

APPLICATION OF NUCLEAR MAGNETIC
RESONANCE TO PROTEIN STRUCTURE AND
PROTEIN-LIGAND INTERACTION STUDIES

By

PENGHUI LIN

Bachelor of Science in Bioengineering
Harbin Institute of Technology
Harbin, China
2006

Master of Science in Pathogen Biology
Fudan University
Shanghai, China
2009

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
July, 2015

APPLICATION OF NUCLEAR MAGNETIC
RESONANCE TO PROTEIN STRUCTURE AND
PROTEIN-LIGAND INTERACTION STUDIES

Dissertation Approved:

Dr. Donghua H Zhou

Dissertation Adviser

Dr. Aihua Xie

Dr. YingMei Liu

Dr. Junpeng Deng

ACKNOWLEDGEMENTS

First of all I would like to express my sincere gratitude to my advisor Dr. Donghua Zhou for being supportive and kind through my entire PhD study. His hard-working spirit, creative thinking and enthusiasm in science deeply influenced me. I also thank all my committee members, Drs. Aihua Xie, Yingmei Liu and Junpeng Deng for their efforts in guiding me through each step in the PhD period. My collaborators Drs. José Luis Soulages, Estela Arrese, K. Darrel Berlin, Doris Benbrook and Steven Hartson all helped me in various ways including designing, interpreting and troubleshooting the experiments, they are appreciated. Special thanks to our NMR facility manager Dr. Margret Eastman for her kind help during the NMR experiments. Also I would like to thank all my lab members: Dr. Liying Wang, Future Drs. Lian Duan, Amanda Taylor, Maryam Mashayekhi and Hem Moktan. The cooperation, friendship and fun we had together will be my treasure of a lifetime.

I am grateful to my parents for being understanding and supportive all the time. At last, I would like to express my deepest appreciation to my wife, Huan Song, for her love, dedication and support as well as my two beautiful angels, Pinkie and Rainie, who always bring sunshine into my daily life.

Name: PENGHUI LIN

Date of Degree: JULY, 2015

Title of Study: APPLICATION OF NUCLEAR MAGNETIC RESONANCE TO
PROTEIN STRUCTURE AND PROTEIN-LIGAND INTERACTION
STUDIES

Major Field: PHOTONICS

Abstract: NMR (Nuclear Magnetic Resonance) has been expanding its application since it was first discovered in the early 50s. In biophysics, it is a very powerful tool complement to X-ray crystallography for protein structure studies. In this dissertation, two different projects studied by two different NMR methods will be presented. In the first part, a membrane protein PLIN1 which resides on the surface of the cellular organelle lipid droplet is investigated by solid-state NMR (ssNMR). It has been shown that PLIN 1 exclusively locates on the surface of lipid droplet and upon phosphorylation, recruits lipase to digest the triglycerides stored in the lipid droplet. However, due to its membrane protein nature, its insolubility resists to crystallization. ssNMR is a perfect tool to study this membrane protein. PLIN1 was reconstituted into DMPG liposomes and multi-dimensional ssNMR spectra were acquired. According to proton spin diffusion data, a membrane interaction model was proposed and later verified by both molecular dynamics (MD) simulation and experimental reconstitution data. In the second part, a soluble protein Mortalin was studied by solution NMR. Mortalin is a heat shock protein 70 (HSP70) family membrane primarily located in the mitochondria. In cancer cells, Mortalin is released to the cytoplasm and forms a complex with p53, sequestering it in the cytoplasm thus inhibiting its translocation to nucleus to induce cell apoptosis. The flexible heteroarotinoid (Flex-Het) SHetA2 with promising anti-cancer activity can bind to Mortalin and release p53 from the complex to induce cell apoptosis. We successfully determined the substrate binding pocket of Mortalin to be the interacting sites with SHetA2 by chemical shift perturbation (CSP). Using AutoDock as the prediction tool, at least two binding configurations of SHetA2 are generated with high binding affinity. According to these results, more SHetA2 analogs were designed and tested in AutoDock. We find that the analogs with longer linkers that can occupy both configurations of SHetA2 yield the highest binding affinity (lowest binding energy). These results will guide future drug designs to increase the efficiency of Flex-Het anti-cancer activity.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
II. HISTORY OF NMR	15
III. THE BASICS OF NMR	22
Section 1 Introduction.....	22
Section 2 The Bloch Equation	28
Section 3 Relaxation	34
Section 4 Chemical Shift	40
Section 5 Limitation of Classical Model	44
Section 6 Density Matrix Formalism.....	47
IV. NMR INSTRUMENTATION	54
Section 1 Magnet	56
Section 2 Transmitter.....	57
Section 3 Duplexer.....	57
Section 4 Probe	58
Section 5 Receiver	60
Section 6 Data Processing.....	63
V. THE LIPID DROPLET PROTEIN PLIN1	65
Section 1 Abstract.....	65
Section 2 Introduction.....	66
Section 3 Materials and Methods.....	69
3.1 Protein expression and purification	69
3.2 Protein reconstitution	70
3.3 NMR spectroscopy.....	72
3.4 MD simulation and spectral simulation	73
3.5 Trypsin protection assay and mass spectrometry.....	74
3.6 Truncated PLIN1 mutant clone and reconstitution.....	75

Section 4 Results and Discussion	76
4.1 Spectra of ¹³ C, ¹⁵ N-PLIN1	76
4.2 Membrane insertion by proton spin diffusion.....	78
4.3 Model building.....	83
4.4 Model verification by NMR data.....	89
4.5 Model verification by experimental data	90
Section 5 Conclusion	93
 VI. INTERACTION BETWEEN MORTALNI AND SHETA2.....	 94
Section 1 Abstract.....	94
Section 2 Introduction.....	95
Section 3 Materials and Methods.....	95
3.1 Protein expression and purification	101
3.2 NMR spectroscopy.....	102
3.3 SHetA2 titration	102
3.4 Protein assignments	103
3.5 AutoDock docking simulation	103
Section 4 Results and Discussion	104
4.1 Recombinant Mortalin SBD expression and sample preparation	104
4.2 Solution NMR of Mortalin SBD and backbone assignments	105
4.3 SHetA2 titration and chemical shift perturbation	110
4.4 Multi-configurations of interactions between SHetA2 and Mortalin by AutoDock.....	112
4.5 Improvements on small molecule structure to enhance binding.....	115
Section 5 Conclusion	125
 REFERENCES	 127
 APPENDICES	 137

LIST OF TABLES

Table	Page
Table 1	15
Table 2	16
Table 3	92
Table 4	119
Table 5	121
Table 6	129

LIST OF FIGURES

Figure	Page
Figure 1	2
Figure 2	18
Figure 3	20
Figure 4	22
Figure 5	23
Figure 6	24
Figure 7	26
Figure 8	27
Figure 9	32
Figure 10	34
Figure 11	38
Figure 12	42
Figure 13	53
Figure 14	56
Figure 15	59
Figure 16	71
Figure 17	76
Figure 18	78
Figure 19	80
Figure 20	81
Figure 21	82
Figure 22	83
Figure 23	84
Figure 24	87
Figure 25	88
Figure 26	90
Figure 27	90
Figure 28	91
Figure 29	93
Figure 30	109
Figure 31	110

Figure 32	112
Figure 33	113
Figure 34	114
Figure 35	115
Figure 36	116
Figure 37	117
Figure 38	119
Figure 39	120
Figure 40	134
Figure 41	134
Figure 42	134

CHAPTER I

INTRODUCTION

Nuclear Magnetic Resonance (NMR) spectroscopy has been continuously expanding its application areas since it was discovered after World War II when Bloch [1] at Stanford and Purcell [2] at Harvard reported NMR in bulk water and paraffin, respectively. The growth accelerated even more after the developments of pulsed Fourier transform NMR spectroscopy and the conception of multidimensional NMR spectroscopy. Despite its application in physics first, it is widely used in chemistry, biology, pharmaceutical and clinical diagnosis (MRI) nowadays. NMR theory is based on the spin angular momentum, an intrinsically quantum mechanical property that does not have a classical analog. However, precession can be used to help visualize the behavior of the nucleus under external magnetic fields. The precession of the nuclear spin results an oscillating current that can be readily detected as free induction decay (FID). NMR has been a relatively insensitive method due to the fact that it detects the population difference between two states of the interested nuclei. In protein structure determination, X-crystallography is the dominant method. When the first NMR structure of a small protein BPTI was published, the reaction from the well-established community from X-ray crystallography was disbelief. They argued that the structure had been modeled using other proteins whose structures were previously determined by X-ray crystallography. But, with years of developments, such as higher magnetic field, cryoprobes, stability enhancements, and more complicated processing methods, nowadays,

NMR is a major structure determination technique complementary to X-ray crystallography. The number of structure determined by NMR has been increasing steadily according to the Protein Data Bank statistics.

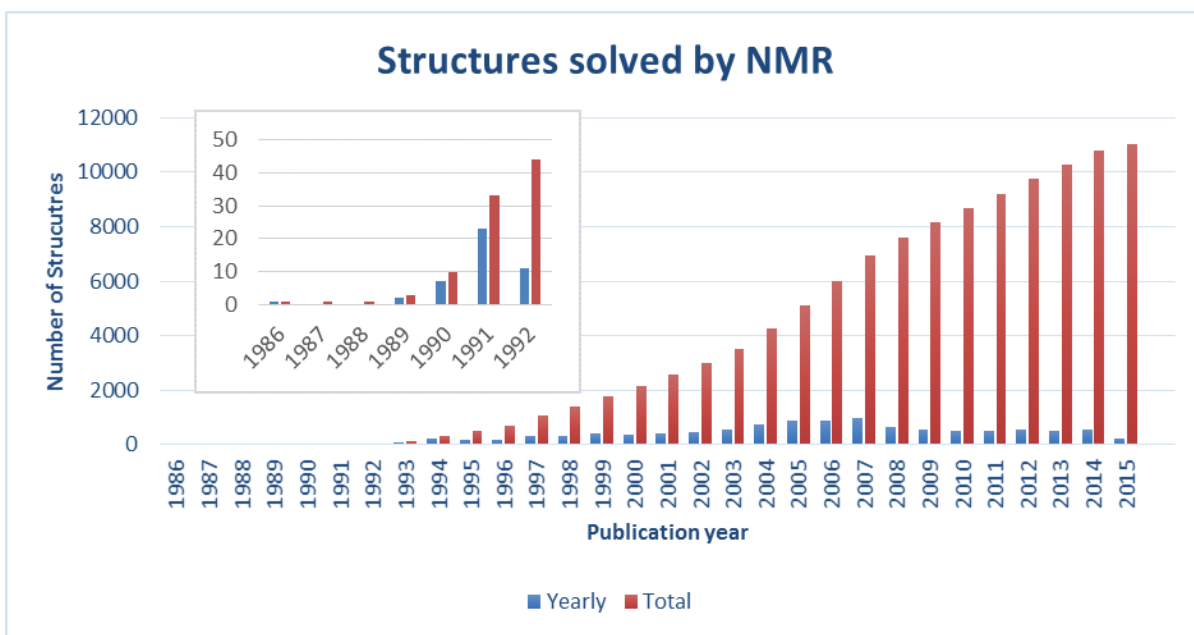


Figure 1. The growth of protein structure solved by NMR according to RCSB protein data bank statistics. The number of protein structures solved by NMR grows steadily each year since 1986. The maroon bars represent the total structures accumulated every year while the blue bars indicate the structures solved by NMR each year. The inset is the enlarged portion of earlier years with extended scale for clarification.

Although comparing to X-ray crystallography, the number of total protein structures determined by NMR is still very small, NMR can provide other types of information that is hardly amenable by crystallography. For example, a wide range of time-scale dynamics can be investigated by different NMR techniques, from slow exchange by magnetization exchange spectroscopy (EXSY) to fast exchange by spin relaxation [3]. Chemical shift mapping can also provide

information on which part of the protein is interacting with substrates or ligands. This is very important for small molecule screening and optimization in drug discovery. Protein-protein or protein-ligand interactions play a key role in numerous cellular processes. However, even with defined structures for both partners in an interaction, co-crystallization of the complex might be difficult due to low affinity or local disorder. NMR can complement these interaction studies by chemical shift perturbation (CSP), paramagnetic relaxation enhancement (PRE), intermolecular NOE, H/D exchange rates and residual dipolar coupling.

Another major advantage using NMR is that the sample flexibility. Not like protein crystallography, where the sample has to be crystallized and it is generally the bottleneck of applying crystallography for structure determination, NMR can be applied directly to protein in solutions or various solid forms such as powders, frozen solutions, microcrystals, gels or proteoliposomes [4-6]. This is extremely important for membrane proteins, which by nature is not soluble in aqueous solution and resist to crystallization.

The membrane proteins consisted almost 30% of the whole protein in living cell [7] yet only about 500 structures are solved to date comparing to the total of over 100,000 structures in PDB. Membrane proteins are functionally important in cells as they connect the outside world and the interior of the cell. Numerous diseases are related to membrane proteins such as hyperinsulinemia, nephrogenic diabetes insipidus, congestive heart failure, liver cirrhosis, cystic fibrosis, lung edema, *etc* [8]. Solving the structures of these membrane proteins will guide the drug design for curing these diseases. To achieve this, two major problems in structure

determination of membrane proteins must be overcome. The first is to get sufficient amount of membrane proteins expressed either *in vivo* or *in vitro*. The second is to apply the suitable technique to study the native state of the membrane protein. Recombinant protein expression has been used for decades and also readily applied to membrane proteins. However, they always tend to be expressed as inclusion bodies and require solubilization by denaturing conditions and then refold to its native structure. Detergents and/or lipids must be applied to make the protein soluble during the purification procedures. The final sample state is usually a mixture of proteins and detergents or lipids.

Solid state NMR (ssNMR) is particularly suitable for studying samples like this. In solid samples, the nuclear spinning experience three major interactions: dipolar, chemical shift anisotropy, and quadrupolar interactions, which will cause very broad and featureless lines. By applying magic angle spinning (MAS) at 54.74° [9], these interactions will be averaged out leaving a much narrower spectrum. Normally, a high degree of ^{13}C and ^{15}N isotope labeling is required if molecular structure is to be investigated under MAS conditions. More importantly, the samples could be prepared in lipid bilayer, which is the very native environment as they are *in vivo*. Then the structure analysis can be preceded using a series of multi-dimensional correlation experiments. For example, the local dihedral angle constraints defined by backbone chemical shifts will provide the secondary structure information of the peptide [10]. The sequential ^{15}N - ^{15}N distances can be used to refine the backbone topology [11]. The 3D structure will be finally determined if medium and long-range distance constraints can be derived from MAS NMR data. Selective labeling could also be applied if particular special interactions of atoms of interest are to be investigated.

The ssNMR is still at the very young stage. From all 541 unique membrane proteins in database, only 42 are determined by ssNMR. But it does provide a bright future for membrane protein structure determination. The sample requirements are not as constrained as other methods and the protein will stay in its very native state (in lipid bilayer) during the experiment.

In this dissertation, I will first briefly review the history of NMR spectroscopy in Chapter II. The basic principles and instrumentation of NMR will be introduced in Chapter III and IV, respectively. In Chapter V, I will show my work on a membrane protein PLIN1, which is the characteristic protein of lipid droplet located on the surface of this organelle. ssNMR was applied to try to decipher the structure of this protein. According to the NMR data, we proposed a membrane interacting model for this protein and this model was verified by both molecular dynamics simulation and experimental data. In Chapter VI, the soluble substrate binding domain of Mortalin was studied by solution NMR. We revealed its binding sites to a small anti-cancer molecule SHetA2. Through this interaction, SHetA2 competitively binds to Mortalin and release p53 from the Mortalin-p53 complex and the later enters cell nucleus, regulating cell apoptosis in tumor cells. According to our results, we also further proposed several SHetA2 analogs with greater binding affinities to Mortalin, which potentially will provide more stable interaction and better anti-cancer activity.

