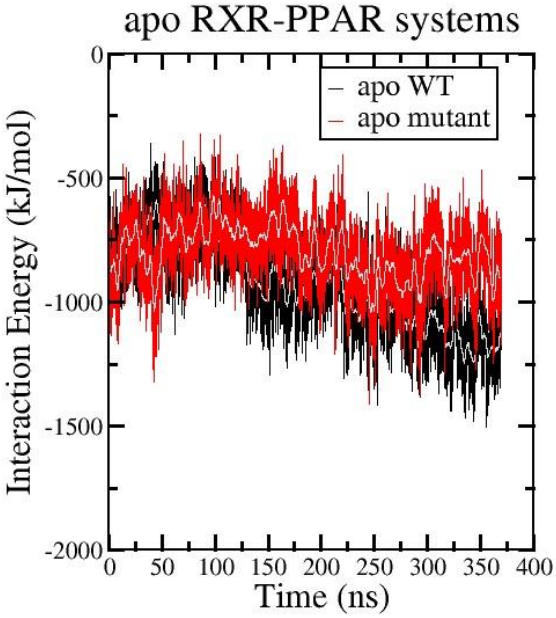

Figure 32: Interaction energy between the monomers of the apo RXR/PPAR systems. Grey lines are running averages over time.

The same behavior is present for the interaction energy between the monomers (Figure 32). In the apo mutant system the interaction energy is slightly increased; the average value for the apo WT is -810 ± 135 kcal/mol, whereas for the apo mutant is -987 ± 167 kcal/mol.

3.2.6 Solvent accessible surface area of the binding site

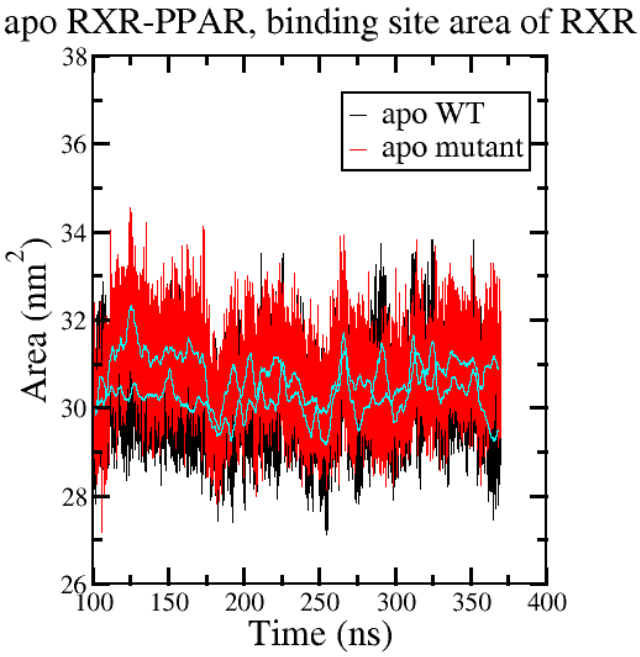

Figure 33: Solvent Accessible Surface Area of the binding site of the RXRα subunit for the RXRα-PPARγ systems. Grey lines are running averages over time.

Figure 33 shows that the solvent accessible surface area of the binding site of the RXRα subunit is the same for the WT and mutant systems.

3.2.7 Dynamical Network Analysis

The Dynamical Network Analysis for the apo RXR-PPAR system was used to identify the optimal paths and their weights (Equation 2.38) between the residues SER/PHE 427 and LEU455 of Helix 12, in order to explore how the mutant residue affects helix 12, which is the structural element of that controls the activation of RXRα (see Section 1.3.3).

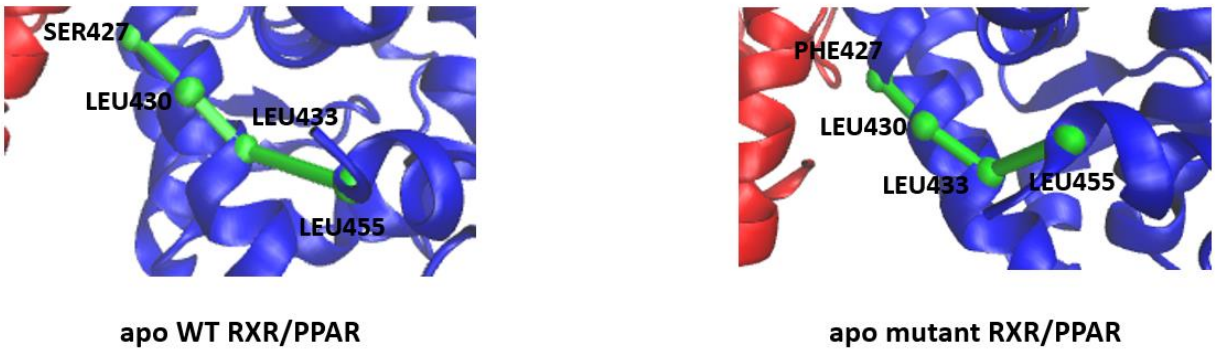

Figure 34: Optimal path between SER/PHE427 and Helix 12 (LEU455).

We observed that the optimal path between the residues SER/PHE427 and LEU455 of Helix 12 remains the same (Figure 34) and the total weight of the paths is 103 and 122 for WT and mutant respectively. Also, the optimal path between the critical node that connects the RXRα subunit with PPARγ and helix 12 of PPARγ is longer in the mutant (Figure 35). The total weight for the paths is 117 and 197 for WT and mutant, respectively. The weight of the WT path is significantly lower meaning that WT RXRα correlates more strongly to helix 12 of PPARγ, while the mutant RXRα has less control over helix 12 of PPARγ.

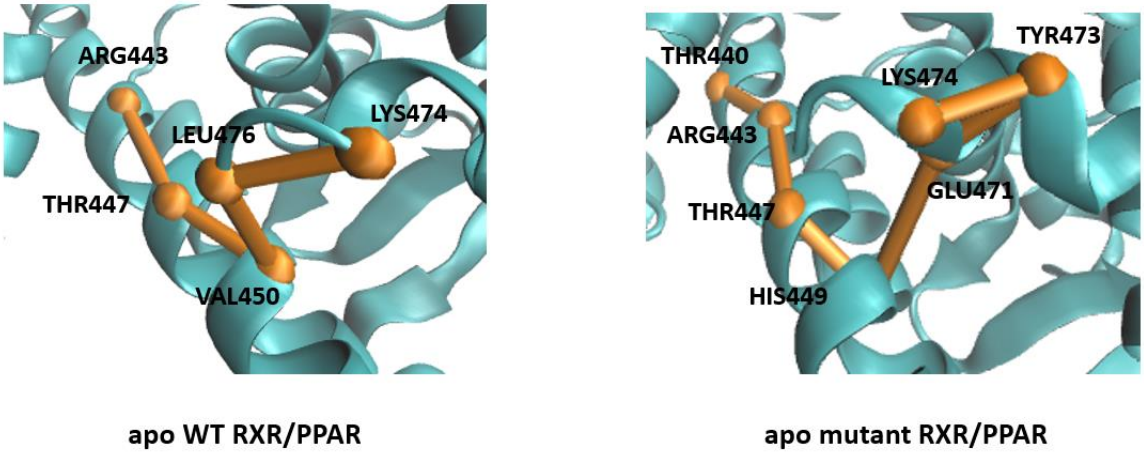

Figure 35: Optimal path between RXRα SER/PHE427 (critical node that connects RXRα with PPARγ) and PPARγ Helix 12 (LYS474).

Moreover, after visualizing the trajectory, we observed that for the first 100 ns the most dominant conformation between the mutant residue PHE427 and the tyrosine terminal of the PPARγ (TYR477) is a π-π interaction (Figure 36). The same interaction was found in the literature [35][36] for the mutant heterodimer RXRα-PPARγ.

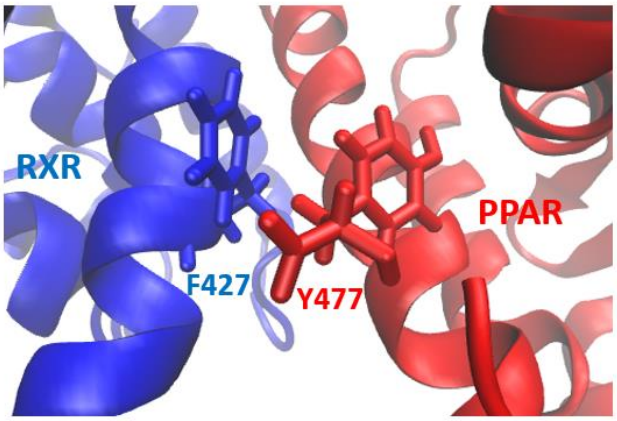

Figure 36: π-π interaction between the two benzene rings of residues PHE427 of RXRα and TYR477 of PPARγ. The conformation shown is the dominant conformation extracted from the first 100 ns of the simulation.

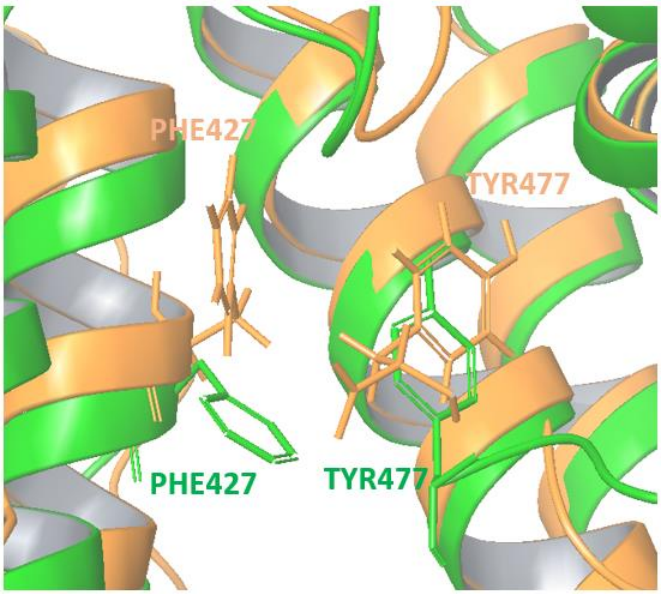

Figure 37: Comparison with the crystal structure of the π-π interaction between the residues PHE427 of RXRα and TYR477 of PPARγ. Simulated apo mutant RXRα-PPARγ is shown in orange and crystal structure of the apo mutant RXRα-PPARγ [36] **(PDB 5JI0) is shown in green.**

In Figure 37 is shown the comparison of the π-π interaction between the simulated apo mutant and the mutant crystal structure from the literature [36].

3.3 Trajectory Analysis for the apo – RXRα Homodimers (WT and Mutants)

In this section, the results from the trajectory analysis of the apo RXRα homodimer for WT, mutant and double mutant systems are explained. The effect of the S427F mutation at homodimerization interface, the solvent accessible surface area of the binding pocket of 9-cis RA, helix 12 responsible for the activation of the complex and residues SER/PHE427 is presented.

3.3.1 Convergence Analysis (density, pressure, temperature)

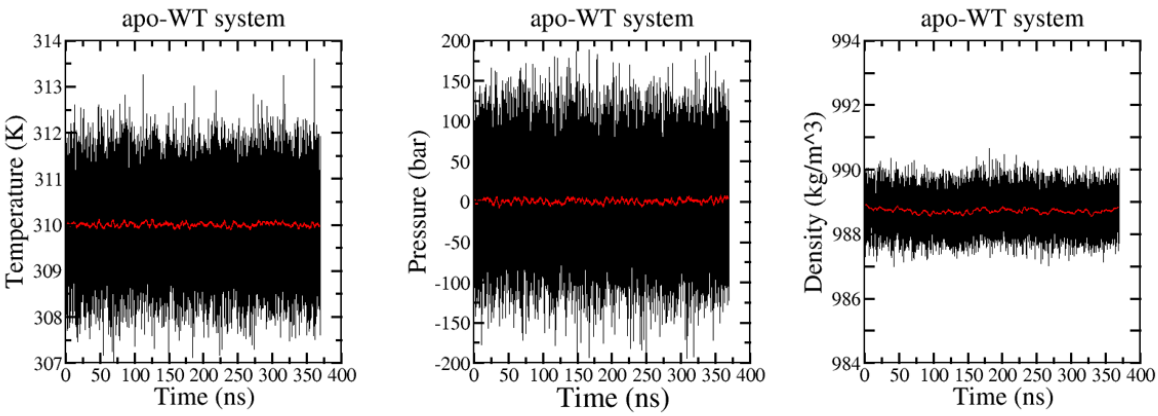

Figure 38: Temperature, Pressure and Density time series for the apo – WT RXRα homodimer. Red lines are running averages over time.

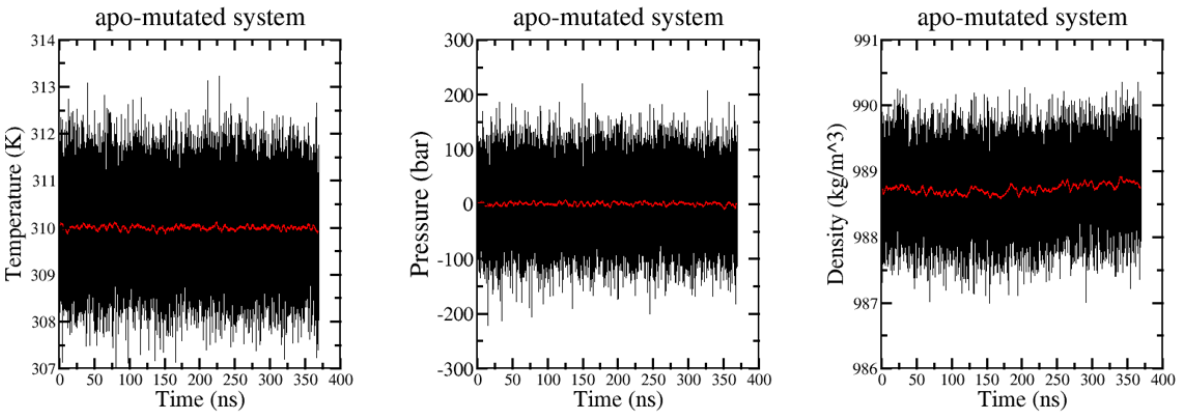

Figure 39: Temperature, Pressure and Density time series for the apo – mutant RXRα homodimer. Red lines are running averages over time.

