Fall 12-14-2018

The Drug Discovery of Potential Anti-Cancer Agents to Inhibit the Demethylation Catalytic Activity of the JmjC domain of KDM3A

Balgees Alshanqeti
Pittsburg State University, balshanqeti@gus.pittstate.edu

THE DRUG DISCOVERY OF POTENTIAL ANTI-CANCER AGENTS TO INHIBIT THE DEMETHYLATION CATALYTIC ACTIVITY OF THE JMJC DOMAIN OF KDM3A

A Thesis Submitted to the Graduate School
In Partial Fulfillment of the Requirements for the Degree of
Master of Science

Balgees Alshanqeti

Pittsburg State University

Pittsburg, Kansas

December 2018

THE DRUG DISCOVERY OF POTENTIAL ANTI-CANCER AGENTS TO INHIBIT THE DEMETHYLATION CATALYTIC ACTIVITY OF THE JMJC DOMAIN OF KDM3A

Balgees Alshanqeti

APPROVED:

Thesis Advisor _____
Dr. Irene Zegar, Department of Chemistry

Committee Member_____
Dr. James McAfee, Department of Chemistry

Committee Member_____
Dr.Santimukul Santra, Department of Chemistry

Committee Member_____
Dr. Benjamin Tayo, Department of Physics

ACKNOWLEDGEMENTS

First of all, I would like to thank Allah for helping get through with this research after lots of suffrage and sleepless nights. Secondly, I would like to thank my government for giving me a chance to achieve my goal and make my dream come true by taking care of all the financial aspects of my study.

Special thanks to my advisor Dr. Zegar for her guidance and advisory throughout this process and for being very helpful and responsive whenever I needed her. I also would like to thank Dr. McAfee for his vital advisory rule and leaving an open door to my constant questioning. Furthermore, special thanks for Dr. Tayo and Dr. Santra for agreeing to be part of my master defense committee.

Additionally, I would like to thank my brother Abdulrahman for his constant support and for every beautiful moment, we lived in our time here in the United States. Not to forget the vital rule my mother and grandmother played in my support to achieve this, and their believe in me even throughout tough times. I also would like to thank my best friend Nehad Nawfawi for helping me complete this research and her massive support through. Further, I would like to thank all my friends for believing in me and supporting me all the time.

THE DRUG DISCOVERY OF POTENTIAL ANTI-CANCER AGENTS TO INHIBIT THE DEMETHYLATION CATALYTIC ACTIVITY OF THE JMJC DOMAIN OF KDM3A

An Abstract of the Thesis by
Balgees Alshanqeti

Lysine demethylase 3A (KDM3A) is an enzyme that specifically catalyzes the removal of 2 or three methyl groups from lysine 9 of histone 3 (KH3). It belongs to the family of histone demethylases that contain the Jumonji C (JmjC) domain, which means cruciform in Japanese. KDM3A also belongs to a family of hydroxylases that are alpha-ketoglutarate dependent. The importance of the role of KDM3A demethylation in the cell is not completely understood, however, studies have showed that its expression is elevated in embryonic stem cells in humans and in mice. In general, patterns of elevated expression of the KDM3A gene have been associated with cell differentiation and cell proliferation in several tissues. Studies have also shown that KDM3A was found to be overexpressed in breast cancer cells, and it is believed to be associated with the regulation of estrogen receptors and with breast cancer growth and metastasis. KDM3A is also found to be upregulated in small cell lung cancer cells which are the least common of lung cancer cells but the most aggressive. Other cancer cells that KDM3A has been associated with and is found to undergo enhanced expression are prostate cancer and the nervous system cancer cells. In all of those cancers, KDM3A was found to be linked to the role of promoting cancer stem cell progression and resistance to chemo and radiation therapy. This role is primarily attributed to the role that KDM3A in demethylating KH3 which is known to result in gene repression of the tumor suppression protein, p53 and in the formation of inactive heterochromatin. The goal of this thesis is to begin a drug discovery process in order to determine potential therapeutic agent(s) that can bind the JmjC domain of KDM3A and ultimately inhibit its catalytic activity in the demethylation of KH3. The first step in this process is to conduct a virtual screening study where a large library of small molecule structures is examined for their interactions with the JmjC domain of KDM3A. To do this the coordinates of the three-

dimensional structure of the JmjC domain along with the approximate coordinates that outline possible binding sites to be occupied are inputted in a computer software as the receptor molecule, and its binding search space. A large number of structures of ligands from small molecule databases are then used to find a set of ligands that form the most energetically favorable complexes with the protein receptor site. Once a set of protein ligand complexes have been determined, we screened the whole library of complexes for those with the highest binding affinities to the protein receptor. Further analysis was also conducted regarding the types of interactions between the ligand and the receptor and how close their binding sites are to the crucial catalytic region of KDM3A. The objective would be to choose the ligands that binds the tightest to KDM3A and act to block the catalytic domain from carrying out its function. A total of six small molecules that fulfilled one or both of the above requirements were chosen. Examinations of the interactions of these small molecules to the various amino acids, and the cofactors in the catalytic region were conducted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES PAGES

Chapter I

Introduction

DNA is the genetic material and consists of over 3 billion base pairs and is located in the nucleoid of prokaryotes and the nucleus of eukaryotes cells. Because of the macroscopic length of DNA relative to the volume of the cell DNA is highly condensed through the interaction with a group of proteins known as histones to form chromatin. The lowest order of chromatin structure is the nucleosome which consist of 147 nucleotides of DNA wrapped 1.7 times around an octamer of histones. The octamer of histones consists of two dimers of H2A and H2B, and a heterotetramer of H3 and H4. One core nucleosome is connected to another by a DNA linker of up to 80 base-pairs . A fifth histone H1 is associated with a linker DNA and is believed to be necessary for the second level of chromatin condensation. The next level of condensation involves the wrapping of the nucleosomes into a solenoidal structure referred to as the 30 nm chromatin fiber. Higher order chromatin structures are believed to result from the interaction of the 30 nm chromatin fiber with an insoluble matrix present in the nucleus that further condenses the chromatin. Higher levels of condensation are required to accommodate the condensed chromatin structure observed in metaphase chromosomes.

In addition to packaging chromosomes inside the nucleus, chromatin structure plays a vital role in protecting the DNA from damage and in regulating gene expression [1]. Histones complexed with chromatin undergo a number of post-translational modifications including acetylation, phosphorylation, methylation, sumoylation, ubiquitination, and several other modifications. Most of these modifications occur on the amino-terminal ends of the histones that project away from

the nucleosome surface (i.e they are exposed). These post-translational processes are essential epigenetic modifications that impact multiple biological processes such as chromosomal packaging which in turn influences gene expression through either promoting or inhibiting transcriptional complexes.

The effect of acetylation by histone acetyl transferases (HAT) and deacetylation by histone deacetylases on gene expression is well understood with the former modification inhibiting transcription and the latter stimulating it. However, methylation of histone tails can both stimulate or inhibit transcription. In particular, the methylation of H3 on lysine 9 (H3K9) represses gene expression while H3K4 activates transcription. To further complicate our understanding, lysines may undergo mono-, di-, or trimethylation.

