STUDIES OF HEAT SHOCK PROTEIN A9 (HSPA9/MORTALIN/GRP75) SUBSTRATE BINDING DOMAIN AND ANTICANCER FLEX-HETS INTERACTIONS USING NUCLEAR MAGNETIC RESONANCE AND MOLECULAR MECHANICS TOOLS

By

MARYAM MASHAYEKHI

Bachelor of Science in Physics
University of Arak
Arak, Iran
2008

Master of Science in Photonics
Iran University of Science and Technology
Tehran, Iran
2011

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2019

STUDIES OF HEAT SHOCK PROTEIN A9 (HSPA9/MORTALIN/GRP75) SUBSTRATE BINDING DOMAIN AND ANTICANCER FLEX-HETS INTERACTIONS USING NUCLEAR MAGNETIC RESONANCE AND MOLECULAR MECHANICS TOOLS

Dissertation Approved:

Dr. Donghua H. Zhou

Dissertation Adviser

Dr. Mario Borunda

Dr. Gil Summy

Dr. Junpeng Deng

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude toward:

My advisor, Dr. Donghua Zhou, who shared his experience and knowledge with me. His support, kindness, and enthusiasm has always been an inspiration.

All my committee members, Drs. Junpeng Deng, Mario Borunda, and Gil Summy for their valuable time and insightful comments. Specially Dr. Deng and his group who generously provided me with the technical support for some of my experiments.

Dr. Benbrook and her group in Stephenson Cancer Center, OUHSC for biological tests and Drs. Berlin and Bunce and their groups in Chemistry Department, OSU for compound synthesis part of this study.

My colleagues, Penghui Lin, Amanda Taylor, Lian Duan, Hem Moktan, and Dipendra Bhandari who made the journey enjoyable.

My family, especially Mom, Dad, and My brothers for their unconditional love and support.

Name: MARYAM MASHAYEKHI

Date of Degree: MAY, 2019

Title of Study: STUDIES OF HEAT SHOCK PROTEIN A9 (HSPA9/MORTALIN/GRP75) SUBSTRATE BINDING DOMAIN AND ANTICANCER FLEX-HETS INTERACTIONS USING NUCLEAR MAGNETIC RESONANCE AND MOLECULAR MECHANICS TOOLS

Major Field: PHOTONICS

Abstract: Ovarian cancer is the most life-threatening of all gynecologic malignancies with survival rate of about 30% that often has few symptoms prior to late stage diagnosis. Ovarian cancer causes 5 percent of cancer deaths among US women. In 2018 approximately 22,240 new cases of ovarian cancer were diagnosed in the United States and 14,070 died. Cisplatin and taxane drugs usually have unacceptable side effects. SHetA2 is an anticancer compound that has exhibited strong activity against ovarian cancer. It has discriminatory activity against malignant cells versus normal cells and has very low toxicity. SHetA2 is a sulfur containing Flex-Het with the highest anti-cancer activity among the tested Flex-Hets. SHetA2 has been shown effective against all of over 60 cancer cell lines in the US National Cancer Institute. SHetA2 interacts with heat shock protein HSPA9/mortalin. Mortalin was shown to associate with the tumor protein p53 in a concentration-dependent manner and leads to inactivation of p53 apoptotic function in stressed cells. Since p53 has a major role in cell cycle arrest and apoptosis, the loss of p53 function leads to immortalization of cells. Protein p53 is nonfunctional in some tumors by cytoplasmic sequestration of its binding proteins, such as Bcl-2, hsp70/mortalin, Parc, PML, and cytoskeleton proteins. One pathway for the compounds to inhibit the growth of cancer cells in vitro is by displacing p53 from mortalin. Then, p53 migrates to the nucleus and initiates apoptosis. Binding of SHetA2 to mortalin interrupts p53-mortalin interaction, and releases protein p53 to nucleus where it initiates apoptosis. Study of the interactions between mortalin and SHetA2 as well as the interactions between mortalin and SHetA2 analogs is the subject of this study. We used solution state NMR spectroscopy technique (chemical perturbation method) and molecular mechanics approaches (Autodock tools and molecular dynamics via GROMACS) to study protein-ligand interactions. We investigate the binding site and binding affinity using these techniques. Our findings helped designing improved SHetA2 analogs that were tested for their biological activities.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Nuclear Magnetic Resonance (NMR) spectroscopy addresses the behavior of nuclear spin angular momentum, which is an intrinsic quantum mechanical property, in the presence of magnetic fields. To help visualizing the concept, this behavior is described as precession of a magnetic dipole. In the absence of the rf radiation, the spin angular momentum precesses in the external magnetic field. The rf radiation with proper frequency is absorbed by the spin, which makes a transition to a higher energy state. The re-emitted electromagnetic radiation from the spin while it is relaxing back to the original state forms a detectable Nuclear Magnetic Resonance (NMR) signal (Free Induction Decay, FID), which is a characteristic of the static magnetic field, the chemical environment of the nuclei, and magnetic properties of the nuclei. NMR spectroscopy has wide applications in physics, biology, chemistry, etc. Although X-ray crystallography is the dominant method in protein structure determination, NMR spectroscopy has made drastic improvements and is used widely in structure determination of biomolecules. NMR spectroscopy provides information on a wide range of time scale dynamics thus, it is powerful in studying protein-protein and protein-ligand interactions, small molecule screenings, and optimization in drug discovery and development. Sample preparation is flexible and relatively easy in NMR spectroscopy.

When studying samples in solution, the rapid tumbling of the molecules will give rise to a high resolution NMR spectra. In these cases, the orientationally dependent anisotropic interactions are averaged out on NMR experiment time scale. However, in samples where anisotropic interactions cannot be neglected, such as large proteins or solid samples, considerations should be taken into account. In such cases Solid-State NMR spectroscopy (SSNMR) is used to overcome the loss of resolution in the spectrum of solid samples, which is due to line broadening effect of anisotropic interactions. SSNMR spectroscopy is a very useful technique in studying membrane proteins, which are not naturally soluble in aqueous solution and resistant to crystallization. In this study, we focus of solution NMR spectroscopy and we apply it to determine protein-ligand interactions.

We studied the anti-cancer molecule, SHetA2, and its interaction with the heat shock protein 70 family member, mortalin, using computational and experimental methods. Our findings inspired us to design several SHetA2 analogs in attempt to optimize properties of SHetA2. Several oxygen and nitrogen containing analogs have been tested *in vitro* and by computational tools and showed promising results.

Chapter II. NMR Spectroscopy and its Application in Protein-Ligand Binding Studies

In chapter II, we first review the basics of NMR Spectroscopy. We briefly review other available methods in studying the protein-ligand complexes. NMR spectroscopy applications in studying protein-ligand interactions is reviewed next.

Chapter III. Molecular Modeling Studies of Protein-Ligand Complexes

We review the application of molecular modeling techniques in studying the protein-ligand interactions in chapter III, with the emphasize on Molecular Docking tool, Autodock, and Molecular Dynamic tool, GROMACS.

Chapter IV. SHetA2 and mortalin Interaction

In chapter IV we review the evolution of heteroarotinoids, and their biological activities with the emphasis on SHeA2 and its chemopreventive role. Mortalin, a receptor for SHetA2 ligand, p53 protein, and p53-mortalin interaction is also reviewed.

Chapter V. NMR Studies of the Interaction between mortalin and SHetA2

This chapter includes our solution state NMR studies on the substrate binding domain of mortalin in complex with SHetA2 which helped identifying the binding site of SHetA2 on mortalin. The residues forming the hydrophobic binding pocket have been identified from the chemical shift perturbation data.

Chapter VI. SHetA2 Molecular Modeling Studies of Mortalin-SHetA2 using GROMACS

Chapter VI covers molecular docking (using Autodock) and molecular dynamic studies (using GROMACS) of SHetA2-mortalin and mutant V482F. Binding configuration and binding free energy has been calculated. We showed that residue V482 participates in the interaction of mortalin with SHetA2. V482F reduced the available space for binding of SHetA2 and changed the configuration of the ligand in the binding pocket.

Chapter VII. Oxygen-Containing Analogs of the Flex-Het anticancer Agent SHetA2

As our results from molecular mechanics studies suggests, hydrophobicity of the ligand is an important factor in its binding performance. Thus, we propose several analogs for SHetA2 and investigated the effect of hydrophobicity on their efficacy and potency. Chapter VII contains our published work on the activity of these oxygen-containing analogs of the SHetA2.

Chapter VIII. Nitrogen-Containing Analogs of the Flex-Het anticancer Agent SHetA2

Including a nitrogen atom in the structure of SHetA2 is proposed with the goal to increase the binding capacity of the molecule to its protein-binding partners. Chapter VIII is our work on nitrogen containing Flex-Hets and evaluating their activities. Several analogs have been tested and their free binding affinity, potency and efficacy values have been found and compared with those of SHetA2

CHAPTER II

NMR SPECTROSCOPY AND ITS APPLICATION IN PROTEIN-LIGAND BINDING STUDIES

2.1. Introduction

The first applications of NMR spectroscopy trace back to observing the resonance spectrum of paraffin (solid phase) by Bloch et al. [1] and water (liquid phase) by Purcell et al. [2] in 1946. Bloch and Purcell shared the Nobel Prize in Physics for discovery of NMR spectroscopy in 1952. Varian built the first commercial NMR spectrometer and the first NMR based 3D structure of a 57 residue protein was identified in 1985 by Wüthrich [3]. Since its discovery, improvements in software, hardware, and NMR techniques has broadened the applications of NMR spectroscopy. Some major breakthroughs include development of Fourier Transform NMR spectroscopy (by Ernst and Anderson), development of multidimensional NMR spectroscopy (by Jeener), and using solvent signal suppression that allows the protein spectra to be acquired in solution. A more recent improvement is development of the cryogenic probes with much higher sensitivities and higher signal to noise ratio that has lowered the protein concentration requirement to 0.5 mM.

Improvements in powerful magnets with better field strength, homogeneity, and stability, together with the use of pulsed rf excitations instead of the traditional continuous wave scan, isotopic labelling techniques, and line narrowing techniques in solid state NMR, etc. have made NMR spectroscopy a powerful tool to study the structures, dynamics, and real-time reactions of

complicated macromolecules.

NMR spectroscopy allows the studying of time-dependent chemical phenomena e.g. reaction kinetics and intramolecular dynamics. It provides information on proteins in more native aqueous environment, where molecules can tumble freely. Thus, NMR spectroscopy is the premier method in studying flexible proteins and intrinsically unfolded proteins. It can be used to study protein-folding intermediates and conformation changes. NMR spectroscopy can also be applied to variety of solid form samples such as powder, frozen samples, micro-crystals, proteoliposome, or gels. Protein-protein and protein-ligand interactions can be studied via NMR spectroscopy. Which makes it a powerful tool for drug screening and optimization studies. Very weak interactions (K_D~1 mM), where protein-protein or protein-ligand complexes cannot even be isolated, can be detected by solution NMR. Not only the overall molecular tumbling but also the internal motions of the structural segments can be characterized by solution NMR spectroscopy. NMR spectroscopy obtains high-resolution structures. In solution state the directionally dependent anisotropic interactions are averaged and can be neglected on NMR experiment time scale, the rapid tumbling of the molecules will give rise to a high resolution NMR spectra. However, in large or solid samples where anisotropic interactions cannot be averaged out, Solid State NMR spectroscopy (SSNMR) is used to overcome the loss of resolution in the spectrum due to line broadening effect of anisotropic interactions. SSNMR is a very useful technique in studying membrane proteins, which are not naturally soluble in aqueous solution and resist to crystallization. In this study, we focus of solution state NMR spectroscopy and utilize it to investigate the interaction between protein and ligand.

Protein-ligand binding is important in signal transmission, cellular metabolism, etc. Study of protein-ligand interactions is essential for understanding the regulation of biological functions and for designing more potent bioactive molecules that can inhibit proteins, modulate their function, and serve as drug candidates. Some examples of protein-ligand interactions with

biological functions are interaction of hormone receptors with hormones, enzyme inhibitor interactions, interaction of proteins with nucleic acid sequences, and antigen interaction with receptors. One important step in identifying and optimizing a drug candidate is identifying compounds with reasonable binding activities to specific protein receptors. Initially a very large number of compounds is screened and the *leads* are identified. NMR spectroscopy is a powerful tool in identifying the binding activity of leads/ligands with protein receptors to assess and identify drug candidates.

In this chapter, we first review the basics NMR Spectroscopy. Then other available methods in studying protein-ligand complexes are briefly reviewed. Followed by introducing NMR spectroscopy applications in identifying protein-ligand interactions.

2.2. Other Available Methods for Studying Protein-Ligand Interactions

X-ray crystallography, NMR spectroscopy and 3D electron microscopy (3DEM) are the three main techniques used to determine the three-dimensional structures of macromolecules at atomic resolution. The first one-dimensional protein spectrum was recorded about four decades ago. To date (January 2019) 12485 protein structures has been release from NMR experiments. While 132182 structures have been released from X-ray Crystallography experiments and 2746 structures from 3D electron microscopy [https://www.rcsb.org/].

In X-ray crystallography, purified and crystalized protein is subjected to X-ray beams and diffracts the beam into characteristic diffraction pattern. This diffraction pattern is then analyzed to determine the distribution of electrons in the protein sample. From this map of electron density, the location of each atom will be determined. The quality of the crystals is the determining factor in accuracy of the predicted atomic structure. While X-ray crystallography provides very detailed atomic information and is suitable for rigid proteins, crystallization is a difficult process and

makes the study of the flexible proteins difficult. Flexible parts can be invisible in electron density maps.

Femtosecond crystallography using *X-ray Free Electron Lasers (XFEL)* is revolutionizing the X-ray crystallography. In this method a stream of crystals (nanometers to micrometers size) passes through a beam of extremely short (femtosecond long) and intense pulses (up to 10^{13} photons) created by the XFEL. Each X-ray pulse produces a diffraction pattern from a crystal. Tens to hundreds of thousands of snapshot diffraction patterns will be created. This powerful method allows study of molecular processes that happen over very short time, e.g. absorption of light by biological chromophores. Conventional sources of beams can break the bonds and change the structure during exposure. XFEL however outruns the radiation damage.

In electron microscopy a high energy electron beam bombards the sample. Electron lenses are used to provide an image from the scattering pattern electrons. The most commonly used technique to get a 3D structure from 2D projection image is Cryo-EM. In this method, thin films of samples preserved on none-crystalline ice are placed on the electron microscopy grid and thousands of 2D images capture the molecule in different orientations. Then a computational approach is used to obtain a 3D mass density map. An atomic model of the macromolecule then is fitted into 3DEM maps.

Under cryogenic temperatures, the biological sample rapidly freezes in its aqueous environment thus sample damage due to the electron irradiation reduces, making it possible to use higher doses of electrons to increase the signal to noise ratio. This also avoids ultrastructure changes and redistribution of elements. Moreover, the structure is similar to the native or liquid state. Molecules are in solution and are not distorted by being attached or flattened against the surface of the supporting film. In Cryo-EM the contrast between the nucleic acids, proteins and

lipids is distinguishable. Thus, Cryo-EM makes it possible to study the dynamics of biological macromolecular structures in near native condition at close to atomic resolution.

Single particle 3DEM is used to study structures at resolution limits comparable to macromolecular crystallography and can be used to visualize amino acid sidechains, surface water molecules, and bound ligands. Cryo-electron tomography (CET) yields slightly lower resolution structural information such as protein domains and secondary structures.

Some disadvantages of this technique include low image contrast and signal to noise ratio due to very low electron absorption of biological macromolecules. Difficulty in obtaining good images due to the thick ice cross section of a tilted frozen sample, and the need to keep samples at temperatures below 135 degree Celsius.

NMR spectroscopy studies the behavior of nuclear spin angular momentum. The rf radiation provides sufficient energy for the nuclei to resonance. The re-emitted electromagnetic radiation from the nuclei while the nuclei spin is relaxing back to original state forms a detectable NMR signal. The NMR spectrum is a signature of a target molecule and it gives us information about nuclei type, concentration, chemical environment, and molecule structure and dynamics. NMR spectroscopy provides information on a wide range of time scale dynamics thus is powerful in determining protein-ligand interactions, small molecule screenings, and optimization in drug discovery and development. It provides easier study of protein-protein interactions when co-crystallization of the complex is difficult. Sample preparation is easier and more flexible in NMR spectroscopy. Solid State NMR spectroscopy is a very useful technique in studying membrane proteins, which are not naturally soluble in aqueous solution and difficult to crystallize.

X-ray and electron microscopy are based on Fourier optics and image reconstruction from diffraction data. NMR spectroscopy however is based on measuring structural restraints such as distances and angles and finding structural solutions that satisfy those restraints.

Studying the protein complexes via X-ray crystallography and NMR spectroscopy is hindered by the limitations of these methods. Integrative or Hybrid Methods (I/HM) are very useful and fast growing ways of studying large macromolecular assemblies such as ribosomes complexes, tRNA and protein factors and muscle actomysins, membrane proteins, membrane embedded ion channels and receptors. 3DEM, Transmission Electron Microscopy (TEM) and cryo-TEM ,NMR spectroscopy, X-ray crystallography, mass spectroscopy, chemical cross linking, fluorescence resonance energy transfer, and computational methods can be combined to study very large macromolecular assemblies.

2.3. Theory of NMR Spectroscopy

The theory of NMR includes quantum mechanics of nuclear spin angular momentum, denoted by **I**. The magnitude of this vector quantity is given by

$$|\mathbf{I}| = [\mathbf{I} \cdot \mathbf{I}]^{1/2} = \hbar[\mathrm{I}(\mathrm{I}+1)]^{1/2} \qquad (2.1)$$

Where I is the *nuclear spin angular momentum quantum number* and ħ is the Planck's constant divided by 2π.

Nuclei with even mass number and even atomic number thus zero spin quantum number, are NMR inactive. These nuclei are not affected by external magnetic field. Nuclei with even mass numbers and odd atomic numbers have integral spin numbers (I = 1, 2, 3, …) and those with odd mass number, regardless of atomic number being odd/even, have half–integral spin numbers (I = 1/2, 3/2, 5/2, …). These nuclei are affected by external magnetic field thus are NMR active.

Based on principles of quantum mechanics only one Cartesian component of **I** can be specified simultaneously with $\mathbf{I}^2$, which by convention that component is the z-component and is defined by:

$$I_z = \hbar m \tag{2.2}$$

Where m is the *magnetic momentum quantum number* and takes 2I+1 values from –I to I thus m=(-I, -I+1 … I-1, I). For a nucleus with quantum spin number I the number of possible spin states is given by m = 2I+1. Thus, z-component of **I** takes discrete values, meaning the orientation of the spin angular momentum vector is quantized. In the absence of an external magnetic field the quantum states corresponding to all 2I+1 values of m, have same energy values and there is no preferred orientation. The isotopes of interest in this study are spin ½ nuclei, ^{1}H, ^{13}C and ^{15}N which have two allowed spin states, m = ±1/2.

Nuclei also have *nuclear magnetic momentum* **μ**, defined by:

$$\boldsymbol{\mu} = \gamma \mathbf{I} \tag{2.3}$$

$$\mu = \gamma\hbar\sqrt{I(I+1)}$$

$$\mu_z = \gamma I_z = \gamma\hbar m \tag{2.4}$$

Where γ is the *gyromagnetic ratio*, a characteristic constant of a given nucleus related to the receptivity (sensitivity) of a nucleus in NMR spectroscopy. Some of the properties of the most commonly used nuclei in biological NMR spectroscopy, including the gyromagnetic ratio are listed in table 2.1.

Nuclear magnetic moment is quantized. From the equation 2.1 it is perceivable that the nuclei with the same quantum number I have the same spin angular momentum. However, the magnetic momentum is different for each nucleus depending on their γ values.

Table 2.1 Properties of the most commonly used nuclei in biological NMR spectroscopy.

Isotope	Spin	Gyromagnetic ratio γ (10^7 rad s^{-1} T^{-1})	Natural Abundance (%)
^{1}H	½	26.752	99.99
^{2}H	1	4.107	0.012
^{13}C	½	6.728	1.07
^{14}N	1	1.934	99.63
^{15}N	½	-2.713	0.37
^{17}O	5/2	-3.628	0.038
^{19}F	½	25.18	100
^{23}Na	3/2	7.081	100
^{31}P	½	10.839	100
^{113}Cd	½	-5.961	12.22

Natural abundance of proton and its large gyromagnetic ratio has made it the most popular spin studied in NMR spectroscopy.

2.3.1. Energy States and Boltzmann Distribution

Energy of the spin states of the nucleus in presence of an external magnetic field **B** with the strength equal to B_0 is given by:

$$\mathbf{E} = -\boldsymbol{\mu} \cdot \mathbf{B} \tag{2.5}$$

$$E_m = -\gamma I_z B_0 = -m\hbar\gamma B_0 \tag{2.6}$$

By convention, the external magnetic field is applied along z-axis of laboratory coordinate system. From the equation 2.6 it is understood that the energy is quantized and takes discrete values. In the presence of an external static magnetic field, 2I+1 equally spaced energy levels, known as *Zeeman levels,* exist (Figure 2.1). By applying electromagnetic radiation, transitions between Zeeman states can be stimulated. According to the selection rule, only transition between two Zeeman states with $\Delta m = \pm 1$ is allowed. The required energy to stimulate a transition is then:

$$\Delta E = \hbar \gamma B_0 \tag{2.7}$$

According to Plank's law the transition frequency, i.e. the frequency of electromagnetic radiation required to excite transitions between Zeeman states, which is called *Larmor frequency,* is given by:

$$\omega = \frac{\Delta E}{\hbar} = \gamma B_0 \tag{2.8}$$

In s^{-1} units. Or

$$\nu = \frac{\omega}{2\pi} = \frac{\gamma B_0}{2\pi} \tag{2.9}$$

In Hertz.

When a spin ½ nucleus is placed in an external magnetic field (B_0), the nuclear magnetic moment experiences a torque trying to align it to the field. Nuclei in spin state m = +1/2 are in the lower energy state (α–state (E_1)) and align parallel to the field, and those in the negative spin state m = -1/2 are in higher energy state (β–state (E_2)) and align anti–parallel to the field (Figure 2.1).

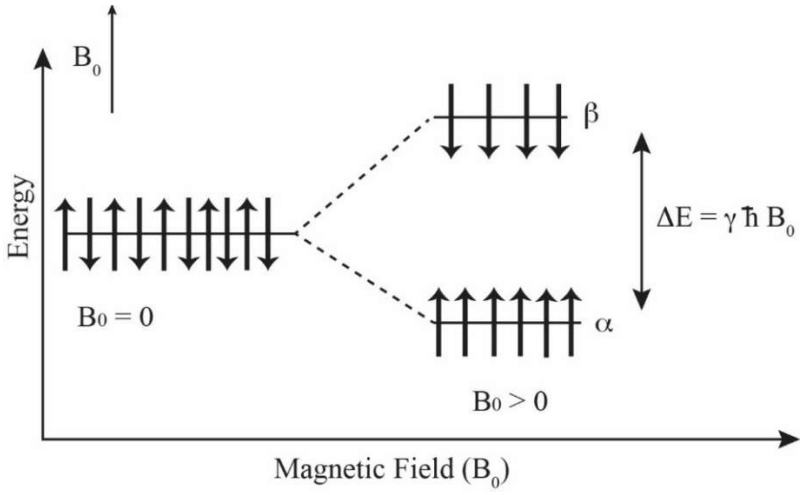


Figure 2.1 Zeeman splitting for a spin ½ nucleus.

The sensitivity of NMR spectroscopy depends on the population difference between Zeeman transition states. At equilibrium, the lower energy states are more populated. The relative population of the Zeeman states is determined by the Boltzmann equation as follows:

$$\frac{N_m}{N} = \exp\left(-\frac{E_m}{k_B T}\right) / \sum_{m=-I}^{I} \exp\left(-\frac{E_m}{k_B T}\right) \tag{2.10}$$

Which can be simplified to:

$$\frac{N_m}{N} \approx \left(1 + \frac{m\hbar\gamma B_0}{k_B T}\right)/(2I+1) \tag{2.11}$$

Where N is the total population of the spins from which N_m nuclei are in the *m*th state, k_B is the Boltzmann constant and T is the absolute temperature in Kelvin. The population of states depends on both nucleus type and the external magnetic field strength. In thermodynamic equilibrium, the populations of nuclei in each energy state are nearly equal and the nuclei in parallel state (lower energy) slightly more. By increasing the external magnetic field strength, the energy differences and the population differences between the energy states increase. For ^{1}H spins in an 11.7 T

magnetic field, the population difference is in the order of 1 in 10^5 and only a small fraction of the spins contributes to the signal intensity. This characterizes NMR as an insensitive spectroscopic technique and emphasizes the necessity of having stronger magnets.

2.3.2. The Bloch Equations

The Bloch equations formulate a semiclassical vector model to describe a sample of nuclei and how the sample/bulk magnetic moment (**M**) evolves with time in the static external magnetic field; **M(t)**.

Vector sum of the magnetic moment **μ** and angular momentum **I** over all the nuclei, give the macroscopic magnetic momentum **M** and angular momentum **J** for the sample.

$$M = \gamma \Delta N \tag{2.12}$$

Where ΔN is the small population difference of spins between the higher and lower energy spin states.

At the thermal equilibrium sum of the transverse components thus the net moment in the x–y plane is zero and **M** is parallel to the external magnetic field, i.e. $\mathbf{M}=M_0\mathbf{k}$. The magnetic moment of a nucleus precesses at an angle θ around the z–axis, and the frequency of the precession is the Larmor frequency. If the nuclei is now in an external magnetic field, the torque on the sample/bulk magnetic moment, **M(t),** will be:

$$\frac{d\mathbf{J}(t)}{dt} = \mathbf{M}(t) \times \mathbf{B}(t) \tag{2.13}$$

Multiplying both sides by γ we get:

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma\mathbf{B}(t) \tag{2.14}$$

Now we re-write the above equation in a rotating reference with an angular velocity of ω with respect to the laboratory frame.

$$\left(\frac{d\mathbf{M}(t)}{dt}\right)_{rot} = \left(\frac{d\mathbf{M}(t)}{dt}\right)_{lab} + \mathbf{M}(t) \times \omega \tag{2.15}$$

$$= \mathbf{M}(t) \times (\gamma\mathbf{B}(t) + \omega)$$

If we define the effective field, $\mathbf{B}_{eff}$ as:

$$\mathbf{B}_{eff} = \mathbf{B}(t) + \omega/\gamma \tag{2.16}$$

The equations of motion for magnetization in the rotating frame is going to have the same form as in laboratory frame. For $\omega = -\gamma\mathbf{B}(t)$, the effective field $\mathbf{B}_{eff}$ is 0 thus $\frac{d\mathbf{M}(t)}{dt} = 0$ which means that $\mathbf{M}(t)$ is time independent in the rotating frame and precesses around $\mathbf{B}(t)$ with a constant frequency $\omega = -\gamma B$ in laboratory frame. For a static external field with the strength of B_0, $\omega_0 = -\gamma B_0$. ω_0 is known as Larmour frequency or precessional frequency and is same as the magnitude of the transition frequency between Zeeman levels.

By convention, γB is referred to as the magnetic field strength in unit of frequency (s^{-1} or Hertz) instead of B with units of Gauss or Tesla, e.g. spectrometers with an 11.7 T and 21.2 T magnet are called 500 MHz and 900 MHz respectively.

As mentioned earlier, in thermal equilibrium the bulk magnetization is parallel to the external magnetic field. Radiofrequency (rf) electromagnetic radiation which can be applied in the form of *pulses* with duration of several microseconds, can disturb the equilibrium of the bulk magnetization from its parallel position. Once the bulk magnetization is displaced from the equilibrium, it precesses around the external magnetic field at the Larmor frequency. This creates

a time-varying magnetic field, which according to Faraday's induction law induces a detectable electromotive force (emf), i.e. a signal in a coil.

The linearly polarized rf field along the x-axis can be decomposed into two circularly polarized field rotating in opposite directions about the z-axis (Figure 2.2).

$$\mathbf{B}_{rf}(t) = 2B_1 \cos(\omega_{rf} + \phi)\,\mathbf{i} \tag{2.17}$$

$$= \mathbf{B}_1\{\cos(\omega_{rf} + \phi)\mathbf{i} + \sin(\omega_{rf} + \phi)\mathbf{j}\}$$

$$+ \mathbf{B}_1\{\cos(\omega_{rf} + \phi)\mathbf{i} - \sin(\omega_{rf} + \phi)\mathbf{j}\}$$

Where B_1 is the amplitude of the applied field, ω_{rf} is the angular frequency of the rf field known as *carrier frequency or transmitter frequency*, and ϕ is the phase of the field.

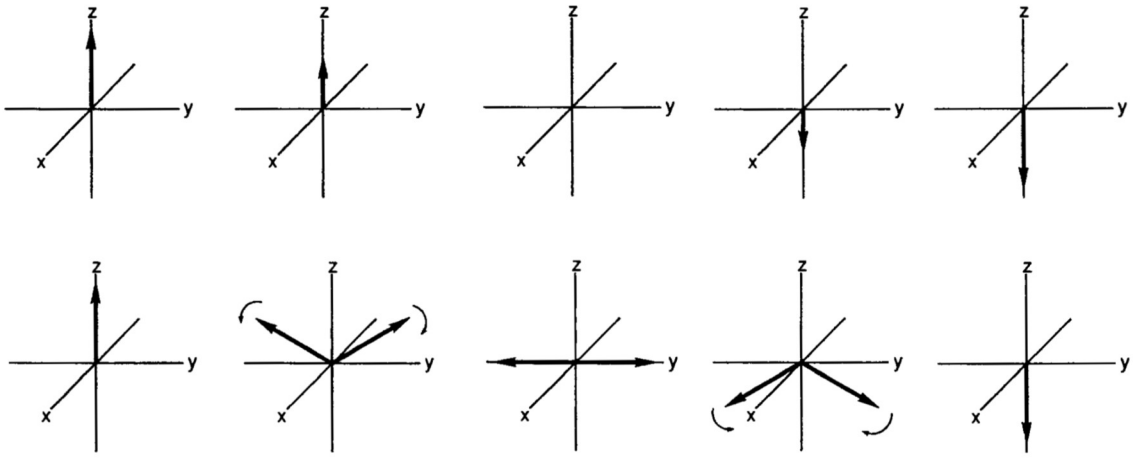


Figure 2.2 The decomposition of radio frequency radiation into two circularly polarized fields rotating in opposite directions about the z axis.

Since only the component of the field rotating in the same direction as the sample magnetic moment interacts significantly with the sample magnetic moment, the counter-rotating component of the field can be neglected. Thus:

$$\mathbf{B}_{rf}(t) = B_1\{\cos(\omega_{rf} + \phi)\,\mathbf{i} + \sin(\omega_{rf} + \phi)\mathbf{j}\} \quad (2.18)$$

We then can use the *rotating frame transformation* by moving to a rotating frame with angular frequency ω_{rf} to solve the time-dependent field and obtain the following equation:

$$\frac{d\mathbf{M}^r(t)}{dt} = \mathbf{M}^r(t) \times \gamma\mathbf{B}^r \quad (2.19)$$

$\mathbf{B}^r$ is the effective field in the rotating frame and is given by:

$$\mathbf{B}^r = B_1 \cos\phi\,\mathbf{i}^r + B_1 \sin\phi\,\mathbf{j}^r + \Omega/\gamma\,\mathbf{k}^r \quad (2.20)$$

$\Omega = -\gamma B_0 - \omega_{rf} = \omega_0 - \omega_{rf}$ is the *offset*. If the carrier frequency ω_{rf} is equal to the Larmor frequency ω_0, then the offset will be 0. In this case the irradiation is called *on-resonance* and $B^r = B_1$. When on resonance, low–energy (parallel) nuclear spins get excited into higher energy levels. Followed by the transition of excited spins back to original energy levels, measureable current will be induced in the receiver coil that forms the NMR signal.

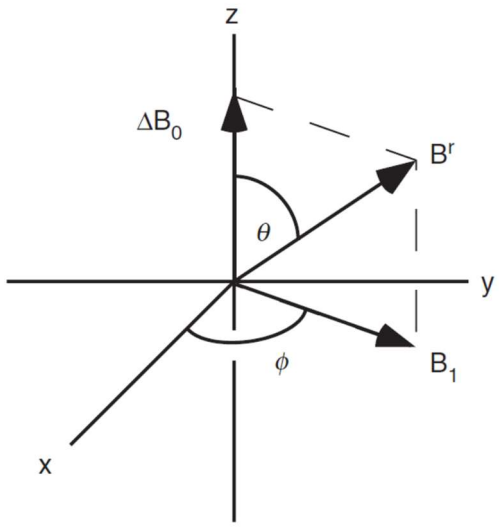


Figure 2.3 ΔB_0, B_1 and B^r in the rotating frame

The magnitude and the angle of the effective field with respect to the z-axis is:

$$B^r = \sqrt{(B_1)^2 + (\Delta B_0)^2} = B_1\sqrt{1 + (\tan\theta)^{-2}} = B_1/\sin\theta \tag{2.21}$$

$$\tan\theta = \frac{B_1}{\Delta B_0} = \frac{-\Upsilon B_1}{\Omega} = \frac{\omega_1}{\Omega} \tag{2.22}$$

The $\mathbf{M}^r(t)$ precesses around the effective field $\mathbf{B}^r$ with an angular frequency of $\omega^r = -\gamma B^r$.

On resonance, the bulk magnetization $\mathbf{M}^r(t)$ precesses around the B_1 field with frequency $\omega^r = -\gamma B^r = -\gamma B_1 = \omega_1$. Where ω_1 is the angular frequency of applied pulse.

2.3.3. Relaxation

As mentioned earlier, following the rf pulse, the sample magnetization precesses around the external magnetic field with frequency equal to Larmor frequency. Magnetization eventually decays and returns to the Boltzmann equilibrium distribution. There are two relaxation processes through which magnetization decays back to equilibrium. *Longitudinal/spin-lattice relaxation* and *transverse/spin-spin relaxation.* From Bloch equations, we obtain the relaxation terms as follows:

$$\frac{dM_z(t)}{dt} = R_1[M_0 - M_z(t)] \tag{2.23}$$

$$M_z(t) = M_0 - [M_0 - M_z(0)]e^{-R_1 t}$$

$$\frac{dM_{x/y}}{dt} = -R_2 M_{x/y} \tag{2.24}$$

$$M_{x/y} = M_{x/y}(0)e^{-R_2 t} = M_{x/y}(0)e^{-t/T_2}$$

$R_1=1/T_1$ is the *spin-lattice relaxation rate constant* (T_1 is the *spin–lattice relaxation time*), $R_2=1/T_2$ is the *spin-spin relaxation rate constant* (T_2 is the *spin–spin relaxation time*).

The energy absorbed from an applied rf electromagnetic field can be either dissipated to the lattice (T_1 relaxation), or can be exchanged between spins of the system (T_2 relaxation).

T_1 relaxation is the time required for the net magnetization to reach its equilibration state that allows the nuclear spins to dissipate their energy to the lattice. The net magnetization is restored along the z–axis and T_1 relaxation is in the direction of static magnetization field, B_0. Generally, an experiment is repeated after a period of $5T_1$ to allow enough time for the spins to fully recover the net magnetization, M_0.

Interactions between spins result in the loss of phase coherence in the x–y plane or *transverse plane* and is described by time constant, T_2. T_2 relaxation is in the transverse plane perpendicular to the static magnetization field and is the time required for the net magnetization in the transverse plane to dephase. Dephasing of the vectors results in an exponential decay of the net transverse magnetization. Thus T_2 is a measure of how fast the transverse magnetizations (M_x or M_y) decays to zero. Spin–spin interactions arise from local magnetic field inhomogeneity due to an inhomogeneous B_0 field from the spectrometer magnet or due to varying magnetic susceptibilities within the sample. Measuring T_2 is an indicative of the nuclear environments in a sample. Rigidly bound protons have shorter T_2 relaxation times.

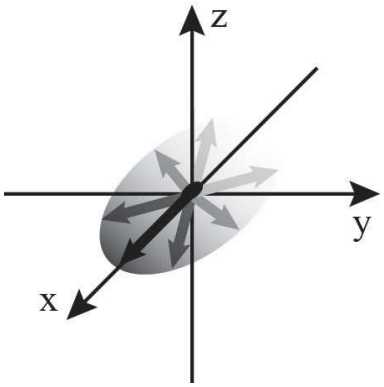


Figure 2.4 Spin–spin relaxation
modelled as decoherence of magnetization in the transverse plane due to local magnetic field inhomogeneity.

Ideally, transverse relaxation is inversely proportional to the spectral linewidth. However due to the spin diphase in x-y plane, inhomogeneous local magnetic fields increase the decay rate.

$$LW_{1/2} = 1/\pi T_2 * \tag{2.25}$$

$LW_{1/2}$ is the spectral peak width at half height, T_2* is the relaxation value and has two components. One is the real T_2 value of the sample and the other is contribution of field inhomogeneity.

$$1/T_2* = 1/T_2 + 1/T_{inhom} \tag{2.26}$$

Back to the Bloch equations, substituting $\frac{dM_x}{dt}, \frac{dM_y}{dt}, \frac{dM_z}{dt}$ from (2.23) and (2.24) into (2.14), Bloch equations can be written in the rotating reference frame. For convenience, we write the Bloch equations in a matrix form.

$$\frac{d\mathbf{M}(t)}{dt} = \begin{bmatrix} -R_2 & -\Omega & \omega_1 \sin\phi \\ \Omega & -R_2 & -\omega_1 \cos\phi \\ -\omega_1 \sin\phi & \omega_1 \cos\phi & -R_1 \end{bmatrix} \mathbf{M}(t) + R_1 M_0 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.27}$$

If the pulse B_1 is only applied for a short time $\tau_p \ll 1/R_1$ and $1/R_2$, then the above equation simplifies to:

$$\frac{d\mathbf{M}(t)}{dt} = \begin{bmatrix} 0 & -\Omega & \omega_1 \sin\phi \\ \Omega & 0 & -\omega_1 \cos\phi \\ -\omega_1 \sin\phi & \omega_1 \cos\phi & 0 \end{bmatrix} \mathbf{M}(t) \tag{2.28}$$

If B_1 and ϕ are time independent then the solution to this equation can be represented as a series of rotation matrices being applied on the magnetization $\mathbf{M}(0)$:

$$\mathbf{M}(\tau_p) = \mathbf{R}_z(\phi)\mathbf{R}_y(\theta)\mathbf{R}_z(\alpha)\mathbf{R}_y(-\theta)\mathbf{R}_z(-\phi)\mathbf{M}(0) \tag{2.29}$$

$\mathbf{R}_i(\beta)$ represents a right handed rotation of angle β around i-axis. A positive rotation is a counterclockwise rotation viewed down the x-axis.

This rotation matrix can be used to calculate the effect of rf pulses on isolated spins.

Rotation matrices are represented as follows:

$$\mathbf{R}_x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix} \tag{2.30}$$

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\mathbf{R}_z(\beta) = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Transverse sample magnetization results in the maximum NMR signal thus rf pulses that cause 90° rotation of $\mathbf{M}(t)$ from the z-axis are of great importance in NMR spectroscopy.