CHAPTER II

History of NMR

NMR originated from the ideas that spins and magnetic moments are the intrinsic properties of both electrons and nuclei. The first was established by several studies among which was the Stern-Gerlach experiment [12], where beams of atoms were separated in an inhomogeneous magnetic field according to the electron magnetic moment. Rabi *etc* [13] found later in 1933 that deflection of beams of hydrogen atoms through a homogeneous magnetic field subjected to radio frequency electromagnetic energy which was absorbed by the hydrogen atom in a very sharp frequency. This is the first time NMR was observed. Gorter [14] tried to detect the heat produced by the absorption of rf resonance energy and anomalous dispersion of the rf field to observe magnetic resonance in LiF. But the sample choice caused a fruitless attempt because of the long relaxation time of LiF. Instead of trying to detect the heat or dispersion, Bloch took another way to detect the resonance. According to physics law, he knew that the macroscopic magnetic moment would be rotated away from equilibrium position that was parallel to the applied magnetic field when the rf energy was applied. This displacement of the magnetic moment would then precess about the applied magnetic field at a well-defined frequency. This frequency could be detected if a resonance coil was applied, which is in the rf range. Bloch, Hansen and Packard performed the experiment on bulk water, and the first NMR or nuclear induction as they called was born [15]. Purcell, Torrey and Pound [2] fortunately and successfully detected the

absorption of rf energy by the magnetic moment of proton in bulk paraffin because not like LiF, proton has a relatively much shorter relaxation time. Both Bloch and Purcell were awarded Noble prize in 1952 as they discovered NMR independently.

In early days, NMR was not easy to perform as researchers had to build everything from scratch like magnets, coils, amplifiers. Russell Varian caught this opportunity and built the first commercial NMR spectrometer based on homogeneous magnet. The NMR technique developed rapidly thereafter as researchers can buy a basic NMR system rather than building everything from scratch. Bloembergen, Purcell and Pound [16] explained the concept of nuclear relaxation. They also stated the Brownian motion of liquids were the main reason that liquid NMR signals were magnitude narrower than solids because their nuclear magnetic dipole-dipole interaction averaged to zero. With the improvements of homogeneity of the magnet, more precise measurement on liquid were achievable. However, according to the theory, resonance of a particular atom should have a fixed frequency in a given magnetic field regardless of which chemical compounds the nucleus was in as shown in equation 1.

$$\omega = \gamma B_0 \tag{1}$$

However, the experimental results showed variations in of the signal from ^{19}F and ^{31}P beyond error level. It was then postulated that this is caused by the electromagnetic properties of the surrounding electrons of the nucleus. These electrons provided the so-called shielding effect and it can be denoted as coefficient σ . So the modified equation 1 was:

$$\omega = \gamma B_0(1 - \sigma) \tag{2}$$

Where the value of σ depended on local environment of the nucleus, specifically the density and configuration of electrons. This shift was very annoying to physicists at first because they could not precisely measure the anticipated resonance frequency to determine the gyromagnetic ratio. However, it was later found to be the cornerstone to apply NMR to chemistry, as chemical shift.

With the improvements of homogeneity and stability of magnetic field, ^1H chemical shifts were determined because the shielding for proton was magnitude smaller than those for other nuclei. The demonstration of proton chemical shifts for ethanol in 1951 was the first time to show to the chemists what NMR spectroscopy could do to analytical chemistry [17]. Improvements in resolution further showed that even in the chemical shift resonance lines, there were a collection of resonance lines. And it was found that neighboring nuclei were responsible for these multiplets [18] and thus the idea of spin-spin coupling (scalar coupling) were conceived [19]. However, the lack of multiplets for OH group in ethanol drove the development of chemical exchange.

Traditional continuous wave (cw) method varied the strength of the magnetic field of the spectrometer to find the resonance condition, which would show a deflection in the oscilloscope. However, this method was very time-consuming (several minutes per scan) and required high stability of the spectrometer during the entire acquiring period. Typical ^1H resonance scan could take 5 to 10 minutes per scan, if longer time was needed, the random drift of the magnetic field might be the dominant factor. Due to the nature of NMR, the sensitivity was always a problem, especially when the sample was very diluted or limited amount, or even less sensitive nuclei were studied like carbon or nitrogen. Coherent time averaging was applied to deal with the problem because after N consecutive scans, the signal level could be N times stronger, while noise level only increased $N^{1/2}$, thus increase the signal to noise ratio by a factor of $N^{1/2}$. However, this is

sometimes not practically possible as increasing 100 times signal to noise ratio requires 10,000 times of scans while each scan took 5 to 10 min to finish. Besides, most of the time only a small region of the whole cw spectrum was needed. The cw method was indeed time-consuming and inefficient. Bloch proposed the idea using PULSED rf excitation [1] as an alternative method rather than the traditional cw scan. It was first applied by mostly physicists to study systems with a single line, such as the broad line of a solid sample. It did not attract the chemists' attentions mainly due to the complexity of the FID. Lowe and Norberg [20] showed the first time that the FID following an rf pulse could in principle be transformed to a spectrum that a normal cw scan would achieve. Ernst and Anderson [21] made it real in their publication. It was the development of small sized computer that made Fourier transform more real than on paper. The FT-NMR revolutionized the field. Not only the sensitivity was enhanced but also the timing window for the experiments are much shorter so that the dynamics and real-time reactions can be studied by NMR spectroscopy.

Another Major breakthrough in NMR history is the development of 2D and multiple dimension spectroscopy. In 1971, Jeener originated the idea of 2D NMR spectroscopy, however, it was Ernst [22] that developed it into a practical method. In the 2D NMR spectroscopy, the nuclear magnetizations are allowed to precess during an initial time period, then various pulse sequences are applied to manipulate the coherence in this mixing period, and at last an FID is recorded. Applying Fourier transform on both dimension could transform the spectrum to be displayed on two orthogonal frequency dimensions. This method not only greatly increase the resolution comparing to 1 dimension spectrum, but also various interactions between different nuclei can be studied by applying different resonance frequencies. It is very powerful in assigning peaks in

complex spectra and studying interactions mediated by cross relaxation, chemical exchange or other physical factors [23].

Besides the wide application of NMR spectroscopy in organic chemistry, the developments of new powerful spectrometers, computer hardware, software, FT techniques, new multiple-dimension methods, and line narrowing in solids, NMR spectroscopy finally can reach the far more complicated biological macromolecules. For small molecules, double resonance is the major tool to solve the chemical bonding schemes through the investigation of spin-coupling connectivity by selective decoupling. And the NOE measurements could give valuable information on internuclear distances. However, when such methods applied to proteins and other complex biopolymers, the complexity of the spectrum and the tedious acquisition time for hundreds of spectra lead to very limited success. The breakthrough was when CORrelated SpectroscopY (COSY) [24], which permitted the establishment of spin-coupling connectivities, and Nuclear Overhauser Enhancement SpectroscopY (NOESY), which allowed the relaxation effects to be used to estimate internuclear distances, were invented by Nagayama, Wüthrich, Bachmann and Ernst in 1977. Although the information obtained from COSY and NOESY could be extracted from a large number of sequential 1D spin-decoupling and NOE experiments, the 2D approach permits simultaneous measurements throughout the entire spectrum, increasing the efficiency by magnitudes. The enormous data acquired from the 3D experiments require fast speed and large capacity of modern computer as well as sophisticated software to support the Fourier transformation. The development of modern computer was essential for the application of NMR methods [25]. Solvent signal suppression is another major breakthrough in NMR history so that protein spectra can be acquired in water, permitting the NH resonances to be included in the

spin-coupling and NOE pathways. With this crucial addition of experiment, complete 3D protein structure can be determined solely on NMR spectra and the first structure of a 57-residue protein was published in 1985 by Wüthrich [26]. Therefore, an alternative method was finally established besides X-ray crystallography for biopolymers.

The routes split into two at this point. In one direction, higher magnetic field and more sophisticated 3D methods were invented to get higher resolution thus more information can be obtained from the 3D spectra. In the other direction, some investigators were exploiting the speed and sensitivity of FT-NMR on small molecules in living cells, organisms or whole body. Bloch once joked about he was the first to apply *in vivo* study of NMR by putting his finger in the probe and observed water signals. However, water was the only signal they had at those years as they were so abundant in their spectra, little information could be obtained for other substances. Later, people were able to separate ^{31}P signals from the bulk water [27]. Many groups succeeded growing tissues in the NMR tubes so that they can study cellular systems in living cells [28, 29]. Excised organs were perfused in the NMR tubes with nutrients so the metabolism could be investigated. Surface coils were developed to obtain signals from localized volumes near the skin of experimental animals. Depth pulses were designed to tailor rf excitation and define more precisely the volume of interest. The rf coils were even planted in animals to examine the internal organs. However, there is another way to use NMR non-invasively by applying magnetic field gradient. The Larmor equation became modified by the position thus the NMR frequency became a measurement of position. By repeating the NMR measurements in different directions, the 2D or even 3D image could be reconstructed by combining the spectra [30, 31]. And this became a well-known method called NMR imaging (MRI). Paul Lauterbur at Stony Brook University

generated the first MRI image of a living mouse in 1974. People immediately realized that NMR imaging had great potential for investigating human and animal anatomy since different water content and relaxation times in different tissues can be distinguished. The images of normal and pathological tissues can be readily obtained. In the late 1970s, Peter Mansfield, a physicist and professor at the University of Nottingham, England, developed a mathematical technique that would allow scans to take seconds rather than hours and produce clearer images than Lauterbur had. Damadian, along with Larry Minkoff and Michael Goldsmith, performed the first MRI body scan of a human being on July 3, 1977. Commercial development of MRI began ever since and quickly grew exponentially. Nowadays, almost every hospital installs MRI as a standard equipment. In 2003, Nobel Prize in Physiology or Medicine was awarded to Paul Lauterbur and Peter Mansfield for their "discoveries concerning magnetic resonance imaging". Although MRI is most commonly performed at 1.5 T, higher fields such as 3T are gaining more popularity because of their increased sensitivity and resolution. In research laboratories, human studies have been performed at up to 9.4 T and animal studies have been performed at up to 21.1T.

NMR technology was quickly picked up by biologists and dramatic advances have been made in the application of high-resolution NMR spectroscopy in the determination of the 3D structure of large biopolymers, especially proteins. Earlier studies has been restricted to a molecular weight of 10 kDa due to the complexity of the 2D spectra and the increased linewidth as the increasing molecular size. The combination of isotope labeling (with ^{13}C and ^{15}N) and new experimental methods (transfer ^1H magnetization through the peptide link from one amino acid residue to another to permit backbone sequential assignment) overcome the problem and 3D even 4D NMR spectra could be obtained. The precision of NMR methods is as good as that of X-Ray

crystallography while the NMR has the advantage of determining the structure in solution or in membrane which mimic the native state of the protein and dynamic processes can also be examined.

CHAPTER III

THE BAISCS OF NMR

Section 1: Introduction

Spin is an intrinsic property of elementary particles. Quarks possess spins, thus the protons and neutrons which are made of quarks have spins. The combination of protons and neutrons, aka, nucleus, also possesses spins. The spin of a nucleus can be shown as a nuclear spin angular momentum \mathbf{I} ,

$$|\mathbf{I}| = [\mathbf{I} \cdot \mathbf{I}]^{1/2} = \hbar[I(I + 1)]^{1/2} \quad (3)$$

Where I is the angular spin quantum number, \hbar is the Planck's constant ($6.63 \times 10^{-34} \text{ m}^2\text{kg/s}$) divided by 2π . According to the angular spin quantum number I , all nuclei can be categorized into three groups as shown in Table 1, 1) nuclei with odd mass have half-integral spins. 2) nuclei with even mass but odd atomic number have integral spins. 3) nuclei with even mass and even atomic number have 0 spins. Since nuclear magnetic resonance rely on nuclear angular spin momentum. All nuclei in the last category are silent to NMR spectroscopy. Nuclei with a spin quantum number greater than $\frac{1}{2}$ possess electric quadrupole momentum due to the non-spherical distribution of electrons, which will significantly broaden the resonance lines and complicate the spectrum.

Table 1

Mass number	Atomic number	Nuclear spin angular momentum quantum number I
Odd	Odd or Even	$\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \dots$
Even	Even	0
Even	Odd	1, 2, 3, ...

Typical nuclei that are most relevant to biological studies are listed in Table 2, among which the most interesting ones are ^1H , ^{13}C , ^{15}N , ^{19}F , ^{31}P and ^2H , where the ^2H is an integer spin quantum number nucleus.

Due to the quantum limits, when $\mathbf{I}^2 \equiv \mathbf{I} \cdot \mathbf{I}$ is defined, only one of the three Cartesian components can be specified. By convention, z direction component will be specified as

$$I_z = \hbar m \quad (4)$$

, where m is the magnetic momentum quantum number and its value can be from $(-I, -I+1 \dots I-1, I)$. Since I_z can only be $2I+1$ discrete values and I is a constant, the spin angular momentum is quantized in space. It has $2I+1$ quantum states corresponding to the $2I+1$ values of I_z . In the absence of external field, all the quantum states have the same energy level and the spin angular momentum has no preferred orientation. The m also define a value called nuclear magnetic momentum, $\boldsymbol{\mu}$, which can be expressed by the nuclear spin angular momentum vector \mathbf{I} ,

$$\boldsymbol{\mu} = \gamma \mathbf{I} \quad (5)$$

$$\mu_z = \gamma I_z = \gamma \hbar m, \quad (6)$$

where γ is the gyromagnetic ratio. It is an intrinsic characteristic constant of the nucleus and it determines the receptivity of NMR spectroscopy. The γ value of some biologically important nuclei are shown in Table 2.