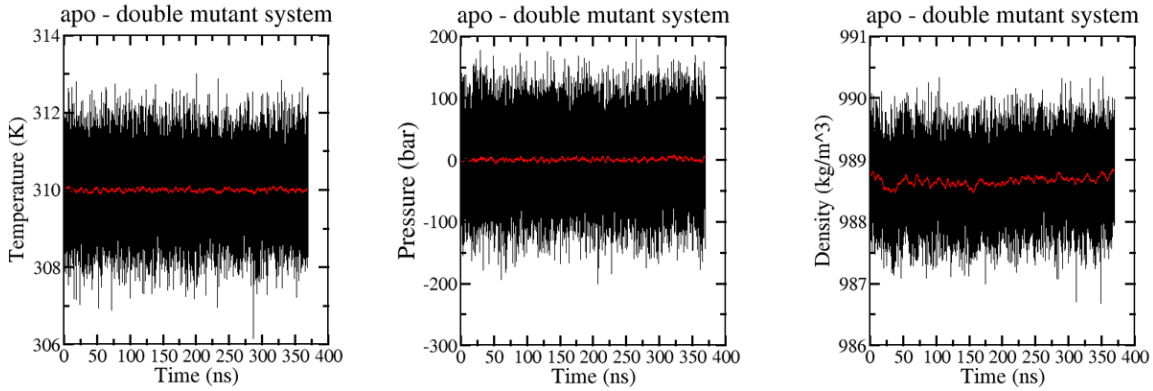

Figure 40: Temperature, Pressure and Density time series for the apo – double mutant RXRα homodimer. Red lines are running averages over time.

The time series of the temperature (Figure 38), pressure (Figure 39) and density (Figure 40) for the apo RXRα-RXRα WT and mutant systems show that these properties have reached stable values. This means that these properties have converged, and the system has probably reached thermodynamic equilibrium; thus, we are ready to proceed with the analysis.

3.3.2 Root mean square deviation of the monomers

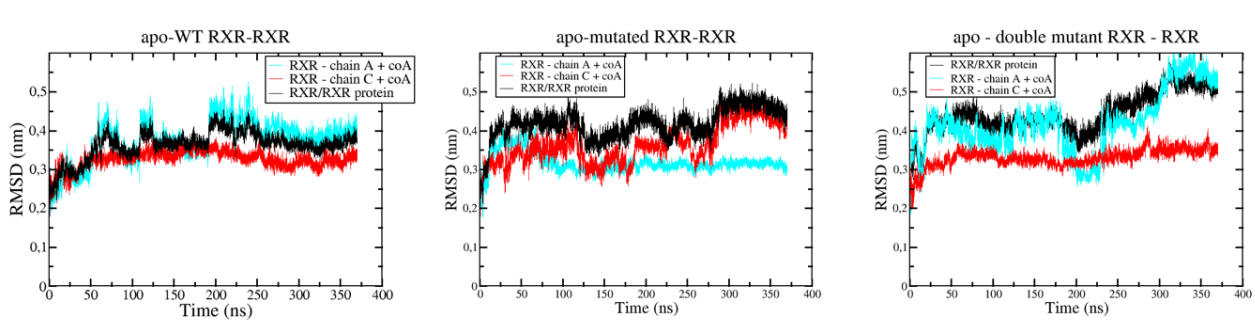

Figure 41: RMSD time series for the apo RXRα systems. (From left to right) – apo WT system, apo mutant system, apo double mutant system. In all graphs the RMSD of the protein is shown in black color, chain A in cyan and chain C in red.

The RMSD time series shows the performance of the apo monomers as well as of the whole protein (Figure 41) over time. It seems that the simulation time was enough to continue the analysis of the trajectory.

3.3.3 Distance of the center of mass between the monomers over time

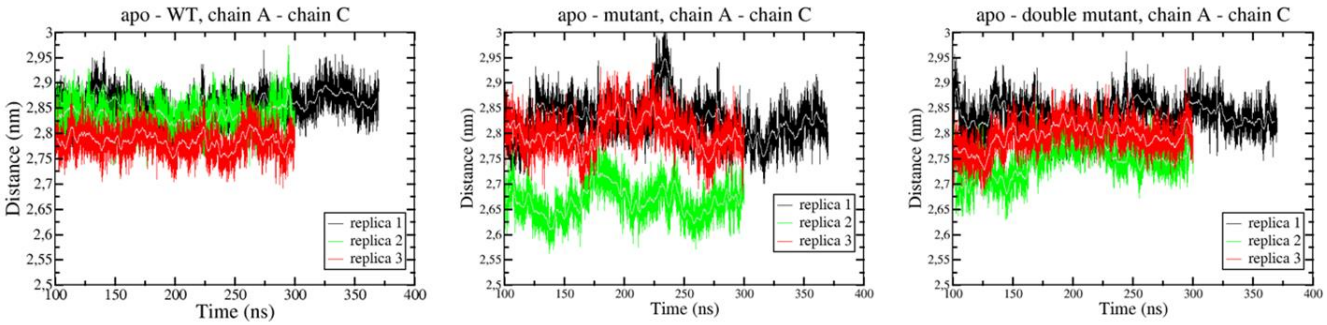

Figure 42: Time series of the distance of the center of mass between the monomers for the apo RXRα systems (all replicas). Grey lines are running averages over time. (From left to right) – apo WT systems, apo mutant systems, apo double mutant systems. In all graphs the distance between the monomers of replica 1 is shown in black color, replica 2 in green and replica 3 in red.

According to the graph (Figure 42) of the distance between the center of mass of the monomers, in the mutant systems the distance is slightly decreased. Specifically, Table 4 shows all the average distance values of all the apo systems.

Table 4: The average values of the distance of the center of mass between the monomers for the apo RXRα systems (all replicas).

System	Center of Mass of Monomers Distance (nm)	
Apo – Wild Type	Replica 1	2.85 ±0.02
	Replica 2	2.83 ± 0.02
	Replica 3	2.78 ± 0.02

Apo – Mutant	Replica 1	2.83 ± 0.04
	Replica 2	2.68 ± 0.03
	Replica 3	2.80 ± 0.03
Apo – Double Mutant	Replica 1	2.83 ± 0.02
	Replica 2	2.74 ± 0.03
	Replica 3	2.79 ± 0.03

3.3.4 Hydrogen bond analysis for the dimer

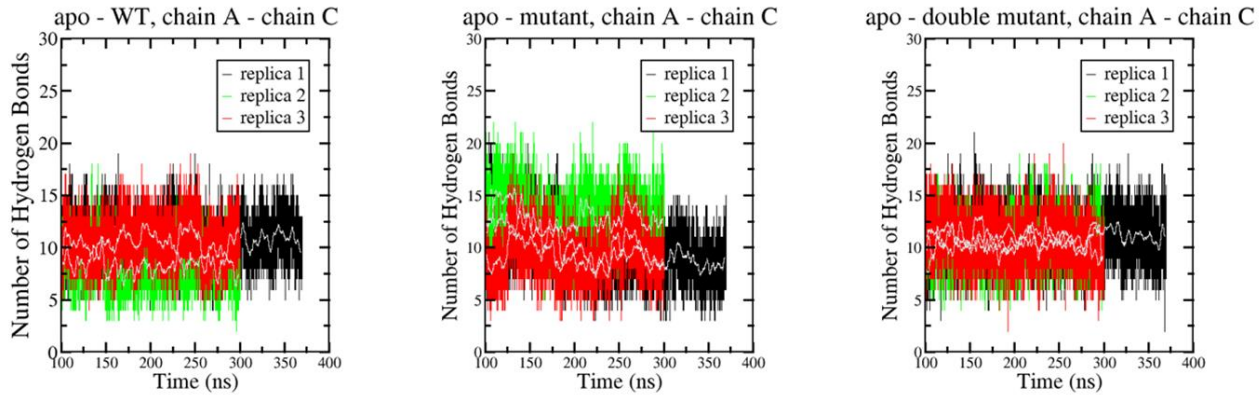

Figure 43: Calculation of the average number of the hydrogen bonds between the monomers of the apo RXRα systems. Grey lines are running averages over time. (From left to right) – apo WT systems, apo mutant systems, apo double mutant systems. In all graphs the average number of hydrogen bonds between the monomers of replica 1 is shown in black color, replica 2 in green and replica 3 in red.

The hydrogen bond analysis (Figure 43) along with Table 5, showed that the apo WT and mutant systems present small differences between the average values of the number of hydrogen bonds between the monomers.

Table 5: The averages values of the average number of hydrogen bonds between the monomers for the apo RXRα systems (all replicas).

System	Average Number of Hydrogen Bonds	
Apo – Wild Type	Replica 1	10.7 ± 1.8
	Replica 2	8.3 ± 1.9
	Replica 3	10.5 ± 1.9
Apo – Mutant	Replica 1	9.93 ± 2.5
	Replica 2	12.8 ± 2.3
	Replica 3	9.5 ± 2.0
Apo – Double Mutant	Replica 1	10.9 ± 2.0
	Replica 2	10.4 ± 1.9
	Replica 3	10.5 ± 1.9

3.3.5 Interaction energy between the monomers

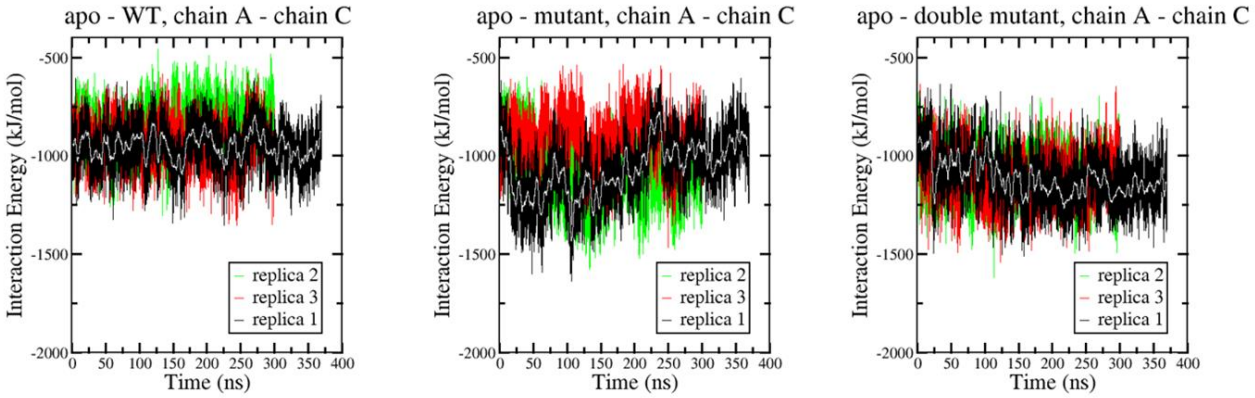

Figure 44: Interaction energy between the monomers for the apo RXRα systems. Grey lines are running averages over time. (From left to right) – apo WT systems, apo mutant systems, apo double mutant systems. In all graphs the interaction energy between the monomers of replica 1 is shown in black color, replica 2 in green and replica 3 in red.

Regarding the interaction energy (Figure 44), the apo mutant systems depict a slightly increase in the average values. Table 6 presents all the average values of all apo systems.

Table 6: The averages values of the interaction energy between the monomers for the apo RXRα systems (all replicas).

System	Interaction Energy (kcal/mol)	
Apo – Wild Type	Replica 1	-952 ± 100
	Replica 2	-796 ± 101
	Replica 3	-943 ± 106
Apo – Mutant	Replica 1	-1043 ±150
	Replica 2	-1094 ± 148
	Replica 3	-921 ± 120
Apo – Double Mutant	Replica 1	-1145 ± 105
	Replica 2	-1094 ± 100
	Replica 3	-1124 ± 109

The next section illustrates all the calculations of the solvent accessible surface area of the binding site, helix 12 and residues SER/PHE 427.

3.3.6 Solvent accessible surface area of the binding site, Helix 12 and residues SER/PHE 427

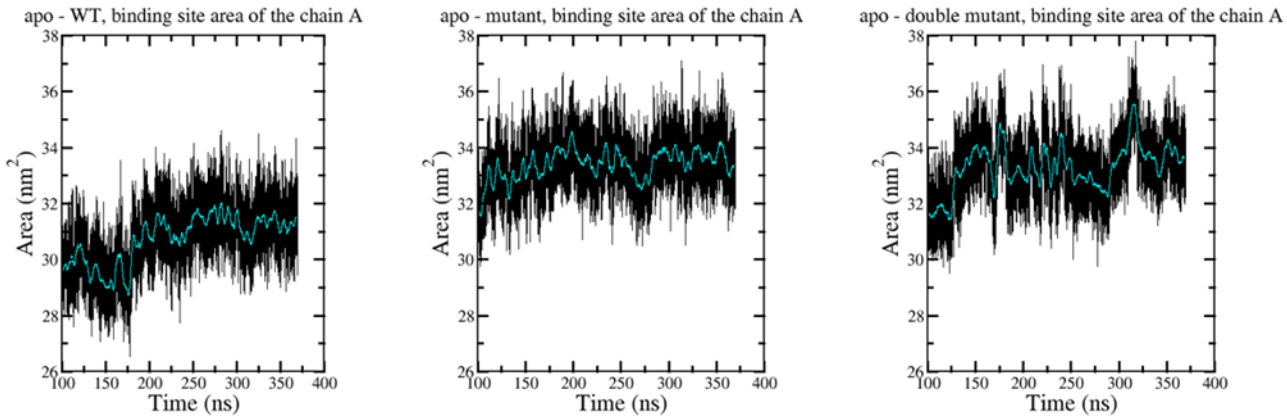

Figure 45: Solvent Accessible Surface Area for the binding site of the chain A of the apo RXRα systems. Cyan lines are running averages over time. (From left to right) – apo WT system, apo mutant system, apo double mutant system.

Concerning the binding site in chain A, Figure 45 shows that the solvent accessible surface area is increased. The average values for the systems are: 30.74 ± 1.1 nm^2, 33.39 ± 0.8 nm^2 and 33.29 ± 1.0 nm^2 for the apo WT, mutant and double mutant respectively.

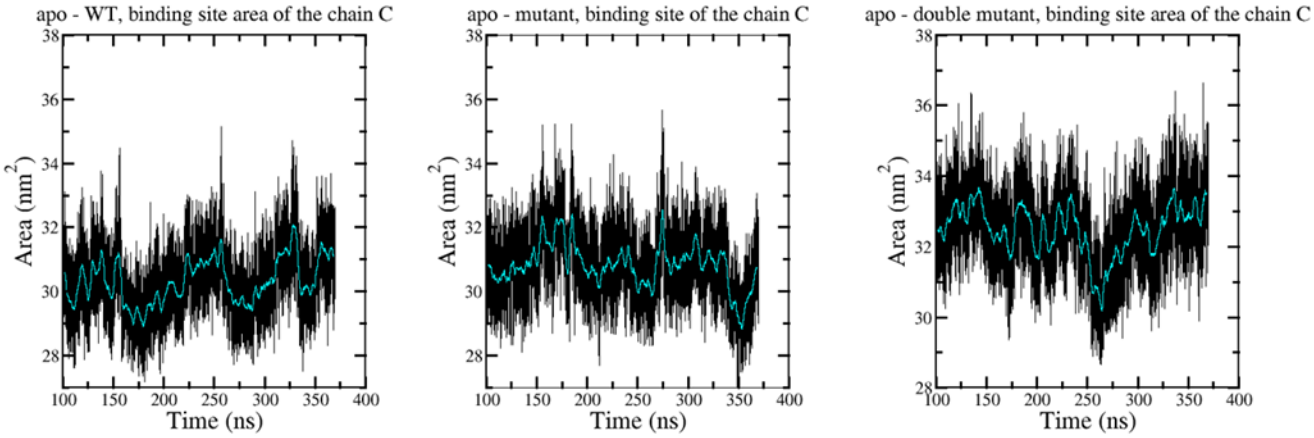

Figure 46: Solvent Accessible Surface Area of the binding site of the chain C of the RXRα systems. Cyan lines are running averages over time. (From left to right) – apo WT system, apo mutant system, apo double mutant system.

