Development of inhibitors of TCF/β-catenin complex and their validation using Molecular Modeling techniques

Thesis submitted in partial fulfilment of the requirements of the

BS-MS Dual Degree Program at IISER, Pune

Submitted by

M B Harsha

20141057

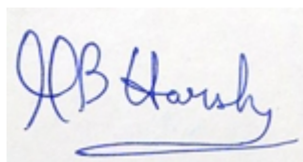
Under the guidance of

Dr. Arnab Mukherjee

Department of Chemistry, IISER Pune.

CERTIFICATE

This is to certify that this dissertation entitled **"Development of inhibitors of TCF/β-catenin complex and their validation using Molecular Modeling techniques"** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents research carried out by **M B Harsha** at the Indian Institute of Science Education and Research, Pune under the supervision of **Dr. Arnab Mukherjee,** Associate Professor of Chemistry Department, IISER Pune, during the academic year 2018-2019.

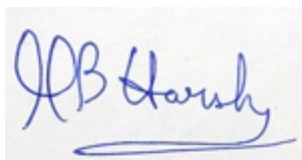


M B Harsha
BS-MS Student
IISER Pune

Dr. Arnab Mukherjee
Associate Professor
Dept. Of Chemistry, IISER Pune

DECLARATION

I hereby declare that the matter embodied in the report entitled "**"Development of inhibitors of TCF/β-catenin complex and their validation using Molecular Modeling techniques"** are the results of the investigations carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of **Dr. Arnab Mukherjee** and the same has not been submitted elsewhere for any other degree.



M B Harsha
BS-MS Student
IISER Pune

Dr. Arnab Mukherjee
Associate Professor
Dept. Of Chemistry, IISER Pune

I dedicate this thesis to my mother S R Lalitha

Acknowledgement

I would like to begin by thanking my guide **Dr. Arnab Mukherjee** for being an extremely patient and supportive supervisor. His constant encouragement and support throughout the project helped me carry out this work very smoothly. I learnt a lot of things from the discussions that we had, especially given the fact that I was new to this field.

Next, I would like thank my TAC member and my former project supervisor **Dr. Harinath Chakrapani** for his constant support throughout the project as well as my life at IISER Pune. I will always be thankful for all the encouragement that I received whenever I did well and all the pieces of advice that I received whenever I faltered. I wil always be grateful for everything that I have learnt through our interactions.

Special thanks to my lab members **Rituparno, Atul, Reman, Debasis, Hridya** and **Abhijit** for helping me out throughout the project with their teachings, suggestions and discussions. I am extremely grateful to all of them for patiently answering all my queries and teaching me whenever I needed their help

I would like to thank my mother, **S. R. Lalitha,** without whose efforts and struggle, none of these would be possible. She has always been very encouraging and supportive in whatever endeavours I want to pursue.

I would like to thank **IISER Pune** for giving me such a wonderful opportunity to carry out an extensive and involved research project as part of the BS-MS program. I would like to thank **DST-INSPIRE** for providing the financial support to carry out my project.

I would like to thank **all my friends** who have been there for me through my ups and downs. All the coffee table discussions we had about science and life taught me a lot of things.

Thank you all!

Contents

List of abbreviations..(i)

List of figures...(ii)

Abstract..(iii)

1. Introduction
   - 1.1. Canonical Wnt signalling pathway..1
   - 1.2. β-catenin and protein-protein interactions..................................2
   - 1.3. List of some commercially available inhibitors and their discovery..4
   - 1.4. Outline of the methodology..7
2. Methodology
   - 2.1. Molecular Docking...9
   - 2.2. Molecular Dynamics Simulations..10
   - 2.3. Well-tempered Metadynamics...11
3. Results and Discussion
   - 3.1. Docking..13
   - 3.2. Well-tempered Metadynamics
     - 3.2.1. BC1 and β-catenin..14
     - 3.2.2. BC2 and β-catenin..16
     - 3.2.3. CHEMBL1334062 and β-catenin..18
     - 3.2.4. BC4 and β-catenin..20
   - 3.3. Comparitive study...22
4. Conclusion..25
5. Reference...25

List of abbreviations

1) FES – Free-Energy Surface

2) RMS – Root-Mean Squared

3) HTS – High-Throughput Screening

4) CV – Collective Variable

5) MD – Molecular Dynamics

List of Figures

Figure	Title	Page No.
1	Canonical Wnt signalling pathway	1
2	β-catenin	3
3	β-catenin/TCF3 complex	3
4	PNU-74654	4
5	PNU-74654 binding mode	5
6	2,4 – Diamino-quinazoline derivative	5
7	Quercetin	6
8	Inhibitors of β-catenin discovered through HTS	6
9	Chosen molecules from known and new pool	7
10	Docking region in β-catenin	10
11	Docking energies comparison	13
12	MINDIST CV for BC1	14
13	Snapshots from trajectory for BC1	14&15
14	FES for BC1	15
15	MINDIST CV for BC2	16
16	Snapshots from trajectory for BC2	16&17
17	FES for BC2	17
18	MINDIST CV for CHEMBL1334062	18
19	Snapdhots from trajectory for CHEMBL1334062	18&19
20	FES for CHEMBL1334062	19
21	MINDIST CV for BC4	20
22	Snapshots from trajectory for BC4	20
23	FES for BC4	21
24	FES of BC1, BC2, CHEMBL1334062 and BC4 in same plot	22
25	Protein-Ligand interaction energies (Coulombic & LJ)	23
26	Water-Cl-Ligand interaction energies (Coulombic & LJ)	23
27	H-bond number for BC1 and BC2	24

Abstract

The Wnt signalling pathway plays a vital role in cell proliferation and hence plays a crucial role in the embryonic development stage of the cells. This pathway is usually inactive in well grown cells. But, it is seen that in cancerous cell lines, especially colorectal cancer cells, the Wnt signalling pathway is active, due to which there is an increased concentration of the protein β-catenin in the cell. B-catenin then enters the nucleus and replaces the groucho protein, to interact with T-Cell Factor. This complex acts as a transcriptional activator and hence leads to cell proliferation. Given the importance of discovering strategies to inhibit this pathway, one of the important ways of achieving this is to inhibit the β-catenin/TCF complex. There are a number of commercially available molecules that are known to inhibit this complex but there are not enough computational studies that demonstrate the binding affinities of these molecules to the different "Hotspots" present in this complex. The goal of this project was to calculate the free-energy of interactions for different molecules, some known and some newly designed, and compare them to get a quantitative idea about the binding affinities. The calculations showed that the molecule BC2 has a great affinity towards "Hotspot 2" when compared to BC1 and BC4, where BC1, BC2 and BC4 are moleules that were newly designed. The free-energy of BC2 molecule was found to be very similar to CHEMBL1334062, a commercially known inhibtor.

1.Introduction

1.1 Canonical Wnt signalling pathway

The canonical Wnt signalling pathway is one of the signal transduction pathways, which plays a vital role in a variety of processes, like, cell proliferation, cell differentiation, etc [1]. One of the primary role of the pathway is during the embryonic development stage and in stem cells, where it is really important for cell proliferation to occur. Naturally, the Wnt signalling pathway in developed tissues are expected to be mostly inactive, as the genes that are activated by the pathway correspond to enhanced cell proliferation [2]. To understand the relevance of studying the pathway, it becomes necessary to look into the mechanism of the pathway.

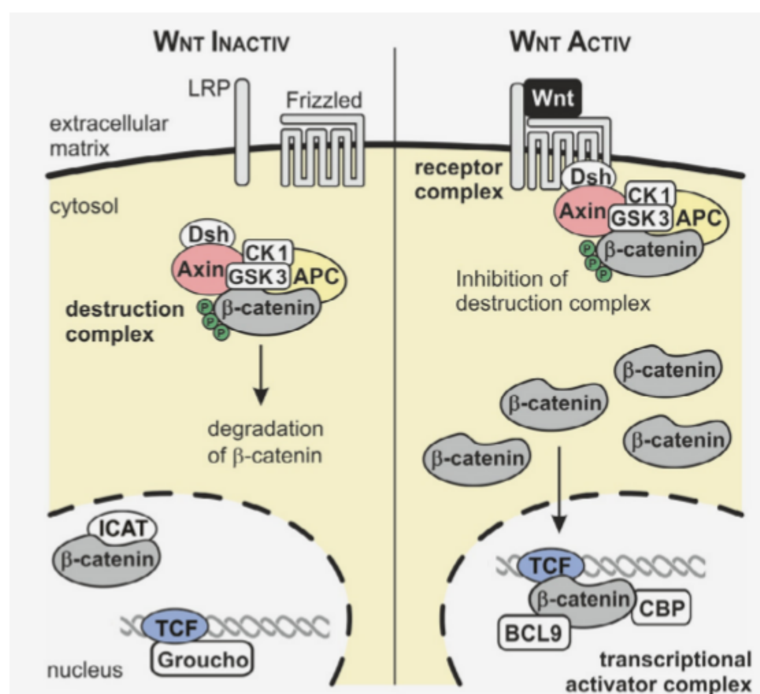


Figure 1: A representation of the active and inactive states of the canonical Wnt signalling pathway

Figure 1 gives a representation of the active and inactive states of the Wnt signalling pathway [3]. One of the most important components of the pathway is the protein β-catenin. As the concentration of the protein increases in the cytoplasm, it gets into the nucleus. Now, it replaces the groucho protein which was bound to the T-Cell Factors (TCF), and forms an activation complex that binds to the activator sequence. This over-expresses the parts of gene that is responsible for cell proliferation [4].

In normal cells the levels of β-catenin is controlled by this elaborate degradation procedure involving the formation of a destruction complex, which occurs in the Wnt inactive state. The β-catenin molecule is first phosphorylated by the destruction complex, followed by ubiquitinylation and at last proteosomal degradation. Whereas, in the Wnt active state, that is, when Wnt binds to the Frizzled, it phosphorylates the cystolic domain and thus facilitates the binding of the destruction complex to the domain through the components, Axin and Dishevelled [5]. Thus, the destruction complex gets inhibited which in turn leads to an increased concentration of the β-catenin in the cell. This in turn leads to the cell proliferation, as discussed before.

It is found that the Wnt signalling pathway is found to be in the active state in certain kinds of cancerous tissues, especially colorectal cancer [6]. Hence, this pathway becomes one of the targets for which one can try to discover inhibtors. As it is an elaborate pathway involving multiple steps, one can think of multiple possibilities of inhibiting the pathway. For example, there have been methodologies developed that activate the destruction complex so that the β-catenin levels are decreased [7]. But this method would not work in the case where the destruction complex is completely inactivated. Another, such strategy is to curb the biosynthesis of the Wnt molecule itself [8]. One of the way that is very common nowadays is to inhibit the β-catenin from binding with TCF, hence preventing the formation of the transcription activator complex.

This thesis will be focusing on using the strategy of inhibiting the β-catenin/TCF interaction in an attempt find inhibitors for the Wnt signalling pathway.