Let us assume rf pulse duration/pulse width is τ_p, pulse strength B_1, tilt angle of the pulse $\theta = \pi/2$ is applied onto the equilibrium sample magnetization, $\mathbf{M_0}$ along the y-axis, thus phase $\phi = \pi/2$. Thus only one rotation matrix, i.e. $\mathbf{R}_y(\alpha)$ is acting on the $\mathbf{M_0}$. In this case equation (2.29) can be written as:

$$\mathbf{M}(\tau_p) = \mathbf{R}_y(\alpha)\mathbf{M_0} = \mathbf{i}M_0 \sin\alpha + \mathbf{k}M_0 \cos\alpha = \begin{pmatrix} M_0 \sin\alpha \\ 0 \\ M_0 \cos\alpha \end{pmatrix} \qquad (2.31)$$

Where α is the rotation angle, $\alpha = -\gamma B^r \tau_p$. A maximum transverse magnetization can be generated with α = 90°. The corresponding rf pulse is called *90°* or a *π/2 pulse*. A π/2 pulse rotates the net magnetic moment vector to the x–y (transverse) plane and equalizes the population of spin states for the spin ½ nuclei. *180° pulse* or *π pulse* on the other hand generates no transvers magnetization, instead it inverts the sample magnetization to be $-M_0\mathbf{k}$, and the population in the two ±½ spin states. For a nucleus with a positive γ, a π pulse inverts the population such that –½ state (β-state) possesses excess population. After the rf pulse is applied, the sample magnetization precesses around the direction of the effective magnetic field in the rotating frame for a period of time called *acquisition period, t* during which a signal is generated by the precessing magnetization and recorded by the NMR spectrometer. This induced signal is an exponentially decaying function of time and is called *free induction decay (FID)*. The FID is then Fourier transformed (FT) to the frequency domain to obtain the characteristic NMR spectrum.

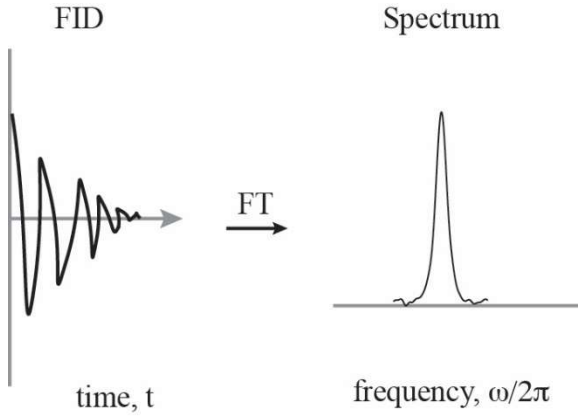


Figure 2.5 Fourier transformation of FID
from time domain to the frequency domain.

Bloch equations for the free induction decay in the rotating frame for a 90° pulse with $B_{x/y} = 0$ and $B_z = B_0$, give two components for M, M_x and M_y, which can be combined and be written as M^+ as follows:

$$M^+(t) = M_x(t) + iM_y(t) \tag{2.32}$$

$$= M_0 \sin\alpha \cos(\Omega t)\, e^{-R_2 t}$$

$$+ M_0 \sin\alpha \sin(\Omega t)\, e^{-R_2 t}$$

$$= M_0 \sin\alpha\, e^{i\Omega t - R_2 t}$$

As it is seen in the above equation the transverse components of the sample magnetization decay with a factor of $e^{-R_2 t}$ in the spin-spin relaxation process.

Both the real and imaginary parts of the complex signal are detected by the NMR spectrometer with an experimental proportional factor λ. This is a complex time-domain signal which can then be converted to a complex frequency-domain signal through Fourier transformation. Let us define $s^+(t) = \lambda M^+(t)$. Then:

$$S(\omega) = \int_0^\infty s^+(t) e^{-i\omega t} dt \tag{2.33}$$

$$= \lambda M_0 \frac{R_2}{R_2^2 + (\Omega - \omega)^2} + i\,\lambda M_0 \frac{\Omega - \omega}{R_2^2 + (\Omega - \omega)^2}$$

$$= \nu(\omega) + iu(\omega)$$

The real part, $\nu(\omega)$ represents a signal with an absorptive Lorentzian lineshape and the imaginary part, $u(\omega)$ represents a signal with dispersive Lorentzian lineshape. Normally the real part of the signal is displayed as NMR spectrum.

2.3.4. Linewidth

Linewidth of the Lorentzian lineshape $\nu(\omega)$, $\Delta\nu_{FWHM}$, is defined as full-width at half maximum and can be shown to be $\Delta\nu_{FWHM} = \frac{R_2^*}{\pi} = \frac{R_2 + R_{inhom}}{\pi}$. R_2 in the homogeneous part of the linewidth is due to intrinsic molecular properties and R_{inhom} in the inhomogeneous part of the linewidth can be due to instrumental imperfections. Homogenous linewidth is proportional to the rotational correlation time of the protein, thus to the molecular mass and shape of the protein. If the linewidth is significantly larger than expected, this can imply that aggregation is increasing or that chemical exchange effects are contributing to the linewidth.

2.3.5. Chemical Shift

The static external magnetic field induces motion of electrons which then generates local secondary magnetic fields. So the net magnetization is sum of both static external magnetic field and local secondary magnetic fields (The latter is called *nuclear shielding*). The observed resonance frequencies are not the same for all nuclei, instead they depend on the local electronic and structural environment of each individual nucleus. The differences in resonance frequencies for different nuclei is called *chemical shifts*.

$$B = (1 - \sigma)B_0 \qquad (2.34)$$

For convenience and because it is difficult to measure absolute value of the chemical shift of a resonance, chemical shifts are measured in *parts per million (ppm)* relative to a reference resonance signal.

The chemical shifts of ^{1}H, ^{13}C and ^{15}N are referenced to DSS, DSS and liquid NH3, respectively.

$$\delta = \frac{\Omega - \Omega_{\text{ref}}}{\omega_0} \times 10^6 = (\sigma_{\text{ref}} - \sigma) \times 10^6 \quad (2.35)$$

Ω_{ref} and Ω are the offset frequencies of the reference signal and the signal of interest. Chemical shifts in the unit of ppm are independent of the static magnetic field strength. Average chemical shifts of nuclei in amino acids of proteins (in ppm) is listed in Table 2.2.

Table 2.2 Average chemical shifts of nuclei in amino acids of proteins (in ppm)[4].

Residue	$^{1}H^{N}$	^{15}N	^{13}C	$^{13}C\alpha$	$^{1}H\alpha$	$^{1}H\beta$
Ala	8.15	122.5	177.6	52.2	4.33	1.39
Arg	8.27	120.8	176.6	56.0	4.35	1.89, 1.79
Asn	8.38	119.5	175.6	52.7	4.74	2.83, 2.75
Asp	8.37	120.6	176.8	53.9	4.71	2.84, 2.75
Cys	8.23	118.0	174.6	56.8	4.54	3.28, 2.96
Gln	8.27	120.3	175.6	56.0	4.33	2.13, 2.01
Glu	8.36	121.3	176.6	56.3	4.33	2.09,1.97
Gly	8.29	108.9	173.6	45.0	3.96	.
His	8.28	119.1	174.9	55.5	4.60	3.26, 3.20
Ile	8.21	123.2	176.5	61.2	4.17	1.90
Leu	8.23	121.8	176.9	55.0	4.32	1.65
Lys	8.25	121.5	176.5	56.4	4.33	1.85, 1.76
Met	8.29	120.5	176.3	55.2	4.48	2.15, 2.01
Phe	8.30	120.9	175.9	57.9	4.63	3.22, 2.99
Pro	.	128.1	176.0	63.0	4.42	2.28, 2.02
Ser	8.31	116.7	174.4	58.1	4.47	3.88
Thr	8.24	114.2	174.8	62.0	4.35	4.22
Trp	8.18	120.5	173.6	57.6	4.66	3.32, 3.19
Tyr	8.28	122.0	175.9	58.0	4.55	3.13, 2.92
Val	8.19	121.1	176.0	62.2	4.12	2.13

In an anisotropic electronic charge distribution (which is true in most cases), the effect of nuclear shielding is described by second rank shielding tensors $\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix}$. In the principle coordinate system the shielding tensor is a diagonal matrix with components σ_{xx}, σ_{yy}, σ_{zz}.

Chemical shifts in protein NMR spectroscopy consist of the random coil chemical shift component (which is observed in a conformationally disordered peptide) and the secondary

chemical shift component. Secondary chemical shifts are useful in prediction the protein structure and have information regarding the secondary and tertiary structure of the protein.

2.3.6. Limitations of the Bloch Equations

The splitting of the resonance signals is observed in high resolution NMR spectra of liquids. Direct dipolar interactions through space between magnetic dipole moments, *dipolar coupling,* is an anisotropic quantity that in solutions averages out to zero. Thus, it cannot explain the splitting of signals in NMR spectra of solutions. S*pin-spin coupling* or *scalar coupling* however can explain the splitting of the resonance signals. Scalar coupling is indirect dipole-dipole interaction between the two spins mediated by bonding electrons and is orientation dependent. As the number of bonds separating the nuclei increases, the strength of scalar coupling decreases. The interaction between the nuclei is measured by $^{n}J_{ab}$, *scalar coupling constant*, which has the units of Hertz. n is the number of covalent bonds separating the two nuclei a, b.

The three-bond scalar coupling constant ^{3}J has a relationship with the dihedral angle and is a quantitative method to obtain dihedral angles in NMR spectroscopy via Karplus equation:

$$J = A\cos^2\varphi + B\cos\varphi + C \tag{2.36}$$

φ is the dihedral angle and A, B, C are constants

Let us consider a sample of two spin ½ nuclei. We show the spins by I and S. with the resonance frequencies ω_I and ω_S.

$$\omega_I = -\gamma_I B_0(1 - \sigma_I) \tag{2.37}$$

$$\omega_S = -\gamma_S B_0(1 - \sigma_S) \tag{2.38}$$

The system of two spins can be described by four wave functions as follows

$$\psi_1 = \psi\left(\frac{1}{2},\frac{1}{2}\right), \psi_2 = \psi\left(\frac{1}{2},-\frac{1}{2}\right), \tag{2.39}$$

$$\psi_3 = \psi\left(-\frac{1}{2},\frac{1}{2}\right), \psi_4 = \psi\left(-\frac{1}{2},-\frac{1}{2}\right)$$

And the energy of the system is the sum of energies for each spin which is shown by:

$$E_1 = \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S, E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S, \tag{2.40}$$

$$E_3 = -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S, E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S$$

The total magnetic quantum number is the sum of four magnetic numbers for each state.

$$m_1 = \frac{1}{2} + \frac{1}{2} = 1, m_2 = \frac{1}{2} - \frac{1}{2} = 0, \tag{2.41}$$

$$m_3 = -\frac{1}{2} + \frac{1}{2} = 0, m_4 = -\frac{1}{2} - \frac{1}{2} = -1$$

The transition between states obeys the selection rule $\Delta m = \pm 1$. Thus only transitions between states 1-2, 1-3, 2-4 and 3-4 can happen, with the energy differences between these transitions to be:

$$|\mathrm{E}_2 - \mathrm{E}_1| = \hbar\omega_\mathrm{S}, |\mathrm{E}_3 - \mathrm{E}_1| = \hbar\omega_\mathrm{I}, \tag{2.42}$$

$$|\mathrm{E}_4 - \mathrm{E}_2| = \hbar\omega_\mathrm{I}, |\mathrm{E}_4 - \mathrm{E}_3| = \hbar\omega_\mathrm{S}$$

In the absence of scalar coupling, the transition frequency between energy states 1-2 and 3-4 is ω_S, and the transition frequency between energy states 1-3 and 2-4 is ω_I. So two resonance lines are predicted.

In the presence of the scalar coupling between spins I and S however:

$$E(m_I, m_S) = m_I\omega_I + m_S\omega_S + 2\pi m_I m_S J_{IS} \tag{2.43}$$

Thus

$$E_1 = \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS}, \tag{2.44}$$

$$E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS},$$

$$E_3 = -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS},$$

$$E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS}$$

The transitions still obey the selection rules thus only the transitions between energy states 1-2, 1-3, 3-4, 2-4 can happen with the following resonance frequencies:

$$\omega_{12} = \omega_S + \pi J_{IS}, \omega_{13} = \omega_I + \pi J_{IS}, \omega_{34} = \omega_S - \pi J_{IS}, \omega_{24} \tag{2.45}$$
$$= \omega_I - \pi J_{IS}$$

Thus, four resonance spectral lines are predicted as it is observed in practice. Bloch equations fail to predict this observation and are only applicable in noninteracting spin-1/2 nuclei.

2.3.7. Quantum Mechanical Representation of NMR Spectroscopy

Quantum mechanics provides a thorough approach to analyze the dynamics of the nuclear spin system via *density matrix formulation*. In Bloch equations, the evolution of magnetization vector was observed. And as it was mentioned, this approach has limitations when it comes to coupled/interactive nuclei. In the quantum mechanical approach, however the evolution of the density matrix is studied and gives a more complete description of the spin system over time.

The time evolution of a quantum mechanical system is given by the Schrodinger equation.

$$\frac{\partial \Psi(t)}{\partial t} = -\frac{i}{\hbar}\hat{H}\Psi(t) \tag{2.46}$$

$\hat{H}$ is the Hamiltonian of the system, $\Psi(t)$ is the wavefunction of the system and has all the information about the state of the sytem. Knowing the wave function all the observable properties of the system can be determined. Using the separation of variables, for a time-independent Hamiltonian, the Schrodinger equation can be solved.

$$\Psi(t) = \psi(\tau)e^{-i\left(\frac{E}{\hbar}\right)t} = \psi(\tau)e^{-i\omega t} \tag{2.47}$$

Where $E=\hbar\omega$ and $\Psi(\tau)$ contain the time dependent spatial and spin variables.

Expectation value of a property A is the average magnitude of A and is obtained from the integral of wave function as follows:

$$\langle A \rangle = \int \Psi^*(t) A \Psi(t)\, d\tau \tag{2.48}$$

For a spin ½ system in the static magnetic field, the wavefunction can be written as the linear combination of the two basis states with m=1/2 and m= -1/2:

$$\Psi = c_\alpha \psi_\alpha + c_\beta \psi_\beta = a e^{-i\omega_\alpha t}\psi_\alpha + b e^{-i\omega_\beta t}\psi_\beta \tag{2.49}$$

ψ_α and ψ_β are the two states, E_α and E_β are the energy of the states, and $a = |a|e^{i\phi a}$, $b = |b|\ e^{i\phi b}$ are complex numbers which satisfy the normalization equation $|a|^2 + |b|^2 = 1$.

The expected values are going to be:

$$\langle \mu_x \rangle = \int \Psi^* \mu_x \Psi d\tau = \gamma\hbar|a||b|\cos(\omega_0 + \phi) \tag{2.50}$$

$$\langle \mu_y \rangle = \int \Psi^* \mu_y \Psi d\tau = \gamma\hbar|a||b|\sin(\omega_0 + \phi) \tag{2.51}$$

$$\langle \mu_z \rangle = \int \Psi^* \mu_z \Psi d\tau = \frac{\gamma\hbar}{2}(|a|^2 - |b|^2) \tag{2.52}$$

where $\omega_0 = \omega_\alpha - \omega_\beta = -\gamma B_0$ is the Larmor frequency and $\phi = \phi_\alpha - \phi_\beta$ is a phase angle.

The angular momentum operators are going to be:

$$I_x\psi_\alpha = \frac{\hbar}{2}\psi_\beta,\ I_x\psi_\beta = \frac{\hbar}{2}\psi_\alpha \tag{2.53}$$

$$I_y\psi_\alpha = i\frac{\hbar}{2}\psi_\beta,\ I_y\psi_\beta = -i\frac{\hbar}{2}\psi_\alpha$$

$$I_z\psi_\alpha = \frac{\hbar}{2}\psi_\alpha,\ I_z\psi_\beta = -\frac{\hbar}{2}\psi_\beta$$

These results are similar to the results that was previously obtained from the Bloch model.

2.3.8. The Density Matrix

The density matrix formulation of quantum mechanics facilitates the calculations on commonly used quantum mechanical operators like scalar products and expectation values. A further simplification is using a notational system called *Dirac Notation.* In Dirac notation, the wave function and its conjugate are shown by $|\psi\rangle$ and $\langle\psi|$ and the scalar product is shown by $\langle\varphi|\psi\rangle$. The expectation value is written as $\langle A\rangle = \langle\psi|A|\psi\rangle$.

Using $|\alpha\rangle$ and $|\beta\rangle$ basis functions to represent the states of spin in a spin ½ nucleus, where $|\alpha\rangle = \begin{pmatrix}1\\0\end{pmatrix}, |\beta\rangle = \begin{pmatrix}0\\1\end{pmatrix}$ are eigen vectors, the matrices of the angular momentum operators will be shown by the Pauli spin matrices as follows:

$$I_x = \frac{1}{2}\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}, I_y = \frac{1}{2}\begin{bmatrix}0 & -i\\i & 0\end{bmatrix}, I_z = \frac{1}{2}\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix} \tag{2.54}$$

Which satisfy the commutation relation of:

$$[I_x, I_y] = iI_z, \qquad [I_y, I_z] = iI_x, \qquad [I_z, I_x] = iI_y \tag{2.55}$$

An arbitrary ket can then be shown by a linear combination of the eigen kets and eigen bras:

$$|\Psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle = \begin{pmatrix}c_\alpha\\c_\beta\end{pmatrix} \tag{2.56}$$

A pulse sequence consists of rf pulses (shown by time dependent Hamiltonian) and delays (shown by time independent Hamiltonian). Let us neglect the interaction between the spin system and the lattice. Equivalent to the rotating frame transformation in Bloch model, here we can transform the time dependent Hamiltonian to **H**, the time independent Hamiltonian.

Consider a spin ½ nucleus in the static external magnetic field B_0. If a linearly polarized rf pulse is applied with the magnitude of $2B_1$ and angular frequency of ω_{rf}, The Hamiltonian has two terms, the Zeeman Hamiltonian and the rf pulse Hamiltonian terms:

$$H = -\boldsymbol{\mu}.\mathbf{B}(t) = H_z + H_{rf} \tag{2.57}$$

$$= \omega_0 I_z + \omega_1[I_x \cos(\omega_{rf}t + \phi) + I_y \sin(\omega_{rf}t + \phi)]$$

Where $\omega_0 = -\gamma B_0$ and $\omega_1 = -\gamma B_1$

Using the transformation matrix **U**=exp ($i\omega_{rf}I_z t$), and after some calculations we get the time independent Hamiltonian as

$$\mathbf{H} = \Omega I_z + \omega_1[I_x \cos(\phi) + I_y \sin(\phi)] \tag{2.58}$$

The solution of this Hamiltonian describes how the density operator evolves in the rotating frame.

We define the exponential rotation operators as:

$$\mathbf{R}_x(\alpha) = e^{-i\alpha I_x} \tag{2.59}$$

Then for a rf pulse with duration of τ_p, $\alpha=\omega_1 \tau_p$, and rotation of the initial magnetization $\sigma(0)$ is

$$\sigma(t) = \mathbf{R}_j(\alpha)\sigma(0)\mathbf{R}_j^{-1}(\alpha) \tag{2.60}$$

Where $\mathbf{R}_j(\alpha)$ and $\mathbf{R}_j^{-1}(\alpha)$ are pulse rotation operators.

After expansion of $\mathbf{R}_x^{-1}(\alpha)$ and using the Pauli spin matrices we get:

$$\mathbf{R}_x^{-1}(\alpha) = \mathbf{E}\cos\left(\frac{\alpha}{2}\right) + 2iI_x \sin\left(\frac{\alpha}{2}\right) \tag{2.61}$$

The matrix representation of the pulse operators will then be:

$$\mathbf{R}_x^{-1}(\alpha) = \begin{pmatrix} c & is \\ is & c \end{pmatrix}, \qquad \mathbf{R}_x(\alpha) = \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \tag{2.62}$$

$$\mathbf{R}_y^{-1}(\alpha) = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}, \qquad \mathbf{R}_x(\alpha) = \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \tag{2.63}$$

$$\mathbf{R}_z^{-1}(\alpha) = \begin{pmatrix} c + is & 0 \\ 0 & c - is \end{pmatrix}, \mathbf{R}_z(\alpha) = \begin{pmatrix} c - is & 0 \\ 0 & c + is \end{pmatrix} \tag{2.64}$$

Where c = cos(α/2), s = sin(α/2).

Now we consider the simplest NMR experiment, which is a single pulse experiment. For an x-phase pulse with rotation of angle α, applied on I_z:

$$\mathbf{R}_x(\alpha) I_z \mathbf{R}_x^{-1}(\alpha) = \frac{1}{2}\begin{pmatrix} c & -is \\ -is & c \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} c & is \\ is & c \end{pmatrix} \tag{2.65}$$

$$= \frac{1}{2}\begin{pmatrix} c & is \\ -is & -c \end{pmatrix}\begin{pmatrix} c & is \\ is & c \end{pmatrix} =$$

$$\frac{1}{2}\begin{pmatrix} c^2 - s^2 & 2ics \\ -2ics & s^2 - c^2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \cos\alpha & i\sin\alpha \\ -i\sin\alpha & -\cos\alpha \end{pmatrix} = I_z \cos\alpha - I_y \sin\alpha$$

For α=180° and α=90°, we will get:

$$\mathbf{R}_x(\pi) I_z \mathbf{R}_x^{-1}(\pi) = -\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = -I_z \tag{2.66}$$

This is compatible with the result from Bloch model and shows that this pulse rotates the magnetization from z direction to –z direction and causes population inversion.

$$\mathbf{R}_x\left(\frac{\pi}{2}\right) I_z \mathbf{R}_x^{-1}\left(\frac{\pi}{2}\right) = \frac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} = -I_y \tag{2.67}$$

Which is again compatible with the Bloch model showing the change of magnetization from + z direction to –y axis followed by the precession of the $-I_y$ magnetization under the time independent Zeeman Hamiltonian in the rotating frame.

$$\mathbf{H} = (\omega_0 - \omega_{rf})I_z = \Omega I_z \tag{2.68}$$

$$\sigma(t) = -e^{-i\Omega t I_z} I_y e^{i\Omega t I_z} = I_x \sin(\Omega t) - I_y \cos(\Omega t) \tag{2.69}$$

$\sigma(t)$ changes from $-I_y$ at $t = 0$ to $+I_x$ at $\Omega t = \pi/2$.

2.3.9. Multi Spin Systems

The density operator applications can be extended to coupled spin systems where the Bloch model fails. Wave functions and operators are shown with higher dimension matrices. The spin systems can consist of N spin in general but we first look at a two spin ½ system of spins I and S. The system rotation is derived by applying the rotation operator on each individual spin and then multiplying the results.

$$\mathbf{R}_x(\alpha)[I] = \mathbf{R}_x(\alpha) \otimes \mathbf{E} = \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.70}$$

$$= \begin{pmatrix} c & 0 & -is & 0 \\ 0 & c & 0 & -is \\ -is & 0 & c & 0 \\ 0 & -is & 0 & c \end{pmatrix}$$

$$\mathbf{R}_x(\alpha)[S] = \mathbf{E} \otimes \mathbf{R}_x(\alpha) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \tag{2.71}$$

$$= \begin{pmatrix} c & -is & 0 & 0 \\ -is & c & 0 & 0 \\ 0 & 0 & c & -is \\ 0 & 0 & -is & c \end{pmatrix}$$

Multiplying the results, we get:

$$\mathbf{R}_x(\alpha) = \mathbf{R}_x(\alpha)[I]\mathbf{R}_x(\alpha)[S] = \begin{pmatrix} c^2 & icsu & icsv & -s^2 \\ icsu & 1 - s^2u^2 & -s^2uv & icsu \\ icsv & -s^2uv & 1 - s^2v^2 & icsv \\ -s^2 & icsu & icsv & c^2 \end{pmatrix} \tag{2.72}$$

$$\mathbf{R}_x^{-1}(\alpha) = \begin{pmatrix} c^2 & -icsu & -icsv & -s^2 \\ -icsu & 1-s^2u^2 & -s^2uv & -icsu \\ -icsv & -s^2uv & 1-s^2v^2 & -icsv \\ -s^2 & -icsu & -icsv & c^2 \end{pmatrix} \quad (2.73)$$

Where $u = \cos\theta + \sin\theta$, $v = \cos\theta - \sin\theta$ and $c = \cos(\alpha/2)$, $s = \sin(\alpha/2)$.

This approach is useful in heteronuclear NMR experiments.

Now we apply the following approach to a two spin system after applying a single pulse.

The equilibrium matrix representation of the density operator for a two weakly coupled spin system is:

$$\sigma(0) \approx \omega_I I_z + \omega_S S_z = \quad (2.74)$$

$$= \frac{1}{2}\begin{pmatrix} \omega_I + \omega_S & 0 & 0 & 0 \\ 0 & \omega_I - \omega_S & 0 & 0 \\ 0 & 0 & -\omega_I + \omega_S & 0 \\ 0 & 0 & 0 & -\omega_I - \omega_S \end{pmatrix}$$

A x-pulse with rotating angle of $\alpha=\pi/2$ (for simplicity) rotates the initial density operator.

$$\mathbf{R}_x\left(\frac{\pi}{2}\right) = \frac{1}{2}\begin{pmatrix} 1 & -i & -i & -1 \\ -i & 1 & -1 & -i \\ -i & -1 & 1 & -i \\ -1 & -i & -i & 1 \end{pmatrix}, \mathbf{R}_x^{-1}\left(\frac{\pi}{2}\right) = \frac{1}{2}\begin{pmatrix} 1 & -i & -i & -1 \\ -i & 1 & -1 & -i \\ -i & -1 & 1 & -i \\ -1 & -i & -i & 1 \end{pmatrix}$$

and $\sigma(0) = \frac{1}{2}\begin{pmatrix} \omega_I + \omega_S & 0 & 0 & 0 \\ 0 & \omega_I - \omega_S & 0 & 0 \\ 0 & 0 & -\omega_I + \omega_S & 0 \\ 0 & 0 & 0 & -\omega_I - \omega_S \end{pmatrix}$

Thus

$$\sigma(t) = \mathbf{R}_x\left(\frac{\pi}{2}\right)\sigma(0)\mathbf{R}_x^{-1}\left(\frac{\pi}{2}\right) = \quad (2.75)$$

$$= -\omega_I I_y - \omega_S S_y$$

Each term in the initial density operator is transformed and rotated to –y axis.

Considering the coupling constant J_{IS}, the matrix representation of the exponential operator will be:

$$\exp\left[i(\Omega_I I_z+\Omega_S S_z+2\pi J_{IS} I_z S_z)t\right] = \frac{i}{2}\begin{pmatrix} 0 & 0 & 0 & -e^{-i(\Omega_I+\pi J_{IS})t} - e^{-i(\Omega_S+\pi J_{IS})t} \\ 0 & e^{i(\Omega_S+\pi J_{IS})t} & e^{i(\Omega_S+\pi J_{IS})t} & 0 \\ 0 & e^{i(\Omega_I+\pi J_{IS})t} & e^{i(\Omega_I+\pi J_{IS})t} & 0 \\ 0 & 0 & 0 & e^{i(\Omega_I-\pi J_{IS})t} + e^{i(\Omega_S-\pi J_{IS})t} \end{pmatrix} \quad (2.76)$$

The spectrum has two doublets, total of four signals with the frequencies equal to $\Omega_I \pm \pi J_{IS}$ and $\Omega_S \pm \pi J_{IS}$. This observed magnetization comes from the fact that the trace of the above matrix has four terms $\exp[i(\Omega_I+\pi J_{IS})t]$, $\exp[i(\Omega_I-\pi J_{IS})t]$, $\exp[i(\Omega_S+\pi J_{IS})t]$, $\exp[i(\Omega_S-\pi J_{IS})t]$.

2.4. Application of NMR Spectroscopy in Protein-Ligand Interactions

Study of the dynamics processes of protein-ligand interaction can improve our understanding of their mechanism. E.g. Kinetic and thermodynamic parameters, dissociation constants and conformational changes. Surface plasmon resonance, isothermal titration calorimetry, circular dichroism, fluorescence polarization assay, fluorescence resonance energy transfer, quartz crystal microbalance, ultra/visible absorption spectroscopy, enzyme linked immunosorbent assay, differential scanning fluorimetry, analytical ultracentrifugation, microscale thermophoresis, small angle X-ray scattering, atomic force microscopy, electron ionization mass spectroscopy, and NMR spectroscopy are some of the methods used to study protein-ligand interactions.

NMR spectroscopy provides site-specific quantitative information and is the most powerful technique among the above mentioned techniques for investigating the dynamics of the protein-ligand interactions [5]. NMR approaches can be applied either in fast exchange, or in intermediate and slow exchange regime. First category uses the population-averaged NMR

observable quantities of the free and bound states of receptor or ligand and includes population averaging of chemical shift, longitudinal relaxation rate R_1, and transverse relaxation rate R_2. This method is suitable for studying very weak binding. The second category of NMR experiments uses the relaxation dispersion experiment or ZZ-Exchange spectroscopy to analyze micromolar or sub-micromolar dissociation constant and not very weak interactions. Appropriate approach should be chosen according to the chemical shift regime [6].

Most NMR studies of protein-ligand interactions are described by a two-state exchange model:

$$A + B \underset{k_{off}}{\overset{[B]k_{on}}{\rightleftharpoons}} A:B \tag{2.77}$$

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[A][B]}{[A:B]} \tag{2.78}$$

$$k_{ex} = [B]k_{on} + k_{off} \tag{2.79}$$

A, is the protein, B is the ligand, A:B is the complex. k_{on} and k_{off} are association and dissociation rates. K_D is the dissociation constant. K_D <0.1 μM is considered strong interactions and K_D >10 μM is considered weak. k_{ex} is the exchange rate in units of s^{-1} and varies by changing the concentration. k_{ex} and Δω (the chemical shift difference between the free and bound states), identify the chemical exchange regime of the interaction. If $k_{ex} \gg \Delta\omega$ the interaction is considered in the fast exchange regime, if $k_{ex} \ll \Delta\omega$ the interaction is considered in the slow exchange regime, and if $k_{ex} \approx \Delta\omega$ the interaction is considered in the medium exchange regime.

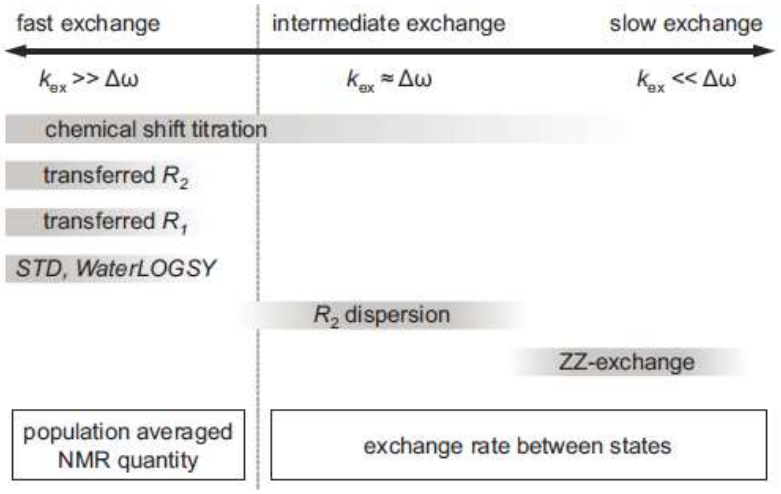


Figure 2.6 NMR methods
for quantitative analysis of protein-ligand interactions with their chemical shift timescales [6].

Transferred R_1, transferred R_2, chemical shift titration, saturation transfer difference (STD) and water-ligand observed via gradient spectroscopy (Water-LOGSY) are methods used in the fast exchange regime.

In transferred R_1 method, relative changes in the R_1 relaxation rate of a ligand is measured in free and bound state. And the longitudinal relaxation rate R_{1av} is measured by population-averaging the free and bound states:

$$R_{1av} = \rho_F R_{1F} + \rho_B R_{1B} \tag{2.80}$$

Transferred R_2 is similar to transferred R_1 method but more sensitive and is applied only when the interaction is sufficiently fast.

In the fast exchange regime, saturation transfer difference (STD) and water-ligand observed via gradient spectroscopy (Water-LOGSY) are the most commonly NMR methods used to quantitatively analyze protein-ligand interactions. Because in these methods the resonance signals of ligand are observed, they are also called *ligand-based or ligand-observed* NMR techniques.

Some other major ligand-observed NMR methods are transferred-NOE spectroscopy (NOESY), transient transferred NOE spectroscopy, and cross saturation [7, 8].

In the ligand-observed approach, a spectrum of free ligand is acquired and then the protein is added. In protein-observed approach, the spectrum of the free protein is obtained and then the ligand is titrated into the protein sample. These two approaches are complementary and ideally, use of both together gives a complete picture of the binding.

Although in theory all NMR spectroscopic parameters can be used as measures of protein-ligand binding, in practice only the parameters that can be easily obtained and with high sensitivity are of interest, e.g. the changes in chemical shifts, relaxation times, diffusion constants, NOEs, or saturation exchange [8].

In protein-observed methods, normally the parameter of interest is chemical shift changes of the target protein upon ligand binding. The 3D structure of the complex can be identified by heteronuclear experiments of isotopically labelled protein. However, this approach is size limited, requires highly stable and soluble protein, and is time consuming [7].

2.4.1. Ligand-Observed Methods

Because the protein NMR signals are not observed, the ligand-observed methods are not limited by the size of the protein. Large proteins (>50kD) can be studied via these techniques. These techniques require less protein (micromolar range) without needing isotope labeling. Ligand-observed methods can probe weak interactions and are used in obtaining information about affinities, specificity and 3D structures of protein-ligand complexes. One type of ligand observed methods is transferred NOE.

NOE is an effective tool in determining the 3D structure of molecules in solution and in the structural analysis of proteins. Upon binding of ligand to protein, NOEs drastically change and transferred NOEs are observed. These changes are used to detect and characterize the binding. The observation of trNOEs is based on different tumbling times τ_c of molecules in the free and bound states. Molecules with low or medium MW (<1000-2000) have short correlation time τ_c, and depending on their MW and shape and the field strength, they have positive/zero/very small negative NOEs. Large molecules however have strong negative NOEs. When a ligand is bound to a large receptor it adopts the receptor NOE behavior and shows strong negative NOEs, called *trNOEs*. TrNOEs reflect the bound conformation of the ligand thus the sign and size of it characterizes the binding. Another discriminating factor between trNOEs and ligand NOEs is the build-up rate, i.e. the time required to achieve maximum intensity. For trNOEs, this value is in the range of 50 to 100 ms however for free molecules it is four to ten times as long.

Intermolecular trNOE allow the determination of bound-ligand conformations thus intermolecular trNOEs can be used to determine the orientation of bound ligands in binding pocket of the receptor.

Some experimental schemes that rely on the trNOEs include: Saturation Transfer Difference (STD) NMR Spectroscopy Experiments, Water-LOGSY, Cross Saturation, Transient Transferred NOE.

Saturation Transfer Difference (STD) NMR Spectroscopy Experiments: If a ligand has two different signals because of a slow exchange between the bound and free states, by irradiating one of the ligand states, for example the free ligand signals, the signals of the other ligand state, here the bound ligand may be identified as a result of a transfer of saturation. In a method developed based on the transfer of saturation, saturation is transferred from the protein to the bound ligand. Dissociation of the ligands will transfer this saturation to the solution where the free ligands give

resonance signals with narrow line width. Then the off resonance spectrum is collected and finally subtracting the spectrum in which the protein is saturated (on-resonance spectrum) from the one without protein saturation (off-resonance spectrum) gives only the signals that are from the bound ligands and all molecules without binding activity are cancelled out [8, 9].

The STD technique can be combined with any NMR pulse sequence giving powerful NMR experiments such as STD TOCSY, STD HSQC, etc. Another significant advantage of STD NMR spectroscopy is that it makes it possible to investigate the binding to membrane-bound proteins.

STD can determine binding kinetics and binding constants of ligand-protein in both cases of one primary ligand and one primary ligand plus a secondary ligand which can have three schemes: independent binding sites for the primary and secondary ligands, allosteric inhibition and allosteric enhancement of the secondary ligand.

Water-LOGSY (water-ligand observation with gradient spectroscopy): A variation of STD NMR spectroscopy uses the bulk water to detect the protein-ligand interactions. In Water-LOGSY the ^{1}H resonance of bulk water are excited and magnetization transfers from the bound water to the bound ligand.

Intermolecular water-ligand interactions create a negative NOE peak. Just like STD NMR, on-resonance and off-resonance spectra are obtained. One approach is using steady state NOE experiments in which on-resonance saturation is applied at the water chemical shift and the off-resonance is applied outside of any ligand /protein resonances. Then the two spectra are subtracted and if the binding exists, a negative NOE is observed.

Cross Saturation/Transferred-Cross Saturation: Transferred cross-saturation is an extension of cross-saturation and is used in finding the location of the interface between protein and ligand in large complexes (>150 kD) with low binding affinity. In cross saturation one protein is specifically isotope labelled. Utilizing the steady state NOE-difference experiment mapping the interaction sites in the complex is done. In this method if the two components are bound to each other, after irradiating the complex, only the labelled protein is affected. Saturated ^{1}H resonance signals of unlabelled components leads to observation of magnetization of NH protons of the second deuterated and ^{15}N-labelled protein. Saturation is further transferred on to the binding area of the labelled protein via cross-saturation. ^{1}H- ^{15}N HSQC spectra are recorded before and after irradiation. From these spectra, the intensity of the residues of the labelled protein located at the binding interface reduces [10].

Transient Transferred NOE: In transient NOE experiments high frequency pulses are used to generate a nonequilibrium state which will go back to equilibrium by relaxation in a mixing time period during which NOEs are generated. During the mixing time NOEs/trNOEs build up to a maximum value and then because of T_1 relaxation, decrease to zero. Small molecules show positive NOEs and large molecules show negative NOEs which build up during mixing time. Two facts than can discriminate between ligands with binding activity from those without, is the difference in the sign of the trNOEs and NOEs and the difference times of reaching the maximum enhancement.

2.4.2. Protein-Observed Methods

Binding of a ligand to a protein changes the chemical shifts of the both molecules. Proton resonance signals can be affected by the local electron density, spatial proximity of groups with magnetic susceptibility anisotropies (such as aromatic rings), non-covalent interactions with

ligand and solvent, etc. The largest affected protein nuclei are the ones located in the binding pocket. One of the most common approached to study the protein-ligand interaction is the chemical shift mapping (CSM)/chemical shift perturbation (CSP)/complexation induced changes in chemical shifts (CIS). In chemical shift titration method, normally an unlabeled ligand is titrated into an isotope labeled (^{15}N and/or ^{13}C) protein and a 1D or 2D NMR spectrum is obtained after each titration to map a ligand binding site on a target and estimate K_D. In theory a 2D homonuclear NMR experiment, such as TOCSY, can provide a complete ^{1}H NMR chemical shift map for the bound and free states and then the two can be compared. However in practice, for protein protons, signal overlap and difficulty in identifying the amino acids involved in the interaction results in ambiguous assignments. Thus, heteronuclear correlation NMR spectroscopy experiments are used. For ligand protons however signals are less overlapping and spectra is less complex thus the changes of ligand proton resonance signals can be used to study the binding. One of common approaches is determining the dissociation constant from titration experiments [8].