Regarding the binding site in chain C, it presents the same behavior (Figure 46). The average values are: 30.33 ± 1.0 nm^2, 30.82 ± 1.0 nm^2 and 32.49 ± 1.0 nm^2 for the apo WT, mutant and double mutant systems respectively.

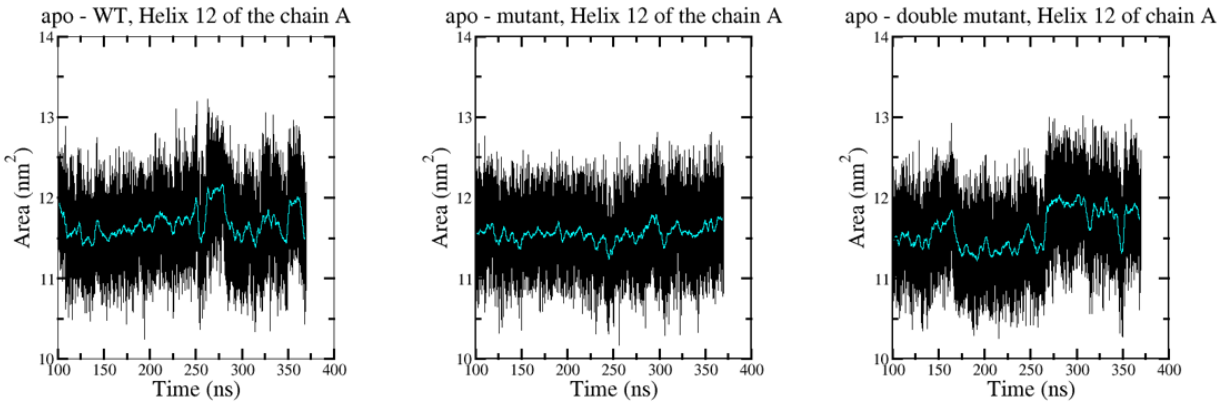

Figure 47: Solvent Accessible Surface Area of the Helix 12 of the chain A of the RXRα systems. Cyan lines are running averages over time. (From left to right) – apo WT system, apo mutant system, apo double mutant system.

Moreover, the calculation of the solvent accessible surface area of helix 12 shows small differences in the mutant systems (Figure 47). For chain A, the average values are: 11.67 ± 0.3 nm^2, 11.55 ± 0.3 nm^2 and 11.60 ± 0.4 nm^2 for the apo WT, mutant and double mutant systems, respectively.

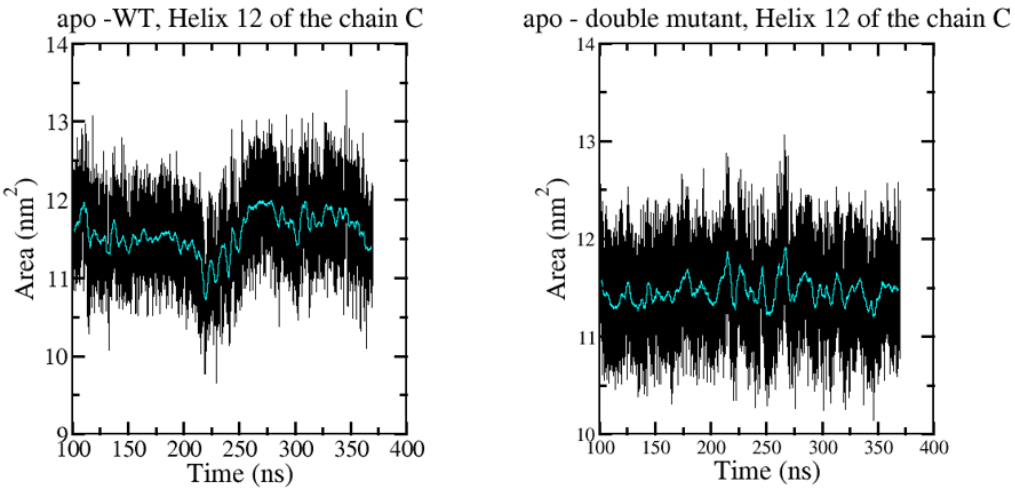

Figure 48: Solvent Accessible Surface Area of Helix 12 of chain C of the RXRα apo WT (left) and double mutant systems (right). Cyan lines are running averages over time.

For chain C (Figure 48) in the apo WT the average value is 11.58 ± 0.4 nm^2, whereas in the apo double mutant system is 11.46 ± 0.3 nm^2.

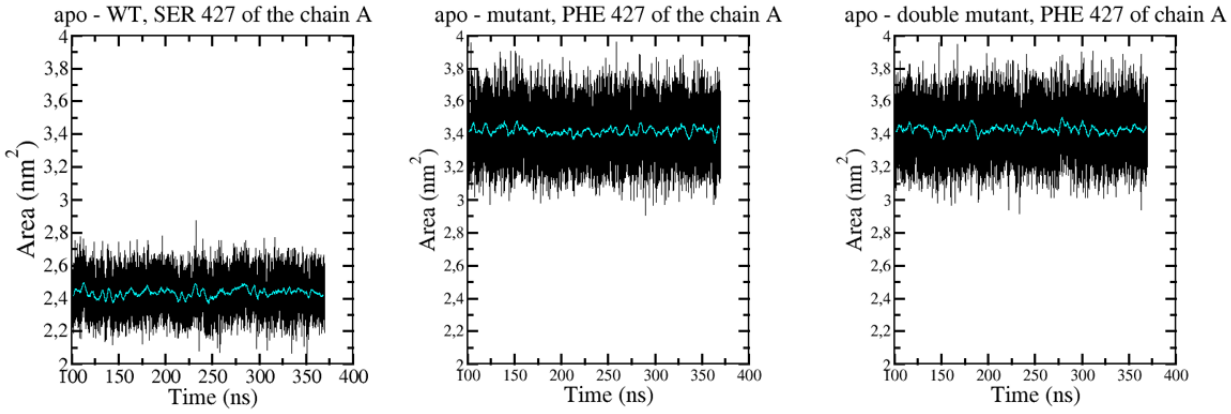

Figure 49: Solvent Accessible Surface Area of the residue SER/PHE427 of chain A of the RXRα apo systems. The cyan lines are the average values over time. (From left to right) – apo WT system, apo mutant system, apo double mutant system.

Finally, the calculation of the solvent accessible surface area of residues SER/PHE427 in chain A (Figure 49 shows that the solvent accessible surface area of SER427 is lower than that of PHE427). The average values are: 2.43 ± 0.09 nm^2, 3.42 ± 0.1 nm^2 and 3.43 ± 0.1 nm^2, for the apo WT, mutant and double mutant systems respectively.

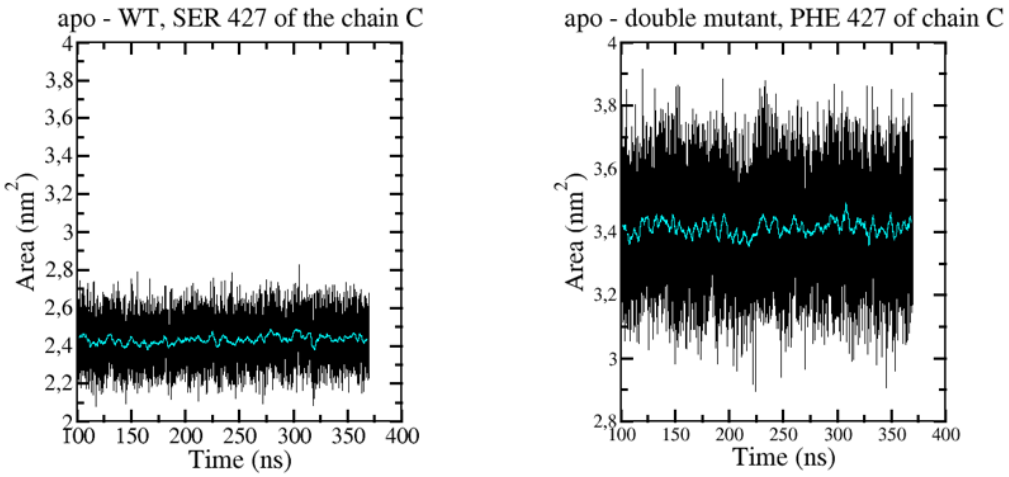

Figure 50: Solvent Accessible Surface Area of residues SER/PHE427 (chain C) of the RXRα apo WT (left) and double mutant systems (right). The cyan lines are the average values over time.

The same behavior is present for residues SER/PHE427 in chain C (Figure 50), with the average value for the apo WT to be 2.43 ± 0.09 nm^2 and for the apo double mutant 3.41 ± 0.1 nm^2.

3.3.7 Dynamical Network Analysis

In this section, a comparison of the results of the Dynamical Network Analysis for the apo RXRα-RXRα systems is demonstrated.

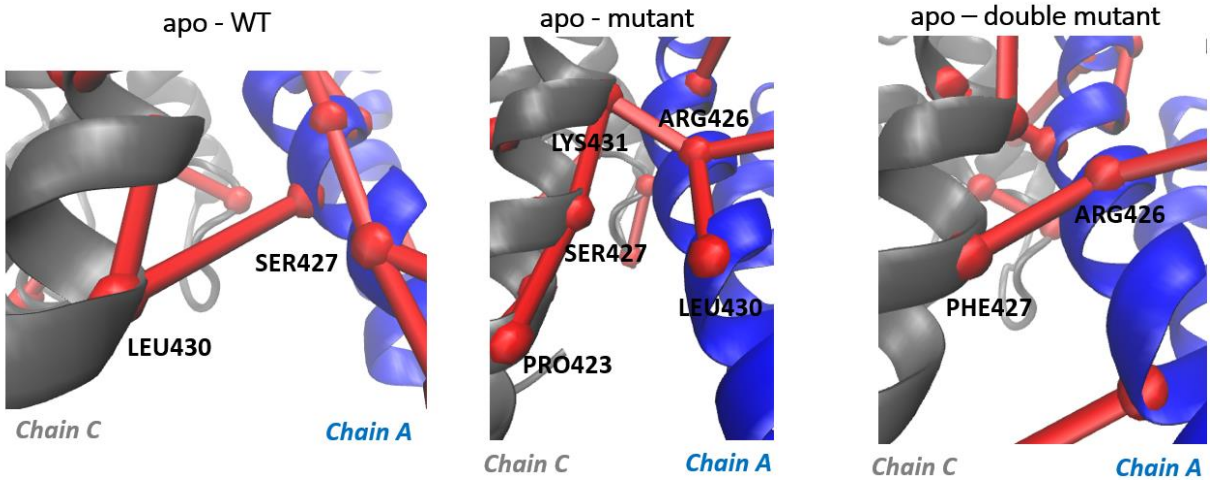

Figure 51: Comparison of the critical nodes between the RXRα apo systems in their interface area. (From left to right) – apo WT dimer, apo mutant dimer, apo double mutant dimer.

Firstly, comparison of the critical nodes of the dimer interface between the apo WT and mutant systems shows that residues SER/PHE427 are involved in the communication between the two monomers (Figure 51). In the apo WT residue SER427 of the chain A is present as a critical node. In the apo mutant PHE427 of chain A is no longer a critical node, but SER427 of chain C is. In the apo double mutant PHE427 of chain C remains a critical node.

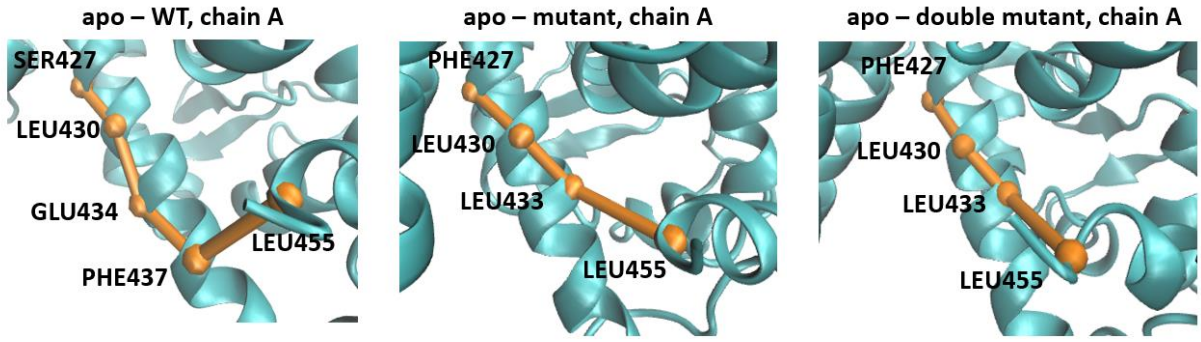

Figure 52: Comparison of the optimal paths between SER/PHE427 and Helix 12 (LEU455) of the chain A for the apo RXR systems. (From left to right) – apo WT dimer, apo mutant dimer, apo double mutant dimer.

Furthermore, comparison of the optimal paths (Figure 52) between the residues SER/PHE427 and LEU455 of helix 12 in chain A shows that there are some differences between the mutant and WT systems. The residues GLU434 and PHE437 that appear to be involved in the optimal path of the apo WT system, disappear in the mutant systems,

thus shortening the path of communication between PHE427 and helix 12. The total weight of the paths is 178, 112 and 66 for the WT, mutant and double mutant respectively (see Equation 2.38). The weight of the paths in the mutant dimers is decreased indicating that the residues are more correlated in the mutant proteins compared to WT.

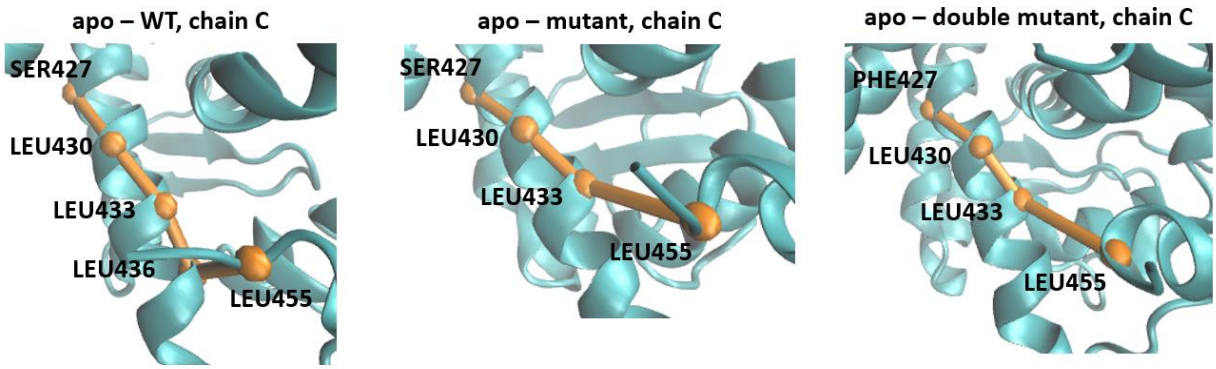

Figure 53: Comparison of the optimal paths between S427F and Helix 12 (LEU455) of the chain C for the apo RXR systems. (From left to right) – apo WT dimer, apo mutant dimer, apo double mutant dimer.