Demethylation involves the removal of one, two or three methyl groups from lysine and arginine. Recent studies indicate that demethylation of histone is the opposite process of histone methylation. In other words, histone demethylases that remove the H3K4 methylation silence gene transcription, and the demethylation of H3K9 unwraps the DNA and activates gene expression [13][2][3].

Demethylation is catalyzed by a host of demethylases that are classified into two major groups. The LSDI group contains the amine oxidase catalytic domain and the JHDM group contains the jumonji (JmjC) catalytic domain. Histone demethylases that belong to the LSD1 group require the flavin adenine dinucleotide (FAD) amine to mediate the crucial oxidation step in the demethylation process. Whereas, the histone demethylases that belong to the JHDM group require Fe (II) and α-ketoglutarate-dependent dioxygenase (α-KG) to mediate their crucial oxidation step in the demethylation process. [4][5].

This thesis focuses on one particular histone demethylase, namely, KDM3A, which belongs to the JHDM group that contains the JmjC catalytic domain. KDM3A is also a member of the Fe (II) and α-ketoglutarate-dependent dioxygenase demethylases. It is known to catalyze the removal of the second and third methyl groups on H3K9, a process that is linked to the activation of gene

transcription. The KDM3A sequence also contains several copies of the contiguous five- amino acid motif sequence, LXXLL (L for leucine) which is an important motif present on most transcriptional coactivators and is responsible for the interaction between KDM3A and hormonal receptors. [4][5].

Recently KDM3A has been shown to play a role in the development and progression of cancer, specifically in mammary gland cancer. In cancer cells, KDM3A has been linked to the demethylation of H3K9 and to the demethylation of the tumor suppressor protein, p53. The demethylation of H3K9 in cancer cells was linked to the initiation of stem-based cancer cells which are believed to be responsible for the proliferation, drug -resistant and the metastatic properties of late stage cancers. The demethylation of p53 was found to lead to the suppression of the pro-apoptotic properties of the protein p53, a crucial function of p53 that leads to cell death in cells that are pre-cancerous. [5][6].

KDM3A has not only been found to actively take a part in the growth and development of ovarian cancer, but it has also been implicated in the development and maintenance of the malignant stage ovarian cancer cells. Its ability to promote the metastasis of ovarian cancer is primarily linked to its role in the initiation and propagation of cancer stem cells (CSC), believed to be resistant to the stress induced by chemotherapy and radiotherapy [7] [8] [9].The role of KDM3A in prostate cancer has been linked to its enhancement of the c-Myc transcriptional activation by a concerted effort of the LXXLL motif used to recruit the androgen receptor to the c-Myc gene enhancer along with the activation of the demethylation of H3K9, a process that leads to an increase in the transcription of the c-Myc mRNA. [5].

KDM3A has also been found to be markedly upregulated in small cell lung cancer cells, Small cell lung cancer is the less frequent lung cancer effecting patients in the USA, but it is the most aggressive and deadliest lung cancer known. KDM3A was also liked to the progression and migration of small cell lung cancer cells *in vitro* [10]. Finally, the role of KDM3A in promoting

the aggressive expression of central nervous system cancer cells stem cell. has been linked to the demethylation activity of the JmjC domain [11].

The aforementioned multiple roles of KDM3A in cancer make it a primary target for drug cancer therapy. The goal of the research outlined in this thesis is to use the JmjC domain of KDM3A as a cellular target to conduct virtual screening of large sets of small-molecule databases in an effort to find potential therapeutic agent(s) that target the catalytic domain of KDM3A in cancer cells in order to deactivate the enzyme and thus hinder it from carrying out its activity in cancer cells. This was accomplished, first, by developing a three-dimensional structure of the JmjC domain of KDM3A using a technique called homology modelling. This is necessary because virtual docking protocols require the availability of structures of biological target molecules. The development of the three-dimensional structure of the JmjC domain of KDM3A was done using homology or comparative modelling, a technique that is widely used in pharmaceutical research for drug discovery purposes. Virtual docking was then carried using large databases of small molecules. Exhaustive screening and analysis were finally conducted to develop a small number of potential therapeutic agents that bind strongly to the JmjC domain of KDM3A and will hopefully inhibit, the KDM3A demethylation activity in cancer cells.

Chapter II

Materials and Methods

Computer Programs for KDM3A Tertiary Structure Prediction:

Chimera is a software that was developed by the Resource for Biocomputing, Visualizing and informatics (RBVI) at University of California, San Francisco (UCSF) with a supported contribution from the National Institutes of Health [14]. Chimera is an interactive visualizing program. It is used to analyze molecular structures, provide docking results, sequence alignments, trajectories, etc. [15]. Chimera also provides a graphical interface to remotely access the program Modeller which is used in this work to generate a prediction model of the tertiary structure of KDM3A using comparative or homology modeling with limited incorporation of ab initio loop region structure prediction protocols. This was conducted through a webserver maintained by RBVI at UCSF [16].

Computer Programs Used for the Virtual Screening of Small Molecule Libraries for the Targeting of KDM3A in Cancer Cells:

Autodock vina an open source computer program developed at the Molecular Graphics Laboratory at the Scripps Research Institute [17]. AutoDock Vina was used to simulate the interactions between the Modeller generated tertiary structure of KDM3A and small molecules using large numbers of small molecule databases [17]. Autodock Vina was access remotely through the opal2 webserver, nbcr-222.ucsd.edu. Three small molecule databases available on the opal webserver were used for the vitural docking study. These databases were as follows: the NCIDS_SC database, which represents the NCI Diversity Set 1 with rotamers, prepared 2006-

2007 and conains 2282 compounss, NCI_DS3 database, which represents the NCI Diversity Set 3, prepared 2012 and contains 1563 molecules, and the Steriods database prepared 2014 and contains 652 molecules [18] [19]. AutoDock Vina was also used locally for some small molecule docking through PyRx , an all-in-one software that performs processes from receptor structure preparation to final job submission in a few user-friendly steps [20].

AutoDock Tools (ADT) (also developed at the Molecular Graphics Laboratory at the Scripps Research Institute) is a free software used to perform structure editing, the assignment the receptor's binding search space, viewing the docking results, and computing interactions between the small molecules (ligands) and KDM3A (receptor) [21].

Yasara view is the freeware part of the Yasara suite. This software was developed by Elmer Krieger, the founder of Yasara Biosciences in Vienna Austria [22]**.** Yasara View was primarily used to interactively view and analyze protein structures and to create some of the structural images used in this thesis. A free download of YASARA View was obtained through the website: http://www.yasara.org/viewdl.htm

FoldX is an empirical force field developed in the Serrano lab at the Heidelberg Laboratory of the EMBL by Raphael Guerois, Jens Nielsen, Jesper Ferkinghoff-Borg, Joost Schymkowitz, Frederic Rousseau, Francois Stricher and Luis Serrano [23]. It was later implemented for use as a plugin in Yasara View by Joost Van Durme at the Switch Laboratory. FoldX was used in repairing the homology modelling generated JmjC domain of KDM3A from steric clashes and to perform energy minimization on the resulting structure.