One of the most common experiments used for the chemical shift mapping is ^{15}N HSQC. For this purpose, proteins need to be uniformly labelled by ^{15}N. Shifting of the resonance position of cross peaks in the spectra of the free protein as the reference versus the ligand bound protein samples, is indicative of binding. HSQCs are recorded during the titration and displaced peaks in the overlaid spectra shows the binding. Being involved in the binding is not the only cause of peak shifts. Conformational changes also can lead to peak shifts. Although there is no direct way to distinguish the two sources, peak shifts that are due to a conformational change are usually observed in a region of protein which is buried inside or away from the interaction site [10].

Chemical shift titration can be used to find the dissociation constant of the weakly bound ligands and provides information about the binding strength. It is also used to find the binding site on the protein thus, it is very useful in ligand screening. Peak shifting pattern can help identifying

if the ligand binds to multiple binding sites with different affinities. In this case, peaks in ^{15}N HSQC spectrum shift linearly until the primary (stronger) binding site is saturated by the ligand and then changes the direction during the binding of the ligand to the secondary (weaker) binding site, producing a nonlinear pattern [10].

Chemical shift mapping is limited by the size of the protein since for larger proteins, slower molecular tumbling will cause faster transverse relaxation and the linewidth will increase. Thus ^{15}N HSQC does not have good qualities for proteins beyond 40-50 kDa [10]. However, several techniques have been developed to push the size limit of protein, such as transverse relaxation optimized spectroscopy (TROSY), deuteration, stereoarray isotope labeling (SAIL), direct ^{13}C detection or methyl-TROSY.

Another improvement to protein-observed approach is adding an inert, freely soluble paramagnetic agent to the protein solution so increase the relaxation (both T_1 and T_2) of protein. This method is known as s*olvent paramagnetic relaxation enhancements (sPREs)*. A bound ligand protects the binding site from sPREs and is detected by reduced sPREs thus changes in linewidth.

2.5. Limitations of NMR Spectroscopy

Dipolar coupling which is due to the interaction between magnetic dipole moments of two close nuclear spins, has different manifestation in solid versus solutions. In solids very strong dipolar coupling of spins results in multiple levels of resonance splitting hence why spectra of solids has low resolution. In solution however due to fast tumbling of molecules dipolar couplings cancel out and on the timescale that NMR spectra is recorded it average to zero. This is why NMR peaks of proteins in solution state are sharp normally with perfect Lorentzian line-shape. However, the presence of dipole-dipole interaction has large effect on the decay of spin

coherence described by transverse relaxation rate R_2. And this gives rise to the theoretical size limit of molecules that can be studies by solution NMR. Since coherence relaxes faster for large, slow tumbling molecules. One way to overcome this is *protein deuteration*, which due to its smaller dipole moment compared to proton, can greatly slow down the transverse relaxation. Moreover relaxation-optimized NMR experiments such as ^{1}H-^{15}N TROSY HSQC and methyl ^{1}H-^{13}C TROSY HMQC has been developed to push the size limit.

Another limitation of NMR spectroscopy is overlapping peaks in the NMR spectra of large proteins by which results in ambiguous assignments. In order to address this limitation, isotope-labelling schemes has improved.

CHAPTER III

MOLECULAR MODELING STUDIES OF PROTEIN-LIGAND COMPLEXES

3.1. Introduction

Molecular modeling is a powerful tool in drug discovery and development. Incorporating experimental feedback in molecular modeling results is necessary to validate the computationally obtained results and to get the optimum results. Experimental methods for determining the complex structures are difficult and expensive thus using computational methods is a helpful tool in predicting complex structures. We have incorporated molecular docking via Autodock as well as molecular dynamic via GROMACS to investigate the ligand-protein interactions.

Molecular docking is a computational tool to predict and rank the structure of the complexes formed between two or more molecules with known 3D structure. Binding modes and binding affinities of a complex can be predicted by molecular docking. Molecular dockings can be applied in molecular recognition studies including enzyme-substrate, drug-nuclei acid, protein-nucleic acid, protein-protein, and protein-drug which is of great importance in structure-based drug design and is the focus of this study.

42 years ago the first molecular dynamics simulation of a macromolecule (pancreatic trypsin inhibitor-BPTI) was published. Since then molecular dynamics studies has been improved

greatly and now there are thousands of publications on molecular dynamics. Molecular dynamic give details of individual particles motions over time and can answer questions regarding the system that are otherwise hard to address via experiments. Comparing the simulation and experimental data can then be used to test the accuracy of the results and provide a gauge for improving the method. Improvements in the methodology and computer speed has extended the application of molecular dynamics to larger systems, greater conformational changes, and longer time scales. This can make it possible to get information beyond the access of experiments.

In this chapter we review molecular dynamic tools, specifically GROMACS software and molecular docking tools, with the emphasize on AutoDock software.

3.2. Molecular Docking Tools

3.2.1. Introduction

Study of ligand-receptor alignments and their binding sites traces back to 1982 when Kuntz et al. developed a geometric approach to study protein ligand interactions of heme-myoglobin and thyroid hormone analogs with Prealbumin [11]. As mentioned, molecular docking, especially protein-ligand docking has an important role in drug discovery and development. In the initial applications of molecular docking, both the ligand and the receptor were assumed rigid bodies, thus only considering 6 degrees of translational and rotational freedom. New advancements lead to models in which ligand is assumed flexible and protein is assumed rigid. Some methods have been developed that consider both ligand and protein flexible. Flexibility assumption can improve affinity, binding site and binding orientation and gives more compatible results with experiment results, X-ray, etc. Lock and key theory has given its place to the models that implement the receptor/ligand flexibility assumptions. In current models instead of assuming one dominant and more stable conformation, a protein is described as an ensemble of

several differently populated conformations in equilibrium. However, the highest unbound populated conformation is not usually the same as the highest populated in the complex.

Docking protocols are a combination of *sampling* (also called *search/pose*) and *scoring algorithms*. In sampling, ligand-binding conformations in binding site of a protein are predicted. In scoring binding strength for these binding conformations are predicted and ranked with a physical or empirical energy function. The conformation with the lowest energy score will be predicted as the *binding mode*. The performance of most docking tools depends on the characteristics of the binding partners and there is no universal best algorithm/scoring function. The number of degrees of freedom (related to flexibility) in the conformational search is determinant of the searching efficacy. Molecular docking can be done fast and on large data and give reasonable solutions, which can be considered as a starting point for further accurate analysis.

3.2.2. Sampling Algorithms

Sampling algorithms can be divided into two main categories. Flexible ligand-search and Flexible protein-search.

Flexible Ligand-Search: Flexible ligand-search algorithms include shape matching methods, systematic methods, random/stochastic, and simulation methods.

Shape matching method is one of the simplest sampling/search algorithms. The initial placement of the ligand in the binding site of the protein is based on molecular surfaces being complement of each other. Six degrees of freedom (three translational and three rotational) is assumed for the ligand. DOCK, FRED, Ligand Fit, and FLOG are some the docking programs that implement this method. However since the ligand is normally conformationally fixed during this process, shape matching method takes an ensemble of pre-generated ligand conformations to dock into protein and then the docked poses will be ranked according to their energy scores.

Systematic docking algorithms, generates all possible ligand binding conformations by exploring all the degrees of freedom in ligand, can be further divided into 1) conformational/ exhausive search methods, 2) fragmentation search methods, and 3) database or conformation ensemble search methods.

In conformational/exhausive search methods, all rotatable bonds rotate 360° with a fixed increment until all the possible combinations have been probed. One drawback of this method that makes its applications limited is "combinatorial explosion", which is the dramatic increase of the number of generated structures with number of rotatable bonds. Thus, normally geometric/chemical constraints are applied and further optimizations takes place to make the docking practical. Glide and FRED use this method.

In fragmentation method, ligand incrementally grows into the active site. This can be done via "the place-and-join" approach in which rigid ligand fragments are docked into the active site and covalently linked to make a ligand structure. It also can be done via dividing the ligand into a rigid core and flexible sites. Initially the core and subsequently the flexible sites are added. FlexX, DOCK, ADAM, Hammerhead, eHiTs, and LUDI use this.

In database method, libraries of pre-generated conformations (conformation ensembles) are rigidly docked and then ligand binding modes from different docking runs are ranked according to their binding energies. FLOG, phDOCK, MDock, Q-Dock, MS-DOCK use this method.

In *random/stochastic algorithms* random changes are applied to ligand(s). The changes will be accepted/rejected based on a predefined probability function. Four methods based on random algorithm are: Monte Carlo methods (MC), Genetic Algorithm methods (GA) or Evolutionary Algorithms (EAs), and Tabu.

In Monte Carlo methods accepting/rejection criteria is based on Boltzmann probability function.

$$P \sim \exp\left[\frac{-(E_1 - E_0)}{k_B T}\right] \tag{3.1}$$

E_0 and E_1 are energy scores of ligand before and after the random change, k_B is the Boltzmann constant and T is the absolute temperature of the system. DockVision, ICM, MCDOCK, QXP, Prodock use this method.

Genetic algorithms search for the most favorable binding mode is based on the idea of evolutionary process in biological systems. Starting from an initial population of conformations, by altering the variables that are indicative of translation, orientation, and conformation (genes) through applying genetic operators (mutations, crossovers, migrations) a final population of conformations that optimize a predefined fitness function will be obtained eventually. Autodock, GOLD, DIVALI, MolDock, PSI-DOCK, EADock and DARWIN.

Tabu Search prevents the search from revisiting conformations that have been previously explored. RMSD of new ligand binding conformations is compared with the previous conformations and if the RMSD between the new and previously recorded conformations is less than a cutoff, the new conformations will be rejected. PRO_LEADS, and PSI_Dock use this method.

Simulation methods, which can be divided into molecular dynamics (MD), and pure energy minimization (EM) methods, find the solutions to Newton's equations of motion.

Energy minimization methods can be direct searches, gradient methods (e.g. steepest descend), conjugate-gradient methods, second derivative methods, and least squares methods. Autodock uses energy minimization methods.

Flexible Receptor-Search: In order to have more accurate docking predictions, protein movements upon binding of the ligand should be accounted for. The basic principles here are same as flexible ligand search but with increased dimensions and search space. Assumption of fully flexible protein is computationally expensive. Thus in some methods limited protein motion

has been considered, or an ensemble of protein conformations has been addressed instead of a single conformation.

Methods that used to account for protein flexibility include: soft docking, side-chain flexibility, molecular relaxation, and protein ensemble docking [12].

Softening the interatomic van der Waals interactions, *soft docking,* allows for some overlap between the ligand and the protein normally by using tolerant scoring function called *soft core potentials*. This method is computationally efficient, fast, easy to implement, and can detect subtle conformational changes on the receptor. However they only partially introduce flexibility.

Side-chain flexibility keeps the backbones fixed and samples the side-chain conformations.

Molecular relaxation initially docks the rigid-body of ligand into the binding site and then relaxes the surrounding protein backbone and side-chain residues. The formed complexes can then be minimized by Monte Carlo, Molecular Dynamics simulations, etc. Advantage of this method is allowing some backbone flexibility in addition to side-chain flexibility. However, in comparison to side-chain flexibility method, it is time consuming and more sensitive to the scoring functions and inaccuracies in the scoring function can cause artifacts.

Protein ensemble docking is the mostly used method and it includes using an ensemble of protein structures to represent various possible conformations. The energy grids from each experimentally determined protein structure is combined to create a weighted average energy grid. This method is implemented in AutoDock. However, the averaging nature of the method may rise some inaccuracy.

Applying a combination of different methods gives promising results.

3.2.3. Scoring Functions

Scoring functions, ΔG (in kcal/mol) are used to evaluate and rank the ligand conformations. Scoring function directly determines the accuracy of the algorithm. Two

important aspect of a scoring function is speed and accuracy. They can be categorized as 1) Force field-based, 2) Empirical, 3) Knowledge-based, and 4) Consensus.

In *force field-based* scoring functions, ligand binding energies are decomposed into sum of several terms that identify interaction energies between the binding partners and the internal energy of the ligand. These interaction energies can be van der Waals energies (with a Lennard-Jones potential), electrostatic energies (with a Coulomb potential), bond stretching/ bending/ torsional energies, etc. and apply a set of derived force-field parameters e.g. AMBER or CHARMM.

Some force field scoring functions include Autodock scoring functions (which are based on Amber force field), G- and D- Scores and GoldScore.

One of the major considerations in force field scoring is including the solvent effect. This can be achieved by using a distance-dependent dielectric constant, $\varepsilon(r_{ij})$. Some simplifications should be assumed to make scoring functions computationally less expensive. In a less computationally expensive approach, water can be treated as a continuum dielectric medium via implicit solvent models such as the Poisson-Boltzmann/Surface Area (PB/SA) or the Generalized-Born/Surface Area (GB/SA) models.

In *Empirical Scoring Functions* such as Bohm's, F-Score, and Chem-Score, binding energy score of a complex is approximated by a sum of uncorrelated empirical energy terms to reproduce experimentally determined data. E.g. VDW energy, hydrogen bonding, electrostatic interactions, entropy term, desolvation term, hydrophobic interactions, etc. [12].

$$\Delta G = \sum_{i} W_i \,.\, \Delta G_i \qquad (3.2)$$

Empirical scoring functions are much more computationally efficient compared to force field scoring functions. This is because of simple energy terms. However, because of fitting to known

binding affinities of training set, the general applicability of an empirical scoring function depends of the choice of training set, i.e. the number of complexes in the training set. GlideScore, LigScore, F-Score, X-Score, SCORE, LUDI, ChemScore are some examples of empirical scoring functions.

Knowledge-based scoring functions try to reproduce experimentally determined structures by taking the potential parameters directly from experimentally determined structural information. Compared to the last two scoring functions, the knowledge-based scoring functions are extracted from a large number of structures thus are relatively robust and general and have a good balance between accuracy and speed. Some challenges of knowledge-based scoring functions include extension of the pairwise interactions to many-body interactions (seen in hydrogen binding and directional interactions), accurate calculations of entropy, etc. Some examples are SMoG score, BLEEP, MScore, KScore, and DRUGScore.

Consensus scoring combines different scores from several scoring function to compensate for errors from individual scoring functions and get a more accurate solution. The main issue is choosing the combination rule such that the true binders can be discriminated from others. X-CScore and MultiScore are two example.

3.2.4. Challenges and Limitations of Molecular Docking

Some main challenges for protein-ligand docking include the treatment of protein flexibility, solvation effects and structural water molecules, and entropy of binding.

Flexibility of protein side-chain and backbone can affect characteristics of the binding site and binding conformation. Currently most docking methods account for ligand flexibility. However, protein flexibility is still a challenge. Huge number of degrees of freedom make it a

difficult task to account for protein flexibility. As mentioned earlier, several techniques to account for flexibility of receptors include soft docking, side-chain flexibility, molecular relaxation, and protein ensemble docking.

Solvation effect can affect the binding ability of a ligand. Solvation effect treatments have been included in many scoring functions. E.g. force field-based scoring methods employ a distant dependent dielectric constant, or empirical based scoring methods include desolvation energy term.

Regarding the treatment of water molecules, an implicit representation of the solvent is not enough and an explicit atomic level description of water is needed. This can be done in several ways. Typical molecular mechanical force field include water models such as 3-point water models (e.g. TIP3P, SPC, SPC/E), or even more complicated models such as TIP4P and TIP5P. In case of strongly-bound or conserved structural molecules (observed in a variety of X-ray structures), water molecule can be treated as an integral part of the protein/ligand for docking purposes.

Entropic effects have an important effect to binding energy. Due to complexity of calculations, the entropy contribution is mostly neglected. Entropy contribution can be due to the reduction of the degrees of freedom upon binding, changes in normal modes of the binding partners upon binding, protonation and deprotonation events and arrangement of water molecules upon binding. Several attempts has been done to account for entropy in various ways such as clustering or including entropy in scoring function [13].

Most docking programs have average accuracies of about 1.5-2 Angstrom with 70-80% success rates [14]. The major limiting factor is in the scoring functions. Shortcoming of score functions in employing accuracy and speed is the biggest drawback. Simplifications used in scoring functions to lower the computational expenses, result in the loss of accuracy. For example not accounting for entropy and electrostatic interactions. Some other sources of complexity include solvent effect and protein flexibility.

3.2.5. Comparison of Available Docking Softwares

Over the last two decades more than 60 academic/commercial docking tools has become available such as AutoDock, AutoDock Vina, FlexX, Ligand Fix, Glide, FRED, DOCK, GOLD, and Surflex. Performance docking tools generally varies for different targets. In addition, the properties, which examine the quality of docking tools, may differ. Such as the prediction of binding free energy, the accuracy of virtual screening, and pose qualities. The assumption of approximation levels of different docking tools varies too. Thus, it is very difficult to make generally applicable conclusions. The power of a docking tool lies in its ability to reproduce experimental (X-ray, NMR) poses and in the liability of their scoring functions in predicting the free energy. Many studies have compared different programs based on different gauges e.g. accuracy in reproducing the X-ray conformations, capacity to predict binding free energies from the best scored conformation, etc. AutoDock, GOLD and FlexX are the most popular docking programs.

In this study we have used AutoDock [15] and a newly designed version of AutoDock, AutoDock Vina [16]. AutoDock is a commonly used docking program developed by Morris & co-workers at the Scripps Research Institute [17, 18]. AutoDock 4.0 accounts for side chain flexibility on selected residues. Similar to GOLD, AutoDock creates ligand poses by using genetic algorithm. Autodock search algorithms include Monte Carlo Simulated Annealing algorithm, Genetic Algorithm, hybrid local search GA known as Lamarkian Genetic Algorithm (LGA).

In AutoDock, a force field based on the AMBER force field is implemented with a Lennard-Jones dispersion, a hydrogen bonding, coulomb electrostatic potential, an entropic, an intermolecular pairwise desolvation terms. Scaling factors for force field terms are empirically calibrated from a set of 30 known complexes. Docking results from AutoDock can be done with help of a visual interface called AutoDock Tools (ADT).

In the last decade, AutoDock was the most cited protein-ligand docking program with around 500 citations per year (Figure 3.1). AutoDock Vina predicted top ranking poses with best scores and predict poses with RMSD averaging from 1.5 to 2 Angstrom from experimental poses [19].

AutoDock uses an empirical scoring function and the optimizer is GA-based. But AutoDock Vina uses a knowledge-based scoring function with a Monte Carlo sampling technique and Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is used for local optimization. As a result of these modifications, both prediction accuracy and docking time is improved in AutoDock Vina [19]. AutoDock Vina was developed by Trott & Olson at the Scripps Research Institute, La Jolla, California [16]. It predicts more accurate binding modes and is also up to two orders of magnitude faster than AutoDock. Using new search and scoring algorithms, high computational efficiency and ability to use multiple CPUs, makes it a good candidate and a competitive alternative for virtual screening.

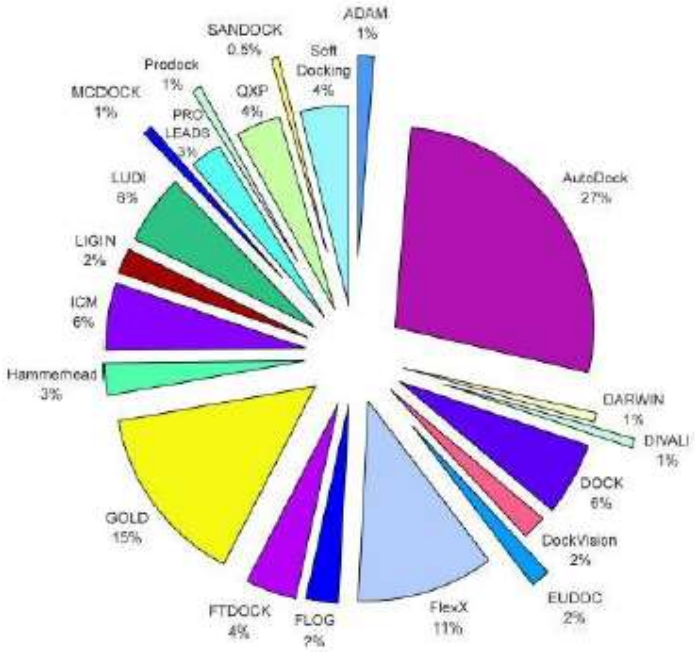


Figure 3.1 Number of citations for some of the most common docking programs [14]

3.3. Molecular Dynamic Simulations using GROMACS

3.3.1. Introduction

The software suite GROMACS (Groningen Machine for Chemical Simulation) was developed at the University of Groningen, in the early 1990s. GROMACS is a free, very fast, and reliable software for molecular dynamic simulations and energy minimization. GROMACS is written in C. It is compatible with all operating systems and makes parallel computing possible. The goal of molecular modeling is to understand and predict macroscopic physical properties based on detailed knowledge on an atomic scale. Macroscopic properties can be static (equilibrium) such as binding constants, or dynamic (non-equilibrium) properties, such as reaction kinetics. GROMACS is capable of investigating non-equilibrium dynamics. Macroscopic properties are ensemble averages over a representative statistical ensemble of system. Because single structure is not sufficient to compute thermodynamic equilibrium properties such as binding constants. Moreover averaging over the details that are not relevant for the macroscopic properties of interest simplifies the problem in hand. Although time-dependent Schrodinger equation describes the properties of the molecular system with high degree of accuracy, in practice any system more complex than the equilibrium state of a few atoms cannot be handled with such high accuracy and approximations are necessary. Two methods that are used to generate a representative equilibrium ensemble. Monte Carlo simulations and Molecular Dynamics simulations. MC simulations do not require the computation of forces thus are simpler than MD simulations. However, they do not yield significantly better statistics than MD. In cases where the forces are very large such as when the starting configuration is very far from equilibrium, MD simulations fail. Thus, a robust energy minimization is required in those cases.

3.3.2. Molecular Dynamics Simulations

A *system* in GROMACS is defined by its size, shape, number, and type of molecules, as well as coordinates and velocities of all atoms it contains. Systems are simulated as triclinic boxes with periodic boundary conditions, encompassing all possible constructs.

GROMACS supports GROMOS, Encad, OPLS-AA, Amber, CHARMM force fields. The forces are computed based on bonded and non-bonded interactions and special interactions.

The bonded interactions include harmonic, cubic, Morse, periodic, Lennard-Jones potentials, etc. Bonded interactions are defined by one or more atoms, an enumerated index that defines the interaction and a set of parameters such as bonds, angles, and torsion, position and distance restraints, etc. Instead of applying interaction functions, SHAKE, LINCS, or SETTLE (for water) algorithms can be applied to constrain bond lengths and angles to given values.

Non-bonded interactions consist of a Lennard-Jones 6-12 potential or a Buckingham exponential term potential and a Coulomb term. One can use arbitrary tabulated functions for non-bonded interactions. Long-range interactions can be addressed by employing cutoffs.

Special interactions can be defined to impose position, angle, and distance restraints on the system to add a penalty to the potential energy while some threshold is exceeded.

The influence of electrons in a molecular mechanics (MM) force field is expressed by empirical parameters from experimental data or from the results of high level quantum calculations. The accuracy of MM approximation is sufficient for the ground state processes of a covalent structure where the overall connectivity of atoms does not change. However, in case of chemical reactions or processes that involve multiple electronic states where the connectivity of atoms changes, electrons cannot be ignored and a quantum mechanical description is needed. One approach is to use a combination of quantum mechanics and molecular mechanics, QM/MM in which the reacting parts of the system are treated quantum mechanically and the rest of the system is modelled using the force field.

In studying the conformational dynamics of the macromolecules, for example in protein folding, it is important to get the longest possible time step without losing much accuracy.

One can find free energy of the system using GROMACS. Two types of free energy can be obtained from the simulations. One is a difference in Gibbs free energy between two thermodynamic states of a system, e.g. bound and free states of protein-ligand complex. We can get the binding constant of the ligand this way. The other type is a potential of mean force and includes a free energy profile as a function of reaction coordinate(s), which can be served as a restraint or constraint.

MD simulations finds the solutions of Newton's equations of motion for a system consisting of N interacting atoms:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i \, , i = 1 \ldots N \tag{3.3}$$

And

$$F_i = -\frac{\partial V}{\partial r_i} \tag{3.4}$$

The coordinates of the system is recorded to an output file at regular intervals while average temperature and pressure is kept constant at specific defined values. The recorded coordinates form the *trajectory* of the system. After enough time the system reaches an equilibrium state. Many macroscopic properties can be found by averaging over an equilibrium trajectory. Basic units used in GROMACS are given in Table 3.1. and a scheme of GROMACS scheme is given in Figure 3.2.

Table 3.1 Basic units used in GROMACS

Mass	u (unified atomic mass unit, ~1.66 e^{-27} kg)
Length	nm
Time	ps
Charge	e (elementary charge, ~1.60 e^{-19} C)
Temperature	K

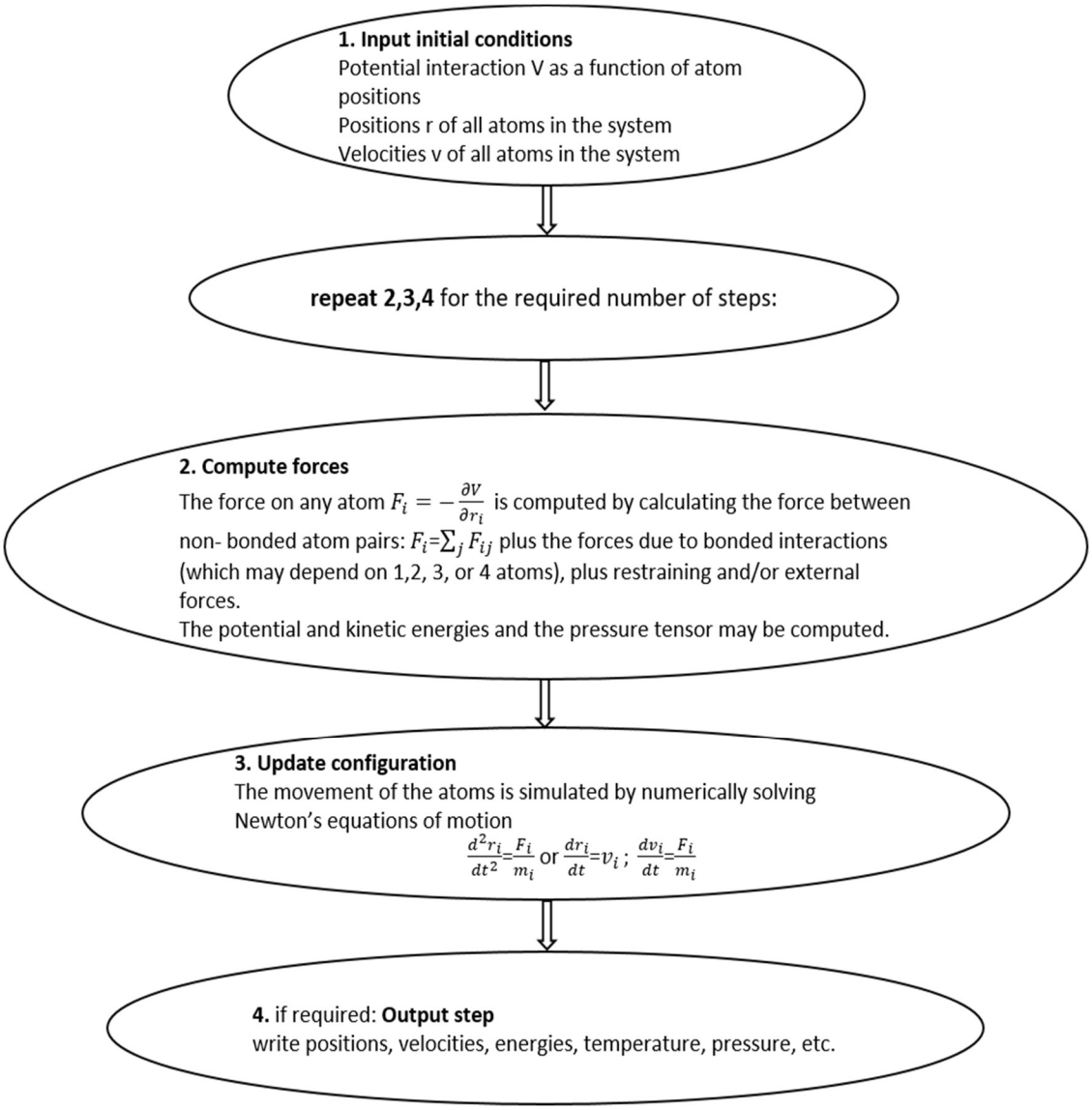


Figure 3.2 A global flow scheme for MD

GROMACS interfaces with quantum mechanical packages e.g. Gaussian and Mopac thus can perform mixed MM/QM simulations. It implements microcanonical Hamiltonian mechanics, stochastic dynamics (SD) including Langevin and Brownian dynamics and energy minimization. GROMACS allows various coupling methods to temperature and pressure baths, anisotropic pressures, triclinic box changes to address pressure tensor fluctuations. It is also possible to organize atoms in special groups and study their selective participation in dynamics or for energy calculations. User defined atom selections can be stored in *index files* by using *make_ndx* interactive program. This provides a very flexible way of performing operations on a subset of a system. Some example of such analysis which are used in investigating protein-ligand interactions include [20]:

- Protein secondary structure analysis
- Computing chemical shifts
- Calculating Relaxation times for molecular motion
- Ramachandran plots
- Dihedral angle calculations
- Analysis of energy components between default/user defined groups of atoms
- Root mean square deviations (RMSD) of a trajectory against itself, against another trajectory, or between a reference structure and a trajectory.
- RMSD of pairs of distances
- Analysis of hydrogen bond and hydrogen bond lifetime
- Analysis of formation/breaking salt bridges
- Analysis if bind length distributions
- Computing solvent-accessible surface of macromolecules

3.3.3. Algorithms

Periodic boundary conditions: Periodic boundary conditions are applied to minimize edge effects in a finite system. Space-filling boxes/unit cells of atoms is surrounded by copies of itself thus there are no boundaries. Periodic boundary conditions is combined with the *minimum image convention* in GROMACS. According to minimum image convention, only the nearest image of each particle is considered for short-range non-bonded interaction terms. The cut-off radius that is used to truncate non-bonded interactions cannot be greater than half of the shortest box vector. Because otherwise more than one image will be in the cut-off distance.

The Group Concept: In GROMACS certain actions are performed on user-defined groups of atoms. Each atom can only belong to 6 different groups. Possible groups can be temperature-coupling group, freeze group, accelerate group, energy monitor group, center of mass group, and compressed position output group.

3.3.4. Finding the Binding Free Energy

Molecular Mechanics can be combined with Poisson Boltzmann Surface Area (MM/PBSA) to estimate free energy of bonding of ligand to macromolecule. In this method, binding free energy of a protein-ligand complex is calculated as:

$$\Delta G_{bind} = G_{PL} - (G_P + G_L) \tag{3.5}$$

G is the free energy. G_{PL}, G_P and G_L are average free energies of the complex, the isolated protein, and isolated ligand respectively. The free energy G for each individual entity is given by:

$$G = E_{MM} - TS + G_{solvation} = \tag{3.6}$$

$$(E_{bond} + E_{nonbond} + G_{pol} + G_{nonpol}) - TS$$

where S is solute entropy and $E_{nonbond} = E_{vdw} + E_{elec}$ modelled with Coulomb and Lennard-Jones potential functions. Free energy of solvation, $G_{solvation}$, is the energy required to transfer a solute from vacuum into the solvent. Therefore,

$$\Delta G_{bind} = \Delta E_{bond} + \Delta E_{vdw} + \Delta E_{elec} + \Delta G_{pol} + \Delta G_{nonpol} - T\Delta S \qquad (3.7)$$

The first three terms are energy terms from bonded (including bond, angle, and dihedral), van der Waals, and electrostatic interactions. ΔG_{pol} and ΔG_{nonpol} are polar and non–polar contributions to the change in solvation free energy, T is the absolute temperature and ΔS is the change in solute entropy [21-23]. The polar solvation energy addresses the electrostatic interaction between the solute and the solvent. Three extra energy terms of cavitation, dispersion, and repulsion are needed to obtain solvation energies. These energy terms represent the cost of making a cavity in the solvent, the attractive and repulsive parts of van der Waals interactions between the solvent and salute. $G_{nonpol} = G_{cavity} + G_{vdw}$. G_{cavity} is the work done by the solute to create a cavity in the solvent. This term depends on the geometry of the solute. G_{vdw} is the attractive energy between the solvent and solute.

In order to calculate these MMPBSA energy terms we used GORMACS tool, g_mmpbsa [24].

3.3.5. Limitations of MD Simulation

Due to limitations of MD simulations, it is necessary to check the known experimental properties to test the accuracy of the simulations. Some of these limitations are as follows.

The MD simulations are classical: The fact that MD simulations used classical mechanics to solve Newton's equation of motion, does not cause a problem for most atoms at normal temperatures, however, there are exceptions. E.g. Hydrogen atoms and tunneling of proton through a potential barrier that might happen in the course of a transfer over a hydrogen bond or

the high frequency vibrations of covalent bonds that are quantum oscillators rather than classical harmonic oscillators.

Force fields are approximate: The GROMACS force fields which provide the forces, is pair-additive, they cannot incorporate polarizabilities and do not contain fine-tuning of bonded interactions.

Electrons are in the ground state: The electronic motions of the atoms are not considered in MD simulations, rather the force field is a function of the position of the atoms only and atoms remain in their ground state upon position changes (the *Born-Oppenheimer* approximation).

Long-range interactions are cut off: A cut-off radius is considered for the Lennard-Jones and Coulomb interactions. According to the minimum image convention of GROMACS, only one image of each particle in the periodic boundary condition is considered for a pair interaction. Thus, the cut off radius has to be less than half the box size.

Boundary conditions are unnatural: Periodic boundary conditions are used to avoid real phase boundaries. The GROMACS system size is small and in small systems the GROMACS boundary conditions is a more serious problem by enhancing the internal spatial correlation.

CHAPTER IV

SHETA2 AND MORTALIN INTERACTION

4.1. Introduction

Retinoids are natural or synthetic vitamin A derivatives with clinically proven anti-cancer activities. However, the toxicity of retinoids has limited their application *in vivo*. Attempts to improve anticancer activities of retinoids while reducing their toxicity have resulted into variations of retinoids. Heteroatom substitution on the ring in retinoids structure leads to generation of a new class called *heteroarotinoids* with dramatically lower toxicity compared to retinoids. Further modification led to generation of *flexible heteroarotinoids (Flex-Het)*. Flex-Hets contain a flexible linker between the heteroatom ring and aryl ring. A variety of Flex-Hets has been tested and among all, *N-(3,4-dihydro-2,2,4,4,-tetramethyl-2H-1-benzothiopyran-6-yl)-N'(4-nitrophenyl)thiourea*, also known as *SHetA2,* is the most promising compound with the highest anti-cancer activity. SHetA2 is effective against all of over 60 cancer cell lines in the US National Cancer Institute [25] and is a candidate for clinical trials for the treatment of ovarian cancer. It induces both intrinsic (mitochondrial mediated) and extrinsic (death receptor mediated) apoptosis pathways in cancer cells [26] and has low toxicity and differential activity against malignant versus normal cells [27, 28]. SHetA2 has been shown to interfere and interact with the activity of the heat shock protein HSPA9 (mortalin) [26]. Mortalin on the other hand interacts with the tumor protein p53 in a concentration-dependent manner [27] and inactivates p53. One

pathway for the compounds to inhibit the growth of cancer cells *in vitro* is by displacing p53 from mortalin. Binding of SHetA2 to mortalin interrupts p53-mortalin complex, allowing p53 migrates to the nucleus where it initiates apoptosis. However, interactions of mortalin with other client proteins, such as Bcl-2 and p66shc, may also impact cell apoptosis. Here we review the chemistry evolution, anti-cancer activity, and biological mechanism of heteroarotinoids, specifically SHetA2. Mortalin-SHetA2 interactions and its role in induction of apoptotic is reviewed next. Other mortalin-p53 inhibitors are reviewed briefly.

4.2. Review of Flex-Hets

4.2.1. Flex-Hets Chemistry Evolution

SHetA2 belongs to a relatively new chemical class, flexible heteroarotinoids (Flex-Hets), which have been evolved progressively from retinoids, to arotinoids, and to heteroarotinoids.

Retinoids are natural or synthetic vitamin A analogues/derivatives with at least one aromatic ring. The study on retinoids originates back from the exploration of Vitamin A metabolism [28]. Retinoids' mechanism of action is through their interaction with two types of nuclear receptors, the retinoic acid receptors (RARs) and the retinoid X receptors (RXRs). As the regulators of these retinoid receptors, retinoids have chemotherapeutic and chemopreventive properties in a variety of cancers in animal models and human cell lines. Retinoids regulate gene expression, control developmental, metabolic, and differentiation mechanisms, and induce growth inhibition and apoptosis in cancer cells in vivo and in vitro [29].

A retinoid molecule consists of a cyclic group, a polyene side chain, and a polar end group (Figure 4.1). Naturally occurring retinoids include Retinol (aka Vitamin A1), Retinal (Retinaldehyde), all-*trans*-retinoic acid (*t*-RA), 9-*cis*-retinoic acid (9-*cis*-RA), and 13-*cis*-retinoic

acid (13-*cis*-RA, 2). They are shown to be potential chemotherapeutic and chemopreventive agents [30-33].

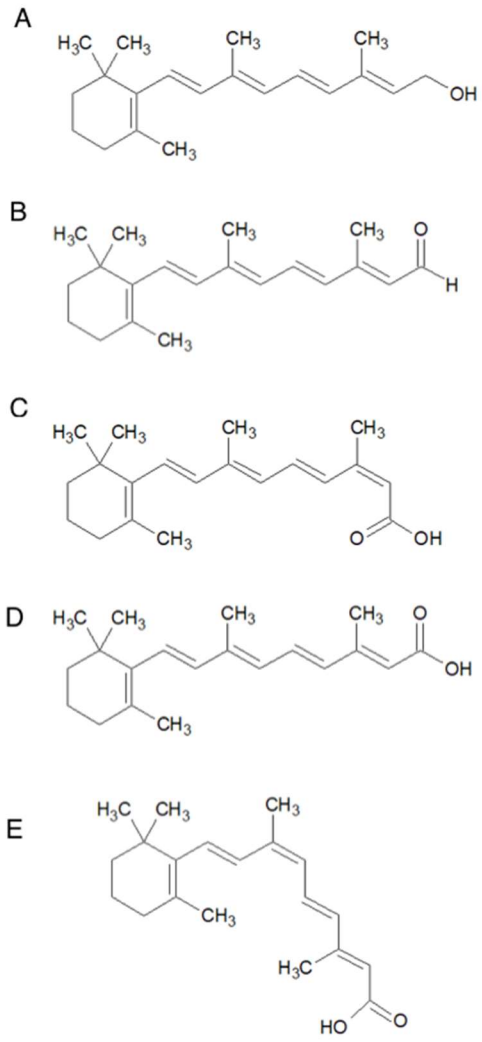


Figure 4.1 Natural retinoids.
(A) retinol/vitamin A, (B) retinal/retinaldehyde, (C) isotretinoin/13-cis-RA, (D) tretinoin/all-*trans*-retinoic acid (*t*-RA), and (E) alitretinoin/9-cis-RA

However, the clinical applications of retinoids are limited due to innate and acquired resistance in some malignant cells and significant side effects associated with their activation of nuclear receptors, such as teratogenicity and toxicity to skin, mucous membranes, gastrointestinal system, liver, kidneys, hair, eyes, endocrine system, and bone [34, 35]. Efforts to minimize

retinoids toxicity, to increase the potency (half-maximal inhibitory concentration IC_{50}), and to overcome resistance have led to design and synthesis of novel retinoids [36], such as 6-[3-(1-adamantyl)-4-hydroxyphenyl]-2-naphtalene carboxylic acid (AHPN/CD437), N-(4-hydroxyphenyl)retinamide (Fenretinide or 4-HPR), bexarotene (LGD1069), arotenoids, heteroarotenoids, and Flex-Hets. CD437 is a RARγ activator that showed significant apoptosis-inducing activity in a variety of cancer cells in vivo and in vitro [37, 38]. The synthetic amide of retinoic acid, Fenretinide, is a widely studied chemopreventive agent that inhibits the growth of several human cancer cell lines. Fenretinide induces apoptosis through both retinoid acid receptor-dependent and receptor-independent mechanisms [39]. It has good efficacy (maximal inhibition of cancer cell growth) compared with other retinoids, with no acute or severe toxicity and causes no liver function abnormalities. However, Fenretinide showed a few side effects in clinical trials including impaired dark adaptation (night blindness) resulted mainly from reduction of circulating retinol [40, 41]. LGD1069 is a synthetic derivative of 9-cis retinoic acid that showed more potency and less toxicity than its parent compound. However, the compound showed limited therapeutic efficacy [42, 43].