Also, the paths in chain C demonstrate the same behavior (Figure 53) albeit. In this case only one residue of the apo WT, LEU436, disappears in the mutant systems and the optimal path remains the same. The total weight for the paths is 188, 100 and 70 for the WT, mutant and double mutant. The same reduction of the weight is shown to these paths, thus residues in mutants are more correlated.

Visualizing the last frame of all simulations, helix 12 seems to remain in the active conformation for both monomers of the apo RXRα-RXRα (Figure 54, 55), mostly because of the presence of the coactivator.

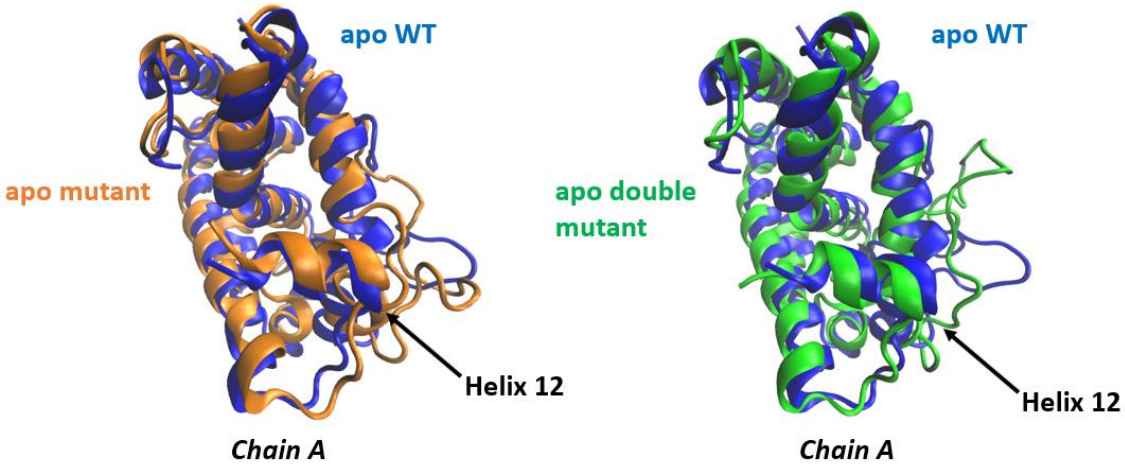

Figure 54: Comparison of the conformation of helix 12 in chain A between WT, mutant and double mutant using the last frame of the apo RXR-RXR simulations. Apo WT is shown in blue, apo mutant in orange and double mutant in green.

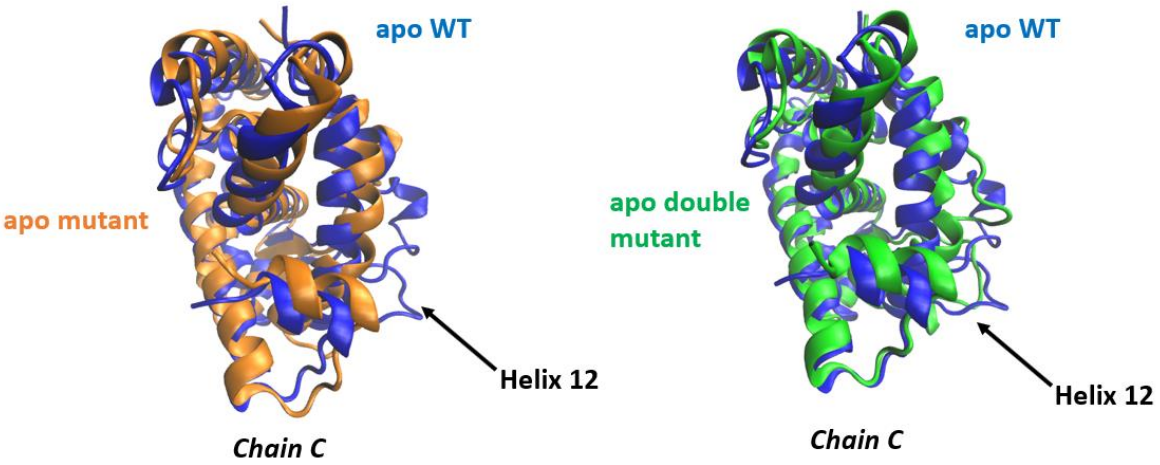

Figure 55: Comparison of the conformation of helix 12 in chain C between WT, mutant and double mutant using the last frame of the apo RXR-RXR simulations. Apo WT is shown in blue, apo mutant in orange and double mutant in green.

Moreover, after docking [87] the ligand 9-cis RA in the binding site of mutant systems it seems that 9-cis RA cannot bind correctly into the pocket. Specifically, it appears to not have any interactions with the binding site residues in the mutant (Figure 56) compared to WT (Figure 57).

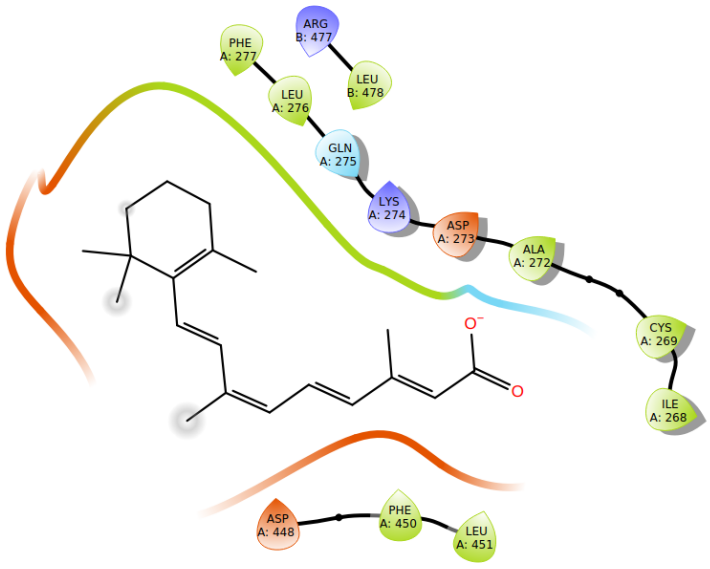

Figure 56: Ligand interaction diagram of the docked 9-cis RA to the binding pocket of the mutant RXRα-RXRα homodimer.

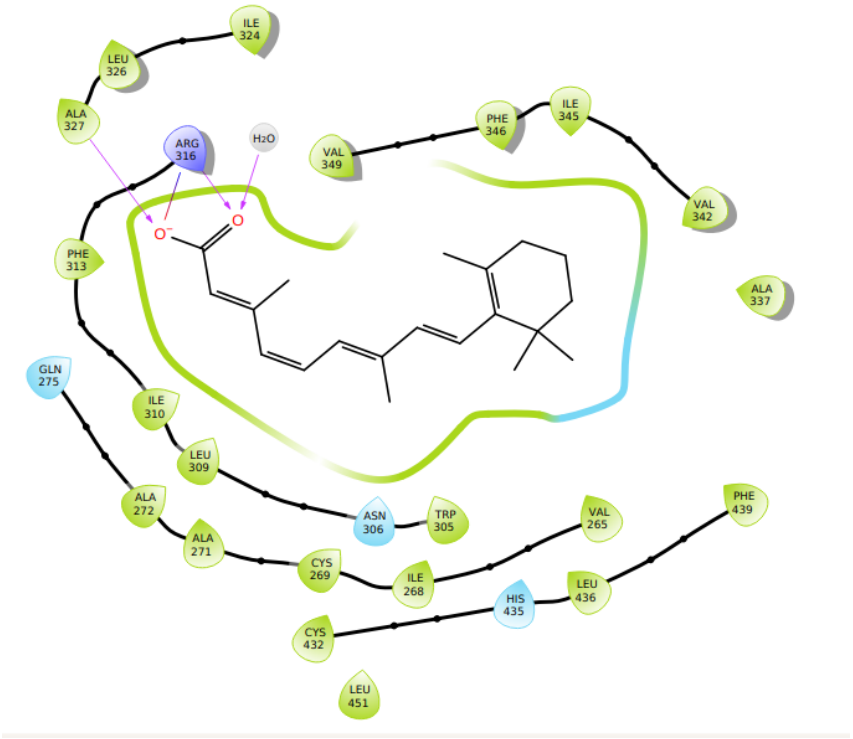

Figure 57: Ligand interaction diagram of the docked 9-cis RA to the binding pocket of the WT RXRα-RXRα homodimer.

3.4 Trajectory Analysis for the holo – RXRα Homodimers (WT and Mutants)

In this section, the results of the trajectory analysis of the holo – RXRα systems, WT, mutant and double mutant, is illustrated. The way that this mutation affects the homodimerization interface, the solvent accessible surface area of the binding pocket of 9-cis RA, helix 12 responsible for the activation of the protein and residues SER/PHE427 is explored.

3.4.1 Convergence Analysis (temperature, pressure, density)

Time series of the temperature (Figure 58), pressure (Figure 59) and density (Figure 60) for the holo RXRα-RXRα WT and mutant systems show that these properties have reached stable values. This means that these properties have converged, and the system has probably reached thermodynamic equilibrium; thus, we are ready to proceed with the analysis.

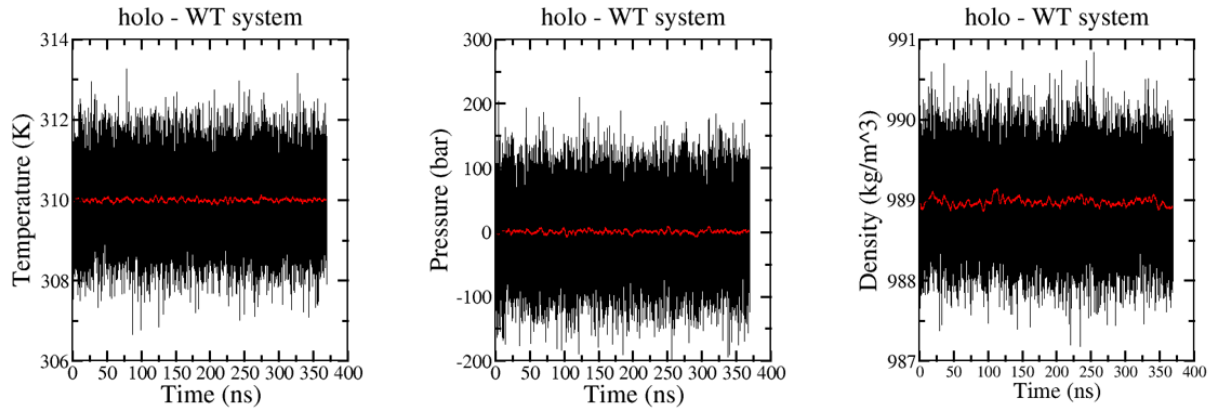

Figure 58: Temperature, Pressure and Density time series for the holo – WT RXRα homodimer. Red lines are running averages over time.

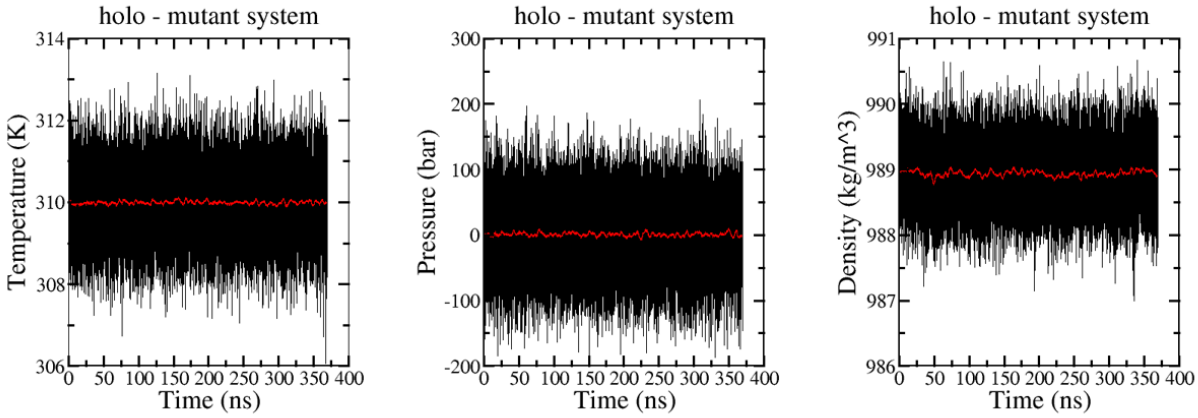

Figure 59: Temperature, Pressure and Density time series for the holo – mutant RXRα homodimer. Red lines are running averages over time.

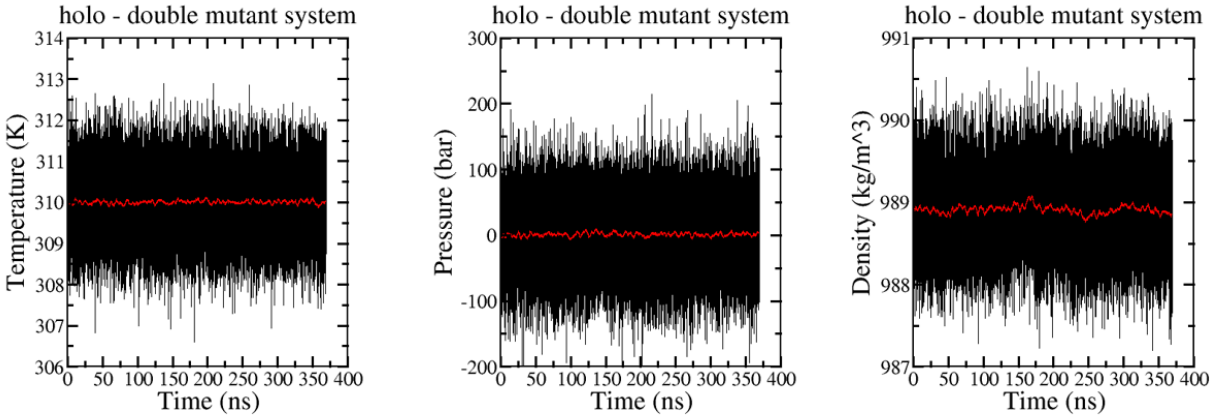

Figure 60: Temperature, Pressure and Density time series for the holo – double mutant RXRα homodimer. Red lines are running averages over time.