Creating a Virtual Structure of KDM3A Using Homology Modelling by Modeller:

The complete amino acid sequence of the homo sapiens KDM3A protein was downloaded from Medline via the website: www.pumbed.gov. The part of the sequence that corresponds to the JmjC domain of KDM3A was excised and a protein blast was conducted using Chimera by choosing the Tools pulldown menu followed by choosing "Sequence" and "Blast Protein". The JmjC amino acid sequence was then pasted in the "Plain Text" window, and a Blast Protein was

initiated by choosing \" ok" .The "4C8D_A" query from the Blast: query window was selected followed by selecting the "Load Structure" option. The crystal structure of the PDB ID: 4C8D which belongs to the JmjC domain of KDM3B was accessed from www.pubmed.gov. To begin the homology modelling of the JmjC domain of KDM3A using the structure of the JmjC domain of KDM3B, a sequence alignment between the two domains was carried out by selecting "Sequence" from the "Tools" menu followed by selecting "Sequence" This is followed by adding the amino acid sequence of the JmjC domain of KDM3A by selecting "Add Sequence" from the pulldown menu "Edit" of the displayed sequence window. In the "Add Sequence to Chain A:" window, the name KDM3A was added to Sequence name: and the sequence for the KDM3A JmjC domain was pasted in the sequence window. To remove gaps in the two sequences, the Realign sequences selection from the edit pulldown menu was chosen followed by "ok" to realign the two sequences. From the "Realigned Sequences" window, the Modeller (homology) plugin was selected from the "Structure" pulldown menu. The KDM3A sequence is then selected from the "Choose sequence to be modeled" menu, and the "4C8D" structure from "Sequence" was selected as the template structure to be used for homology modelling. The "Run Modeller from the Webserver" choice was selected and the Modeller License Key: MODELIRANJE was used and from "Advanced Options" the "Include Non-Water HETATM from Template" choice was selected. Homology modelling was then carried out by selecting "OK". Once the structure is generated, Modeller (loops/refinement) was chosen from the "Structure" pulldown menu for more accurate prediction of flexible loop structures. The resulting structure was then saved and used as is in the subsequent AutoDock Vina calculations.

Virtual Docking of Small Molecules in the JmjC Domain of KDM3A Using Vina:

To carryout virtual docking vina was used through the opal2 webserver (www.ncbr-222.ucsd.edu/) which required the submission of the generated structure of JmjC domain of KDM3A, and the submission of approximate coordinates of the search space of the desired binding site on the generated structure. Those coordinates were determined using the AutoDock

Tools software. This was done by first loading the generated KDM3A structure using the “Read Molecule” from the File pulldown menu in AutoDock Tools. The coordinates for the search space of the potential binding site in the JmjC domain of KDM3A were chosen using Gridbox from the “Grid” pulldown menu. In the Grid options window, the “Spacing (Angstroms) is set to 1.000, the x, y, z coordinates for the box center were set near the middle of the region that represents the desired location of the ligands binding site. The x, y, z dimensions that border the binding site were also set making sure that enough space is allowed to accommodate bulky ligands.

Chapter III

Results

The Generation of the Three-Dimensional Structure of JmjC Domain of KDM3A Using HomologyModelling:

Recent findings conclusively demonstrated that the lysine specific demethylase, KDM3A , an enzyme that is found to be upregulated in breast cancer cells, plays a crucial role in promoting metastatic breast cancer by removing a methyl group from lysine 9 of histone H3, which functions to suppress pro-invasive genes and by demethylating the tumor suppressor p53 protein which is normally associated with the chromatin and it induces pro-apoptotic events that cells with damaged DNA undergo to prevent the initiation and progression of cancer [8].

Thereby the main goal of the work conducted in this thesis is to discover a potential therapeutic agent to be used as anti-cancer drug to prevent breast cancer cells from metastasizing by targeting the catalytic domain (JmjC domain) of KDM3A. An efficient way to go about this is to carry out a virtual screening protocol where the three-dimensional structure of the cellular target (protein or nucleic acid) is used as a receptor for the potential binding of small molecules (ligands) found in a large small-molecule structural database. However, a crystallographic or an NMR solution structure of the JmjC domain of KDM3A is not yet available. Therefore, to generate the structure we used homology or comparative modelling, a technique that is used extensively in pharmaceutical research as a first step in drug discovery and drug design. This technique first involves the search for a protein that has an existing structure that contains high sequence homology to the protein of interest. A sequence alignment between the two proteins is conducted.

The structure of the target protein is then developed using Modeller by backbone and side-chain generation and optimization, structure building of missing parts, along with the use of ab initio protocols for the generation of loop structures which are usually difficult to predict even in experimentally determined structures. The steps outlined above were conducted on KDM3A using the program Modeller. Using this program, KDM3B, an enzyme from the same demethylase family as KDM3A was identified as a highly homologous protein with an existing three-dimensional structure. The amino acid sequence homology between the JmjC domains of both enzymes was determined by aligning the two sequences as shown in Figure 1 where the amino acid sequence of the JmjC domain of KDM3A is shown on the bottom and the JmjC domain of KDM3B is shown on top. The alignment scheme showed that the two sequences have 80% amino acid identity. This high level of identity is crucial for the successful generation of protein structures using homology modellig.

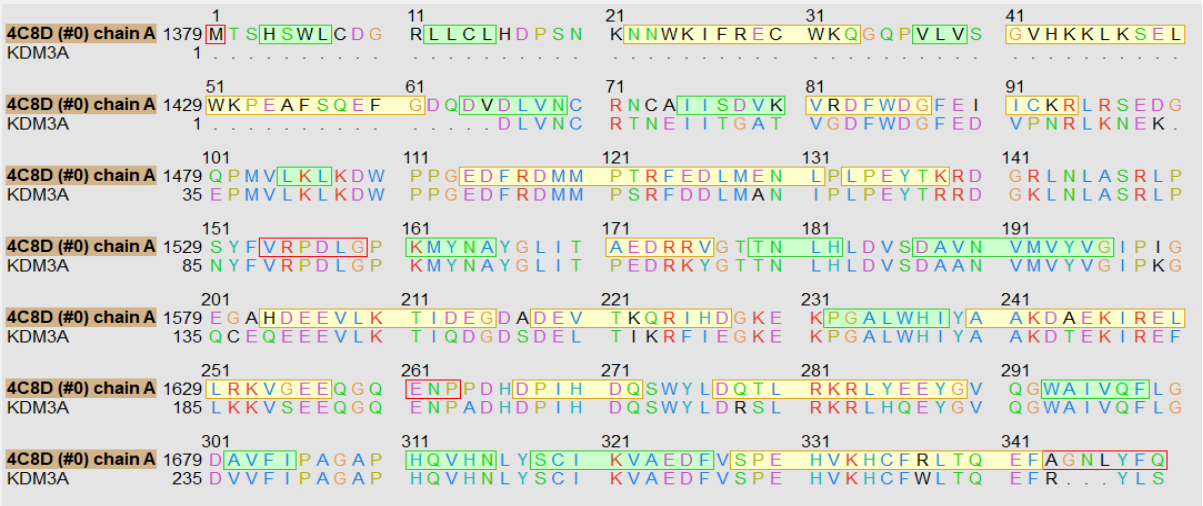

Figure 1. the sequence alignment of the KDM3B and KDM3A.