Table 2 magnetic properties of nuclei of biological interests

Isotope	Spin	Natural Abundance (%)	Quadrupole moment Q (10^{-28} m^2)	Gyromagnetic ratio γ ($10^7 \text{ rad s}^{-1} \text{ T}^{-1}$)	Relative Sensitivity	NMR frequency (MHz) at 14.1 Tesla
^1H	$\frac{1}{2}$	99.98	—	26.7522	100	600
^2H	1	1.5×10^{-2}	2.87×10^{-3}	4.1066	0.965	92.106
^3H	$\frac{1}{2}$	0	—	28.5350	121	639.978
^7Li	$\frac{3}{2}$	92.58	-3.7×10^{-2}	10.3976	29	233.178
^{11}B	$\frac{3}{2}$	80.42	4.1×10^{-2}	8.5847	17	192.504
^{13}C	$\frac{1}{2}$	1.108	—	6.7283	1.59	150.864
^{14}N	1	99.63	1.67×10^{-2}	1.9338	0.101	43.344
^{15}N	$\frac{1}{2}$	0.37	—	-2.7126	0.104	60.798
^{17}O	$\frac{5}{2}$	3.7×10^{-2}	-2.6×10^{-2}	-3.6280	2.91	81.342
^{19}F	$\frac{1}{2}$	100	—	25.1815	83	564.462
^{23}Na	$\frac{3}{2}$	100	0.10	7.0704	9.25	158.706
^{25}Mg	$\frac{5}{2}$	10.13	0.22	-1.6389	0.267	36.717
^{31}P	$\frac{1}{2}$	100	—	10.8394	6.63	242.886

³⁵ Cl	3/2	75.53	-8.2×10 ⁻²	2.6242	0.47	58.788
³⁹ K	3/2	93.1	5.5×10 ⁻²	1.2499	0.0508	28.002
⁴³ Ca	7/2	0.145	-5×10 ⁻²	-1.8028	0.64	40.386
⁵¹ V	7/2	99.76	2.17×10 ³	-5.2×10 ⁻²	38	157.734
⁵⁷ Fe	½	2.19	—	0.8687	3.37×10 ⁻³	19.386
⁷⁵ As	3/2	100	0.29	4.5961	2.51	102.756
⁷⁷ Se	½	7.58	—	5.1214	0.693	114.402
¹¹³ Cd	½	12.26	—	-5.9609	0.109	133.092

When there is an external magnetic field, the energies on each state is given by

$$E = -\boldsymbol{\mu} \cdot \mathbf{B}, \quad (7)$$

in which \mathbf{B} is the magnetic field vector. The vector $\boldsymbol{\mu}$ cannot collinear with the magnetic field \mathbf{B} due to the fact that $|\mathbf{I}| > I_z$. So the minimum value of energy can be obtained when the projection of $\boldsymbol{\mu}$ onto \mathbf{B} is maximized. The energy of each states are quantized and proportional to their projections onto \mathbf{B} . In NMR spectroscopy, external field are placed conveniently to the z directions, so the energy levels on each states are

$$E_m = -\gamma I_z B_0 = -m\hbar\gamma B_0, \quad (8)$$

where B_0 is the strength of the external magnetic field. So the projection of \mathbf{I} onto z axis resulted in $2I+1$ equally spaced energy levels, which are known as Zeeman splitting levels. It is a very important feature in NMR spectroscopy that the excited state has a lifetime on the order of 10^9 longer than that of an electronic excited state due to the Einstein's law of spontaneous emission

$$\tau \propto \frac{1}{\omega} \quad (9)$$

where τ is the lifetime of the excited state and ω is the transition frequency from ground state to excited state. According to Heisenberg's uncertainty principle, the long lifetime could result in an extremely narrow spectral lines in the NMR spectrum. So very subtle changes in the population differences between the two states can be readily detected and the long lifetime enable multi-dimensional NMR spectroscopy where the magnetization is transferred from one nucleus to another before detection.⁵

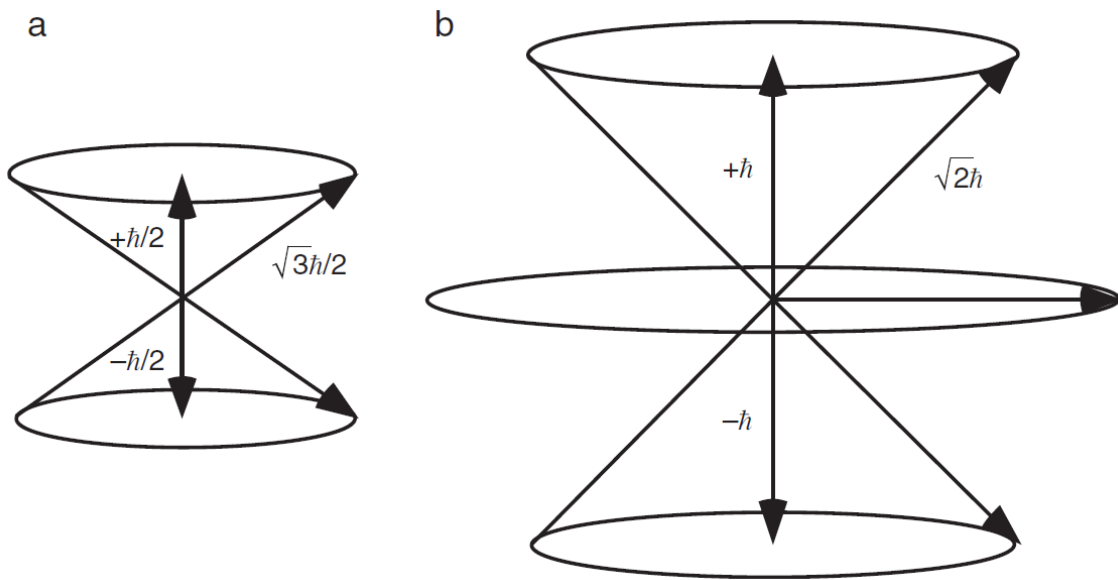


Figure 2 The nuclear spin quantum angular momentum I shown as the vector on the surface of the cone. It cannot be specified due to the quantum uncertainty of the x and y components. Its projection on the z axis are the quantized value for I_z . a) a spin $\frac{1}{2}$ nucleus. b) a spin 1 nucleus.

At equilibrium, according to Boltzmann distribution theory, the two energy states are unequally populated with the lower energy state being the preferred state. The population on each state is given by

$$\frac{N_m}{N} = \exp\left(-\frac{E_m}{k_B T}\right) / \sum_{m=-I}^I \exp\left(-\frac{E_m}{k_B T}\right) \quad (10)$$

where N_m is the population of the nuclei in the m th state, N is the total population of the spins, E_m is the energy on the m th state, k_B is the Boltzmann constant ($1.3806488 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$) and T is the absolute temperature in Kelvin. Substituting equation (8) into (9) will give

$$\frac{N_m}{N} = \exp\left(\frac{m\hbar\gamma B_0}{k_B T}\right) / \sum_{m=-I}^I \exp\left(\frac{m\hbar\gamma B_0}{k_B T}\right) \quad (11)$$

Expand equation (10) into first order Taylor series results in

$$\frac{N_m}{N} \approx \left(1 + \frac{m\hbar\gamma B_0}{k_B T}\right) / \sum_{m=-I}^I \exp\left(1 + \frac{m\hbar\gamma B_0}{k_B T}\right) \approx \left(1 + \frac{m\hbar\gamma B_0}{k_B T}\right) / (2I + 1) \quad (12)$$

So, the population on each state depends on both the nucleus type (γ) and the applied external field strength (B_0). The larger the external field, the larger the energy gap between two adjacent states, the larger differences in population between the two states. The energy difference between m th and $(m+1)$ th states is given by

$$\Delta E = \hbar\gamma B_0 \quad (13)$$

It is directly proportional to the external magnetic field for a given nucleus. For a spin $\frac{1}{2}$ nucleus, which is the most studied nuclear type in NMR, the electromagnetic frequency required to excite the spins from lower state or α state ($m = +1/2$) to upper state or β state ($m = -1/2$) is given by

$$\omega_0 = \gamma B_0 \quad (14)$$

$$\nu_0 = \frac{\omega_0}{2\pi} = \gamma B_0 / 2\pi \quad (15)$$

And the ratio of the spin population in the two energy states is given by

$$\frac{N_\beta}{N_\alpha} = \exp\left(-\frac{\Delta E}{k_B T}\right) = \exp\left(-\frac{\hbar\gamma B_0}{k_B T}\right) \quad (16)$$

For ^1H nuclei in a 14.1 Tesla magnetic field at room temperature (~ 300 K), the population difference between the two states is in the order of 1 in 10^5 . This tiny surplus from the lower energy state will be manipulated and detected as the NMR signal. Therefore, NMR is a relative insensitive spectroscopic technique comparing to visible or ultraviolet spectroscopy. This is also why building high magnetic field magnet is very critical in NMR development.

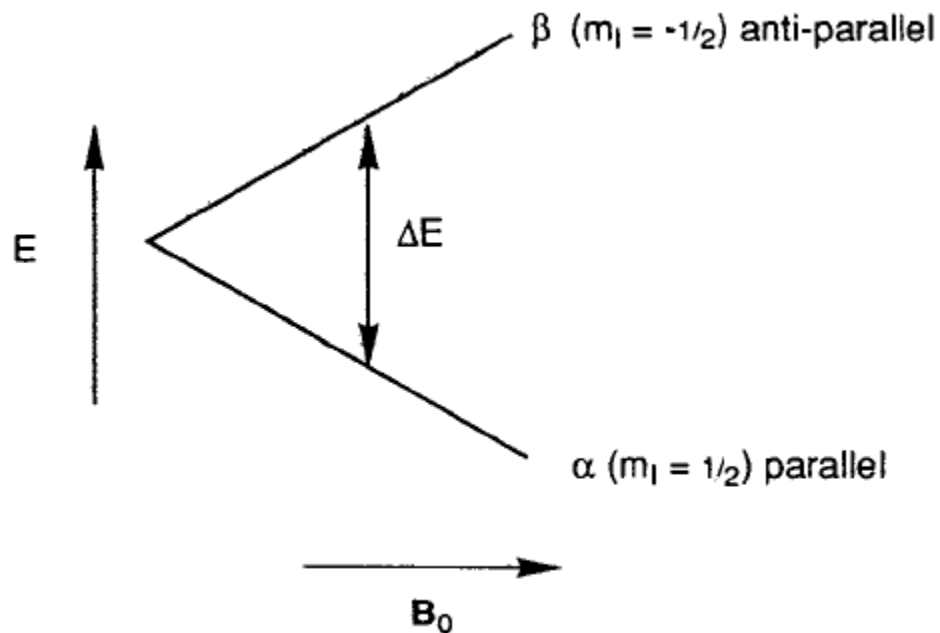


Figure 3 The two Zeeman levels for a spin $\frac{1}{2}$ nuclei. The nuclei in the lower energy state (α state for a positive γ nucleus) spin parallel with the external magnetic field, while those in the higher energy state (β state) spins anti-parallel. As the external static magnetic field strength increases, the energy gap between the two states increases.

There are two ways to view nuclear magnetic resonance, classical vector formalism and quantum mechanical density operator formalism.

Section 2: The Bloch Equation

When there is an external magnetic field, the spins states are unequally populated, the excess population in the lower energy state give rise the individual magnetic moment $\boldsymbol{\mu}$ and angular moment \mathbf{I} . The magnetic moments and angular moments from different nuclei in a given sample add up to a macroscopic bulk magnetic moment \mathbf{M} and bulk angular moment \mathbf{J} . At thermal equilibrium, the transverse components from the individual moments are not correlated and average out to 0. Both bulk magnetizations of the sample are along the same direction as the external field, z . So the bulk magnetization can be expressed as $\mathbf{M} = M_0\mathbf{k}$, where \mathbf{k} is the unit vector in z direction in a laboratory frame.

Bloch formulated a simple semi classical vector model to describe the evolution of the bulk magnetic moment \mathbf{M} over time. When an external magnetic field is present, the bulk magnetic moment experiences a torque that is equal to the time derivative of the angular momentum as shown

$$\frac{d\mathbf{J}(t)}{dt} = \mathbf{M}(t) \times \mathbf{B}(t) \quad (17)$$

Multiply both sides by γ yields

$$\frac{d\mathbf{M}(t)}{dt} = \mathbf{M}(t) \times \gamma\mathbf{B}(t) \quad (18)$$

This expression shows how the bulk magnetic moment evolves with time and can be shown using a rotating frame with an angular velocity ω . Initially, the rotating frame is superimposed with the laboratory frame, all vectors are represented exactly the same. While the time differentials are represented differently in the two frames. Their relations can be shown in

$$\left(\frac{d\mathbf{M}(t)}{dt}\right)_{rot} = \left(\frac{d\mathbf{M}(t)}{dt}\right)_{lab} + \mathbf{M}(t) \times \omega = \mathbf{M}(t) \times (\gamma\mathbf{B}(t) + \omega) \quad (19)$$

Comparing the two equations (17) and (18), it is shown that the expression of the rotation are identical for rotating frame and laboratory frame except that in rotating frame, the magnetic field is replaced with an effective magnetic field \mathbf{B}_{eff} :

$$\mathbf{B}_{eff} = \mathbf{B}(t) + \omega/\gamma \quad (20)$$

When $\omega = -\gamma\mathbf{B}(t)$, the effective field \mathbf{B}_{eff} is 0. The bulk magnetization $\mathbf{M}(t)$ is time independent in the rotating frame and it precesses around $\mathbf{B}(t)$ at a constant frequency $\omega = -\gamma B$ in laboratory frame, which is known as the Larmor frequency. For an external static field B_0 , the bulk magnetization precesses around the main static field at its Larmor frequency $\omega_0 = -\gamma B_0$ which is exactly the same frequency required to excite the transitions between Zeeman levels.

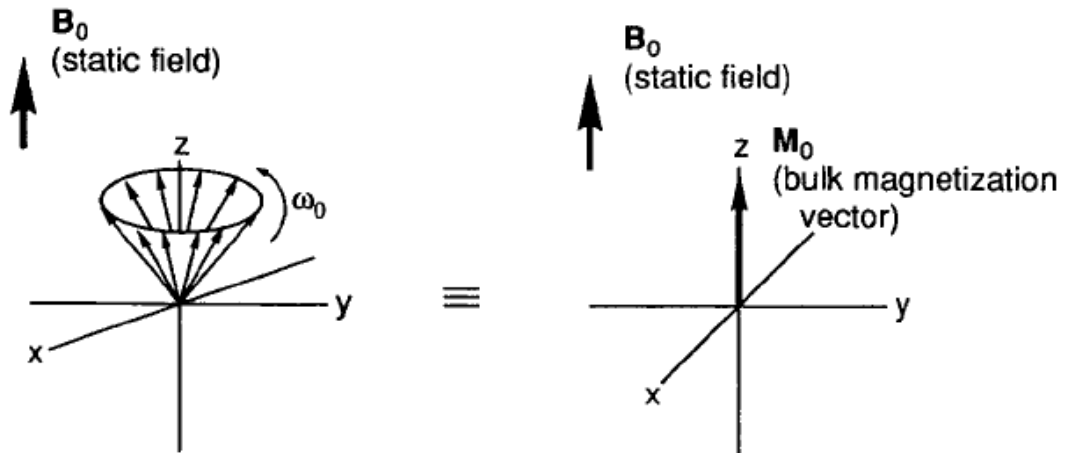


Figure 4 Classical view of the bulk magnetization \mathbf{M} in an external static magnetic field \mathbf{B}_0 . It precesses around the B field at its Larmor frequency. The net result is a bulk magnetization vector M_0 along the same direction as the B_0 field, z direction.

Classical description only applies to isolated spins as each spin can be treated as a vector precessing around the main static magnetic field.

As the bulk magnetization precesses, the varying magnetic field should create an induced electromotive force due to Faraday's induction law. However, at thermal equilibrium, the bulk magnetization vector is collinear with the external magnetic field thus no induced current or emf can be detected due to the overwhelming external static magnetic field. But electromagnetic radiation can be represented as an oscillating magnetic field that interact with the bulk magnetization (as shown in Figure 5). With suitable frequency, normally in the radio frequency (rf) range for NMR, the bulk magnetization can be manipulated and the signal from the precession of the bulk magnetization can be readily detected.

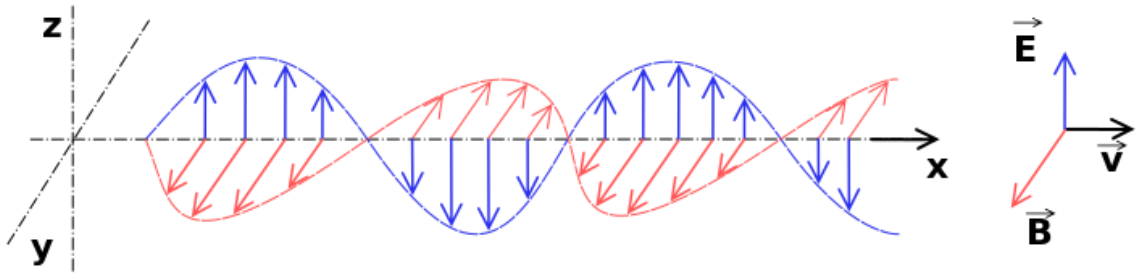


Figure 5 Electromagnetic waves can be treated as oscillating magnetic field.

This short burst of radio frequency electromagnetic radiation is referred as a pulse. For an rf field linearly polarized to the x axis of the laboratory frame can be written as

$$\mathbf{B}_{rf}(t) = 2B_1 \cos(\omega_{rf} + \phi) \mathbf{i} \quad (21)$$

$$= \mathbf{B}_1 \{ \cos(\omega_{rf} + \phi) \mathbf{i} + \sin(\omega_{rf} + \phi) \mathbf{j} \} + \mathbf{B}_1 \{ \cos(\omega_{rf} + \phi) \mathbf{i} - \sin(\omega_{rf} + \phi) \mathbf{j} \} \quad (22)$$

where B_1 is the amplitude of the applied field, ω_{rf} is the angular frequency of the rf field or carrier frequency, ϕ is the phase of the field. As shown in equation (21), the rf pulse can be decomposed into two circularly polarized fields rotating in opposite directions about the z axis. The decomposition of the rf field at difference phases are shown in Figure 6.