3.4.2 Root mean square deviation of the monomers

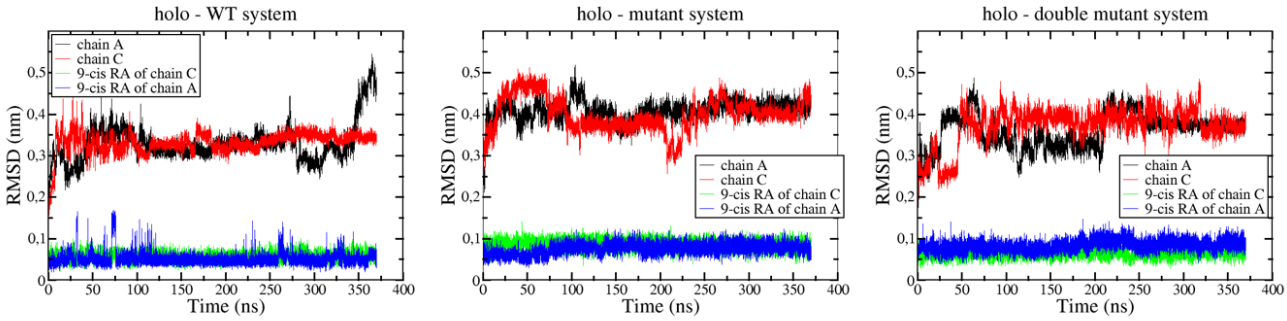

Figure 61: RMSD time series for the holo RXRα systems. (From left to right) – holo WT system, holo mutant system, holo double mutant system. In all graphs the RMSD of the chain A is shown in black color, chain C in red, 9-cis RA of chain A in blue and 9-cis RA of chain C in green.

Regarding the RMSD time series, Figure 61 shows that the monomers of all holo systems, along with their ligands, 9-cis RA, present uniform behavior over time. Perhaps, the holo WT needs more time in order to chain A be stabilized.

3.4.3 Distance of the center of mass between the monomers over time

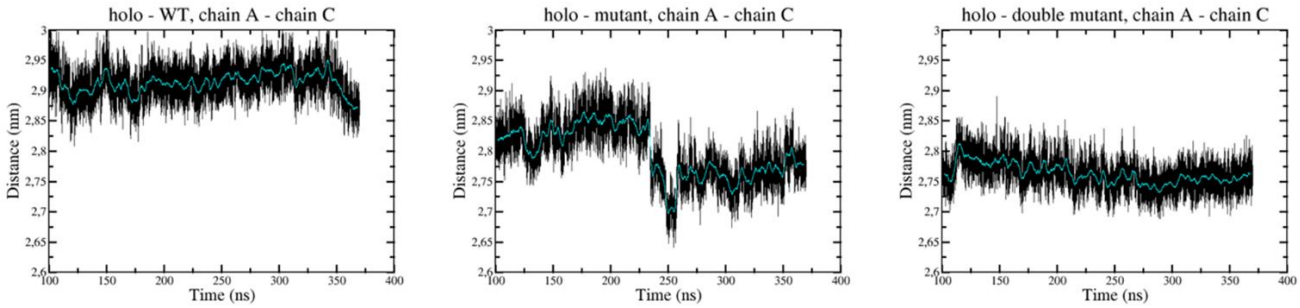

Figure 62: Time series of the distance of the center of mass between the monomers for the holo RXRα systems. Cyan lines are running averages over time. (From left to right) – holo WT system, holo mutant system, holo double mutant system.

The graphs of the distance of the mass between the monomers (Figure 62) as well as Table 7 show that there are small differences between the values of the holo WT and mutant systems. The distance between the monomers does not seem to be affected by the mutation.

Table 7: The average values of the distance of the center of mass between the monomers for the holo RXRα systems.

System	Center of Mass of Monomers Distance (nm)	
Holo – Wild Type	Replica 1	2.91 ± 0.02
Holo – Mutant	Replica 1	2.79 ± 0.04
Holo – Double Mutant	Replica 1	2.76 ± 0.02

3.4.4 Hydrogen bond analysis for the dimer

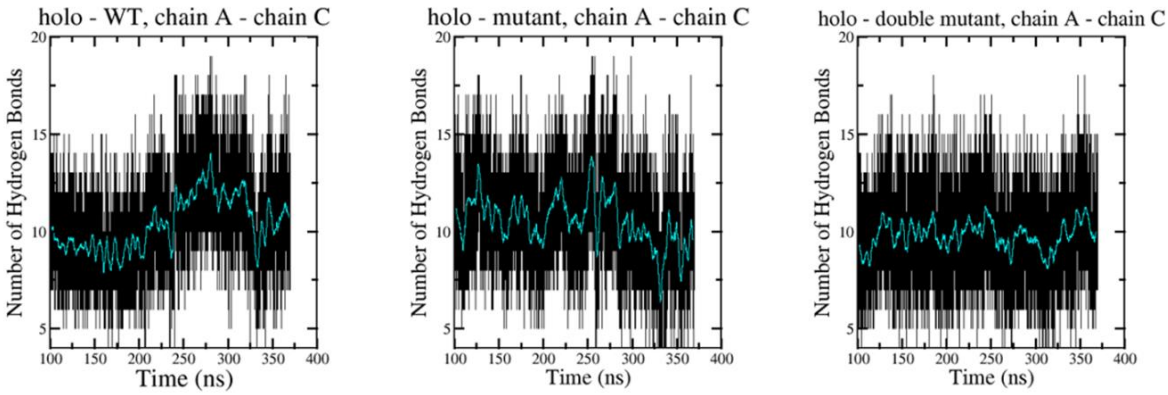

Figure 63: Calculation of the average number of the hydrogen bonds between the monomers of the holo RXRα systems. Cyan lines are running averages over time. (From left to right) – holo WT system, holo mutant system, holo double mutant system.

From the hydrogen bond analysis (Figure 63), insignificant differences between the average values of the holo systems are presented. As Table 8 shows, the average number between holo WT and mutant is approximately the same, whereas in the double mutant is slightly decreased.

Table 8: The averages values of the average number of hydrogen bonds between the monomers for the holo RXRα systems.

System	Average Number of Hydrogen Bonds	
Holo – Wild Type	Replica 1	10.3 ± 2.1
Holo – Mutant	Replica 1	10.5 ± 2.2
Holo – Double Mutant	Replica 1	9.7 ± 1.9

3.4.5 Interaction energy between the monomers

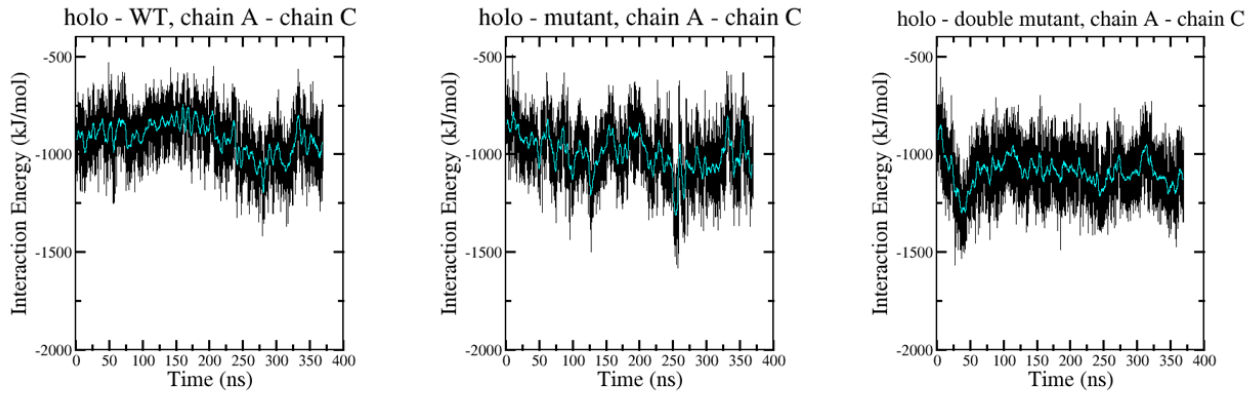

Figure 64: Interaction energy between the monomers for the holo RXRα systems. Cyan lines are running averages over time. (From left to right) – holo WT system, holo mutant system, holo double mutant system.

Moreover, the interaction energy between the monomers of the holo systems (Figure 64) does not present large deviations. Also, the average values are shown in Table 9.

Table 9: The averages values of the interaction energy between the monomers for the holo RXRα systems.

System	Interaction Energy (kcal/mol)	
Holo – Wild Type	Replica 1	-923 ± 122
Holo – Mutant	Replica 1	-1015 ± 129
Holo – Double Mutant	Replica 1	-1088 ± 101

3.4.6 Solvent accessible surface area of the binding site, Helix 12 and residue SER/PHE 427

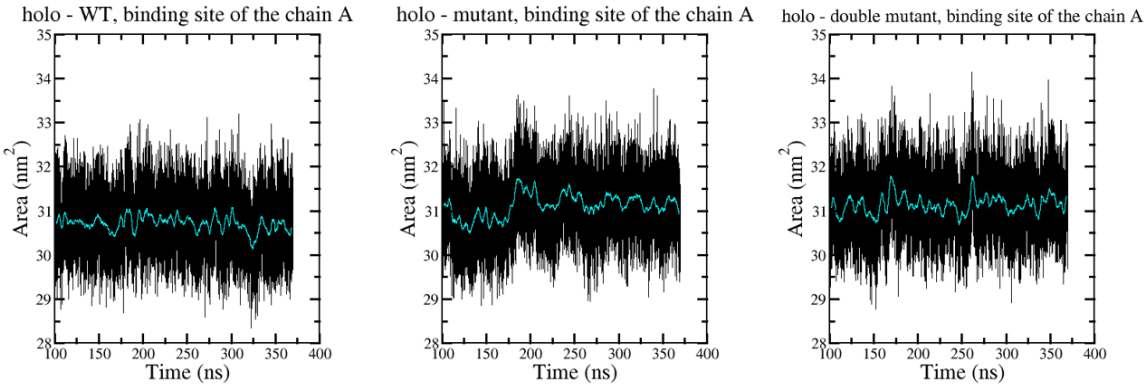

Figure 65: Solvent Accessible Surface Area for the binding site area (chain A) of the holo RXRα systems. Cyan lines are running averages over time. (From left to right) – holo WT system, holo mutant system, holo double mutant system.

According to the graphs of the solvent accessible surface area of the binding site in chain A (Figure 65), there is a small increase of the values. Specifically, the average value of the holo WT is 30.71 ± 0.5 nm^2, the holo mutant is 31.11 ± 0.6 nm^2 and the holo double mutant is 31.14 ± 0.6 nm^2.

On the other hand, there is a small reduction in the solvent accessible surface area of the binding site in chain C (Figure 66) between the holo WT and the double mutant. The average value from 31.13 ± 0.6 nm^2 of the holo WT decreases to 28.68 ± 0.6 nm^2.

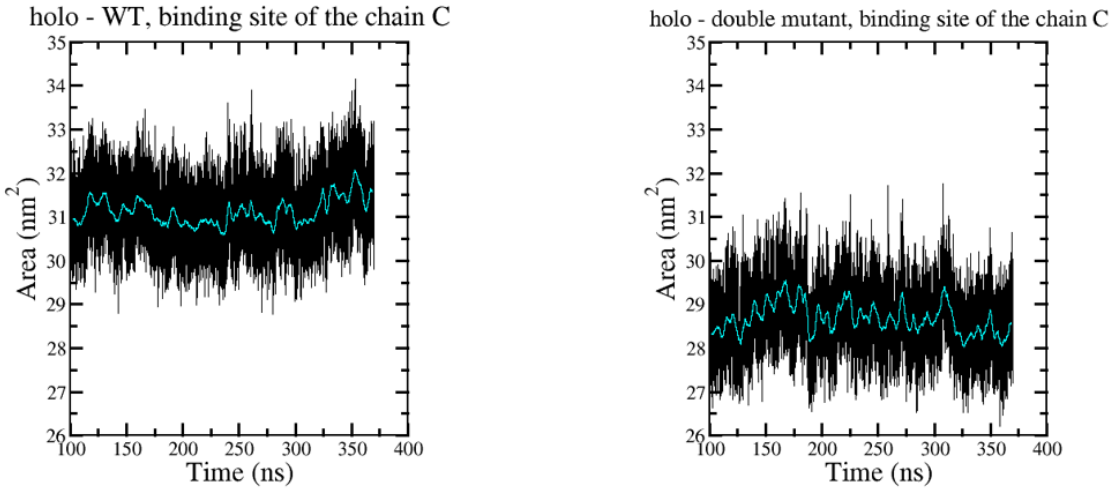

Figure 66: Solvent Accessible Surface Area for the binding site area (chain C) of the holo WT (left) and double mutant RXRα systems (right). Cyan lines are running averages over time.

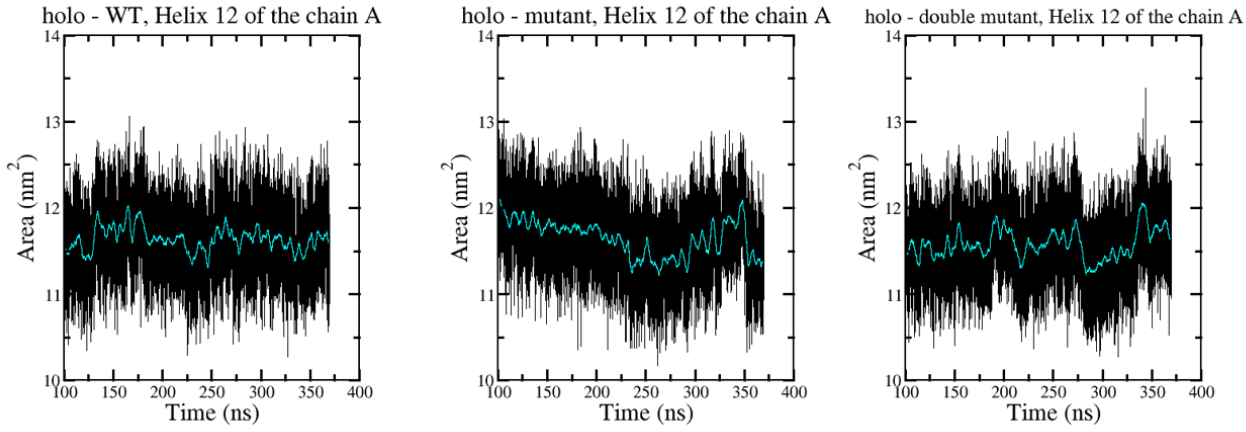

Figure 67: Solvent Accessible Surface Area of Helix 12 (chain A) of the holo RXRα systems. Cyan lines are running averages over time. (From left to right) – holo WT system, holo mutant system, holo double mutant system.

Regarding the helix 12 of chain A, the graphs of the solvent accessible surface area (Figure 67) show that there is no significant difference. The average values are: 11.63 ± 0.3 nm^2, 11.65 ± 0.3 nm^2 and 11.58 ± 0.3 nm^2, for the holo WT, mutant and double mutant respectively.