The crystal structure of the JmjC domain of KDM3B (PDB ID: 4C8D) was accessed from the protein databank archives using the website , www.RCSB.org, and was used by Modeller to construct a low resolution structure of the JmjC domain of KDM3A using Modeller-homology modelling and Modeller Loops refinements schemes to insure optimal modeling of the structured

regions (α helices and α sheets) and the loosely structured loops regions. To determine the degree of success of homology docking the final optimized KDM3A-JmjC domain structure was compared to the structure of the JmjC domain of KDM3B. This was done using the Match and Align protocol from the Structure Comparison menu of Chimera, which is done by aligning the corresponding backbone α-carbons of both structures. The aligned structures are shown in Figure 2, with the KDM3B structure is colored pink and the homology modeled structure of KDM3A is colored blue. Qualitatively, the two structures appear to align fairly closely. A more quantitative assessment of the alignment of both structures is done by determining the root mean square deviation (RMSD) of the positions of the α carbons in both structures. Normally, if the template and the target protein have amino acid sequence homology > 70%, a homology modelling calculation that results in a RMSD value of 0.5Å is considered successful. The RMSD value obtained from the comparison of the KDM3B structure and the homology modelling generated structure of KDM3A was found to be 0.303Å which gave us added confidence in the quality of the generated structure.

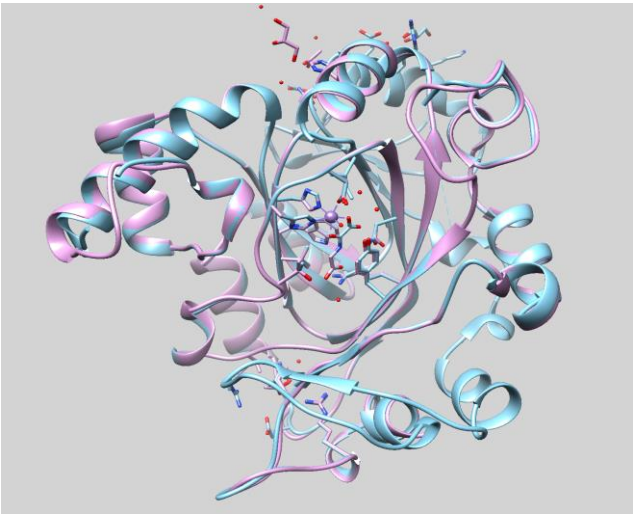

Figure 2: The alignment of the backbone α-carbons of the structure of the JmjC domain of KDM3B and the generated structure of the JmjC domain of KDM3A.

To further assess, the validity of the generated structure, we examined the coordination bonds between the manganese ion, the cofactor OGA and crucial amino acids in the catalytic domain of

JmjC of KDM3A as compared to those found in the crystal structure of the JmjC domain of KDM3B. As shown in Figure 3, both structures show manganese with a six-coordination number. In the KMD3B structure (panel A), two coordination bonds are between the metal ion and a carbonyl oxygen and a hydroxyl oxygen of OGA. The metal ion formed two coordination bonds to the imidazole nitrogens of histidines 1560 and 1689, and one coordination bond with the carboxylic oxygen on aspartate 1562. The metal ion also formed two coordination bonds with the carboxylic oxygen and the amide oxygen of NGO. The sixth coordination bond is with a nearby water molecule. Similar coordination is found for the manganese ion in the generated structure of the JmjC domain KDM3A where the amino acids involved in the coordination bonds are found to be histone 100 and 229 and asparate102. The two coordination bonds to the carboxylic oxygen and the amide oxygen are also present, and the sixth bond was also found to be between the metal ion and a nearby water molecule. Given this level of similarity between the two structures, we strongly believe that there is enough precision in the molecular and structural details in the JmjC domain of KDM3A to proceed with the virtual screening of small molecule databases in search of a potential therapeutic agent that specifically binds the main components of the catalytic domain of KDM3A and can potentially lead to inhibiting the enzyme from carrying out its deleterious functions in cancer cells.

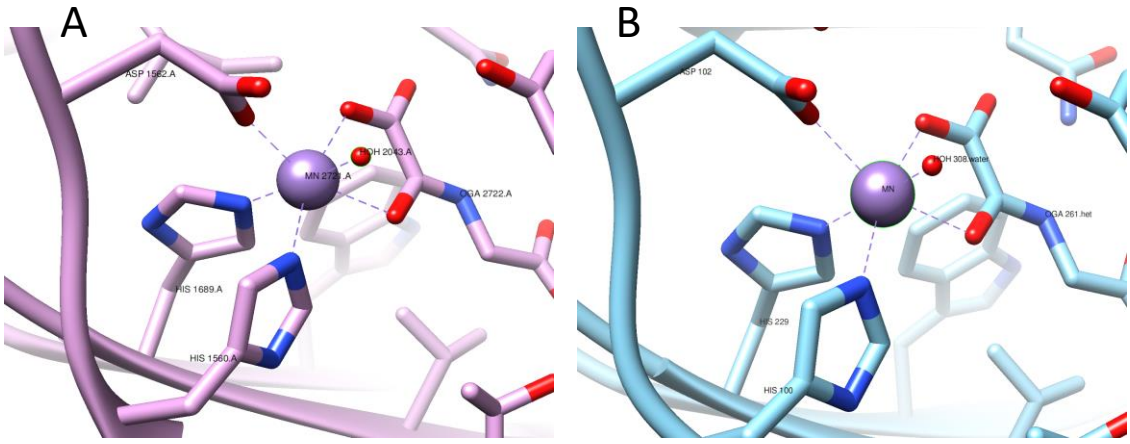

Figure 3: Coordination complex formation between the manganese metal ion and OGA and amino acids and the catalytic domain of KDM3B (panel A) and KDM3A (panel B).

The Search for Potential Inhibitors for H3-lysine 9 demethylase, KDM3A

The virtual screening of three databases of small molecule structures containing hundreds of thousands of ligands were used to find candidates of potential inhibitors of the catalytic domain of KDM3A using the open source VINA software. In our search, we limited the screening by loosely specifying the general coordinates the search space designating the important region of the catalytic pocket of the enzyme, which includes the region occupied by the manganese ion and the OGA cofactor, and the amino acids that are involved in their interactions. These coordinates along with the PDB file of the generated KDM3A were inputted in a VINA data window and the calculations were launched using aforementioned databases of small molecules. Once the docking results were obtained, we conducted further screening of the large body of data using a cutoff value of the binding free energy parameter, ΔG_{bind}, of -8.5 kcal/mol. This screening led to the discovery of 6 ligands that have ΔG_{bind} ranging from -8.8 kcal/mol to -12.3 kcal/mol. The names of these compounds, chemical formulae and their binding free energies are listed in Table1. These binding energies indicate strong interactions with the JmjC domain of KDM3A with binding constants (K_{eq}) ranging from 3×10^6 to 1×10^9. To assess the significance of these binding affinities and whether or not these compounds can be potential inhibitors, we examine the structures of these ligands in an effort to compare their functional groups to functional groups that are common in established anti-cancer therapeutic agents. Among the most prevalent atoms that one finds in compounds that have been proven to be effective anti-cancer drugs is nitrogen. Therefore, nitrogen containing functional groups are looked upon favorably among drug discovery researchers. Below is a list of some of the most prevalent nitrogen containing functional groups that are frequently found in therapeutic agents. Specifically, indoles, purines, and, amino or imino containing rings are very common groups to find in drugs. The potential for H-bonding is also an important factor to keep in mind while examining molecules as potential therapeutic agents. Hydroxyl groups, along with nitrogen containing groups will increase the likelihood of discovering a water-soluble drug that can ultimately reach its biological target.