1.2 β-catenin and protein-protein interactions

The structural feature of β-catenin that is relevant to study its interaction with other proteins and small molecules is the repeating armadillo motifs. Armadillo motifs are typically 40 amino acid long and is found in a number of proteins like β-catenin, α-importin, plakoglobin, etc. These armadillo repeats are found between residue numbers 140-700 [9]. For example, the human β-catenin has a total of 781 residues and the armadillo repeats are found between 149-691. As seen in **Figure 2,** the

protein comprises of α-helices, with armadillo repeat comprising of three α-helices each.

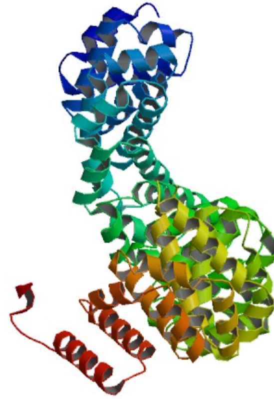


Figure 2: β-catenin protein (PDB ID: 2Z6H)

To find inhibitors for β-catenin it becomes important to study its interactions with other proteins like TCF. Typically, protein-protein interactions are spread over a large surface area. In the case of β-catenin/TCF3 complex, the interacting surface area is around 4800 $Å^2$. Though the interacting surface area is large, the interaction is not uniform in all the regions. There are some regions which contribute considerably to the overall interaction between the two protein molecules. These regions are called the “hotspots”. A total of six such “hotspots” were found for the β-catenin/TCF3 complex by Trosset et. al., upon studying the system with the PASS program [10].

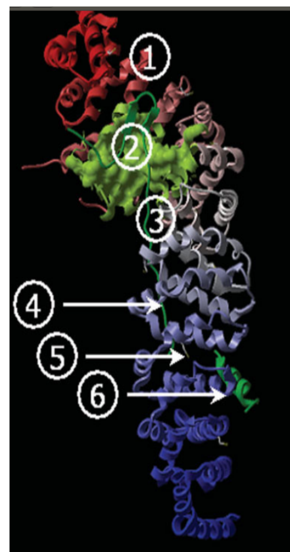


Figure 3: β-catenin/TCF3 complex with the 6 detected hotspots marked. Green ribbon is the TCF3 protein and the surface representation depict “Hotspot 2”, which is of our interest.

Figure 3 shows the β-catenin/TCF3 complex with all the hotspots marked. These, hotspots were studied further to determine the contribution of each of these hotspots to the overall interaction strength. Trosset et. al., determined the solvent-accessible surface using the FLO_QXP program. This revealed multiple clefts that are present in "Hotspot 2" proposing that it is the strongest hotspot. It was further ascertained by the results obtained by docking a molecule, PNU-74654, shown in **Figure 4,** according to which the molecule bound 1000 times better "Hotspot 2" than "Hotspot 3". So, it becomes very important to study the interaction of small molecules with "Hotspot 2".

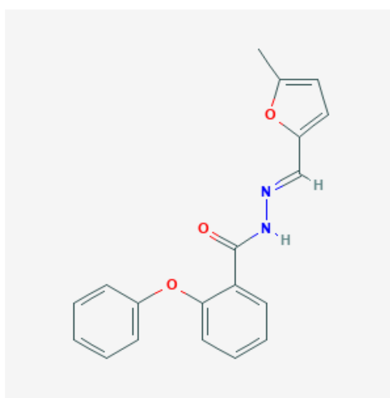


Figure 4: PNU-74654 (Image source: PubChem)

Let us look at some of the known and commercially available inhibitors for the β-catenin/TCF complex.

1.3 List of some commercially available inhibitors and their discovery

The comound shown in **Figure 4,** was discovered through a combination of virtual-screening and biophysical screening techniques like NMR WaterLOGSY and ITC measurements. Proposed mode of interaction of the drug with the hotspot is shown in **Figure 5.**

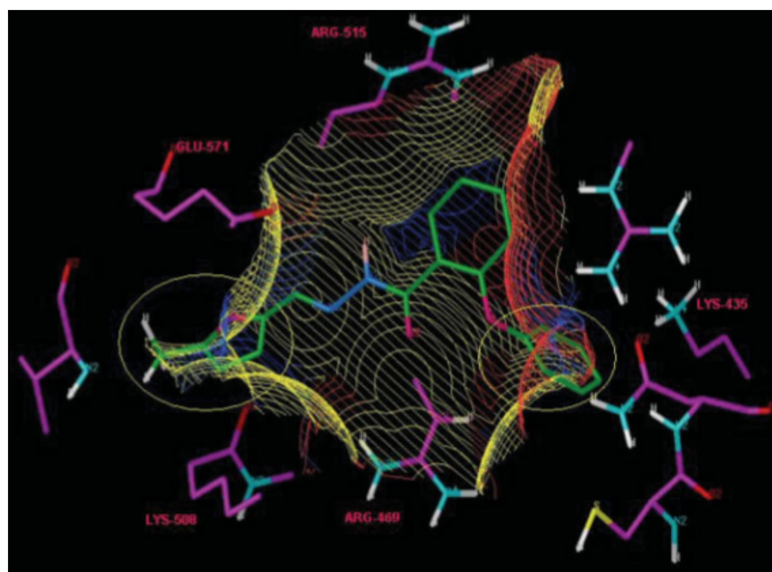


Figure 5: Proposed binding of PNU-74654 with hotspot “2”

Next set of compunds are 2,4-Diamino-quinazoline derivatives[11]. **Figure 6** shows both the lead compound (**1)** and the most efficient derivative (**16**).

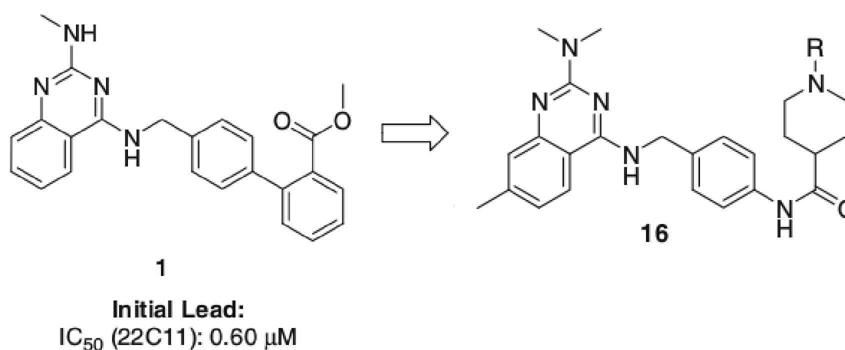


Figure 6: Initial lead compund **1.** When R group was 4-fluorobenzyl IC_{50} was 0.22um.

The lead compound was obtained by screening a huge compound library. After some set of modifications, compound **16** was obtained. The mechanism of action of this compound is not known as its activity was found through biological cell-based reporter assay.

The next compound is Quercetin, shown in **Figure 7,** which belongs to a class of compounds called Flavanoids. This compound was also discovered to be active by conducting studies on SW480 and HEK293 cell lines [12].

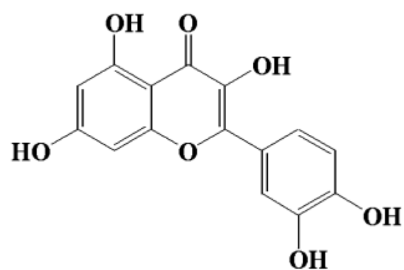


Figure 7: Quercetin

Figure 8, shows a set of compunds which was discovered by screening over 7000 compounds using a GST based screening technique [13].

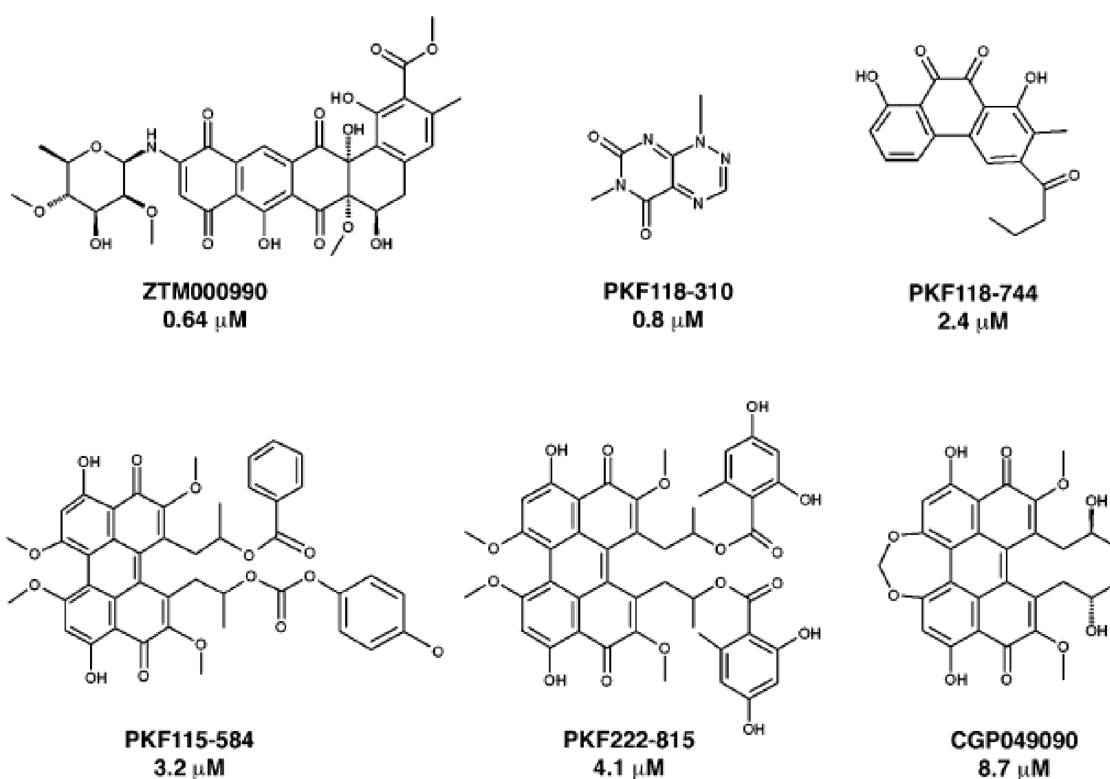
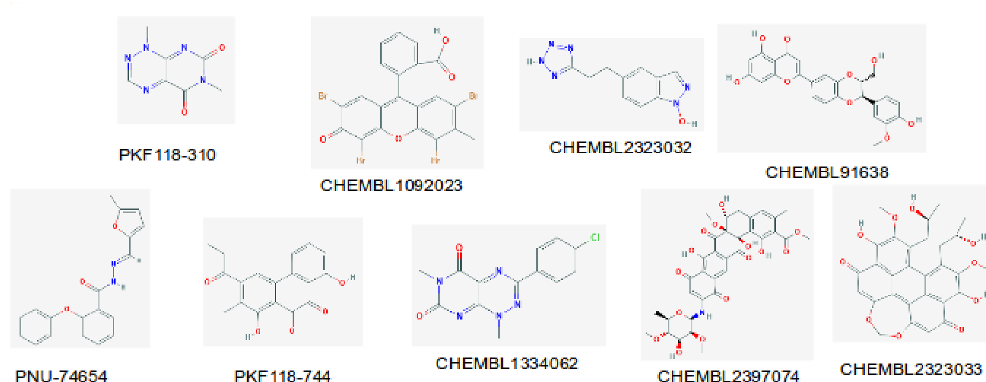


Figure 8: Structures of compounds that were discovered by a GST based screening technique.