In contrast, the analysis for the helix 12 of chain C (Figure 68), depict a decrease between the holo WT and the double mutant. The average value for the WT is 11.71 ± 0.4 nm^2 and for the double mutant is 11.16 ± 0.3 nm^2.

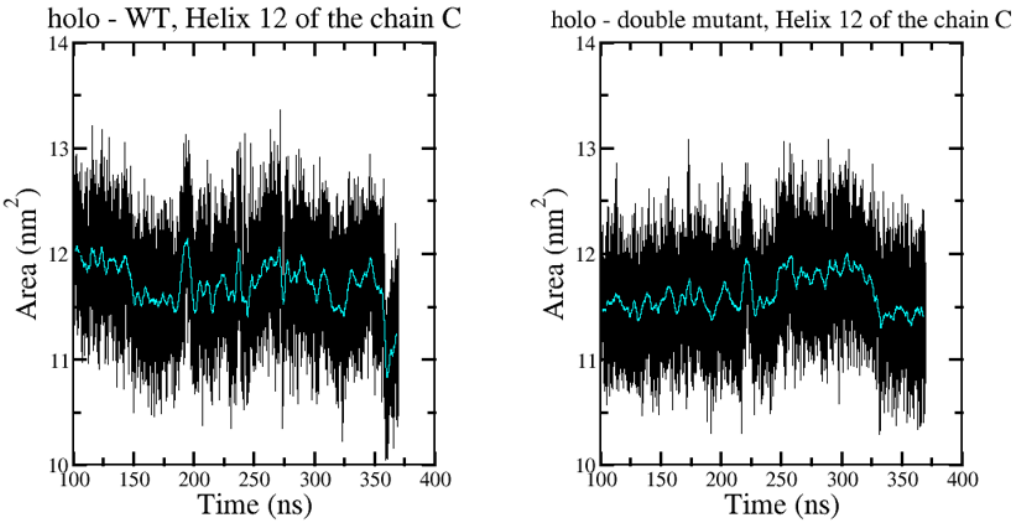

Figure 68: Solvent Accessible Surface Area of Helix 12 (chain C) of the RXRα holo WT (left) and double mutant systems (right). The cyan lines are the average values over time.

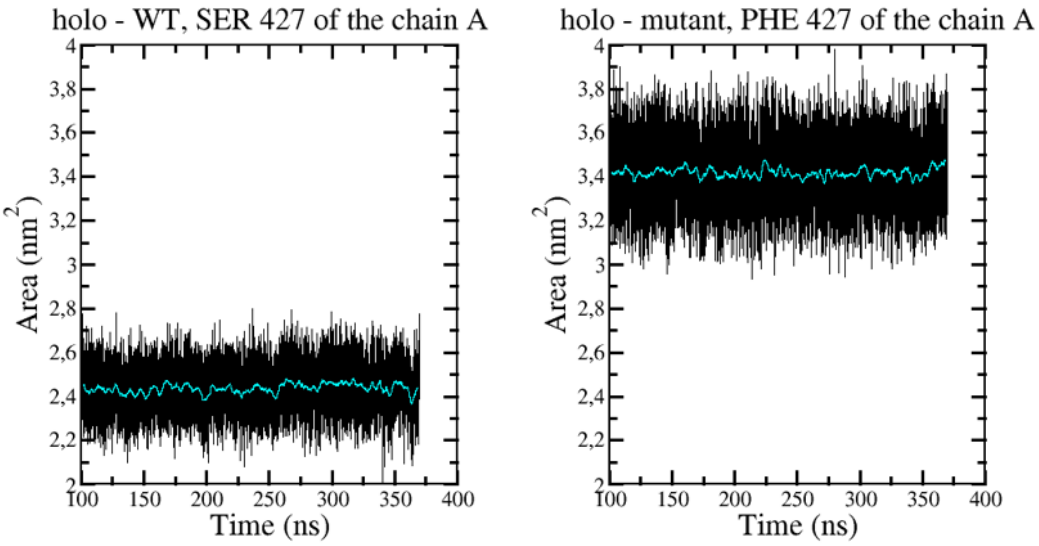

Figure 69: Solvent Accessible Surface Area of residues SER/PHE427 (chain A) of the RXRα holo WT (left) and mutant systems (right). Cyan lines are running averages over time.

Finally, the graphs of the solvent accessible surface area of the residues SER/PHE427 in chain A (Figure 69) display, again, a small difference, as the average value of the holo WT is 2.43 ± 0.09 nm^2, whereas the value of the holo mutant is 3.41 ± 0.1 nm^2. In chain C (Figure 70), the average values deviate from 2.43 ± 0.09 nm^2 in holo WT to 3.40 ± 0.1 nm^2 in holo double mutant.

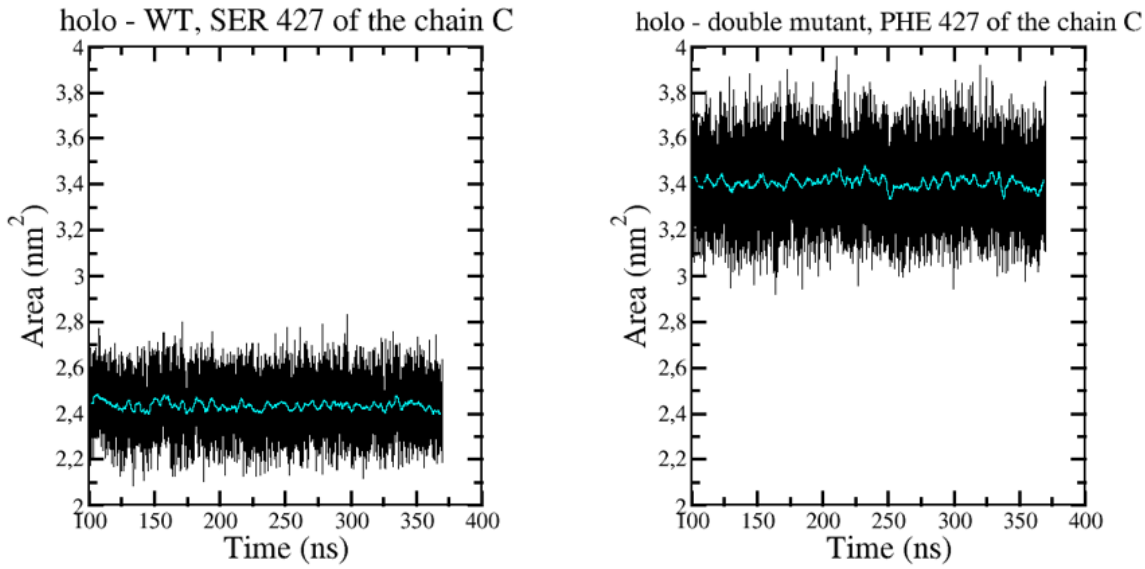

Figure 70: Solvent Accessible Surface Area of residues SER/PHE427 (chain C) of the RXRα holo WT (left) and double mutant systems (right). Cyan lines are running averages over time.

3.4.7 Dynamical Network Analysis

In this section, a comparison between the results from the Dynamical Network Analysis for the holo RXRα systems is demonstrated.

The comparison between the critical nodes of the dimer interface between the monomers of the holo systems (Figure 71) presents that in holo WT residue SER427 is not a critical node. However, in mutant system both SER427 of chain C and PHE427 of chain A are present, whereas in double mutant disappear again.

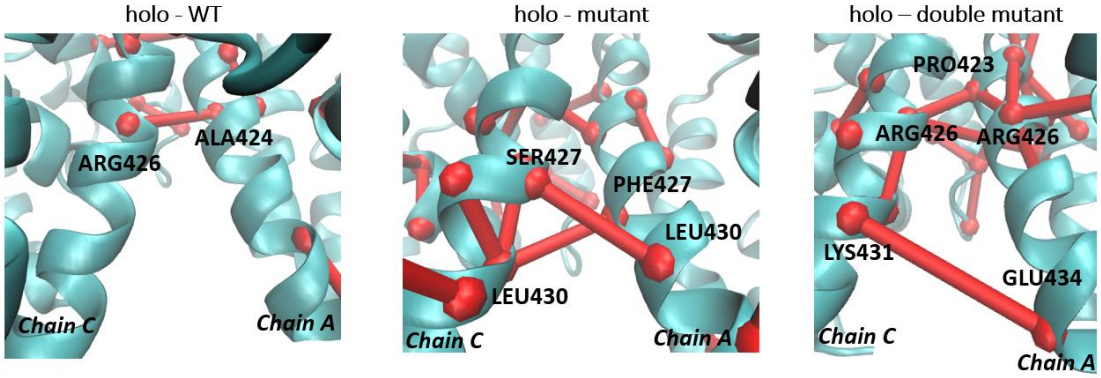

Figure 71: Comparison of the critical nodes between the RXRα holo systems in their interface area. (From left to right) – holo WT dimer, holo mutant dimer, holo double mutant dimer.

The comparison between the optimal paths of residues SER/PHE427 and LEU455 of helix 12 in chain A (Figure 72) shows that more residues are added in the mutant paths. In holo mutant, residues such as GLU434, PHE437 are added, whereas in holo double mutant residues GLU434 and LEU436 appear. The total weight of the paths is 188, 91 and 229 for WT, mutant and double mutant respectively. The weight of the paths shows that in holo mutant residues are more correlated compared to WT and double mutant.

In chain C (Figure 73), the only thing that is changed is that in holo mutant the residue PHE437 is replaced by LEU436. Also, the total weight of the paths is 196, 114 and 144 for WT, mutant and double mutant respectively. This shows that residues of the paths in holo mutant and double mutant are more correlated compared to WT.

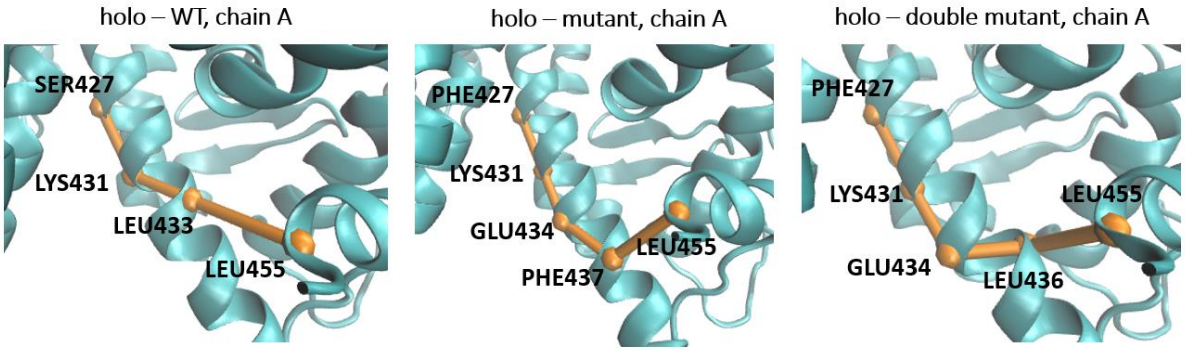

Figure 72: Comparison of the optimal paths between SER/PHE427 and Helix 12 (LEU455) of the chain A for the holo RXR systems. (From left to right) – holo WT dimer, holo mutant dimer, holo double mutant dimer.

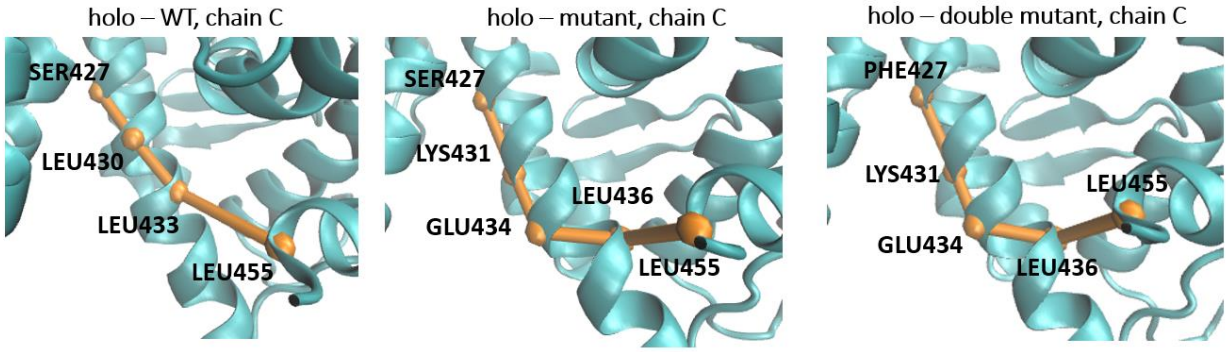

Figure 73: Comparison of the optimal paths between SER/PHE427 and Helix 12 (LEU455) of the chain C for the holo RXR systems. (From left to right) – holo WT dimer, holo mutant dimer, holo double mutant dimer.

3.5 Trajectory Analysis for the apo – RAR/RXRα Heterodimer

Finally, Dynamical Network Analysis was performed for the RXRα-RARα heterodimer and confirmed the previously results from the same analysis in Dr. Cournia lab [23].

First, the comparison of the critical nodes of the dimer interface (Figure 74) shows that PHE427 is no longer a critical node in the mutant structures. This could mean a disruption in the communication, in terms of motional correlation.

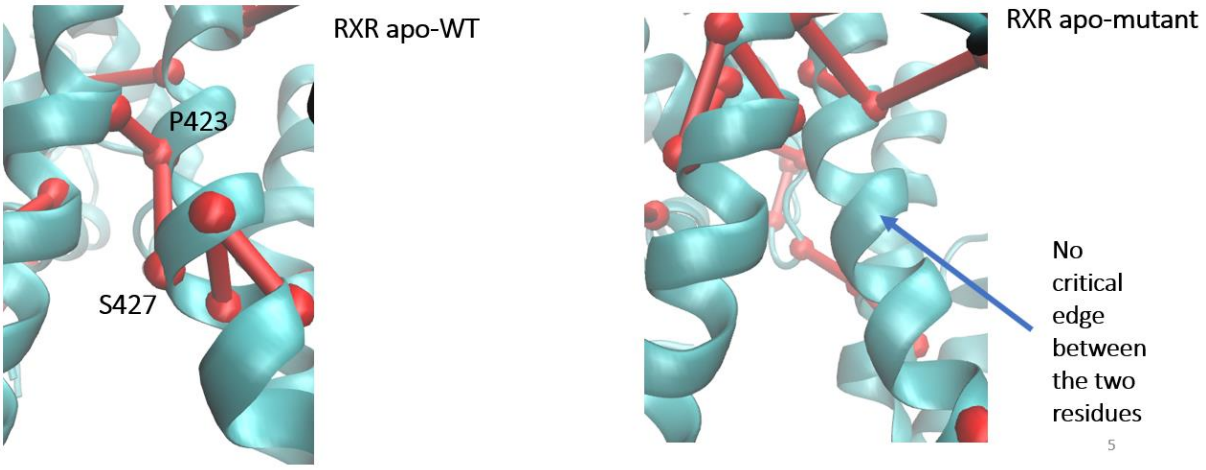

Figure 74: Comparison of the critical nodes between the RXR-RAR systems in their interface area.