Examining the structural formulae of the compounds we selected (see Figure 4), it is noted that most of these compounds are rich in nitrogen and have strong potentials for H-bonding to amino acids in the enzyme's catalytic domain. Another important observation is that most of these compounds contain significant amount of flexibility as determined by the large number of torsions that the structures contain. That is an important point to consider since as potential drugs, accessing a target site in a usually very bulky biomolecule (protein or nucleic acid) is essential and often requires the flexibility of a potential drug that allows it bend in order to fit within a specific binding pocket.

Table 1. Names, chemical formulae and the binding free energy values of top Vina docking Picks.

Names of Possible KDM3A inhibitors	Chemical formula	ΔG values (kcal/mol)	K_{eq}
4-(2-Amino-6-naphthalen-1-yl-pyrimidin-4-yl)-6-naphthalen-1-yl-pyrimidin-2-amine **(compound 1)**	$C_{28}H_{20}N_6$	-11.1	$1x10^8$
4-[2-Amino-6-(2-methyl-1H-indol-3-yl)pyrimidin-4-yl]-6-(2-methyl-1H-indol-3-yl)pyrimidin-2-amine **(compound 2)**	$C_{26}H_{22}N_8$	-10	$2x10^7$
1-[(7-Oxobenzo[a]phenalen-3-yl)amino]anthracene-9,10-dione **(compound 3)**	$C_{31}H_{17}NO_3$	-12.2	$9x10^8$
Indanthren Corinth RK **(compound 4)**	$C_{35}H_{20}N_2O_5$	-12.1	$7x10^8$
Vat Yellow 20 **(compound 5)**	$C_{30}H_{15}N_3O_4$	-11.1	$1x10^8$

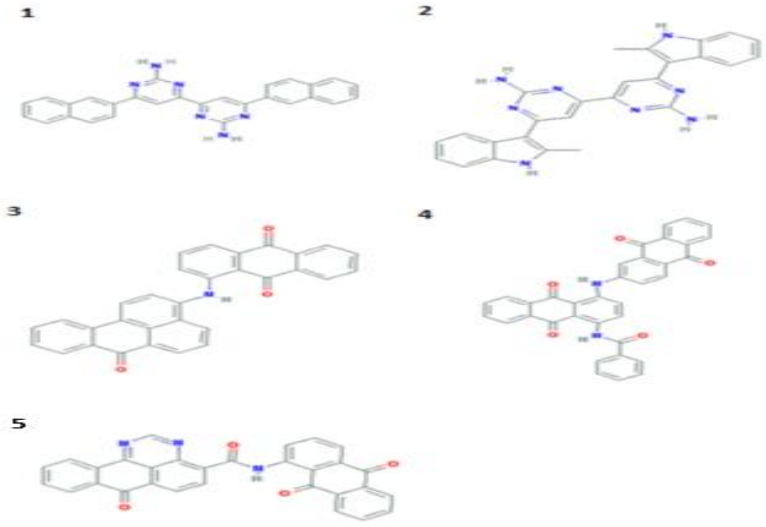

Figure 4: The structural formulae of the 5 ligands selected using VINA virtual screening calculations.

KDM3A and Ligands Interactions and Docking Result Through Autodock vina and Pyrx:

The six compounds chosen from the three large sets of ligand libraries were selected based on their high binding affinities as assessed from by their ΔG_{bind} to the JmjC domain of KDM3A determined by Autodock VINA calculations. The calculation of ΔG_{bind} takes into account, electrostatic interactions or interactions due to partial charges (ΔG_{ele}), Van der Waals interaction or dispersive and repulsive interactions (ΔG_{VW}), hydrogen-bonding interactions (ΔG_{HB}), entropic interactions due to ligand flexibility (ΔG_{ent}), and disolvation interactions which assumes the protein as a solvent, (ΔG_{solv}), Our objective is to find a ligand that binds to the region of the JmjC domain of KDM3A that is most crucial to the catalysis of the demythelation of lysine, and the stronger the interaction at that position the better the action of the potential inhibitor. The search algorithm used by Autodock Vina (the Lamarchean Genetic Algorithm) treats each site outlined by the Grid box as a gene, as the ligand moves to a new position, that is called a mutation and the value of the ΔG_{bind} of the mutated position if higher will lead to a lower energy complex that will ultimately converge into the output that Vina provides. In some cases, the binding site of the ligand will be too far from the desired binding site and the structure is rendered useless. To determine which of the top six complexes chosen exclusively based on the value of ΔG_{bind} satisfy our binding location, we examined the micro-environment of the complexation between each ligand and the JmjC KDM3A structure. Figures 5-10 show a close up of the binding sites of compounds 1-5 in the JmjC domain of KDM3A, where the ring structured potential inhibitors are shown in grey (with the blue colored atom representing nitrogen and the red colored atom representing oxygen). The purple sphere represents the manganese ion in the protein's catalytic domain, and the branched small molecule (also colored grey with red oxygens) representing the OGA cofactor. Examination of the results shown in Figures 5-10 indicates that each of the potential inhibitors forms at least one hydrogen-bond (shown as a blue line) and as much as four hydrogen-bonds to amino acids that are directly involved in the catalytic demethylation reaction, or to OGA, a crucial cofactor that mediates the removal of methyl groups form lysine. A

summary of the observed interactions between compounds 1-5 and the catalytic domain residues and cofactor is shown in Table 2. By analyzing the data shown in the figures and summarized in the corresponding table gave us to the confidence of predicting that compounds 1 and 2 would have the most significant effect as inhibitors of the catalytic activity of KDM3A as they both appear to form hydrogen-bonds to amino acids that are involved in the manganese coordination complex and to OGA, an significant cofactor that the enzyme needs for the catalysis of lysine demethylation. We predict that all six molecules will be shown to inhibit the catalytic activity of KDM3A simply due to the fact that they all bind within the catalytic domain of KDM3A which would provide a steric factor that should adversely affect the catalysis process.