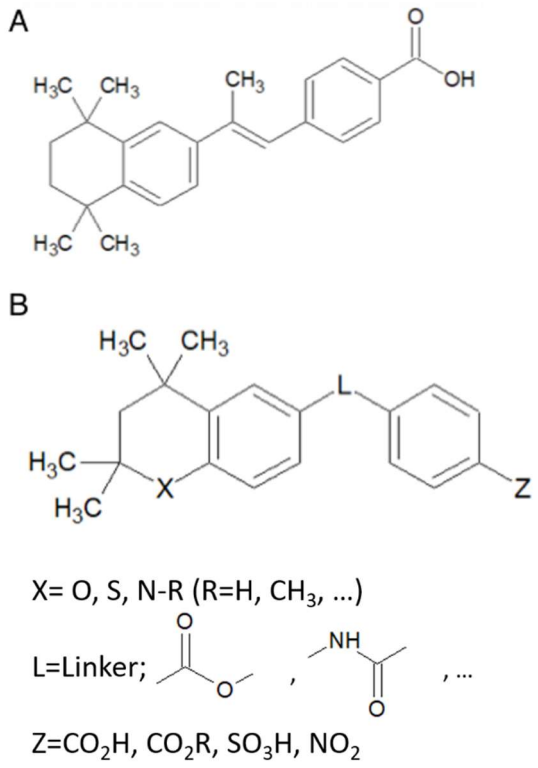


Figure 4.2 Synthetic arotinoid (A) TTNPB and (B) Heteroarotinoids

To improve retinoids, arotinoids (Figure 4.2A) were synthesized by adding an aromatic ring to constrain the retinoic acid double bonds of retinoids. The first arotinoid 4-[(E)-2-(5, 6, 7, 8-Tetrahydro-5, 5, 8, 8-tetramethyl-2-naphthalenyl)-1-propenyl]benzoic Acid known as TTNPB activates RARs instead of RXRs. Arotinoids have better anticancer activity than retinoids but intolerably higher toxicity [44, 45].

Further attempts were made to develop mimics of *trans*-retinoic acid and arotinoids with similar clinical effects but less toxicity and side effects. To reduce toxicity, oxygen or sulfur replaced a carbon atom in the tetrahydronaphthalene ring of TTNPB. The resulting compounds were named heteroarotinoids (Hets) (Figure 4.2B) [46]. Heteroarotinoids have structural relationship to diarylureas. Diarylurea and its derivatives are important pharmacophores with extensive application in design of anticancer compound [47].

The partially saturated aromatic ring can be five or six-membered, with at least one heteroatom (O, S, N, etc.). The heteroatom can block the benzylic oxidation of the parent arotinoids to toxic metabolites, reducing *in vivo* toxicity by up to 1,000-fold, comparable to the level of a natural retinoid, all-*trans* retinoid (*t*-RA) [46]. The linker between the heterocyclic ring and the aryl ring can be a two or more atoms moiety and it can be an amide, alkene ester, urea, thiourea, etc. Heteroarotinoids have been tested against various cancer cell lines such as kidney, lung, head and neck, and breast cancers and showed high activity and low toxicity [48-51]. Hets have similar anti-cancer activity as retinoids by activating RARs and inducing gene regulation [46, 52]. Arotinoids and *t*-RA only activate RAR receptors while 9-*c*-RA and some heteroarotinoids activate both RAR and RXR receptors [53, 54].

The specificity for retinoid receptors can be changed by alterations in the structure of the compound. Benbrook and colleagues. synthesized several heteroarotinoids and studied the effect of structural characteristics like heteroatom, ring size, number of aryl groups, and terminal side chain on the retinoid receptor [55]. In general six-membered ring had greater efficacy and specificity for RARβ than five-membered rings. Monoaryl compounds had lower RARα receptor specificity but greater efficacy. The effective concentration (EC_{50}) values of sulfur containing heteroarotinoids for RARα were lower in diaryl, whereas the opposite relationship was observed in oxygen containing compounds.

Dhar et al. synthesized three heteroarotinoids with a nitrogen heteroatom and a two atom C-O linker group [49] and studied their RAR and RXR activation, biological activity, and growth inhibition against cervix, vulvar, ovarian, and head/neck tumor cell lines. In comparison with the positive control, 4-HPR, the modified nitrogen heteroarotinoids had low apoptosis effect like *t*-RA and were classified as nonapoptotic retinoids.

Zacheis et al. evaluated growth inhibition activity of 14 heteroarotinoids against two HNSCC cell lines in vitro and compared their activities to that of 9-*c*-RA [48]. The results

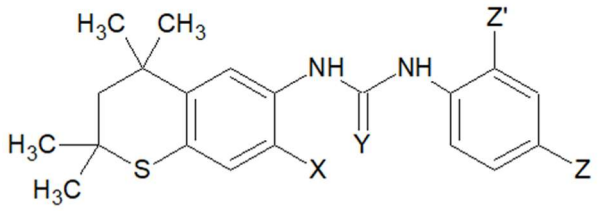
suggested that heteroarotinoid induced growth inhibition through the regulation of gene expression via the nuclear retinoic acid receptors. For most tested heteroarotinods, the tumor size had statistically significant reduction while no significant toxicity was observed.

Brown at al. ranked the antibacterial activity of 15 heteroarotinoids against *Mycobacterium bovis* according to their MIC values (defined as the lowest concentration of drug in μg/mL of drug that results in 99% reduction in the number of bacterial colonies compared to the control drug free plate) [56]. The result revealed the inhibitory ability of these first examples of heteroarotinoids against the growth of *M. bovis.*

4.2.2. Flex-Hets

Further exploration led to a more potent class of Hets called Flexible Heteroarotinoids (Flex-Hets) (Figure 4.3). Flex-Hets have a three-atom flexible linker of urea or thiourea linker group between the heterocyclic ring and the aryl ring.

In vitro experiments on cancer cells have shown that Flex-Hets regulate growth, differentiation and apoptosis, similar to conventional Hets. However, unlike conventional conformationally restricted two-atom linker Hets, Flex-Hets induce apoptosis independent of retinoic acid receptors RAR/RXR, hence not causing toxicity like skin irritation, genotoxicity, and teratogenicity [25, 51, 57-59].



X = H, CH_3
Y=O, S
Z=CO_2Et, SO_2NH_2, NO_2
Z'=H, CH_3

Figure 4.3 Structure of thiourea/urea containing heteroarotinoids, Flex-Hets [58]. SHetA2 (X=H, Y=S, Z= NO_2, Z'=H), SHetA3 (X=H, Y=S, Z= CO_2Et, Z'=H), and SHetA4 (X=H, Y=O, Z= CO_2Et, Z'=H).

Although Flex-Hets regulate growth and induce differential apoptosis in cancer cells, in normal cells they induce growth inhibition only [25]. Benbrook and colleagues [60] have studied the hypothesis that Flex-Hets induce differential apoptosis by directly targeting the mitochondria. Mitochondria is involved in cell growth and death. B-cell lymphoma 2 (Bcl2) family of proteins, which reside on the mitochondria outer membrane, are responsible for maintaining the integrity of mitochondria. This family consists of anti-apoptotic (Bcl-xl, Bcl-2, Bcl-w, Mcl-1, Bcl-b, A1) and pro-apoptotic (Bak, Bim, Bad, Bik, Bax, Bid, Bcl-xs, Noxa) protein members. The pro-apoptotic members of this family destabilize mitochondria by forming pores in mitochondria and inducing apoptosis. While anti-apoptotic members interrupt pore formation in mitochondria. Thus in order to regulate cell apoptosis the balance between members of Bcl-2 family is important. Mitochondrial mutations and disturbance of the balance between pro-apoptotic and anti-apoptotic mitochondria proteins could lead to mitochondria destabilization. Having unstable mitochondria due to their higher metabolism and mitochondria mutations, cancer cells are more susceptible to Flex-Hets than normal cells. Flex-Hets affects expression of Bcl2 family of proteins. They decrease expression of anti-apoptotic Bcl2 family members in outer mitochondria membrane in

cancer cell and increase their expression in normal cells in a time dependent manner. Meanwhile the pro-apoptotic expression levels are not changed as shown in ovarian, kidney, or lung cells lines [60-62]. The receptor-independent differential apoptosis induced by Flex-Hets occurs due to the loss of mitochondrial membrane integrity. In this process ROS (Reactive Oxygen Species) is generated, which consequently can alter membrane potential and release caspase activating factors such as cytochrome *c* [58, 60, 63].

The major concerns regarding the toxicity of any mitochondria-targeting drug are effect on the mitochondria in normal cells and increased ROS levels. Increase of ROS levels to levels exceeding the capacity of detoxifying enzymes can lead to ROS-induced cellular damage and cause potentially carcinogenic change in cells. A good drug candidate should have least effect on normal cells, and maximum effect on cancer cell. As mentioned, normal cell lines show more resistance to Flex-Hets than cancer cell lines.

4.2.3. Sulfur-Heteroatom Flex-Hets

Studies have been done in search of the most promising Flex-Hets. Several Flex-Hets with urea and/or thiourea linking groups and an ester or nitro group or sulfonamide group at the para position of the single phenyl ring were synthesized and studied [58].

Chun et al. performed a preclinical study in which the activity of six synthetic heterorenoids against eight HNSCC human cell lines were tested [63]. In this study, a sulfur containing heteroarotinoid with thiourea linker (SHetA2) was the most potent among the six Hets and natural RAs (all-*trans*- and 9-*cis*-RA). *In vitro* studies confirmed greater activity of compounds with a sulfur heteroatom as previously was shown in other studies [46, 50].

Studies have shown reduced toxicity of some sulfur heteroarotinoids is associated with activation of nuclear RARs and RXRs [46, 57]. Classification of some apoptotic rotinoids based on their receptors activity is given in Figure 4.4. Flex-Hets SHetA2, SHetA3, and SHetA4 (Figure

4.4A) show no activity towards nuclear retinoid receptors when used at doses up to 10 μM and they show much greater growth inhibitory effects towards cancer cells than normal or benign cells. SHet50(Figure 4.4B, with 2-atom amide linker) , however, is both RAR and RXR receptor pan-agonist. This induced differentiation and apoptosis was observed *in vitro* in cancer cell lines [57, 58].

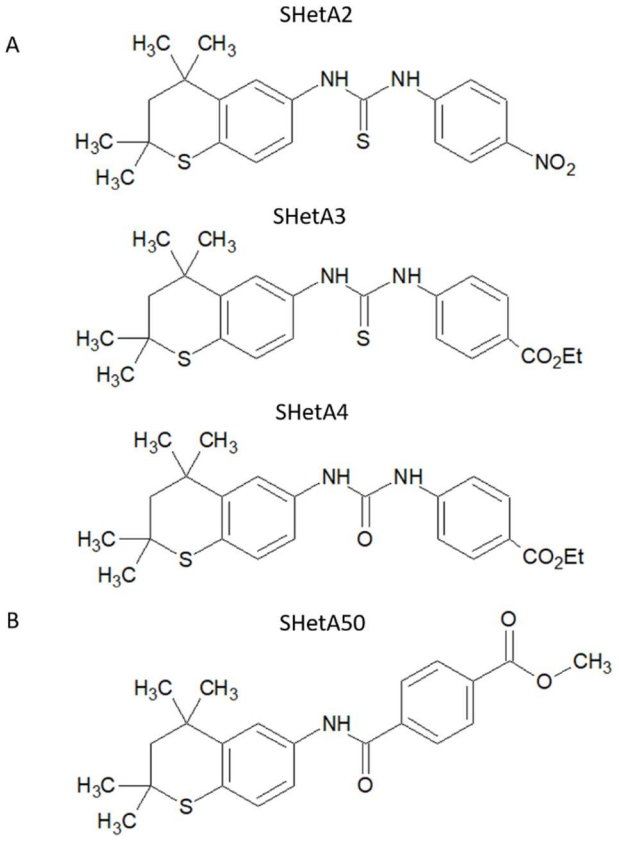


Figure 4.4 (A) Receptor-independent heteroarotinoid, SHetA2, SHetA3, and SHetA4. (B) Receptor-active heteroarotinoid, SHetA50.

Berlin et al. [58] synthesized sulfur-containing Flex-Hets with urea/thiourea linker and studied their growth inhibition effect against some ovarian cancer cell lines. They compared the efficacy and potency of these Flex-Hets with each other and with 4-HPR. Results showed that sulfur-containing heteroarotinoids have similar range of EC_{50} as 4-HPR and that Flex-Hets do not activate RARγ receptor. Moreover they have high growth inhibition of borderline and cancer cells

compared to normal cells, and enhanced the growth inhibitory ability of Flex-Hets against ovarian cancer cell lines compared to 4-HPR.

Le et al. reported that NO_2 group was also slightly more potent than CO_2Et, urea linker analogs were slightly more potent than thiourea linker compounds (e.g. SHetC2 with SHetA2), and a flexible thiochroman ring unit instead of planar quinolone, increases inhibitory activity and apoptosis significantly [64]. Compounds having nitro substitute on the phenyl ring, like SHetC2 and SHetA2, have greater growth inhibition than those having methyl or ethyl ester substitute like SHetA3, SHetA4, SHetD3, and SHetD4 (Figures 4.4 and 4.5) [62].

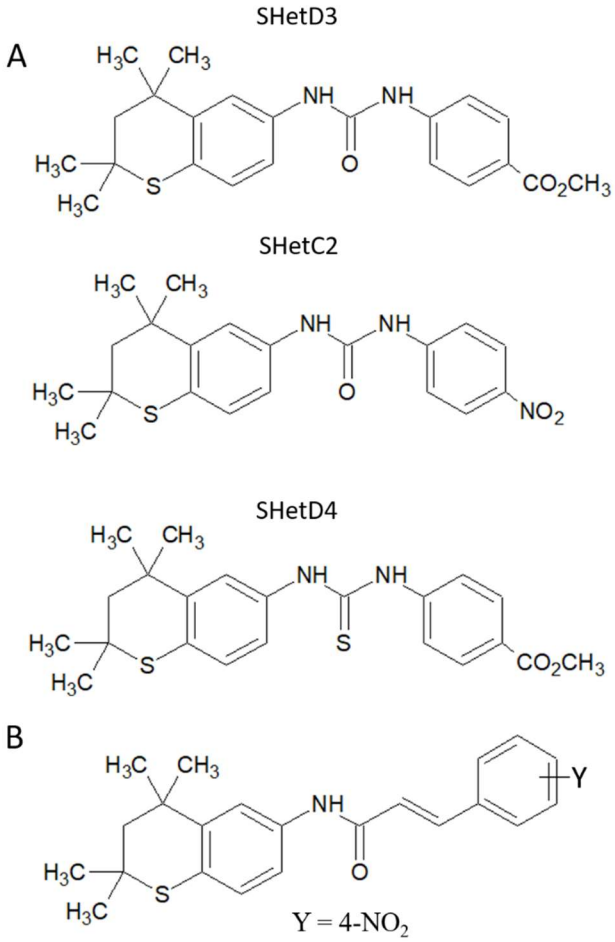


Figure 4.5 Structure of some urea and thiourea containing Flex-Hets

Gnanasekaran et al. [65] synthesized a series of 11 compounds with a 4-atom acrylamide linker (Figure 4.5B) and studied the biological activity against the A2780 ovarian cancer cell line. In general, electron-withdrawing groups on the linker, showed enhanced activity compared to electron donating groups. Among these new Flex-Hets, the most active candidate is still only at comparable level to SHetA2.

Among all the sulfur-hetroatom Flex-Hets, SHetA2 has exhibited the highest efficacy and potency without toxicity *in vivo* [25, 66] and strong chemopreventive activity in murine tumor and in human cell culture. SHetC2 has almost equal efficacy and potencies as SHetA2 for inhibiting the growth of renal cancer cells and inducing apoptosis. Therefore it is a reasonable alternative for SHetA2 [64]. SHetA2 showed growth inhibitory activities against all 60 cell lines in NCI human tumor panel at micromolar concentrations [25]. Animal studies and preclinical studies did not show evidence of any general toxicity, bone, skin or liver toxicity, teratogenicity, carcinogenicity, mutagenicity, and skin irritancy for SHetA2 [25]. Genetic toxicology assay showed SHetA2 is not mutagenic or genotoxic [67, 68]. Growth inhibition and apoptosis effects of SHetA2 has been studied in NSCLC cells both *in vitro* and *in vivo* [69], in organotypic cultures of ovarian cancer cell lines [57, 64], cervical cancer cell lines[25], in head and neck cell carcinoma [59, 63], lung cancer [61, 70], kidney cancer [62], colon and small intestinal cancer in $APC^{min/+}$ mice [71], etc.

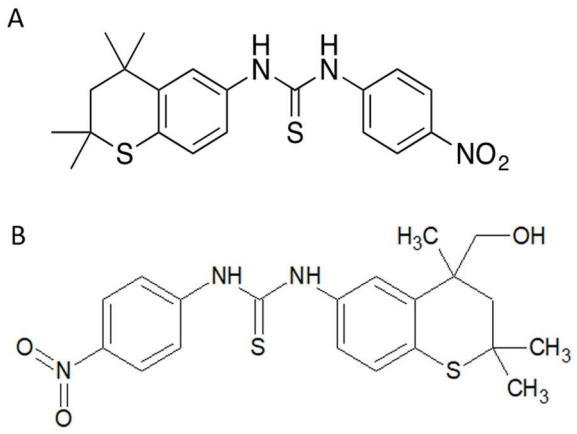


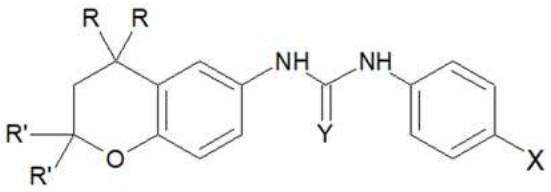
Figure 4.6 Structure of (A) SHetA2, (B) primary metabolite of SHetA2

Structural similarity of SHetA2 and its metabolite (Figure 4.6) to diarylureas, helped preliminary assessments of their binding using molecular docking (using Glide docking program) [72]. Two well-known diarylureas, Sorafenib and Imatinib, were employed as the reference and their kinase interactions was studied by molecular docking to B-Raf and KIT and also experimentally confirmed. Sorafenib and SHetA2 had similar orientation in binding site of B-Raf. Moreover, thiourea in Sorafenib and urea in SHetA2 formed similar H-binding with Glu501 and Asp549 of B-Raf.

Imatinib and the SHetA2 metabolite had the similar binding site in KIT 4 H-bonds with Asp810, and Glu640 and hydrophobic contact with hydrophobic binding pocket of KIT. The results suggest that SHetA2 and its metabolite have the potential to be inhibitor of B-Raf and KIT.

As we will mention, later on mortalin was identified as binding receptor for SHetA2.

4.2.4. Oxygen-Heteroatom Flex-Hets



R=H, Et
R'=H, Et
Y=S, O
X=NO_2, CO_2Et, CF_3, SO_2, CN, OCF_3

Figure 4.7 Oxygen analogs of SHetA2

In a recent study [73] and as we mention in chapter 7, in order to enhance aqueous solubility, oxygen analogs of SHetA2, with both urea and thiourea linkers, in presence of germinal dimethyl/diethyl in ring A, and different substituents on ring B with H-bonding capability (NO_2, CF_3, CN, OCF_3, CO_2Et, SO_2NH_2 have been studied (Figure 4.7). Their potency, efficacy against human A2780 ovarian cancer cells *in vitro*, and binding free energy values *in silico* are compared. A positive correlation between potency-binding free energy and a negative correlation between efficacy-binding free energy and potency-efficacy is observed. Among the tested compounds, the group of compounds with NO_2 substituent and urea linker, had efficacy values of 91-94% and IC_{50} values of 2.0-2.4 μM which is significantly better than the corresponding values for SHetA2 being 84% and 3.2 μM respectively. Compounds where docked into substrate binding domain of mortalin using Autodock 4.2 [74]. Docking results which was then confirmed by cellular dose-response data showed improved activity of 1) urea linker-containing compounds over thiourea counterparts, 2) germinal diethyl-containing compounds at C_2 position of ring A over compounds with germinal diethyl at C_4 or compounds with germinal diethyl at both C_4 and C_2 position of ring A, and 3) NO_2-containing compounds over CF_3 or other ring B substituents.

4.3. Biological Activities of Flex-Hets

As mentioned SHetA2 differentially induces apoptosis in cancer cells. Differential effect of SHetA2 has been observed in vivo, e.g. in ovarian cancer xenografts [25], in ovarian cancer cell lines versus normal HOSE cells [75], and in the Caki-1 kidney cancer cell lines versus normal kidney cells [62]. Resistance of mitochondria to SHetA2 in normal cells has been confirmed by lack of cardiac toxicity in treated wild type mice versus untreated mice at the highest dose used [71].

Study of the role of RNA and protein synthesis in SHetA2's mechanism of action has showed that SHetA2 works independently from regulation of gene expression [60]. Similar to other Flex-Hets, SHetA2 induces differential cell apoptosis independent of the retinoic acid receptor-signaling pathway and by directly targeting mitochondria [63].

4.4. Apoptosis Pathways

There are two major apoptosis pathways: extrinsic death receptor (DR)-induced pathway and intrinsic mitochondria-mediated pathway, which involve activation of caspase-8 and caspase-9 respectively [76]. Intrinsic and extrinsic pathways converge through Bid protein [77]. SHetA2 induces both intrinsic and extrinsic apoptosis [78].

Lin et al. [61] studied the effect of SHetA2 on expression levels of some apoptosis regulator proteins. They found that in Human NSCLC cells, SHetA2 down-regulates both c-FLIP (a major inhibitor of extrinsic apoptosis) and Survivin (a major inhibitor of intrinsic apoptosis) in all 6 testes cell lines and that it reduces the levels of several other apoptosis regulator proteins like XIAP, Bcl-2, Bcl-xl in some tested cell lines. Overexpression of c-FLIP makes cells resistant to SHetA2-induced apoptosis alone or combined with TRAIL. Overexpression of Survivin does not affect the sensitivity of cells to SHetA2 or SHetA2-TRAIL combination. Down-regulation of

Survivin and c-FLIP enhances SHetA2 and TRAIL-induced apoptosis with the later having more effect.

Myers and colleagues [66] demonstrated anti-angiogenic activity of SHetA2 through microarray analysis of ovarian cancer organotypic cultures and showed that Flex-Hets regulate expression of angiogenic cytokines. Thus by blocking cytokine release from cancer cells, SHetA2, inhibits angiogenesis. In this study, 44 genes were significantly regulated by SHetA2. SHetA2 reduces the expression of Thymidine Phosphorylase (anti-angiogenic growth factor) in cancer cells but not normal cells. Angiogenic inhibition of SHetA2 *in vivo* has been confirmed in Caki-1 renal cancer xenografts.

As mentioned above, SHetA2 can induce intrinsic apoptosis [63]. Alteration in the mitochondrial membrane potential, suppression of mitochondrial membrane permeability transition (MPT), and loss of mitochondria integrity enhances the release of caspase-activating factors like cytochrome c from mitochondria to the cytosol. The released cytochrome c interacts with cytosolic factors like Apaf-1 and dATP, thus activates caspase-9 and caspase3-like activity in a time dependent manner. SHetA2 treatment can cause mitochondria swelling and loss of mitochondria membrane potential, resulting in activation of intrinsic apoptotic pathways [79]. SHetA2 directly affects on mitochondria and reduces anti-apoptotic proteins, Bcl-2 and Bcl-xl in ovarian, kidney, lung cells lines, and in colon polyps [71] leading to apoptosis. In normal ovarian and endometrial cells, SHetA2 increases Bcl-x2 and Bcl-xl protein levels, blocking apoptosis in normal cells [60]. *In vitro* study showed that SHetA2 did not change Bax level in either normal or cancer cells [60]. Also SHetA2 binding to heat shock protein A (HSPA9, HSPA8, and HSPA5) [80] in both normal and cancer cells leads to cyclin D1 degradation thus G1 cell cycle arrest (with greater effect on cancer cells compared to normal cells) [62, 81, 82].

SHetA2 reduces the expression of Bcl-2 in cancer cells thus promotes apoptosis [60, 62] while mortalin can prevent this effect leading to inhibition of apoptosis. Thus Mortalin and

SHetA2 can regulate the Bcl-2 protein family and affect the stability of mitochondria. (Activation of Bax is in favor of apoptosis and activation of Bcl-2 inhibits initiation of apoptosis).

Interestingly, Liu and colleagues tested the hypothesis that SHetA2 induces mitochondrial effects through ROS generation and found that ROS generation is a consequence of Flex-Het action on mitochondria and not a cause of it [60, 62].

SHetA2 can activate DRLs (Death Receptor Ligands) and induce extrinsic caspase-8-dependent apoptosis by differentially sensitizing cancer cells while not affecting normal cells. DRLs (such as Tumor necrosis factor α (TNFα) and TNF-related apoptosis inducing ligand (TRAIL)) and death receptor activating antibodies have been used in cancer therapy. Compounds that can increase sensitivity of resistant cells to death receptor activating ligands without increasing toxicity are of great importance in cancer therapy. Clinical trials on TNFα failed due to its resistance and toxicity [83, 84]. In attempt to reduce the toxicity of TNFα, tumor-selective delivery of TNFα via SHetA2 has been studied. SHetA2 represses nuclear factor κB (NF-κB) activity [62, 82]. Some NF-κB inhibitors sensitize other types of cancer to death receptor ligands [66, 85]. Similarly SHetA2 sensitizes ovarian cancer cells to TNFα-induced apoptosis effects in a mechanism that involves repression of NF-κB [86, 87]. SHetA2 may also sensitize cancer cells to other death receptor ligands [69] .

Lin and colleagues [70] showed for the first time that SHetA2 induced a caspase-8-dependent apoptosis mechanism, in which SHetA2 induced CHOP expression and up-regulated death receptor 5 (DR5) in human lung cancer in cell cultures and in mice. DR5 induction is thus a key factor of SHetA2-iduced apoptosis or SHetA2-TRAIL-induced apoptosis.

Moxley et al. [86] studied the effect of chemotropic agents cisplatin, paclitaxel, and SHetA2 with TNFα and TRAIL on ovarian cell lines A2780 and SKOV3 and normal human endometrial cells. The result revealed normal and cancer cells are resistant to the death receptor ligands, TNFα and TRAIL. Combination of cisplatin and paclitaxel with TNFα and TRAIL did

not eliminate this resistance. However, dose responsive apoptosis effect of SHetA2 was enhanced when combined with TNFα and TRAIL. SHetA2 when combined with TNFα and TRAIL sensitized cancer cells and induced an extrinsic casapase-8 and -3 apoptotic pathways. Moreover, SHetA2 did not affect the sensitivity of normal cells. Simultaneous treatment of SHetA2 and TNFα not only induces caspase 8-dependent extrinsic apoptosis but also SHetA2 enhances caspase 9-dependent intrinsic apoptosis activity of SHETA2 as a single agent [78, 86].

SHetA2 induction of cell cycle arrest, can be effective in broad type of cancer cells, regardless of presence or absence of p53 in cells [82].

The National Cancer Institute's Rapid Access to Intervention Development (RAID) program (Application 196, Compound NSC 721689) and Rapid Access to Preventive Intervention Development (RAPID) program completed preclinical testing on SHeA2. Not showing toxicity effects in preclinical trials, SHetA2 is now in Phase-0 clinical trial.

4.5. Mortalin: a Receptor for SHetA2

Benbrook et al. [80] applied an affinity chromatography combined with mass spectroscopy analysis to identify SHetA2 target protein(s). They synthesized a metabolite of SHetA2 by replacing the methyl group on the heteroatom ring with a hydroxyl group. This metabolite was attached to a long linker which then was conjugated to NanoLink Amino-Magnetic Microspheres through an amide linkage. ShetA2-attached microspheres and free microspheres (negative control for non-specific binding) were incubated with the whole cell protein extracts from A2780 human ovarian cancer cell line. Microspheres were washed and then eluted with excess SHetA2. In SHetA2-conjugated microsphere, a band was observed around 75 kDa. The specific bands excised from dried SDS-PAGE gels of the microspheres with and without SHetA2 and straight eluents were passed to QStar and Shotgun Orbitrap mass spectrometers. Three heat shock protein family members, HSPA5, HSPA8, and HSPA9/mortalin

were identified in both mass spectroscopy analysis. Among which mortalin was significantly present in both experiments and was considered as the client protein for SHetA2. Moreover, western blot analysis of co-immunoprecipitates of mortalin with its client proteins in protein extracts from A2780 and SK-OV-3 human ovarian cancer cell lines revealed that SHetA2 binds to mortalin and disturbs the interaction of mortalin with p66shc and p53 in cancer cells. Protein p53 then translocates to mitochondria and nucleus, Bcl-2 degrades and p66shc is released and generates ROS. Thus, the SHetA2-Mortalin interaction affects intrinsic apoptosis induced by these molecules.

Mortalin is a 73,913-kDa protein with 679 amino acids and a member of the heat shock protein (HSP) family. Heat shock proteins are ubiquitous, highly conserved proteins that are expressed constitutively and located in all cell compartments [88]. The role of HSPs in cell homeostasis was first observed by F. Ritossa when he found the overexpression of proteins of 70 and 26 kD after exposing *Drosophila* larvae to heat stress [89]. In cancer cells, the conditions mimic stress thus HSPAs are upregulated.

According to their molecular weight, HSPs are classified into several subfamilies among which HSP70 is the most important HSP involved in protein folding and tumor progression. Therefore HSP70 proteins are good targets for cancer therapy [90]. HSP70 subfamily consists of at least eight homologous chaperone proteins. Among which six are mainly located in cytosol and nucleus, the other two reside in endoplasmic reticulum (Grp78) and mitochondria (Grp75).

Mortalin, is also known as glucose-regulating protein 75 (Grp75), mot-2, hmot-2, p66mot-1, mitochondria stress-70 protein (mtHSP70), HSPA9/HSPA9B, peptide-binding protein 74 (PBP74), CSA, Mot, Mot2, and MGC4500. Mortalin was first cloned as an HSP70 family member in cytoplasmic of normal fibroblasts from CD1-ICR mouse [91, 92]. It was first identified as a cellular mortality factor as it was found present in cytosolic fractions of mortal cells while being absent in cytosolic fractions of immortal cells [91]. Immunocytochemical studies of the antibody against mortalin showed the immunofluorescence in perinuclear region of

immortal cells and pancytoplasmic distribution of the protein in normal cells [93]. Thus, mortalin staining pattern is a sensitive marker of normal and cancerous cell types.

It was shown that mouse has two types of mortalins; mot-1 and mot-2 alleles, which are assigned to chromosome 18 [92, 94]. These two types have different subcellular localization, secondary structures, and contrast functions although they only differ in two amino acids in carboxy-terminus; mot-1 has V618 and R624, and mot-2 has M618 and G624 [95]. These two residues are determinant factors in identifying secondary structure of mot-1 and mot-2. Mot-1 is originally cloned from normal cells, has pancytosolic distribution and is associated with senescence in NIH 3T3 cells [96]. Mot-2 however, is originally cloned from in immortal cells with perinuclear distribution [93] and is associated with malignant transformation of NIH 3T3 cells [97]. Each of the two differentiating residues and three motif regions, thus the secondary structure of mot-1, can be essential in determining the ability to bind p53 in vivo [98]. Unlike mot-1, mot-2 binds to p53 in vivo, abrogates p53 nuclear translocation and inactivates p53 [98, 99].

Cloning of human mortalin cDNA from different cells showed that human mortalin has identical sequences with same functions as mot-2 in mouse [100]. Due to this similarity, it was first known as *hmot-2*.

Human mortalin is translated in the cytoplasm and transported to mitochondria. Although mortalin is primarily found in mitochondria, it has also been found in several other subcellular sites including endoplasmic reticulum (ER), Golgi apparatus, plasma membranes, cytoplasmic vesicles, and cytosol [88].

Mortalin is not a heat activated protein and can be induced by various stressors such as accumulation of unfolded or misfolded protein, low levels of ionizing radiation, glucose deprivation (GD), loss of calcium homeostasis, and metabolic stress [101-103]. Mortalin is also found overexpressed in several transformed and tumor cell lines [104, 105].

The three-dimensional structure of full-length mortalin is not available so far. Evolutional conservation suggests a similar structure for mortalin as other HSP70s. Mortalin has a conserved ~42 kD nucleotide-binding domain (NBD, ATPase domain), residues 1-443 and a ~25 kD substrate-binding domain (SBD, peptide-binding domain as PBD), residues 444-697, joined by a highly conserved protease sensitive linker connecting them (Figure 4.8). The NBD consists of subdomains I and II, each dividable into a and b regions. The SBD can be further divided into two subdomains, the ~13 kD β-sandwich domain (SBDβ), consisting of 2 sets of 4-stranded antiparallel beta sheets and the ~12 kD C-terminal helical 'lid' domain (SBDα) consisting of five helices (A-E). The substrate-binding site is a hydrophobic region between the two β-sheets forming a deep binding pocket for hydrophobic residues.

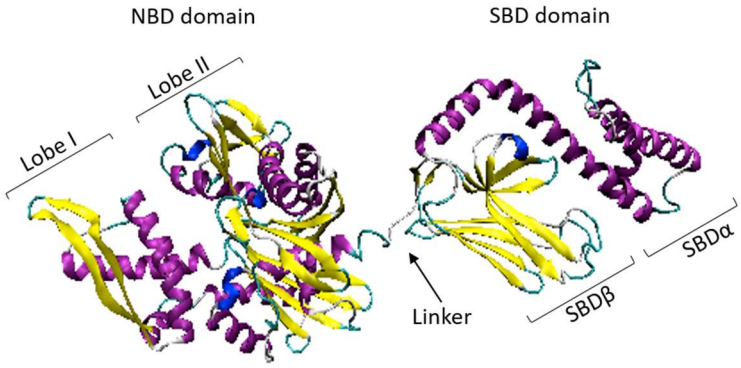


Figure 4.8 Structure model of full-length mortalin in ADP state constructed with I-TASSER server [106].
Known structures of NBD (PDB code 4HBO) and SBD (PDB code 3N8E) were used in construction of full length.

NBD and SBD do not interact in ADP-bound state. In ADP-state, the lid covers the substrate-binding pocket in SBD and it opens up on exchange of ADP for ATP. In ATP-bound state, however the two domains are coupled [107]. As observed from the X-ray on Hsp110 [108], NMR on *E.coli* DnaK [109], and NMR on *T. thermophiles* [110], ATP binding induces the opening of polypeptide binding channel, thus facilitates the release of the substrate from SBD and on the other hand, substrate binding to SBD stimulates the closing of polypeptide binding channel

and ATP hydrolysis [111] . Thus in an allosteric effect, binding of ATP improves the release of substrate while substrate binding increases the rate of ATP hydrolysis in NBD. Based on the narrow resonances in NMR TROSY spectrum, a flexible structure for the linker is suggested and it is assumed that linker region works like a switch and stimulates the ATPase activity [112]. ATP hydrolysis provides the energy required for protein folding. However, some misfolded proteins cannot be recovered and need to be degraded.

4.6. Protein p53 and its Interaction with Mortalin

The homotetramer p53 is known as the "guardian of the genome". It is a stress response factor that activates with intrinsic or extrinsic stress. Protein p53 consists of an amino-terminus transactivation domain (TAD residues 1-42), a proline rich domain (PRD 60-93), a sequence specific DNA binding domain (DBD or core domain CD residues ~100-300), and a C-terminal domain (p53CTD, residues 293-393) (Figure 4.9). CTD is the main site for protein-protein interactions. The p53CTD domain contains the tetramerization domain (TD residues 326-356) and the negative regulatory domain (p53NRD, residues ~361-393). Binding of proteins to p53NRD activates the DNA-binding of p53 [113]. The structure of full length P53 is not available.

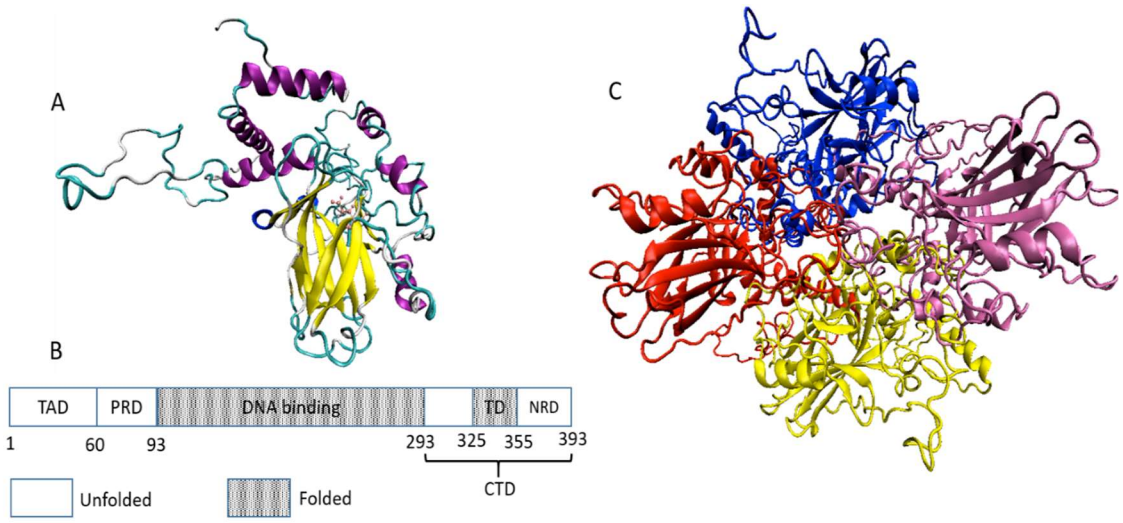


Figure 4.9 (A) Structure model of full-length p53 constructed with I-TASSER server. (B) Domain structure of p53. (C) Tetramer model of p53 obtained by aligning the predicted full-length model to the crystal structure (PDB code 2AC0) of four p53 DNA-Binding domains in complex with two DNA half-sites
[114, 115].