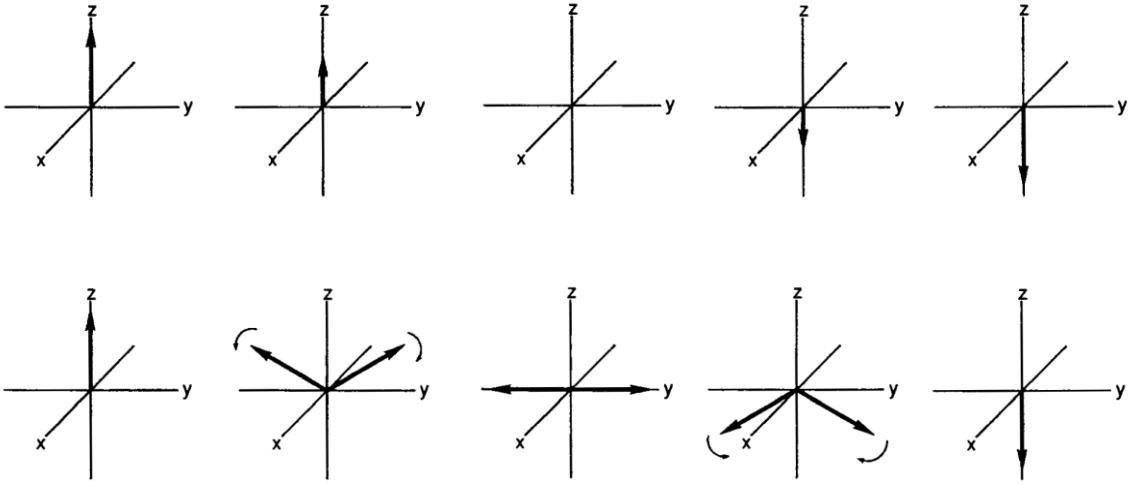


Figure 6 The illustration of the decomposition of radio frequency radiation into two circularly polarized fields rotating in opposite directions about the z axis.

Only the field rotating in the same direction as the bulk magnetic moment will interact significantly with the bulk magnetization. The counter-rotating, non-resonant field only interact with the bulk magnetization to the order of $(B_1/2B_0)^2$. Typically this value is very small and this interaction will be discarded. Thus, the simplification form of equation (21) will be

$$\mathbf{B}_{rf}(t) = B_1 \{ \cos(\omega_{rf} + \phi) \mathbf{i} + \sin(\omega_{rf} + \phi) \mathbf{j} \} \quad (23)$$

In this case, where it is time dependent, the rotating frame transformation can be applied that makes the rotating frame rotates at angular frequency of ω_{rf} about the z axis so that the equation of motion form the magnetization in the rotating frame $\mathbf{M}'(t)$ can be written as

$$\frac{d\mathbf{M}^r(t)}{dt} = \mathbf{M}^r(t) \times \gamma\mathbf{B}^r \quad (24)$$

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

$$\mathbf{B}^r = B_1 \cos \phi \mathbf{i}^r + B_1 \sin \phi \mathbf{j}^r + \Omega/\gamma \mathbf{k}^r \quad (25)$$

in which the $\Omega = -\gamma B_0 - \omega_{rf} = \omega_0 - \omega_{rf}$ is the offset, the difference between the effective field and resonance field. ΔB_0 is known as the reduced static field affected by the B_1 field and is equivalent to the z component of the effective field, namely $-\Omega/\gamma$. And the phase angle is defined such that for an rf field of fixed phase ϕ , $B_x = B_1 \cos \phi$, $B_y = B_1 \sin \phi$. The magnitude of the effective field is given by

$$B^r = \sqrt{(B_1 \cos \phi)^2 + (B_1 \sin \phi)^2 + (\Delta B_0)^2} = B_1 \sqrt{1 + (\tan \theta)^2} \quad (26)$$

The θ is the angle which the effective field is tilted from the z axis and it is defined as

$$\tan \theta = \frac{B_1}{\Delta B_0} \quad (27)$$

So, the direction of the effective field is defined by both θ , which is in turn defined by the strength of rf field B_1 , and the offset, the difference between the carrier frequency and Larmor frequency, and ϕ , the phase of the rf field in the laboratory frame. Their relations are shown in Figure 7.

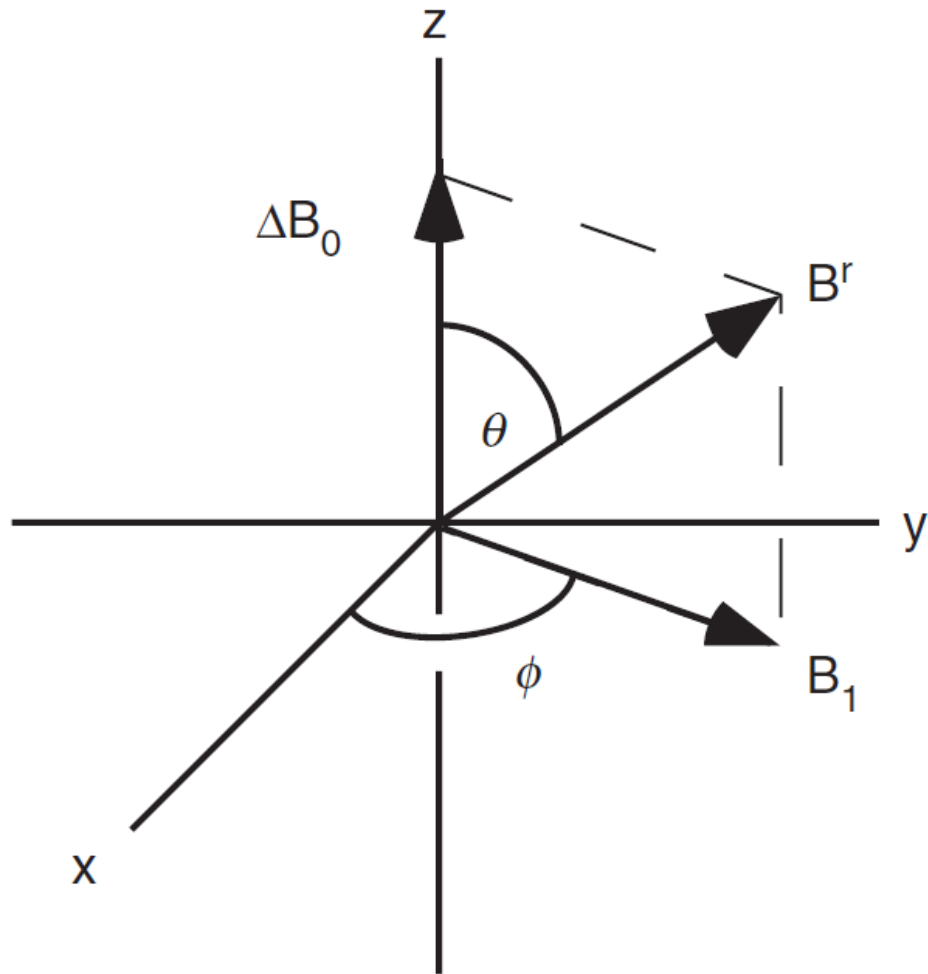


Figure 7 The orientation of reduced static magnetic field B_0 , radio frequency field B_1 and effective field B^r in the rotating frame.

The $M^r(t)$ precesses around the effective field B^r with an angular frequency of ω^r where

$$\omega^r = -\gamma B^r \quad (28)$$

If the carrier frequency ω_{rf} is equal to the Larmor frequency ω_0 , then the offset Ω which is $\omega_0 - \omega_{rf}$ will be 0. At this frequency, the impact of the external static magnetic field B_0 is removed. The

effective field will be collinear with the applied field \mathbf{B}_1 . This is a so called on-resonance. When on resonance, the bulk magnetization $\mathbf{M}(t)$ precesses around the \mathbf{B}_1 field with a frequency defined by the \mathbf{B}_1 field, $\omega^r = -\gamma B^r = -\gamma B_1 \equiv \omega_1$.

If the radio frequency field is turned on for a period time of τ_p , which is called the pulse width or pulse length. The effective flip angle α of $\mathbf{M}(t)$ from effective field is given by

$$\alpha = \omega^r \tau_p = -\gamma B^r \tau_p \quad (29)$$

The precession of the bulk magnetization around the effective field in the rotating frame can be illustrated in Figure 8.

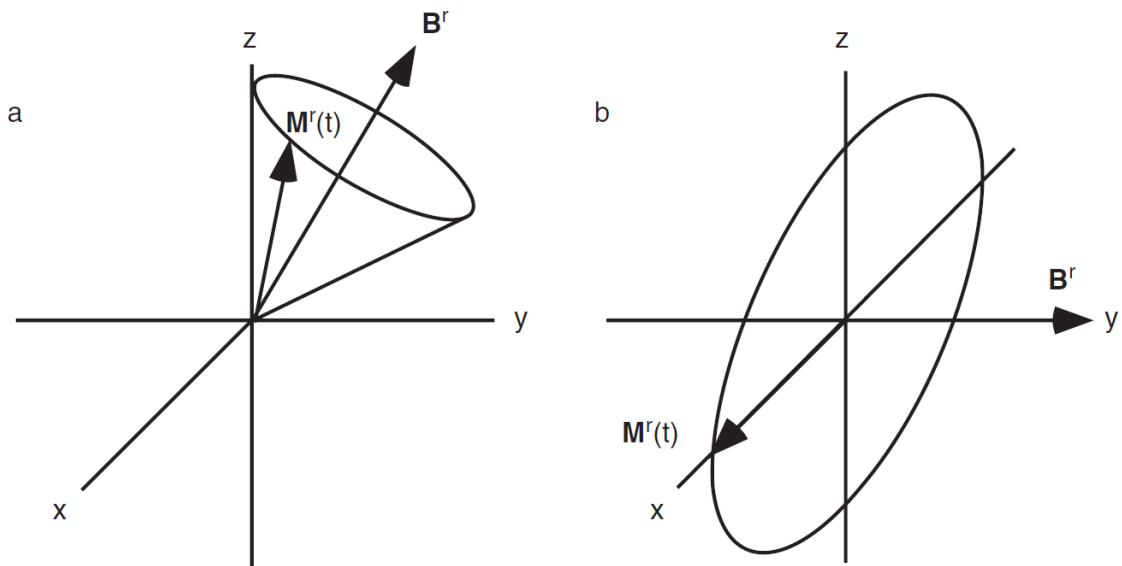


Figure 8 The bulk magnetization precessing around the effective field in the rotating frame after a y phase radio frequency pulse is applied. a) Non-resonance case, where the effective field is in the y-z plane. b) On-resonance case, the effective magnetic field is collinear with the applied field in the y axis, the bulk magnetization is flipped 90° from \mathbf{B}_1 and precess in the x-z plane.

Section 3: Relaxation

The magnetization won't stay in the x-y plane to spin forever after the pulse. However, it will slowly come from Zeeman splitting back to thermal equilibrium Boltzmann distribution. This process is defined as the spin-lattice relaxation or longitudinal relaxation. It is the mechanism by which the component of the magnetization vector along the direction of the static magnetic field reaches thermodynamic equilibrium with its surroundings (the "lattice"). R_1 is defined to characterize the rate at which the longitudinal M_z component of the magnetization vector recovers towards its thermodynamic equilibrium, according to equation:

$$\frac{dM_z(t)}{dt} = R_1[M_0 - M_z(t)] \quad (30)$$

Where M_0 is the magnetization at thermal equilibrium. Solving this equation gives

$$M_z(t) = M_0 - [M_0 - M_z(0)]e^{-R_1 t} \quad (31)$$

$M_z(0)$ is the magnetization at $t = 0$ where it is the beginning point of the decay process. So the longitudinal magnetization recovers to thermal equilibrium in the exponential fashion. In a special case, where $M_z(0) = 0$, which means the magnetization is flipped to the x-y plane after the rf pulse, the equation can be written as

$$M_z(t) = M_0(1 - e^{-R_1 t}) \quad (32)$$

So we can define a time $T_1 = 1/R_1$, which means the time it takes for the longitudinal magnetization to recover approximately 63% of its initial value after being flipped into the magnetic transverse plane by a 90° radiofrequency pulse.

Another process account for the NMR decay is the spin-spin relaxation, which the transverse components of the magnetization vector exponentially decay to their equilibrium states. It can be shown as

$$\frac{dM_{xy}}{dt} = -R_2 M_{xy} \quad (33)$$

where R_2 is defined as the spin-spin relaxation rate constant of the transverse components of the magnetization vector and its reverse T_2 , which is $1/R_2$, is the spin-spin relaxation time constant.

Solving this equation would result in

$$M_{xy} = M_{xy}(0)e^{-R_2 t} = M_{xy}(0)e^{-t/T_2} \quad (34)$$

So the T_2 is defined as the transverse components of the magnetization decay to 37% of its initial value ($t = 0$) which is caused by the rf pulse. Different material have dramatically different T_2 s.

For proton as an example, it can vary from sub milliseconds (in crystals) to seconds (in fluids). So it can be used as a detecting method to study the different components in a given sample. Putting both relaxation rates back into equation (17) generates the famous Bloch equations in the laboratory frame

$$\frac{dM_x(t)}{dt} = \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_x - R_2 M_x(t) \quad (35)$$

$$\frac{dM_y(t)}{dt} = \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_y - R_2 M_y(t) \quad (36)$$

$$\frac{dM_z(t)}{dt} = \gamma[\mathbf{M}(t) \times \mathbf{B}(t)]_z - R_1[M_z(t) - M_0] \quad (37)$$

where $\mathbf{M}(t)$ is the bulk magnetization, $\mathbf{B}(t)$ is the magnetic field experienced by the nuclei and it is equal to $B_x(t)\mathbf{i} + B_y(t)\mathbf{j} + (B_0 - \Omega/\gamma)\mathbf{k}$. Opening up the brackets would give

$$\frac{dM_x(t)}{dt} = \gamma[M_y(t)B_z(t) - M_z(t)B_y(t)] - R_2 M_x(t) \quad (38)$$

$$\frac{dM_y(t)}{dt} = \gamma[M_z(t)B_x(t) - M_x(t)B_z(t)] - R_2 M_y(t) \quad (39)$$

$$\frac{dM_z(t)}{dt} = \gamma[M_x(t)B_y(t) - M_y(t)B_x(t)] - R_1[M_z(t) - M_0] \quad (40)$$

In the rotating frame where $\mathbf{B}^r = B_1 \cos \phi \mathbf{i}^r + B_1 \sin \phi \mathbf{j}^r + \Omega/\gamma \mathbf{k}^r$, if neither B_1 nor ϕ is time-dependent and the pulse is only applied for a short time so that $\tau_p \ll T_1$ or T_2 , these equations can be written as

$$\frac{dM_x(t)}{dt} = \Omega M_y(t) - \omega_1 \sin \phi M_z(t) \quad (41)$$

$$\frac{dM_y(t)}{dt} = \omega_1 \cos \phi M_z(t) - \Omega M_x(t) \quad (42)$$

$$\frac{dM_z(t)}{dt} = \omega_1 \sin \phi M_x(t) - \omega_1 \cos \phi M_y(t) \quad (43)$$

It can also be written in a matrix form

$$\frac{d\mathbf{M}(t)}{dt} = \frac{d}{dt} \begin{bmatrix} M_x(t) \\ M_y(t) \\ M_z(t) \end{bmatrix} = \begin{bmatrix} 0 & \Omega & -\omega_1 \sin \phi \\ -\Omega & 0 & \omega_1 \cos \phi \\ \omega_1 \sin \phi & -\omega_1 \cos \phi & 0 \end{bmatrix} \mathbf{M}(t) \quad (44)$$

The solution to this equation can be represented by a series of rotations of the magnetization

$\mathbf{M}(0)$:

$$\mathbf{M}(\tau_p) = \mathbf{R}_z(\phi) \mathbf{R}_y(\theta) \mathbf{R}_z(\alpha) \mathbf{R}_y(-\theta) \mathbf{R}_z(-\phi) \mathbf{M}(0) \quad (45)$$

Each of these rotation operators can be written as

$$\mathbf{R}_x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix} \quad (46)$$

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \quad (47)$$

$$\mathbf{R}_z(\beta) = \begin{bmatrix} \cos \beta & -\sin \beta & 0 \\ \sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (48)$$

And rotation operator $\mathbf{R}_y(\beta)$ means the right hand rotation about y axis for an angle of β . In an ideal pulse NMR experiment, an on-resonance rf pulse of duration τ_p , strength B_1 , tilt angle $\theta = \pi/2$, phase $\phi = \pi/2$, the evolution of the magnetization after the pulse can be found by using the rotation operators as

$$\mathbf{M}(\tau_p) = \mathbf{R}_y(\alpha)M_0 = \mathbf{i}M_0 \sin \alpha + \mathbf{k}M_0 \cos \alpha = \begin{pmatrix} M_0 \sin \alpha \\ 0 \\ M_0 \cos \alpha \end{pmatrix} \quad (49)$$

where $\alpha = -\gamma B^r \tau_p$, depending on the magnetic field strength and duration of the pulse. A maximum transverse magnetization can be generated with $\alpha = 90^\circ$. This rf pulse is called a 90° or a $\pi/2$ pulse, which equalizes the population of both spins states for the spin $1/2$ nuclei. And a 180° pulse or π pulse reverse the population in the two spin states as shown in Figure 9.

