Computational Design of a Pentapeptide Inhibitor for Fibroblast Growth Factor Receptor (FGFR3) IIIB

by

Mehmet Ali ÖZTÜRK

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Computational Sciences and Engineering

Koc University

January 2011

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Mehmet Ali Öztürk

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Burak Erman, Ph. D. (Advisor)

_____

Atilla Gürsoy, Ph. D.

_____

Seda Keskin, Ph. D.

Date: _____

ABSTRACT

Fibroblast growth factor receptor (FGFR) is a cell membrane protein, a member of tyrosine kinase family, which has extracellular domains activated by ligand binding, followed by receptor dimerization. FGFR3 has two isoforms, IIIb and IIIc. The IIIb isoforom of FGFR3 is a highly expressed epithelial cell protein, whose R248C mutation causes different kinds of dermatological diseases like seborrheic keratoses (SK), acanthosis nigricans (AN) and epidermal nevi (EN). This mutation leads to ligand independent receptor dimerization, which increases intracellular signaling, resulting in skin diseases. In order to prevent R248C mutation-caused cellular signaling, a pentapeptide ligand is designed that recognizes the mutation and binds to the receptor dimerization site. For this purpose molecular docking and molecular dynamics simulations are conducted. Binding free energy is calculated with Steered Molecular Dynamics (SMD) and Molecular mechanics – Generalized Born Surface Area (MM-GBSA) methods. The found pentapeptide sequence appears to be a possible drug candidate for FGFR3 IIIb R248C mutation related skin diseases.

ÖZET

Fibroblast büyüme faktörü reseptörü (FGFR) ligand bağlanması ile aktive olan hücre dışı bileşeninin dimerleştiği, tayrozin kinaz ailesine mensup bir hücre zarı proteinidir. FGFR3 IIIb ve IIIc olmak üzere iki izoform yapısına sahiptir. FGFR3 IIIb izoformu epitel hücrelerde yüksek miktarda üretilen bir protein olup; R248C mutasyonu seborrheik keratoz, deri nevüsü ve akantozis nigrikans gibi cilt hastalıklarına yol açar. Bu mutasyon, hücre içi sinyalizasyonu arttırarak deri hastalıklarına neden olan ligand bağımsız reseptör dimerleşmesine yol açar. R248C mutasyonunun neden olduğu hücresel sinyalizasyonu engellemek için, mutasyonu tanıyan ve dimerizasyon bölgesine bağlanan beşli bir peptid inhibitör tasarlanmıştır. Bu amaca yönelik olarak moleküler yerleştirme (docking) ve moleküler dinamik (MD) simülasyonları gerçekleştirilmiştir. Bağlanma serbest enerjileri, çekme moleküler dinamik (SMD) ve Moleküler Mekanik Genelleştirilmiş Born Yüzey Alanı (MM-GBSA) metodları kullanılarak hesaplanmıştır. Bulunan beşli peptid, FGFR3 IIIb R248C mutasyonunun yol açtığı cilt hastalıkları için etkili bir ilaç adayı özelliği taşımaktadır.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor Prof. Dr. Burak Erman for guiding, helping and supporting my thesis and teaching independent research in the field of drug design, which gives chance to work for all human being.

Additionally, I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for financial sport during my graduate studies.

I would like to thank Musa Özboyacı and Gözde Eskici for their precious friendships and endless support. I will never forget our good times, all fun were with them.

Special thanks to Evrim Besray Ünal, Deniz Şanlı, Mert Gür and Özge Engin for answering my endless questions, which they always replied very kindly.

For being good friends, I would like to thank all my office, flat and graduate mates, whom I always enjoyed to be with.

Being a member of this wonderful campus, I really thank for all KU staff working hard and kindly.

Having spent very beautiful two years together, I would like to thank Sedef Dinçer and her family for their support.

Finally and most importantly, I would like to express my gratitude to my family members for always being with me. I am very grateful to Emel Öztürk, Fuat Öztürk and Ozan Utku Öztürk for their endless love, without them this thesis can never be completed.

TABLE OF CONTENTS:

LIST OF FIGURES:

LIST OF TABLES:

1. INTRODUCTION

Seborrheic keratoses (SK), acanthosis nigricans (AN) and epidermal nevi (EN) are dermatologiacal diseases caused by fibroblast growth factor receptor (FGFR3) R248C mutations. As they occur at the outer surface of the dermis, transdermal drug delivery can be used for candidate drugs. By being mostly treated via surgery, topical cure can be cost-effective.

FGFR is a member of tyrosine kinase family which has extracellular domains activated by ligand binding, followed by receptor dimerization. FGFR is a four membered (1, 2, 3 and 4) protein family which is generally involved in normal angiogenesis and embryonic development. FGFR3 has two isoforms called IIIb and IIIc. The IIIb isoform is expressed in epithelial cells, in which its mutations cause dermatological diseases, described above.

R248C mutation of FGFR3 IIIB causes ligand independent receptor dimerization, which increases intracellular signaling. In order to prevent this situation, peptide based drugs can be used for receptor inhibition. Because of the fact that peptide drugs are easy to produce and modify, they are chosen as candidate inhibitors in this study

Computational drug discovery is an active research area, aiming to obtain faster results for drug design, which usually takes more than 10 years for a drug to reach the market. Docking and molecular dynamics simulations are the major tools that are recently used in drug discovery.

In this thesis, the peptide drug design process for FGFR3 IIIb R248C mutation-caused diseases is investigated in six different chapters.

At the literature overview part, dermatological diseases caused by FGFR3 IIIb R248C mutations are explained briefly. Then activation and disease relating mechanism of the FGFR3 is explained.

In the methodological background part, molecular docking techniques, online loop prediction tools are introduced. The molecular dynamics methodology is explained in detail followed by steered molecular dynamics (SMD) and molecular mechanics generalized Born surface area (MM-GBSA) explanation.

In the methods part, the technical basis of the study is given in the order of, preparation of initial coordinate files, molecular dynamics with CHARMM 27 and AMBER force fields, peptide preparing and docking methodology. Binding free energy calculation methods are explained by SMD and MM-GBSA in detailed format.

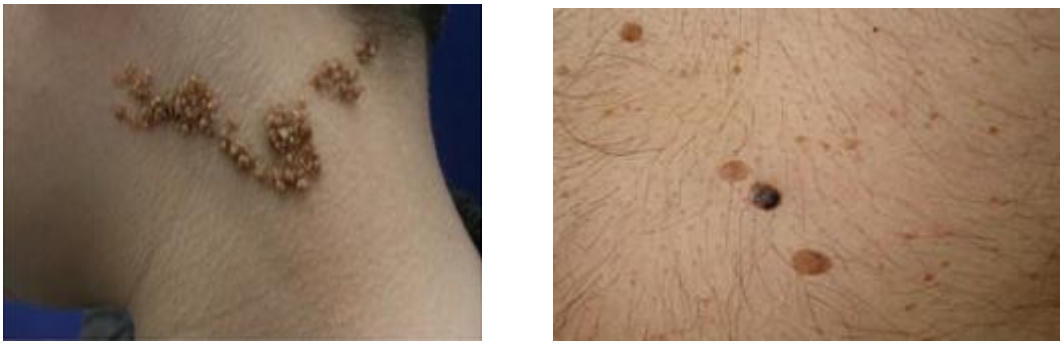
In results and discussion part, the following terms are emphasized in the following order, MD simulations before docking, docking results, SMD simulations and MM-GBSA analysis with RMSD values and calculations done by Jarzynski identity. Additionally drug-likeliness is discussed.

In the conclusion part, a general overview is given about the work done and the effective usage of the candidate drug is explained. Finally, planned future work for this drug design process is explained.

2. LITERATURE OVERVIEW

2.1 FGFR3 related diseases

Seborrheic keratoses (SK), acanthosis nigricans (AN) and epidermal nevi (EN) have common histopathologocial issues like hyperpigmentation, hyperkeratosis, papillomatosis and acanthosis (Figure 2.1.1). Some EN patients can have urothelial (urinary tract) carcinoma at young ages, related to the conjunct FGFR3 mutations [1]. Work completed by Logie et al. has shown that somatic mutations of FGFR3 are important for SK occurrence [2]. Additionally, a transgenic mouse model showed that S249C mutations of the FGFR3 cause skin to look like SK histologically [1]. Also, Hernández concluded that R248C mutation of FGFR3 is an important hot spot resulting in EN [3] .

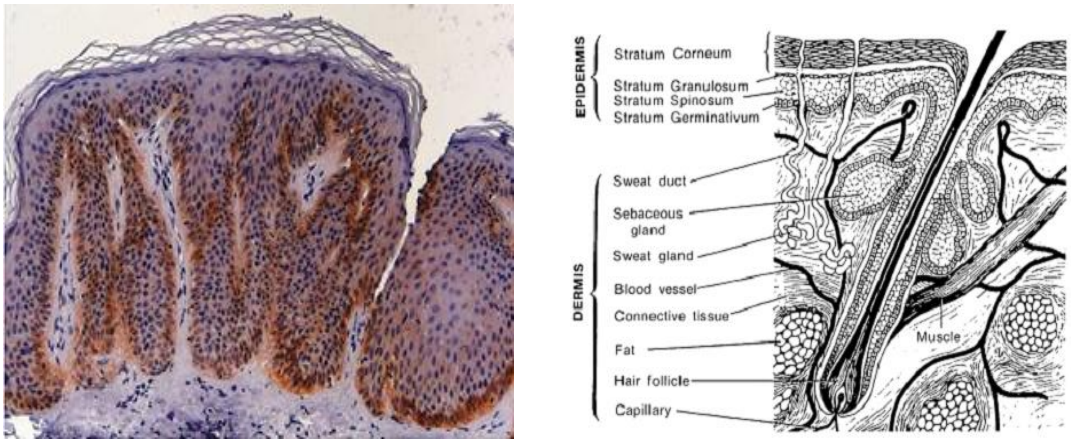


A B

Figure 2.1.1: A- Epidermal nevi on the neck [4] B- Seborrheic keratoses on the back [4]

There is still not enough data explaining the detailed mechanisms of the FGFR3 mutations at urinary tract tissues and skin. Even though exposure to UV is related to SK, it is not enough to explain urethelial tumors [5]. Additionally all SK histological subtypes do not include FGFR3 mutations [3].

An immunohistochemical staining analysis showed that (Figure 2.1.2), expressed FGFR3 proteins are involved at the dermis part of the skin. Since FGFR3 expressed region is too close to the outer surface of epidermis, theoretically it is possible to use topical drugs for treatment.



A B

Figure 2.1.2: A-Immunohistochemical staining for the FGFR3 protein from a seborrheic keratoses region. Brown stained parts are the FGFR3 protein [6]. B- Cross section of the human skin [7].

Currently, epidermal nevi and seborrheic keratoses are treated by surgery [1]. As the mutation patterns are already described for FGFR3, the use of topical peptide drugs is possible to cure these kinds of tumors.

Protein based drugs like insulin and thyroid hormones have been widely in use since mid 1900s. Today, there are more than 200 proteins and peptides approved as a drug by the US Food and Drug Administration (FDA). Peptide drugs can be classified according to their source and activation mechanism as follows: hormones and growth factors, clotting

factors and anticoagulants, bacterial or plant toxins, drug-activating enzymes and antibody-based drugs [8].

For skin disorders transdermal drug delivery can be achieved by chemical enhancers, iontophoresis, microneedles, microdermabrasion, thermal ablation and ultrasound [9]. These recent findings have increased FDA approved transdermal drug rate since 1980s as seen in (Figure 2.1.3). For a peptide based drug, natural pore forming peptide, magainin, can be used as an effective agent for drug permeation through the skin [10].

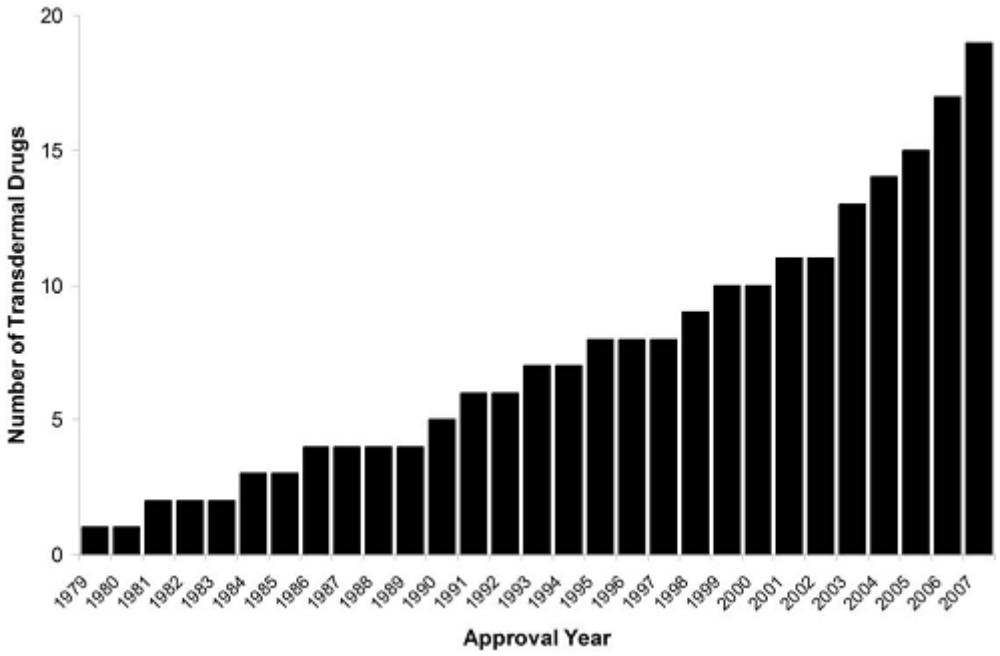


Figure 2.1.3: Cumulative number of FDA approved transdermal drugs [9].

Peptides used in cancer treatment target specific protein-protein interactions which are also seen in the tumor mechanism of FGFR3 R248C mutation. Compared to other chemical compounds, peptide drugs are more preferable as they are small, less

immunogenic, and can be easily modified to avoid degradation and improve bioavailability [8].

2.2 Mechanism of FGFR3 activation

Receptor tyrosine kinases are a large family of proteins that have extracellular domain activated by ligand binding-induced receptor dimerization. Depending on the cell type and its environment, different responses can be obtained like cell migration, proliferation, differentiation and apoptosis. Fibroblast growth factor receptor (FGFR) is a four membered protein family which is generally involved in normal angiogenesis and embryonic development. Structurally the FGFRs have an extracellular ligand binding domain composed of two or three immunoglobulin-like domains, a transmembrane region and a cytoplasmic domain showing intracellular activity (Figure 2.2.1) [11]. Alternative splicing is a process in which the exons of RNA (translated parts of RNA) are reorganized with spliceosome enzyme to produce different proteins. By the alternative splicing of the second half of the Ig domain, different isoforms of FGFRs (FGFR 1-4) (Figure 2.2.2) are produced. This splicing causes ligand binding specificity for different isoforms. FGFR3 has two isoforms, FGFR3 IIIb is an alternatively spliced form of exon 8, FGFR3 IIIc is an alternatively spliced form of exon 9 [12].

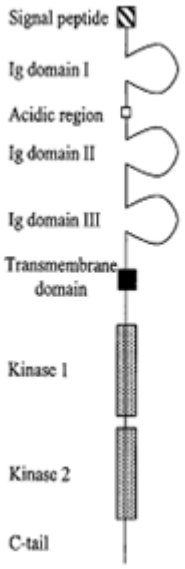


Figure 2.2.1: General structure of FGFR [13]

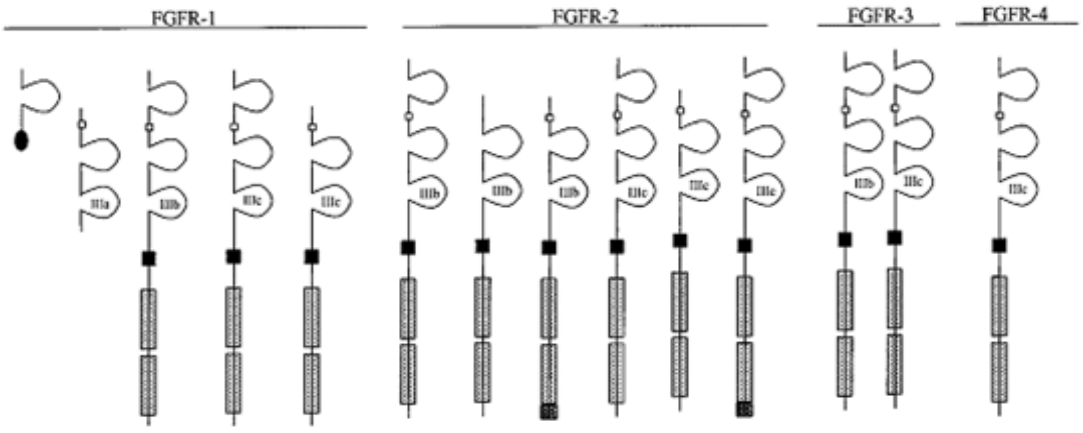


Figure 2.2.2: Structural representatives of FGFR isoforms and variants of RNA splicing [13]

There are about 23 fibroblast growth factor receptors (FGFs) activating the FGFRs. This activation is achieved with the help of sulphated glycosaminoglycans (GAGs) such as heparin or heparan sulfate. Crystallization studies showed that heparin both interacts with FGF and its receptor. This indicates that there is a complex activation mechanism of FGFR receptors. FGFR3 IIIb and IIIc are both activated by FGF1 and FGF9 ligands. Additionally FGF2, 4, 6 and FGF8 also bind to the FGFR3 IIIc isoform. By having a wide range of ligands there is a high number of combinations for receptor activation and possible effects (Table 2.2.1) [13] .