1.4 Outline of the methodology

This thesis will focus on evaluating the free energy of interaction of selected known and newly generated molecules, to “Hotspot 2” and thus compare the binding affinities.

Firstly, the molecules shall be chosen from the known pool of commercially available inhibitors. Then, five molecules with the least binding energies from the pool of molecules generated from our group, shall also be chosen. **Figure 9,** shows all these molecules.



a)

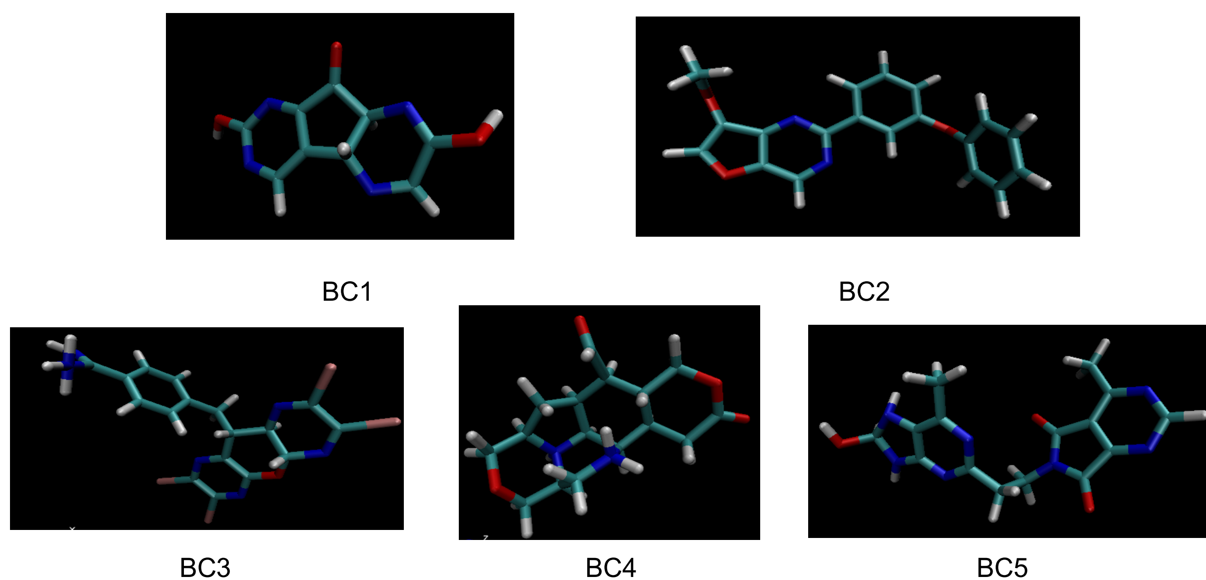


Figure 9: The chosen molecules from the pool of known commercially available inhibitors and the new molecules developed.

These molecules shall be docked at “Hotspot 2” of the β-catenin protein. Once the molecules are docked, the system shall be equilibriated using regualr Molecular Dynamics simulations. After this, well-tempered metadynamics shall be used to evaluate the free-energy change involved in the unbinding process. To monitor this process, the Collective Variable (CV), that maybe used is MINDIST. This CV, in principle, calculates all the pair-wise distances between two atoms from the protein and the ligand group and takes the minimum of those distances. Among other possible CVs like distance between COM, distance along with a distance vector, etc. MINDIST is rather less complex to define and more accurate. For example, defining a distance and a vector may do the job, but since it has two degrees of freedom, it is more prone to error propagation.

A simple run, monitoring the above mentioned CV would give an idea about the σ of the wells that can be deposited during the metadynamics run. Then, the system shall be subjected to a well-tempered metadynamics run, at the end of which we shall be able to procure the free-energy surface. Once the surface is obtained, other properties like, interaction energies between protein-lagand, protein-water and ligand-water, the RMS Fluctuation of the protein, the number of hydrogen bonds, etc can be calculated to get more insight into the unbinding event.

In order to get some indirect information about the entropy change, one can use the Q6 and Q4 CVs found in PLUMED driver.

2. Methodology

2.1 Molecular Docking

Molecular docking is the procedure which is used to predict the binding modes of a molecule to a specified binding pocket in the macromolecule, like protein, DNA, Lipids, etc, through a computational algorithm. Given the large number of conformations that the computer has to explore before arriving at the minimum energy conformation, it will be computationally very expensive to sample all these using Molecular Dynamics. So, it becomes easier to use docking softwares if the goal is to just predict the binding modes.

Typically, these softwares use some scoring function that helps in ranking the different conformations obtained based on the affinity. These scoring functions are can be broadly categorized as Force-Fields based scoring function, Empirical scoring function and Knowledge-based scoring function. There is also this new category of scoring functions known as Consensus scoring functions. These scoring functions employ more than one strategies from the above mentioned list. The software that was used to carry out the docking studies for this thesis is Vina AutoDock [14].

The scoring function implements experimental affinity measurements and also the preferred conformations from empirical observations. As a result, the predictions are not only computationally cheap, but also accurate to a good extent.

The shortcomings of docking is that it is not quantitatively accurate. As the scoring functions do not include all the force fields, the docking energy obtained will not correspond to the actual free energy of binding of the ligand to the receptor. So, the docking studies can be applied to just to predict the binding modes. This has to be further studied using other advanced Molecular Modelling techniques to get a better quantitative picture.

All the dockings were carried out holding the receptor region regid and making all the possible rotatable bonds in the ligand flexible. The region that was chosen to carry out the docking is shown in **Figure 10.**

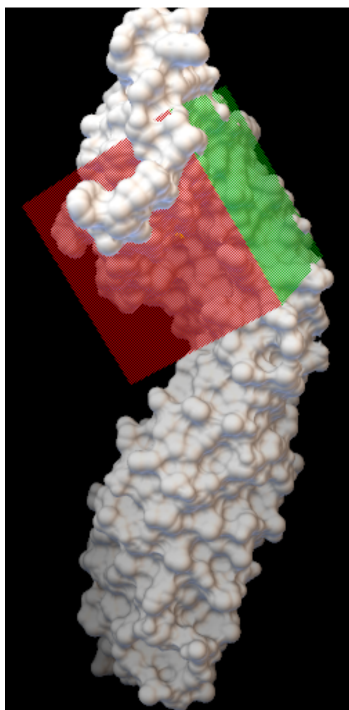


Figure 10: β-catenin protein with the docking region marked

2.2 Molecular Dynamics simulations:

Regular Molecular Dynamics simulations were carried out to equilibriate the docked structures before carrying out the well-tempered Metadynamics run. The software GROMACS 4.5.5 [15-19] patched with PLUMED 1.3.0 [20] was used to run these simulations.

Since, over 100 residues were missing in the β-catenin PDB file that was used, these residues were predicted using the software MODELLER [21] and the obtained complete PDB file was used for further studies. It was made sure that the new PDB wasn't structurally different from the obtained PDB, around the hotspot of interest.

The force field that was used for the protein was AMBER99SB. The force field paramters for the molecules were generated using the software REDIII [22], which in

turn used GAUSSIAN 03 [23] for charge calculation. The generated tripos mol2 file was supplied to amber tools to generate the topology file for the ligand [24].

The system consisting of β-catenin and ligand molecule was taken in a cubical simulation box of side 18 nm. After addition of TIP3P water molecules and addition of ions, the system's energy was minimized by using the steepest descents method. Then the sytem was heated to 300K in steps, using a berendsen thermostat, followed by the release of restraint in 5 steps. Then, the sytem was equilibriated under NPT ensemble condition using the Parinello-Rahman barostat and berendsen thermostat for a duration of 5 ns. This was followed by NVT equilibriation with a Nose-Hoover thermostat for a duration of 5 ns. Then, NVT ansemble production run was carried out with a Nose-Hoover thermostat, for a duration of 5 ns. The final configuration that was obtained was used to carry out further metadynamics calculations.

2.3 Well-Tempered Metadynamics:

Well-tempered metadynamics is one of the adaptive nonquilibrium methods which employs external biasing potential so that the system is able to sample even the higher barrier regions. This aids in the estimation of the Free Energy Surface (FES) of a given process and the Collective Variable (CV). Unlike in a Metadynamics simulation, the deposited gaussian potentials' height decrease with time according to the **Equation 1**. In case of a Metadynamics simulation there is an issue of overfilling of the potential wells. This is avoided to an extent in welltempered Metadynamics.

$$W(k\tau) = W_0 \exp\left(-\frac{V(\vec{s}(q(k\tau)), k\tau)}{k_B \Delta T}\right) \qquad \text{...Equation 1}$$

Since the height of the well deposited decreases with time, some of the irreversible high energy states won't be sampled in the case of well-tempered Metadynamics simulation. In **Equation 1**, the situation in which $\Delta T = 0$ would correspond to a regular Molecular dynamics simulation and $\Delta T \rightarrow \infty$ would correspond to a regular Metadynamics. A right choice of the the bias factor γ (shown in **Equation 2**), would ensure good sampling of all the states and thus improves the accuracy of the free-energy of calculation [25].

$$\gamma = \frac{T + \Delta T}{T}.$$ ...**Equation 2**

All the well-tempered Metadynamics simulations for this thesis were performed using PLUMED 1.3.0. The simulations were run for 20 ns.

The CV that was used to carry out the well-tempered Metadynamics simulation was the MINDIST CV. As the name suggests, this CV determines all the distances between the atoms from the protein and the ligand pairwise, and takes the minimum value of all of them. The distance CV would need to be supplied with the vector as well to study an unbinding event. Since, the definition of a vector for a protein is not straightforward, the MINDIST was decided to be used as the CV.

All the metadynamics simulations were carried out for a duration of 20 ns, where the hills height was 0.3, the simulation temperature T for well-tempered metadynamics as 300 and the bias factor γ as 10. One-third of the sigma that was obtained as the standard deviation for each of the system, was used to carry out the well-tempered metadynamics simulations.

3. Results and Discussion

3.1 Docking

Figure 11, represents the docking energies obtained after docking the chosen known and new molecules to “Hotspot 2”. As we can see, the new inhibtors are quite comparable to the existing commercially available inhibitors when it comes to the docking energy. Though the docking data, in general, is not quantitative, it still gives a good preliminary idea about the feasibilty of binding. Hence, it becomes important to take these docked structures forward to calculate the free-energy of unbinding.