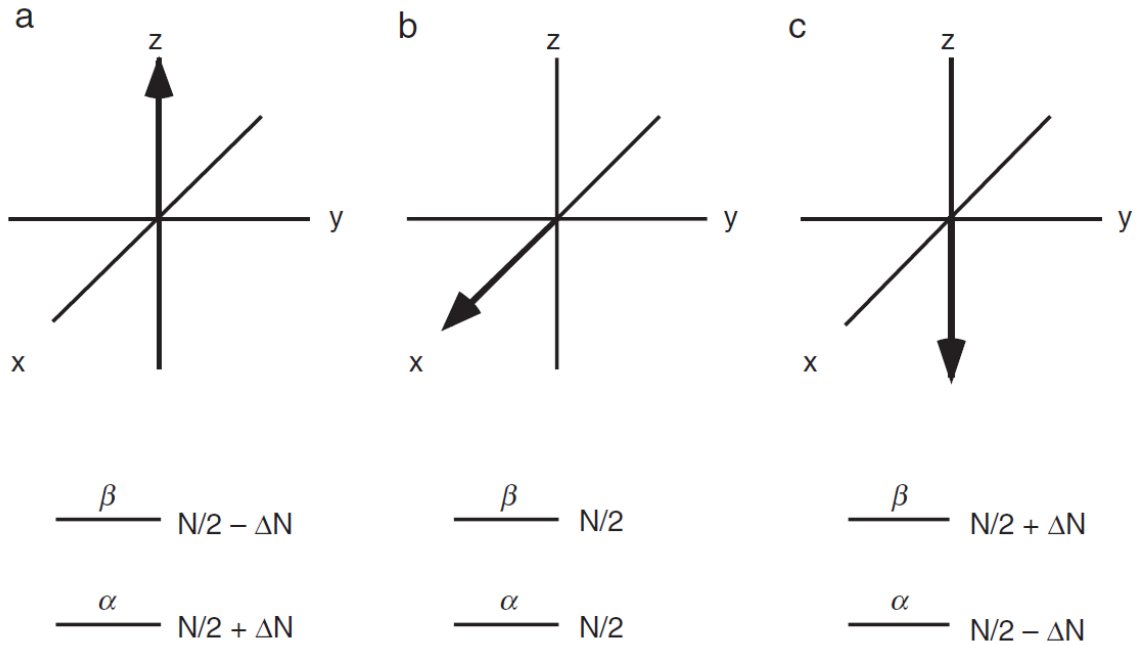


Figure 9 The effects of on-resonance pulses. a) at thermal equilibrium, the two spin state are unequally populated according to Boltzmann distribution where N is the total number of spins and $\Delta N = N\hbar\gamma B_0/(4k_B T)$. b) after a 90° pulse, the population in both states are equal. c) after a 180° pulse, the population in the two states are reversed.

The relaxations after the pulse can be detected during the so called acquisition time t which generates the signal called free induction decay (FID). This is the signal detected by the NMR spectrometer. Simplify the Bloch equation following the 90° pulse, $B_{xy} = 0$, $B_z = B_0$, the free precession Bloch equation can be written as

$$\frac{dM_x(t)}{dt} = \Omega M_x(t) - R_2 M_x(t) \quad (50)$$

$$\frac{dM_y(t)}{dt} = -\Omega M_y(t) - R_2 M_y(t) \quad (51)$$

$$\frac{dM_z(t)}{dt} = -R_1 [(M_z(t) - M_0)] \quad (52)$$

The solution can be derived from these equations and they can be shown as

$$M_x(t) = M_0 \sin \alpha \cos(\Omega t) e^{-R_2 t} \quad (53)$$

$$M_y(t) = M_0 \sin \alpha \sin(\Omega t) e^{-R_2 t} \quad (54)$$

The combined complex notation of the solutions can be expressed as

$$M^+(t) = M_x(t) + iM_y(t) = M_0 \sin \alpha e^{i\Omega t - R_2 t} \quad (55)$$

So the precession of the magnetization after the rf pulse is damped by the spin-spin relaxation exponential factor $-R_2 t$. For the on-resonance 90° pulse, where $\Omega = 0$, $\sin \alpha = 1$, the signal is an exponentially decay curve. The off-resonance pulses, the decaying curve will be damped by the offset as illustrated in Figure 10.

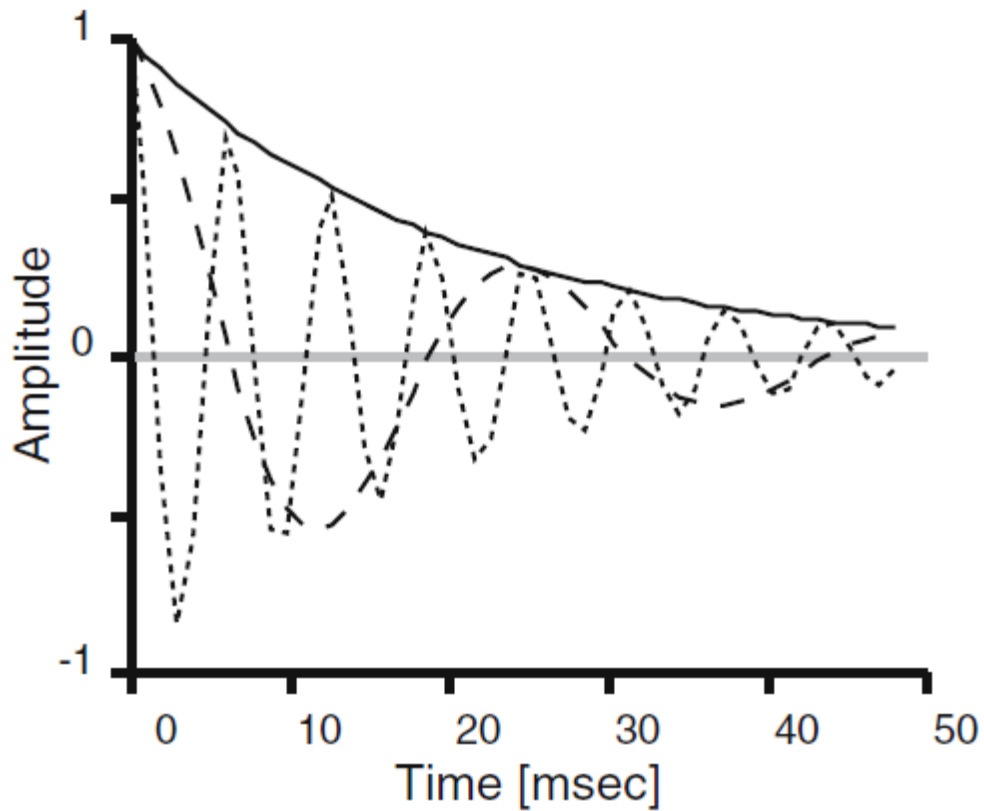


Figure 10 The on-resonance and off-resonance decay of the NMR signal. Solid line indicates the decay after an on-resonance pulse, while dashed and dotted lines represent the signal of spins that are 150 Hz and 650 Hz off-resonance, respectively. All three resonances share the same T_2 so they have the same envelope.

Both the real and imaginary parts of the signal typically will be detected in the real NMR experiments with a proportional factor λ

$$s^+(t) = \lambda M^+(t) \quad (56)$$

In modern NMR, the decaying signal will be Fourier transformed into the frequency domain and give

$$S(\omega) = \int_0^{\infty} s^+(t)e^{-i\omega t} dt = v(\omega) + iu(\omega) \quad (57)$$

where

$$v(\omega) = \lambda M_0 \frac{R_2}{R_2^2 + (\Omega - \omega)^2} \quad (58)$$

$$u(\omega) = \lambda M_0 \frac{\Omega - \omega}{R_2^2 + (\Omega - \omega)^2} \quad (59)$$

The real part represents an absorptive Lorentzian lineshape and the imaginary part is the corresponding dispersive Lorentzian lineshape. Typically the real part of the signal will be displayed in the frequency domain of NMR spectrum.

Section 4: Chemical Shift

As shown before, magnetization would precess around the external magnetic field with their respective Larmor Frequencies and the precession can be detected after it was flipped to the x-y plane by an rf pulse. However, if all same nuclei have the same resonance transition frequency, NMR spectroscopy would not much useful except telling which elements are in the sample. In reality, the observed resonance frequencies depend on the local environments of each individual nucleus as the motion of electrons induced by the external magnetic field generates a secondary magnetic field that cause a shielding effect on the target nucleus. This difference, defined as chemical shift, enables the possibilities to distinguish same nuclear type in different environments thus make NMR useful in detecting different structures of the molecule.

The local field experienced by the nucleus with the shielding effect is given by

$$B = (1 - \sigma)B_0 \quad (60)$$

where σ represents the shielding effect of the nucleus spin. For an isotropic electron distribution, the shielding is given by the Lamb Formula

$$\sigma = \frac{e^2}{3mc^2} \int \frac{\rho(r)}{r} dr \quad (61)$$

When the electron density around the nucleus increases, the shielding effect decreases, leading to a lower resonance frequency.

For an anisotropic electron distribution, which is in most cases, the shielding effect is described by the second-rank nuclear shielding tensor, represented by a 3×3 matrix.

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \quad (62)$$

In the principle axis system (PAS), where the z axis coordinates with respect to the spin of nucleus. The electric field surrounding the nucleus can be described by the principle components of the chemical shift tensor σ_{xx} , σ_{yy} , σ_{zz} . The isotropic chemical shift tensor can be written as the average of the three principle components

$$\sigma_{iso} = \frac{\sigma_{xx} + \sigma_{yy} + \sigma_{zz}}{3} \quad (63)$$

In solution, the rapid reorientation of the molecule caused by the collisions average out the chemical shift anisotropy that only a single isotropic chemical shift value can be observed. But in solids, the molecules will normally be oriented in all possible directions with respect to the external magnetic field and so a superposition of all possible chemical shifts is observed for each nucleus. Each magnetic moment of the system can be related for its PAS to the z axis by an angle,

θ , and its position in the x-y plane given by angle ϕ . The magnetization experienced at the nucleus on the basis of the PAS is given by

$$B_0^{PAS} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) \quad (64)$$

The chemical shift is written as

$$\omega = -\omega_0(\sigma_{xx} \sin^2 \theta \cos^2 \phi + \sigma_{yy} \sin^2 \theta \sin^2 \phi + \sigma_{zz} \cos^2 \theta) \quad (65)$$

This equation can be reduced to

$$\omega = -\omega_0 \sigma_{iso} - \frac{1}{2} \omega_0 \Delta (3 \cos^2 \theta - 1 + \eta \sin^2 \theta \cos 2\phi) \quad (66)$$

in which the $-\omega_0 \sigma_{iso}$ is the isotropic chemical shift that can be observed in liquid state, Δ is the chemical shift anisotropy and η is the asymmetry parameter.

This gives rise to broad peaks with a very characteristic shape as shown in Figure 11. In

Haerberlen convention, the chemical shift anisotropy and asymmetry parameter are defined as

$$\Delta = \sigma_{zz} - \sigma_{iso} \quad (67)$$

$$\eta = \frac{\sigma_{xx} - \sigma_{yy}}{\sigma_{zz} - \sigma_{iso}} \quad (68)$$

$$\delta = \frac{3\Delta}{2} \quad (69)$$

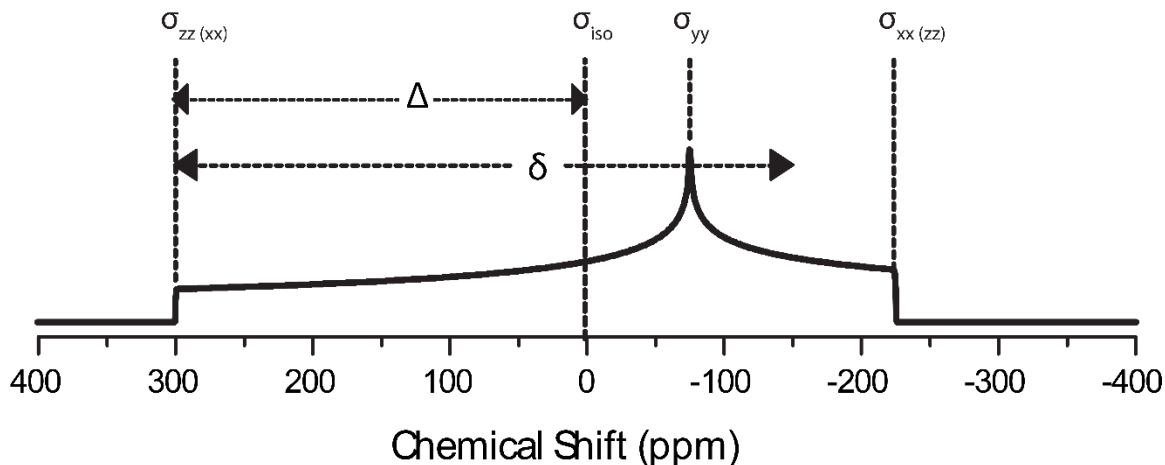


Figure 11 Haeberlin convention of chemical shift anisotropy. The three principle components of the chemical shift tensors are shown as indicated in the dashed lines. Notice the huge spam of the chemical shift tensors in solid.

The resonance frequency is directly proportional to the external magnetic field. So the difference in chemical shift between two resonances measured in frequency will also change when the external magnetic field changes. It is very inconvenient to compare the results of the same sample obtained from different spectrometers that have different magnet field strengths. Besides, any little drift from the external field would cause a different resonance frequency so the magnet has to be measured very precisely and it is almost impossible even in a single experiment, let alone from one measurement to another. So, in practical, the chemical shift is measured in parts per million (ppm) relative to a reference resonance signal of standard molecule instead of direct frequency. The relative chemical shift can be expressed as

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 = (\sigma_{ref} - \sigma) \times 10^6 \quad (70)$$

The Ω and Ω_{ref} are the offsets of the sample and the reference, respectively, and ω_0 is the Larmor frequency for this type of nucleus. Chemical shifts measure in this way are independent of the

external static magnetic field strength and signals from different spectrometers and experiments can be readily compared using ppm.