Table 2.2.1: FGF families and their receptors [13].

Ligand Name	**Receptor Name**
FGF1	FGFR1 IIIB, IIIC; FGFR2 IIIB, IIIC;FGFR3 IIIB, IIIC; FGFR4
FGF2	FGFR1 IIIB, IIIC; FGFR2 IIIC; FGFR3 IIIC; FGFR4
FGF3	FGFR1 IIIB FGFR2 IIIB
FGF4	FGFR1 IIIC, FGFR2 IIIC, FGFR3 IIIC, FGFR4
FGF5	FGFR1 IIIC, FGFR2 IIIC
FGF6	FGFR1 IIIC, FGFR2 IIIC
FGF7	FGFR2 IIIB
FGF8	FGFR1, FGFR2 IIIC, FGFR3 IIIC, FGFR4
FGF9	FGFR2 IIIC, FGFR3 IIIB, IIIC; FGFR4
FGF10	FGFR1 IIIB, FGFR2 IIIB
FGF11-14	Unknown
FGF15	Unknown
FGF16-19	FGFR1 IIIC, FGFR2 IIIC
FGF20	Unknown

The FGFRs are expressed variably during the embryonic development. The FGFR3 expression is detected in the kidneys, lungs, brain, cartilage, intestine, pancreas and testis. Tissue specific expression of IIIb and IIIc isoforms are also described. The IIIb isoform is expressed in the epithelial cells whereas IIIc is generally expressed in the mesenchymal cells [14].

FGFR signaling is directed to the cell by a phosphorylation mechanism. Following the receptor dimerization in residues 167-171, autophosphorylation of tyrosine residues 653 and 654, which are highly conserved between FGFR family, occurs. These tyrosine residues are the binding sites of phosphotyrosine binding intracellular signal proteins. Upon activation, Ras-MAPK and STAT pathways are activated (Figure 2.2.3) [12].

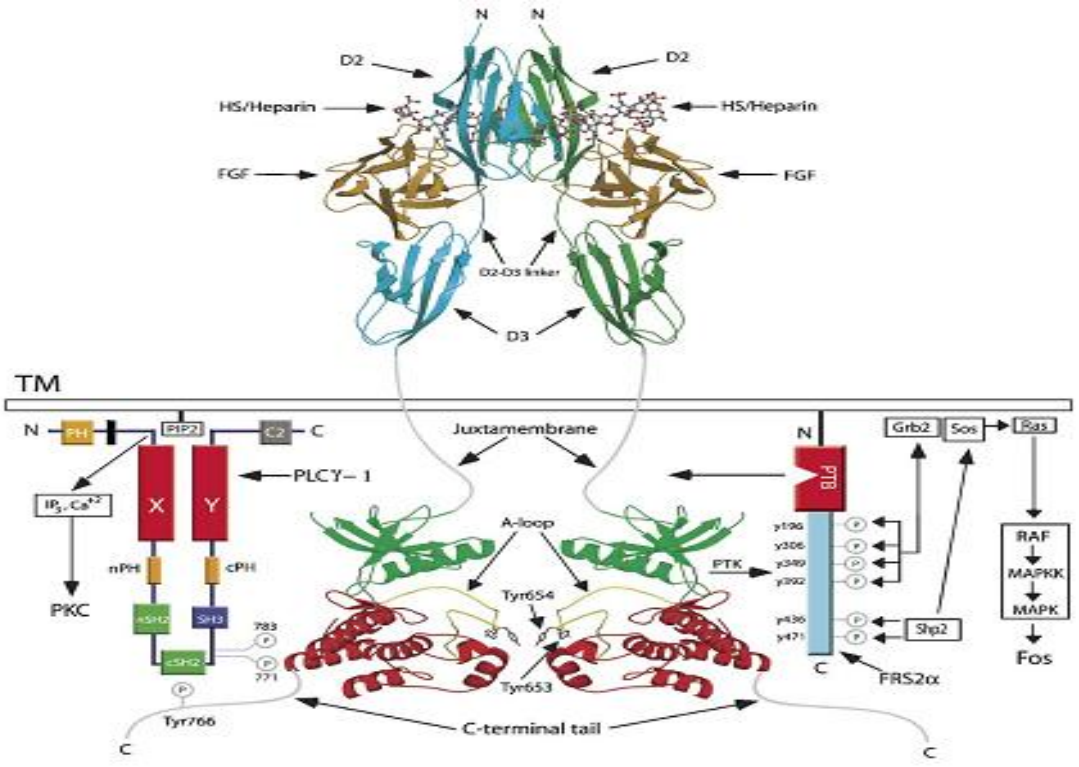


Figure 2.2.3: Structural mechanism of the FGFR3 activation [15]

Specific germline mutations of the FGFR3 cause autosomal dominant human skeletal disorders that include craniosynostoses and chondrodysplasias. The most severe causes of dwarfism are intracellular and extracellular mutations of FGFR3. Mouse model studies show that activated FGFR3 IIIc mutations inhibit normal bone growth which leads to dwarfism.

The over-expression and activated point mutations of FGFR3 give rise to intracellular signaling that results in severe forms of cancer. Point mutations are mostly identified in haematological cancer, multiple myeloma, bladder and cervix cancer [11].

The extracellular missense mutations to cysteine cause ligand independent dimerization of the FGFR3, possibly due to a conformational change, which increase cellular signaling (Figure 2.2.4). Previous studies done by Hartmann and his colleagues show that R248C mutation, mostly observed in benign skin cancers [16], causes FGF independent receptor activation by forming intermolecular disulfide bond [17]. Previously designed FGFR targeted drugs (SU 5402, CHIR-258, PD 173074) are effective on FGFR3, but they are not receptor specific [18-19].

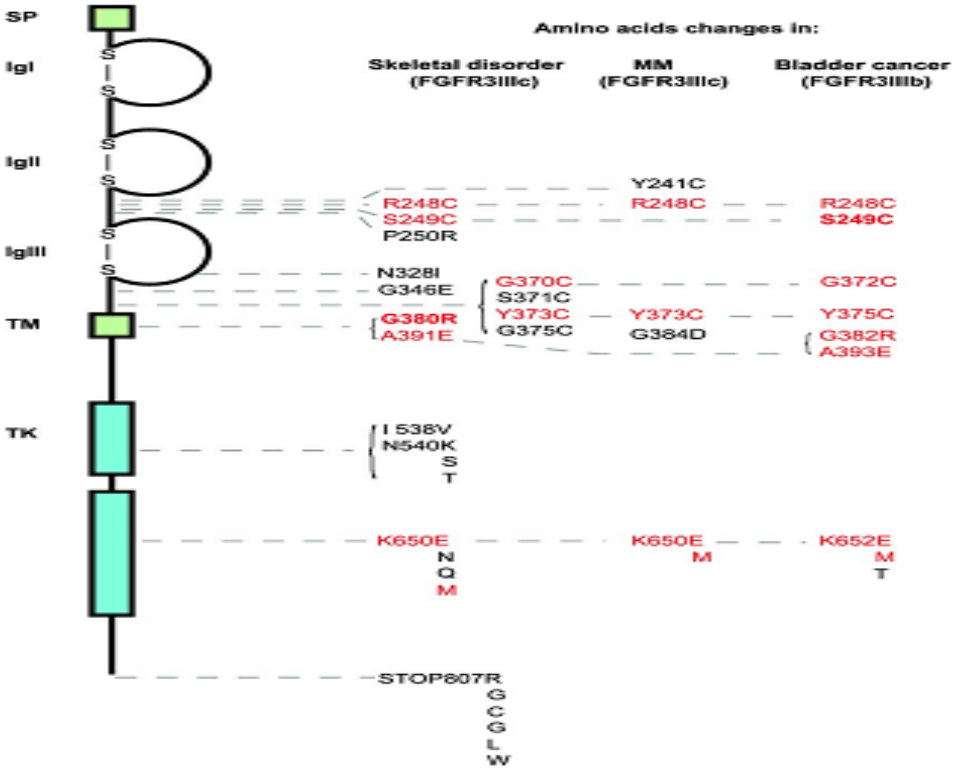


Figure 2.2.4: Most common forms of the FGFR3 mutations

By targeting the residues (248 and 167-171) that are responsible for receptor dimerization and signal transmission, theoretically it is possible to inhibit R248C caused cancers by using receptor specific inhibitors. We are planning to create peptide based inhibitor to cure R248C mutated FGFR3 caused benign skin tumors by the help of computational tools, molecular docking and molecular dynamics.

3. METHODOLOGICAL BACKGROUND

3.1 Docking

Currently, the number of protein structures determined experimentally and obtained by structural genomics is increasing and becoming accessible online [20-21]. Additionally computational tools are increasingly used in the drug design process with the help of the continuous rise in computer power and available methodologies [22]. Molecular docking is a method used to define the structure of the bound drug to active site of a target protein. Since its discovery in the 1980s, docking has been the leading method in the primary step of the drug design process [22-23]. Docking involves two steps: First, correct binding conformation of the drug should be estimated. Second, binding affinity is quantified by a scoring function. Thus, docking protocol can be summarized as a combination of a search algorithm and a scoring function. Reaching effective guess of the binding mode prediction and high speed is the key issue for docking [24].

The bound mode of a docking algorithm is evaluated by a root mean square deviation (RMSD) from experimental results, lower RMSD compared to experimental mode indicates a better docking result. Ligands have high degree of freedom, which results in flexibility as a major point in binding mode prediction [25]. In early works, both ligand and protein were considered rigid bodies. Later several approaches are generated in which the conformational space of the ligand or ligand-protein is calculated as seen in (Table 3.1.1) [24].

Table 3.1.1: Types of flexible ligand, flexible receptor search algorithms [24]

FLEXIBLE-LIGAND DOCKING	**FLEXIBLE-PROTEIN DOCKING**
Systematic	Molecular dynamics (MD)
Conformational	Monte Carlo (MC)
Database	Rotamer libraries
Random/Stochastic	Protein ensemble grids
Monte Carlo (MC)	Soft receptor modeling
Genetic algorithm	
Tabu Search	
Simulation Methods	
Molecular dynamics (MD)	
Energy minimization	

To search the conformational space of a ligand, either systematic, random or stochastic simulation methods are used. Systematic algorithms rotate all rotatable bonds 360°, so all possible search space is evaluated. This type of usage is limited due to a high dimension problem. In a random search, several random samples are taken from conformational space of the ligand. Monte Carlo, Genetic Algorithm and Tabu search are examples of this method. Simulation methods depend on the calculation of Newton's equation of motion. Molecular dynamics and energy minimization are considered members of the simulation method. Crossing the energy barriers and sampling the conformational space in a feasible time are the drawbacks of simulation methods [24].

Flexible ligand search methods have given reliable results in almost half of the studies in which they were used [26]. Many proteins have a conformational change in some side chains and loops upon ligand binding. This kind of receptor flexibility is considered in recent work for docking, which is achieved by MD and MC methods [27], rotamer libraries [28], protein ensemble grids [29] and soft-receptor modeling [30].

Scoring functions are used for determining the ranking of the binding free energies of the ligand bound to the receptor. However, having correct binding mode is not enough; it is also necessary to separate false positives for a successful docking process. Since the computational time increases with a very detailed scoring function, using an acceptable simplified method is a limiting step for the docking process. To achieve this, different scoring functions are used, such as force- field based, empirical and knowledge-based scoring [24].

Docking is a well established method, but it is far from being excellent [31]. In recent studies having an RMSD difference from the experimental structure about 1.5–2 Å is reported with a 70-80 % success rate [32]. However, problems at incorporation of an exact scoring function, considering the solvent effect, low resolution of targets and protein flexibility upon binding are key issues to be solved in a docking process [33].

Although many promising results have been obtained from docking computations, the algorithms are still far from perfect due to inefficient combination of accuracy and speed. Hence, docking results should be accepted as a starting point for the drug design process and further validation methods should be taken into account like molecular dynamics [24].

There are many docking programs with different skills and performance. As seen in (Figure 3.1.1) AutoDock [34], GOLD [35], FlexX [36], DOCK [37], and ICM [38] are the most common docking programs according to citation rates up to 2005.

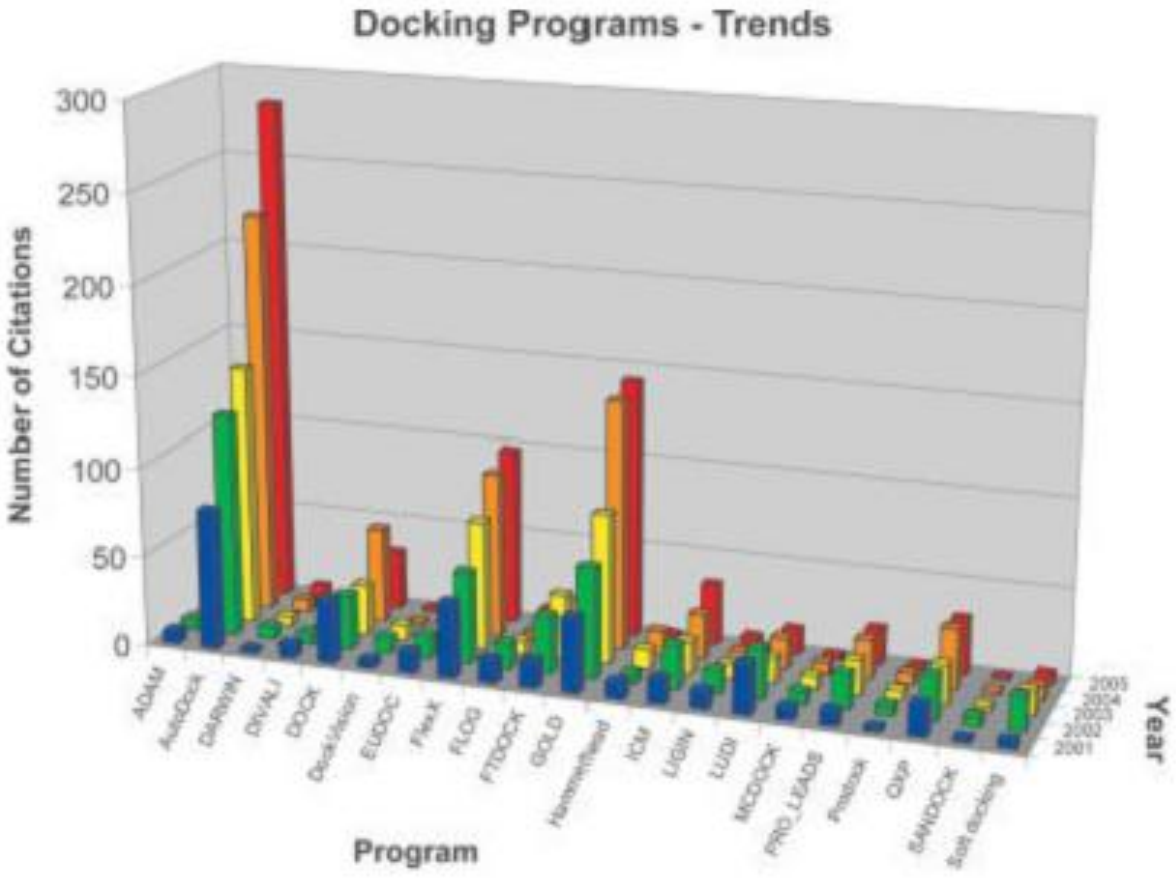


Figure 3.1.1: Docking programs and the number of citations per year for the most common docking programs, analyzed from ISI Web of Science (2005) by the first references taken into account [24].