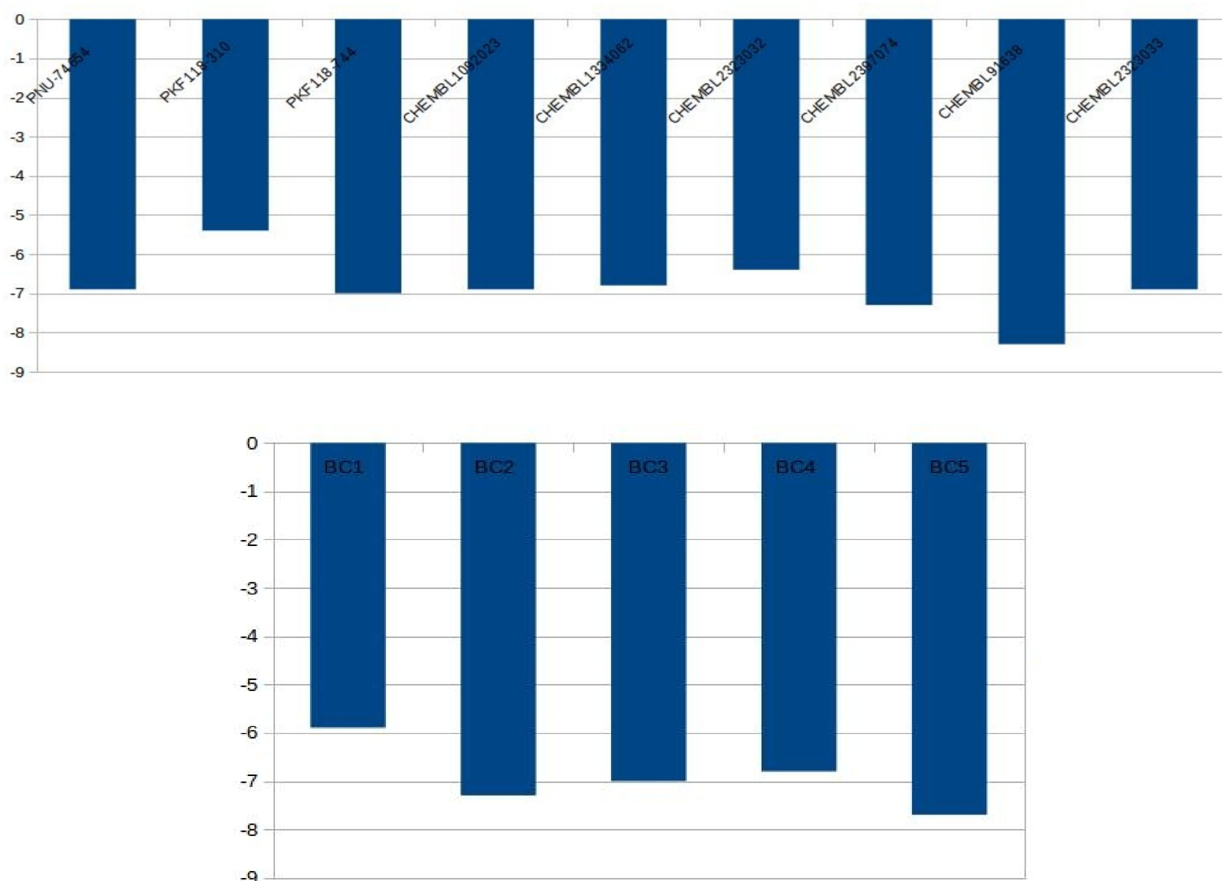


Figure 11: Bar chart representation of the docking energies obtained for the known molecules and new molecules repsectively.

3.2 Well-tempered metadynamics

3.2.1 BC1 and β-catenin

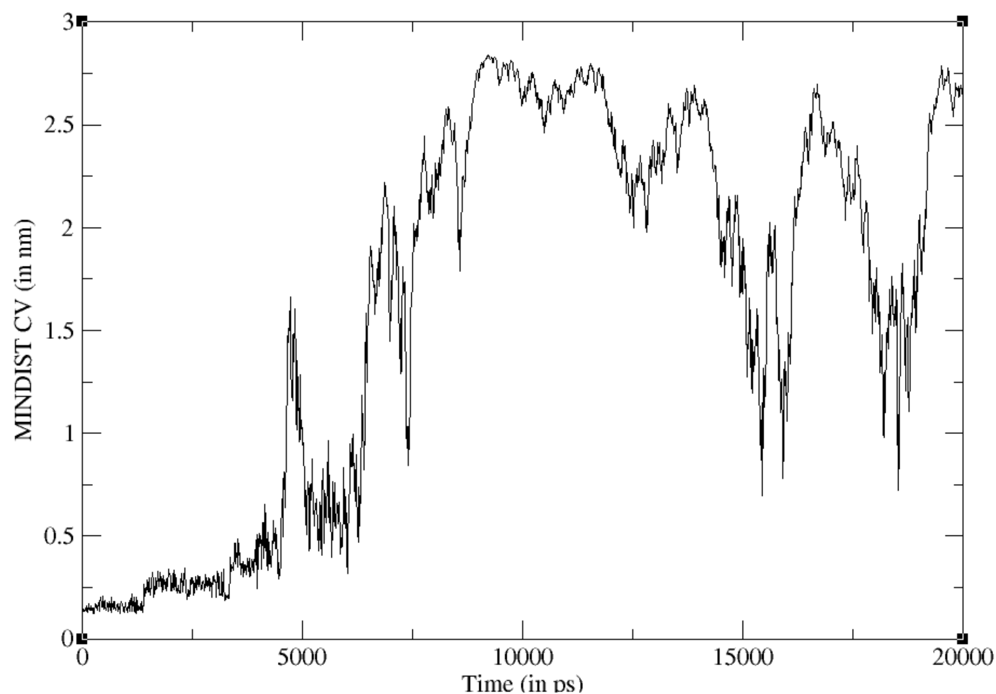
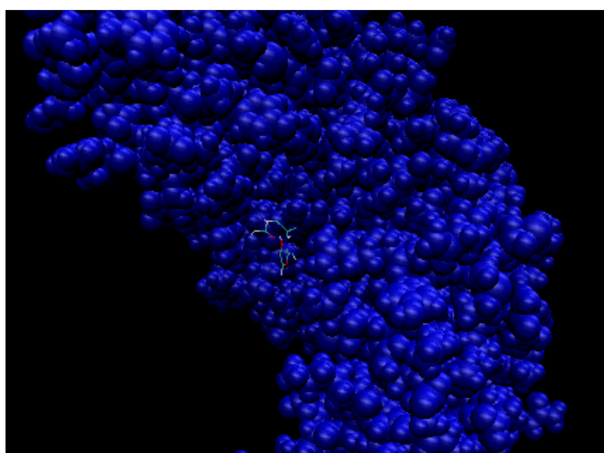
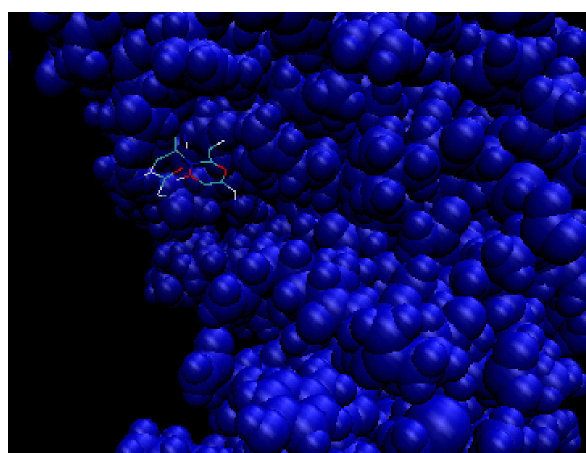


Figure 12: MINDIST CV (in nm) plotted vs time (in ps) for the BC1 and β-catenin system

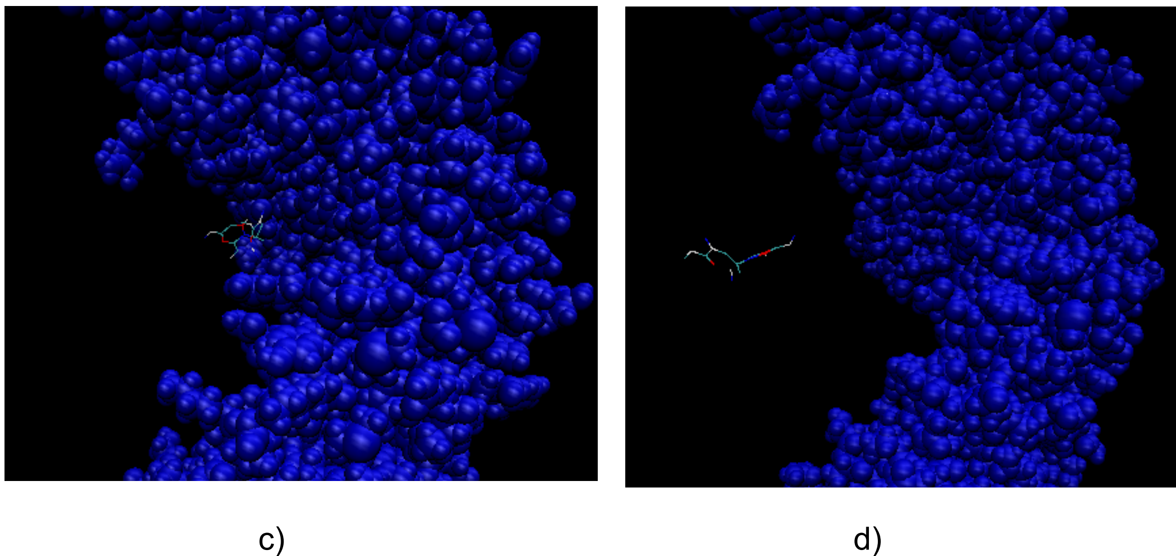
In **Figure 12,** we can observe that the MINDIST is oscillating around the value 0.18 nm till around 1 ns and then oscillating around the value 0.3 nm till around 3 ns before leaving the hotspot. **Figure 13** shows some snapshots from the metadynamics simulation trajectory, which might help in investigating the reason for this observation.



a)



b)



c) d)

Figure 13: Snapshots of the system (solvent molecules removed for visibility) at different time points in the well-tempered metadynamics simulation a) 0.5 ns b) 2.6 ns c) 4.0 ns d) 4.7 ns

As we can observe in the first two snapshots, 0.5 ns and 2.6 ns there are two relatively feasible binding modes near "Hotspot 2". This can be observed even in the free-energy surface shown in **Figure 14.** Around 4 ns, the BC1 molecule starts leaving the hotspot.