Section 5 Limitation of classical model

Classical vector model is very useful in displaying the bulk magnetization of isolated spins. However, it is very limited considering the interactions between spins. There are mainly two types of interactions between two spin nuclei: dipolar coupling and scalar (spin-spin) coupling.

Dipolar coupling is extremely important in solids, however, in solution, it was averaged out to 0 by the fast tumbling of the molecules. The splitting of the signals in the solution is due to the through bond coupling, which is mediated by the paired electrons forming chemical bond between the two nuclei. The interaction, named scalar coupling constant, is measured by ${}^nJ_{ab}$, where n is the number of chemical bonds separating the two nuclei and a, b indicate the two nuclei. Typically, the J coupling within 4 chemical bonds will be considered significant in NMR spectroscopy.

J coupling constant will modify the energy level separated by the Zeeman splitting. Without J coupling, the two spin system for nuclei I and S can be shown by four wave functions

$$\psi_1 = \psi\left(\frac{1}{2}, \frac{1}{2}\right), \psi_2 = \psi\left(\frac{1}{2}, -\frac{1}{2}\right), \psi_3 = \psi\left(-\frac{1}{2}, \frac{1}{2}\right), \psi_4 = \psi\left(-\frac{1}{2}, -\frac{1}{2}\right), \quad (71)$$

in which the first number in the parenthesis indicates the state of spin nucleus I, while the second of S. The energy levels on these four level can be shown as

$$\begin{aligned}
E_1 &= \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S, E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S, \\
E_3 &= -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S, E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S
\end{aligned} \tag{72}$$

Given $\gamma_I > \gamma_S > 0$, the energy levels can be illustrated as shown in part a of Figure 11. The transition between states must obey the selection rule where $\Delta m = \pm 1$. So the transition can only happen between states 1-2, 1-3, 2-4 and 3-4. All other transitions are forbidden. The energy differences between these transitions are

$$|E_2 - E_1| = \hbar\omega_S, |E_3 - E_1| = \hbar\omega_I, |E_4 - E_2| = \hbar\omega_I, |E_4 - E_3| = \hbar\omega_S \tag{73}$$

Thus, the transition between energy states 1-2 and 3-4 share the same resonance frequency ω_S , while the transition between energy states 1-3 and 2-4 share ω_I . So a total of two resonance lines are predicted as shown in Figure 12.

However, if the scalar coupling between nuclei I and S exists, the coupling constant J_{IS} will modify each energy levels as (assuming weak coupling with $2\pi|J_{IS}| \ll |\omega_I - \omega_S|$)

$$E(m_I, m_S) = m_I\omega_I + m_S\omega_S + 2\pi m_I m_S J_{IS} \tag{74}$$

So the resulting energy levels on each states are given by

$$\begin{aligned}
E_1 &= \frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS}, E_2 = \frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS}, \\
E_3 &= -\frac{1}{2}\hbar\omega_I + \frac{1}{2}\hbar\omega_S - \frac{1}{2}\pi\hbar J_{IS}, E_4 = -\frac{1}{2}\hbar\omega_I - \frac{1}{2}\hbar\omega_S + \frac{1}{2}\pi\hbar J_{IS}
\end{aligned} \tag{75}$$

And the transitions would still obey the selection rules such that only the transitions between energy states 1-2, 1-3, 3-4, 2-4 can happen. And the energy differences between these transitions are given by

$$\begin{aligned}
|E_2 - E_1| &= \hbar\omega_S + \pi\hbar J_{IS}, |E_3 - E_1| = \hbar\omega_I + \pi\hbar J_{IS}, \\
|E_4 - E_3| &= \hbar\omega_S - \pi\hbar J_{IS}, |E_4 - E_2| = \hbar\omega_I - \pi\hbar J_{IS}
\end{aligned} \tag{76}$$

And the resonance lines would be shown as a modified resonance close to its original position

$$\omega_{12} = \omega_S + \pi J_{IS}, \omega_{13} = \omega_I + \pi J_{IS}, \omega_{34} = \omega_S - \pi J_{IS}, \omega_{24} = \omega_I - \pi J_{IS} \tag{77}$$

So, none of these transitions share the same resonance frequency but rather has their own distinct resonance lines with a modification term πJ_{IS} as shown in part d of Figure 11.

The classical vector model would only predict two resonance lines for the two spins system.

However, considering scalar coupling, there are actually 4 resonance spectral lines can be observed. So, strictly speaking, classical vector model and Bloch equations can be only applied to isolated spins which are independent magnetization vectors. A more complicated methods must be developed to describe and predict the behaviors of spins. This is where the quantum mechanical density matrix could shine.

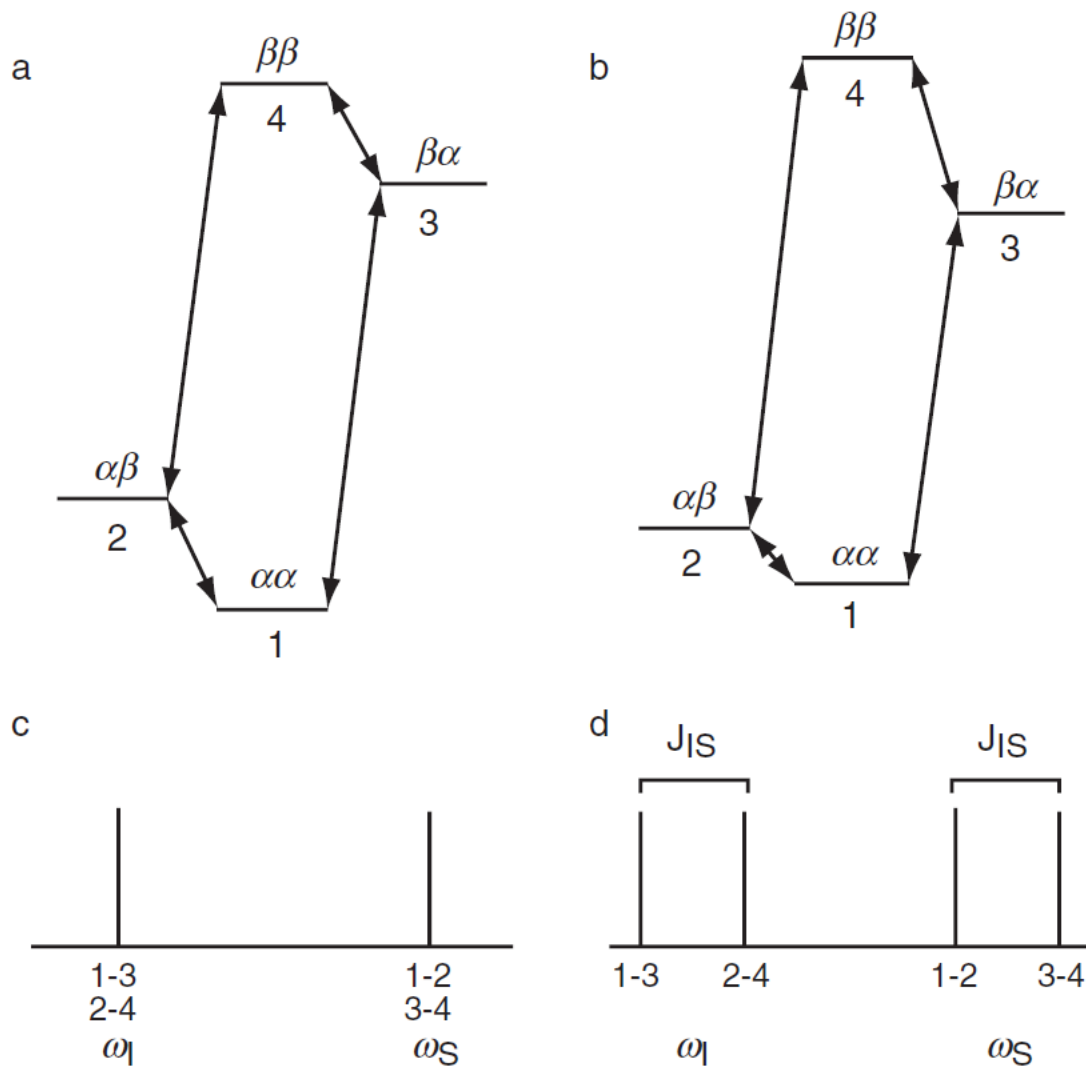


Figure 12 Energy levels of the two spin system (nuclei I and S) and the transition resonance frequency between states. Part a and c show the energy levels and transitions without scalar coupling. The transitions between energy levels 1-3 and 2-4 share the same resonance frequency ω_I while the transition between energy levels 1-2 and 3-4 share ω_S . Part b and d show that when scalar coupling applies, the energy levels on each state are modified by the J coupling constant J_{IS} , so the transition between those states are modified and centered on ω_I or ω_S but separated by $2\pi J_{IS}$.

Section 6: Density matrix formalism

Also the classical vector model really helps in visualizing how the bulk magnetization process, it is cannot deal with coupled spins. A more sophisticated system that not only follow the bulk magnetization, but rather a complete description of the evolution of the spin system at any time points is needed. That is the quantum mechanical density matrix formalism.

Quantum mechanically, the two spin states for a spin $\frac{1}{2}$ nucleus can be represented by two eigenkets

$$|\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\beta\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (78)$$

And of course the eigenbras would be as

$$\langle\alpha| = (1 \ 0), \quad \langle\beta| = (0 \ 1) \quad (79)$$

Any magnetization states ψ at any given time in the NMR experiment can be expressed in this vector space by the linear combination of these two basis kets with α and/or β coefficients as

$$|\Psi\rangle = c_\alpha|\alpha\rangle + c_\beta|\beta\rangle = \begin{pmatrix} c_\alpha \\ c_\beta \end{pmatrix} \quad (80)$$

$$\langle\Psi| = c_\alpha^*\langle\alpha| + c_\beta^*\langle\beta| = (c_\alpha^* \ c_\beta^*) \quad (81)$$

While the rotation operators can be represented by the Pauli spin matrices as

$$I_x = \frac{\hbar}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, I_y = \frac{\hbar}{2} \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, I_z = \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (82)$$

All these operators are Hermitian and satisfy the commutation relations as

$$[I_x, I_y] = iI_z, \quad [I_y, I_z] = iI_x, \quad [I_z, I_x] = iI_y \quad (83)$$

Now, the Schrödinger equation for a quantum mechanical system is given by

$$\frac{\partial \Psi(t)}{\partial t} = -\frac{i}{\hbar} \hat{H} \Psi(t) \quad (84)$$

Where the \hat{H} is the Hamiltonian of the system. This equation can be solved by separation of variables if the Hamiltonian is time-independent as

$$\psi \frac{d\varphi(t)}{dt} = -i\hat{H}\psi\varphi(t) \quad (85)$$

Where ψ is the time independent part of $\Psi(t)$. The solution of it can finally be written as

$$\Psi(t) = \psi e^{-i\omega t} \quad (86)$$

Where $\omega = E/\hbar$ and the energy of the system E is defined by

$$E = \int \psi^* \hat{H} \psi d\tau \quad (87)$$

Thus, the expectation values of the magnetic moments can be solved by using these equations.

For a spin $1/2$ magnetic spin, the wavefunction in the static magnetic field can be written as the linear combination of the two basis states as

$$\Psi = c_\alpha \psi_\alpha + c_\beta \psi_\beta = a e^{-i\omega_\alpha t} \psi_\alpha + b e^{-i\omega_\beta t} \psi_\beta \quad (88)$$

Again, the $\omega_\alpha = E_\alpha/\hbar$, $\omega_\beta = E_\beta/\hbar$ and $a = |a|e^{i\phi_a}$, $b = |b|e^{i\phi_b}$ are phase complex numbers which also are normalized so that $|a|^2 + |b|^2 = 1$. The expected values of the spin $1/2$ nucleus with $\mu = \gamma I$ on each dimension would be given as

$$\begin{aligned} \langle \mu_x \rangle &= \int \Psi^* \mu_x \Psi d\tau = \gamma \int \Psi^* I_x \Psi d\tau \\ &= \frac{\gamma \hbar}{2} \int (a e^{i\omega_\alpha t} \quad b e^{i\omega_\beta t}) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a e^{-i\omega_\alpha t} \\ b e^{-i\omega_\beta t} \end{pmatrix} d\tau \end{aligned} \quad (89)$$

$$\begin{aligned}
&= \frac{\gamma\hbar|a||b|}{2} \{e^{-i[(\omega_\alpha-\omega_\beta)t+\phi]} + e^{i[(\omega_\alpha-\omega_\beta)t+\phi]}\} = \gamma\hbar|a||b| \cos(\omega_0 + \phi) \\
\langle \mu_y \rangle &= \int \Psi^* \mu_y \Psi d\tau = \gamma \int \Psi^* I_y \Psi d\tau \\
&= \frac{\gamma\hbar}{2} \int (ae^{i\omega_\alpha t} \quad be^{i\omega_\beta t}) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} ae^{-i\omega_\alpha t} \\ be^{-i\omega_\beta t} \end{pmatrix} d\tau \quad (90) \\
&= \frac{\gamma\hbar|a||b|}{2i} \{e^{-i[(\omega_\alpha-\omega_\beta)t+\phi]} + e^{i[(\omega_\alpha-\omega_\beta)t+\phi]}\} = \gamma\hbar|a||b| \sin(\omega_0 + \phi)
\end{aligned}$$

$$\begin{aligned}
\langle \mu_z \rangle &= \int \Psi^* \mu_z \Psi d\tau = \gamma \int \Psi^* I_z \Psi d\tau \\
&= \frac{\gamma\hbar}{2} \int (ae^{i\omega_\alpha t} \quad be^{i\omega_\beta t}) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} ae^{-i\omega_\alpha t} \\ be^{-i\omega_\beta t} \end{pmatrix} d\tau \quad (91) \\
&= \frac{\gamma\hbar}{2} (|a|^2 - |b|^2)
\end{aligned}$$

where ω_0 is the Larmor frequency and ϕ is a phase angle. These equations demonstrate the probabilities of each states of the spin which is exactly the same as showed in the classical vector model where the bulk magnetization precesses around the external magnetic field.