AutoDock 4.0 uses a gradient optimization and local optimization procedure to find minimum energy binding conformation of a ligand to a target protein. By using molecular

mechanics force fields Autodock 4.0 gives a scoring function for each docking simulation. The scoring function used at simulations defined by a semiempirical free energy force field, which consists of the following evaluations in Appendix 1 [34]:

AutoDock 4.0 can be used with supercomputers and it is the most cited docking tool as indicated at work by Sousa S. F. et al. (Figure 3.1.1) [24]. Being free for academic users, AutoDock 4.0 is chosen as the docking tool of this study.

3.2 Tools

Proteins' 3D structures are important for calculation of the functional characterization in computational studies. While experimental structure is not available, comparative modeling and threading are generally used with sequence information [39]. Even comparative methods give reliable structure results by using the high resolution experimental studies as a template, the loop structures often are not the same. Additionally loops are found at the surface and mostly have a function at the activity of the proteins [40]. The major functional differences between the same protein families are achieved by the variety of the loop structures. Hence, it is important to define missing loops of the experimental structures while investigating ligand–protein interactions [41].

ArchPRED is an online loop prediction server that uses fragment based search method. By the server, it is possible to predict loops of length 4,8 and 12 with the covarage of 98, 78 and 28% with at least 0.22, 1.38 and 2.47 Å RMSD accuracy, respectively [41].

In our study, ArchPRED server is used to generate missing loops of our protein FGFR3.

3.3 Molecular dynamics

From the first molecular dynamics simulation of a bovine trypsin inhibitor at 1977, there are rapid improvements in computational sources and the force fields. Since then, simulation time has increased from picoseconds to microseconds. There are different applications of the simulations at the macromolecular area like down below:

- Determining configuration space of the molecule to refine experimental data.
- Obtaining thermodynamic, structural and motional properties of the system.
- Getting actual dynamic properties of the system

Steered molecular dynamics (SMD) applies time dependent external forces to define various aspects of proteins like ligand binding and elastic properties. This method is used in biological systems to investigate mechanical properties of proteins, antigen antibody interaction, free energy analysis and conduction of ions through membrane channels [42].

Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) and molecular mechanics-generalized Born surface area (MM-GBSA) are used to calculate free energy of binding by considering ligand, receptor and complex [43]. There are articles reporting correlations with the experimental studies of inhibitor design by using these methods [44].

MD is a powerful technique to get information about particle motions of a given macromolecule. By using well defined energy functions it is possible to determine position of each atom for a given time period. Atoms' positions are calculated based on Newtonian equation of motion from the following in Equation (3.3.1):

$$m_\alpha \ddot{\vec{r}}_\alpha = -\frac{\partial}{\partial \vec{r}_\alpha} U_{tot}(r_1, r_2, ..., r_N), \qquad \alpha = 1,2,...N \qquad \text{Equation (3.3.1)}$$

In Equation (3.3.1) m_α is the mass of the particle, r is the position of the particle, U_{tot} is the total potential energy which depends on positions of all atoms. N is the number of atoms in the system. U_{tot} is calculated by a potential energy function given as in Equation (3.3.2):

$$U_{tot} = U_{bond} + U_{angle} + U_{dihedrals} + U_{Coulomb} + U_{vdW} \qquad \text{Equation (3.3.2)}$$

First three terms in Equation (3.3.2) come from intramolecular interactions:

$$U_{bond} = \sum_{bonds} K_b (b - b_{eq})^2 \qquad \text{Equation (3.3.3)}$$

$$U_{angle} = \sum_{angle} K_\theta (\theta - \theta_{eq})^2 \qquad \text{Equation (3.3.4)}$$

$$U_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \qquad \text{Equation (3.3.5)}$$

Here K represents a constant for bonds and bond angels. b is bond length, θ bond angle. beq and θeq represent bond length and equilibrium bond angle. V is a force constant, φ is dihedral angle and γ is the phase angle. Last two terms in Equation (3.3.2) come from non-bonded interactions between atoms:

$$U_{Coulomb} = \sum_{i<j} \left[\frac{q_i q_j}{\varepsilon R_{ij}} \right] \quad \text{(Equation 3.3.6)}$$

$$U_{vdW} = \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] \quad \text{(Equation 3.3.7)}$$

In Equation (3.3.6) and Equation (3.3.7) qi and qj are partial charges of atoms i and j. ε is the dielectric constant. Aij and Bij are constants differing for different atom interactions and Rij is the distance between the i_{th} and j_{th} atoms. In general, nonbonded interactions are calculated for all atom pairs by using Lennard Jones potential. Nonbonded interactions are computed for all non bonded atom pairs using and Coulomb potential for electrostatic interactions.

During molecular dynamics simulations constraints are applied for the following purposes: Increasing user defined potential energy surface searching, enforce boundary forces to stabilize water molecules, prevent high frequency vibrations [45].

Molecular dynamics simulations can be done at different system conditions. At NVT simulations, also known as canonical ensemble, the number of particles (N), the volume (V) of each system in the ensemble is the same. Additionally the ensemble has a well defined temperature (T), given by a heat bath in which is in equilibrium.

At NPT simulations, also known as isothermal–isobaric ensemble, temperature (T), pressure (P) and number of particles (N) are kept constant.

3.3.1 Steered Molecular Dynamics:

Steered molecular dynamics are widely used for investigating binding mechanism of a ligand, protein unfolding and protein – protein interactions. In a non equilibrated system the external work done on the system at time interval 0 to τ is defined as:

$$W(t) = \int_{z(0)}^{z(t)} F dz, \qquad \text{(Equation 3.3.1.1)}$$

In the Equation (3.3.1.1), $W(\tau)$ represents the work done by the pulling external force. F = F (t) is the force applied to the spring and z = z (t) is the distance between the two force centers.

The second law of thermodynamics states that the average work recorded cannot be smaller than the difference of free energies between the initial and the final states, $\Delta F \leq \langle W \rangle$. The equality in this expression is valid only under the conditions, which the process is quasistatic or reversible, work is independent of the path. A detailed representation is achieved by the extended form of Jarzynski identity (JI) as follows [46] :

$$\exp(-\beta \Delta F) = \langle \exp(-\beta W) \rangle \qquad \text{Equation (3.3.1.1)}$$

Equation (3.3.1.1) states that the free energy difference can be expressed as the exponential average of the work done in the simulation. Here ΔF represents the binding free energy, W is the work done during simulation. B equals to $1 / k_B T$, in which k_B is the Boltzmann's constant. The angular brackets $\langle\ \rangle$ shows an average over an ensemble of nonequilibrium process.

ΔF can be obtained from the equation above by using the average work of steered molecular dynamics simulation of different converged trajectories [47].

3.3.2 Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) Method:

The binding equilibrium of ligand to a receptor in a solution can be shown as follows:

$$R_{aq}^{-} + L_{aq}^{-} \Leftrightarrow C_{aq}^{-} \qquad \text{Equation (3.3.2.1)}$$

In this equilibrium in Equation (3.3.2.1), R stands for receptor, L stands for ligand and C stands for the complex formed upon binding. Binding reaction results in a free energy difference which comes from the binding free energy of the ligand. Experimentally binding energy of a ligand is calculated like down below:

$$\Delta G = RT \ln(K_d) \qquad \text{Equation (3.3.2.2)}$$

In Equation (3.3.2.2), R represents the molar gas constant, T is the temperature and Kd is the dissociation constant experimented. The binding free energy of the ligand can be also computed [48] as the free energy difference between complex, receptor and ligand is shown with the following Equation (3.3.2.3):

$$\Delta G_{\text{binding}} = \Delta G_{\text{water}}\,(\text{complex}) - [\Delta G_{\text{water}}\,(\text{protein}) + \Delta G_{\text{water}}\,(\text{peptide})] \qquad \text{Equation (3.3.2.3)}$$

The binding free energies can be estimated from the change in molecular mechanical energies (ΔE_{MM}), the solvation free energies ($\Delta G_{PB} + \Delta G_{nonpolar}$) and the vibration, rotation and translation entropies. Each of these terms can be restated as follows in Equation (3.3.2.4):

$$\Delta G_{water} = \Delta E_{MM} + \Delta G_{solvation} - T\Delta S \qquad \text{Equation (3.3.2.4)}$$

Here ΔE_{MM} is the average energy of the solute, composed of bonded and non-bonded molecular mechanics interactions:

$$\Delta E_{MM} = \Delta E_{Bond} + \Delta E_{Angle} + \Delta E_{Torsional} + \Delta E_{vdW} + \Delta E_{Coulomb} \qquad \text{Equation (3.3.2.5)}$$

In Equation (3.3.2.5) ΔE_{Bond} ΔE_{Angle} and $\Delta E_{Torsional}$ contributions come from the change in bonded interactions. ΔE_{vdW} and $\Delta E_{Coulomb}$ come from change in non bonded interactions.

The term $T\Delta S$ in Equation (3.3.2.4) represents the entropic contribution of the solute. T shows temperature and ΔS describes change in the entropy, obtained from molecular mechanics. ΔS can be written in a detailed form as follows in Equation (3.3.2.6):

$$\Delta S = \Delta S_{Rotational} + \Delta S_{Translational} + \Delta S_{Vibrational} \qquad \text{Equation (3.3.2.6)}$$

Here $\Delta S_{Rotational}$, $\Delta S_{Translational}$ and $\Delta S_{Vibrational}$ stand for the change in rotational, translational and vibrational motions of the solute respectively.

In Equation (3.3.2.4) the term $\Delta G_{solvation}$ represents the solvent contribution to free energy. It has two components as follows:

$$\Delta G_{solvation} = \Delta G_{Polar} + \Delta G_{Nonpolar} \qquad \text{Equation (3.3.2.7)}$$

Here ΔG_{Polar} stands for the polar contribution of the solvent and is calculated by Generalized Born (GB) method in this study. The $\Delta G_{Nonpolar}$ represent the non-polar contribution and is also computed from solvent accessible surface area (SASA) [49]:

$$\Delta G_{Nonpolar} = \gamma SASA + \beta \qquad \text{Equation (3.3.2.8)}$$

Here γ stands for surface tension, SASA stands for the solvent accessible surface area of the solute and β stands for an offset value.

4. METHODS

4.1 Preparation of Initial Coordinate Files:

Currently the structure of FGFR3 is released in the Protein Data Bank (PDB) at 2.10 Å with code 3GRW [50]. The complex contains 3 chains; FGFR3, Fab antibody chains L and H. In order to design a specific inhibitor to FGFR3, L and H chains were deleted from the PDB file. Disease forming R248C mutant is obtained by deleting the residue R248 from the PDB file and using ArchPRED online server to predict coordinates of the missing 248C residue.

4.2 Molecular Dynamics:

NAMD molecular dynamics package [51] is used with two different sets of force fields, as follows:

Firstly the NAMD simulation package is used with the CHARMM 27 parameter set [52-53]. The mutated structure is solvated at waterbox by using Tk console at VMD. TIP3P atomistic water model [54] is used for the solvation. In order to have enough pulling space for SMD simulations, appropriate number of water molecules is added to the system. Three Cl^- ions are added to neutralize the system.

Minimization is achieved by using the steepest descent energy minimization procedure followed by the molecular dynamics simulation gradually annealed from 10 K to 310 K in a simulation period of 10 ps.

Periodic boundary conditions are used during equilibration and run period of the system. In order to keep the bond length of the water molecules fixed, the SHAKE algorithm [55] is applied with a rigid bond tolerance of 10^{-5} Å. The long-range Coulombic

interactions are calculated using the Particle Mesh Ewald (PME) summation method with box size 64 x 64 x 120 [56]. The non-bonded cutoff distance is applied as 12 Å during the simulation. Integration times of the simulations are chosen as 1 fs and coordinates and energies were obtained at every 1 ps. Temperature is controlled by using a Langevin thermostat with a coefficient of 5 / ps. Simulations are performed at isobaric-isothermal ensemble (NPT), constant pressure control is applied to the systems. Constant pressure is obtained by using a barostat Langevin piston pressure at 1.01325 bar on the basis of Nose-Hoover method [57-58] with oscillation time of 100 fs and a barostat damping time of 50 fs and a barostat noise temperature of 310 K.

In order to obtain a stable conformation of R248C mutated form of FGFR3, a 13.6 ns molecular dynamic simulations is performed with CHARMM 27 parameter set.

Secondly, the NAMD simulation package is used with the AMBER force field for bioorganic molecules (ff03) [59]. The mutated structure is loaded to the Leap module of AMBER 10 molecular dynamics simulation package [60-61] in PDB file format. The molecule is solvated in a water box with the TIP3P atomistic water model where the distance between the edges of the water box and the protein was set to 12 Å [54]. The system is neutralized like CHARMM 27 parameter set by adding three Cl^- ions.

Minimization is achieved by using the steepest descent energy minimization procedure followed by the molecular dynamics simulation gradually annealed from 10 K to 310 K in a simulation period of 25 ps. Other parameters are applied similar to the simulation with CHARMM 27 parameter set.

In order to obtain a stable conformation of R248C mutated form of FGFR3, 3.6 ns molecular dynamic simulations are performed with the AMBER force field.

4.3 Preparation of the peptide sequences:

Firstly, the antibody used at previous work, which successfully inhibits FGFR3 and mutated forms of FGFR3 [50], is carefully investigated. Antibody's FGFR3 interacting residues are extracted as IYDLY. Different residues of the peptide are mutated and ten different sets of peptide sequences are obtained for docking procedure.

HyperChem molecular building tool [62] is used for generation of the peptides. All structures are minimized during 2 ps at 310 K.

4.4 Docking

The receptor structures are obtained from two different molecular dynamics simulations which are conducted with the CHARMM 27 parameter set and the AMBER force field for bioorganic molecules (ff03).

Residues 167-171 are chosen as the binding site residues which are indicated at work by Qing et al. [50]. AutoDock Tools 1.5.2 is used for generation of simulation files. The following docking parameters are used for each peptide: The torsional degree of freedom is 24, the population size is 150, the maximum number of energy evaluations is 250000, the maximum number of generations is 27000, the rate of gene mutation is 0.02 and the rate of crossover is 0.8.

The lowest binding mode structures are used for the following steered molecular dynamics and MM-GBSA binding energy calculations.

4.5 Binding free energy calculation methodology:

The following procedural details are used for molecular dynamics binding free energy calculations with the CHARMM 27 parameter set and the AMBER force field for bioorganic molecules (ff03) on the NAMD simulation package after adding –COOH group to N terminus and -CH_3 to C terminus of the peptides to prevent degradation and to increase stability:

4.5.1 Steered Molecular Dynamics with the CHARMM 27 force field:

For the CHARMM 27 force field, minimization is achieved with the same parameters used with the molecular dynamics simulations before the docking process. At the run period of the simulations the following parameters are different from the pre-docking process. PME is used with box size 96 x 90 x 144. Harmonic constraint is used to fix the protein and to allow water molecules to converge without causing any effect on the protein. Harmonic constraint is decreased by setting k constant from 1 to, 0.5, 0.25, 0.125 for each 0.5 ns NPT molecular dynamics run.

The system's volume is converged by 14.88 ns NPT molecular dynamics run. Then system is run for an additional 5.12 ns NVT simulation with the same parameters. The system is converged after 19.4 ns of total MD run. From the converged trajectory different snapshots are recorded as PDB at time steps 19.4, 19.55, 19.70, 19.85 and 20 ns for the steered molecular dynamics calculations.

For each snapshot the residues 247 and 248 of the protein are fixed and the residues 3 and 5 of the peptide are chosen for SMD simulation pulling. Vector crossing through these atoms is defined for the pulling direction of the SMD simulation. Constant velocity SMD simulations are used with a force constant of k = 7 pN / Å and and velocity 0.00001 Å/fs until work graphics are converged.

For reference peptide, similar parameters are used for MD simulation before SMD process.

The system's volume is converged after 19.47 ns NPT molecular dynamics run. Additionally the system is run for 12.53 ns NVT simulations with the same parameters. Totally 32 ns simulation is done for reference peptide-protein complex. From the converged trajectory five different snapshots are taken for the SMD simulation at time steps 31.4, 31.55, 31.70, 31.85 and 32 ns.

SMD simulation parameters for the reference peptide are the same for the drug candidate sequence.

4.5.2 Steered Molecular Dynamics with Amber force field

For the AMBER force field, minimization is achieved with the same parameters used with the simulations done before the docking process during 25 ps. Only the PME box size is different, used as 75 x 75 x 125 for the system. Additionally, the simulation parameters are also the same with the simulation parameters used before the docking process.