Protein p53 is a nuclear transcription factor and controls cell cycle. To prevent genetic instability p53 induces cell cycle arrest (G1 arrest) and apoptosis as a response to endogenous and exogenous stress signals, through transcriptional regulation or direct interaction with apoptotic proteins [116]. Through its interaction with Bcl2 and mortalin, p53 regulates mitochondrial membrane potential [117]. Protein p53 is inactive in more than half of human cancers. Inactivation of tumor suppressor p53 is involved in transformation and immortalization of cells and is the most common genetically abnormality in human cancer. Three main mechanisms responsible for inactivation of p53 are mutations of p53, post-translational modifications, and cytoplasmic sequestration/nuclear exclusion [118, 119].

Protein p53 is synthesized in the cytoplasm and translocated to the nucleus in order to perform its transcription factor function [120]. Cytoplasmic sequestration of p53 causes loss of the function and can lead to immortalization of cells and tumor resistance to radiotherapy and chemotherapy. The residues 323-337 of the carboxyl terminus of p53 were shown to be required for its cytoplasmic translocation from nucleus [121]. Cytoplasmic sequestration of p53 can be due

to overexpression of its binding proteins, such as Bcl-2, PML, Parc [122] or mortalin [123-127], and cytoskeleton proteins .

Mortalin and p53 have been shown to colocalize in perinuclear region in several cancer cells, e.g. NIH 3T3, MCF7, COS7, NT-2, A172, A2182, HeLa, Balb/3T3, YKG-1, SY-5Y, and U2OS. Kaul et al. [128] showed for the first time that overexpression of mot-2 in normal human diploid fibroblasts (MRC-5) leads to decreased p53 function, thus temporary escape from senescence [128]. Also in human foreskin fibroblasta (HFF5) overexpression of mortalin together with telomerase hTERT extended lifespan substantially [129] indicating reduced p53 function. Life span extension of human lung fibroblasts (MRC-5), increased malignancy of human cancer cells [104], malignant transformation of NIH3T3 cells in mice [97], attenuation of HL-60 leukemia cells by mot-2 [130] are some evidences that suggest the inactivation of p53 by overexpression of mortalin [97, 99, 128]. By interacting with the mitochondrial proteins Bcl2, mot-2, and HSP60, p53 regulates mitochondrial potential [117]. The stress-induced mitochondrial p53 localization is followed by changes in mitochondrial membrane potential, chromosome c release, procaspase-3 activation and apoptosis. The presence of mortalin-p53 complexes in mitochondria during p53-induced apoptosis suggests the transcriptionally independent role of p53 in apoptotic signaling [131]. Different western blot patterns for mutant and wild type p53, suggested that loss of p53 function in human diploid fibroblasts transfected by mot-2, is not caused by mutation [128]. It is reported by the same group [98] that by sequestering p53 in the cytoplasm, mot-2 inhibits transcriptional activation function of the protein in mouse and human transformed cells.

Regulatory function of p53 is independent of its localization and is performed by interaction of p53 with mortalin and Bcl2. Mortalin regulates the DNA-binding ability of p53 [132].

When binding to p53 in cancer cells, mortalin inhibits apoptotic properties of p53 and leads to survival of cancer cells [101]. It sequesters p53 in the cytoplasm fraction through

physical contact and blocks p53-mediated transcription, downregulates p53-target genes expression, e.g. $p21^{SD11/WAF}$, and enhances degradation of p53 by MDM2 in cytoplasm, disturbs the G1-associated nuclear translocation of p53, that leads to uncontrolled cell proliferation, a hallmark of cancer cells [101, 121, 125]. Mortalin breaks the p53-Bcl-xL/2 complex and causes proteasomal degradation pathway of p53 by its association with E3 ligase CHIP (Carboxy terminus of Hsc70 Interacting Protein) [133]. Moreover, mortalin–p53 interactions causes deregulation of centrosome duplication, which is an indicator for cancer. Overexpression of mortalin can reverse the p53-dependent suppression of centrosome duplication indicating the interaction was localized to centrosomes [134].

It is shown that mortalin-p53 interaction depends on the stress levels in cells. In unstressed or weakly stressed cells, p53 and mortalin do not interact [127]. Thus, knock down of mortalin does not induce apoptosis. In contrast, in stressed cancer cells, whether physiologically stressed or under the effect of stress-inducing chemicals, mortalin-p53 interaction exists and knock down of mortalin induces p53-mediated apoptosis.

Many studies has been performed to determine the binding sites of p53 and mortalin. However, the results do not identify the similar interaction sites. Carboxyl terminus of p53 was considered its cytoplasmic sequestration domain [99, 135]. This is in agreement with the previously shown results that breaking mortalin-p53 complex for example by p53 C-terminal peptides leads to restoration of p53 activities, abrogation of this cytoplasmic sequestration and relocation of p53 to the nucleus. Thus, the sequestration is reversible.

Losefson et al. used the method of pull-down with heparin-sepharose beads to track p53-mortalin interaction [132]. They found that this interaction is nucleotide sensitive and that in contrast with the result from cell lysates, the purified SBD of Mortalin, (and not NBD), binds to p53 in a concentration-dependent manor. They suggested that the contrast between *in vitro* results and results from cell lysate can be because of modifications or additional proteins in cell lysates that might affect p53-mortalin interaction. They also found two binding sites for mortalin on p53.

One on TET domain and the other on CTD domain. Any of which is sufficient for binding. Another result was that the mutant p53 (L344P) which fails to form tetramer, binds to mortalin similar to the wild type tetrameric p53. Kaul et al. studied the effect of mortalin mutant in binding to p53 [98]. C-terminus deleted mutants represses p53 activity, suggesting that C-terminus of mortalin is not needed for interaction with p53. The mutation of V482F in the substrate-binding pocket abrogated the binding between mortalin and p53 [132], consistent with a similar mutation in DnaK, a bacterial homolog of mortalin, which also disrupts the substrate binding [136]. Accordingly, pre-incubation of mortalin with a short substrate peptide greatly reduced its binding to p53 [121, 136]. The interaction between mortalin and full-length p53 can be abrogated by adding ATP [132]. All these evidences suggest that mortalin binds to p53 through its substrate-binding pocket on SBD. However, so far, no direct evidence such as co-crystallization show the exact interacting domains between the two proteins.

Gabizon et al. identified and characterized ten binding peptides by screening proteins that interact with tetrameric p53 C-terminal domain (p53CTD). They identified residues 266-280 in mortalin as the binding site for NRD (negative regulatory domain) of p53 [114].

In vivo co-immunoprecipitation of mortalin with p53 and its mutants in Human osteosarcoma (U2OS) and breast carcinoma (MCF7) cells revealed carboxyl terminus of p53 (resides 312-352) as the binding site for N-teminal region of mortalin mot2 [99, 125, 137].

4.7. Protein p53-Mortalin Complex Inhibitors

Mortalin-p53 complex is a selective target for cancer therapy [127]. Some short peptides [121, 135], small molecules, small interfering RNA (siRNA) [138], mortalin small hairpin shRNA_(in hepatocellular carcinoma), UBXN2A (Ubiquitin-like (UBX)-domain-containing protein) can hinder p53-mortalin binding, reactivate p53 and induce apoptosis in cancer cells.

Some small molecule mortalin inhibitors include withanone [139] (a withanolide, one of the major components of the alcoholic extract of Ashwagangha leaves) and MKT-077 [140-142]

(an anti-tumor compound, a water soluble delocalized lipophilic cationic/rhodacyanine dye analogue). Withaferin A (Wi-A) and Withanone (Wi-N) are two structurally similar withanolides. Wi-A is toxic to both cancer and normal cells, Wi-N however has milder effect and does not have toxic effects on normal cells. It suggests that Wi-N can be a safer cancer drug but higher doses are needed for the same effect as Wi-A. Mortalin was docked with both of these compounds and it was found that Arg513 in carboxyl terminus of mortalin interacts with them [143]. The result of our molecular modeling and NMR spectroscopy studies has confirmed the involvement of this residue in binding of mortalin and SHetA2.

Data shows stronger binding of Wi-A to mortalin. Wi-A and Wi-N also bonds with Arg282 and Leu111 residues of mortalin. Wi-A binds more effectively to p53. Attempts has been made to enhance the production of withanolides and elucidate their molecular mechanisms [144].

MKT-077 is selectively toxic to cancer cells and IC_{50} of MKT077 for normal cells is more than 100 times greater than that of cancer cell lines thus it was considered as cancer chemo-therapeutic in a Phase I trial. However, the trial was stopped due to excessive renal toxicity in phase I clinical trials against solid tumors [145, 146]. MKT-077 interacts with HSPA8 [140] and HSPA9 [141]. The binding site of MKT-077 (positively charged) to HSPA8 ADP-state has been determined to be a negatively charged pocket close to residues in nucleotide binding site in the interface of subdomain Ia and IIa, which are 100% conserved between 13 human HSPAs [140]. The binding site of MKT-077 to HSPA9 is within residues 252-310, which in some studies is shown to include residues needed for binding of p53 [141]. Immunoaffinity studies showed that MKT-077 binds to mortalin at alpha-helix within sub-domain IIb of catalytic cleft and abrogates mortalin function, causes pancytoplasmic distribution of mortalin (characteristic of normal cells) and nuclear translocation and activation of p53. Thus induces growth arrest of cancer cells [141].

Molecular dynamic simulations and co-immunoprecipitation studies for both MKT-077 and withanone ligands showed the abrogation of p53-mortalin interaction and re-activation of p53 upon treatment with the ligands [139, 141]. There are several potential ways for MKT-077 to

inhibit mortalin binding to p53. MKT-077 may inhibit binding of Tid1 by destabilizing the ATP-bound conformation. Or it might inhibit binding of p53 by perturbing the conformation of mortalin at its β strand (residues 267-271) [147]. It is shown that growth arrest function of MKT-077 is mediated by a p53-dependent pathway. In cancer cells, and not in normal cells, abrogation of p53-mortalin interaction leads to release of p53 from cytoplasm and translocation of p53 to nucleus and restoration of its transcriptional activation [141]. Similar effect for MKT-077 has been observed in a nonmammalian cancer model in soft shell clam *Mya arenaria* [126].

4.8. Pharmocokinetics Studies

SHetA2 absorption can be affected by factors like dose, species (thus different dietary habits, etc.), and formulation [148]. Effect of SHetA2 is time dependent and dose dependent. For example in ovarian cancer cell lines, cell cycle arrest contributes to SHetA2 growth inhibition, time-dependently and dose-responsively [149]. At lower concentrations, differentiation is observed however in higher concentrations apoptosis in dominate [57, 59]. Thus, it is important to determine concentrations, bioavailability, and stability of SHetA2 in order to measure the sufficient dosing of SHetA2 in clinical trials to target apoptosis inducing concentrations and maximize the activity of the drug.

Kabirov et al. performed some preclinical pharmacokinetics to determine the toxicokinetics of SHetA2 in animals [67]. HPLC-UV studies have been applied and pharmacokinetic profile of SHetA2 has been obtained to quantify SHetA2 in plasma in mouse and human by Sharma et al. [150]. The result of a study performed by Zhang et al. [151] indicates a linear and predictive pharmacokinetics in mouse plasma. The terminal half-life, total body clearance value, oral absorption, and bioavailability were measured. This study also showed high tissue uptake. It measured the observed total body clearance of SHetA2 to be 1.81 l/h/kg and suggested that the clearance of SHetA2 is due to degradation/metabolism.

Liu et al.[152] applied liquid chromatography-ultraviolet/multi-stage mass spectrometry (LC-UV/MS^n) to evaluate SHetA2 metabolites in human and rat liver microsomes in vitro and in mice and rats in vivo.

In another study Sharma et al. [153] developed a novel method for quantification of SHetA2 in human and mouse plasma. In this HPLC-UV bioanalytical method, lower limit of quantification (LLOQ) is =5 ng/ml and lower limit of detection (LLOD) is 2.5 ng/ml, which was lower than previous study in which LLOQ was 10 ng/ml, thus this method has higher sensitivity. Moreover, this method showed specificity, accuracy, precision, stability, and reproducibility.

It appeared that once the absorption saturation reaches, increasing doses will decrease bioavailability and absorption [81] for higher doses. Absorption saturation in mice, rats and dogs is about 100 mg/kg body weight/day [67, 151, 152].

Highly hydrophobic SHetA2 has low water solubility and low gastrointestinal (GI) absorption, and expressed low oral bioavailability in rats (<1%) and dogs [67]. In dogs, suspension of SHetA2 in 30% aqueous Kolliphor HS15 enhances the bioavailability.

Optimal dose is a dose at which chemoprevention occurs without toxicity. The dose that reduces cyclin D1 in neoplastic and not in normal cells. The AIN76A diet formulation of SHetA2, 187 mg/kg/day SHetA2 suspended in Kolliphor HS15 for bioavailability enhancement has been reported in mice and effective tissue drug levels was observed at this dose with no toxicity [81]. SHetA2 showed very low toxicity, with a no-observed-adverse-effect level (NOAEL) of 500 mg/kg/day in rats and over 1500 mg/kg/day in dogs [67]. Based on bioavailibility-dose studies NOAEL in human dose was chosen 100 mg/kg [153]. On the other hand, the oral administered dose of SHetA2 that reduces the colon and small intestinal polyps incidence and sizes is 30 mg/kg/day [71]. While a dose of 10 to 60 mg/kg/day can inhibit xenograft tumor growth [25, 62, 67]. This large therapeutic window makes SHetA2 an ideal chemopreventive drug.

Responses to SHetA2 doses in $APC^{min/+}$ Mice was found to be gender related. This suggests that male and females to be dosed differently [71].Some administration routes of SHetA2 include oral [25, 153] and intravenous in rats, mice and dogs, dietary formulation [81], vaginal [154] administration.

4.9. Conclusion

SHetA2 is a Flex-Het with promising anti-cancer activities. It contributes in both intrinsic apoptosis by directly targeting mitochondria associated with reduction of Bcl-2, and extrinsic apoptosis by enhancing death receptor activation through inhibition of nuclear factor κB (NF-κB) and inducing the expression of the CHOP protein. SHetA2 reduces the expression of Bcl-2 in cancer cells thus promotes apoptosis [60, 62] while mortalin can prevent this effect leading to inhibition of apoptosis. Thus Mortalin and SHetA2 can regulate the Bcl-2 protein family and affect the stability of mitochondria.

Inactivation of p53 through its interaction with mortalin, can disturb p53 apoptosis function. Once p53 is released from the complex with mortalin, it can translocate to the nucleus where it induces multiple genes that contribute to apoptosis. The interaction between the two proteins can be disturbed by SHetA2. SHetA2 abrogation of mortalin binding to p53 augments SHetA2 induced apoptosis.

SHetA2 has low water solubility. In chapter 7 and 8 we study SHetA2 analogs with the goal to improve the properties of the compound.

CHAPTER V

NMR STUDIES OF THE INTERACTION BETWEEN MORTALIN AND SHETA2

5.1. Introduction

As mentioned in the previous chapter, Heat shock protein HSPA9/mortalin interacts with the tumor protein p53 and is also a client for SHetA2. Mortalin sequesters p53 in cytoplasm and hinders its apoptotic function. Binding of SHetA2 to mortalin will interrupt p53-mortalin complex, and releases protein p53 to nucleus where it initiates apoptosis. Thus, one pathway for the compounds to inhibit the growth of cancer cells in vitro is by displacing p53 from mortalin. However, interactions of mortalin with other client proteins, such as Bcl-2 and p66shc, may also affect cell apoptosis.

The substrate-binding domain (SBD) of mortalin was shown to associate with the tumor protein p53 in a concentration-dependent manner [27]. A mutation of V482F in the substrate-binding pocket abrogated the interaction with p53 [27], consistent with a similar mutation in DnaK, a bacterial homolog of mortalin, which also disrupts the substrate binding [136]. Moreover, pre-incubation of mortalin with a short substrate peptide greatly reduced the interactions between mortalin and p53 [136]. These are some evidences that suggest that p53 binds to the substrate-binding pocket in the substrate-binding domain of mortalin.

In this chapter, we study the interaction of the mortalin substrate binding domain (Protein Data Bank ID: 3N8E) with SHetA2 via solution NMR. We use chemical shift perturbation experiments to study these interactions.

5.2. Materials and Methods

5.2.1. Chemicals and Reagents

Purified SHetA2 powder is kindly provided by Dr. Berlin. Plasmid pET-52 with Mortalin SBD (AA#439-597) encoding gene is kindly provided by Dr. Benbrook. Lysozyme was purchased from MP Biomedicals (Santa Ana, CA). Benzonase was purchased from Semba Biosciences (Madison, WI). Protease inhibitor cocktail, ampicillin, kanamycin, imidazole, biotin, DTT, chloramphenicol and other chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Isotope-enriched (^{13}C, ^{15}N) ISOGRO supplement, $^{15}NH_4Cl$ and uniformly labeled ^{13}C glucose were purchased from Cambridge Isotope Laboratories (Andover, MA). The Profinity™ IMAC Nickel Charged Resin was purchased from Bio-Rad Laboratories (Hercules, CA). Thrombin is purchase from Bio Pharm Laboratories (Alpine, UT). Ni-NTA purification kit is purchased from Bio-Rad (Hercules, CA).

5.2.2. Sample Preparation

A. Protein expression and purification

The mortalin SBD encoding gene was introduced into pET-52 vector through NcoI and BamHI double restriction cleavage sites with a customized stop codon right after the sequence. The recombinant protein was constructed with 10×His tag on the N terminus, a HRV-3C protease cleavage site (LEVLFQ/GP), and a linker (SGPASPR). The theoretical pI of the recombinant protein after treatment with HRV-3C protease is 5.31 and molecular weight (MW) is 18129.47 Da [155]. Plasmid pET-52 encoding mortalin SBD was transformed into Escherichia coli BL21

(DE3) competent cells. The glycerol supplemented stock cells were prepared and stored in -80°C. 5 µl of transformed cell stock was then spread on an ampicillin treated Lauria agar plate, incubated overnight at 37°C. A colony was transferred to 100 mL ampicillin supplemented LB medium and incubated at 37 °C, 250 rpm until optical density at 600 nm (OD600) reaches 0.8 or overnight. The cells were pelleted at 4000 rpm for 20 min, the pellet was re-suspended in 5 times volume of ^{13}C-glucose and $^{15}NHCl_4$ supplemented M9 medium (Table 5.1) . The culture was incubated at 37 °C 250 rpm until OD600 reaches 0.8. Final concentration of 1 mM IPTG was added and the cells were induced at 20°C, 250 rpm for 20 hours. The cells were then harvested by centrifugation at 4000 rpm for 30 minutes. The cell pellets were re-suspended in the lysis buffer (Table 5.2). Lysozyme was added to a final concentration of 0.5 mg/ml. DNase was added to reduce viscosity and clumpy cell suspension. 1 mM tris (2-carboxyethyl) phosphine (TCEP) and 1× protease inhibitor cocktail (to prevent protein degradation) was added. We then either vortexed the cells every 15 minutes for 1 hour on ice or put on rocker at 4°C for 2 hours. The lysed cells were sonicated and centrifuged at 30,000 kg for 40 min at 4°C. The protein is in the supernatant. The supernatant was filtered by a 0.2 µm pore size filter to remove any remaining cells for the next step, that is protein purification. The supernatant was incubated with Ni-NTA resin for 1 hour at 4ºC and loaded onto a column. In parallel, the column was washed with Buffer B (20 mM Tris, 500 mM NaCl, 250 mM Imidazole, at pH 7.8), water, and Buffer A (20 mM Tris, 500 mM NaCl, 20 mM Imidazole, at pH 7.8). The resin-supernatant mixture was then loaded on the column. The flow through, wash and elute eluents were collected. The column was washed with 10 CV Buffer A and eluted by 10 CV Buffer B and then Buffer C (20 mM Tris, 500 mM NaCl, 1M Imidazole, at pH 7.8). The eluents were run on a SDS page gel to identify the eluent(s) containing the protein. Next, the protein-containing eluents were cleaved by HRV-3C protease to remove His-tag and buffer exchanged to dialysis buffer via dialysis overnight. Dialysis buffer contains 20 mM Tris, 100 mM NaCl, at pH 7.5. The cleaved product was then

equilibrated with Ni-NTA resins for 1 hour at 4ºC and loaded on to the second Ni column to remove the cleaved His-tag. The goal is to have the final sample in the dialysis buffer. FT eluent was collected. Then the column was washed with 5 CV of wash-I (with Buffer D), followed by 5 CV of Buffer A, 5 CV of Buffer B, and 5 CV of Buffer C. The protein is mostly in FT and some in wash I. SDS-PAGE was run to test the purity of the protein. We centrifuge the sample to the desired concentration for NMR (~10-20 mg/ml) and stored at 4 °C before use. In case of the need for longer term storage, samples should be flash freezed and store at -80°C.

Table 5.1 Supplemented minimal growth media (M9) for ^{13}C,^{15}N-labeled protein expression.

M9 (1L)	
5x M9 salts	200 ml
^{13}C-glucose	2 g
$^{15}NH_4Cl$	2 g
1000x Ampicillin	1 ml
Biotin	8 mg
100x BME	10 ml
100x Thiamine	1 ml
$MgSO_4$(1M)	400 µl
$CaCl_2$(1M)	40 µl
$FeCl_3$(.01M)	200 µl
^{13}C,^{15}N-Isogro	2 ml

Table 5.2 Lysis buffer composition

Lysis buffer for 1L of cells	
20 mM Tis, 500 mM NaCl, pH=8.0	24 ml
50 mg/ml Lysozyme	480 µl
100x protease inhibitor cocktail	240 µl

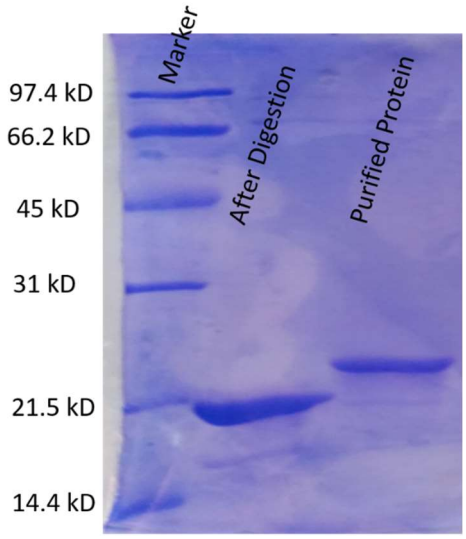


Figure 5.1 SDS-PAGE of purified Mortalin and digestion by HRV 3C protease.
The lane on the left is the marker proteins with molecular weight labeled. The lane in the middle shows the protein after treated with HRV-3C and removal the His-tag. The right lane is the purified protein through Ni-NTA.

B. NMR Sample Preparation

SHetA2 is water insoluble therefore we dissolved SHetA2 in DMSO and used the dissolved SHetA2 sample for titration with the protein sample. To prepare the ligand sample we dissolved 4.5 mg SHetA2 into 25 µl DMSO. This gives a 180 mg/ml or 420 mM concentration sample consistent with our previous tests. Protein sample was prepared with adding 450 µl protein solution (which contains 16 mg protein in 650 µl NMR Buffer, thus 22.2 mg/ml= 1.23 mM) in 50

µl D2O (10% D2O). D2O is added to lock the NMR signals. The ligand sample was then titrated into the protein sample to make our NMR samples. Molar ratio of protein to ligand was 1:0.2, 1:0.4, 1:0.6, 1:0.8, 1:1, 1:2, 1:4, 1:8, and 1:16. In one tube, we had the protein sample which we titrated with ligand sample, and in the other tube we had the protein sample titrated with DMSO as control. The same NMR experiments with same parameters were performed on both samples.

5.2.3. NMR Spectroscopy

NMR experiments were performed at 25 ºC on a Agilent INOVA 600 MHz spectrometer with a Nalorac 5-mm ^{1}H, ^{13}C, ^{15}N PFG triple resonance probe and VnmrJ software with Biopack suite of pulse programs to collect the data. A 1D experiment was performed to have a quick check on the status of the sample and the signal. If desired quality is observed, then a 2D ^{15}N–heteronuclear single quantum coherence (NHSQC) is acquired to further check the protein sample quality. HNCA, HNCOCA, HNCO, HNCACO, HNCACB, and HNCOCACB spectra are collected for backbone assignment and CCONH and HCCONH for side chain assignment.

NMR spectrum were collected and data were processed using NMRPipe [156], a software for processing and analyzing multidimensional NMR spectroscopic data. The processed data from NMRPipe was then used as the input for SPARKY [157], a graphical interface that is used to assign, pick, and integrate NMR peaks of macromolecule. Output of the SPARKY is text peak lists with assignments, chemical shifts, linewidth, etc. Then the sequence of the protein and the spectral peak list were uploaded to the PINE server [158]. Then we explored the auto-assigned spectra in SPARKY and made some controls and corrections. The data was then analyzed with TALOS+ [159], a software in NMRPipe software package that predicts the secondary structure by establishing a relation between ^{13}C, ^{15}N and ^{1}H chemical shifts and backbone dihedral angles ϕ and ψ.

The chemical shifts were assigned to each individual amino acids. The chemical perturbation was calculated from:

$$\delta = \sqrt{\left(\delta_{N_l} - \delta_{N_s}\right)^2 + \left[10 \times \left(\delta_{H_l} - \delta_{H_s}\right)\right]^2} \quad (5.1)$$

where the δN_l and δH_l are the ^{15}N and ^{1}H chemical shift difference between the last SHetA2 titration and *apo*-state. The δN_s and δH_s are the ^{15}N and ^{1}H chemical shift difference between the corresponding last DMSO titration and *apo*-state.

5.3. Result and Discussion

5.3.1. Solution NMR of Mortalin SBD and Backbone Assignments

The NMR sample was loaded into a 5 mm solution NMR tube. A 1D experiment was performed to have a quick check on the status of the sample and the signal. If desired quality is observed, then a 2D ^{15}N–heteronuclear single quantum coherence (NHSQC) is acquired to further check the protein sample quality (Figure 5.2). In NHSQC one axis is proton ^{1}H and the other is the heteronucleus N. Through J coupling magnetization is transferred from H to the attached N, and then transfers back to the H and will be detected. The NHSQC spectrum mainly shows the backbone amide H-N correlations. Well dispersed peaks with proton chemical shifts spreading around 7 ppm to 9.7 ppm is an indicator of a well-folded protein and we can continue acquiring further NMR spectrum which includes sequential 3D backbone assignments experiments i.e. HNCA, HNCOCA, HNCO, HNCACO, HNCACB, HNCOCACB.

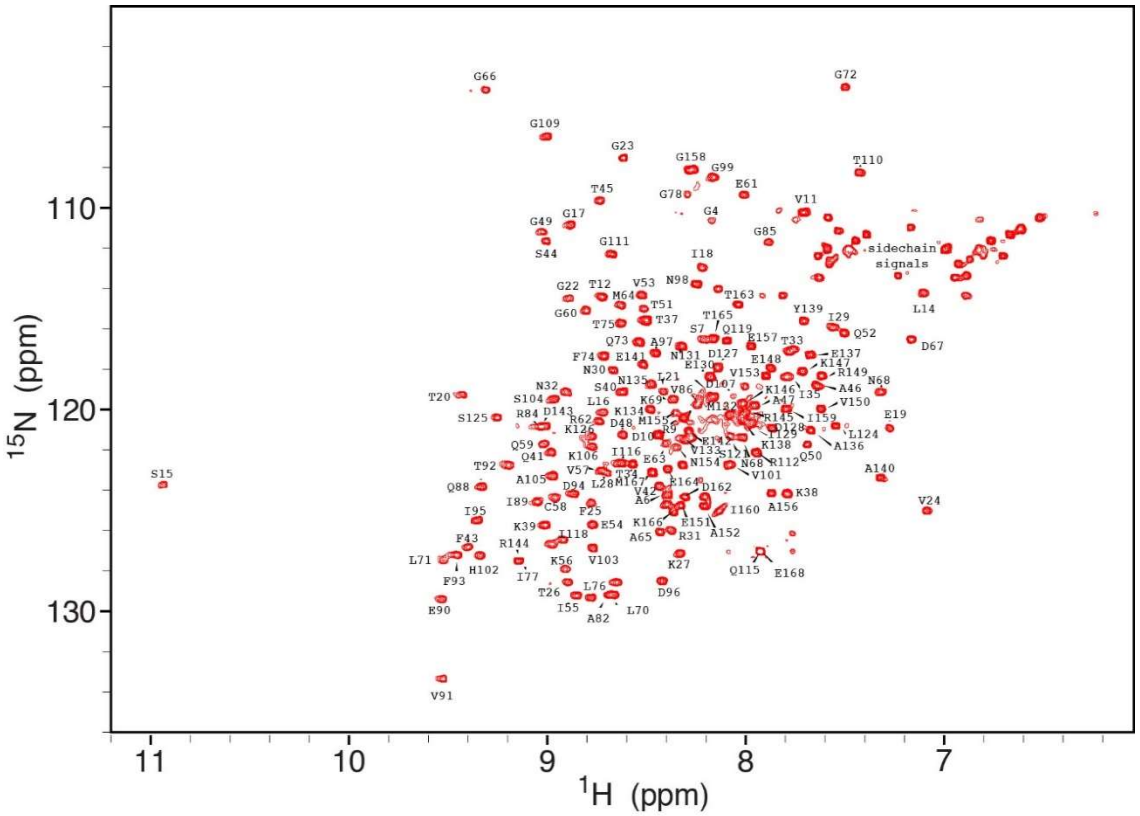


Figure 5.2 2D ^{1}H-^{15}NHSQC spectrum
of the labeled mortalin sample with assignments.

Pairs of two complementary 3D NMR (or triple resonance) spectrums, HNCA-HNCOCA, HNCO-HNCACO, and HNCACB-HNCOCACB are acquired to obtain sequential 3D backbone assignments. HCCONH-CCONH pair of spectrum can give more side chain residues assigned. The peaks in NHSQC were selected and used as the guiding position for peak selection in the 3D data set. The HNCA, HNCACO and HNCACB correlates inter- and intra-residues, while HNCOCACB, HNCOCA and HNCO only give the intra-residue correlation. HNCACB for example gives the chemical shift information for each amide of residue i and the chemical shift information of the alpha and beta carbon of residue i and i-1. Thus, HNCACB correlates each NH group of residue i with the Cα and Cβ chemical shifts of residue i and Cα and Cβ of residue i-1. HNCOCACB however give the chemical shift information for each amide of residue i and only the chemical shift information of the alpha and beta carbon of residue i-1, Figure 5.3. The values

of Cα and Cβ chemical shifts are characteristic of the amino acid type. Then the assigned chain can be matched to the sequence of amino acids in the protein. Resonances within a residue and between neighboring residues allows for complete chemical shift assignment through *backbone walk.* In Figure 5.4 the HNCACB and HNCOCACB spectra for residues 102 to 106 are displayed. Each HNCACB slice has the Cα and Cβ from residue i and the Cα and Cβ from residue i-1 and each HNCOCACB slice has two Cα and Cβ peaks from residue i. All there peaks can be correlated through backbone walk.

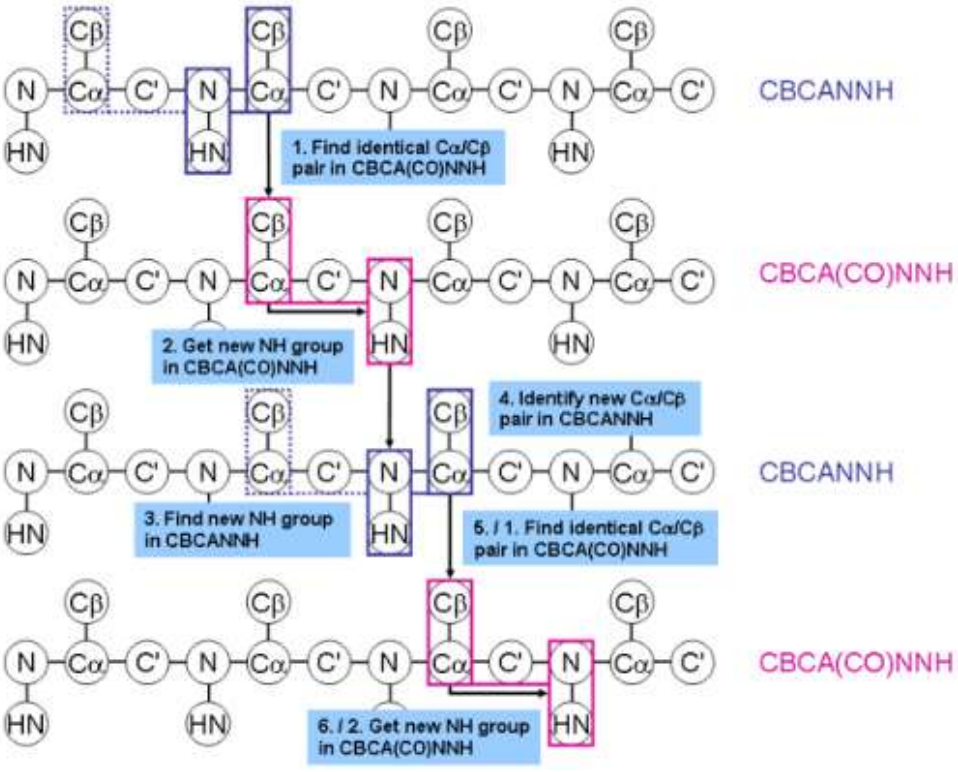


Figure 5.3 Triple resonance backbone assignment
[https://www.protein-nmr.org.uk]

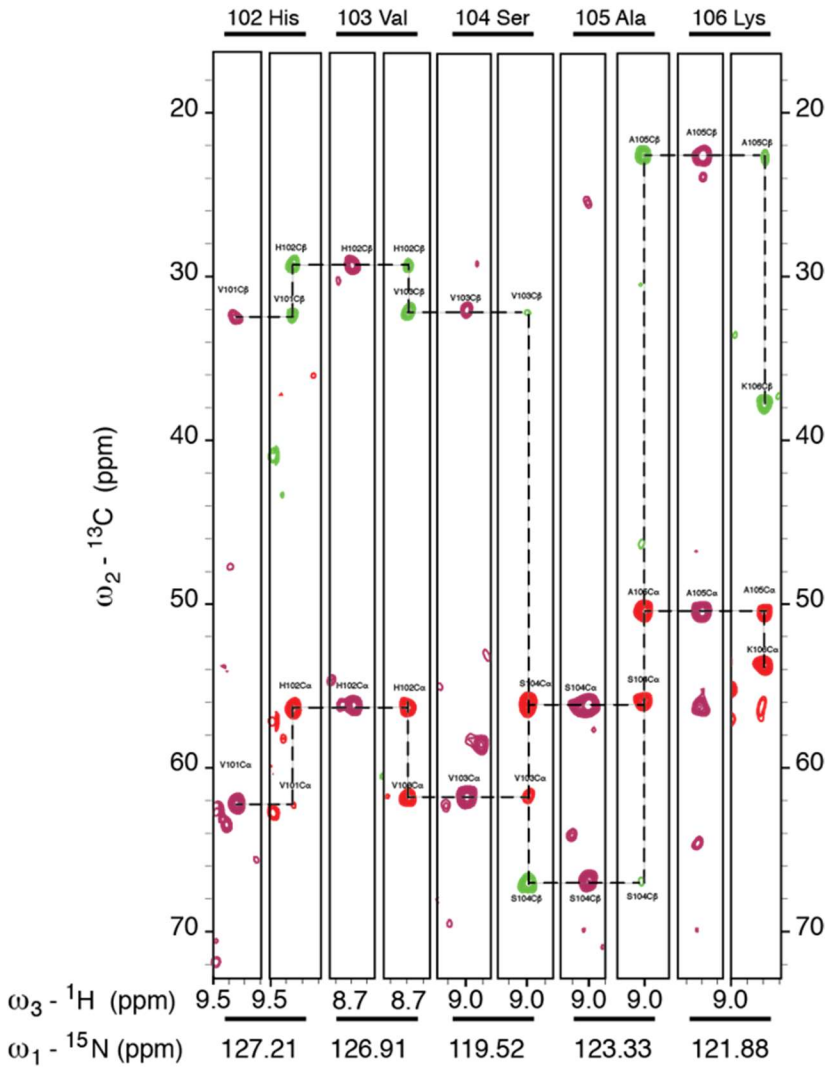


Figure 5.4 The backbone assignments example.

The peaks positions and intensities data were uploaded to the PINE server 2.0 for auto-assignment of the peaks. The assignments are then inspected manually for control. The output result has 90% confidence in assignment (Figure 5.5). Secondary structure probabilities are predicted using PECAN automated tool [160] from the chemical shift assignments with good agreement with the crystal structure of the protein, PDB ID: 3N8E (Figure 5.6). PECAN is a tool

for probabilistic secondary structure determination. The output obtained from PINE server is submitted to PECAN. Input of PECAN has BMRB format.

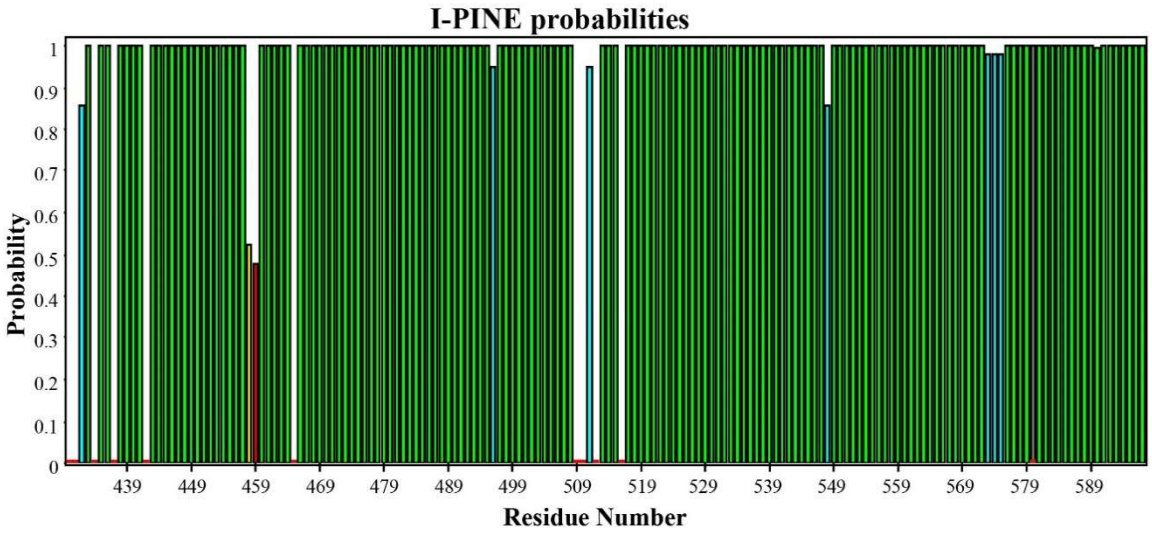


Figure 5.5 Assignments prediction probabilities for each amino acids in mortalin obtained from the I-PINE server.
Green shows 100% confidence, Yellow-green indicates over 90% confidence, Cyan means over 80% confidence, Yellow shows 50%, and Red is less than 50% confidence.