If we defining the rotation operators as

$$\mathbf{R}_j(\alpha) = e^{-i\alpha I_j}, \quad (92)$$

then any rotation of the initial magnetization $\sigma(0)$ can be written as

$$\sigma(t) = \mathbf{R}_j(\alpha)\sigma(0)\mathbf{R}_j^{-1}(\alpha), \quad (93)$$

where j is the direction and it can be x, y or z. Expanding the rotation operator in x direction in Taylor series gives

$$\begin{aligned}
\mathbf{R}_x^{-1}(\alpha) &= \mathbf{E} + i\alpha I_x - \frac{1}{2}\alpha^2 I_x^2 + \dots \\
&= \mathbf{E} \left(1 - \frac{\alpha^2}{2!2^2} + \frac{\alpha^4}{4!2^4} + \dots \right) + 2iI_x \left(\frac{\alpha}{2} - \frac{\alpha^3}{3!2^3} + \frac{\alpha^5}{5!2^5} + \dots \right) \\
&= \mathbf{E} \cos\left(\frac{\alpha}{2}\right) + 2iI_x \sin\left(\frac{\alpha}{2}\right)
\end{aligned} \tag{94}$$

Substituting the matrix representation of I_x into equation this equation results the matrix representation of the pulse operators as follows:

$$\mathbf{R}_x^{-1}(\alpha) = \begin{pmatrix} c & is \\ is & c \end{pmatrix}, \quad \mathbf{R}_x(\alpha) = \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \tag{95}$$

$$\mathbf{R}_y^{-1}(\alpha) = \begin{pmatrix} c & s \\ -s & c \end{pmatrix}, \quad \mathbf{R}_y(\alpha) = \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \tag{96}$$

$$\mathbf{R}_z^{-1}(\alpha) = \begin{pmatrix} c + is & 0 \\ 0 & c - is \end{pmatrix}, \quad \mathbf{R}_z(\alpha) = \begin{pmatrix} c - is & 0 \\ 0 & c + is \end{pmatrix}. \tag{97}$$

Thus, any pulse with any flip angle can be derived using these pulse operators. For example, an x-phase 180° pulse applied to I_z can be calculated as

$$\begin{aligned}
\mathbf{R}_x(\pi)I_z\mathbf{R}_x^{-1}(\pi) &= \frac{1}{2} \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} c & is \\ is & c \end{pmatrix} = \frac{1}{2} \begin{pmatrix} c & is \\ -is & -c \end{pmatrix} \begin{pmatrix} c & is \\ is & c \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} c^2 - s^2 & 2ics \\ -2ics & s^2 - c^2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \cos \pi & i \sin \pi \\ -i \sin \pi & -\cos \pi \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= -\frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = -I_z
\end{aligned} \tag{98}$$

As we expected by classical vector model, this pulse rotates the magnetization from z direction to $-z$ direction and reverse the population in the two states. If the angle is changed to 90° which is $\pi/2$, the rotation will result

$$\mathbf{R}_x\left(\frac{\pi}{2}\right)I_z\mathbf{R}_x^{-1}\left(\frac{\pi}{2}\right) = \frac{1}{2}\begin{pmatrix} \cos\frac{\pi}{2} & i\sin\frac{\pi}{2} \\ -i\sin\frac{\pi}{2} & -\cos\frac{\pi}{2} \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} = -I_y. \quad (99)$$

It is consistent with the classical vector model prediction, where the x-phase 90° pulse will rotate the magnetization from + z direction to – y axis. Now, if the pulse is removed, the magnetization – I_y will evolve freely guided by the time-independent Hamiltonian

$$\hat{H} = (\omega_0 - \omega_{rf})I_z = \Omega I_z \quad (100)$$

Applying this Hamiltonian onto – I_y would give

$$-e^{-i\Omega t I_z} I_y e^{i\Omega t I_z} = I_x \sin(\Omega t) - I_y \cos(\Omega t), \quad (101)$$

which shows that the magnetization rotates from – y at t = 0 to + x at $\Omega t = \pi/2$, to + y at $\Omega t = \pi$ and to – x at $\Omega t = 3\pi/2$, same as predicted from classical vector model.

Density matrix formalism is particularly useful for multi-spin system, where the classical vector model is hard to use in order to illustrate the evolution of multiple spins. For example, a two spin system consist of I and S, the free-precession Hamiltonian is given by

$$\hat{H} = \omega_I I_z + \omega_S S_z + 2\pi J_{IS} \mathbf{I} \cdot \mathbf{S}, \quad (102)$$

where the J_{IS} is the scalar coupling constant. The strong coupling parameter θ is defined as

$$\tan(2\theta) = \frac{2\pi J_{IS}}{\omega_I - \omega_S} \quad (103)$$

With the defined strong scalar coupling parameter, the eigenfunctions can be found for a system of two coupled spins as

$$\Psi_1 = |\alpha\alpha\rangle, \quad \Psi_2 = \cos\theta|\alpha\beta\rangle + \sin\theta|\beta\alpha\rangle, \quad (104)$$

$$\Psi_3 = \cos \theta |\beta\alpha\rangle - \sin \theta |\alpha\beta\rangle, \quad \Psi_4 = |\beta\beta\rangle,$$

The transformation matrix \mathbf{U} that converts product basis to strong coupling basis is given by

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & \sin \theta & 0 \\ 0 & -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (105)$$

For weak coupling, where $\theta = 0^\circ$ and the wavefunctions of the two energy levels $|\alpha\beta\rangle$ and $|\beta\alpha\rangle$ are independent, the weak Hamiltonian can be simplified as

$$\hat{H} = \omega_I I_z + \omega_S S_z + 2\pi J_{IS} I_z S_z \quad (106)$$

And the transformation matrix \mathbf{U} that converts product basis to weak coupling basis is given by

$$U = \begin{pmatrix} c + is & 0 & 0 & 0 \\ 0 & c - is & 0 & 0 \\ 0 & 0 & c - is & 0 \\ 0 & 0 & 0 & c + is \end{pmatrix}, \quad (107)$$

where $c = \cos(\pi J_{IS} t/2)$ and $s = \sin(\pi J_{IS} t/2)$.

Now, for a 2 spin $1/2$ system I and S, the rotation operator for both spins can be derived by apply rotation operator on the individual spin and then multiply the result, for example, the rotation operator on the x axis applied on spin I can be written as:

$$\mathbf{R}_x(\alpha)[I] = \mathbf{R}_x(\alpha) \otimes \mathbf{E} = \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} c & 0 & -is & 0 \\ 0 & c & 0 & -is \\ -is & 0 & c & 0 \\ 0 & -is & 0 & c \end{pmatrix} \quad (108)$$

Apply the rotation operator on spin S gives

$$\mathbf{R}_x(\alpha)[S] = \mathbf{E} \otimes \mathbf{R}_x(\alpha) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} c & -is \\ -is & c \end{pmatrix} = \begin{pmatrix} c & -is & 0 & 0 \\ -is & c & 0 & 0 \\ 0 & 0 & c & -is \\ 0 & 0 & -is & c \end{pmatrix} \quad (109)$$

Now, multiply both could result the effect of the rotation operator on the weak coupling two spin system as

$$\mathbf{R}_x(\alpha) = \mathbf{R}_x(\alpha)[I] \mathbf{R}_x(\alpha)[S] = \begin{pmatrix} c^2 & icsu & icsv & -s^2 \\ icsu & 1 - s^2 u^2 & -s^2 uv & icsu \\ icsv & -s^2 uv & 1 - s^2 v^2 & icsv \\ -s^2 & icsu & icsv & c^2 \end{pmatrix}, \quad (110)$$

where $c = \cos(\alpha/2)$, $s = \sin(\alpha/2)$, $u = \cos\theta + \sin\theta$, $v = \cos\theta - \sin\theta$. This rotation operator for 2 spin system can be readily applied to any weakly coupled system. For example, an ideal 90° pulse with x-phase is applied to the two spin system. The initial state of the two spin system can be written as

$$\begin{aligned} \sigma(0) &= \omega_I I_z + \omega_S S_z = \omega_I I_z \otimes \mathbf{E} + \mathbf{E} \otimes \omega_S S_z \\ &= \omega_I \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \omega_S \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} \omega_I + \omega_S & 0 & 0 & 0 \\ 0 & \omega_I - \omega_S & 0 & 0 \\ 0 & 0 & -\omega_I + \omega_S & 0 \\ 0 & 0 & 0 & -\omega_I - \omega_S \end{pmatrix} \end{aligned} \quad (111)$$

Apply the 90° x-phase pulse can be shown as

$$\sigma(t) = \mathbf{R}_x\left(\frac{\pi}{2}\right) \sigma(0) \mathbf{R}_x^{-1} = \frac{1}{8} \begin{pmatrix} 1 & -i & -i & -1 \\ -i & 1 & -1 & -i \\ -i & -1 & 1 & -i \\ -1 & -i & -i & 1 \end{pmatrix} \times \quad (112)$$

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

As expected, an ideal 90° x-phase pulse rotate both magnetization into $-y$ axis. Now, taking into account of the coupling constant J_{IS} , executing the exponential operator $\exp[i(\Omega_I I_z + \Omega_S S_z + 2\pi J_{IS} I_z S_z)t]$ onto the magnetization after the flip would result

$$\frac{i}{2} \begin{pmatrix} 0 & 0 & 0 & -e^{-i(\Omega_I + \pi J_{IS})t} - e^{-i(\Omega_S + \pi J_{IS})t} \\ 0 & e^{i(\Omega_S + \pi J_{IS})t} & e^{i(\Omega_S + \pi J_{IS})t} & 0 \\ 0 & e^{i(\Omega_I + \pi J_{IS})t} & e^{i(\Omega_I + \pi J_{IS})t} & 0 \\ 0 & 0 & 0 & e^{i(\Omega_I - \pi J_{IS})t} + e^{i(\Omega_S - \pi J_{IS})t} \end{pmatrix} \quad (113)$$

The observed magnetization will be proportional to the trace of this matrix with four terms: $\exp[i(\Omega_I + \pi J_{IS})t]$, $\exp[i(\Omega_I - \pi J_{IS})t]$, $\exp[i(\Omega_S + \pi J_{IS})t]$, $\exp[i(\Omega_S - \pi J_{IS})t]$. So the final spectrum observed will consist of four resonance lines at frequencies $\Omega_I \pm \pi J_{IS}$ and $\Omega_S \pm \pi J_{IS}$, which has been shown before.

CHAPTER IV

NMR INSTRUMENTATION

The basic NMR spectrometer consists of three main parts: the magnet, the console and the user interface computer. Each of these parts consist of some subsystems that work together to form modern NMR spectrometer. A brief sketch of the NMR spectrometer is shown in Figure 13.

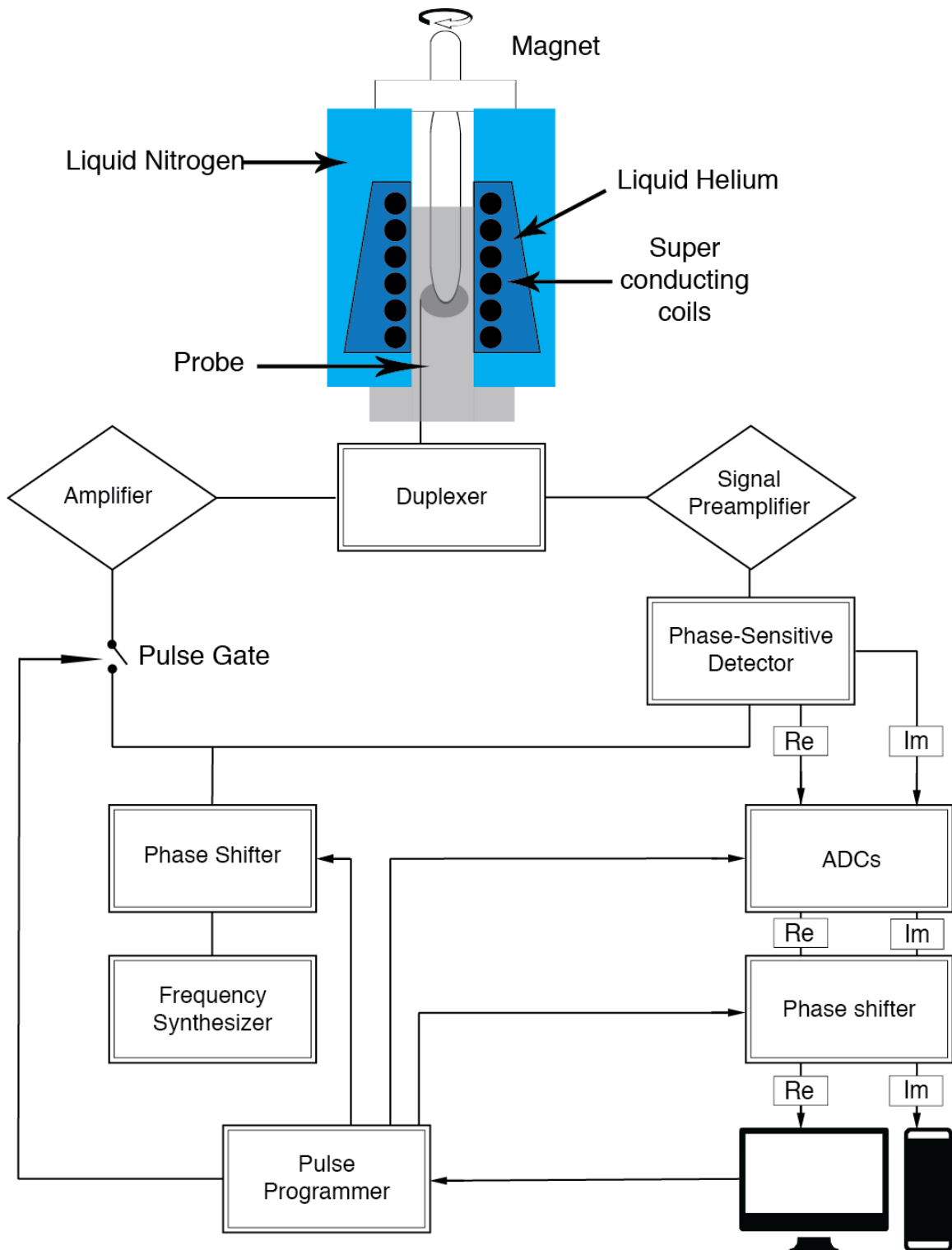


Figure 13 Block diagram of a single channel NMR spectrometer. The main components includes the magnet with the sample probe, the transmitter and the receiver. The detailed description can be found in this each sections.

Section 1: Magnet

According to the NMR theory, the signal is directly proportional to the gyromagnetic ratio of the atom as well as the external magnetic field strength. Since we cannot control the gyromagnetic ratio of the atom as it is the intrinsic characteristic of the atom, the only variable we can manipulate is the magnetic field strength. Over years of innovation, nowadays, almost all magnets inside the NMR spectrometer are made of superconducting solenoid. To date, the highest commercially available spectrometer has a magnetic field of 23.5 Tesla with proton Larmor frequency of 1000 MHz, which is 500,000 times of the magnetic field of the earth. To achieve this kind of superconductivity, the superconducting solenoid is immersed in liquid helium at 4.2 K which is surrounded by thermal radiation shield, a vacuum space, and then a dewar of liquid nitrogen to keep the minimum loss of liquid helium. The radiation shield which typically made of aluminum foil prevents the radiation of heat transfer. The vacuum prevents the conduction and convection of heat transfer. With such high insulation, the bore of the spectrometer, the shimming, excitation and detection coils as well as the sample can be placed at room temperature. Field homogeneity is very crucial to modern NMR spectrometers. When the superconducting coils is installed, the cryoshims will be adjusted to achieve a homogeneity field around the sample positions, typically less than 1 ppm inhomogeneity. After the superconducting coils starts to work, these cryoshims should not be changed. Upon daily uses, the room temperature shim coils, which are around the bore but outside the probe, should be adjusted to achieve better homogeneity with the sample inside the spectrometer bore. Present-generation magnets have field

homogeneity on the order of 1 ppb (part per billion). In addition to field homogeneity, stability of the magnet is also critical. It is affected by the inevitable decay of the magnetic field itself and external magnetic fields. In order to maintain the stability, a mini spectrometer called lock system was used parallel to the main spectrometer to monitor the locking frequency (typically deuterium). The lock system continuously measures the resonance frequency of the deuterium in the sample. If the resonance frequency of deuterium changes, the system will automatically correct it by changing the current in the shimming coil to maintain the lock frequency at the correct value.

Section 2: Transmitter

Radio frequency pulses that are used to excite spin states in NMR experiment are generated by transmitter. In order to achieve desired pulse width and frequency as long as correct phase and power level, the transmitter consist of several parts to execute these functions. The source of the radio frequency signal comes from the Frequency Synthesizer. It provides a stable continuous electromagnetic wave at desired frequency using a standard frequency reference. This source of continuous wave is gated by a transmitter controller that cut off the excess signals to create pulses. The transmitter controller also controls the phase, pulse power and gating switch. The amplitude is controlled by the computer controlled attenuator to desired level. The radio frequency pulse created by the gating is at low power. Before it is irradiated onto the sample, it will pass through the linear amplifier to obtain the pulse power that is required for a particular experiment. For the modern NMR spectrometer, each channel will equip one transmitter. For example, a triple-resonance spectrometer typically equips 3 transmitters for proton, carbon and nitrogen each in addition to the one used for locking. The power required on each channel is

directly proportional to their corresponding detecting nucleus gyromagnetic ratio. For proton, the power of the amplifier is typically at 50-100 W. For heteronucleus, the peak output power of the amplifier is on the order of hundreds of watts.