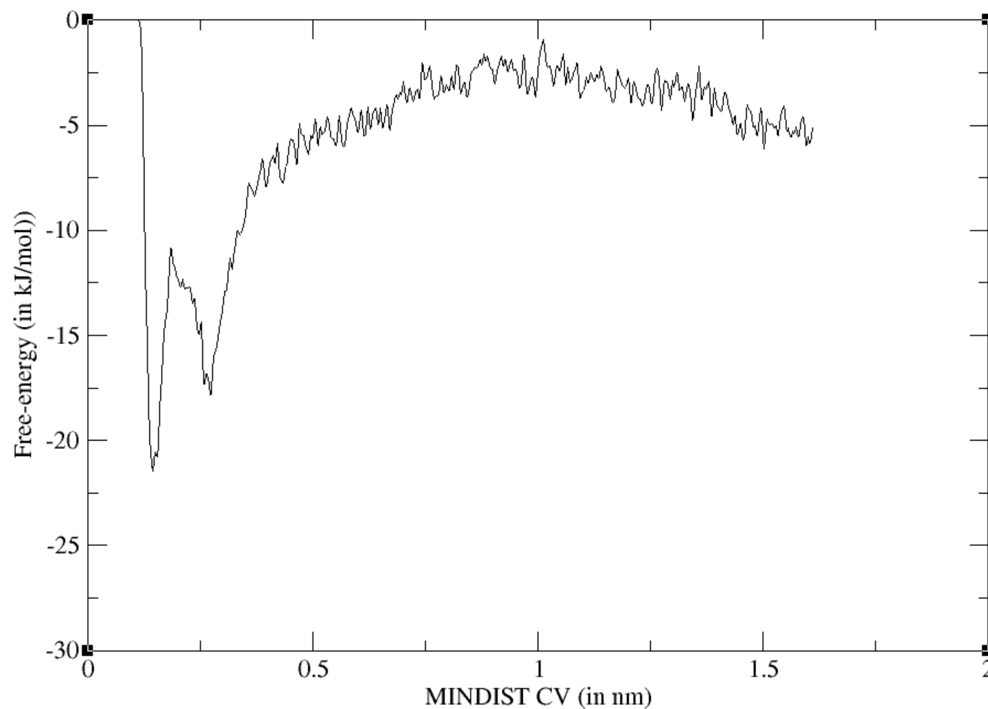


Figure 14: Free-energy surface obtained after well-tempered metadynamics simulation of the BC1 and β-catenin system

3.2.2 BC2 and β-catenin

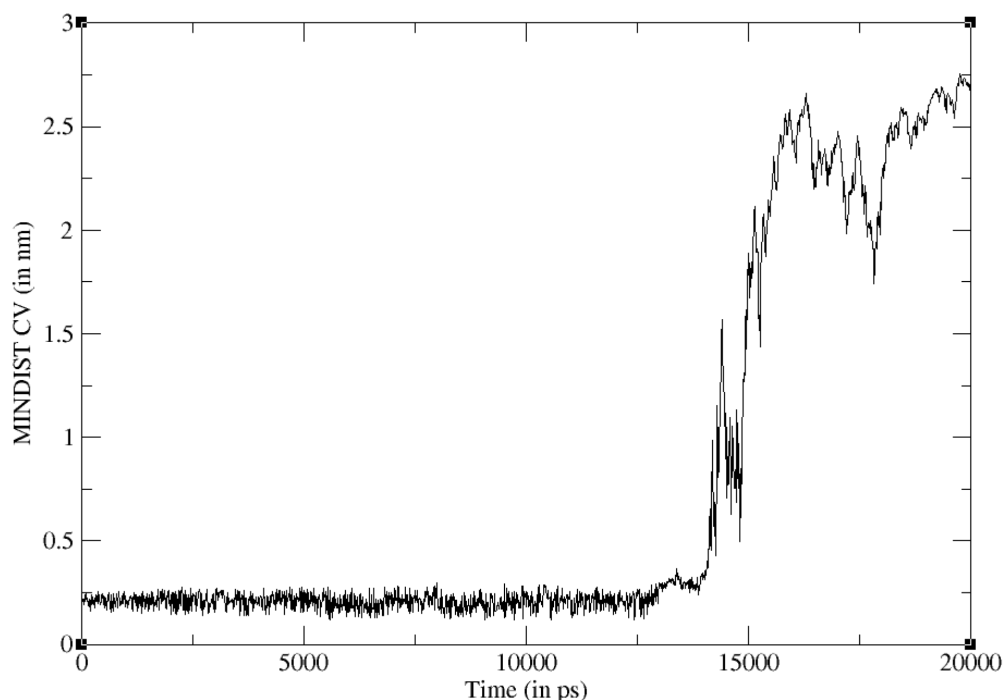
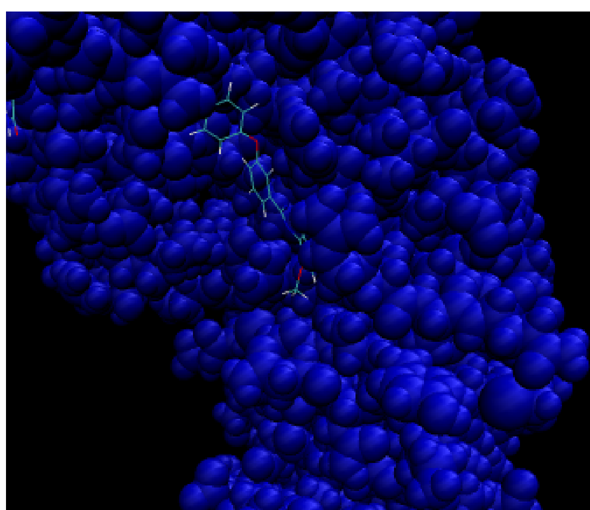
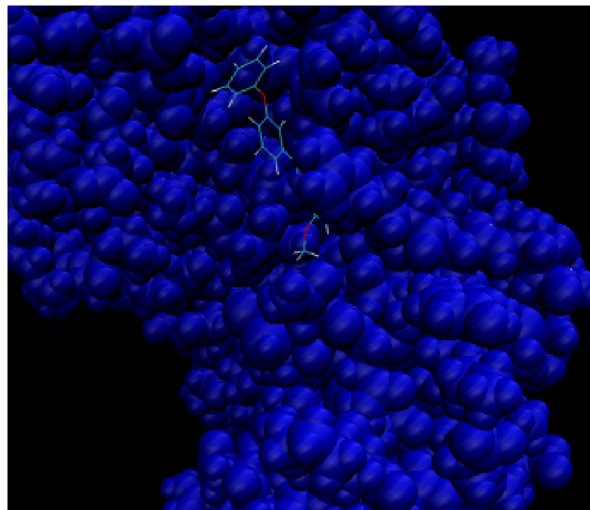


Figure 15: MINDIST CV (in nm) plotted vs time (in ps) for the BC2 and β-catenin system

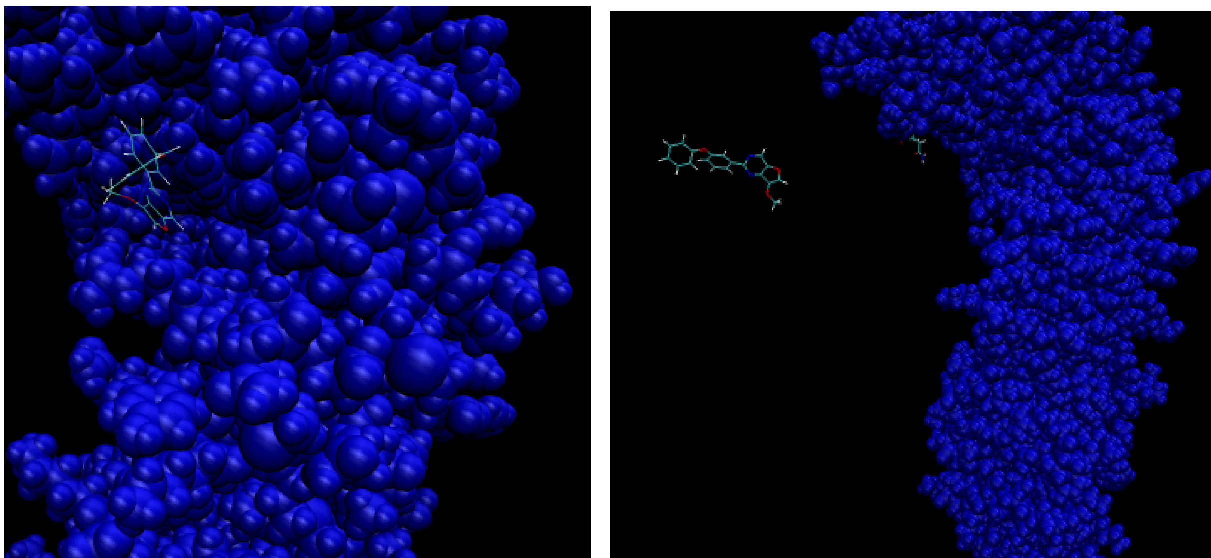
Figure 15 shows the MINDIST CV (in nm) as a function of time (in ps), during the well-tempered metadynamics simulation of the BC2 and β-catenin system. With the run parameters being same, we can see that the this molecule is taking longer time to unbind and leave the hotspot. This could mean that BC2 molecule is better at binding than the other molecules. **Figure 16** shows some snapshots of the system from the well-tempered metadynamics trajectory.



a)



b)



c) d)

Figure 16: Snapshots of the system (solvent molecules removed for visibility) at different time points in the well-tempered metadynamics simulation a) 4.0 ns b) 12.0 ns c) 14.0 ns d) 15.0 ns

Figure 17 shows the free energy plot obtained after the well-tempered metadynamics simulation of the BC2 and β-catenin system. The value obtained is around 60kJ/mol.