Also, the comparison of the optimal paths between the residues PRO423 (it forms constantly hydrogen bond with SER427) and LEU455 of helix 12 (Figure 75) shows that the residue SER427 communicates with helix 12 through a highly correlated motion of helices 10 and 11. In the mutant system the residue PHE427 cuts the communication with helix 11 and directly communicates and regulates the motion of helix 12.

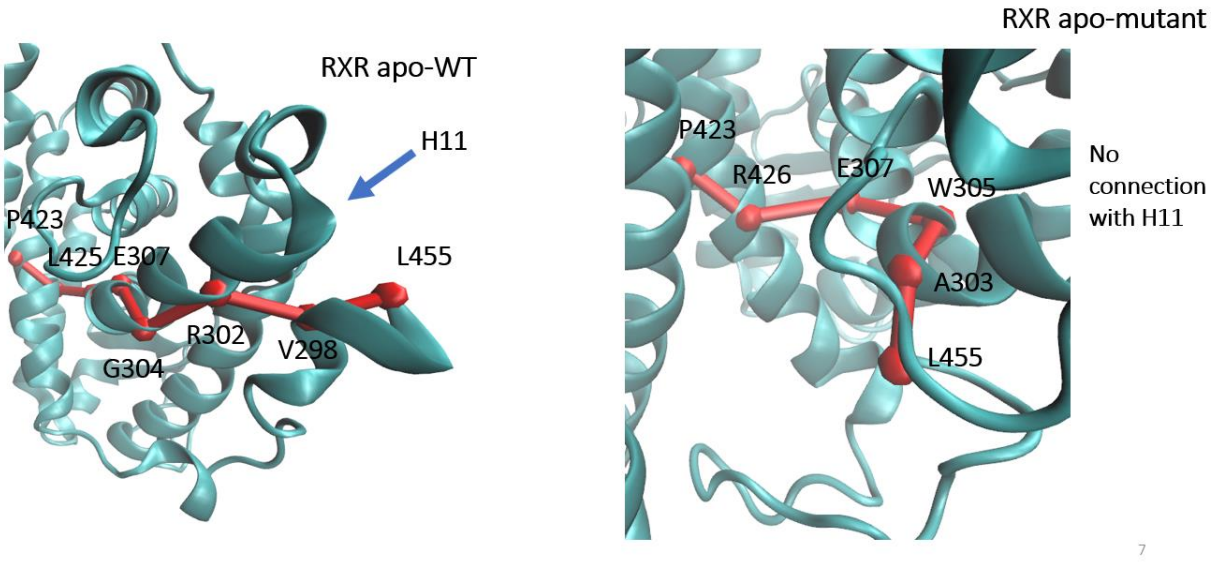

Figure 75: Comparison of the optimal paths between PRO423 and Helix 12 (LEU455) for the RXR-RAR systems.

Considering all the above, the major finding of this master thesis is that the single mutation, from serine to phenylalanine, at position 427 in apo RXRα homodimer leads to an increase of the binding site. This assumption is based on the evidence that the solvent accessible surface area of the binding site in both monomers is expanded (a difference approximately 3-4 units between WT and mutant systems).

4. CONCLUSIONS

The advent of scientific supercomputing brought about advances in molecular simulations, which are now able to complement experimental procedures and shed light in intrinsic details of biomolecular. In particular, Molecular Dynamics simulations have been termed the “computational microscope”, which has provided a unique framework for the study of the phenomena of molecular biology in atomic level detail. Several MD studies [75] [74] have shown the validity of the approach and the successes in relating atomic detail to the function of biomolecular complexes, which cannot be achieved by smaller-scale simulations such as quantum mechanical calculations or existing experimental approaches alone. In this context, with MD simulations we can study the molecular effects of cancer activating mutations on the structure and dynamics of mutated proteins, as it has been suggested that changes in the equilibrium between its active and inactive conformations are linked to its oncogenic potential [103].

In this study, MD simulations were performed in order to elucidate the structural and dynamical effects of the S427F RXRα mutation, which exerts its effects in 5-8% of bladder cancer patients. Our study aims at determining the exact molecular mechanism of how the S427F mutation exerts its action on RXRα in relation to its partners PPARγ and RXRα and compare with available data from the literature. Therefore, MD simulations of the RXRα-RXRα homodimer and the RXRα-PPARγ heterodimer in their WT forms as well as the S427F mutant forms were performed. The trajectory analysis that ensued following the simulations, allowed us to gain valuable insights pertaining to the mechanism of S427F RXRα.

Initially, simulations were checked for convergence by monitoring the time series of the pressure, temperature, density of the systems as well as the root mean square deviation from the experimentally determined crystal structure of each system. The convergence analysis of all systems showed that all systems have been equilibrated and therefore trajectory analysis can be performed.

For the RXRα-PPARγ heterodimer, the calculations of the interaction energy, the average number of hydrogen bonds and the distance of the center of mass between the monomers presented that the S427F mutation does not affect the dimerization interface of the dimer. Moreover, using dynamical network analysis, we identified the optimal paths between the residues SER/PHE427 (mutant) and LEU455 of helix 12 of RXRα as well as helix 12 of PPARγ in both WT and mutant proteins. The comparison for the RXRα connecting the mutated residue with RXRα helix 12 showed that the path is the same, and this could mean that the mutation does not affects helix 12 directly. For the PPARγ subunit, the path connecting the mutant residue and LEU455 of helix12 of PPARγ is longer in the mutant indicating that due to the mutation this heterodimer may be losing control of helix 12 of PPARγ, which is free to adopt the active conformation. Also, visualization of the trajectory revealed a π-π interaction between residues PHE427 of RXRα and TYR477 of PPARγ as dominant conformation for the first 100 ns of the simulation. All these findings, confirm and further support the outcome from previous simulations in the literature [35] [36] that revealed no effect of the mutation in dimerization interface, the aromatic interaction between the mutant residue and the tyrosine terminal of PPARγ as well as the allosteric control of the mutant exerted on helix 12 of PPARγ but not of RXRα.

For the RXRα-RXRα homodimers, the calculations of the interaction energy, the average number of hydrogen bonds and the distance of the center of mass between the monomers revealed that the S427F mutation does not affect the dimerization interface of the dimers. However, the calculation of the solvent accessible surface area (SASA) of the binding

site, helix 12 and residues SER/PHE427 in both monomers demonstrated some differences.

Concerning the apo RXRα homodimers, we observed that the SASA of the binding site of both chains in the mutants is significantly increased. The direct effect of this observation is that the 9-cis RA ligand is not able to bind in the binding site of RXRα and activate it. On the other hand, the SASA calculations of helix 12 and residues SER/PHE427 do not present any statistically significant differences.

For the holo RXRα homodimers, the SASA of the binding site in chain A of the mutants is slightly increased, whereas in chain C it is decreased; this means that the binding site of chain C in the mutants becomes more hydrophobic compared to the binding site of the holo WT. The SASA calculations of helix 12 do not present any statistically significant deviations, while residues SER/PHE427 demonstrate small differences.

The Dynamical Network Analysis for the apo and holo RXRα-RXRα homodimers uncovered that the optimal paths between the residues SER/PHE427 and LEU455 of helix 12 have small differences. Specifically, in apo mutants the paths become shorter, because some residues that existed in WT system, they disappear in the mutants. Also, according to weights of the paths it seems that PHE427 communicates with helix 12 directly. On the other hand, in holo mutants, the paths remain the same, which shows that the communication between the mutation and the helix 12 is unaffected. The comparison between the critical nodes of the dimers interface do not reveal any significant differences.

Comparing the RXRα-RXRα homodimer simulations with previous simulations of RXRα-RARα heterodimer [23], we observe that although in the apo mutant RXRα-RARα heterodimer the binding site volume is decreased, in the apo mutants RXRα-RXRα homodimers the binding site volume is increased. Also, our results confirmed the previous biological experiments (see Section 1.3.6). In the case of RXRα-RAR heterodimer, the decrease of the binding site volume justifies the decrease of the transcriptional activity of the mutant dimer as the ligand may not be able to bind to the pocket. In the case of RXRα-RXRα homodimer, the increase of the binding site may explain the small increase of the transcriptional output of the mutant dimer as the ligand might not be able to bind correctly to the pocket.

Considering all the above, the main conclusion of these MD simulations and analysis is that the single-point mutation, from serine to phenylalanine at position 427 in RXRα, affects its structure and functionality depending on its dimer partner. As the mutant action on the structural dynamics of the RXRα-RXRα homodimer is different from the RXRα-PPARγ and RXRα-RARα heterodimers [35] [23], we assume that depending on the heterodimer partner a new mechanism of RXRα regulation will ensue. This gives a distinctive characteristic in the action of the mutant RXRα when it functions either as a heterodimeric or homodimeric partner, and it presents an exciting system to be studied further in relation to other heterodimer partners.

5. FUTURE PRESPECTIVES

Although these findings provide a significant structural insight of the effects of the S427F mutation on the RXRα, further investigation is needed to cover all the aspects of the function of this mutation.

Due to the major finding that this mutation presents a different mechanism based on the dimer partner, it would be very interesting to explore how it regulates the transcriptional activity of other dimers such as RXR-LXR as well as other heterodimer partners.

Moreover, it would be interesting to investigate how the S427F mutation affects the apo RXRα-RXRα homodimer removing the coactivator, by performing MD simulations. The removal of coactivator will give us the opportunity to observe how the mutants evolve over time without the locked active conformation that the coactivator provides. Through these simulations, we can detect the conformational changes of the helices, and particularly of helix 12 and observe if it will remain in the active conformation for both WT and mutant dimers.

Moreover, it would be advisable for the simulations of the apo and holo RXRα-RXRα homodimers to be further continued. More simulation time would give a better indication of the opening of the 9 cis RA binding pocket and of other structural changes that may occur (Figures 38, 54), although we do not anticipate any significant differences compared to the presented results as the simulations have converged.

Finally, the hydrogen bond analysis (Figure 31) along with the distance of the center of mass monomers (Figure 30) for the RXRα-PPARγ heterodimer showed that, also, this system should be continued in order to converge these properties. It would be interesting to notice whether a simulation of the WT apo RXRα-PPARγ heterodimer would converge helix 12 back to its inactive state in the absence of an activating ligand or coactivator.

ABBREVIATIONS – ACRONYMS

MD	Molecular Dynamics
9-cis RA	9-cis Retinoic Acid
NR	Nuclear Receptor
RXRα	Retinoid X Receptor alpha
DBD	DNA Binding Domain
LBD	Ligand Binding Domain
AF-1	Activation Function 1
AF-2	Activation Function 2
FXR	Farnesoid X Receptor
LXR	Liver X Receptor
PPAR	Peroxisome Proliferator-Activated Receptor
RAR	Retinoic Acid Receptor
TR	Thyroid Receptor
VDR	Vitamin D Receptor
DR	Direct Repeat
WT	Wild Type
PBC	Periodic Boundary Conditions
RMSD	Root Mean Square Deviation
SASA	Solvent Accessible Surface Area

REFERENCES

[1] D. S. Latchman, “Transcription factors: An overview,” *Int. J. Biochem. Cell Biol.*, vol. 29, no. 12, pp. 1305–1312, Dec. 1997.

[2] D. B. Nikolov and S. K. Burley, “RNA polymerase II transcription initiation: a structural view.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 1, pp. 15–22, Jan. 1997.

[3] P. J. Mitchell and R. Tjian, “Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.,” *Science*, vol. 245, no. 4916, pp. 371–8, Jul. 1989.

[4] R. M. Evans, “The steroid and thyroid hormone receptor superfamily,” *Science (80-.).*, 1988.

[5] J. M. Olefsky, “Nuclear Receptor Minireview Series,” *J. Biol. Chem.*, 2001.

[6] R. M. Evans and D. J. Mangelsdorf, “Nuclear receptors, RXR, and the big bang,” *Cell*. 2014.

[7] L. M. Sherwood, E. E. Parris, and B. W. O’Malley, “Mechanisms of Action of Steroid Hormones,” *N. Engl. J. Med.*, vol. 284, no. 7, pp. 370–377, Feb. 1971.

[8] K. Yamamoto, “Steroid Receptor Regulated Transcription of Specific Genes and Gene Networks,” *Annu. Rev. Genet.*, 1985.

[9] D. J. Mangelsdorf *et al.*, “The nuclear receptor superfamily: the second decade.,” *Cell*, 1995.

[10] Á. R. De Lera, W. Krezel, and R. Rühl, “An Endogenous Mammalian Retinoid X Receptor Ligand, at Last!,” *ChemMedChem*. 2016.

[11] P. D. Kiser, M. Golczak, and K. Palczewski, “Chemistry of the Retinoid (Visual) Cycle,” *Chem. Rev.*, 2014.

[12] “Panretin (Alitretinoin): Side Effects, Interactions, Warning, Dosage & Uses.” [Online]. Available: https://www.rxlist.com/panretin-drug.htm. [Accessed: 18-Jun-2019].

[13] P. Huang, V. Chandra, and F. Rastinejad, “Retinoic Acid Actions through Mammalian Nuclear Receptors,” *Chem. Rev.*, 2014.

[14] R. Kumar and E. B. Thompson, “The structure of the nuclear hormone receptors.,” *Steroids*, vol. 64, no. 5, pp. 310–9, May 1999.

[15] A. Wärnmark, E. Treuter, A. P. H. Wright, and J.-Å. Gustafsson, “Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation,” *Mol. Endocrinol.*, 2003.

[16] “File:Nuclear Receptor Structure.png - Wikimedia Commons.” [Online]. Available: https://commons.wikimedia.org/wiki/File:Nuclear_Receptor_Structure.png. [Accessed: 21-Jun-2019].

[17] M. Mark, N. B. Ghyselinck, and P. Chambon, “Function of retinoic acid receptors during embryonic development,” *Nucl. Recept. Signal.*, vol. 7, no. 1, p. nrs.07002, Jan. 2009.

[18] T. Perlmann, P. N. Rangarajan, K. Umesono, and R. M. Evans, “Determinants for selective RAR and TR recognition of direct repeat HREs.,” *Genes Dev.*, vol. 7, no. 7b, pp. 1411–1422, Jul. 1993.

[19] R. Kurokawa *et al.*, “Polarity-specific activities of retinoic acid receptors determined by a co-repressor,” *Nature*, vol. 377, no. 6548, pp. 451–454, Oct. 1995.

[20] F. Rastinejad, “Retinoid X receptor and its partners in the nuclear receptor family,” *Curr. Opin. Struct. Biol.*, vol. 11, no. 1, pp. 33–38, Feb. 2001.

[21] H. Gronemeyer, J.-Å. Gustafsson, and V. Laudet, “Principles for modulation of the nuclear receptor superfamily,” *Nat. Rev. Drug Discov.*, vol. 3, no. 11, pp. 950–964, Nov. 2004.

[22] L. Altucci and H. Gronemeyer, “The promise of retinoids to fight against cancer,” *Nat. Rev. Cancer*, vol. 1, no. 3, pp. 181–193, Dec. 2001.