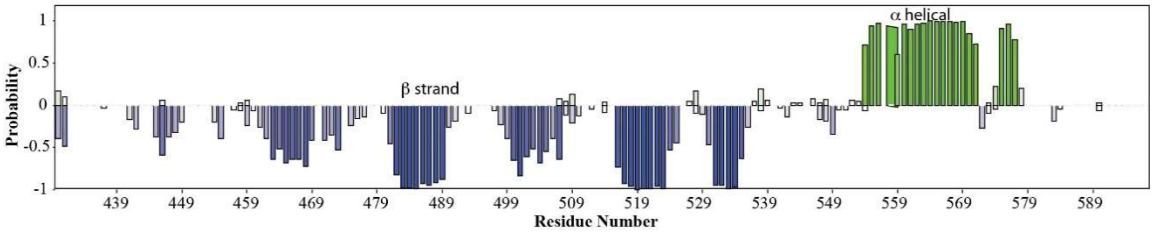


Figure 5.6 Probabilities of the secondary structure predictions obtained using PECAN.
The length of each bar represents the probability.

We can predict the secondary structure by TALOS+ also as it is shown in Figure 5.7. The predicted order parameter S^2 (upper window) and ANN-predicted secondary structure (lower window) plotted for all residues are shown. The predictions have high overall confidence and good agreement with the crystal structure of the protein, especially in structured regions. The low

confidence in the both ends is because these parts are flexible regions in solution. In the crystal environment however, the C-termini of the two units are in close contact and the helices are more stable.

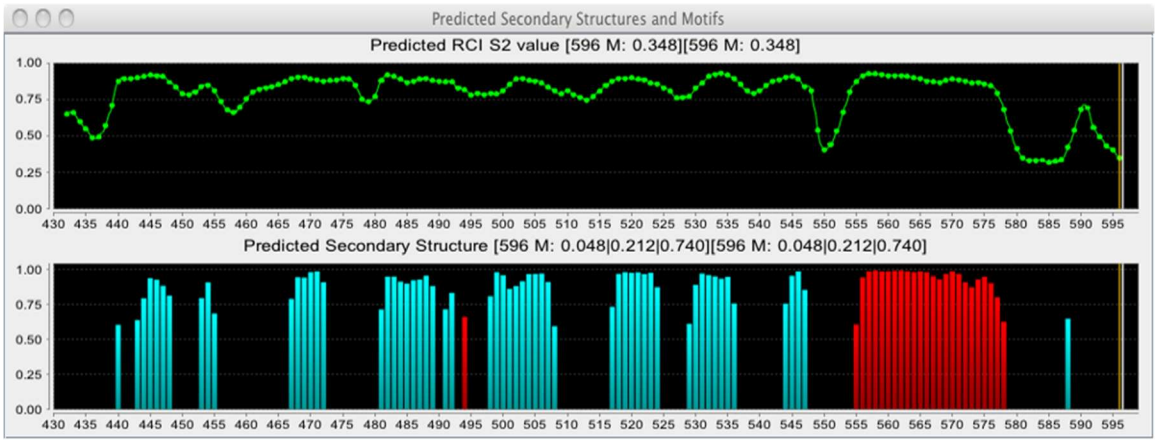


Figure 5.7 Secondary structure predicted by TALOS+.
Aqua bars represent β-sheets and red bars represent helices. The height of the bars is a measure of the probability of the prediction.

5.3.2. Chemical Shift Perturbation

Chemical shift perturbation information are used to identify protein-ligand binding sites. Peak shifts following the titrations indicate conformational changes of the protein upon binding of SHetA2. NHSQC spectrum was acquired for the *apo*-state protein sample (without SHetA2) and titrated SHetA2 sample (Figure 5.8). The protein backbone resonance peaks in the NHSQC spectrum were affected by SHetA2 titration. The top most affected shifts are L457, V471, S473, A475, G514, V515, I518, D572, K576, V579 and E586, which are depicted in Figure 5.10.

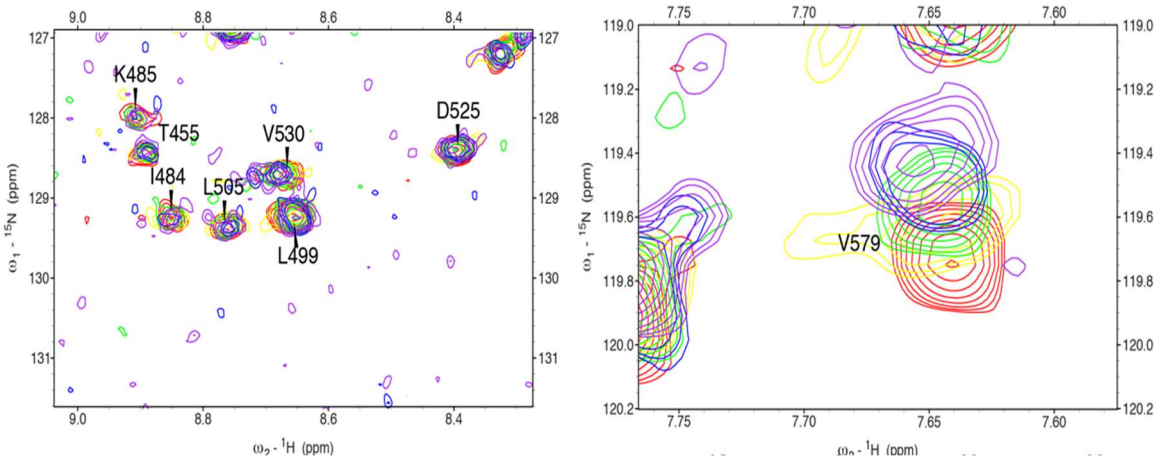


Figure 5.8 Peak shifts caused by adding SHetA2.
Different colors indicate different molecular ratio of protein:SHetA2. (Red 1:0, yellow 1:0.4, green 1:0.8, blue 1:2, purple 1:8)

The chemical shift perturbation plot is found from eq. 5.1 and plotted in Figure 5.9. The height of the bar is indicative of the relative chemical shift changes of SHetA2 titrated spectrum from DMSO titrated spectrum. These residues form the binding pocket for SHetA2 in substrate region of mortalin. Our NMR spectroscopy data together with the result from our molecular dynamic studies identifies this binding site to overlap with the binding site of p53 to mortalin. Thus confirms that SHetA2 may be an inhibitor for the binding of mortalin and p53.

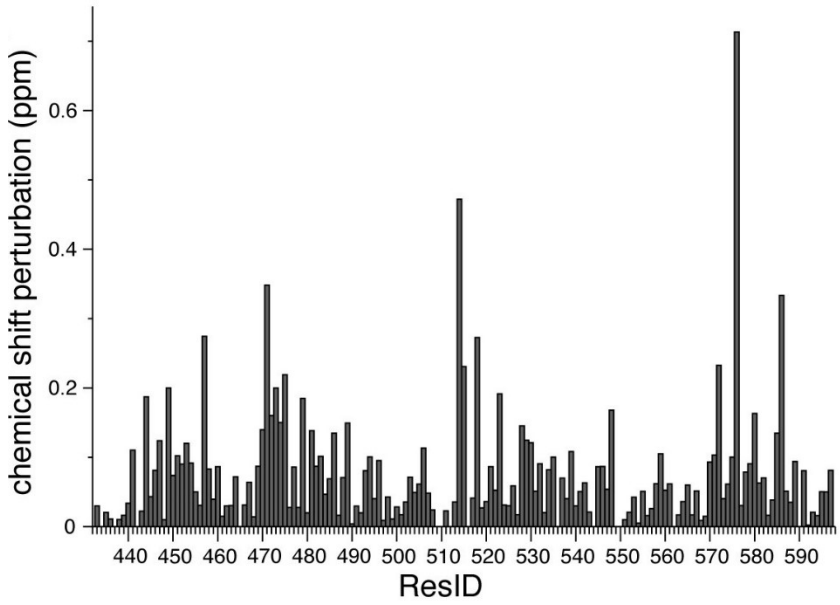


Figure 5.9 Chemical shift perturbation of mortalin residues from the last titration of SHetA2.

The top 11 residues which had the most significant chemical shifts are obtained from chemical shoft assignments and been plotted using VMD Figure 5.10.

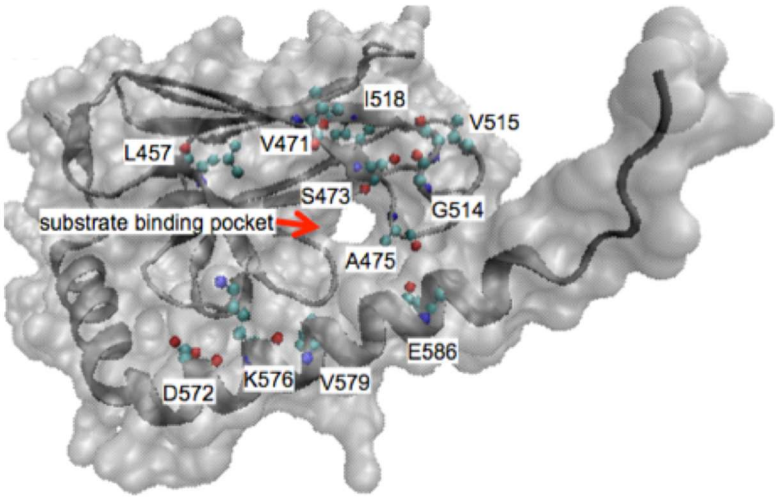


Figure 5.10 The most shifted peaks
upon binding of SHetA2 are shown in the substrate binding domain of mortalin

As it is observed in Figuer 5.10 the top shifted peaks are L457, V471, S473, A475, G514, V515, I518, D572, K576, V579 and E586. These include residues on a hydrophobic pocket in SBD and some residues on the lid helix.

5.4 Conclusion

The predicted structure from chemical shifts has high similarity to the crystal structure except the more flexible unstructured C-terminus. We investigated the interaction of mortalin with SHetA2 using solution NMR. From the chemical shift perturbation data, we observe that SHetA2 binds to a binding pocket in the substrate binding domain of mortalin which in some studies has been shown to be the binding site for p53. Thus, the competitive binding of SHetA2 and p53 to mortalin can disrupt the interactions between p53 and mortalin and release p53. We identified the most perturbed residues upon binding. Among which some form a hydrophobic pocket within the SBD. This suggests the importance of hydrophobicity in the binding pocket . Same conclusion was drawn from our simulation results. In the last two chapters we try to study

the effect of increasing hydrophobicity on the binding strength by designing novel analogs of SHetA2. NMR spectroscopy results can identify the strength of binding (K_d), site of interaction and residues involved in the binding to ligand. Thus NMR spectroscopy can help designing of more preferable binders.

CHAPTER VI

MOLECULAR DYNAMICS STUDIES OF MORTALIN-SHETA2 USING GROMACS

6.1. Introduction

Molecular Modeling has vast applications in drug discovery and design. It provides a fast and powerful tool to investigate the interactions between protein and ligand. The result however needs to be validated and optimized by incorporating experimental findings. We applied molecular docking and molecular dynamics tools to investigate mortalin-SHetA2 interactions. We found the binding configurations and affinities of the complex and provide some evidence that the interaction between mortalin and SHetA2 can abrogate mortalin-p53 binding. More details on our studies in molecular docking approach is presented in chapters VII and VIII. In this chapter, we emphasize on molecular dynamic studies of the wildtype and mutant mortalin in complex with SHetA2 using GROMACS.

6.2. Material and Methods

The crystal structure of protein, mortalin (PDB ID: 3N8E), was downloaded from Protein Data Bank. VMD [161] mutator was used to generate protein mutants. To include the ligand topology, the ligand.itp file is needed. First ChemSketch™ (Advanced Chemistry Development, Inc. ADC/Labs, Toronto, Ontario, Canada) was used to generate the SMILES notations which were then converted to PDB files with initial three-dimensional coordinates using Open Babel

GUI [162]. Next the ligand.pdb file was converted to ligand.mol2 using UCSF Chimera [163]. This was achieved by submitting ligand.mol2 to swissparam server, http://www.swissparam.ch/. The ligand.pdb and ligand.itp were included in system topology file. System preparation, some of simulations, and final analysis were performed on an iMac with a 2.4 GHz Intel Core i3 processor and 4 GB RAM. The main molecular dynamics simulations were performed at the OSU High Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OAC–1126330. The Cowboy linux cluster has 252 standard compute nodes, each with dual Intel Xeon E5-2620 "Sandy Bridge" hex core 2.0 GHz CPUs, with 32 GB of 1333 MHz RAM. We used software GROMACS 4.5.5 [20] and CHARMM27 sets of force field with tip3p water model [164]. Ligand topology was created with swissparam and was included in the system topology. Periodic simulation box was defined to mimic the infinite system to avoid real phase boundaries and to minimize finite boundary effects. Distance of the complex from the walls of the box was at least 1.5 nm. The box was filled with water and 100 mM of NaCl was added to mimic the experimental condition, under which we performed NMR experiments. The initial energy minimization was then performed by using steepest descent algorithm until the system converges to a minimum energy with forces less than 100 kJ/mol/nm (emtol=100 kJ/mol/nm).

The following equilibrium step was conducted in two phases, constant volume (NVT) and constant pressure (NPT). Each with 1 fs time steps for 5 and 10 ns respectively. NVT is performed to ensure that the *average* temperature of a system is correct (300 K in this experiment). In NVT equilibration the system was strongly coupled to a temperature bath using V–rescale coupling [165] with coupling constant of $\tau_T = 0.1$ ps. Equilibration of pressure is done during the NPT phase to maintain the pressure isotropically at 1 bar. A strong Parrinello–Rahman pressure coupling with time constant $\tau_P = 5.0$ ps was used [166], as well as a weak Nose–Hoover temperature coupling with time constant $\tau_T = 0.5$ ps to ensure a true NPT ensemble [167-169].

After equilibration, 100 ns production run of molecular dynamics was performed in 2 fs time steps with weakly coupled temperature and pressure (Nose–Hoover τ_T = 0.5 ps and Parrinello–Rahman τ_P = 2 ps). Leap–frog integrator [170] was applied during NVT, NPT, and MD simulations. During the MD run, system coordinates, velocities, trajectories and energies were recorded every 2 ps. Cut-off distances including neighbor list, electrostatic, and van der Waals cut-offs were 1.2 nm. The Lennard–Jones potential is used to estimate the non–bonded van der Waals interactions, and Particle-Mesh Ewald (PME) algorithm is used for long range electrostatic interactions treatment [171]. The bonds were constrained by LINear Constraint Solver (LINCS) algorithm [172].

6.3. Results and Discussion

Binding of mortalin to p53 causes cytoplasmic sequestration of p53 thus loss of the function of p53 and leads to immortalization of cells and tumor resistance to radiotherapy and chemotherapy. Thus abrogating this binding is important in re-locating p53 to the nucleus where it re-gains its apoptosis functions. Identifying the binding site of mortalin with p53 gives insights on identifying the complex inhibitors, which can potentially have anticancer effects. As mentioned in chapter 4, several studies have been performed to identify the binding site of p53 and mortalin. Some identified NBD and some identified SBD as the binding domain on mortalin. Our result from molecular dynamic simulations, showed both bindings are possible. (Result from NBD domain is not included in the current paper).

MD simulations was carried out for 50 ns followed by distant calculations. The configuration with minimum binding energy is shown in Figure 6.1.

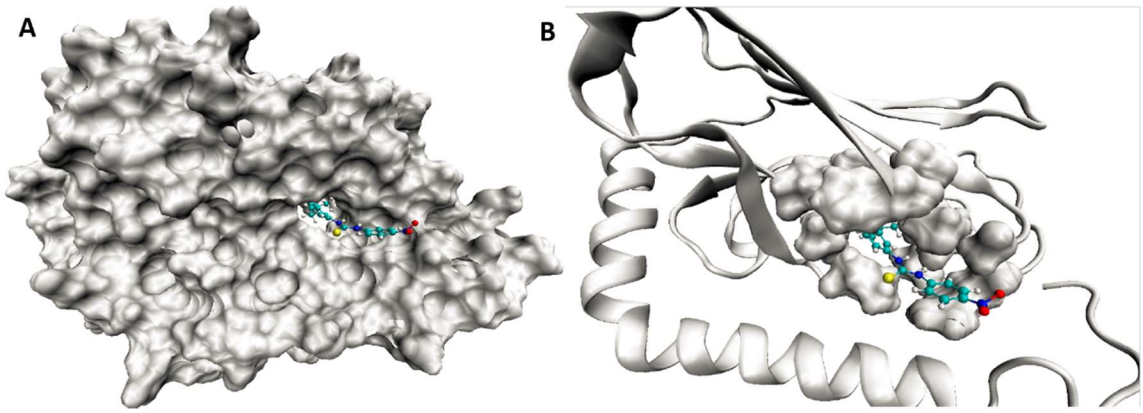


Figure 6.1 SBD mortalin – SHetA2 complex lowest bonding energy frame in 50 ns MD simulations using GROMACS and visualized by VMD.
(A) Surface display and (B) Secondary structure display of mortalin SBD and Surface display of hydrophobic binding pocket

It is observed that SHetA2 is located within a hydrophobic binding pocket in SBD mortalin, surrounded by ILE447, THR449, LEU450, PHE472, SER473, THR 474, ALA475, VAL482, ILE484, GLY514, ILE518 residues. These residues are shown in the Figure 6.1. Moreover, backbone oxygen atom of S473 and T474 form hydrogen bonds with the nitrogen atoms in the ligand (Tables 6.4 and 6.5) and positively charged R513 have electrostatic interaction with NO2 group of the SHetA2.

We performed several mutations on mortalin SBD to identify the essential residues involved in the binding. Including substrate binding domain residues R513E, S473A, T449A, and V482F. Among which we mention V482F in this paper.

In the case of V482F, similar mutation in DnaK, a bacterial homolog of mortalin, suggested defective interactions with substrate binding proteins [136] thus it is expected to be the case in moratlin SBD as well. Experimental results from also show this mutation abrogates p53-mortalin interaction [173]. We calculated the binding energy using GROMACS and g_mmpbsa tool to investigate the effect of this mutant. The structure of the wild type and mutant protein in complex with SHetA2 is displayed in Figure 6.3.

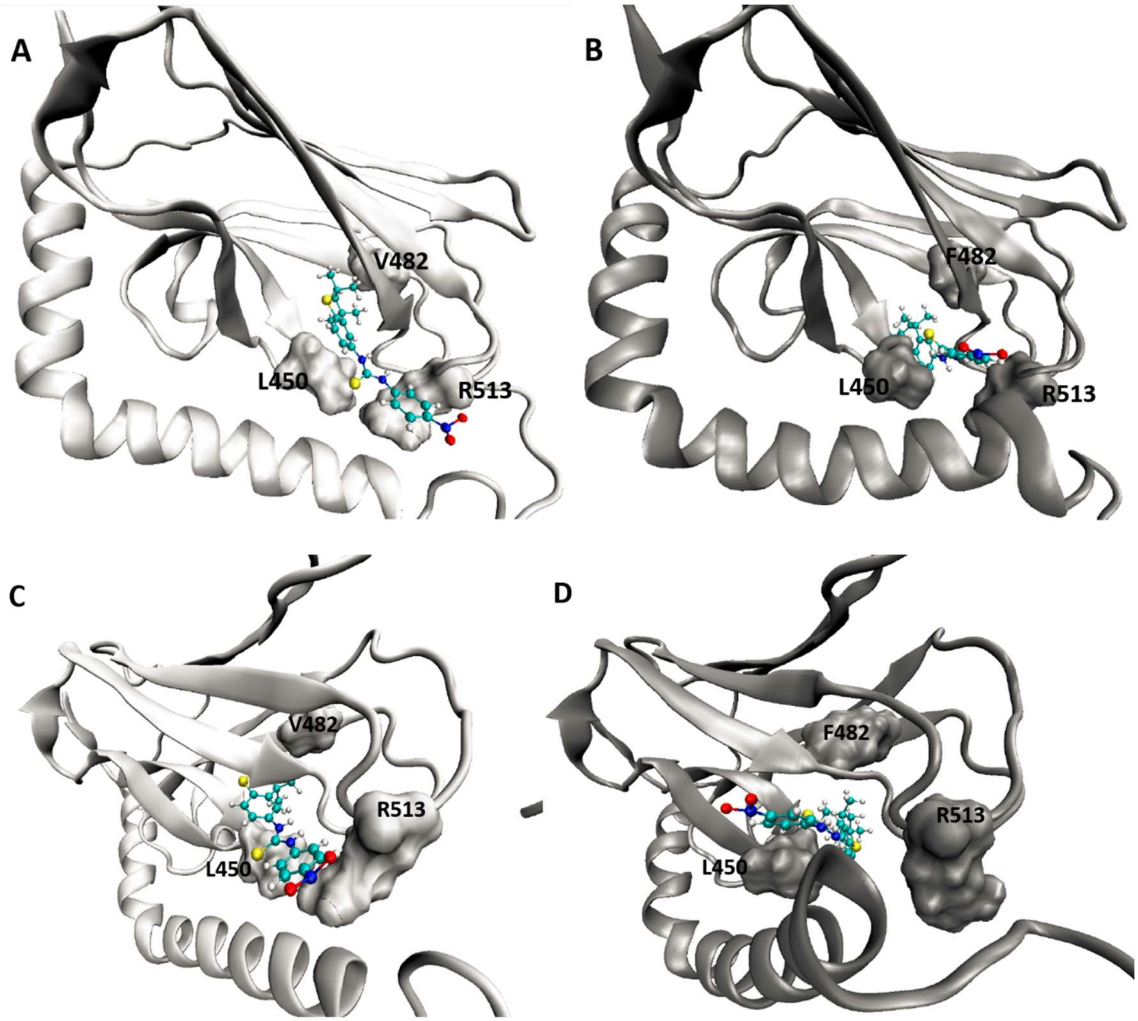


Figure 6.2 WT and mutant SBD mortalin in complex with SHetA2 after 50 ns MD simulations using GROMACS and visualized by VMD
(A) SHetA2 in the binding pocket of the WT protein-Front view
(B) SHetA2 in the binding pocket of the mutant protein (V482F)-Front view
(C) SHetA2 in the binding pocket of the WT protein-Side view
(D)SHetA2 in the binding pocket of the mutant protein (V482F)-Side view

In the front view is observed that in the mutant protein ring A of the ligand is pushed down, away from F482 and closer to L450. In the side view we can see ring A is now further from R513.

Considering the previously found results that showed V482F mutation breaks the interaction of mortalin and p53 and the result from our simulation studies that revealed the change in the conformation of SHetA2 in the binding pocket, it appears V482 can be involved in

both p53 and ShetA2 interactions with mortalin. SHetA2 may compete with p53 to bind to mortalin and abrogate their interactions.

Using g_mmpbsa, we performed energy calculations for the ligand-protein interaction after MD simulations on protein-ligand for 50 ns. Table 6.1 contains several lowest energy residues sorted based on mechanical energy with the corresponding energy terms of MM, polar, nonpolar, and total energy in kJ/mol. The only difference seems to be Arg 513 which is among 7 lowest energy residues in wildtype SBD and is absent from the list of lowest residues for the V482F mutant. And the Gly 451 which is one the lowest energy residues in the mutant and is not among the 7th lowest energy residues of wildtype. R513 is no longer reachable because the NO2 of SHetA2 gets distance from R513 when the other end of the molecule is pushed down by F482 in the mutant.

Table 6.1 Residue-wise contribution to different energy terms for WT and V482F mutant

Residues (WT)	**MM Energy (kJ/mol)**	**Polar Energy (kJ/mol)**	**Apolar Energy (kJ/mol)**	**Total Energy (kJ/mol)**
ALA-475	-7.2900	1.1470	-0.3170	-6.4600
ARG-513	*-7.3290*	*7.0320*	*-0.5420*	*-0.8390*
THR-449	-7.9330	3.4610	-0.3430	-4.8150
PHE-472	-9.8360	1.3820	-0.5480	-9.0020
SER-473	-10.2440	9.4460	-0.7880	-1.5860
THR-474	-10.3930	5.5310	-0.1360	-4.9980
LEU-450	-12.2720	3.1210	-0.9680	-10.1190

Residues (V482F)	**MM Energy (kJ/mol)**	**Polar Energy (kJ/mol)**	**Apolar Energy (kJ/mol)**	**Total Energy (kJ/mol)**
GLY-451	*-6.8240*	*4.5710*	*-0.7830*	*-3.0360*
THR-474	-7.0700	4.1320	-0.1520	-3.0900
ALA-475	-8.8400	3.2710	-1.3340	6.9030
THR-449	-9.4750	4.8310	-0.1110	-4.7550
SER-473	-10.4920	5.2010	-0.7410	-6.0320
PHE-472	-10.7700	2.4760	-0.3520	-8.6460
LEU-450	-17.3880	5.3920	-0.9520	-12.9480

E_{MM} is the molecular mechanics potential energy in vacuum and includes both bonded (such as bond, angle, torsion energies) and nonbonded terms (such as van der Waals ad electrostatic interactions). Moreover, polar and apolar energy terms are the components of free energy of solvation.

In Table 6.2, average binding energy of the two systems (wild type and mutant) is listed along with different nonzero energy terms.

Table 6.2 Comparison of different energy terms

Energy Terms	WT (kJ/mol)	V482F (kJ/mol)
Van der Waal	-175.869±10.327	-206.796±11.723
Electrostatic	-13.497±10.327	-18.579±2.235
Polar Solvation	75.476±5.089	105.476±10.483
SASA	-17.821±0.772	-20±167
Binding Energy	-131.711±10.096	-140.066±10.954

Solvent Accessible Surface Area (SASA) is a model that is used to estimate nonpolar solvation energy counting for the repulsive(G_{cavity}) and attractive (G_{vdw}) forces between the solute and the solvent upon the formation of the cavity.

The lowest energy frame has been obtained from the trajectory file of the system and the corresponding energy of that frame is calculated to be -167.901 kJ/mol for the wildtype and -182.956 kJ/mol for the V482F mutant. As listed in table 6.3 along with the average energy of the system.

Table 6.3 Binding Energy comparison

	WT kJ/mol	V482F
Min BE	-167.901	-182.956
Average BE	-131.711±10.906	-140.066±10.954

Table 6.4 contains percentages of the time when hydrogen bond is formed between the specific acceptor/donor partners. For example more than 90% of the whole simulation time, a hydrogen bond is formed between the carbonyl group of T474 and N1 nitrogen atom of SHetA2 linker while no hydrogen bond was observed to be formed between the T474 and S473 residues in V482F.

Table 6.4 Hydrogen bonding comparison
Lig_N and Lig_N1 are the two nitrogen atoms in SHetA2 thiourea linker, with N being the atom closer to the double ring.

Acceptor/Donor	WT	V482F
S473_CO_Lig_N	51.0%	0%
S473_CO_Lig_N1	2.14%	0%
T474_CO_Lig_N	30.06%	0%
T474_CO_Lig_N1	92.76%	0%

The distance between the H-bond participant atoms is calculated at the lowest binding energy time frame given in the next table. The hydrogen bond cut-off is 0.35 Angstrom.

Table 6.5 Distance (nm) at the point of min BE

Acceptor/Donor	WT	V482F
S473_CO_Lig_N	0.314	0.369
S473_CO_Lig_N1	0.408	0.343
T474_CO_Lig_N	0.332	0.374
T474_CO_Lig_N1	0.310	0.469

The difference in binding energies of the mutant and WT can have different reasons. One reason can be that phenylalanine contains a highly hydrophobic benzene ring so the contribution of this residue to the binding energy is higher in the wildtype. Also the ligand gets closer to the Leucine 450 and this can explain the increased contribution of this residue to the MM energy in the mutant, thus the increase in total binding energy of the mutant.

6.4. Conclusion

We identified the binding configurations and the interacting residues of mortalin with SHetA2 using MD simulations. Mutations in the protein sequence can alter the binding configuration and the binding energies. We studied V482F mutation in this study. Energy calculations have been performed on both mutant and wild type protein and the result have been compared. The bulky side chain of phenylalanine at position of amino acid 482 tends to significantly reduce the available space in the hydrophobic pocket. Thus, the ring A of SHetA2 is pushed down and away from the bulky F482. Ring B (containing NO2) gets tilted away from R513 and gets closer to the Leucine 450. As it is observed in table 6.5, the mutation of Valine482

to Phenylalanine breaks the hydrogen bonds that was formerly observed between T474-LigN,N1 and S473-LigN,N1.

Considering the previously found results that showed V482F mutation breaks the interaction of mortalin and p53 and the result from our simulation studies that revealed the change in the conformation of SHetA2 in the binding pocket, it appears V482 can be involved in both p53 and ShetA2 interactions with mortalin. SHetA2 may compete with p53 to bind to mortalin and abrogate their interactions.

CHAPTER VII

OXYGEN-CONTAINING ANALOGS OF THE FLEX-HET ANTICANCER AGENT SHETA2

7.1. Introduction

The currant chapter is an overview of our published paper on the novel activity of oxygen- versus sulfur-containing analogs of the SHetA2.

As our results from molecular mechanics studies suggests, hydrophobicity of the ligand is an important factor in its binding performance. Thus, we propose several analogs for SHetA2 and investigated the effect of hydrophobicity on their efficacy and potency. In addition, to determine if certain fragments of the lead molecule SHetA2 could be modified to enhance aqueous solubility, targets were synthesized with an oxygen-containing chroman unit, in place of the thiochroman, and both urea and thiourea containing compounds were studied. The aim of this work was to develop synthesis of oxygen counterparts of SHetA2 (structures **2-8** in Figure 7.1), with sufficient structural diversity to allow a modest structure-activity relationship (SAR) analysis of this family of structures. Five series of chromans with urea and thiourea bridges connecting a chroman unit (ring A) and a single 4-substituted benzene unit (ring B) have been prepared and evaluated relative to SHetA2 for activity against the human A2780 ovarian cancer cell line. Substituents on the chroman unit were modified to appraise the importance of hydrophobicity in this moiety, which had been suggested by molecular modeling analyses.

The 2-Me-4-Me series (two sets of geminal dimethyl groups at C2 and at C4 on the ring A unit) permitted direct comparison with SHetA2. Ring B in this series was evaluated with specific functional groups at C4 on the ring, including NO_2, CO_2Et, CF_3, OCF_3, CN and SO_2NH_2.

The 2-H-4-Me series (only one geminal dimethyl group at the C4 position on ring A) permitted structure-activity relationship analysis to relate the critical importance of the hydrophobic geminal dimethyl groups on ring A to the activity of SHetA2. The remaining three series 2-Et-4-Me, 2-Me-4-Et and 2-Et-4-Et (ring A methyl groups replaced with ethyls at C2, at C4 and at both C2 and C4, respectively) offered the opportunity to modulate the hydrophobicity of the chroman moiety. Additionally, in all these series, the influence of a urea versus a thiourea bridge was also investigated. The results of these modifications are summarized below. The exact analog of SHeA2 with oxygen substituted for sulfur in ring A (**2a**) showed comparable efficacy but a significant reduction of inhibitory prowess against the ovarian cancer cell line. However, the urea bridged analogues bearing CF_3, CN, and OCF_3 at C4 of ring B (**3c**, **d**, and **f**) had higher IC_{50} values and greater efficacy than SHetA2. Removing the geminal dimethyl group at C2 (**4a-c**, **5a-c**) caused a significant lowering of the efficacy and the percent of growth inhibition for all of the derivatives screened, indicating that the hydrophobic geminal dimethyl group at C2 in ring A is crucial for activity. In general, ethyl groups at C2 and C4 of ring A improved potency of the compounds. Replacing the geminal dimethyl groups with geminal diethyls on ring A in the urea derivatives gave **6b-c**, **7c-d** and **8b**, all of which outperformed SHetA2 with respect to efficacy and IC50. The results for compounds **6-8** are in concurrence with modeling studies, which predicted that greater hydrophobicity in ring A would be beneficial. Binding energies were determined for compounds docked *in silico* to mortalin. The urea bridge showed activity comparable to or, in some cases, greater than compounds with a thiourea bridge. Several compounds achieved 94% efficacy and an IC_{50} of 2 μM, which were better than SHetA2 (84%, 3 μM).

Preliminary data revealed that substituting a urea linker for a thiourea linker demonstrated that the former exhibited an intramolecular stabilizing H-bonding effect while the thiourea did not [174]. Consequently, it is not unreasonable to expect stronger H-bonding capabilities by the urea linker compared to the thiourea linker when docked in mortalin. Moreover, the oxygen analogues would likely be more stable to oxidative degradation both *in vivo* and in long-term storage.

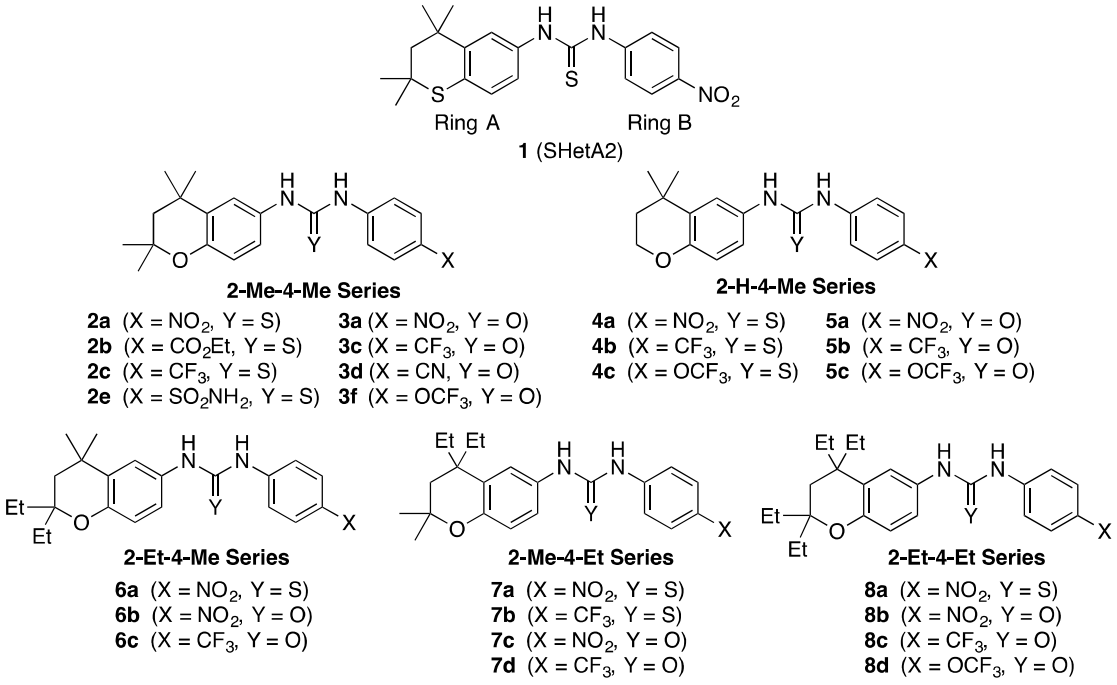


Figure 7.1 Target compounds

7.2. Material and Methods

7.2.1. Chemistry

In view of the established interaction of SHetA2 with mortalin [26], the nature of the heteroatom in ring A and the substituents in ring B were both deemed important. Consequently, the selection of O in ring A and highly polar substituents in ring B were investigated as related to biological activity. The influence of larger alkyl groups on ring A was also examined. In an effort to determine the influence of the position and size of hydrophobic groups on the activity,

syntheses were initiated on three additional sets of targets, **6a-c** (2-Et-4-Me Series), **7a-d** (2-Me-4-Et Series), and **8a-d** (4-Et-4-Et Series). Detailed discussion on the synthesis of compounds is out of the scope of this study. For a detailed information refer to [73].

All of the substituents attached to ring B in **6-8** had H-bonding capability with polar groups in the binding site of mortalin. Members of **6** retained the geminal dimethyl group at C4, but a geminal diethyl group was added at C2. Replacement of geminal dimethyl groups at C4 with geminal diethyl groups in **7** allowed an assessment of the geometric effects of the larger ethyl groups on the activity. System **8**, with geminal diethyl groups at both C2 and C4, should increase the overall capacity to interact with hydrophobic centers in mortalin. Increasing the hydrophobic properties of the systems was envisioned with the 2,2,4,4-tetraethyl analogues (2-Et-4-Et, Series **8**).

7.2.2. Modeling Studies

Autodock 4.2 [15] was used to dock the compounds to the substrate binding domain of mortalin (Protein Data Bank ID: 3N8E). For the individual compounds ChemSketch™ (Advanced Chemistry Development, Inc. ADC/Labs, Toronto, Ontario, Canada) was employed to generate the SMILES notations which were subsequently converted to PDB files with initial three-dimensional coordinates using Open Babel GUI [162]. AutoDockTool [175] was then used to prepare the protein with partial charges and ligand with rotatable bonds. Only polar hydrogens were retained in the molecules. Gasteiger united atom partial charges and solvation parameters were assigned. The search space of 44 Å*47 Å*41 Å was slightly larger than the protein molecule, and the grid spacing was 0.375 Å. Autogrid was run first to prepare the coordinate system and the Lamarckian genetic algorithm was applied with a population size of 150 and 5 million maximum evaluations. The minimum empirical binding free energy between each ligand

and receptor was recorded. All dockings were performed on an iMac computer with a 2.4 GHz intel Core i3 processor and with 4 GB RAM.

7.3. Results and Discussion

To verify whether compounds shown in Figure 7.1. have the biological effects expected from the chemical structural changes to SHetA2, all compounds were screened against A2780 human ovarian cancer cells by Tim Pouland and Doris Benbrook in the Department of Obstetrics and Gynecology, Stephenson Cancer Center, University of Oklahoma Health Sciences Center. The potency (half-maximal inhibitory concentration, IC_{50}) and efficacy (the maximal inhibition of cancer cell growth), have been listed in Table 7.1.

The cell count and ligand concentrations obtained from MTT assay, has been plotted by fitting a sigmoidal function. Figure 7.2A shows this sigmoidal fit for SHetA2 and Figure 7.2B is the plot for compound **7c.** IC50 values are found from this plot and depicted in Figure 7.4. the horizontal axis shows the concentration of the ligand and the vertical axis is the cell counts.