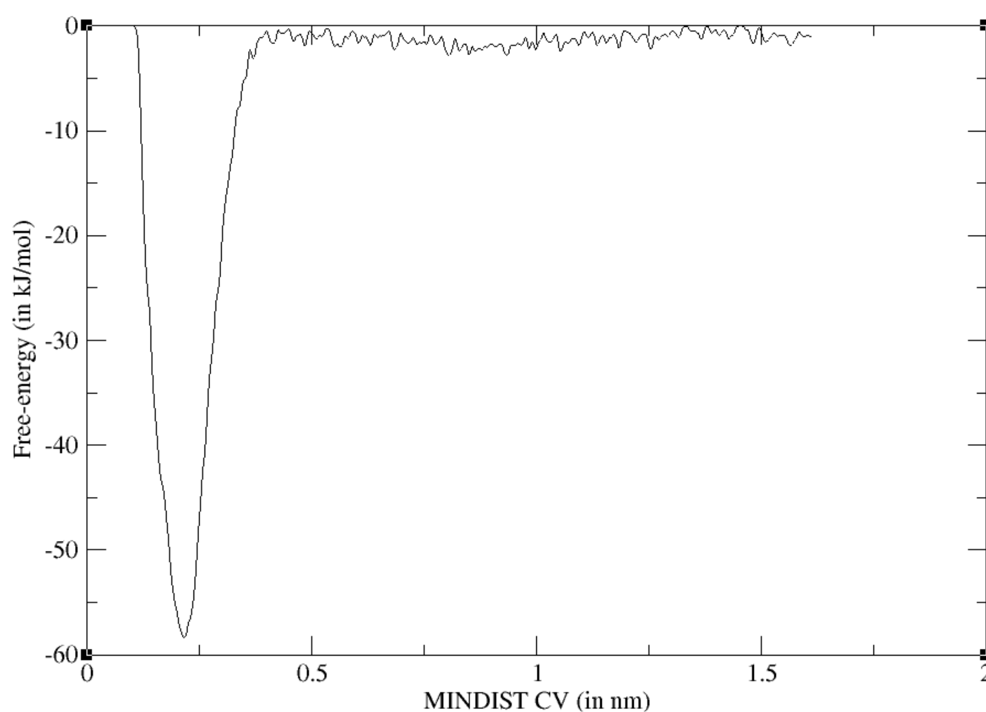


Figure 17: Free-energy surface obtained after well-tempered metadynamics simulation of the BC1 and β-catenin system

3.2.3 CHEMBL1334062 and β-catenin system

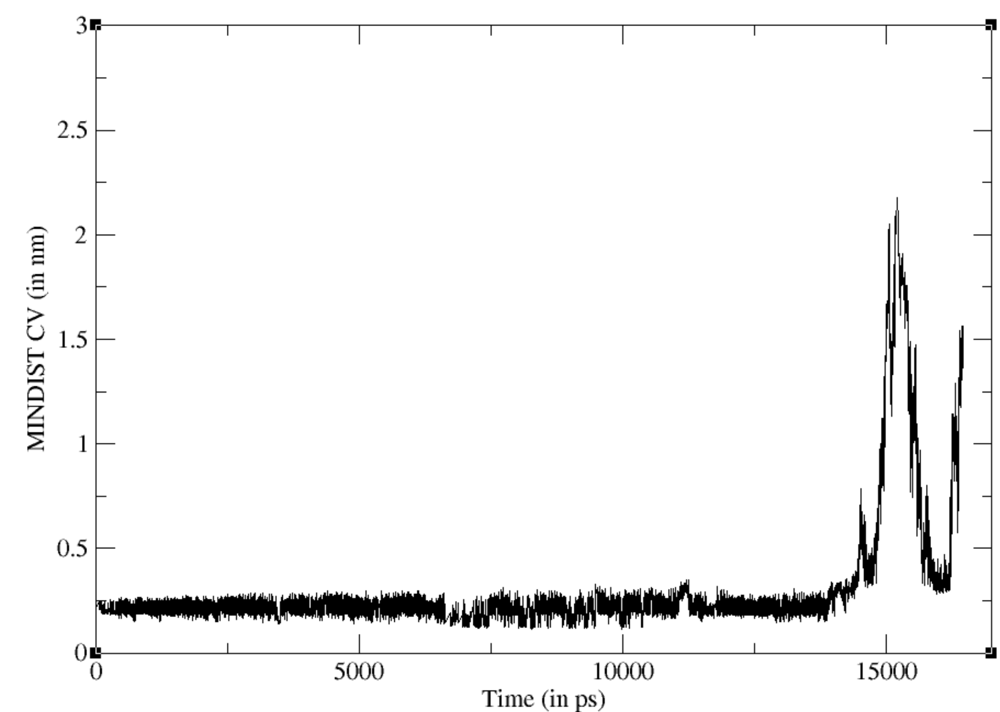
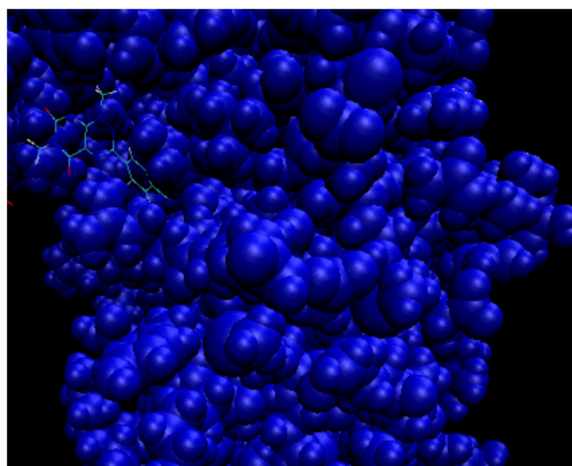
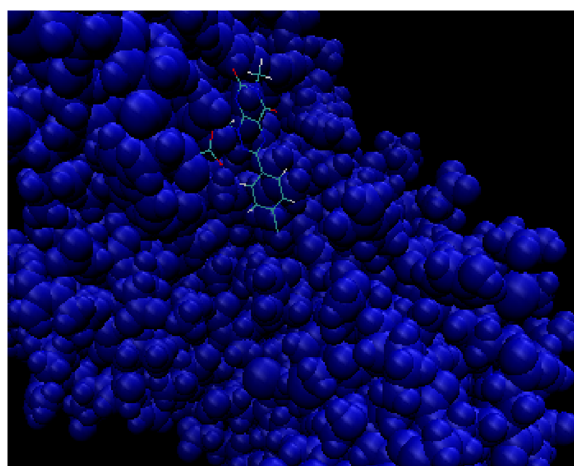


Figure 18: MINDIST CV (in nm) plotted vs time (in ps) for the CHEMBL1334062 and β-catenin system

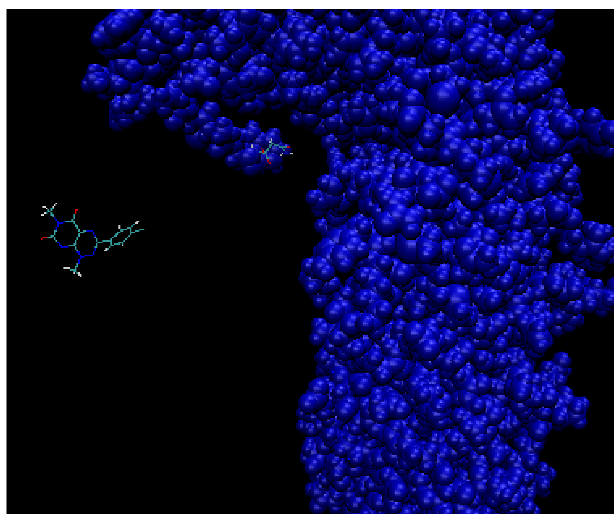
As we can see from **Figure 18,** the molecule is leaving the hotspot at around 14 ns. Hence, we can say that the affinity is on the stronger side. **Figure 19,** shows some snapshots along the metadynamics simulation.



a)



b)



c)

Figure 19: Snapshots of the system (solvent molecules removed for visibility) at different time points in the well-tempered metadynamics simulation a) 10.0 ns b) 14.0 ns c) 15.0 ns

Figure 20 shows the free-energy surface obtained after the well-tempered metadynamics simulation of the drug CHEMBL1334062 with β-catenin system.

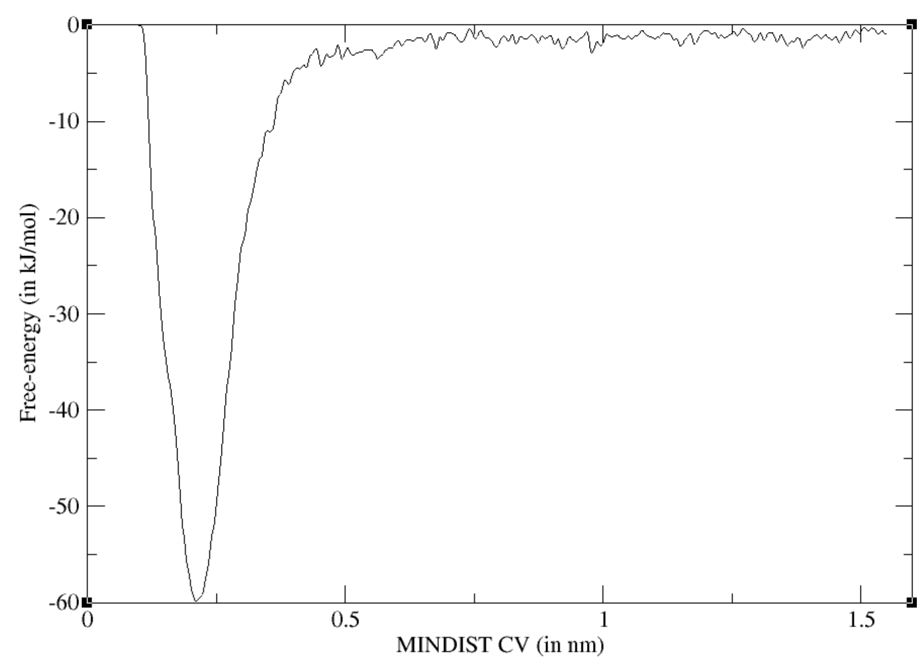


Figure 20: Free-energy surface obtained after well-tempered metadynamics simulation of the CHEMBL1334062 and β-catenin system