The system's volume converged after 25.7 ns NPT molecular dynamics simulation. Then system is simulated an additional 5.8 ns NVT run with the same parameters. The peptide receptor complex is converged after 5 ns NVT run. From the converged trajectory, different snapshots are recorded as PDB at time steps 30.7, 30.9, 31.1, 31.3 and 31.5 for the steered molecular dynamics calculations.

For each snapshot the residues 247 and 248 of the protein are fixed and the residues 3 and 5 of the peptide are chosen for steered MD simulation as it is done with the CHARMM 27 force field. Vector crossing through these atoms is defined for the pulling

direction of the steered MD simulation. Constant velocity SMD simulations are used with a force constant of k=7 pN / Å and velocity 0.00001 Å / fs until work graphics are converged.

4.5.3 MM-GBSA analysis

For the MM-GBSA analysis, the converged trajectory of the peptide-protein complex between 15 ns and 24.5 ns from NPT run is recorded. From the recorded trajectory, 177 snapshots are generated with equal intervals.

GB analysis is conducted by using AMBER MM-PBSA depending on the modified GB model which is proposed by Onufriev et al. [63]. The internal dielectric constant is taken as 2 and the external dielectric constant is as 80. Parse radii [49] and Duan et al. charges are used and modified Bondi radii [64] is used as 2 Å for GB analysis.

SASA calculations are conducted with the Molsurf [63], which is a part of the AMBER simulation package. LCPO method is used with γ value taken as 0.005 and β value taken as 0 for nonpolar contributions.

Entropic terms of binding are calculated by NMODE module of the AMBER simulation program. All 177 converged snapshots at the NPT run are used with the distance dependent dielectric constant $\varepsilon = 4Ri$.

5. RESULTS:

5.1 Steered molecular dynamics (SMD) simulations:

The R248C mutated form of FGFR3 is simulated with the CHARMM 27 force field at NAMD during 13.6 ns, in order to prepare the FGFR3 receptor for docking. As there are global motions of the protein during simulations, the C_α RMSD value is calculated for domain II and domain III separately. First, domain III is fixed and the RMSD of domain II is calculated. Second, domain II is fixed and the RMSD of domain III is calculated. The RMSD graphics are obtained as seen in (Figure 5.1.1 and 5.1.2):

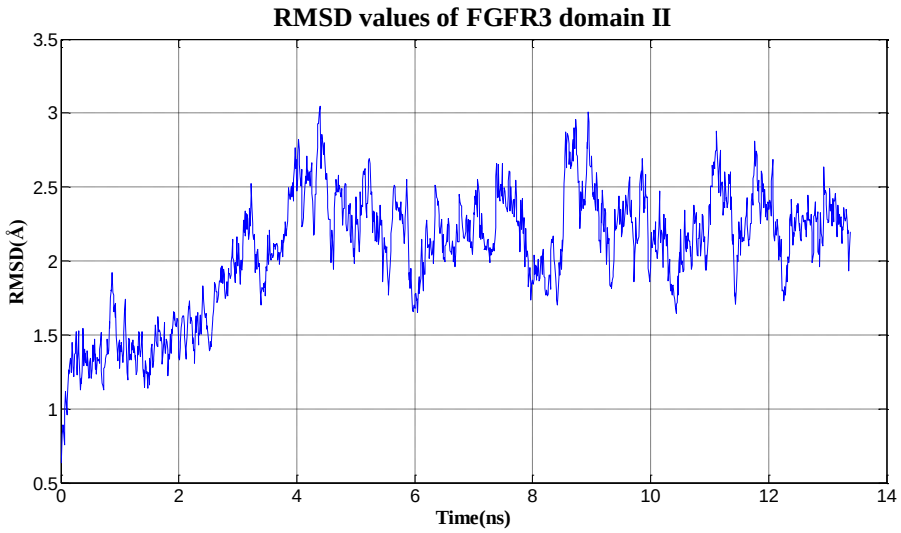


Figure 5.1.1: RMSD values of the FGFR3 domain II during MD simulations before docking

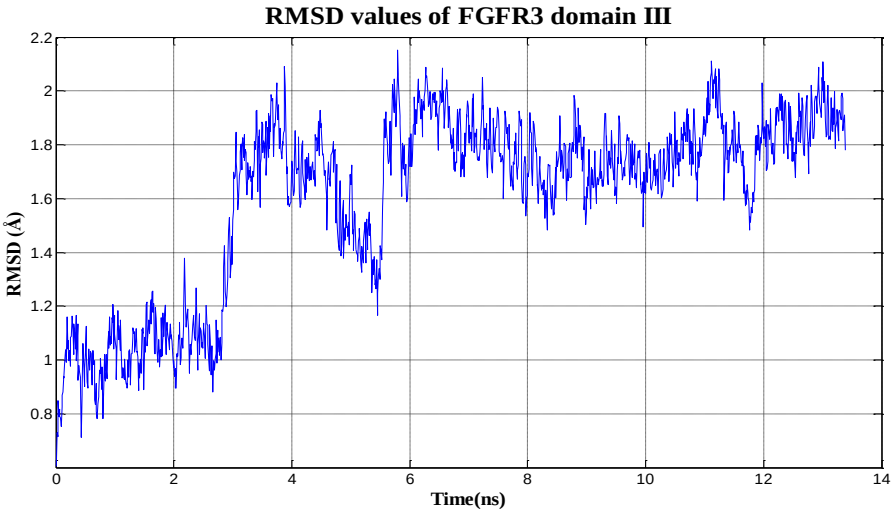


Figure 5.1.2: RMSD values of the FGFR3 domain III during MD simulations before docking

As the general structure of the protein is converged, the final structure is recorded as PDB for docking simulations.

From the docking simulations semi-empirical binding energies are obtained for the mutated forms of the peptide IYDLY which is the part of the antibody inhibiting the FGFR3 as indicated in Qing et al.'s investigation [50]. Additionally**,** the work of Naski et al. [17] indicates that R248C mutation forms intramolecular disulfide bond causing ligand independent signal transmission. Thus for docking simulations, indicated sites (249, 167-171) are chosen for docking grid. The docking results are illustrated in (Table 5.1.1):

Table 5.1.1: Docking results for R248C mutated form of FGFR3 – pentapeptide sequences.

Peptide Number	**Sequence**	**Binding Energy (kcal/mol)**
1	IYDLY	-2.96
2	PYDLY	-4.07
3	IWDLY	-3.88
4	IYELY	-2.6
5	IYDMY	-10.04
6	IYDLF	-3.71
7	GFDLY	-4.06
8	VYNLY	-2.09
9	IYCHY	-12.67
10	IRDAY	-8.03
11	IYDMW	-10.71

Peptides having cysteine or methionine have lowest binding scores compared to the others. Lowest binding modes of docking results of the peptide sequences are investigated for 5, 9 and 11. The sequence 5 is chosen as it interacts with 167-171 and 248th residues of the FGFR3 as shown in Figure 5.1.3:

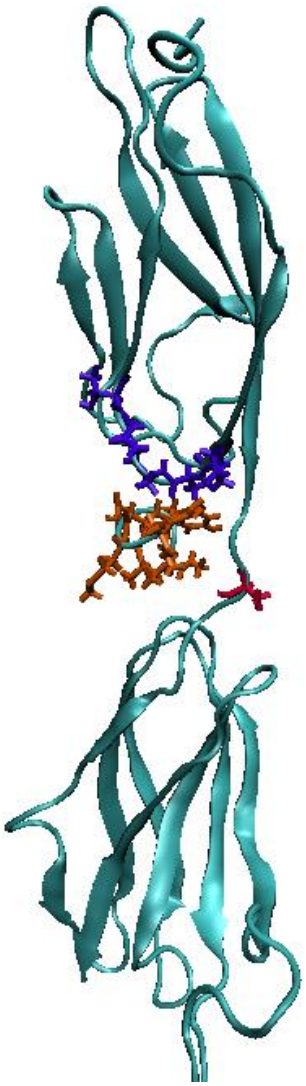


Figure 5.1.3: The docked conformation of the peptide 5 (shown in orange) to the R248C mutated form of the FGFR3 at VMD. Violet and red bond representatives show residues 167-171 and residue 248 respectively.

In order to increase cell membrane permeability and to prevent degradation –N terminus of the peptide is acetylated and –C terminus is methylated. The RMSD of the complex is converged after being simulated during 14.88 ns NPT and 5.12 ns NVT MD run. After 20 ns simulation, the RMSD values indicating a converged structure of the FGFR3 and for the whole complex is shown in (Figure 5.1.4), (Figure 5.1.5), (Figure 5.1.6) and (Figure 5.1.7).

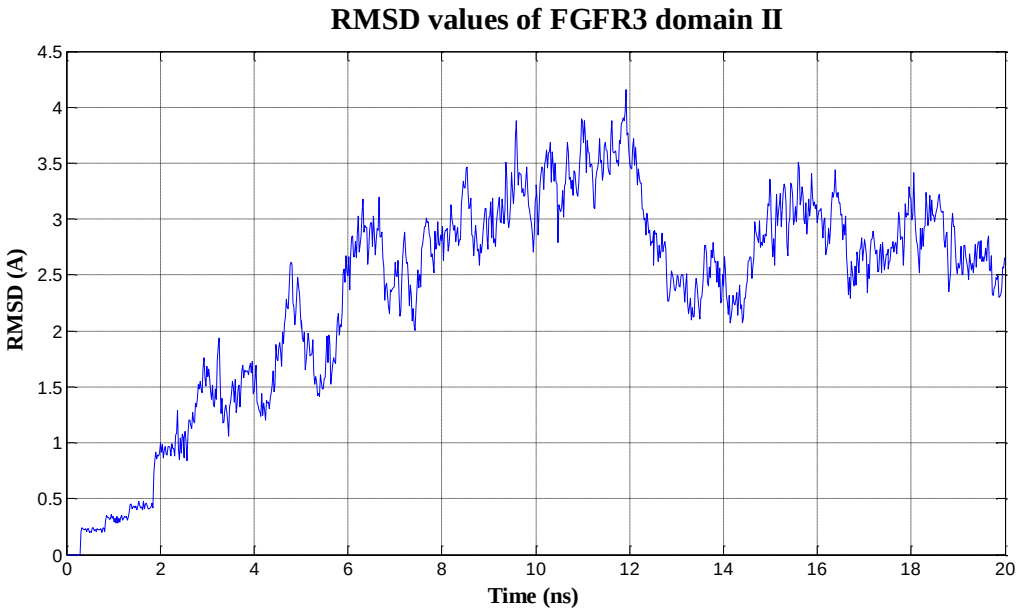


Figure 5.1.4: RMSD values of FGFR3 domain II during MD simulations after docking

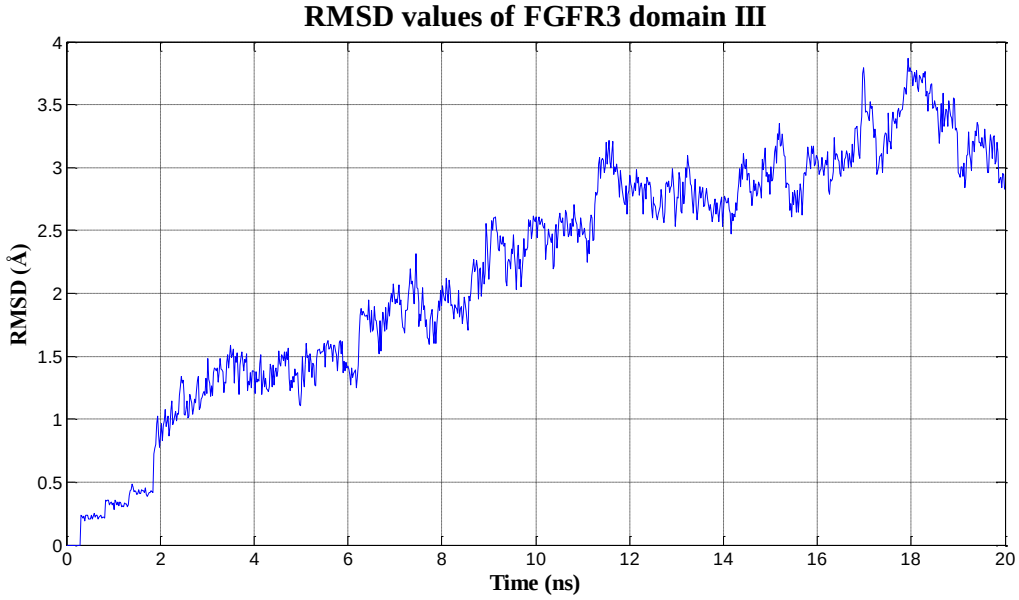


Figure 5.1.5: RMSD values of FGFR3 domain III during MD simulations after docking

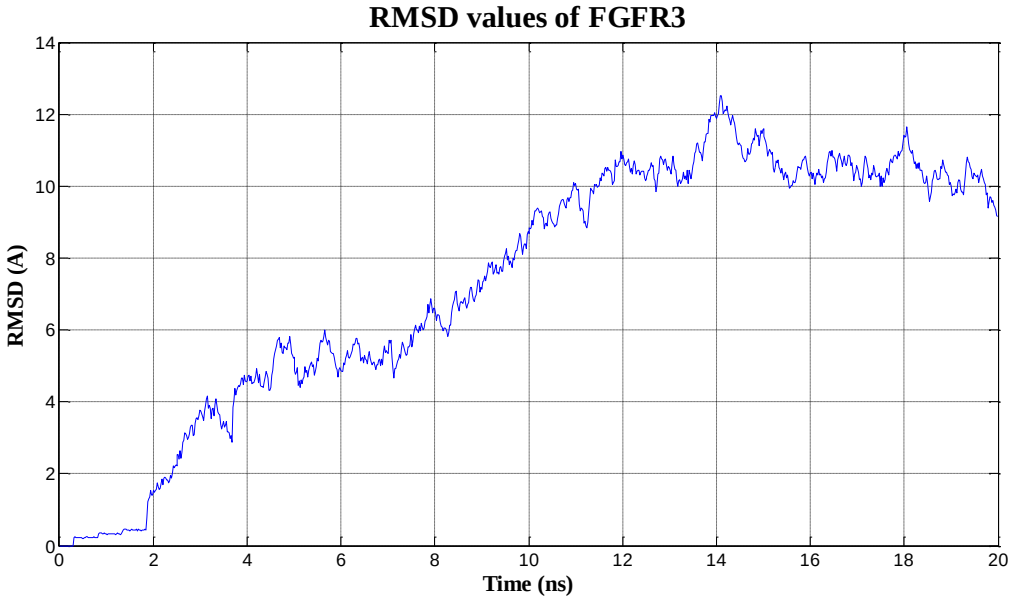


Figure 5.1.6: RMSD values of FGFR3 during MD simulations after docking.

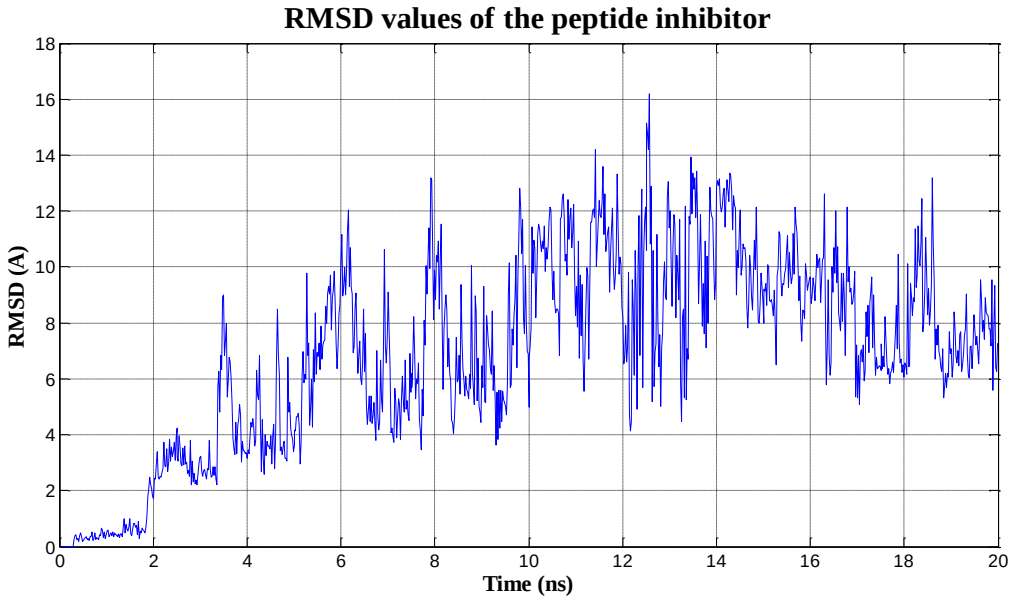


Figure 5.1.7: RMSD values of peptide inhibitor during MD simulations after docking.