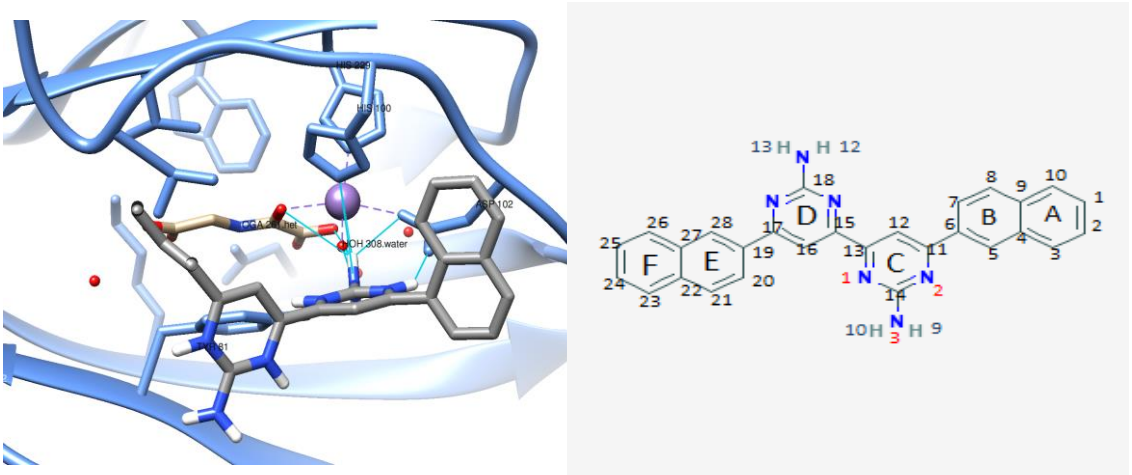

Figure 5. KDM3A and compound 1 interactions. 4 hydrogen bonds appeared in blue lines between imidazole nitrogen of his100 and H10, carbonyl oxygen of OGA and H9, carbonyl oxygen of asp102 and H9, and hydroxyl oxygen of asp102 and H8 of compound 1.

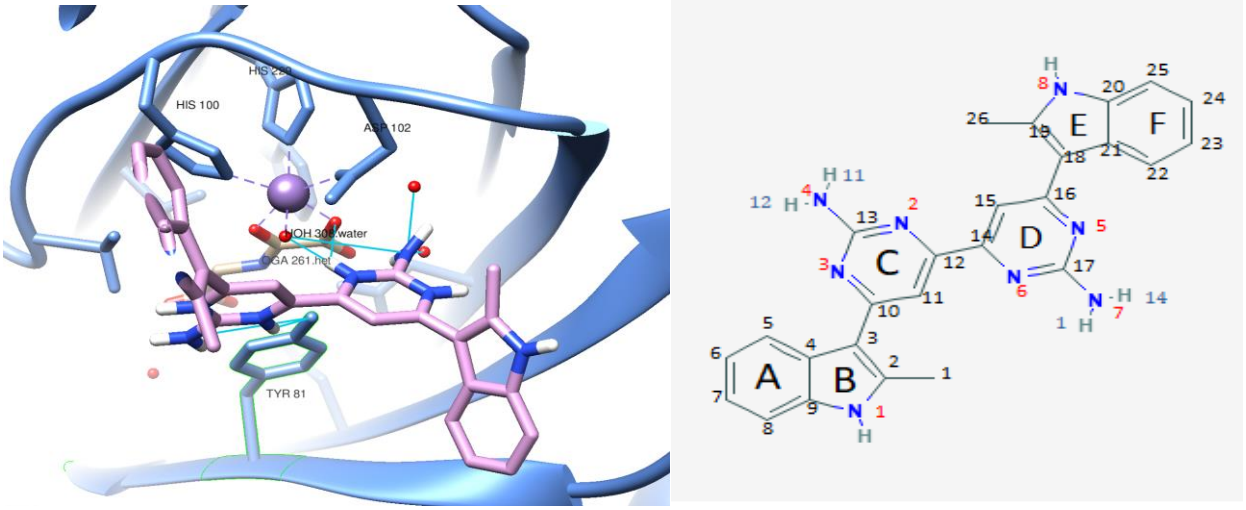

Figure 6. KDM3A and compound 2 interactions. Three hydrogen bonds represented in blue lines bind hydroxyl oxygen of OGA with H10 of compound 1 and hydroxyl oxygen of tyrosine 81 with H12 and H14 of compound 1.

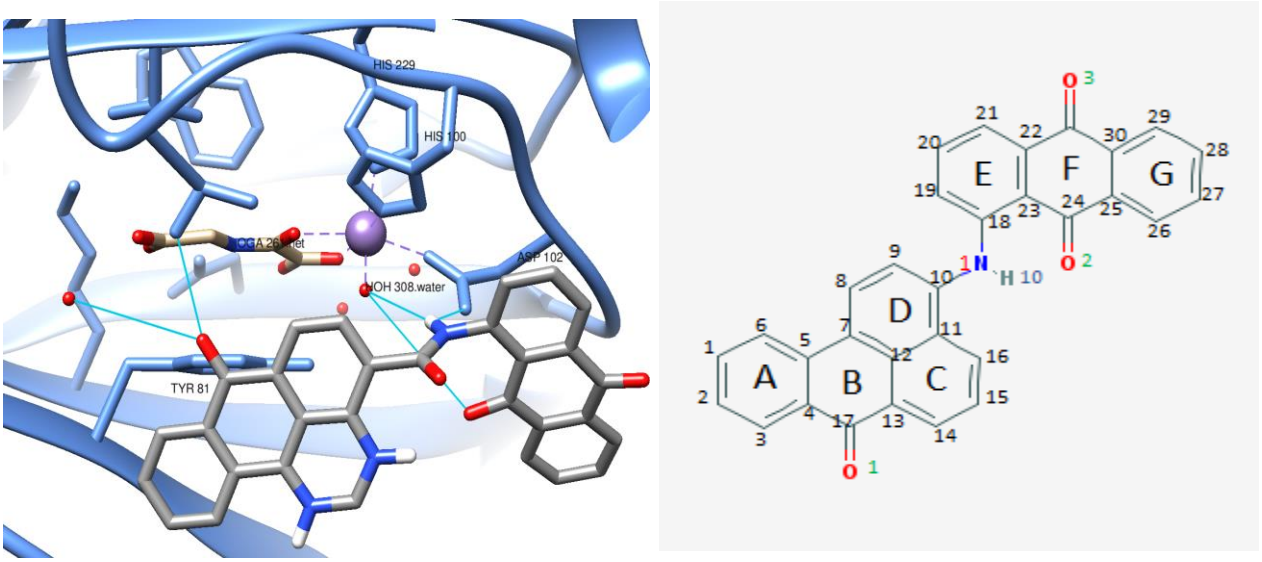

Figure 7. KDM3A and compound 3 interactions. Two hydrogen bonds appeared in blue lines between hydroxyl oxygen of asp102 and H10, and between methyl hydrogen of valine and O1 of compound 3.