3.2.4 BC4 and β-catenin system

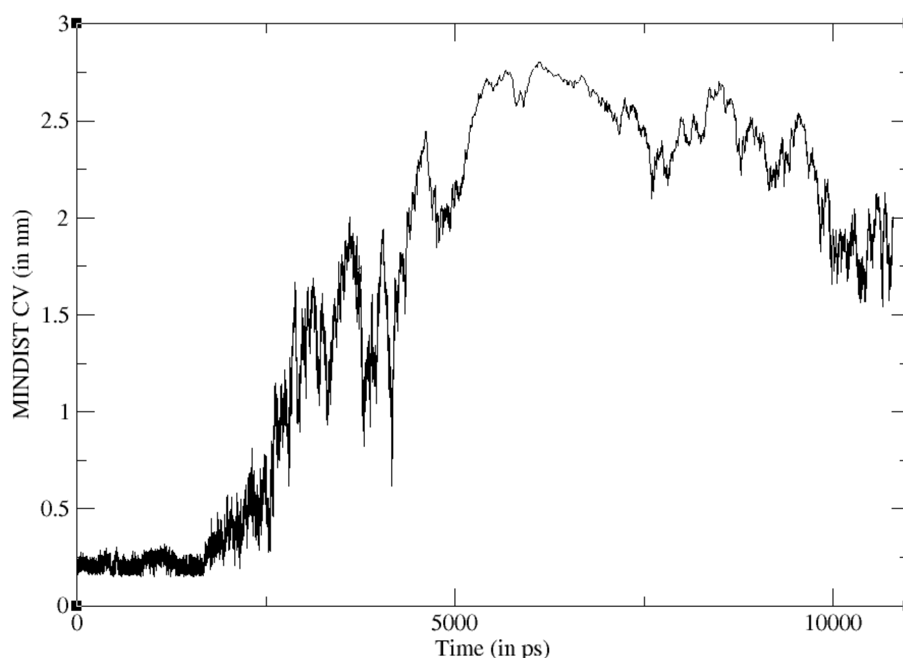
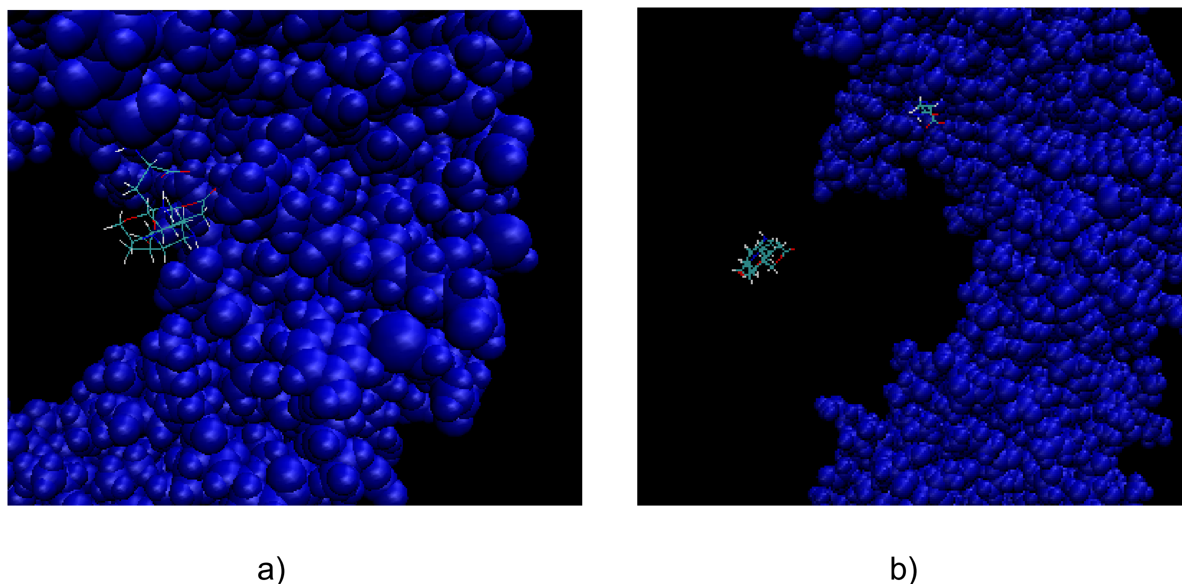


Figure 21: MINDIST CV (in nm) plotted vs time (in ps) for the BC4 and β-catenin system

From **Figure 21** we can conclude that the BC4 molecule is readily leaving the hotspot and thus is not a great binder to “Hotspot 2” of β-catenin. Some snapshots of the system at different time points through the well-tempered metadynamics simulations, is shown in **Figure 22.**



a) b)

Figure 22: Snapshots of the system (solvent molecules removed for visibility) at different time points in the well-tempered metadynamics simulation a) 1.0 ns b) 3.0 ns

Figure 23 shows the free energy surface obtained on running the well-tempered metadynamics simulation with the BC4 and β-catenin system. We can observe that it very low, further justifying the fact that it is a poor binder to “Hotspot 2” of β-catenin.

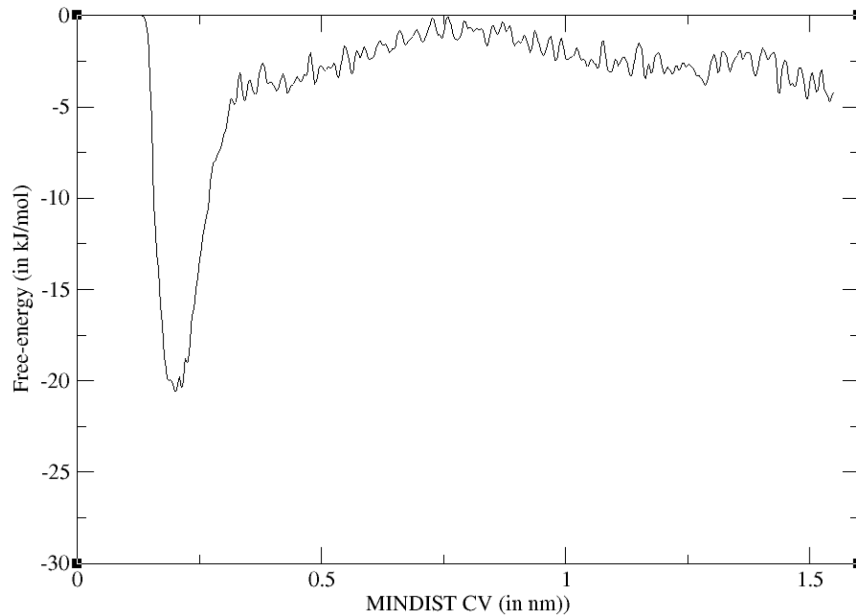


Figure 23: Free-energy surface obtained after well-tempered metadynamics simulation of the BC4 and β-catenin system

3.3 Comparitive study:

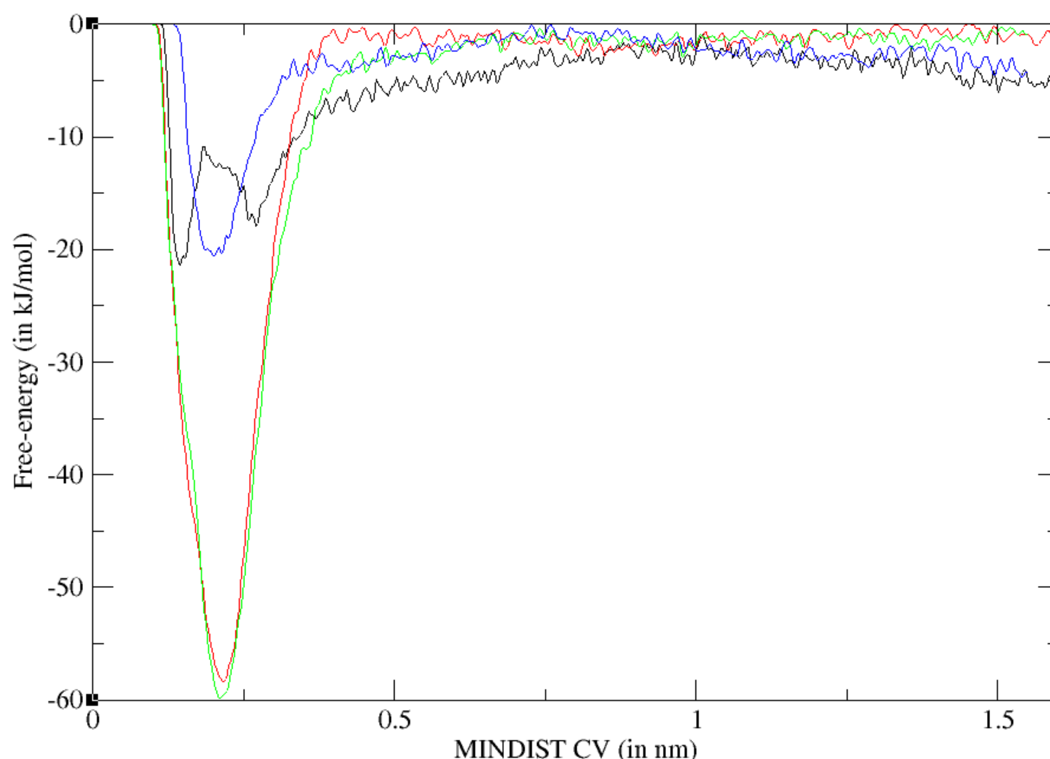


Figure 24: A plot contaning all the free-energy surfaces obtained, with black corresponding to BC1 molecule, red corresponding to BC2 molecule, green corresponding to CHEMBL1334062 molecule and blue corresponding to BC4 molecule

The free energy profiles shown in **Figure 24**, clearly show that the binding of BC2 is much more favoured than BC1 and BC4 among the new molecules. The free-energy profile of BC1 even shows two local minimas implying that the binding of BC1 to Hotspot 2 is not very specific, unlike in the case of BC2 and BC4 molecule which has a single minima. We can also observe that the free-energy profile of BC2 molecule is very similar to that of CHEMBL1334062 which is a well-known inhibitor of the β-catenin/TCF complex. This result shows that the molecule BC2, among the others, could be used as a potential inhibitor to conduct further studies.

Since, the metadynamics is a non equilibrium process it becomes necessary to calculate average energies for each value of the CV, which is MINDIST in this case. To do this, all the frames from the trajectory that belong to every particular value of

MINDIST and within a narrow range around that value, were collected so that all the average energies can be calculated.

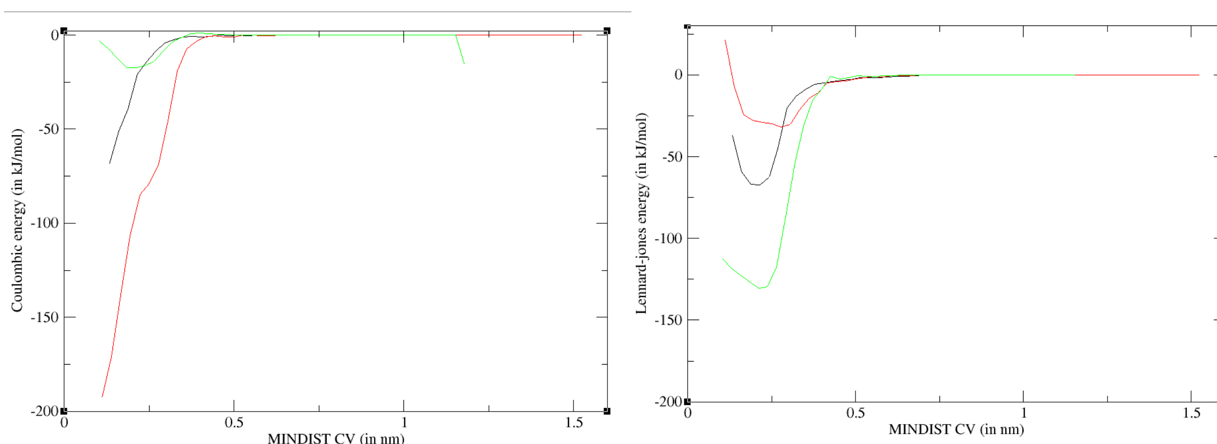


Figure 25: The plots show the coulombic and lennard-jones components of the interaction energy between β-catenin and the molecules. (Black – BC4, Red – BC1 and Green – BC2)

From **Figure 25,** we can see that the predominant interaction between the protein and BC1 is coulombic and in the case of BC2 it is the covalent interaction, i.e. the lennard-jones component. Compound BC4 is found in between the two other molecules with respect to both the coulombic and the Lennard-Jones component. This fact is further justified by the plots shown in **Figure 26**, which shows the interaction of the molecules with water and chloride ions. We observe a similar trend in the interaction energies between the molecule, water and chloride ions.