Form **(**Figure 5.1.4), (Figure 5.1.5), (Figure 5.1.6) and (Figure 5.1.7) it is observed that the RMSD of the system is converged after 4.5 ns NVT, 19.4 ns total run. From the converged trajectory different snapshots are recorded as PDB at time steps 19.4, 19.55, 19.70, 19.85 and 20 ns for the steered molecular dynamics calculations. In (Figure 5.1.8), conformation of the complex recorded as PDB after 20 ns simulation is shown.

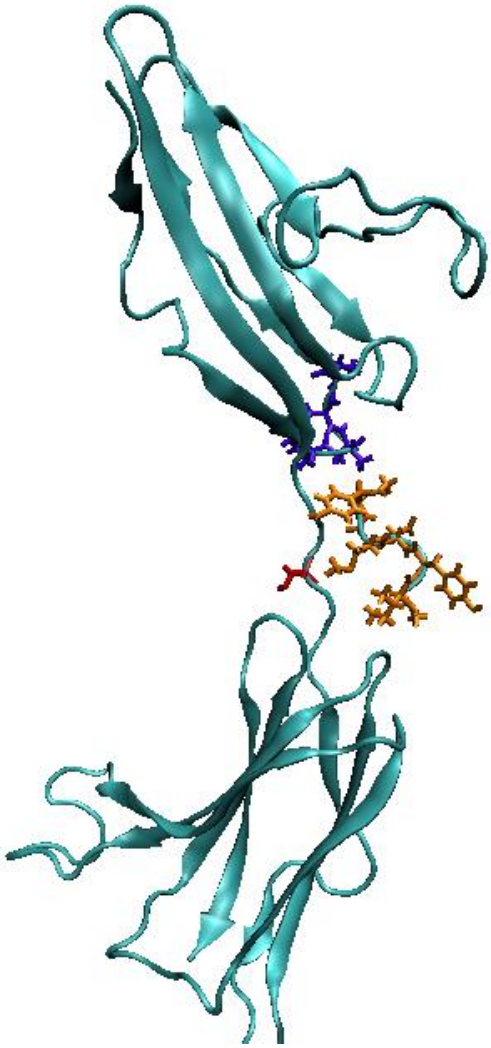


Figure 5.1.8: Conformation of the peptide 5 (shown in orange) bound to the R248C mutated form of the FGFR3 after 20 ns MD simulation. Violet and red bond representatives show residues 167-171 and residue 248 respectively.

Obtained PDB structures from MD simulations above are used for SMD simulations. By using the velocity constant as 0.00001 Å/fs and a force constant of k = 7 pN / Å SMD simulations are done for each snapshot.

Simulations are conducted until work values are stabilized for each simulation. The force and time step values are extracted from the output files for each simulation by normalizing the force vector through pulling direction via Tk console at VMD.

From the obtained time step and force values, work values are plotted by using $W = \int F \cdot v \cdot dt$ equation. Results obtained from 5 different SMD simulations are shown in (Figure 5.1.9):

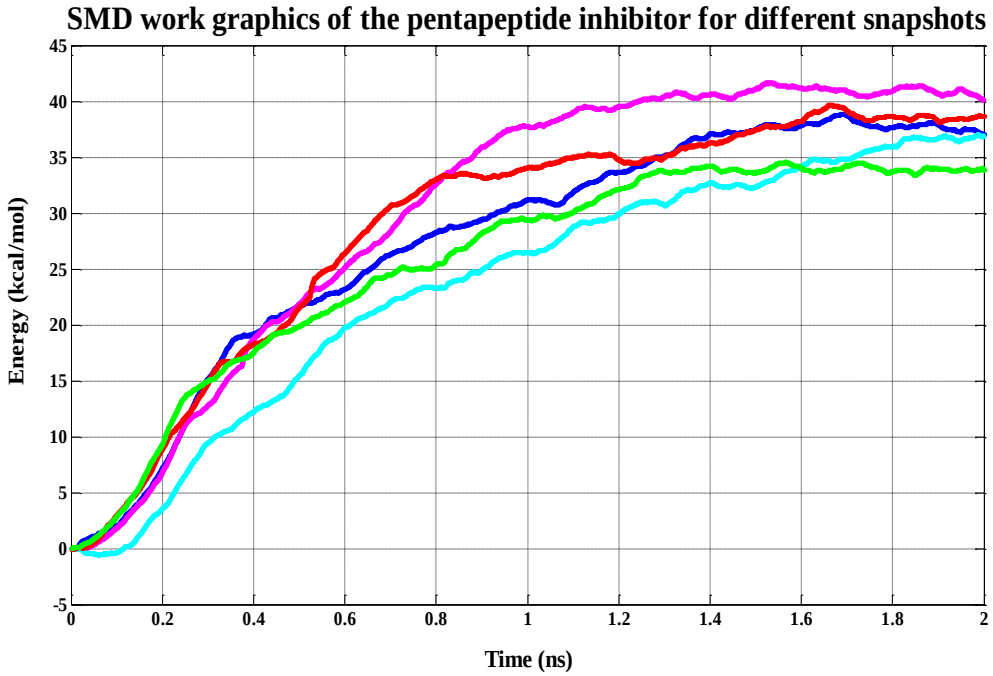


Figure 5.1.9: SMD work graphics of the peptide inhibitor from different snapshots. Here blue represents the snapshot at 19.4 ns, red for the snapshot at 19.55 ns, cyan for the snapshot at 19.70 ns, green for the snapshot at 19.85 ns and magenta for the snapshot at the 20 ns.

The five different work values obtained for each snapshot are used for free energy calculations. Jarzynski identity, which states that $\exp(-\beta\Delta G) = \langle \exp(-\beta W) \rangle$, is used for the calculations. Calculation results are shown in (Figure 5.1.10):

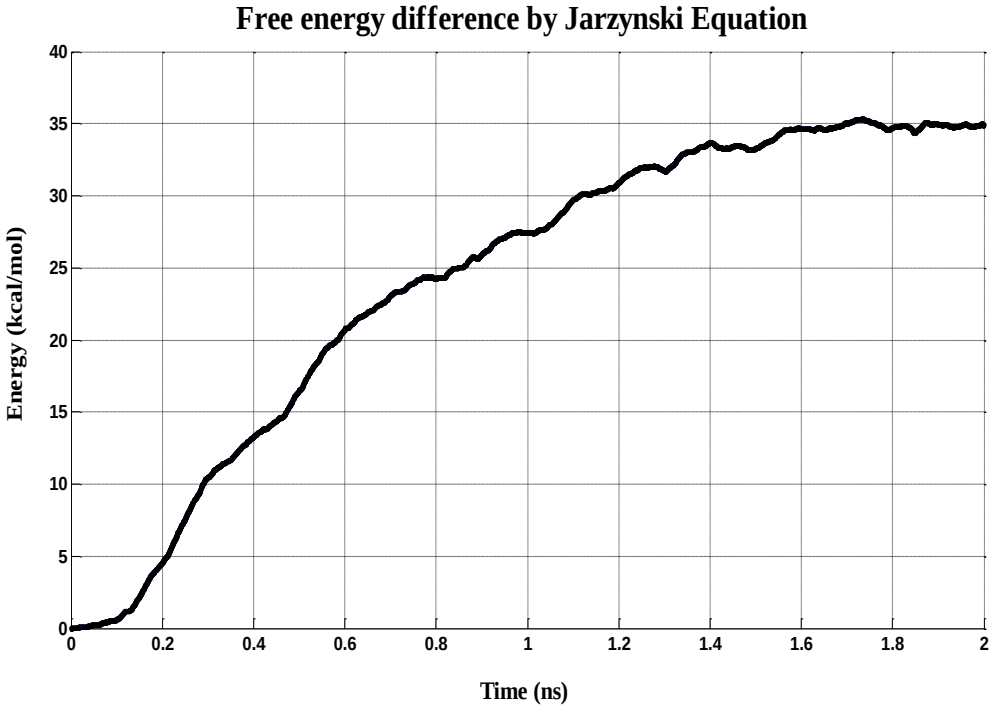


Figure 5.1.10: Free energy difference of the pentapeptide bound to the receptor, calculated from 2ns SMD simulation of the snapshots taken from 19.4, 19.55, 19.70, 19.85 and 20 ns of the NVT + NPT simulations.

From the graph in (Figure 5.1.10), after 2 ns simulation of each snapshot, it is found that the pentapeptide's binding free energy to the receptor is ΔF= -35 kcal/mol. Since there is ignorable volume change during simulations, binding free energy can be approximated to Gibbs free energy as ΔG=-35 kcal/mol.

5.2 Steered molecular dynamics with Amber force field:

The AMBER force field is used for SMD and MM-GBSA analysis to calculate free energy of the peptide inhibitor bound to R248C mutated form of the FGFR3.

Similar with previous runs done with the CHARMM force field, FGFR3 has global motions during simulations as it is done with the AMBER force field. Thus C_α RMSD values are calculated for domain II and domain III separately as seen in (Figure 5.2.1 and 5.2.2):

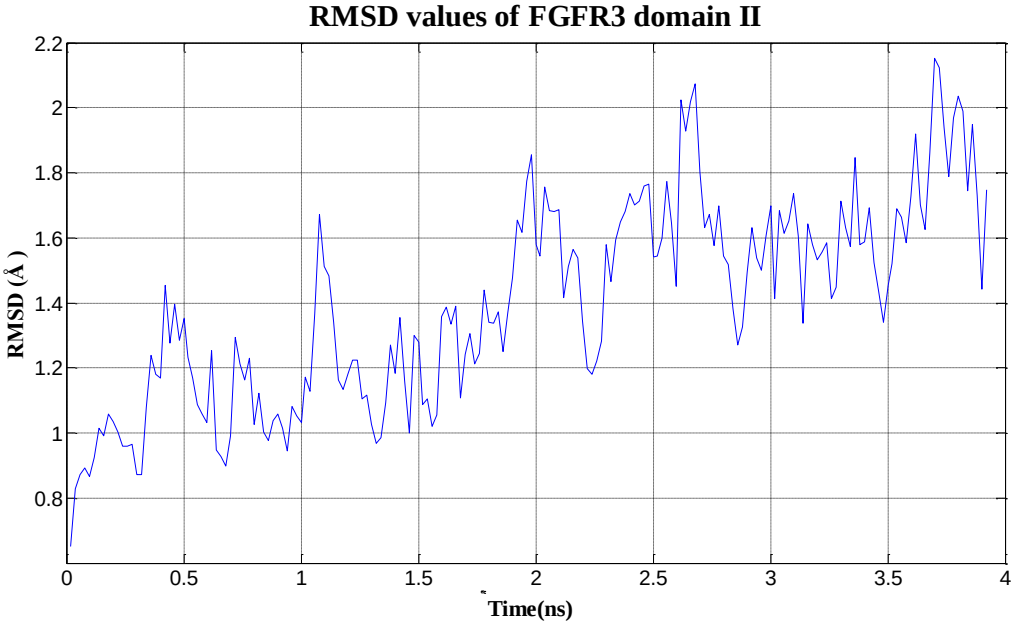


Figure 5.2.1: RMSD values of FGFR3 domain II during MD simulations before docking.

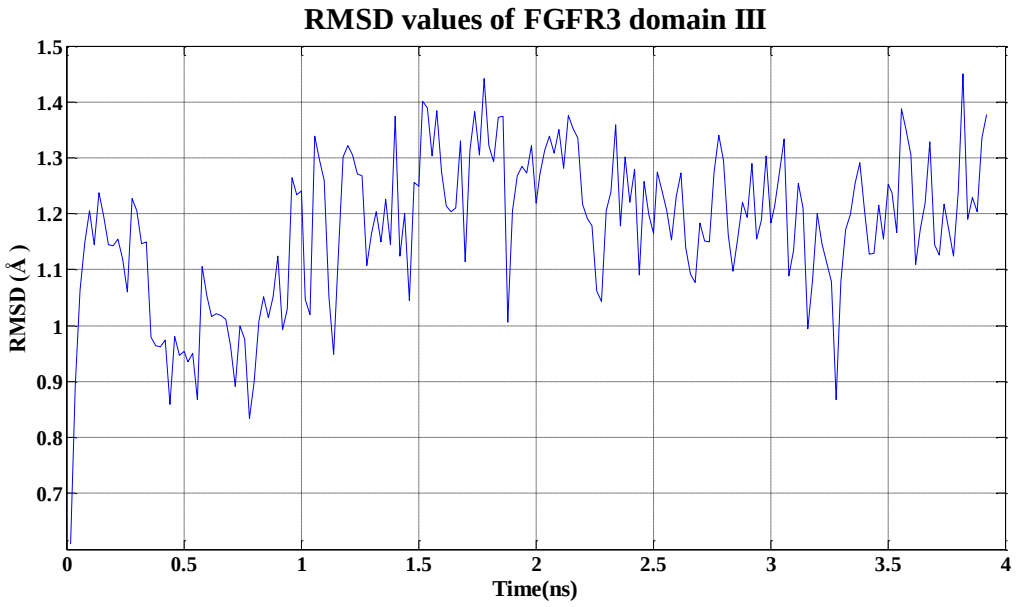


Figure 5.2.2: RMSD values of FGFR3 domain III during MD simulations before docking.

As the RMSD of the system is converged, the last snapshot is recorded as PDB to use for the docking calculations. Similar to the CHARMM force field, the docking grid is chosen as residues 167-171 and 249. The docking results of the peptides are

given in (Table 5.2.1). Compared with previously reported antibody's binding site sequence IYDLY, peptide 5 (IYDMY) again give better results in docking simulations.

Table 5.2.1: Docking results for R248C mutated form of FGFR3 - pentapeptide sequences.

Peptide number	Sequence	Binding Energy (kcal/mol)
1	IYDLY	-6.33
2	PYDLY	-6.01
3	IWDLY	-8.17
4	IYELY	-7.78
5	IYDMY	-7.71
6	IYDLF	-8.54
7	GFDLY	-8.93
8	VYNLY	-7.62
9	IYCHY	-8.27
10	IRDAY	-9.51
11	IYDMW	-7.46

Peptides 4, 5, 6, 7, 8, 9, 10 and 11 have lower binding free energies than peptide 1. Peptide 5 is again chosen and capped like the CHARMM 27 force field case in order to obtain free energy from SMD simulations with the AMBER force field and from MM-GBSA analysis with AMBER. By this approach, three different binding energy calculations are done by CHARMM 27 and AMBER force fields, also with MM-GBSA at AMBER for the inhibitor peptide.

The docked conformation of the peptide 5 for the AMBER force field SMD and MM-GBSA simulations is shown in (Figure 5.2.3).

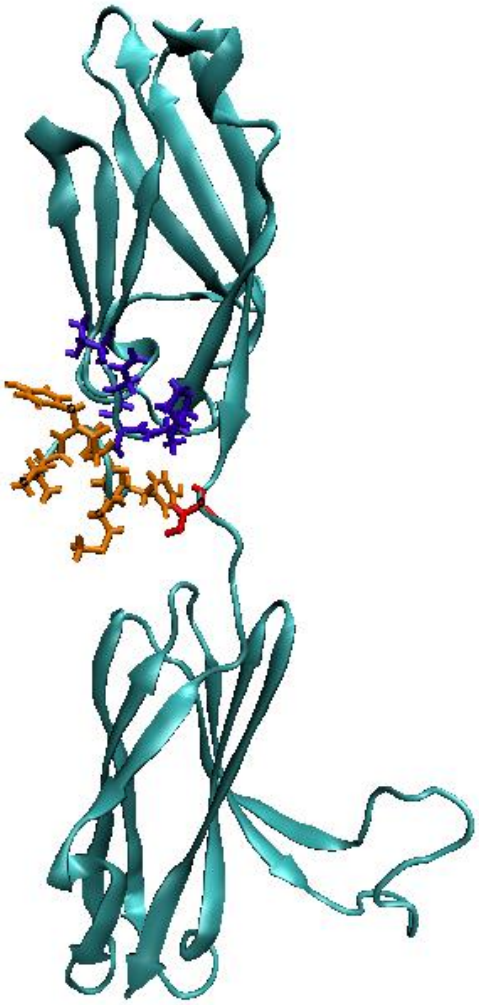


Figure 5.2.3: Docked conformation of peptide 5 (shown in orange) to R248C mutated form of the FGFR3 at VMD. Violet and red bond representatives show residue 167-171 and residue 248 respectively.

RMSD of the complex is converged after being simulated during 24.5 ns NPT and 5.8 ns NVT MD run. SMD simulations are conducted after the system converged to an equilibrium state.