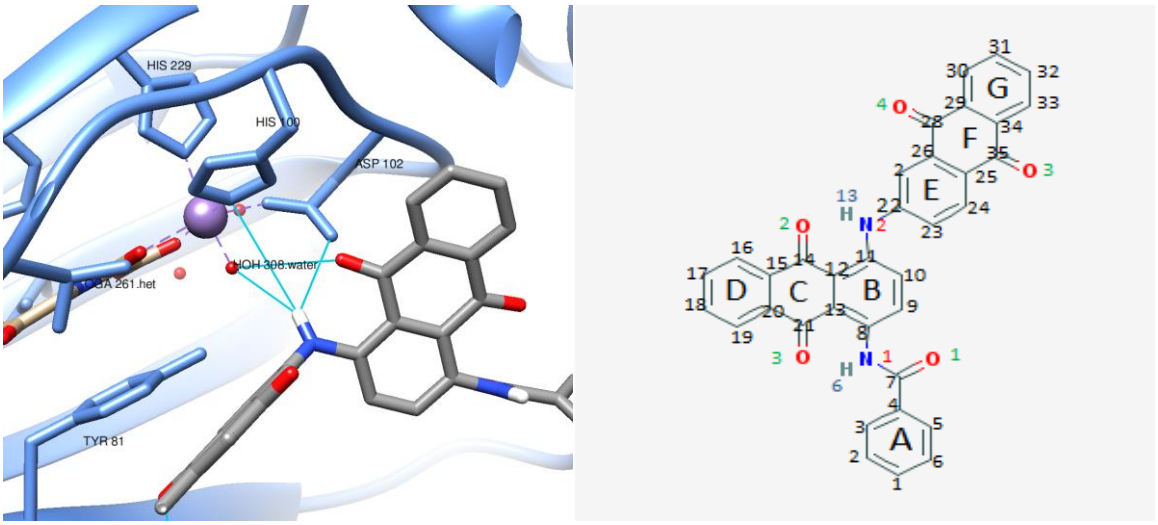

Figure 8. KDM3A and compound 4 interactions. In blue lines appear one hydrogen bond between imidazole N of his100 and H13, and another hydrogen bond between hydroxyl oxygen of asp102 and H13 of compound 4.

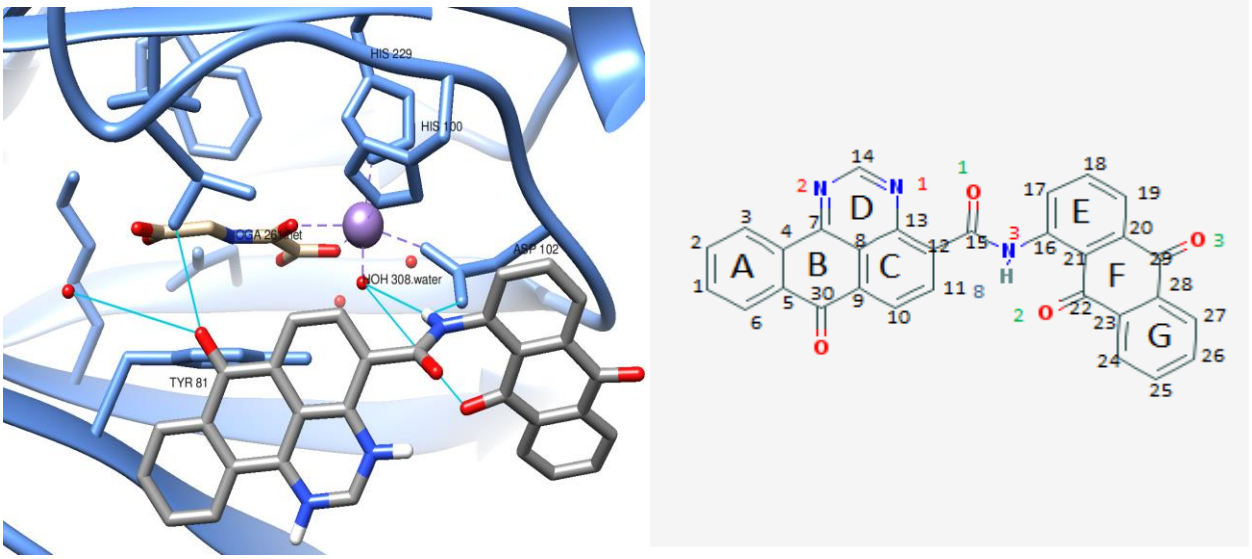

Figure 9. KDM3A and compound 5 interactions. Two hydrogen bonds represented in blue lines between hydroxyl O of asp102 and H8 and between methyl H of valine and O4 of compound 6.

Table 2. Summary of interactions between compounds 1-5, catalytic domain residues and cofactor GOA.

Compound #	Number of H bonds	Active domain amino acids of KDM3A	Compound atoms	Cofactor (OGA)
1	4	Imidazole N of his100 Carbonyl O of asp102 Hydroxyl O of asp102 -	H10 H9 H8 H9	- - - Carbonyl O
2	3	Hydroxyl O of tyr81 -	H13 and H14 H10	- Hydroxyl O
3	2	Hydroxyl O of asp102 Methyl H of valine (val).	H10 O1	- -
4	2	Hydroxyl H of asp102 Imidazole N of his100	H13 H13	-
5	2	Methyl H of valine Hydroxyl O of asp102	O4 H8	- -

Chapter IV

Discussion

The demethylation of lysine 9 of H3 is catalyzed by the JmiC domain of KDM3A. This demethylation process is mediated by α-ketoglutarate (α-KG) and Iron (II) cofactors. The general mechanism of the JmjC domain-containing demethylases as outlined in Figure 10 which begins with the reaction of Fe(II) with molecular oxygen (O_2) leading to the generation of a superoxide radical form of Fe(II). This radical attacks the second carbon atom of the alpha ketoglutarate and produces succinate and CO_2. This decarboxylation process of the cofactor leads to the formation of Fe(VI)-oxo intermediate which is reduced by the hydrogen of the methylated lysine and forms hydroxylated carbinolamine which spontaneously decomposes to forms formaldehyde and the demethylated lysine regenerating the Fe (II) active center.

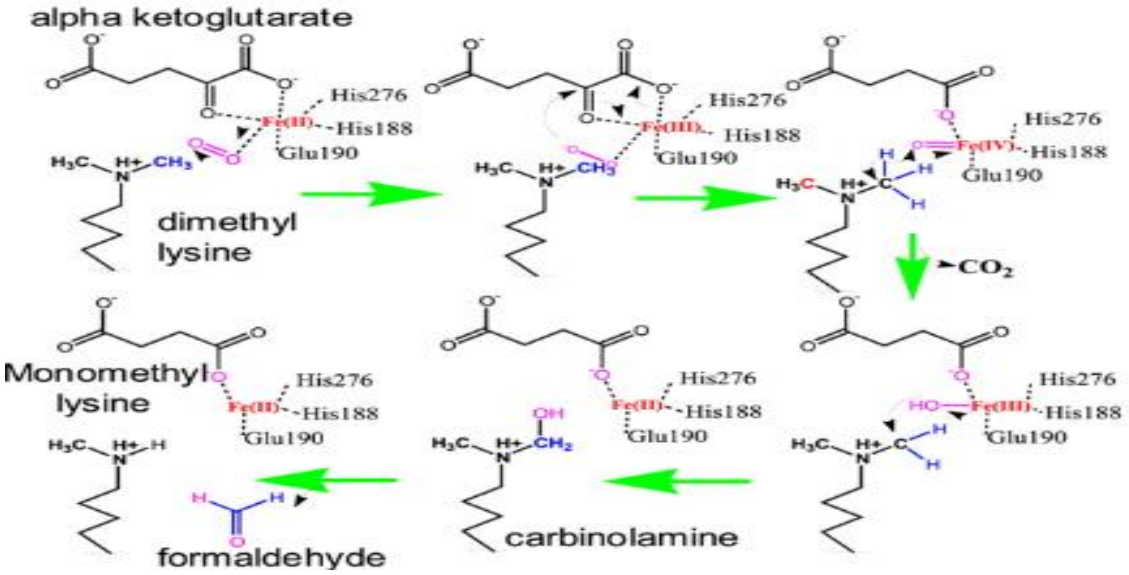

Figure 10. The mechanism of JmjC domain in demethylases [4].