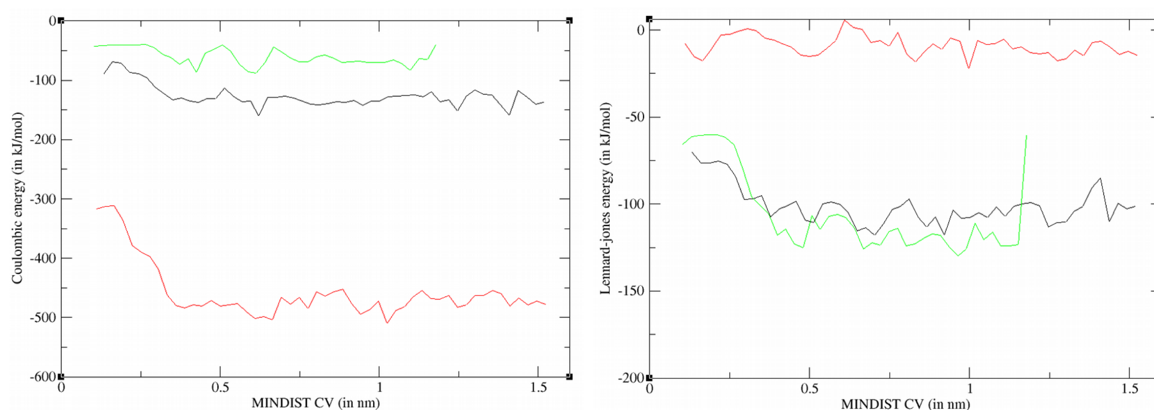


Figure 26: The plots show the coulombic and lennard jones components of the interaction energy between the molecules and water and chloride ions. (Black – BC4, Red – BC1 and Green - BC2)

Given the difference in the nature of interactions between the molecules and the protein molecule, another property that might help in studying the interaction would be hydrogen bonds number. **Figure 27**, shows the plot consisting of the number of hydrogen bonds between the molecules, BC1 and BC2, and the protein in all the frames obtained from the trajectory which are around the minima.

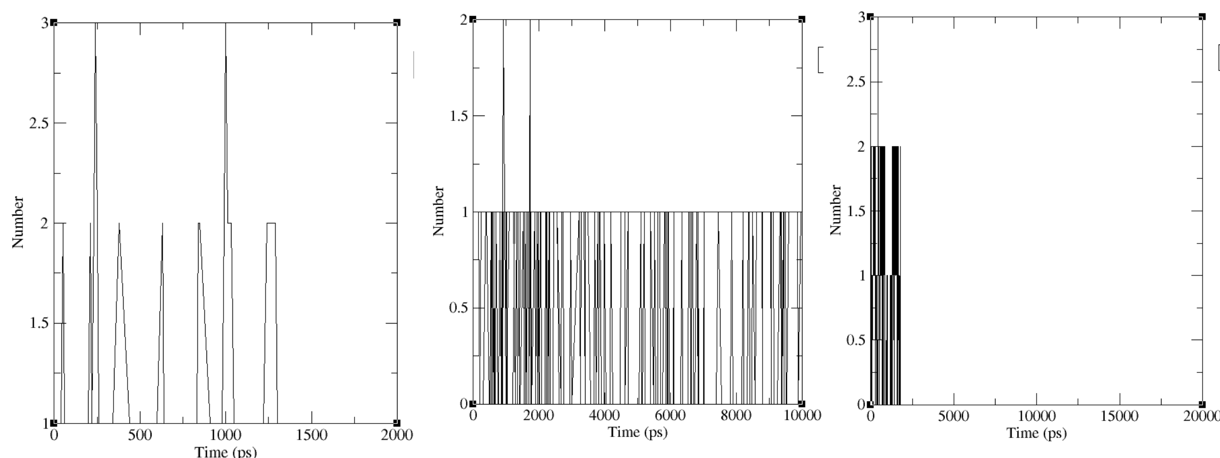


Figure 27: Plots of number of hydrogen bonds vs time for BC1, BC2 and BC4 respectively.

From the hydrogen bond number plots we can observe that in the case of BC1 and BC4 the average number of hydrogen bonds in the bound state is between 2 and 3, whereas in the case of BC2 the number of hydrogen bonds found in the bound state is on an average equal to 1.

Though, the fact that the interaction energies and the hydrogen bond analysis imply that BC1 and BC4 have a stronger interaction than BC2, we also know that the free-energy value obtained for BC2 is much more negative than that of BC1 and BC4. So, to explain this discrepancy one might have to study the entropy changes involved in the process. Though, it might be difficult to evaluate the entropy changes directly, there are other indirect ways of measuring the same. One such way is to determine the 4th order and the 6th order Steinhardt's parameter, which in a way determines the orderliness of atoms around a defined group of atoms. These are refferred to as Q4 and Q6 CVs, respectively, in the plumed driver software.

4. Conclusion

The free-energy studies reveal that the molecule BC2 is definitely a great binder to "Hotspot 2" of the β-catenin molecule and its binding free-energy is found to be comparable to that of CHEMBL1334062 molecule, which is a commercially known inhibitor of β-catenin. Though the interaction energies of BC1 and BC4 seems to be better than BC2, the entropy might be playing a huge role in determining the binding affinity of the molecule. This kind of study can be further carried out with other molecules to get a better picture about all the factors that contribute in determining the binding affinity of a molecule. The free-energy studies can be further strengthened by carrying out umbrella sampling studies to these molecules.

5. References:

1) Logan CY; Nusse R. The Wnt signaling pathway in development and disease. *Annual review of cell and developmental biology,* **2004,** 20, 781–810.

2) Clevers, H.; Nusse, R. Wnt/β-catenin signaling and disease. *Cell,* 2012, 149, 1192.

3) G. Hahne; T. N. Grossmann. Direct targeting of b-catenin: Inhibition of protein–protein interactions for the inactivation of Wnt signaling. *Bioorg. Med. Chem.,* **2013,** 21, 4020–4026.

4) M. Molenaar; M. van de Wetering; M. Oosterwegel; J. Peterson-Maduro; S. Godsave; V. Korinek; J. Roose; O. Destrée; H. Clevers. XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell*, **1996,** 86, 391-399

5) V.S. Li; S.S. Ng; P.J. Boersema; T.Y. Low; W.R. Karthaus; J.P. Gerlach; S. Mohammed; A.J. Heck; M.M. Maurice; T. Mahmoudi; H. Clevers. Wnt signaling inhibits proteasomal β-catenin degradation within a compositionally intact Axin1 complex. *Cell*, **2012,** 149, 1245-1256

6) Liu, W.; Dong, X.; Mai, M.; Seelan, R.S.; Taniguchi, K.; Krishnadath K.K.; Halling K.C.; Cunningham, J.M.; Boardman, L.A.; Qian, C. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signaling. *Nat. Genet.,* **2000,** 26, 146-147.

7) Thorne, C. A.; Hanson, A. J.; Schneider, J.; Tahinci, E.; Orton, D.; Cselenyi, C. S.; Jernigan, K. K.; Meyers, K. C.; Hang, B. I.; Waterson, A. G.; Kim, K.; Melancon, B.; Ghidu, V. P.; Sulikowski, G. A.; LaFleur, B.; Salic, A.; Lee, L. A.; Miller, D. M.; Lee. Structure-Based Design of Potent and Selective CK1γ Inhibitors. *E. Nat. Chem. Biol.* **2010**, 6, 829.

8) Chen, B. Z.; Dodge, M. E.; Tang, W.; Lu, J. M.; Ma, Z. Q.; Fan, C. W.; Wei, S. G.; Hao, W. N.; Kilgore, J.; Williams, N. S.; Roth, M. G.; Amatruda, J. F.; Chen, C.; Lum, L. *Nat. Chem. Biol.,* **2009**, 5, 100.

9) Sampietro J.; Dahlberg C.L.; Cho U.S.; Hinds T.R.; Kimelman D.; Xu W.;. Crystal structure of a beta-catenin/BCL9/Tcf4 complex. *Mol Cell,* **2006**, 24, 293–300

10) Trosset, J. , Dalvit, C. , Knapp, S. , Fasolini, M. , Veronesi, M. , Mantegani, S. , Gianellini, L. M., Catana, C. , Sundström, M. , Stouten, P. F. and Moll, J. K. Inhibition of protein–protein interactions: The discovery of druglike β-catenin inhibitors by combining virtual and biophysical screening. *Proteins*, **2006**, 64, 60-67.

11) Zecheng Chen, Aranapakam M. Venkatesan, Christoph M. Dehnhardt, Osvaldo Dos Santos, Efren Delos Santos, Semiramis Ayral-Kaloustian, Lei Chen, Yi Geng, Kim T. Arndt, Judy Lucas, Inder Chaudhary, Tarek S. Mansour. 2,4-Diamino-quinazolines as inhibitors of β-catenin/Tcf-4 pathway: Potential treatment for colorectal cancer. *Bioorganic & Medicinal Chemistry Letters*, **2009**, 19, Issue 17, 4980-4983.

12) Chi Hoon Park, Joon Young Chang, Eun Ryeong Hahm, Seyeon Park, Hyun-Kyung Kim, Chul Hak Yang. Quercetin, a potent inhibitor against β-catenin/Tcf signaling in SW480 colon cancer cells. Biochemical and Biophysical Research Communications, **2005**, 328, Issue 1, 227-234.

13) Maina Lepourcelet, Ying-Nan P. Chen, Dennis S. France, Huisheng Wang, Phillip Crews, Frank Petersen, Charles Bruseo, Alexander W. Wood, Ramesh A. Shivdasani. Small-molecule antagonists of the oncogenic Tcf/β-catenin protein complex, *Cancer Cell*, **2004**, 5, Issue 1, 91-102.

14) Trott, O. and Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **2010**, 31, 455-461.

15) Bekker, H., Berendsen, H. J. C., Dijkstra, E. J., Achterop, S., van Drunen, R., van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., Renardus, M. K. R.

Gromacs: A parallel computer for molecular dynamics simulations. *Physics Computing,* **1993**, 92.

16) Berendsen, H. J. C., van der Spoel, D., van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.,* **1995**, 91, 43–56.

17) Lindahl, E., Hess, B., van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.,* **2001**, 7, 306–317.

18) van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., Berendsen, H. J. C. GROMACS: Fast, Flexible and Free. *J. Comp. Chem.,* **2005,** 26, 1701–1718.

19) Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comp.,* **2008,** 4(3), 435–447.

20) M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia and M. Parrinello. PLUMED: a portable plugin for free-energy calculations with molecular dynamics, *Comp. Phys. Comm.* **2009,** 180 (10), 1961-1972

21) B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, *John Wiley & Sons*, Inc., 5.6.1-5.6.37, 2016.

22) F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski & P. Cieplak, The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building, *Phys. Chem. Chem. Phys.* **2010**, 12, 7821-7839

23) Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.;

Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.
24) D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. The Amber biomolecular simulation programs. *J. Computat. Chem.,* 26**, 2005**, 1668-1688.
25) A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.,* **2008**, 100, 020603