RMSD values indicating a converged structure of the FGFR3 domain II, FGFR3 domain III, FGFR3 receptor and peptide inhibitor are shown in (Figure 5.2.4), (Figure 5.2.5), (Figure 5.2.6) and (Figure 5.2.7) respectively.

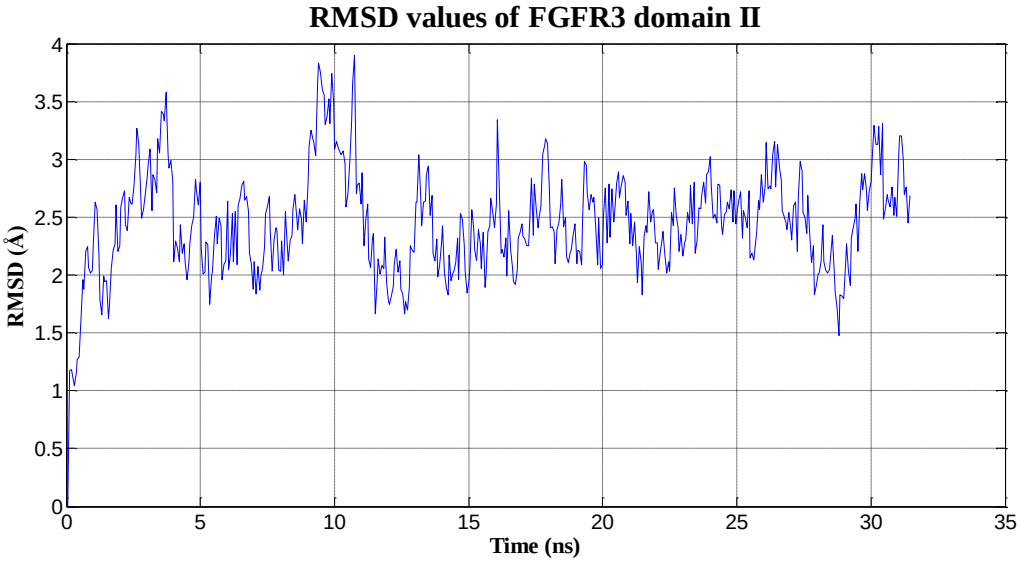


Figure 5.2.4: RMSD values of FGFR3 domain II during MD simulations after docking.

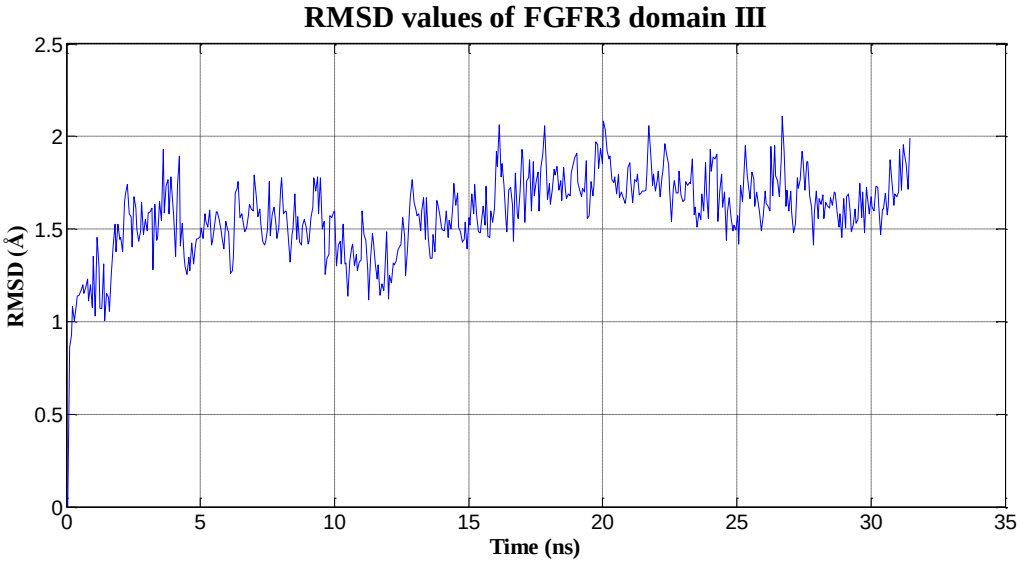


Figure 5.2.5: RMSD values of FGFR3 domain III during MD simulations after docking.

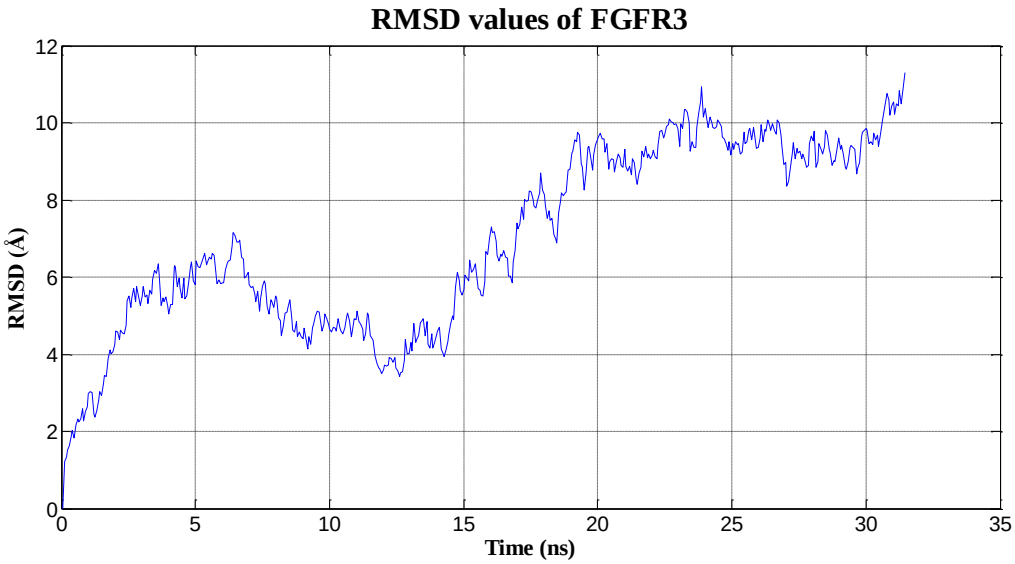


Figure 5.2.6: RMSD values of FGFR3 during MD simulations after docking.

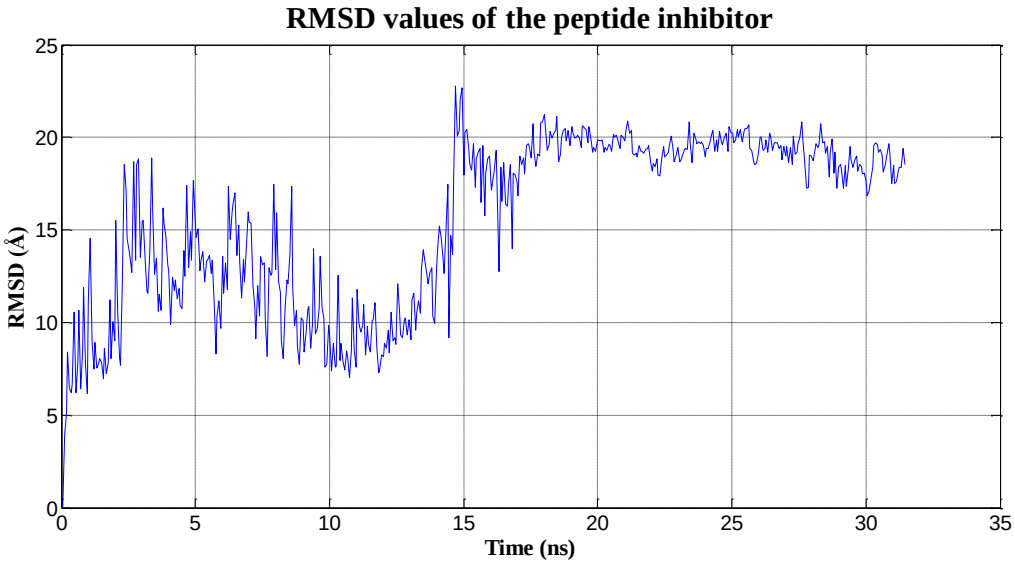


Figure 5.2.7: RMSD values of peptide inhibitor during MD simulations after docking.

The last 30.7, 30.9, 31.1, 31.3 and 31.5 ns steps snapshots are recorded as PDB for the SMD analysis. Snapshot of the 31.5 ns of the simulation is shown in (Figure 5.2.8).

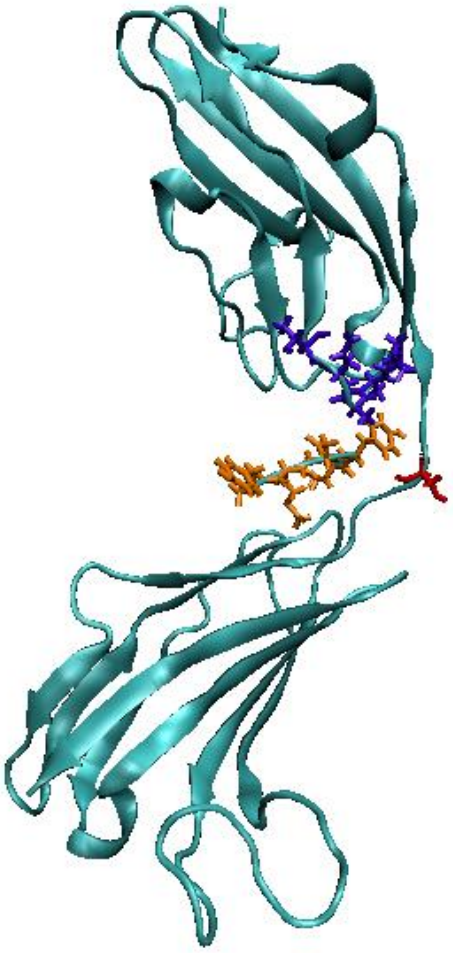


Figure 5.2.8: Conformation of peptide 5 (shown in orange) bound to R248C mutated form of the FGFR3 after 31.5 ns MD simulation. Violet and red bond representatives show residues 167-171 and residue 248 respectively.

PDB structures from MD simulations above are used for SMD simulations. Velocity constant 0.00001 Å/fs and force constant k = 7 pN / Å is used during SMD simulations for each recorded snapshot.

Simulations are done until work values are stabilized for each simulation. Force and time step values are extracted from output files for each simulation by normalizing force vector through pulling direction via Tk console at VMD.

From the obtained time step and force values work values are plotted by using $W = \int F \cdot v \cdot dt$ equation. Results obtained from 5 different SMD simulations are shown in (Figure 5.2.9):

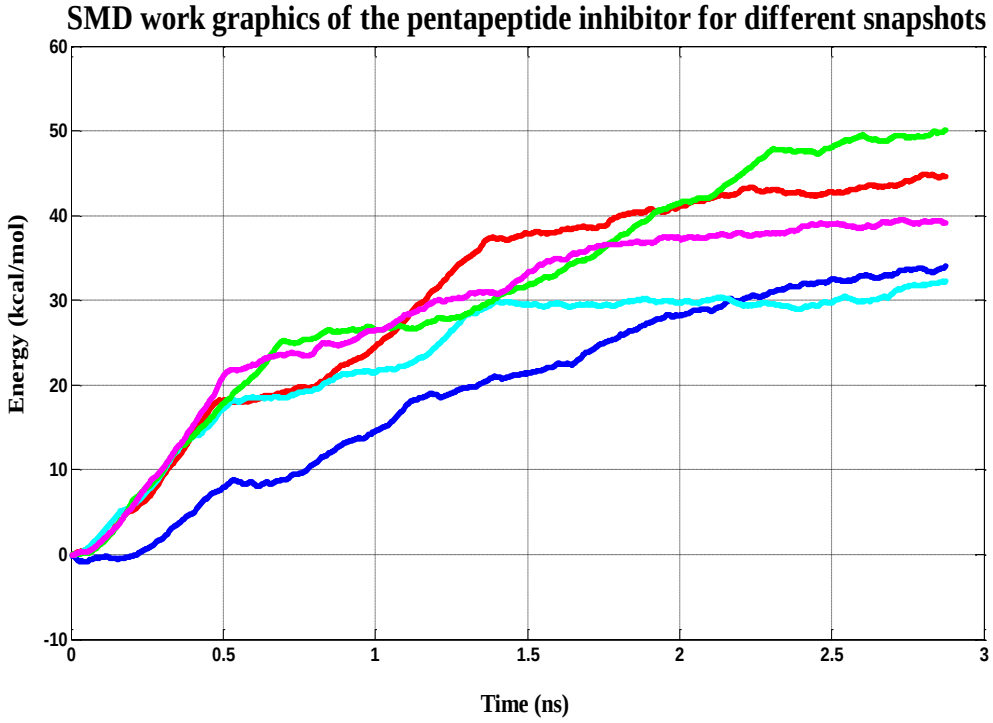


Figure 5.2.9: SMD work graphics of the peptide inhibitor from different snapshots during 2.9 ns simulation. Here blue represents the snapshot at 30.7 ns, red for the snapshot at 30.9 ns, cyan for the snapshot at 31.1 ns, green for the snapshot at 31.3 ns and magenta for the snapshot at the 31.5 ns.

As the receptor is at a closed form (Figure 5.2.8) at the beginning of the SMD simulations with the AMBER force field, peptide pulling SMD simulations lasted 0.9 ns longer than the CHARMM 27 force field SMD simulations.

The five different work values obtained for each snapshot are used for free energy calculations with Jarzynski identity, which is in the form of $\exp(-\beta\Delta G) = \langle \exp(-\beta W) \rangle$. Results shown in (Figure 5.2.10):

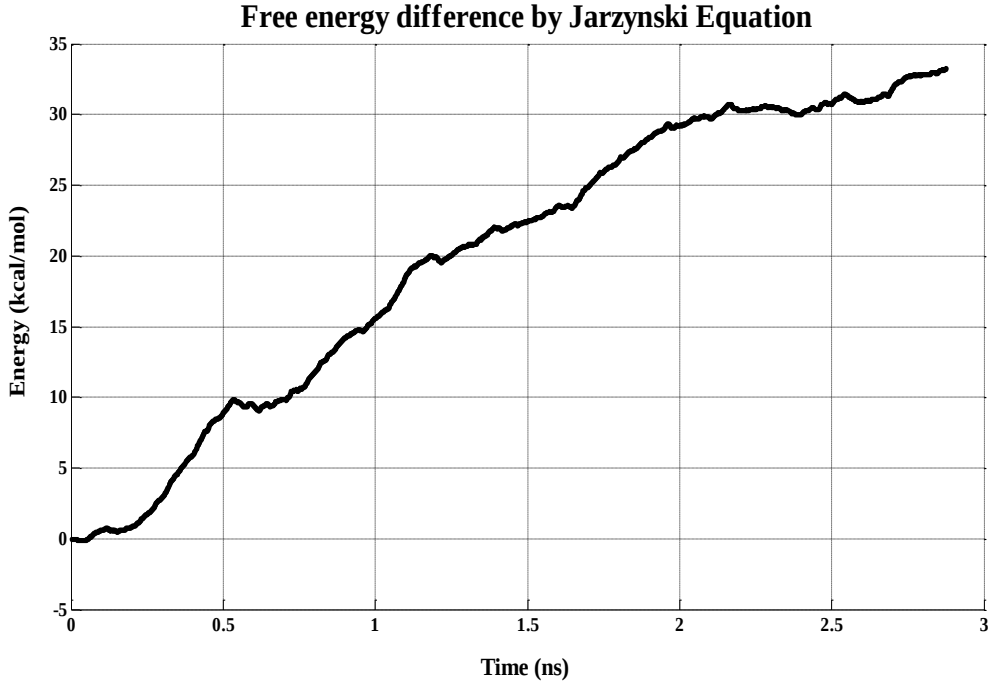


Figure 5.2.10: Free energy difference of the pentapeptide bound to the receptor, calculated from 2.9 ns SMD simulation of the snapshots taken from 30.7, 30.9, 31.1, 31.3 and 31.5 ns of the NVT + NPT simulations.

The graph in (Figure 5.2.10) shows that, after 2.9 ns of SMD simulation, average binding free energy is obtained as ΔF= -30 kcal/mol (at 2 ns), which indicates that the pentapeptide inhibitor favors to bind targeted residues.

The energy difference between docking and SMD results is expectable. MD uses all atomic interactions and the solvent effects during the simulation considering the dynamic nature of the receptor and the ligand. Since Autodock only uses a stable form of the receptor and a constant to represent water effects, such differences are unavoidable.