[23] Z. Cournia and I. Galdadas, “Personal Communication.”

[24] P. Germain, J. Iyer, C. Zechel, and H. Gronemeyer, “Co-regulator recruitment and the mechanism of retinoic acid receptor synergy,” *Nature*, vol. 415, no. 6868, pp. 187–192, Jan. 2002.

[25] S. Kersten, D. Kelleher, P. Chambon, H. Gronemeyer, and N. Noy, “Retinoid X receptor alpha forms tetramers in solution.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 19, pp. 8645–9, Sep. 1995.

[26] R. T. Gampe, V. G. Montana, M. H. Lambert, G. B. Wisely, M. V Milburn, and H. E. Xu, “Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix,” *Genes Dev.*, vol. 14, no. 17, pp. 2229–2241, Sep. 2000.

[27] S. Kersten, L. Pan, and N. Noy, “On the Role of Ligand in Retinoid Signaling: Positive Cooperativity in the Interactions of 9-cis Retinoic Acid with Tetramers of the Retinoid X Receptor,” *Biochemistry*, vol. 34, no. 43, pp. 14263–14269, Oct. 1995.

[28] S. Kersten, P. R. Reczek, and N. Noy, “The Tetramerization Region of the Retinoid X Receptor Is Important for Transcriptional Activation by the Receptor,” *J. Biol. Chem.*, vol. 272, no. 47, pp. 29759–29768, Nov. 1997.

[29] A. IJpenberg *et al.*, “In vivo activation of PPAR target genes by RXR homodimers,” *EMBO J.*, vol. 23, no. 10, pp. 2083–2091, May 2004.

[30] X. Li, P. A. Hansen, L. Xi, R. A. S. Chandraratna, and C. F. Burant, “Distinct Mechanisms of Glucose Lowering by Specific Agonists for Peroxisomal Proliferator Activated Receptor γ and Retinoic Acid X Receptors,” *J. Biol. Chem.*, vol. 280, no. 46, pp. 38317–38327, Nov. 2005.

[31] T. Tanaka, K. S. Suh, A. M. Lo, and L. M. De Luca, “p21 $^{WAF1/CIP1}$ Is a Common Transcriptional Target of Retinoid Receptors,” *J. Biol. Chem.*, vol. 282, no. 41, pp. 29987–29997, Oct. 2007.

[32] V. Nunez *et al.*, “Retinoid X receptor controls innate inflammatory responses through the up-regulation of chemokine expression,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 23, pp. 10626–10631, Jun. 2010.

[33] X. Zhang, Y. Su, L. Chen, F. Chen, J. Liu, and H. Zhou, “Regulation of the nongenomic actions of retinoid X receptor-α by targeting the coregulator-binding sites.,” *Acta Pharmacol. Sin.*, vol. 36, no. 1, pp. 102–12, Jan. 2015.

[34] V. Vivat-Hannah, W. Bourguet, M. Gottardis, and H. Gronemeyer, “Separation of retinoid X receptor homo- and heterodimerization functions.,” *Mol. Cell. Biol.*, vol. 23, no. 21, pp. 7678–88, Nov. 2003.

[35] A. M. Halstead *et al.*, “Bladder-cancer-associated mutations in RXRA activate peroxisome proliferator-activated receptors to drive urothelial proliferation,” *Elife*, vol. 6, Nov. 2017.

[36] M. Korpal *et al.*, “Evasion of immunosurveillance by genomic alterations of PPARγ/RXRα in bladder cancer.,” *Nat. Commun.*, vol. 8, no. 1, p. 103, 2017.

[37] Weinstein *et al.*, “Comprehensive molecular characterization of urothelial bladder carcinoma,” *Nature*, vol. 507, no. 7492, pp. 315–322, Mar. 2014.

[38] E. M. Van Allen *et al.*, “Somatic ERCC2 Mutations Correlate with Cisplatin Sensitivity in Muscle-Invasive Urothelial Carcinoma,” *Cancer Discov.*, vol. 4, no. 10, pp. 1140–1153, Oct. 2014.

[39] G. Guo *et al.*, “Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation,” *Nat. Genet.*, vol. 45, no. 12, pp. 1459–1463, Dec. 2013.

[40] E. Bonis, A.Klinakis, “Personal Communication.”

[41] “Intermolecular forces.” [Online]. Available: http://chemistry.bd.psu.edu/jircitano/IMforces.html. [Accessed: 17-Jun-2019].

[42] “Intramolecular and intermolecular forces (article) | Khan Academy.” [Online]. Available: https://www.khanacademy.org/test-prep/mcat/chemical-processes/covalent-bonds/a/intramolecular-and-intermolecular-forces. [Accessed: 17-Jun-2019].

[43] “Hydrogen Bonding | Chemistry for Non-Majors.” [Online]. Available: https://courses.lumenlearning.com/cheminter/chapter/hydrogen-bonding/. [Accessed: 23-Jun-2019].

[44] “File:Pi-pi interactions.jpg - Wikimedia Commons.” [Online]. Available: https://commons.wikimedia.org/wiki/File:Pi-pi_interactions.jpg. [Accessed: 23-Jun-2019].

[45] “What Is a Covalent Bond in Chemistry?” [Online]. Available: https://www.thoughtco.com/definition-of-covalent-bond-604414. [Accessed: 17-Jun-2019].

[46] B. J. Alder and T. E. Wainwright, “Phase Transition for a Hard Sphere System,” *Cit. J. Chem. Phys*, vol. 27, p. 1208, 1957.

[47] J. A. McCammon, B. R. Gelin, and M. Karplus, “Dynamics of folded proteins,” *Nature*, vol. 267, no.

5612, pp. 585–590, Jun. 1977.

[48] "The Nobel Prize in Chemistry 2013 - NobelPrize.org." [Online]. Available: https://www.nobelprize.org/prizes/chemistry/2013/summary/. [Accessed: 10-Jun-2019].

[49] S. A. Adcock and J. A. McCammon, "Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins," *Chem. Rev.*, 2006.

[50] A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus, "An all-atom empirical energy function for the simulation of nucleic acids," *J. Am. Chem. Soc.*, vol. 117, no. 48, pp. 11946–11975, Dec. 1995.

[51] A. D. MacKerell *et al.*, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †," *J. Phys. Chem. B*, vol. 102, no. 18, pp. 3586–3616, Apr. 1998.

[52] L. Lipid --Peptide, I. Peptide, I. Kevin, and K. M. Majewski, "Molecular Dynamics."

[53] "Parameter Files." [Online]. Available: https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node25.html. [Accessed: 18-Jun-2019].

[54] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids: Second edition*. 2017.

[55] "Lennard-Jones Potential - Chemistry LibreTexts." [Online]. Available: https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Physical_Properties_of_Matter/Atomic_and_Molecular_Properties/Intermolecular_Forces/Specific_Interactions/Le. [Accessed: 18-Jun-2019].

[56] P. G. Huray, *Maxwell's equations*. Wiley, 2010.

[57] D. C. Rapaport and D. C., *The art of molecular dynamics simulation*. Cambridge University Press, 1995.

[58] J. C. Smith Michael Grunze, "Structure and Dynamics of Biomembranes containing Cholesterol and other Biologically-Important Sterols. A computational perspective. Referees."

[59] D. Frenkel and B. Smit, *Understanding molecular simulation : from algorithms to applications*. .

[60] "1.7: The Maxwell Distribution Laws - Chemistry LibreTexts." [Online]. Available: https://chem.libretexts.org/Courses/University_of_California_Davis/UCD_Chem_107B%3A_Physical_Chemistry_for_Life_Scientists/Chapters/1%3A_Properties_of_Gases/1.7%3A_The_Maxwell_Distribution_Laws. [Accessed: 11-Jun-2019].

[61] W. Phadungsukanan, M. Kraft, J. A. Townsend, and P. Murray-Rust, "The semantics of Chemical Markup Language (CML) for computational chemistry: CompChem," *J. Cheminform.*, 2012.

[62] "Theory of Molecular Dynamics Simulations." [Online]. Available: https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part1.html. [Accessed: 07-Jun-2019].

[63] D. A. (Donald A. McQuarrie, *Statistical mechanics*. New York: Harper & Row, 1975.

[64] J. Grotendorst and Winter school (2002.02.25-03.01 : Kerkrade), *Quantum simulations of complex many-body systems: from theory to algorithms : winter school, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, the Netherlands ; lecture notes*. NIC-Secretariat, 2002.

[65] R. W. Hakala, "A new derivation of the Boltzmann distribution law," *J. Chem. Educ.*, vol. 38, no. 1, p. 33, Jan. 1961.

[66] D. A. (Donald A. McQuarrie, *Statistical mechanics*. University Science Books, 2000.

[67] A. (Akira) Isihara, *Statistical physics*. Academic Press, 1971.

[68] J. J. Kelly, "Ideal Quantum Gases from Statistical Physics using Mathematica," 1996.

[69] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, no. 3, pp. 1695–1697, Mar. 1985.

[70] Justin A. Lemkul, "Lysozyme in Water." [Online]. Available: http://www.mdtutorials.com/gmx/lysozyme/07_equil2.html. [Accessed: 04-Jun-2019].

[71] M. E. Tuckerman* and G. J. Martyna, "Understanding Modern Molecular Dynamics: Techniques and Applications," 1999.

[72] D. Dibenedetto, G. Rossetti, R. Caliandro, and P. Carloni, "A molecular dynamics simulation-based

interpretation of nuclear magnetic resonance multidimensional heteronuclear spectra of α-synuclein·dopamine adducts," *Biochemistry*, 2013.

[73] "Periodic Boundary Conditions." [Online]. Available: http://isaacs.sourceforge.net/phys/pbc.html. [Accessed: 18-Jun-2019].

[74] M. Kotzabasaki, I. Galdadas, E. Tylianakis, E. Klontzas, Z. Cournia, and G. E. Froudakis, "Multiscale simulations reveal IRMOF-74-III as a potent drug carrier for gemcitabine delivery," *J. Mater. Chem. B*, vol. 5, no. 18, pp. 3277–3282, 2017.

[75] I. Tremi *et al.*, "Effect of 5-trans Isomer of Arachidonic Acid on Model Liposomal Membranes Studied by a Combined Simulation and Experimental Approach," *J. Membr. Biol.*, vol. 251, no. 3, pp. 475–489, Jun. 2018.

[76] H. Leontiadou, I. Galdadas, C. Athanasiou, and Z. Cournia, "Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations," *Sci. Rep.*, vol. 8, no. 1, p. 15544, Dec. 2018.

[77] J. N. Israelachvili, *Intermolecular and surface forces*. Academic Press, 2011.

[78] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Ann. Phys.*, vol. 389, no. 20, pp. 457–484, 1927.

[79] "Potential Energy Function." [Online]. Available: https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part2.html. [Accessed: 04-Jun-2019].

[80] R. E. Wilde and 1947- Surjit Singh, *Statistical mechanics : fundamentals and modern applications*. Wiley, 1998.

[81] J. P. Changeux, "50th anniversary of the word 'allosteric,'" *Protein Science*. 2011.

[82] A. Sethi, J. Eargle, A. A. Black, and Z. Luthey-Schulten, "Dynamical networks in tRNA:protein complexes," *Proc. Natl. Acad. Sci.*, 2009.

[83] J. Eargle, A. Sethi, L. Li, and Z. Luthey-Schulten, "Dynamical Network Analysis Network Tool Developers Tutorial Authors," 2012.

[84] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks.," *Proc. Natl. Acad. Sci. U. S. A.*, 2002.

[85] R. W. Floyd and R. W., "Algorithm 97: Shortest path," *Commun. ACM*, vol. 5, no. 6, p. 345, Jun. 1962.

[86] P. F. Egea, A. Mitschler, and D. Moras, "Molecular Recognition of Agonist Ligands by RXRs," *Mol. Endocrinol.*, vol. 16, no. 5, pp. 987–997, May 2002.

[87] M. P. Repasky, M. Shelley, and R. A. Friesner, "Flexible Ligand Docking with Glide," in *Current Protocols in Bioinformatics*, vol. Chapter 8, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007, p. Unit 8.12.

[88] "The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC." .

[89] M. P. Jacobson, R. A. Friesner, Z. Xiang, and B. Honig, "On the Role of the Crystal Environment in Determining Protein Side-chain Conformations," *J. Mol. Biol.*, vol. 320, no. 3, pp. 597–608, Jul. 2002.

[90] "Schrödinger Release 2019-2: Maestro, Schrödinger, LLC, New York, NY, 2019." .

[91] M. P. Jacobson *et al.*, "A hierarchical approach to all-atom protein loop prediction\rAccurate and efficient loop selections by the DFIRE-based all-atom statistical potential\rModeling structurally variable regions in homologous proteins with rosetta," *Proteins*, 2004.

[92] D. C. Bas, D. M. Rogers, and J. H. Jensen, "Very fast prediction and rationalization of pKa values for protein-ligand complexes," *Proteins Struct. Funct. Bioinforma.*, vol. 73, no. 3, pp. 765–783, May 2008.

[93] E. Harder *et al.*, "OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins," *J. Chem. Theory Comput.*, 2016.

[94] J. W. Ponder and D. A. Case, "Force fields for protein simulations.," *Adv. Protein Chem.*, vol. 66, pp. 27–85, 2003.

[95] D. J. Price and C. L. Brooks, "A modified TIP3P water potential for simulation with Ewald summation," *J. Chem. Phys.*, vol. 121, no. 20, pp. 10096–10103, Nov. 2004.

[96] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, “GROMACS: A message-passing parallel molecular dynamics implementation,” *Comput. Phys. Commun.*, vol. 91, no. 1–3, pp. 43–56, Sep. 1995.

[97] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “LINCS: A linear constraint solver for molecular simulations,” *J. Comput. Chem.*, vol. 18, no. 12, pp. 1463–1472, Sep. 1997.

[98] M. Di Pierro, R. Elber, and B. Leimkuhler, “A Stochastic Algorithm for the Isobaric–Isothermal Ensemble with Ewald Summations for All Long Range Forces,” *J. Chem. Theory Comput.*, vol. 11, no. 12, pp. 5624–5637, Dec. 2015.

[99] M. Parrinello and A. Rahman, “Crystal Structure and Pair Potentials: A Molecular-Dynamics Study,” *Phys. Rev. Lett.*, vol. 45, no. 14, pp. 1196–1199, Oct. 1980.

[100] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J. Appl. Phys.*, vol. 52, no. 12, pp. 7182–7190, Dec. 1981.

[101] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark, “Peptide Folding: When Simulation Meets Experiment,” *Angew. Chemie Int. Ed.*, vol. 38, no. 1–2, pp. 236–240, Jan. 1999.

[102] K. Humphrey, W., Dalke, A. and Schulten, “VMD - Visual Molecular Dynamics.” J. Molec. Graphics, p. vol. 14, pp. 33–38, 1996.

[103] L. Sutto and F. L. Gervasio, “Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 26, pp. 10616–21, Jun. 2013.