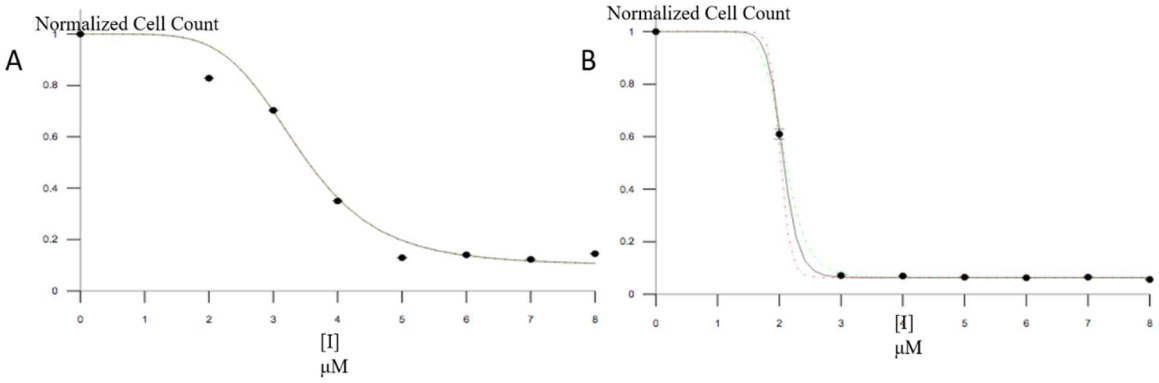


Figure 7.2 Sigmoidal plot for (A) SHetA2 and (B) analog 7c

The MTT Cell Proliferation Assay measures the cell proliferation or the reduction in cell viability in the case of apoptosis or necrosis. Cells are stained by a dye which has different color in viable(living) versus dead cells. Then the absorbance is measured by a spectrophotometer in untreated cells as well as the treated cells with different doses of the agent. Absorbance values are

an indicative of a reduction in the rate of cell proliferation. The effectiveness of the compound in causing death of cells can be measured by a dose-response curve.

The interactions of the compounds with the mortalin substrate-binding domain were studied by virtual docking, with the resulting binding free energy values (-ΔG) also listed in Table 7.1. Potential correlations between binding energy and potency, between binding energy and efficacy, and between potency and efficacy are displayed in Figures 7.3A, B, and C, respectively. All compounds showed various degrees of growth inhibition. In Figure 7.3A, a significant positive correlation (Pearson correlation coefficient $r = 0.52$ with $p = 0.007$) between the energy and potency values was observed, which was consistent with compounds having stronger binding affinities for mortalin (higher -ΔG values) and greater potency (lower IC_{50} values). In Figure 7.3B, a significant negative association ($r = –0.60$, $p = 0.001$) between the energy and efficacy values was observed, which was consistent with compounds having stronger binding affinities for mortalin (higher -ΔG values) and having higher efficacy (higher % inhibition). In Figure 7.3C, the potency and efficacy values for these compounds showed a significant negative correlation ($r = –0.51$, $p = 0.008$), which was consistent with the expectation that a compound with a smaller IC_{50} has higher efficacy in competing for the receptor mortalin versus its partner proteins. In this plot, a sigmoid curve was sketched to capture the main trend in the data. Most compounds with IC_{50} below 5 μM demonstrated efficacy comparable to or better than that of SHetA2. Since a small IC_{50} value necessarily means a stronger affinity for receptor binding, the data seemed to indicate that these compounds needed to have a binding affinity above a certain threshold to effectively inhibit the cancer cells. This threshold is likely determined by interactions between the receptor and its client protein(s), which are competitively displaced by these compounds.

Table 7.1 IC_{50}, Efficacy, Standard Errors of Mean (SEM), and binding free energy (ΔG) of different compounds docked to the mortalin substrate-binding domain (SBD).

	IC_{50} (μM)	**IC_{50} SEM**	**Efficacy (%)**	**Efficacy SEM**	**-ΔG (kcal/mol)**
SHetA2	3.17	0.05	84.3	0.7	8.5
2a	6.97	0.08	87	6	8.1
2b	4.3	3.0	26	10	8.0
2c	4.7	0.2	22	4	7.8
2e	6.9	0.9	32	2	8.3
3a	4.1	0.1	79	4	8.5
3c	3.6	0.1	88.4	1.4	7.9
3d	4.7	0.3	93	3	8.6
3f	4.56	0.05	91.5	0.8	7.7
4a	6.4	0.6	19	5	7.8
4b	5.5	1.5	28	0.6	7
4c	4.3	0.4	16	6	7.3
5a	6.7	0.2	74	11	8.3
5b	5	0.3	47	2	7.8
5c	3.5	0.4	36	3	8.1
6a	2.9	0.1	93.9	0.5	8.5
6b	2.17	0.04	93.2	0.1	8.3
6c	2.45	0.04	92.4	0.1	8.1
7a	3.69	0.04	95.7	0.4	8.4
7b	4.7	0.5	11.8	0.9	8.2
7c	2.05	0.02	93.66	0.05	8.4
7d	2.43	0.05	93.3	0.1	8.6
8a	4.6	0.2	94.00	1.00	8.7
8b	2.09	0.02	91.4	0.3	8.7
8c	2.0	0.1	86.2	0.6	8.8
8d	3.00	0.06	67	3	8.8

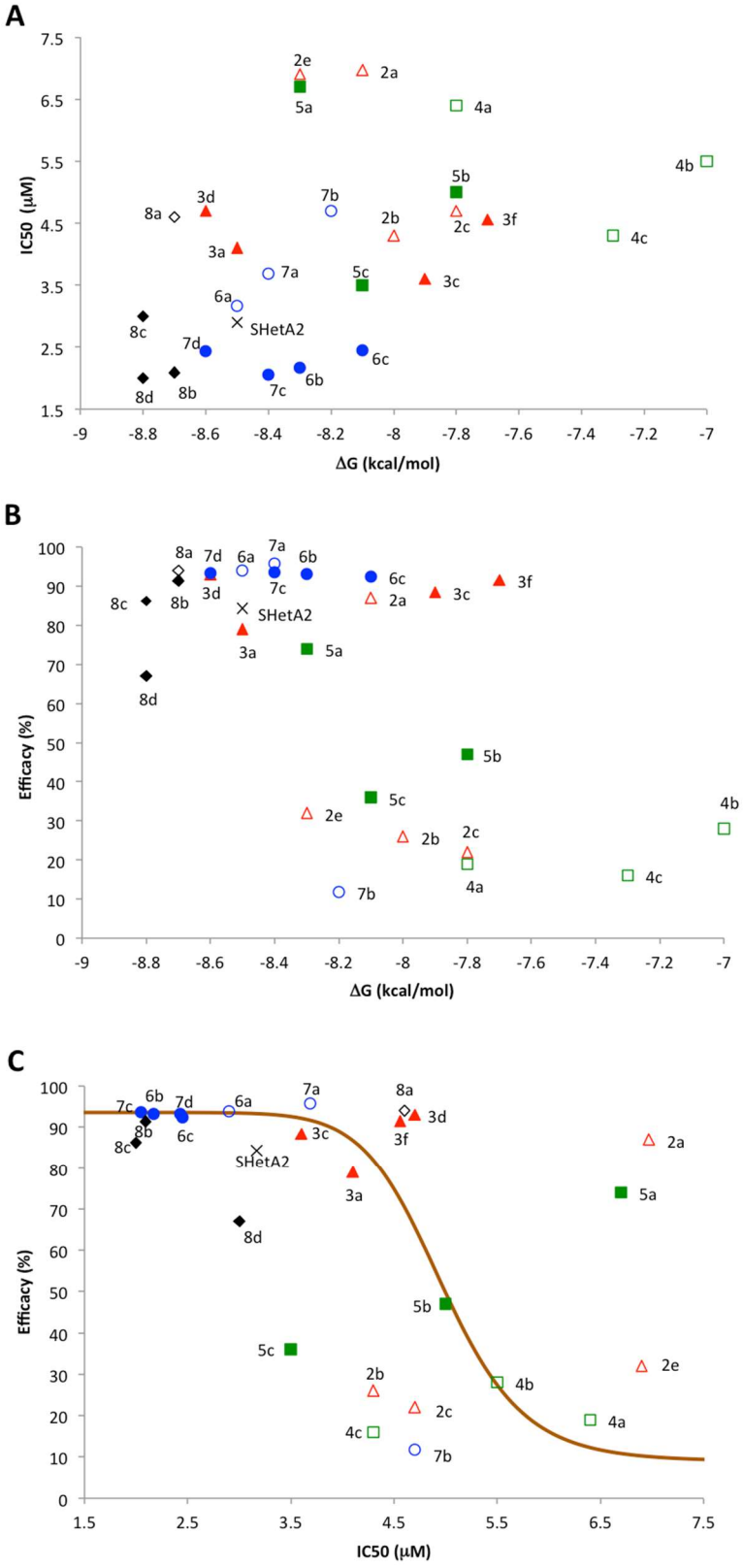

A
IC50 (μM)
ΔG (kcal/mol)
B
Efficacy (%)
ΔG (kcal/mol)
C
Efficacy (%)
IC50 (μM)
SHetA2

Figure 7.3 Assessment of (A) binding energy-IC_{50}, (B) binding energy-efficacy, and (C) IC_{50}-efficacy correlation for all compounds investigated in this study. Data for the lead compound SHetA2 was represented by "X", the 2-Me-4-Me series by red triangles, the 2-H-4-Me series by green squares, the 2-Et-4-Me and 2-Me-4-Et series by blue circles, and the 2-Et-4-Et series by black diamonds. Moreover, filled symbols were used for compounds with oxygen atoms in both the linker and ring A positions, while empty symbols are used for compounds with oxygen in ring A and sulfur in the linker. Data in (C) was fitted by a sigmoid curve with half-activity concentration of 4.9 μM.

The most desirable properties of the target molecules were observed as a combination of the smallest IC_{50} and the highest efficacy. The group of compounds **6b**, **7c**, and **8b** (X=NO_2 , Y=O) are clearly shown as the best in Figure 7.3C, demonstrating excellent efficacy values of 91-94% and IC_{50} values of 2.0-2.4 μM, both of which are a significant improvement over SHetA2 (84% and 3.2 μM). These three most active compounds had three common chemical features. First, all possessed increased hydrophobicity due to the presence of geminal diethyl groups instead of the geminal dimethyl groups in ring A. Second, they all had a urea linker (Y = O). Third, they had a nitro group (X = NO_2) attached to ring B. Docking studies showed that the NO_2 group is capable of forming two hydrogen bonds with the amino acids on mortalin, enhancing the ligand-receptor binding affinity. The second best group of **6c**, **7d**, and **8c** (X=CF_3 and Y=O) had IC_{50} and efficacy values only slightly lower than the first group. They differed from the first group only in the third chemical feature mentioned above, the identity of the polar group attached to ring B, indicating that CF_3 was slightly less effective than NO_2. The third group of **6a**, **7a**, and **8a** (X = NO_2 and Y = S) varied from the first group (**6b**, **7c**, and **8b**) only in the second chemical feature, having a thiourea instead of a urea linker. Their efficacy values were comparable to or even slightly better than the first groups, which might be attributed to a more flexible thiourea linker compared to the stiff peptide bonds involved in the urea linker. However, the third group had much lower potencies (higher IC_{50} values) than the first group. The importance of the first chemical feature, increased hydrophobicity due to the larger ethyl groups in compounds **6-8** can be further appreciated by observing the dramatically deteriorated performance of compounds **4**

and **5**(2-H-4-Me), both of which had reduced hydrophobicity due to the absence of the C_2 geminal dimethyl.

Nevertheless, compounds **8** (2-Et-4-Et) underperformed relative to compounds **6** (2-Et-4-Me) and **7 (**2-Me-4-Et), which might indicate that the chroman unit in **8** became too large with geminal diethyl groups at both C2 and C4. Compounds **6** appeared to have slightly more consistent performance than **7**, noting that **7a** had relatively large IC_{50}. It seemed more beneficial to place the geminal diethyl groups at C2 (as in compounds **6**) than at C4 (as in **7**), a point to be justified by molecular docking studies (*vide infra*). Therefore, the appropriate amount of hydrophobicity, the size of the chroman unit, and the position of the geminal diethyl groups in ring A might all have influence on the receptor binding and cancer inhibition. The remaining compounds in the **6**-**8** groups [**7b** (X=CF_3 with Y=S) and **8d** (X=OCF_3 with Y=O)] exhibited less valuable performance.

The major difference between the two 2-Me-4-Me series **2** and **3** was that the former had a thiourea linker and the latter a urea linker. Both retained geminal dimethyl groups as in SHetA2, but the sulfur heteroatom in ring A was replaced by oxygen. Compounds in series **2** displayed a weak performance with respect to both efficacy and IC_{50} values. Compounds in series **3** demonstrated efficacy comparable to or slightly better than SHetA2, but all had higher IC_{50} values. Quality performance occurred when both the linker and the chroman heteroatom were oxygen or sulfur (as in SHetA2), but the properties deteriorated when the hereroatoms in these two fragments were different. This peculiar requirement might be connected with the flexibility of the linker, with the thiourea linker being more flexible than urea linker, and the ability to form hydrogen bonding with the receptor (*vide infra*). Structures **2a** and **5a** are also somewhat unusual in the sense that they possess respectable efficacy values, but have modest IC_{50} values. Of course, the possibility exists that a different receptor might conceivably be involved.

Figures 7.3A and 7.3C showed the structure model of **7c**, one of the best compounds, docked to the mortalin in its peptide/substrate-binding channel. Figures 7.3B and 7.3C show urea

oxygen h-bonded with S473 and the nitro group h-bonded with Q517. Therefore, the preference for having a urea linker and a nitro group can be understood as they are strong hydrogen bond acceptors, while sulfur has been shown to be a rather poor H-bond acceptor [176]. Nevertheless, this docking conformation was not static as several other docking conformations with similar binding energies found different amino acid h-bonding partners. This conformational diversity contributed entropically to the stabilization of the protein-ligand complex [177, 178]. However, such a conformational entropic contribution is a challenge to quantify, and it was not considered in the binding energies reported in Table 7.1. The chroman ring A with methyl and ethyl attachments was accommodated in the hydrophobic pocket formed by residues V482, L450, I447, and the methyl of T449 (Figures 7.4B and 7.4C). This is consistent with a report that mutation of V482 to phenylalanine abrogated the interaction with tumor suppressor protein p53 [27], likely because the bulky side chain of phenylalanine at that position tends to significantly reduce the available space in the hydrophobic pocket.

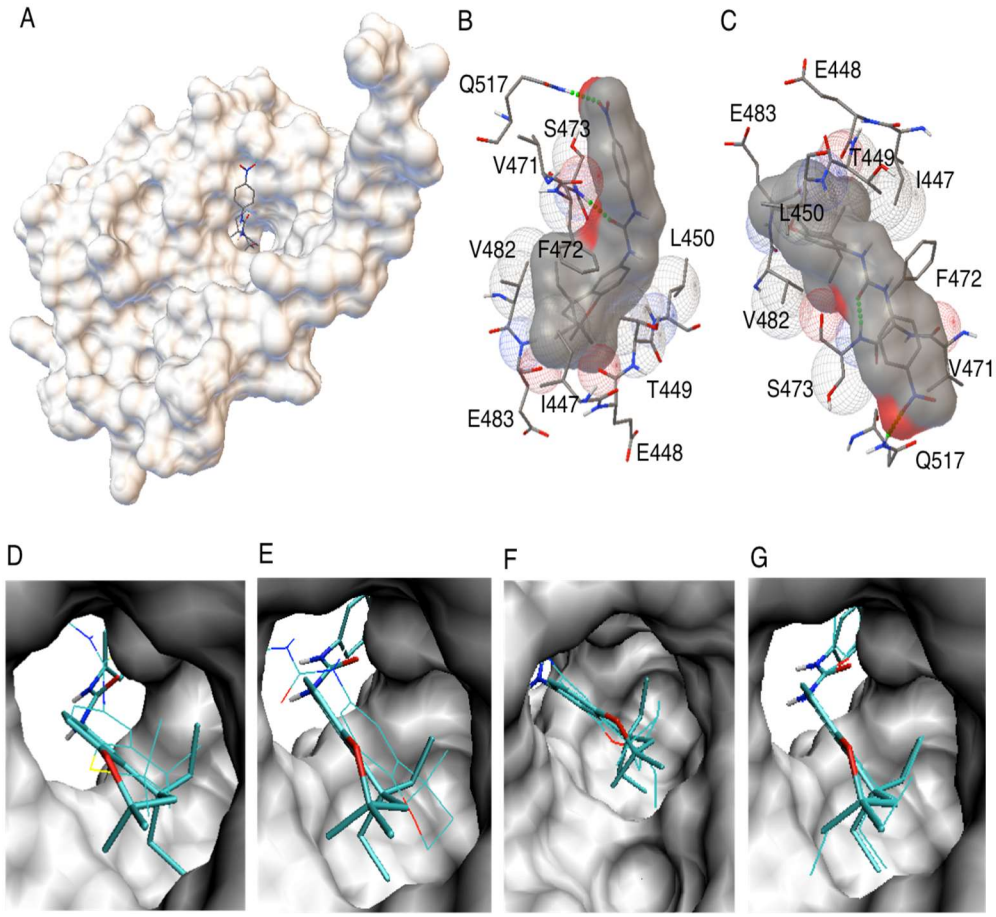


Figure 7.4 Molecular docking of compounds **7c** and several other compounds with the substrate-binding domain of mortalin (PDB ID: 3N8E).
(A) The complex with compound **7c**. (B)-(C) Two views of the same complex highlighting the amino acids in contact with **7c**. The two hydrogen bonds were indicated by green dotted lines. (D)-(G) Comparison of hydrophobic ring A conformations between **7c** and **SHetA2**, **5a**, **6b**, and **8b**, respectively, where **7c** was shown in thicker lines. These four figures were viewed from the other end of the tunnel opposite to Figure 7.3A.

In Figures 7.4D-G, the conformations of the hydrophobic ring A of SHetA2, **5a**, **6b**, and **8b** were compared with that of **7c**. All compounds docked in the same hydrophobic pocket of the substrate-binding channel, yet with subtle but important differences. In comparison to SHetA2, ring A of **7c** (X=NO_2 , Y=O, 2-Met-4-Et) was pushed further toward the opening of the tunnel (closer to the viewer) by the bulky diethyl groups (Figure 7.4D). Therefore **7c** was expected to displace p53 or another client protein more effectively than SHetA2. On the other hand, ring A of

5a with only one dimethyl pair, retreated from the opening since the smaller ring posed less steric hindrance to being drawn by polar groups (including the nitro group) into the binding cavity through interaction with the protein (Figure 7.4E). Therefore, **5a** (X=NO_2, Y=O, 2-H-4-Met) was expected to be less effective in displacing the client protein. Relative to **7c**, in compound **6b** (X=NO_2, Y=O, 2-Et-4-Met) the diethyl pair switched position with the dimethyl pair. In order to pack the more hydrophobic diethyl groups, ring A roughly flipped by 180º, taking the heteroatom oxygen from the channel opening to the interior of the hydrophobic pocket (Figure 7.4F). Such a flip also caused the nitro group to switch its hydrogen bond partner to R513. Consequently, ring A of **6b** was slightly more distant from the channel opening than **7c**. The docking conformation of **8b** (X=NO_2, Y=O, 2-Et-4-Et) nearly overlapped that of **7c**, since in **8b** diethyl groups simply replaced the dimethyl groups next to the heteroatom. As discussed earlier, **6b** and **8b** had nearly the same biological activities as **7c**.

7.4. Conclusions

Guided by the results from our docking studies of SHetA2 analogs that indicate hydrophobicity is an important factor in the binding of the ligand and increasing the hydrophobicity leads to higher affinity to the SBD mortalin, several analogs were synthesized and biological studies were conducted to mainly study the effect of hydrophobicity of the ligand on the potency and efficacy of the binding. The effect of various substituents on ring A and urea and thiourea linker was tested as well.

It was shown that electron-withdrawing groups (NO2, CF3) on the B ring elicited IC50 values and efficacy responses as good as or better than that of SHetA2. Moreover, the best examples also had geminal diethyl groups replacing geminal dimethyl groups at C2, including either O or S in ring A. When geminal diethyl groups were placed at both C2 and C4 of ring A, the efficacy values paralleled that in SHetA2 while the IC50 values varied. Thus, the presence of

two geminal diethyl groups appeared to make ring A too bulky to be accommodated in the binding site of mortalin. Compared to SHetA2, the most exceptional agents were **6b**, **7c**, and **8b**, exhibiting efficacy values of 91-94% with IC50 values of 2.0-2.4 mM (SHetA2 had 84% and 3.2 mM, respectively). In general, the presence of a urea linker between rings A and B gave improved activity compared with a thiourea linker. Binding of **7c** in the substrate binding site of mortalin resulted in H-bonding of the urea oxygen with amino acid S473 and H-bonding of the nitro group with Q517. All compounds docked in the hydrophobic pocket with binding ΔG values ranging from 7.0 to 8.8 kcal/mol. These binding energies were consistent with a common binding site for all compounds. Some fine tuning of related structures is expected to enhance the anticancer activity against the human A2780 ovarian cancer cell line.

CHAPTER VIII

NITROGEN-CONTAINING ANALOGS OF THE FLEX-HET ANTICANCER AGENT SHETA2

8.1. Introduction

The currant chapter is an overview of our paper on studying the Tetrahydroquinoline units in heteroarotinoids and their anti-cancer activities.

Nitrogen occurs in a very high percentage of FDA approved pharmaceuticals [179]. The N atom renders weakly basic compounds with good H-bonding properties, allowing them to bind to a wide range of proteins. Thus, nitrogen analogs of SHetA2 (Figure 8.1) are reasonable candidates for evaluation in ovarian cancer screens. the current work utilized a tetrahydroquinoline ring A series to assess the activity of analogs with a carbonyl group and a hydroxyl group in ring A, with both thiourea versus urea linkers and various substituents in ring B.

A series of compounds having a tetrahydroquinoline (ring A) connected by a urea or thiourea linker to a 4-substituted benzene (ring B) has been prepared and evaluated relative to SHetA2 in terms of inhibition of A2780 ovarian cancer cells. Six of the derivatives had equal or higher efficacy than SHetA2. Series **1**, lacking the methyl on the ring A nitrogen and the *gem*-dimethyls on the adjacent carbon (R, R' = H, G = CH_2), showed only weak activity. Salt **2**, the quaternized *N*,*N*-dimethyl iodide salt analog of **1a**, also possessed very modest growth inhibition

In the cell line studied. Series **3** compounds, which had these two modifications in ring A (G = CO and N-CH_3 replacing the heteroatom S in SHetA2), were most successful. Compound **3a** (R, R' = CH_3, G = C=O, Y = O, Z = NO_2) had much better efficacy (defined as the maximal inhibition of cancer cell growth, 94.8%) than SHetA2 (84.3%), but slightly worse potency (measured by half-maximal inhibitory concentration IC_{50} 3.8 μM) than SHetA2 (3.2 μM). In addition, in series **3**, Z = CF_3 (**3c** and **3d**, Y = O and S, respectively) and Z = OCF_3 (**3e** and **3f**) were also evaluated since these are also electron-withdrawing groups with H-bonding capability. Both displayed excellent potency and efficacy. Compounds **3e** and **3c** showed improvement in both efficacy and potency. In general, when the linker group between rings A and B was urea, efficacy values modestly exceeded those with a thiourea linker in the carbonyl-containing tetrahydroquinoline systems (series **3**). In contrast, when ring A possessed the 1,2,2,4,4-pentamethyltetrahydroquinoline unit with a C3 hydroxyl group (G = CH(OH) series **4**), very poor efficacy and potency were observed. Model compound **5** (G = CH_2, R' = R = CH_3, Y = S, Z=NO_2), which differed from SHetA2 only at the heteroatom (N-CH_3 versus S) in ring A, demonstrated improved efficacy (91.7%) but less potency (4.5 μM IC_{50}).

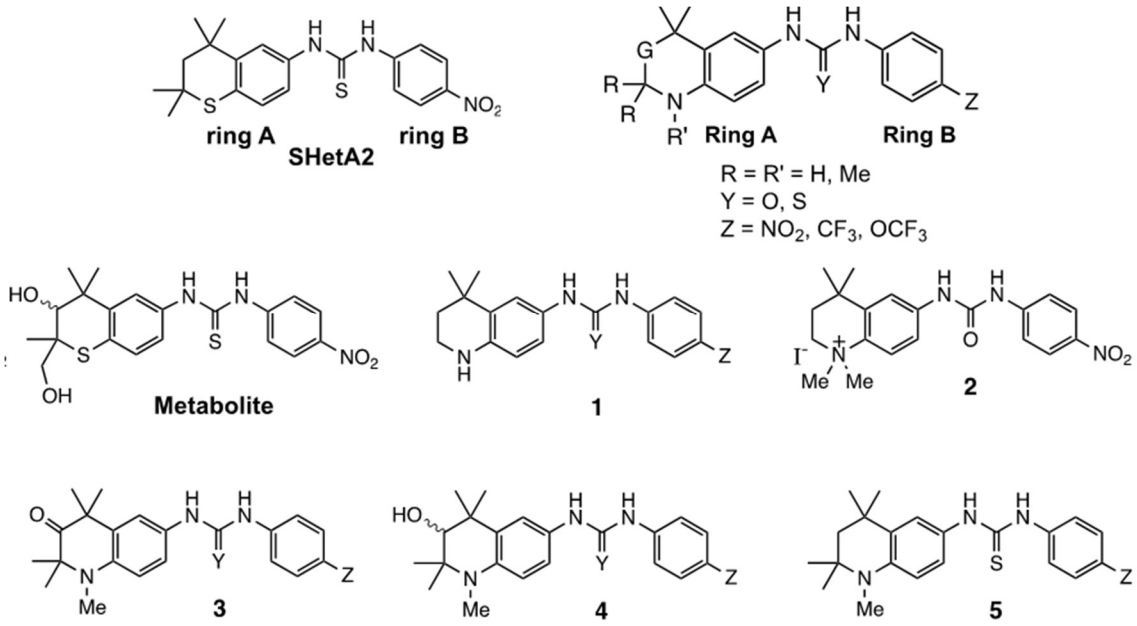


Figure 8.1 Target systems in nitrogen-containing heteroarotinoids.

8.2. Material and Methods

The preparation of members of **1** is out of the scope of this study. The target structures are listed in Figure 8.2.

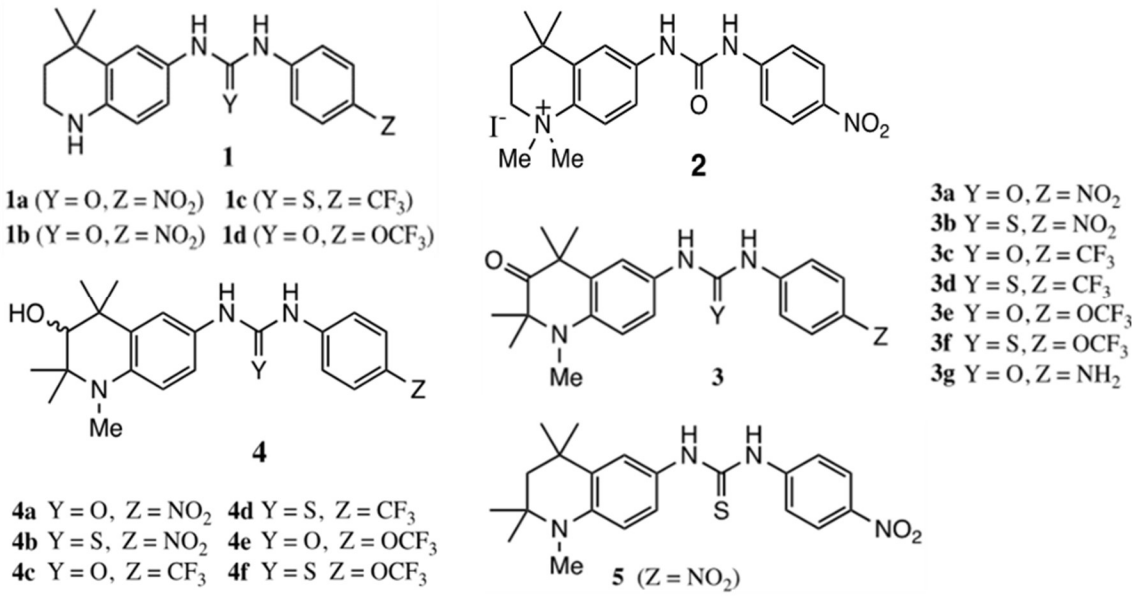


Figure 8. 2 Structure of compound series 1, 2, 3, 4, and 5

All of the compounds were fully characterized by IR, ^{1}H NMR, and ^{13}C NMR analyses. Final products were also confirmed by elemental analyses.

8.2.1. Biological Section

The biological tests have been performed by Doris Benbrook and colleagues in the Department of Obstetrics and Gynecology, Stephenson Cancer Center, University of Oklahoma Health Sciences Center. The compounds were dissolved in DMSO at a concentration of 0.01 M. The human ovarian cancer cell line A2780 was treated and dose-response curves were fitted to extract the IC_{50} and efficacy (the maximal % growth inhibition) parameters.

8.2.2. Theoretical Docking Methods

AutoDock 4.2 [15] was used to dock the compounds to the substrate binding domain of mortalin (Protein Data Bank ID: 3N8E). For the compounds, ChemSketch™ (Advanced Chemistry Development, Inc. ADC/Labs, Toronto, Canada) was used to generate the SMILES notations, which were subsequently converted to PDB files with initial three-dimensional coordinates using OpenBabelGUI [162]. AutoDockTool (ADT) [175] was then employed to prepare the protein and compounds with partial charges and for the latter rotatable bonds too. Only polar hydrogens were retained in the molecules. Gasteiger partial charges and solvation parameter were assigned. The search space of 44 Å × 47 Å × 41 Å was slightly bigger than the protein molecule and the grid spacing was 0.375 Å. Autogrid was run first to prepare the coordinates system and then Lamarckian genetic algorithm was applied with a population size of 150 and 25 million maximum evaluations. The minimum empirical binding free energy (ΔG) between the compound and receptor was reported. The dissociation constant K_d is calculated from binding free energy using the relation $\Delta G = -RT \ln(K_d)$, where R is the gas constant. All dockings were performed on an iMac computer with a 2.4 GHz Intel Core i3 processor and 4 GB RAM.

8.3. Results and Discussion

The biological activities of all compounds were validated by screening with A2780 human ovarian cancer cells (Table 8.1). The half-maximal inhibitory concentration (IC_{50}) and efficacy for all compounds revealed a range of growth inhibition. The Pearson correlation coefficient between the two parameters was –0.82, indicating a highly negative correlation. This is consistent in that a compound with a smaller IC_{50} is usually more effective to compete for the receptor site against its substrate protein(s).

To better assess the dose-response results, data in Table 8.1 were visualized by plotting the efficacy against IC_{50} (Figure 8.3). In this plot, a sigmoid curve was sketched to capture the main trend in the data. Most compounds with IC_{50} below 5 μM demonstrated efficacy comparable to or better than that of SHetA2. Moreover, a small IC_{50} also usually means a stronger affinity for the compound to bind with its protein receptor.

Table 8.1 IC_{50}, efficacy, ΔG, and dissociation constant (K_d) values for different compounds docked to the mortalin substrate-binding domain.

Cpd	Y	Z	IC_{50} (μM)	Efficacy (%)	–ΔG (kcal/mol)	K_d (μM)
SHetA2	S	NO_2	3.17 ±0.05	84.3 ±0.7	8.5	0.6
1a	O	NO_2	6.9 ±0.2	17.1 ±1.2	8.2	1.1
1b	S	NO_2	7.1 ±0.3	17.8 ±1.6	7.9	1.6
1c	S	CF_3	6 ±0.2	42 ±3	7.5	3.3
1d	S	OCF_3	7.1 ±0.8	24 ±2	7.2	5.4
2	O	NO_2	6.6 ±0.3	22 ±4	8.5	0.7
3a	*O*	*NO_2*	*3.8 ±0.1*	*94.8 ±2.2*	*8.9*	*0.3*
3b	S	NO_2	4.4 ±0.2	91.4 ±1.7	8.2	1.1
3c	O	CF_3	2.58 ±0.08	90.1 ±1.4	8.0	1.5
3d	S	CF_3	3.9 ±0.1	90.8 ±2	7.9	1.6
3e	*O*	*OCF_3*	*2.4 ±0.2*	*91.3 ±1.3*	*7.9*	*1.8*
3f	S	OCF_3	5.4 ±0.6	76 ±8	7.7	2.4
3g	S	NH_2	7.7 ±1.4	24 ±4	8.2	1.1
4a	O	NO_2	8.4 ±1.9	26 ±4	8.0	1.6
4b	S	NO_2	10 ±5	25 ±4	7.7	2.3
4c	O	CF_3	6.7 ±0.5	25 ±5	8.4	0.7
4d	S	CF_3	7.8 ±0.2	23.6 ±3.3	8.2	1.1
4e	O	OCF_3	7.6 ±0.7	56.1 ±2.4	7.7	2.3
4f	S	OCF_3	13.1 ±6.1	15.3 ±3.3	7.6	2.9
5	*S*	*NO_2*	*4.5 ±0.1*	*91.7 ±0.4*	*8.7*	*0.5*

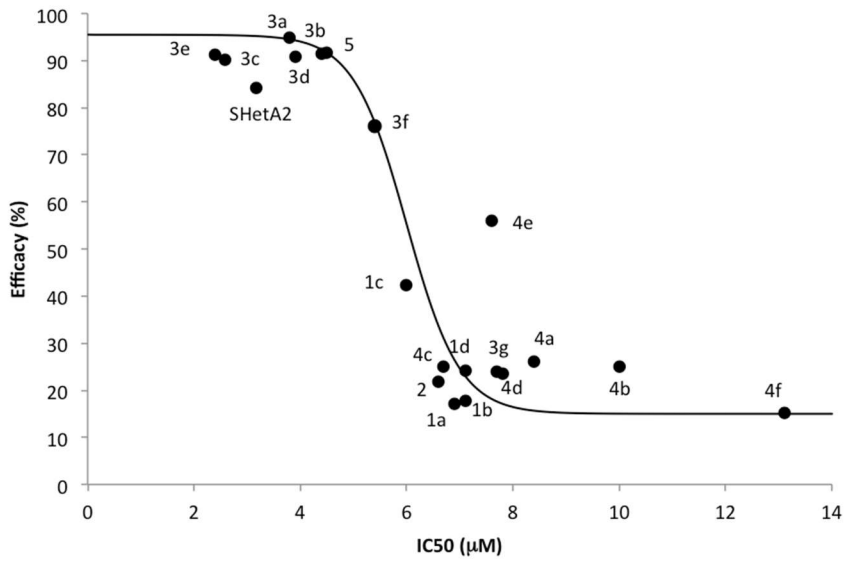


Figure 8.3 Assessment of the efficacy to IC_{50} relation for all compounds.

The data indicate that a compound must have an affinity above a certain threshold for its receptor in order to be effective in initiating the destruction of cancer cells. This threshold is very likely determined by interaction between the receptor and its substrate protein(s), which is competitively displaced by the synthetic ligand. The most desirable properties of a molecule are a combination of the smallest IC_{50} and the highest efficacy. A small IC_{50} is advantageous when translating from single-layered cell cultures to real tissues where a lower concentration of the compound may be available. Compounds **3a-e** showed higher efficacy than SHetA2, with **3a** exhibiting the highest efficacy of 95%. The members of **3** had IC_{50} values of 2.58 μM (**3c**) and 2.4 μM (**3e**), better than that of SHetA2 (3.17 μM).

The interactions of the compounds with the mortalin SBD were studied by virtual docking with Autodock. The binding free energies (ΔG) and dissociation constants (K_d), where a smaller K_d value reflects a stronger binding affinity, are reported in Table 8.1. The Pearson correlation coefficient between the K_d and IC_{50} was 0.32, indicating a weak positive correlation. Partially this weakness in correlation may be due to the relatively low accuracy of ΔG and K_d obtained using virtual docking. Figure 8.3 demonstrated a roughly linear relationship between

IC_{50}, which was obtained from cell-based assays, with the dissociation constant K_d, which was calculated from binding free energy.

IC_{50} depends on the cellular concentration of the receptor (mortalin in this case) and is higher than K_d, obeying $IC_{50} = [R]_0/2 + K_d(1+[S]/K_M)$, where $[R]_0$ is the receptor concentration and [S] the concentration of the substrate protein (for example p53) being competitively displaced by the drug, and K_M the concentration of the substrate protein at which the receptor achieves half activity [180]. The intercept (2.5 ± 1.7 μM) of the fitting line indicated that the mortalin concentration in A2780 cancer cells was roughly 5 μM, and, for this reason, it might be difficult to obtain IC_{50} values much lower than 2.5 μM for these analogs.

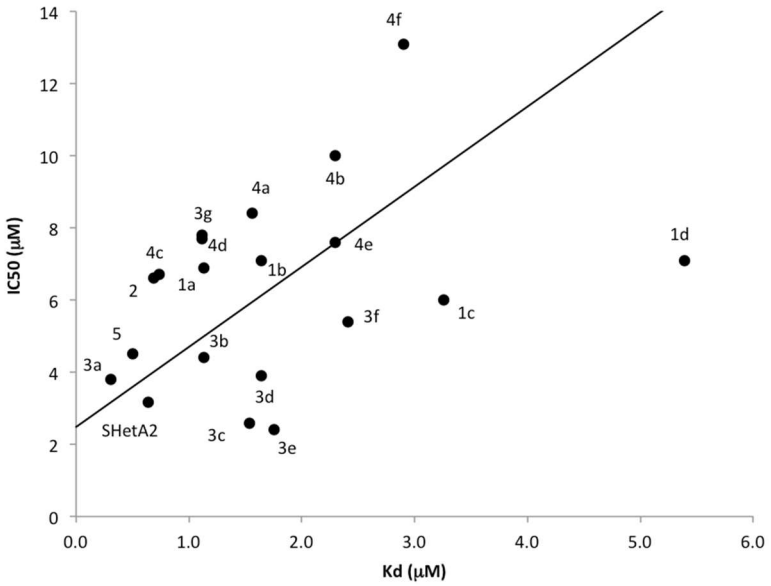


Figure 8.4 IC_{50}- K_d graph.
Assessment of IC_{50} of the compounds and their calculated biochemical dissociation constant (K_d) with the mortalin substrate-binding domain.
The data was linearly fit with a slope (k) of 2.2 ± 0.4 and an intercept (b) of 2.5 ± 1.7 μM

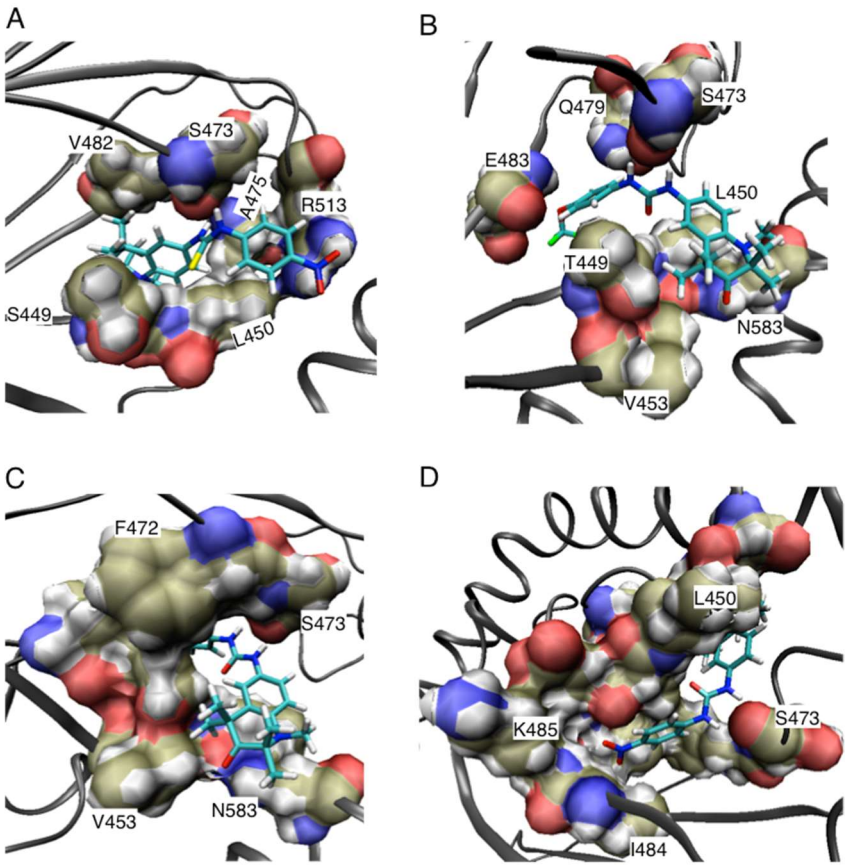


Figure 8.5 Molecular docking of compounds
(A) 5, (B) 3e, and (C,D) 3a to mortalin SBD (PDB ID 3N8E) at the peptide-binding pocket. Amino acids in contact with the compounds were shown in surface drawing method, and the compounds in sticks. Carbon atoms of the amino acids were shown in tan and carbon of compounds in cyan, nitrogen in blue, oxygen in red, sulfur in yellow, and fluorine in green.