SMD simulations with the AMBER force field are 0.9 ns longer than the CHARMM 27 force field, because of the relatively closed form of the inhibitor. This caused 5 kcal/mol energy difference due to interatomic interactions.

Since volume change during MD simulations is relatively small, obtained binding energy value (ΔF) acceptably equals to Gibbs free energy (ΔG) of binding.

5.3 MM-GBSA calculations:

In order to obtain experimentally correlated binding energy results, different internal dielectric constants (1, 2 and 4) have been used in previous studies [44]. In continuum models, value and precise physical meaning of the dilectric constant is system dependent [44].

For the system in this study, internal dielectric constant is chosen as 2, as Moreira et al. used internal dielectric constant binding free energy for non-polar mutated residues [65].

Energy values are computed as described in the methods section. (Table 5.3.1) shows the enthalpic contribution of the different parameters. Electrostatic contribution is found effective as demonstrated from GB calculation, the vdW and electrostatic contributions from the molecular mechanic calculations.

The internal energy and non polar contributions, composed of bond, angle and dihedral energies are composing a small part of the enthalpy.

Table 5.3.1: Binding free energy calculations from MM-GBSA analysis (all values are kcal/mol)

Energy Term	Energy (kcal/mol)
E_{ELE}	-88.40
E_{VDW}	-38.97
$E_{ELE} + E_{VDW}$	-127.37
GB_{SUR}	-3.86
GB	97.51
GB_{TOT}	-33.73

ELE = electrostatic energy as calculated by the MM force field

VDW = van der Waals contribution from MM

GBSUR = nonpolar contribution to the solvation free energy calculated by an empirical model

GB = the electrostatic contribution to the solvation free energy calculated by GB

GBTOT = final estimated binding free energy calculated from the terms above

5.3.1 Entropy calculations:

The entropic effect of the solvent to the binding energy is estimated by the continuum models. The entropic contributions come from rotational, translational and vibrational degrees of freedom of solvent upon complex formation. Translational and rotational degrees of freedom loss is calculated on statistical mechanics basis and vibrational degree of freedom loss is calculated by using normal mode analysis. Entropic terms are computed for all 177 snapshots between 15 ns and 24.5 ns. Results are shown in (Table 5.3.1.1):

Table 5.3.1.1: Results of the entropic terms

Component	Entropy (kcal/mol)
$\mathbf{TS_{TRA}}$	-14.08
$\mathbf{TS_{ROT}}$	-12.29
$\mathbf{TS_{VIB}}$	-0.03
$\mathbf{TS_{TOT}}$	-26.4

TS_{TRA}: Entropy comes from translational degree of freedom

TS_{ROT} : Entropy comes from rotational degree of freedom

TS_{VIB} : Entropy comes from vibrational degree of freedom

TS_{TOT} : Total entropy

Entropy terms are mostly composed of translational and rotational components.

5.3.2 Binding Free Energy Analysis

The enthalpic and entropic terms are calculated with the following formula for all 177 snapshots between 15 ns and 24.

$$\Delta G_{water} = \Delta E_{MM} + \Delta G_{solvation} - T\Delta S \qquad \text{(Equation 5.3.2.1)}$$

The total molecular mechanics and solvation energies are calculated in the MM-GBSA results section. Entropic term is also added to (Table 5.3.2.1):

Additionally as seen in (Table 5.3.2.1), standard deviations given for each component is converted to standard error by the formula:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} \qquad \text{(Equation 5.3.2.2)}$$

Here, s represents standard deviation as 5.81 for GB_{TOT} term and 18.88 for TS_{TOT} term; n is the number of the samples in all calculations as 177.

Table 5.3.2.1: The maximum absolute error is found by adding both errors for GB_{TOT} amd TS_{TOT}.

Component	Energy (kcal/mol	Error (kcal/mol)
GB_{TOT}	-33.73	0.44
$-TS_{TOT}$	26.40	1.42
ΔGB	-7.33	1.86

The results obtained from MM-GBSA and SMD calculations are not exactly equal. Since MM-GBSA does not include protein relaxation and binding related conformational changes, this destabilizing effect may cause different binding free energy results than SMD simulations.

5.4 Steered Molecular Dynamics with CHARMM 27 Force Field for Reference Peptide

To increase the membrane permeability and to prevent degradation, capping is done as before. The RMSD of the complex is converged after being simulated during 19.47 ns NPT and 12.53 ns MD run. SMD simulations are conducted after the RMSD of the system converged to an equilibrium state. After the 32 ns simulation, RMSD values indicating a converged structure for the FGFR3 receptor is shown in (Figure 5.4.1).

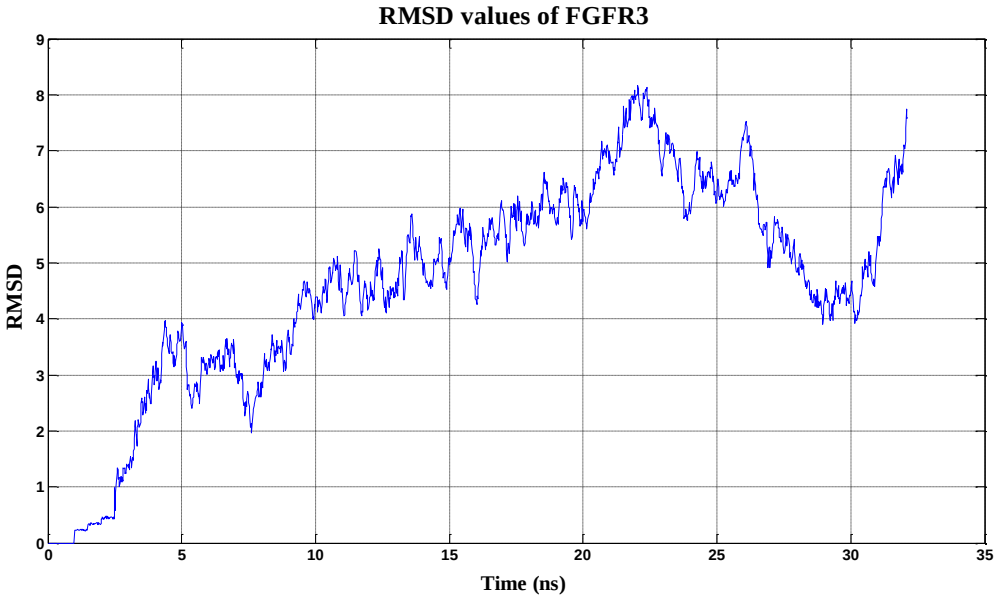


Figure 5.4.1: RMSD values of FGFR3 during MD simulations after docking.

For the SMD analysis from the last 31.4, 31.55, 31.70, 31.85 and 32 ns steps' snapshots are recorded as PDB.

PDB structures from MD simulations above are used for the SMD simulations. Velocity constant is 0.00001 Å/fs and force constant k = 7 pN / Å is used during SMD simulations for each recorded snapshot.

Simulations are done until the work values are stabilized for each simulation. Force and time step values are extracted from the output files for each simulation by normalizing the force vector through the pulling direction via Tk console at VMD.

From the obtained time step and the force values work values are plotted by using the $W = \int F \cdot V \cdot dt$ equation. The results obtained from 5 different SMD simulations are shown in (Figure 5.4.2):

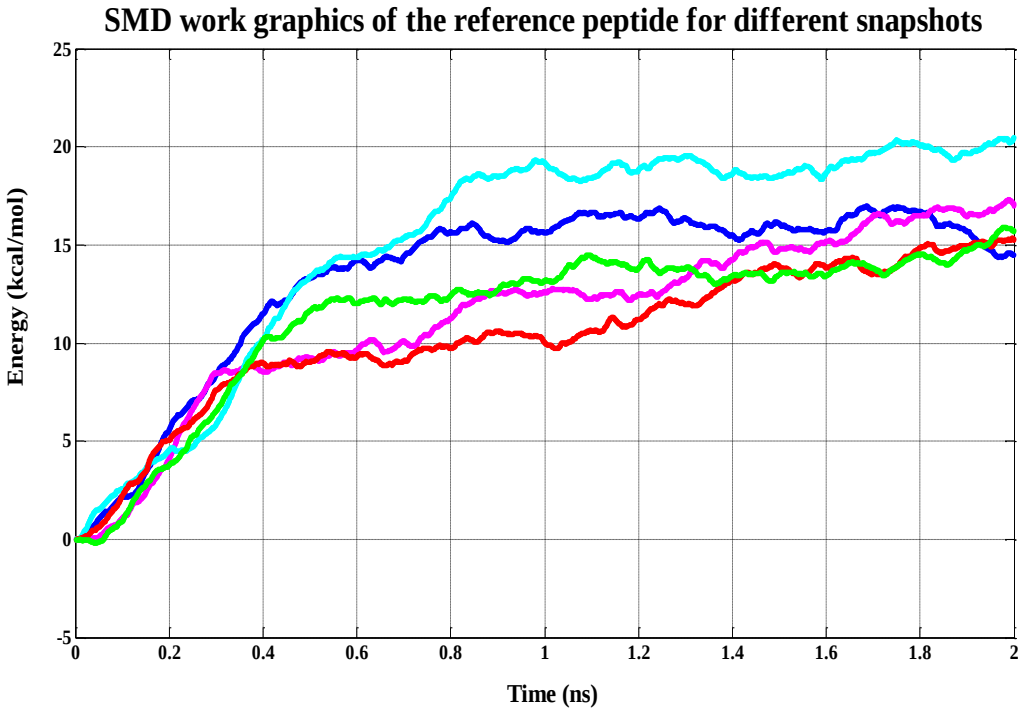


Figure 5.4.2: SMD work graphics of the reference peptide from different snapshots during 2 ns simulation. Here blue represents the snapshot at 31.4 ns, red for the snapshot at 31.55 ns, cyan for the snapshot at 31.70 ns, green for the snapshot at 31.85 ns and magenta for the snapshot at the 32 ns.

The five different work values obtained for each snapshot are used for free energy calculations with the Jarzynski identity which is in the form of $\exp(-\beta\Delta G) = \langle \exp(-\beta W) \rangle$. Results are shown in (Figure 5.4.3):

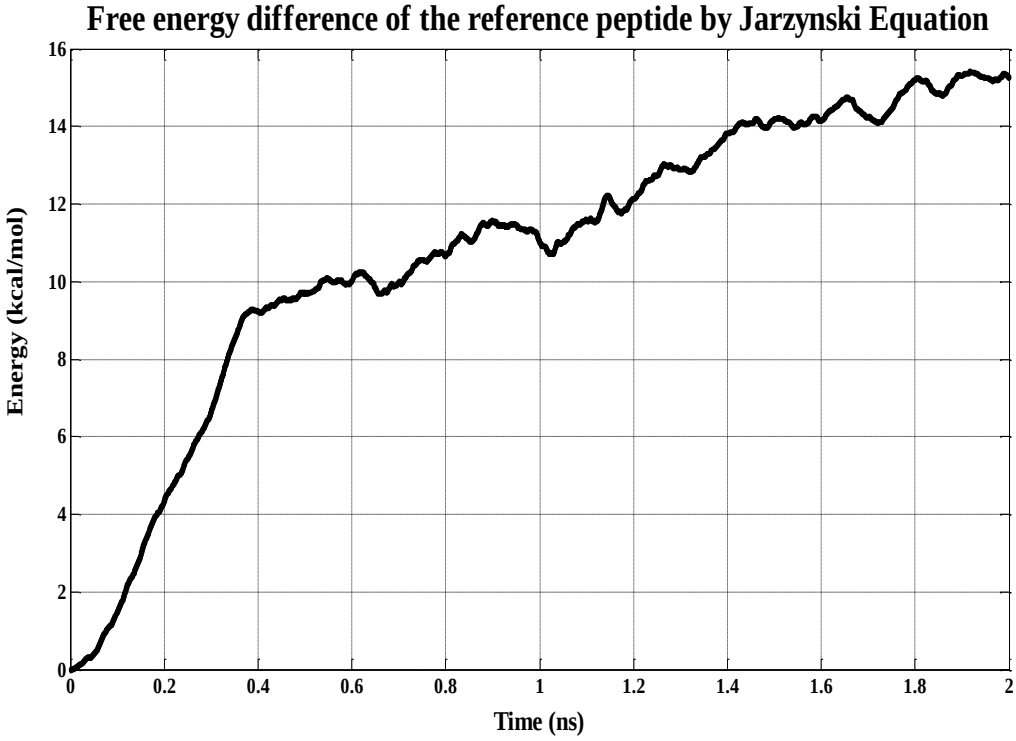


Figure 5.4.3: Free energy difference of the reference pentapeptide bound to the receptor, calculated from 2 ns SMD simulation of the snapshots taken from 31.4, 31.55, 31.70, 31.85 and 32 ns of the NVT + NPT simulations.

The graph in Figure 5.4.3 shows 2 ns of SMD simulation. The average binding free energy is obtained around ΔF= -15 kcal/mol, which indicates that the reference peptide does not favor to bind the receptor as much as the inhibitor peptide.

Comparing the reference peptide's binding free energy with the peptide inhibitor results, two times less binding free energy than the inhibitor peptide is obtained. This shows that the new peptide (IYDMY) is a better inhibitor than the reference peptide. In order to compare the results, inhibitor peptide's and reference peptide's binding free energy graphics are shown in (Figure 5.4.4):

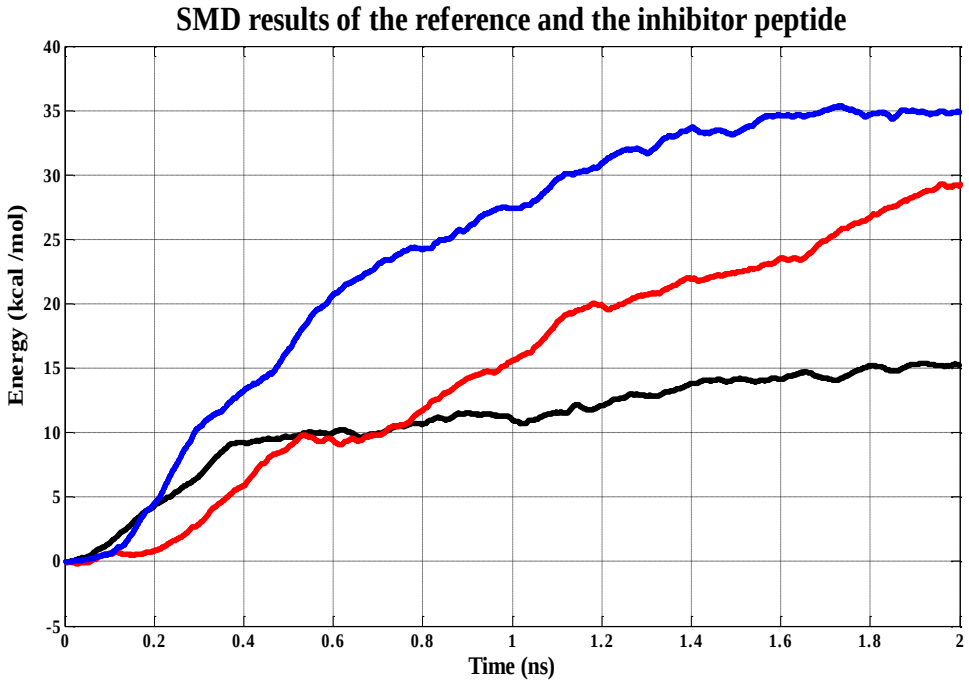


Figure 5.4.4: Free energy difference of the reference pentapeptide with CHARMM 27 force field (in black) and peptide inhibitor with CHARMM 27 force field (in blue) and AMBER ff03 force field (in red).

Inhibitor peptide is SMD simulated with two different force fields as CHARMM 27 and AMBER ff03, whose binding energies are -35 kcal/mol and -30 kcal/mol (in 2.5 ns simulation) respectively. Reference peptide is SMD simulated only with the CHARMM 27 force field and -15 kcal/mol binding free energy is obtained.

Table 5.4.1: Comparison of the binding free energy results obtained from different simulations for reference and inhibitor peptide (results are in kcal/mol)

Simulation Type	Reference Peptide (IYDLY) Binding Energy	Inhibitor Peptide (IYDMY) Binding Energy
SMD with CHARMM 27 force field	-15	-35
SMD with AMBER ff03 force field	Not done	-30
GBSA with AMBER ff03 force field	Not done	-7.33

For inhibitor peptide, binding free energy results from SMD simulations are relatively close. The difference may be because of the small structural differences beginning of the peptide pulling experiments. GBSA result indicates lower binding, probably due to inclusion of structural changes caused by the binding. Additionally, recent work of Hou et al. showed that experimental correlation of the MM-GBSA results are system dependent [66].