The goal of this research is to disrupt this process using a potential inhibitor that can interfere with one of the multiple steps outlined in this mechanism. The homology modeling structure that we generated for the JmjC domain of KDM3A has all the components necessary for the catalysis of the demethylation of lysine with the exception of the use of Mn(II) as a less reactive analogue of Fe(II) and N-oxalylglycine (OGA), a molecule that is an isosteric analogue of (α-KG). Nevertheless, the geometry of the active site is identical to the one discussed in this mechanism. And as can be seen from Figure 3 the manganese metal ion forms six coordination bonds, two belong to the binding of two histidines in the catalytic domain, one to an Aspartate residue in the catalytic domain, two to two oxygens on OGA that have the same configurations as the coordinated oxygens in α-KG and the sixth coordination bond is to molecular oxygen. Examination of the structures obtained for the complexes between the potential inhibitors and the JmjC domain of KDM3A (Figure 4-9) shows that the five compounds we chose from a pool of small compound libraries form 1-4 hydrogen bonds to OGA and/or to the amino acids that are involved in the Mn-coordination bond formation. We predict that these hydrogen bonds would most certainly alter the electronic properties of OGA and the amino acids involved in binding the metal ion which would lead to the disruption of the coordination complex, a necessary component in the catalytic process of the enzyme. Another point to consider is the conformational flexibility afforded to the enzyme in its natural environment as compared to the rigid constraints imposed on it through the calculations which restrict the accessibility of the molecules to potential binding partners in the catalytic region of the enzyme. This idea is posed because the structures generated in this work were obtained by keeping the protein structure rigid and only allowing flexibility to the small molecules. This is done because with added flexibility to large proteins, the very large number of degrees of freedom added will significantly increase the time it would take to generate the structures needed to complete this thesis. Nevertheless, conducting the calculations with the parameters used in this thesis still allowed us to generate structures that are valuable and informative enough to allow us to proceed to the next step in our drug discovery.

References

Maze, I., Noh, K. M., Soshnev, A. A., & Allis, C. D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature Reviews Genetics*, *15*(4), 259.

Wood, A., & Shilatifard, A. (2004). Posttranslational modifications of histones by methylation. In *Advances in protein chemistry* (Vol. 67, pp. 201-222). Academic Press.

Greer, E. L., & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nature Reviews Genetics*, *13*(5), 343.

Anand, R., & Marmorstein, R. (2007). Structure and mechanism of lysine-specific demethylase enzymes. *Journal of Biological Chemistry*, *282*(49), 35425-35429.

Qin, L., Xu, Y., Yu, X., Toneff, M. J., Li, D., Liao, L., ... & Xu, J. (2017). The histone demethylase Kdm3a is required for normal epithelial proliferation, ductal elongation and tumor growth in the mouse mammary gland. *Oncotarget*, *8*(49), 84761.

Ramadoss, S., Guo, G., & Wang, C. Y. (2017). Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. *Oncogene*, *36*(1), 47.

Sharma, B. K. (2017). Synthetic and Natural Compounds as Anti-Cancer Agents–A Review. *Asian Journal of Research in Chemistry*, *10*(5), 699-707.

Zayed, H., & Petersen, I. (2018). Stem cell transcription factor SOX2 in synovial sarcoma and other soft tissue tumors. *Pathology-Research and Practice*, *214*(7), 1000-1007.

Yang, Q., Yang, Y., Zhou, N., Tang, K., Lau, W. B., Lau, B., ... & Wang, X. (2018). Epigenetics in ovarian cancer: premise, properties, and perspectives. *Molecular cancer*, *17*(1), 109.

Cho, H. S., Toyokawa, G., Daigo, Y., Hayami, S., Masuda, K., Ikawa, N., ... & Kelly, J. D. (2012). The JmjC domain-containing histone demethylase KDM3A is a positive regulator of the G1/S transition in cancer cells via transcriptional regulation of the HOXA1 gene. *International journal of cancer*, *131*(3), E179-E189.

Staberg, M., Rasmussen, R. D., Michaelsen, S. R., Pedersen, H., Jensen, K. E., Villingshøj, M., ... & Hamerlik, P. (2018). Targeting glioma stem-like cell survival and chemoresistance through inhibition of lysine-specific histone demethylase KDM2B. *Molecular oncology*, *12*(3), 406-420.

D'Oto, A., Tian, Q. W., Davidoff, A. M., & Yang, J. (2016). Histone demethylases and their roles in cancer epigenetics. *Journal of medical oncology and therapeutics*, *1*(2), 34.

Lin, H., Li, Q., Li, Q., Zhu, J., Gu, K., Jiang, X., ... & Sun, H. (2018). Small molecule KDM4s inhibitors as anti-cancer agents. *Journal of enzyme inhibition and medicinal chemistry*, *33*(1), 777-793.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, *27*(1), 14-25.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Šali, A., Fiser, A., Sánchez, R., Martı-Renom, M. A., Yerkovich, B., Badretdinov, A., ... & Feyfant, E. (2001). Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research Mission Bay Genentech Hall 600 16th Street, Suite N472D University of California, San Francisco.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, *31*(2), 455-461.

Zhang, L. J. (2009). 2009 IEEE Congress on Services (SERVICES-1 2009).

Krishnan, S., Stearn, B., Bhatia, K., Baldridge, K. K., Li, W. W., & Arzberger, P. (2006, September). Opal: Simpleweb services wrappers for scientific applications. In *Web Services, 2006. ICWS'06. International Conference on* (pp. 823-832). IEEE.

Small-molecule library screening by docking with PyRx. *Dallakyan S, Olson AJ Methods Mol Biol. 2015; 1263():243-50.*

Huey, R., & Morris, G. M. (2003). AutoDock tools. *La Jolla, CA, USA: The Scripps Research Institute*.

Krieger, E., Koraimann, G., & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins: Structure, Function, and Bioinformatics*, *47*(3), 393-402.

Van Durme, J., Delgado, J., Stricher, F., Serrano, L., Schymkowitz, J., & Rousseau, F. (2011). A graphical interface for the FoldX forcefield. *Bioinformatics*, *27*(12), 1711-1712.