Protein-ligand interactions were illustrated in Figure 8.5 for several compounds. The docking pose of compound 5 (Figure 8.5A) was very similar to that of SHetA2 [73], with the negatively charged NO_2 group being attracted to the positively charged sidechain of R513, the backbone C=O of S473 being hydrogen-bonded to the thiourea linker, and ring A with methyl groups residing in the hydrophobic part of the pocket. The docking pose of compound 3e was opposite to those of SHetA2 and 5. This might be a consequence of the reduced hydrophobicity in ring A with the addition of oxygen and nitrogen atoms. In addition to the hydrogen bonds involving the linker, there were two more hydrogen bonds formed between OCF_3 oxygen and E483 amide, and between C=O in ring A and N583 sidechain NH_2 group. The docking pose of 3a (Figure 8.5C and

D) was also opposite to SHetA2. The linker C=O also formed hydrogen bond with L450 amide. The NO_2 group formed hydrogen bond with K485 amide group.

8.4. Conclusion

Several sets of compounds having a Ring A tetrahydroquinoline linked by means of a urea or thiourea to a 4-substituted benzene (Ring B) have been prepared and evaluated relative to SHetA2 in terms of inhibition of A2780 ovarian cancer cells. Six of the derivatives had equal or greater efficacy than SHetA2. For compounds **1a**-**1d** [series 1], the nitrogen analogs of SHetA2 without a methyl on the Ring A nitrogen or the *gem*-dimethyls at C2, showed only weak activity. Compound **2**, the quaternized *N*,*N*-dimethyl iodide salt of **1a**, also lacked significant percent growth inhibition and efficacy in the cell line studied. Structures incorporating a permethylated 3-oxo-1,2,3,4-tetrahydroquinolin-6-yl) linked to a 4-nitrobenzene (**3a** and **3b**) were essentially equivalent in percent of growth inhibition to SHetA2, but with the former having better binding affinity [K_d] values. In addition to NO_2 at C4 of Ring B, CF_3, (**3c** and **3d**) and OCF_3 (**3e** and **3f**) were also evaluated and found to exhibit high growth inhibition and efficacy. Surprisingly, for the derivative having Z = NH_2 (**3g**), the IC_{50} and efficacy were weak. In the carbonyl-containing tetrahydroquinoline systems [series 3], efficacy and growth inhibition values were modestly higher when the linker group between ring A and B was a urea versus a thiourea. In contrast, for compounds having a 1,2,2,4,4-pentamethyl-3-hydroxytetrahydroquinoline unit [series 4], the efficacy and the percent growth inhibition varied only slightly, both being less effective than other analogs. An additional compound **5** was also examined for comparison purposes. In this compound ring B with the NO_2-substitutent showed high potency.

CHAPTER IX

CONCLUSION

Making use of computational tools alongside experimental techniques is a very useful approach in drug design. We studied the structure of mortalin using solution NMR and found high similarity to the crystal structure expect the more flexible unstructured C-terminus. We also studied the interaction between SHeA2 and substrate binding domain of mortalin via solution NMR spectroscopy and molecular mechanic tools, GROMACS and Autodock. The result of these studies led us to design improved analogs. Several of these analogs were synthesized and further computational studies and biological tests have been performed on them. By studying the effect of various structural changes on SHetA2, we showed that some of these compounds outperform SHetA2 in terms of efficacy and potency. We performed molecular mechanic studies for several mutants of mortalin SBD and NMR spectroscopy studies for several other analogs of SHetA2 (not included in this paper). All in an attempt to find the most involved residues of the receptor protein in interaction with ligand and the most effective atoms of the ligand in order to design analogs with high efficacy and potency with enhanced ability to competitively bind to mortalin and abrogate mortalin-p53 interactions. Our findings helps designing more promising SHetA2 analogs.

Future work includes studying the effects of mutants in the binding, biological tests of the suggested mutants, designing and testing other SHetA2 analogs to test the efficacy, potency, cytotoxicity of the compounds to find the most promising agent.

REFERENCES

1. Bloch, F., W.W. Hansen, and M. Packard, *The Nuclear Induction Experiment.* Physical Review, 1946. **70**(7-8): p. 474-485.
2. Purcell, E.M., H.C. Torrey, and R.V. Pound, *Resonance Absorption by Nuclear Magnetic Moments in a Solid.* Physical Review, 1946. **69**(1-2): p. 37-38.
3. Williamson, M.P., T.F. Havel, and K. Wuthrich, *Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry.* J Mol Biol, 1985. **182**(2): p. 295-315.
4. Teng, Q., *Structural biology: practical NMR applications*. 2012: Springer Science & Business Media.
5. Takeuchi, K. and G. Wagner, *NMR studies of protein interactions.* Current Opinion in Structural Biology, 2006. **16**(1): p. 109-117.
6. Furukawa, A., et al., *Quantitative analysis of protein-ligand interactions by NMR.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2016. **96**: p. 47-57.
7. Cala, O., F. Guilliere, and I. Krimm, *NMR-based analysis of protein-ligand interactions.* Analytical and Bioanalytical Chemistry, 2014. **406**(4): p. 943-956.
8. Meyer, B. and T. Peters, *NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors.* Angewandte Chemie-International Edition, 2003. **42**(8): p. 864-890.
9. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy.* Angewandte Chemie-International Edition, 1999. **38**(12): p. 1784-1788.
10. Becker, W., et al., *Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy.* Chemphyschem, 2018. **19**(8): p. 894-894.
11. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions.* 1982. **161**(2): p. 269-288.
12. Huang, S.Y. and X.Q. Zou, *Advances and Challenges in Protein-Ligand Docking.* International Journal of Molecular Sciences, 2010. **11**(8): p. 3016-3034.
13. Sousa, S.F., et al., *Protein-ligand docking in the new millennium--a retrospective of 10 years in the field.* Curr Med Chem, 2013. **20**(18): p. 2296-314.
14. Sousa, S.F., P.A. Fernandes, and M.J. Ramos, *Protein-ligand docking: Current status and future challenges.* Proteins-Structure Function and Bioinformatics, 2006. **65**(1): p. 15-26.
15. Osterberg, F., et al., *Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock.* Proteins, 2002. **46**(1): p. 34-40.
16. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-61.

17. Morris, G.M., et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.
18. Morris, G.M., et al., *Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4.* J Comput Aided Mol Des, 1996. **10**(4): p. 293-304.
19. Pagadala, N.S., K. Syed, and J. Tuszynski, *Software for molecular docking: a review.* Biophys Rev, 2017. **9**(2): p. 91-102.
20. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-18.
21. Kollman, P.A., et al., *Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models.* Accounts of Chemical Research, 2000. **33**(12): p. 889-897.
22. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices.* Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
23. Kumari, R., et al., *g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations.* J Chem Inf Model, 2014. **54**(7): p. 1951-62.
24. Kumari, R., et al., *g_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations.* Journal of Chemical Information and Modeling, 2014. **54**(7): p. 1951-1962.
25. Benbrook, D.M., et al., *Flexible heteroarotinoids (Flex-Hets) exhibit improved therapeutic ratios as anti-cancer agents over retinoic acid receptor agonists.* Invest New Drugs, 2005. **23**(5): p. 417-428.
26. Benbrook, D.M., et al., *SHetA2 interference with mortalin binding to p66shc and p53 identified using drug-conjugated magnetic microspheres.* Invest. New Drugs, 2014. **32**(3): p. 412-423.
27. Iosefson, O. and A. Azem, *Reconstitution of the mitochondrial Hsp70 (mortalin)-p53 interaction using purified proteins – Identification of additional interacting regions.* FEBS Lett., 2010. **584**(6): p. 1080-1084.
28. Blomhoff, R., M.H. Green, and K.R. Norum, *Vitamin A: Physiological and Biochemical Processing.* Annu Rev Nutr, 1992. **12**: p. 37-57.
29. Bushue, N. and Y.J. Wan, *Retinoid pathway and cancer therapeutics.* Adv Drug Deliv Rev, 2010. **62**(13): p. 1285-98.
30. Motzer, R.J., et al., *Interferon alfa-2a and 13-cis-retinoic acid in renal cell carcinoma: antitumor activity in a phase II trial and interactions in vitro.* J Clin Oncol, 1995. **13**(8): p. 1950-7.
31. Huang, M.E., et al., *Use of All-Trans Retinoic Acid in the Treatment of Acute Promyelocytic Leukemia.* Blood, 1988. **72**(2): p. 567-572.
32. Hong, W.K., et al., *13-cis-retinoic acid in the treatment of oral leukoplakia.* 1986. **315**(24): p. 1501-1505.
33. Heyman, R.A., et al., *9-Cis Retinoic Acid Is a High-Affinity Ligand for the Retinoid-X Receptor.* Cell, 1992. **68**(2): p. 397-406.
34. Duong, V. and C. Rochette-Egly, *The molecular physiology of nuclear retinoic acid receptors. From health to disease.* Biochim Biophys Acta, 2011. **1812**(8): p. 1023-1031.
35. Oliveira, M.R., *The neurotoxic effects of vitamin A and retinoids.* An Acad Bras Cienc, 2015. **87**(2 Suppl): p. 1361-73.
36. Garattini, E., M. Gianni, and M. Terao, *Retinoid related molecules an emerging class of apoptotic agents with promising therapeutic potential in oncology: pharmacological activity and mechanisms of action.* Curr Pharm Des, 2004. **10**(4): p. 433-48.

37. Oridate, N., et al., *Rapid induction of apoptosis in human C33A cervical carcinoma cells by the synthetic retinoid 6-[3-(1-adamantyl)hydroxyphenyl]-2-naphthalene carboxylic acid (CD437).* Int J Cancer, 1997. **70**(4): p. 484-487.
38. Schadendorf, D., et al., *Treatment of melanoma cells with the synthetic retinoid CD437 induces apoptosis via activation of AP-1 in vitro, and causes growth inhibition in xenografts in vivo.* J Cell Biol, 1996. **135**(6 Pt 2): p. 1889-98.
39. Hail, N., Jr., H.J. Kim, and R. Lotan, *Mechanisms of fenretinide-induced apoptosis.* Apoptosis, 2006. **11**(10): p. 1677-94.
40. Modiano, M.R., et al., *Phase II study of fenretinide (N-[4-hydroxyphenyl]retinamide) in advanced breast cancer and melanoma.* Invest New Drugs, 1990. **8**(3): p. 317-9.
41. Mariani, L., et al., *Chemoprevention of breast cancer with fenretinide (4-HPR): study of long-term visual and ophthalmologic tolerability.* Tumori Journal, 1996. **82**(5): p. 444-449.
42. Miller, V.A., et al., *Initial clinical trial of a selective retinoid X receptor ligand, LGD1069.* J Clin Oncol, 1997. **15**(2): p. 790-5.
43. Rizvi, N.A., et al., *A Phase I study of LGD1069 in adults with advanced cancer.* Clin Cancer Res, 1999. **5**(7): p. 1658-64.
44. Loeliger, P., W. Bollag, and H. Mayer, *Arotinoids, a new class of highly active retinoids.* Chemischer Informationsdienst, 1980. **11**(24): p. no-no.
45. Lindamood, C., 3rd, et al., *Pharmacological and toxicological properties of arotinoids SMR-2 and SMR-6 in mice.* Fundam Appl Toxicol, 1990. **14**(1): p. 15-29.
46. Benbrook, D.M., et al., *Biologically active heteroarotinoids exhibiting anticancer activity and decreased toxicity.* J Med Chem, 1997. **40**(22): p. 3567-83.
47. Garuti, L., et al., *Diaryl Urea: A Privileged Structure in Anticancer Agents.* Curr Med Chem, 2016. **23**(15): p. 1528-48.
48. Zacheis, D., et al., *Heteroarotinoids inhibit head and neck cancer cell lines in vitro and in vivo through both RAR and RXR retinoic acid receptors.* Journal of Medicinal Chemistry, 1999. **42**(21): p. 4434-4445.
49. Dhar, A., et al., *Synthesis, Structure− Activity Relationships, and RARγ− Ligand Interactions of Nitrogen Heteroarotinoids.* J Med Chem, 1999. **42**(18): p. 3602-3614.
50. Spruce, L.W., et al., *Novel Heteroarotinoids - Synthesis and Biological-Activity.* J Med Chem, 1991. **34**(1): p. 430-439.
51. Benbrook, D.M., *Refining retinoids with heteroatoms.* Mini Rev Med Chem, 2002. **2**(3): p. 277-83.
52. Lehmann, J.M., et al., *Identification of retinoids with nuclear receptor subtype-selective activities.* Cancer Res, 1991. **51**(18): p. 4804-9.
53. Mangelsdorf, D., K. Umesano, and R. Evans, *The retinoid receptors. In "The Retinoids: Biology, Chemistry, and Medicine"(M. Sporn, A. Roberts, and D. Goodman, D., Eds.)*. 1994, Raven Press, New York.
54. Zacheis, D., et al., *Heteroarotinoids inhibit head and neck cancer cell lines in vitro and in vivo through both RAR and RXR retinoic acid receptors.* J Med Chem, 1999. **42**(21): p. 4434-45.
55. Benbrook, D.M., et al., *Synthesis and characterization of heteroarotinoids demonstrate structure specificity relationships.* J Med Chem, 1998. **41**(19): p. 3753-7.
56. Brown, C.W., et al., *Novel heteroarotinoids as potential antagonists of Mycobacterium bovis BCG.* J Med Chem, 2004. **47**(4): p. 1008-1017.
57. Guruswamy, S., et al., *Effects of retinoids on cancerous phenotype and apoptosis in organotypic cultures of ovarian carcinoma.* J Natl Cancer Inst, 2001. **93**(7): p. 516-525.

58. Liu, S.Q., et al., *Synthesis of flexible sulfur-containing heteroarotinoids that induce apoptosis and reactive oxygen species with discrimination between malignant and benign cells.* J Med Chem, 2004. **47**(4): p. 999-1007.
59. Chun, K., et al., *Induction of apoptosis in head and neck squamous cell carcinoma (HNSCC) cell lines by heteroarotinoids through a mitochondrial dependent pathway.* 2003. **63**: p. 3826-3832.
60. Liu, T., et al., *Flex-Hets differentially induce apoptosis in cancer over normal cells by directly targeting mitochondria.* Mol Cancer Ther, 2007. **6**(6): p. 1814-22.
61. Lin, Y., et al., *Involvement of c-FLIP and survivin down-regulation in flexible heteroarotinoid-induced apoptosis and enhancement of TRAIL-initiated apoptosis in lung cancer cells.* Mol Cancer Ther, 2008. **7**(11): p. 3556-65.
62. Liu, T.Z., et al., *Development of flexible-heteroarotinoids for kidney cancer.* Mol Cancer Ther, 2009. **8**(5): p. 1227-1238.
63. Chun, K.H., et al., *The synthetic heteroarotinoid SHetA2 induces apoptosis in squamous carcinoma cells through a receptor-independent and mitochondria-dependent pathway.* Cancer Research, 2003. **63**(13): p. 3826-3832.
64. Le, T.C., et al., *Heteroarotinoids with anti-cancer activity against ovarian cancer cells.* Open Med Chem J, 2007. **1**: p. 11-23.
65. Gnanasekaran, K.K., et al., *Synthesis and evaluation of second generation Flex-Het scaffolds against the human ovarian cancer A2780 cell line.* Eur J Med Chem, 2015. **96**: p. 209-217.
66. Myers, T., et al., *Flexible heteroarotinoid (Flex-Het) SHetA2 inhibits angiogenesis in vitro and in vivo.* Investigational New Drugs, 2009. **27**(4): p. 304-318.
67. Kabirov, K.K., et al., *Oral toxicity and pharmacokinetic studies of SHetA2, a new chemopreventive agent, in rats and dogs.* Drug Chem Toxicol, 2013. **36**(3): p. 284-95.
68. Doppalapudi, R.S., et al., *Genotoxicity of the cancer chemopreventive drug candidates CP-31398, SHetA2, and phospho-ibuprofen.* Mutation Research-Genetic Toxicology and Environmental Mutagenesis, 2012. **746**(1): p. 78-88.
69. Lin, Y., S. Chen, and P.J.C.R. Yue, *CHOP-dependent death receptor 5 induction is a major component of SHetA2-induced apoptosis in lung cancer cells.* 2008. **68**: p. 5335-44.
70. Lin, Y.D., et al., *CAAT/enhancer binding protein homologous protein-dependent death receptor 5 induction is a major component of SHetA2-induced apoptosis in lung cancer cells.* Cancer Research, 2008. **68**(13): p. 5335-5344.
71. Benbrook, D.M., et al., *Chemoprevention of Colon and Small Intestinal Tumorigenesis in APCmin/+ Mice By SHetA2 (NSC721689) without Toxicity.* Cancer Prevention Research, 2013. **6**(9): p. 908-916.
72. Nammalwar, B., et al., *Synthesis of N-[3,4-Dihydro-4-(acetoxymethyl)-2,2,4-trimethyl-2H-1-benzothiopyran-6-yl]-N'-(4-nitrophenyl)thiourea and N-[3,4-dihydro-4-(hydroxymethyl)-2,2,4-trimethyl-2H-1-benzothiopyran-6-yl]-N'-(4-nitrophenyl)thiourea, a Major Metabolite of N-(3,4-Dihydro-2,2,4,4-tetramethyl-2H-1-benzothiopyran-6-YL)-N'-(4-nitrophenyl)thiourea.* Phosphorus Sulfur and Silicon and the Related Elements, 2011. **186**(1): p. 189-204.
73. Watts, F.M., et al., *Activity of oxygen-versus sulfur-containing analogs of the Flex-Het anticancer agent SHetA2.* Eur J Med Chem, 2018. **158**: p. 720-732.
74. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-91.

75. Lindamood, C., 3rd, et al., *Toxicologic and immunologic evaluations of N-(all-trans-retinoyl)-DL-leucine and N-(all-trans-retinoyl)glycine.* Toxicol Appl Pharmacol, 1988. **96**(2): p. 279-95.
76. Fulda, S. and K.M. Debatin, *Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy.* Oncogene, 2006. **25**(34): p. 4798-811.
77. Hengartner, M.O., *The biochemistry of apoptosis.* Nature, 2000. **407**(6805): p. 770-776.
78. Chengedza, S. and D.M. Benbrook, *NF-kappa B is involved in SHetA2 circumvention of TNF-alpha resistance, but not induction of intrinsic apoptosis.* Anti-Cancer Drugs, 2010. **21**(3): p. 297-305.
79. Kelly, W.J., et al., *Sheta2 induces mitochondrial swelling, superoxide formation and apoptosis in human cancer cells.* Abstracts of Papers of the American Chemical Society, 2005. **229**: p. U492-U492.
80. Benbrook, D.M., et al., *SHetA2 interference with mortalin binding to p66shc and p53 identified using drug-conjugated magnetic microspheres.* Investigational New Drugs, 2014. **32**(3): p. 412-423.
81. Benbrook, D.M., et al., *Development of a dietary formulation of the SHetA2 chemoprevention drug for mice.* Investigational New Drugs, 2018. **36**(4): p. 561-570.
82. Masamha, C.P. and D.M. Benbrook, *Cyclin D1 degradation is sufficient to induce G1 cell cycle arrest despite constitutive expression of cyclin E2 in ovarian cancer cells.* Cancer Res, 2009. **69**(16): p. 6565-6572.
83. Lans, T.E., et al., *Role of tumor necrosis factor on toxicity and cytokine production after isolated hepatic perfusion.* 2001. **7**(4): p. 784-790.
84. Wu, H., J. Tschopp, and S.-C.J.C. Lin, *Smac mimetics and TNFα: a dangerous liaison?* 2007. **131**(4): p. 655-658.
85. Yoon, H.S., H.A. Kim, and Y.W. Song, *Inhibition of NF-kappaB renders human juvenile costal chondrocyte cell lines sensitive to TNF-alpha-mediated cell death.* Rheumatol Int, 2006. **26**(3): p. 201-8.
86. Moxley, K.M., S. Chengedza, and D. Mangiaracina, *Induction of death receptor ligand-mediated apoptosis in epithelial ovarian carcinoma: The search for sensitizing agents.* Gynecologic Oncology, 2009. **115**(3): p. 438-442.
87. Chengedza, S. and D.M.J.A.-c.d. Benbrook, *NF-kappaB is involved in SHetA2 circumvention of TNF-alpha resistance, but not induction of intrinsic apoptosis.* 2010. **21**(3): p. 297.
88. Ran, Q.T., et al., *Extramitochondrial localization of mortalin/mthsp70/PBP74/GRP75.* Biochem Biophys Res Commun, 2000. **275**(1): p. 174-179.
89. Ritossa, F., *Discovery of the heat shock response.* Cell Stress & Chaperones, 1996. **1**(2): p. 97-98.
90. Shu, C.W. and C.M. Huang, *HSP70s: From Tumor Transformation to Cancer Therapy.* Clin Med Oncol, 2008. **2**: p. 335-45.
91. Wadhwa, R., et al., *Protein Markers for Cellular Mortality and Immortality.* Mutation Research, 1991. **256**(2-6): p. 243-254.
92. Wadhwa, R., et al., *Identification of a novel member of mouse hsp70 family. Its association with cellular mortal phenotype.* J Biol Chem, 1993. **268**(9): p. 6615-21.
93. Wadhwa, R., et al., *Differential Subcellular-Distribution of Mortalin in Mortal and Immortal Mouse and Human Fibroblasts.* Exp Cell Res, 1993. **207**(2): p. 442-448.
94. Kaul, S.C., et al., *Structurally and functionally distinct mouse hsp70 family members Mot-1 and Mot-2 proteins are encoded by two alleles.* DNA Research, 2000. **7**(3): p. 229-231.

95. Deocaris, C.C., et al., *Functional significance of minor structural and expression changes in stress chaperone mortalin.* Molecular Mechanisms and Models of Aging, 2007. **1119**: p. 165-175.
96. Wadhwa, R., et al., *Induction of Cellular Senescence by Transfection of Cytosolic Mortalin Cdna in Nih-3t3-Cells.* J Biol Chem, 1993. **268**(30): p. 22239-22242.
97. Kaul, S.C., et al., *Malignant transformation of NIH3T3 cells by overexpression of mot-2 protein.* Oncogene, 1998. **17**(7): p. 907-11.
98. Kaul, S.C., et al., *Transcriptional inactivation of p53 by deletions and single amino acid changes in mouse mot-1 protein.* Biochem Biophys Res Commun, 2000. **279**(2): p. 602-606.
99. Wadhwa, R., et al., *Inactivation of tumor suppressor p53 by Mot-2, a hsp70 family member.* J Biol Chem, 1998. **273**(45): p. 29586-29591.
100. Xie, H., et al., *Human mortalin (HSPA9): a candidate for the myeloid leukemia tumor suppressor gene on 5q31.* Leukemia, 2000. **14**(12): p. 2128-2134.
101. Kaul, S.C., C.C. Deocaris, and R. Wadhwa, *Three faces of mortalin: a housekeeper, guardian and killer.* Exp Gerontol, 2007. **42**(4): p. 263-74.
102. Flachbartova, Z. and B. Kovacech, *Mortalin - a multipotent chaperone regulating cellular processes ranging from viral infection to neurodegeneration.* Acta Virol, 2013. **57**(1): p. 3-15.
103. Londono, C., et al., *Mortalin, apoptosis, and neurodegeneration.* Biomolecules, 2012. **2**(1): p. 143-64.
104. Wadhwa, R., et al., *Upregulation of mortalin/mthsp70/Grp75 contributes to human carcinogenesis.* Int J Cancer, 2006. **118**(12): p. 2973-2980.
105. Starenki, D., et al., *Mortalin (GRP75/HSPA9) upregulation promotes survival and proliferation of medullary thyroid carcinoma cells.* Oncogene, 2015. **34**(35): p. 4624-4634.
106. Yang, J. and Y. Zhang, *Protein Structure and Function Prediction Using I-TASSER.* Curr Protoc Bioinformatics, 2015. **52**: p. 5 8 1-15.
107. Jackson, S.E., *Allostery in the Hsp70 Chaperone Proteins.* Molecular Chaperones, 2013. **328**: p. 99-153.
108. Liu, Q.L. and W.A. Hendrickson, *Insights into Hsp70 chaperone activity from a crystal structure of the yeast Hsp110 Sse1.* Cell, 2007. **131**(1): p. 106-120.
109. Swain, J.F., et al., *Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker.* Molecular Cell, 2007. **26**(1): p. 27-39.
110. Bhattacharya, A., et al., *Allostery in Hsp70 chaperones is transduced by subdomain rotations.* 2009. **388**(3): p. 475-490.
111. Qi, R.F., et al., *Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP.* Nature Structural & Molecular Biology, 2013. **20**(7): p. 900-907.
112. Swain, J.F., et al., *Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker.* Mol Cell, 2007. **26**(1): p. 27-39.
113. Bai, L. and W.-G. Zhu, *p53: structure, function and therapeutic applications.* J Cancer Mol, 2006. **2**(4): p. 141-153.
114. Gabizon, R., et al., *Specific Recognition of p53 Tetramers by Peptides Derived from p53 Interacting Proteins.* PLoS One, 2012. **7**(5).
115. Kitayner, M., et al., *Structural basis of DNA recognition by p53 tetramers.* Mol Cell, 2006. **22**(6): p. 741-53.
116. Ryan, K.M., A.C. Phillips, and K.H. Vousden, *Regulation and function of the p53 tumor suppressor protein.* Curr Opin Cell Biol, 2001. **13**(3): p. 332-337.

117. Mihara, M., et al., *p53 has a direct apoptogenic role at the mitochondria.* Mol Cell, 2003. **11**(3): p. 577-590.
118. Bieging, K.T., S.S. Mello, and L.D. Attardi, *Unravelling mechanisms of p53-mediated tumour suppression.* Nat Rev Cancer, 2014. **14**(5): p. 359-70.
119. Joerger, A.C. and A.R. Fersht, *The p53 Pathway: Origins, Inactivation in Cancer, and Emerging Therapeutic Approaches.* Annu Rev Biochem, 2016. **85**: p. 375-404.
120. Vousden, K.H. and G.F. Vande Woude, *The ins and outs of p53.* Nat Cell Biol, 2000. **2**(10): p. E178-E180.
121. Kaul, S.C., et al., *Activation of wild type p53 function by its mortalin-binding, cytoplasmically localizing carboxyl terminus peptides.* J Biol Chem, 2005. **280**(47): p. 39373-9.
122. Nikolaev, A.Y., et al., *Parc: A cytoplasmic anchor for p53.* Cell, 2003. **112**(1): p. 29-40.
123. Wadhwa, R., et al., *NIH 3T3 cells malignantly transformed by mot-2 show inactivation and cytoplasmic sequestration of the p53 protein.* Cell Res, 1999. **9**(4): p. 261-269.
124. Gestl, E.E. and S.A. Bottger, *Cytoplasmic sequestration of the tumor suppressor p53 by a heat shock protein 70 family member, mortalin, in human colorectal adenocarcinoma cell lines.* Biochem Biophys Res Commun, 2012. **423**(2): p. 411-416.
125. Wadhwa, R., et al., *Hsp70 family member, mot-2/mthsp70/GRP75, binds to the cytoplasmic sequestration domain of the p53 protein.* Exp Cell Res, 2002. **274**(2): p. 246-253.
126. Walker, C., S. Bottger, and B. Low, *Mortalin-based cytoplasmic sequestration of p53 in a nonmammalian cancer model.* Am J Pathology, 2006. **168**(5): p. 1526-1530.
127. Lu, W., et al., *Mortalin–p53 interaction in cancer cells is stress dependent and constitutes a selective target for cancer therapy.* Cell Death Differ, 2011. **18**(6): p. 1046.
128. Kaul, S.C., et al., *Inactivation of p53 and life span extension of human diploid fibroblasts by mot-2.* Febs Letters, 2000. **474**(2-3): p. 159-164.
129. Kaul, S.C., et al., *Overexpressed mortalin (mot-2)/mthsp70/GRP75 and hTERT cooperate to extend the in vitro lifespan of human fibroblasts.* Exp Cell Res, 2003. **286**(1): p. 96-101.
130. Xu, J., H.H. Xiao, and A.C. Sartorelli, *Attenuation of the induced differentiation of HL-60 leukemia cells by mitochondrial chaperone HSP70.* Oncol Res, 1999. **11**(9): p. 429-35.
131. Marchenko, N.D., A. Zaika, and U.M. Moll, *Death signal-induced localization of p53 protein to mitochondria - A potential role in apoptotic signaling.* J Biol Chem, 2000. **275**(21): p. 16202-16212.
132. Iosefson, O. and A. Azem, *Reconstitution of the mitochondrial Hsp70 (mortalin)-p53 interaction using purified proteins - Identification of additional interacting regions.* Febs Letters, 2010. **584**(6): p. 1080-1084.
133. Deocaris, C.C., et al., *Glycerol stimulates innate chaperoning, proteasomal and stress-resistance functions: implications for geronto-manipulation.* Biogerontology, 2008. **9**(4): p. 269-282.
134. Ma, Z., et al., *Mortalin controls centrosome duplication via modulating centrosomal localization of p53.* Oncogene, 2006. **25**(39): p. 5377-5390.
135. Ostermeyer, A.G., et al., *Cytoplasmically sequestered wild-type p53 protein in neuroblastoma is relocated to the nucleus by a C-terminal peptide.* Proc Natl Acad Sci USA, 1996. **93**(26): p. 15190-15194.
136. Stevens, S.Y., et al., *The solution structure of the bacterial HSP70 chaperone protein domain DnaK(393–507) in complex with the peptide NRLLLTG.* Protein Sci., 2003. **12**(11): p. 2588-2596.

137. Kaul, S.C., et al., *An N-terminal region of mot-2 binds to p53 in vitro.* Neoplasia, 2001. **3**(2): p. 110-114.
138. Yoo, J.Y., et al., *Tumor suppression by apoptotic and anti-angiogenic effects of mortalin-targeting adeno-oncolytic virus.* J Gene Med, 2010. **12**(7): p. 586-595.
139. Grover, A., et al., *Withanone binds to mortalin and abrogates mortalin-p53 complex: computational and experimental evidence.* Int J Biochem Cell Biol, 2012. **44**(3): p. 496-504.
140. Rousaki, A., et al., *Allosteric Drugs: The Interaction of Antitumor Compound MKT-077 with Human Hsp70 Chaperones.* J Mol Biol, 2011. **411**(3): p. 614-632.
141. Wadhwa, R., et al., *Selective toxicity of MKT-077 to cancer cells is mediated by its binding to the hsp70 family protein mot-2 and reactivation of p53 function.* Cancer Res, 2000. **60**(24): p. 6818-6821.
142. Deocaris, C.C., et al., *Mortalin sensitizes human cancer cells to MKT-077-induced senescence.* Cancer Lett, 2007. **252**(2): p. 259-269.
143. Vaishnavi, K., et al., *Differential Activities of the Two Closely Related Withanolides, Withaferin A and Withanone: Bioinformatics and Experimental Evidences.* PLoS One, 2012. **7**(9).
144. Grover, A., *Withanolides: Strategies for Enhanced Production and Mechanistic Insights into their Mode of Action.* Proc Indian Natn Sci Acad, 2015. **81**: p. 599-607.
145. Propper, D.J., et al., *Phase I trial of the selective mitochondrial toxin MKT077 in chemo-resistant solid tumours.* Ann Oncol, 1999. **10**(8): p. 923-7.
146. Britten, C.D., et al., *A Phase I and pharmacokinetic study of the mitochondrial-specific rhodacyanine dye analog MKT 077.* Clin Cancer Res, 2000. **6**(1): p. 42-49.
147. Amick, J., et al., *Crystal structure of the nucleotide-binding domain of mortalin, the mitochondrial Hsp70 chaperone.* Protein Sci, 2014. **23**(6): p. 833-42.
148. Sharma, A., D.M. Benbrook, and S. Woo, *Pharmacokinetics and interspecies scaling of a novel, orally-bioavailable anti-cancer drug, SHetA2.* PLoS One, 2018. **13**(4): p. e0194046.
149. Masamhal, C.P., T. Liu, and D.M. Benbrook, *A116 SHetA2 targets cyclin D1 for proteasomal degradation through a GSK3 beta-independent mechanism leading to G1 cell cycle arrest.* Molecular Cancer Therapeutics, 2007. **6**(12): p. 3372s-3372s.
150. Sharma, A., et al., *Bioanalytical method development and validation of HPLCUV assay for the quantification of SHetA2 in mouse and human plasma: Application to pharmacokinetics study.* J Pharm Technol Drug Res, 2017. **6**.
151. Zhang, Y.L., et al., *High performance liquid chromatographic analysis and preclinical pharmacokinetics of the heteroarotinoid antitumor agent, SHetA2.* Cancer Chemotherapy and Pharmacology, 2006. **58**(5): p. 561-569.
152. Liu, Z.F., et al., *Metabolism of a sulfur-containing heteroarotionoid antitumor agent, SHetA2, using liquid chromatography/tandem mass spectrometry.* Rapid Communications in Mass Spectrometry, 2008. **22**(21): p. 3371-3381.
153. Sharma, A., D.M. Benbrook, and S. Woo, *Pharmacokinetics and interspecies scaling of a novel, orally-bioavailable anti-cancer drug, SHetA2.* PLoS One, 2018. **13**(4).
154. Mahjabeen, S., et al., *Influence of the estrus cycle of the mouse on the disposition of SHetA2 after vaginal administration.* European Journal of Pharmaceutics and Biopharmaceutics, 2018. **130**: p. 272-280.
155. Lin, P., *Applicatioin of nuclear magnetic resonance to protein structure and protein-ligand interation studies*. 2015, Oklahoma State University.
156. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* 1995. **6**(3): p. 277-293.

157. Lee, W., M. Tonelli, and J.L. Markley, *NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy.* Bioinformatics, 2015. **31**(8): p. 1325-1327.
158. Bahrami, A., et al., *Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy.* PLoS Comput Biol, 2009. **5**(3): p. e1000307.
159. Shen, Y., et al., *TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* Journal of Biomolecular Nmr, 2009. **44**(4): p. 213-223.
160. Eghbalnia, H.R., et al., *Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements.* Journal of Biomolecular Nmr, 2005. **32**(1): p. 71-81.
161. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38.
162. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox.* Journal of Cheminformatics, 2011. **3**.
163. Chimera, U.J.J.C.C., *a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE.* 2004. **25**(13): p. 1605-12.
164. Jorgensen, W.L. and J.D. Madura, *Quantum and Statistical Studies of Liquids .25. Solvation and Conformation of Methanol in Water.* Journal of the American Chemical Society, 1983. **105**(6): p. 1407-1413.
165. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* Journal of Chemical Physics, 2007. **126**(1).
166. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method.* Journal of Applied Physics, 1981. **52**(12): p. 7182-7190.
167. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions.* Physical Review A, 1985. **31**(3): p. 1695-1697.
168. Nose, S., *A Molecular-Dynamics Method for Simulations in the Canonical Ensemble.* Molecular Physics, 1984. **52**(2): p. 255-268.
169. Nose, S. and M.L. Klein, *Constant Pressure Molecular-Dynamics for Molecular-Systems.* Molecular Physics, 1983. **50**(5): p. 1055-1076.
170. Hockney, R.W., S.P. Goel, and J.W. Eastwood, *Quiet High-Resolution Computer Models of a Plasma.* Journal of Computational Physics, 1974. **14**(2): p. 148-158.
171. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems.* Journal of Chemical Physics, 1993. **98**(12): p. 10089-10092.
172. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* Journal of Computational Chemistry, 1997. **18**(12): p. 1463-1472.
173. Iosefson, O. and A.J.F.l. Azem, *Reconstitution of the mitochondrial Hsp70 (mortalin)-p53 interaction using purified proteins–identification of additional interacting regions.* 2010. **584**(6): p. 1080-1084.
174. Le, T.C., et al., *Heteroarotinoids with Anti-Cancer Activity Against Ovarian Cancer Cells.* The Open Medicinal Chemistry Journal, 2007. **1**: p. 11-23.
175. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* 2009. **30**(16): p. 2785-2791.
176. Zhou, P., et al., *Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins.* Proteins: Struct., Funct., Bioinf., 2009. **76**: p. 151-163.

177. Zeng, D., Q. Shen, and J.-H. Cho, *Thermodynamic contribution of backbone conformational entropy in the binding between SH3 domain and proline-rich motif.* Biochem. Biophys. Res. Commun., 2017. **484**(1): p. 21-26.
178. Shen, Q., et al., *Molecular Mechanisms of Tight Binding through Fuzzy Interactions.* Biophys. J., 2018. **114**(6): p. 1313-1320.
179. Vitaku, E., D.T. Smith, and J.T. Njardarson, *Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals.* J Med Chem, 2014. **57**(24): p. 10257-74.
180. Cha, S., R.P. Agarwal, and R.E. Parks, Jr., *Tight-binding inhibitors-II. Non-steady state nature of inhibition of milk xanthine oxidase by allopurinol and alloxanthine and of human erythrocytic adenosine deaminase by coformycin.* Biochem Pharmacol, 1975. **24**(23): p. 2187-97.

VITA

Maryam Mashayekhi

Candidate for the Degree of

Doctor of Philosophy

Dissertation: STUDIES OF HEAT SHOCK PROTEIN A9 (HSPA9/MORTALIN/GRP75) SUBSTRATE BINDING DOMAIN AND ANTICANCER FLEX-HETS INTERACTIONS USING NUCLEAR MAGNETIC RESONANCE AND MOLECULAR MECHANICS TOOLS

Major Field: Photonics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Photonics at Oklahoma State University, Stillwater, Oklahoma in May, 2019.

Completed the requirements for the Master of Science in Photonics at Iran University of Science and Technology, Tehran, Iran in 2011.

Completed the requirements for the Bachelor of Science Physics at University of Arak, Arak, Iran in 2008

Experience:

Aug. 2013 – present: Oklahoma State University, Stillwater, OK
Teaching Assistant

January. 2014 – present: Oklahoma State University, Stillwater, OK
Research Assistant