Previously reported articles showed that binding free energy calculations by using Jarzynski equation with SMD simulations can result higher than experimental results up to eight times [67-69]. The high values of SMD simulations may be because of the unfavorable conformational changes occurred at proteins during SMD simulations when pulling the ligand, non-optimal reaction coordinate for the ligand binding or non-equibilirium effects of the receptor as it has flexible Ig domains during the simulation. Even if there is a possible difference from the expected experimental results, forces and free energies have the systematic deviations of the same size for reference and the inhibitor peptide. Thus, it can be claimed that suggested pentapeptide inhibitor has two fold better inhibition rate than the reference peptide considering the binding free energy differences obtained from SMD simulations.

By observing the simulation trajectories carefully, two hydrogen bonds are found effective in inhibitor peptide - receptor interaction caused by aspartic acid and tyrosine residues of the inhibitor peptide. On the contrary, reference peptide does not

have any significant hydrogen bonds with the receptor. Thus, low binding free energy is obtained for the reference peptide in SMD simulations.

The pentapeptide sequence (HOOC-IYDMY-CH_3) used in this study has 28 rotatable bonds, 12 aromatic atoms. logP value is calculated via web server Marvels Space [70] for weighted option, as 1.16, which is relatively hydrophobic, thus the peptide directly has a permeation ability through the skin.

6. CONCLUSION:

Seborrheic keratoses (SK), acanthosis nigricans (AN) and epidermal nevi (EN) are dermatological diseases mostly caused by a R248C mutation of the FGFR3 receptor, which results in ligand independent signal transduction via cysteine bridge. In order to overcome this high signal transduction seen in several dermatological diseases, previously reported effective antibody's receptor interacting sequence (IYDLY) is taken as basis.

The R248C mutated form of FGFR3 is MD simulated for the relaxed conformation and then mutated pentapeptides are docked to the receptor. From Autodock results, a pentapeptide (IYDMY) is found to be a more effective inhibitor compared to the IYDLY sequence. To prevent degradation and the peptidase activity, the found sequence is capped as HOOC-IYDMY-CH_3. The binding free energy of the pentapeptide is calculated by SMD and MM-GBSA methods which have given potent binding free energy. In order to have experimentally correlated binding free energy results; equilibrium binding free energy methods, such as thermodynamic integration, free energy perturbation and weighted histogram analysis method also should be considered. Our results indicate that, binding free energy calculation with Jarzynski equation is independent from choosing rather CHARMM 27 or AMBER force field for bioorganic molecules (ff03).

Additionally the peptide sequence permanently interacts with dimerizing (167-171) and mutated (248) residues of the FGFR3 receptor during MD simulations, which will possibly result in selective inhibition of the mutated receptor only.

As the disease causing target receptor is involved in the skin, the found peptide sequence has a high possibility for topical usage, in which peptidase and instability problems are mostly avoided. Additionally, the membrane permeating peptide, magainin can be used for increasing membrane permeation [10].

7. BIBLIOGRAPHY:

[1] C. Hafner, A. Hartmann, T. Vogt, FGFR3 mutations in epidermal nevi and seborrheic keratoses: lessons from urothelium and skin, J Invest Dermatol, 127 (2007) 1572-1573.

[2] A. Logie, C. Dunois-Larde, C. Rosty, O. Levrel, M. Blanche, A. Ribeiro, J.M. Gasc, J. Jorcano, S. Werner, X. Sastre-Garau, J.P. Thiery, F. Radvanyi, Activating mutations of the tyrosine kinase receptor FGFR3 are associated with benign skin tumors in mice and humans, Hum Mol Genet, 14 (2005) 1153-1160.

[3] S. Hernandez, E. Lopez-Knowles, J. Lloreta, M. Kogevinas, A. Amoros, A. Tardon, A. Carrato, C. Serra, N. Malats, F.X. Real, Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas, Journal of Clinical Oncology, 24 (2006) 3664-3671.

[4] A. Toll, F.X. Real, Somatic oncogenic mutations, benign skin lesions and cancer progression: Where to look next?, Cell Cycle, 7 (2008) 2674-2681.

[5] O.S. Kwon, E.J. Hwang, J.H. Bae, H.E. Park, J.C. Lee, J.I. Youn, J.H. Chung, Seborrheic keratosis in the Korean males: causative role of sunlight, Photodermatol Photoimmunol Photomed, 19 (2003) 73-80.

[6] C. Hafner, A. Hartmann, J.M.M. van Oers, R. Stoehr, E.C. Zwarthoff, F. Hofstaedter, M. Landthaler, T. Vogt, FGFR3 mutations in seborrheic keratoses are already present in flat lesions and associated with age and localization, Modern Pathology, 20 (2007) 895-903.

[7] S. Venkatraman, R. Gale, Skin adhesives and skin adhesion. 1. Transdermal drug delivery systems, Biomaterials, 19 (1998) 1119-1136.

[8] Y.J. Lu, J. Yang, E. Sega, Issues related to targeted delivery of proteins and peptides, Aaps Journal, 8 (2006) E466-E478.

[9] M.R. Prausnitz, R. Langer, Transdermal drug delivery, Nat Biotechnol, 26 (2008) 1261-1268.

[10] Y.C. Kim, P.J. Ludovice, M.R. Prausnitz, Transdermal delivery enhanced by magainin pore-forming peptide, Journal of Controlled Release, 122 (2007) 375-383.

[11] I. Bernard-Pierrot, A. Brams, C. Dunois-Larde, A. Caillault, S.G. Diez de Medina, D. Cappellen, G. Graff, J.P. Thiery, D. Chopin, D. Ricol, F. Radvanyi, Oncogenic properties of the mutated forms of fibroblast growth factor receptor 3b, Carcinogenesis, 27 (2006) 740-747.

[12] C.G. L'Hote, M.A. Knowles, Cell responses to FGFR3 signalling: growth, differentiation and apoptosis, Exp Cell Res, 304 (2005) 417-431.

[13] C.J. Powers, S.W. McLeskey, A. Wellstein, Fibroblast growth factors, their receptors and signaling, Endocr Relat Cancer, 7 (2000) 165-197.

[14] E. Scotet, E. Houssaint, The choice between alternative IIIb and IIIc exons of the FGFR-3 gene is not strictly tissue-specific, Biochim Biophys Acta, 1264 (1995) 238-242.

[15] M. Mohammadi, S.K. Olsen, O.A. Ibrahimi, Structural basis for fibroblast growth factor receptor activation, Cytokine Growth Factor Rev, 16 (2005) 107-137.

[16] C. Hafner, T. Vogt, A. Hartmann, FGFR3 mutations in benign skin tumors, Cell Cycle, 5 (2006) 2723-2728.

[17] M.C. Naski, Q. Wang, J. Xu, D.M. Ornitz, Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia, Nat Genet, 13 (1996) 233-237.

[18] S. Trudel, Z.H. Li, E. Wei, M. Wiesmann, H. Chang, C. Chen, D. Reece, C. Heise, A.K. Stewart, CHIR-258, a novel, multitargeted tyrosine kinase inhibitor for the potential treatment of t(4;14) multiple myeloma, Blood, 105 (2005) 2941-2948.

[19] S. Trudel, S. Ely, Y. Farooqi, M. Affer, D.F. Robbiani, M. Chesi, P.L. Bergsagel, Inhibition of fibroblast growth factor receptor 3 induces differentiation and apoptosis in t(4;14) myeloma, Blood, 103 (2004) 3521-3528.

[20] J. Westbrook, Z. Feng, L. Chen, H. Yang, H.M. Berman, The Protein Data Bank and structural genomics, Nucleic Acids Res, 31 (2003) 489-491.

[21] T.L. Blundell, H. Jhoti, C. Abell, High-throughput crystallography for lead discovery in drug design, Nat Rev Drug Discov, 1 (2002) 45-54.

[22] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: Methods and applications, Nature Reviews Drug Discovery, 3 (2004) 935-949.

[23] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A Geometric Approach to Macromolecule-Ligand Interactions, J Mol Biol, 161 (1982) 269-288.

[24] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: Current status and future challenges, Proteins-Structure Function and Bioinformatics, 65 (2006) 15-26.

[25] R.M. Muegge I, Reviews in computational chemistry, 17 (2001) 1-60.

[26] J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy, J Med Chem, 47 (2004) 45-55.

[27] T.N. Hart, R.J. Read, A multiple-start Monte Carlo docking method, Proteins, 13 (1992) 206-222.

[28] J. Desmet, M. Demaeyer, B. Hazes, I. Lasters, The Dead-End Elimination Theorem and Its Use in Protein Side-Chain Positioning, Nature, 356 (1992) 539-542.

[29] R.M.A. Knegtel, I.D. Kuntz, C.M. Oshiro, Molecular docking to ensembles of protein structures, J Mol Biol, 266 (1997) 424-440.

[30] F. Osterberg, G.M. Morris, M.F. Sanner, A.J. Olson, D.S. Goodsell, Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock, Proteins, 46 (2002) 34-40.

[31] I. Halperin, B. Ma, H. Wolfson, R. Nussinov, Principles of docking: An overview of search algorithms and a guide to scoring functions, Proteins, 47 (2002) 409-443.

[32] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, Comparative evaluation of eight docking tools for docking and virtual screening accuracy, Proteins, 57 (2004) 225-242.

[33] R.D. Taylor, P.J. Jewsbury, J.W. Essex, A review of protein-small molecule docking methods, J Comput Aided Mol Des, 16 (2002) 151-166.

[34] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J Comput Chem, 19 (1998) 1639-1662.

[35] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible ligand docking, Abstr Pap Am Chem S, 214 (1997) 154-COMP.

[36] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J Mol Biol, 261 (1996) 470-489.

[37] T.J.A. Ewing, I.D. Kuntz, Critical evaluation of search algorithms for automated molecular docking and database screening, J Comput Chem, 18 (1997) 1175-1189.

[38] R. Abagyan, M. Totrov, D. Kuznetsov, Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation, J Comput Chem, 15 (1994) 488-506.

[39] A. Fiser, Protein structure modeling in the proteomics era, Expert Rev Proteomics, 1 (2004) 97-110.

[40] A.E. Todd, C.A. Orengo, J.M. Thornton, Evolution of function in protein superfamilies, from a structural perspective, J Mol Biol, 307 (2001) 1113-1143.

[41] N. Fernandez-Fuentes, J. Zhai, A. Fiser, ArchPRED: a template based loop structure prediction server, Nucleic Acids Research, 34 (2006) W173-W176.

[42] B. Isralewitz, J. Baudry, J. Gullingsrud, D. Kosztin, K. Schulten, Steered molecular dynamics investigations of protein function, Journal of Molecular Graphics & Modelling, 19 (2001) 13-25.

[43] J. Kongsted, U. Ryde, An improved method to predict the entropy term with the MM/PBSA approach, Journal of Computer-Aided Molecular Design, 23 (2009) 63-71.

[44] G. Rastelli, A. Del Rio, G. Degliesposti, M. Sgobba, Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA, J Comput Chem, 31 (2010) 797-810.

[45] Setting Constraints and Periodic Boundaries, in.

[46] C. Jarzynski, Nonequilibrium equality for free energy differences, Phys Rev Lett, 78 (1997) 2690-2693.

[47] G. Hummer, A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, P Natl Acad Sci USA, 98 (2001) 3658-3661.

[48] I. Massova, P.A. Kollman, Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, Perspect Drug Discov, 18 (2000) 113-135.

[49] D. Sitkoff, K.A. Sharp, B. Honig, Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models, Journal of Physical Chemistry, 98 (1994) 1978-1988.

[50] J. Qing, X.N. Du, Y.M. Chen, P. Chan, H. Li, P. Wu, S. Marsters, S. Stawicki, J. Tien, K. Totpal, S. Ross, S. Stinson, D. Dornan, D. French, Q.R. Wang, J.P. Stephan, Y. Wu, C. Wiesmann, A. Ashkenazi, Antibody-based targeting of FGFR3 in bladder carcinoma and t(4;14)-positive multiple myeloma in mice, Journal of Clinical Investigation, 119 (2009) 1216-1229.

[51] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: Greater scalability for parallel molecular dynamics, Journal of Computational Physics, 151 (1999) 283-312.

[52] J.B. M.Schlenkrich, A. D. MacKerell, M. Karplus, in Biological Membranes: A Molecular Perspective from Computation and Experiment, Merz, K. M. & Roux, B. Birkhauser, Boston, ed., 1996.

[53] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, Journal of Physical Chemistry B, 102 (1998) 3586-3616.

[54] W.L. Jorgensen, Quantum and Statistical Mechanical Studies of Liquids .10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers - Application to Liquid Water, Journal of the American Chemical Society, 103 (1981) 335-340.

[55] P. Gonnet, P-SHAKE: A quadratically convergent SHAKE in O(n(2)), Journal of Computational Physics, 220 (2007) 740-750.

[56] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A Smooth Particle Mesh Ewald Method, Journal of Chemical Physics, 103 (1995) 8577-8593.

[57] S.E. Feller, Y.H. Zhang, R.W. Pastor, B.R. Brooks, Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method, Journal of Chemical Physics, 103 (1995) 4613-4621.

[58] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant-Pressure Molecular-Dynamics Algorithms, Journal of Chemical Physics, 101 (1994) 4177-4189.

[59] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G.M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J.M. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J Comput Chem, 24 (2003) 1999-2012.

[60] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J Comput Chem, 26 (2005) 1668-1688.

[61] T.A.D. D. A. Case, T.E. Cheatham III, C.L. Simmerling, J. Wang and R.E. Duke et al., AMBER 10, in, University of California, San Francisco, CA, 2008.

[62] M. Froimowitz, Hyperchem(Tm) - a Software Package for Computational Chemistry and Molecular Modeling, Biotechniques, 14 (1993) 1010-1013.

[63] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins-Structure Function and Bioinformatics, 55 (2004) 383-394.

[64] M.L. Connolly, Analytical Molecular-Surface Calculation, Journal of Applied Crystallography, 16 (1983) 548-558.

[65] I.S. Moreira, P.A. Fernandes, M.J. Ramos, Hot spot computational identification: Application to the complex formed between the hen egg white lysozyme (HEL) and the antibody HyHEL-10, Int J Quantum Chem, 107 (2007) 299-310.

[66] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations, J Chem Inf Model, (2010).

[67] M.A. Cuendet, O. Michielin, Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex, Biophys J, 95 (2008) 3575-3590.

[68] F. Grater, B.L. de Groot, H. Jiang, H. Grubmuller, Ligand-release pathways in the pheromone-binding protein of Bombyx mori, Structure, 14 (2006) 1567-1576.

[69] W. Li, J. Shen, G. Liu, Y. Tang, T. Hoshino, Exploring coumarin egress channels in human cytochrome P450 2A6 by random acceleration and steered molecular dynamics simulations, Proteins, 79 (2011) 271-281.

[70] Marvin, Calculator Plugin and Chemical Terms Demo, i

8-APPENDIX 1:

The scoring function used in Autodock simulations defined by a semiempirical free energy force field, which consists of the following evaluations:

$$\Delta G = (V_{bound}^{L\text{-}L} - V_{unbound}^{L\text{-}L}) + (V_{bound}^{R\text{-}R} - V_{unbound}^{R\text{-}R}) + (V_{bound}^{R\text{-}L} - V_{unbound}^{R\text{-}L}) + \Delta S$$

(Equation 8.1.1)

Here L stands for the ligand and R stands for the receptor. ΔG represents the binding free energy, ΔS represents change in the entropy V represents the energy term which includes a van der Waals term, a hydrogen bond term, an electrostatic term and a desolvation term, normalized with constants.

Energy term used at (Equation 8.1.1) can be expressed as follows:

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$
$$+ W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} (S_i) V_j + (S_j) V_i) e^{(-r_{ij}^2 / 2\sigma^2)}$$

(Equation 8.1.2)

ΔS , the entropic term in the (Equation 8.1.1), defines the loss of entropy coming from the torsional degrees of freedom resulted by complex formation. The number of rotable bonds is defined by N_{tors}, which represents the entropic loss contribution as follows in (Equation 8.1.3) [34]

$$\Delta S_{conf} = W_{conf} N_{tors} \qquad \text{(Equation 8.1.3)}$$

9-VITA

Mehmet Ali Öztürk was born in Ankara, Turkey in 1985. He graduated from Ankara Science High School in 2003. He received his B.Sc. degrees in Molecular Biology & Genetics and Chemistry Departments from Istanbul Techical University (ITU) in 2007. At the same year he began his M.Sc. education at ITU then he moved to Computational Sciences and Engineering M.Sc. program at Koç University with lateral transfer in 2009. From 2009 to 2011, he worked as research and teaching assistant at the same university.