Section 3: Duplexer

The amplified radio frequency pulse coming from the transmitter will travel into the duplexer before it reaches the sample in the probe. The duplexer has two main functions. First, it block the strong rf pulse to get into the very sensitive signal detection circuitry. Second, when the NMR FID signal comes from the probe, it diverts it into the signal detection path rather than the path to the amplifier in the transmitter. This function is usually achieved by the arrangements of cables and diodes. But some duplexers use switches that can rapidly change between transmit mode and receive mode controlled by the pulse programmer.

Section 4: Probe

The probe is the most specialized part of the NMR spectrometer and is often the only part that needs to be changed from one experiment to another. It is a complex apparatus that holds the sample and is positioned coaxially in the bore of the magnet. Its main function is to place the sample into the homogeneous region of the magnetic field. The probe also equipped with a device that can rotate the sample in order to reduce the inhomogeneity of the sample so that narrower spectra can be achieved. The VT (variable temperature) control apparatus inside of the probe can also stabilize the temperature of the sample. In the probe, there are rf electronic circuits that serve as both the irradiation source of the rf waves and the detecting instruments for sub sequential rf emissions from the sample. Its design strongly affects the sensitivity of the spectrometer, the

homogeneity of the B_1 rf field, the susceptibility to rf heating of the sample and the quality of the solvent suppression. For the simplest case, the rf circuit consists of a coil and two capacitors as shown in Figure 14.

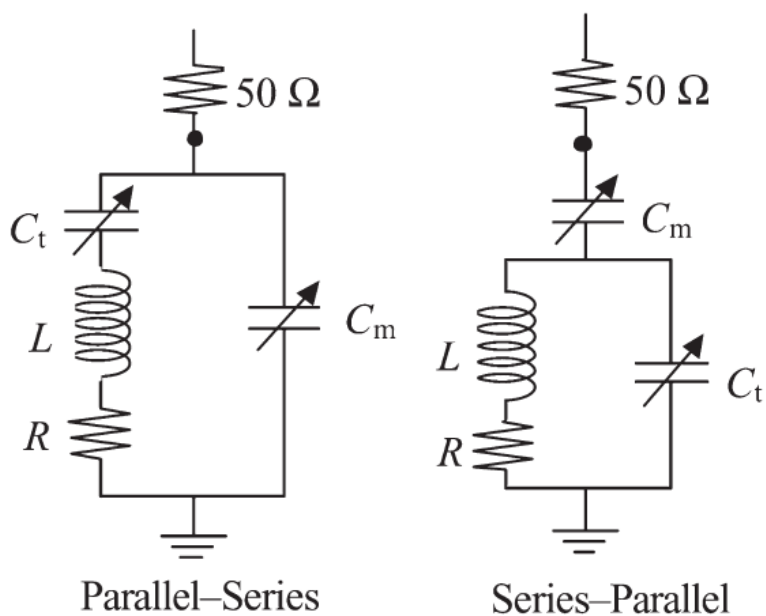


Figure 14 Schematic show of two types of probe coils: Parallel-series and Series-parallel LC circuits. Two capacitors, labeled as C_t and C_m , are adjustable for proper tuning and matching. The 50Ω is the impedance of the circuit.

The function of the coil as mentioned before, is twofold. First, during the pulse, the current coming from transmitter generates an oscillating magnetic field B_1 to the sample. Second, when the pulse terminated, the coil detects the current created by the precession of the sample and send it to the duplexer and receiver. The coil is placed in such a way that the generated magnetic field is predominantly perpendicular to the main filed to create maximum flip angle during the pulse. In theory, two orthogonal coils detecting x-magnetization and y-magnetization should be used. However, it is hard to tune two orthogonal coils to the same frequency. Modern probes usually

recruit a single Helmholtz coil to achieve quadrature detection. The two capacitors in the circuit are labeled as matching capacitor and tuning capacitor. The matching capacitor couples the external signals into the probe circuit with maximum efficiency. The tuning capacitor which is parallel to the coil is to enhance the currents in the coil by electromagnetic resonance. These two capacitors are not independent, in a real experiment, both capacitors should be adjusted iteratively to optimize the NMR probe. The characteristics of the probe are given by two factors, quality factor Q and complex impedance Z , which are given by

$$Q = \omega L/R \quad (114)$$

$$Z = R + i[\omega L - 1/(\omega C)]. \quad (115)$$

The ω is the resonance frequency, L is the inductance of the coil, R is the resistance of the coil and C is the capacitance of the tuning capacitor.

The signal to noise ratio (S/N) is a very important factor to be considered in designing an NMR probe. The factors contributing to the S/N can be shown as

$$S/N \propto \frac{N\gamma_e\gamma_d^{3/2}B_0^{3/2}K}{\sqrt{\Delta f[T_c R_c + T_a(R_c + R_s) + T_s R_s]}} \quad (116)$$

The N is the total number of observed nuclei in the sample, γ_e and γ_d are the gyromagnetic ratios of the excited and detected nuclei, B_0 is the external magnetic field strength, K is a factor related to coil design, Δf is the receiver bandwidth, T is temperature, R is resistance and the subscription of a, c, s represent amplifier, coil and sample. According to this equation, lowering the temperature of the coil and amplifier could potentially increase the signal to noise ratio. This is the idea that cryogenic probes are designed to operate coil at ~ 20 K to enhance the detection.

Although different nuclei resonant at quite different frequencies, it is technically feasible to share

the same coil between separate tuned circuits, each operating on a different frequency. This is called multiple tuning. For example, for a triple-resonance ^1H - ^{13}C - ^{15}N probe, it only contains two coils. One coil is double-tuned for ^1H and ^2H while the other is double-tuned to ^{13}C and ^{15}N .

Section 5: Receiver

The receiver consist of a preamplifier, a phase-sensitive detector and analog-to-digital converter. Immediately after the NMR signal is collected by the coils in the probe, the current will be passed through the duplexer first. The duplexer diverts the weak signal toward the preamplifier in the receiver to be scaled up to a more convenient voltage level. The noise figure of the preamplifier is very critical for the NMR spectrometer since it provides the initial stage of amplification of the signal. Later parts will unavoidably process the real NMR signal as well as the noises coming from the preamplifier. To minimize losses, the preamplifier is located as close to the probe as practical. In order to interpret the NMR data, the electromagnetic signal has to be converted to computerized digital form. This process in done by the analog-to-digital converter. However, the problem for NMR signal is it typically is in the range of several hundreds of megahertz which is too fast for current ADCs. It is very important to down-convert these frequencies to a more feasible level to the ADCs. The down-conversion is achieved by the phase-sensitive detector. It compares the NMR signal with a reference signal coming directly from the rf synthesizer. It detect the offset of the signal which is the difference between the NMR signal around Larmor frequency and the reference signal. This offset signal is less than 1 MHz (typically in kHz range) which is slow enough to be handled by the ADC. The detector also include audio filters designed to restrict the frequency bandwidth of the receiver to minimize the noise. The offset signal could be either positive or negative and it is confusing if only one channel is used in the detector. This

is why a phase-sensitive detector is required to detect the NMR signal to achieve quadrature detection. The signal is first split into two by a splitter. Each of the signal is mixed with the reference signal to achieve the offset audio frequency in the mixer. The two reference signals in the two mixer are 90° out of phase so that the resulting signals coming out of the mixers are one cosine modulated signal at offset frequency while the other is sine modulated signal at the same frequency. Combining the signals coming from both mixers could tell the differences between positive and negative offset since the signals that are 90° out of phase will be canceled out in the sine modulated signals but doubled in cosine modulated signals as shown in Figure 15.

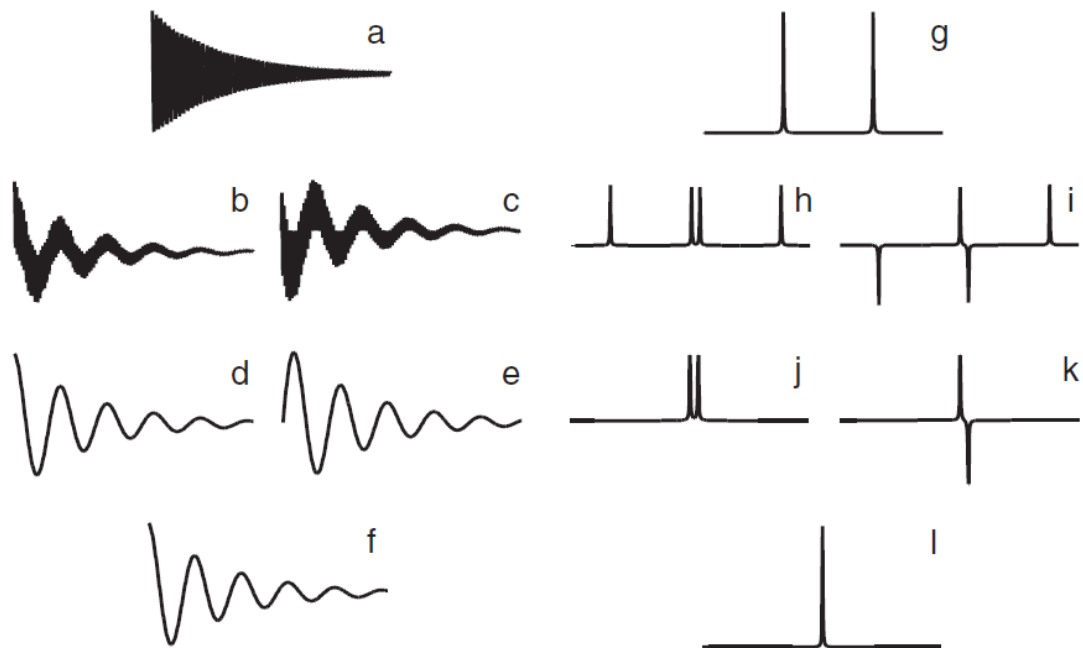


Figure 15 The quadrature detection. The time domain signals are shown on the left panel and the Fourier-transformed frequency domain signals are shown on the right panel. For simplification, the signal coming from a single coil is shown as cosine modulated $\cos(\omega_0 t)$ in (a). After Fourier transformation, two peaks will arise in the frequency domain at $\pm\omega_0$. In order to find the correct resonance frequency, the signal is split into two. The first will be mixed with a $\cos(\omega_{ref} t)$ modulated reference signal as shown in (b). The other is mixed with a $\sin(\omega_{ref} t)$ modulated reference signal as shown in (c). Their corresponding frequency domain spectra are shown in (h) and (i) which resonant at $\pm(\omega_0 - \omega_{ref})$ and $\pm(\omega_0 + \omega_{ref})$. The audio frequency filters will remove high frequency signals that are too far away from the Larmor

frequency. So the time domain signals after the filter are shown as in (d) and (e) and the corresponding frequency domain signals are shown in (j) and (k). In both, only the $\pm(\omega_0 - \omega_{ref})$ signal still remain. The signals from the two channels are now combined to give the correct resonance as shown in (f) and (l).

So the NMR signal is in complex form with both real and imaginary part. After the phase-sensitive detector, each part of the signal is connected to their own ADC. The ADC measures the voltage level of the input signal and convert it into a binary codes. This process is called digitization. The time separation between the sampling points of the ADC is called the sampling interval or dwell time. The inverse of the dwell time is the sampling bandwidth or spectral width. Depending on the experimental requirements, difference ADCs can be used. If a wide range of frequencies is required as typically do in solid-state NMR, the more rapid sampling is needed so comparing to solution NMR. The ADC used in solid state NMR is typically around 4 MHz while in solution NMR around 250 kHz. The resolution of the ADC depends on the bit size of the ADC. However, sampling speed will be reduced if a higher resolution is needed. As a compromise, a 16-bit ADC is typically used in modern NMR spectrometer and it can represent numbers between -2^8 and 2^8-1 (-32,768 to 32,767). The magnitude of the analog signal must stay within the dynamic range of the ADC otherwise it is either not accurately read (too high) or being ignored (too small less than $\frac{1}{2}$ bit).

In many experiments, the phases of the rf pulses and the NMR signals are changed in a dynamic way. This allows one type of NMR signals to be distinguished from experimental artifacts or from different types of NMR signals. So a post-digitization phase shifter can be applied after the ADC. The digitized complex signal is passed into the phase shifter and multiplied by a complex phase factor before it is passed to the computer. This phase shift after the quadrature detection is immune to instrumental imperfections such as misbalance of the two sections of the quadrature detector.

Section 6: Data Processing

Direct Fourier transformation of raw recorded NMR data rarely yield optimal frequency domain spectrum. Typically, some manipulations will be done before or after the Fourier transformation to maximize the information that can be withdrawn from the spectrum.

Zero-filling is a technique that adding a series of zeroes into a data sequence prior to Fourier transformation. Fast Fourier transformation require the data points to be an integral power of 2. So, for a set of data points that is not an integral power of 2, zeroes will be automatically added to make up to the closest number that is an integral number power of 2. Beyond this, since the NMR signal comprises of both real and imaginary parts and the imaginary part is usually discard in multi dimension NMR spectroscopy, adding zeroes could compensate the loss of imaginary part as the imaginary part can be recovered from the real part. So if you have a set of 2^n data points, adding another set of 2^n zeroes could increase the spectral resolution. Additional zero fillings cannot gain additional information beyond this point.

Linear prediction is another way to increase the resolution of the spectrum. This technique is very useful for long T2 relaxation atoms where the FID is truncated before fully decayed out. It is especially useful in multi-dimension NMR where the indirect domain typically don't have enough time to completely decay. It is very important to use linear prediction for the indirect dimension to avoid sinc wiggles created by the truncated FID.

Apodization is another very important technique to get a smooth satisfactory lineshape in the spectrum. The Apodization process multiply a filter function in the time domain before Fourier transformation. The time domain signal needs to be reduced to zero smoothly and a matched

function has to be applied to the FID prior Fourier transformation in order to achieve maximum signal-to-noise ratio. The matched function typically should have the same decay envelop as the FID signal. However, apodization results in more line broadening. There has to be a compromise between the apodization and line width. It is worth to notice that apodization does not increase the peak area under the Fourier transformed spectra.

Phasing is crucial in NMR data processing. The digitized signal usually displays a frequency dependent phase error. This phase error can be corrected by applying a zero-order phase correction function and first-order phase correction function. Zero-order phase correction is frequency independent while the first-order phase correction contributes to a linearly frequency-dependent phase. By adjusting both phase correction functions, a nicely absorption lineshape can be achieved throughout the whole spectra. Another phase error usually comes from a clipped FID where the first several data points exceed the dynamic range of the ADC. This can cause a baseline distortion in the spectrum called baseline roll. By adjusting the time between the observed pulse and the start of sampling, the distortion can be reduced. The most famous pulse sequence Hahn echo is usually applied in the ^1H NMR spectroscopy to reduce baseline distortion. Linear prediction of the first several data points can also be applied to correct baseline distortion.

CHAPTER V

THE LIPID DROPLET PROTEIN PLIN1

Section 1 Abstract

Neutral lipid triglycerides, a main reserve for fat and energy, are stored in organelles called lipid droplets. The storage and release of triglycerides are actively regulated by several proteins specific to the droplet surface, one of which in insects is PLIN1. PLIN1 plays a key role in the activation of triglyceride hydrolysis upon phosphorylation. However, the structure of PLIN1 and its relation to functions remain elusive due to its insolubility and crystallization difficulty. Here we report the first solid-state NMR study on the *Drosophila melanogaster* PLIN1 in combination with molecular dynamics simulation to show the structural basis for its lipid droplet attachment. NMR spin diffusion experiments were consistent with the predicted membrane attachment motif of PLIN1. The data indicated that PLIN1 has close contact with the terminal methyl groups of the phospholipid acyl chains. Structure models for the membrane attachment motif were generated based on hydrophobicity analysis and NMR membrane insertion depth information. Simulated NMR spectra from a trans-model agreed with experimental spectra. In this model, lipids from the bottom leaflet were very close to the surface in the region enclosed by membrane attachment motif. This may imply that in real lipid droplet, triglyceride molecules might be brought close to the surface by the same mechanism, ready to leave the droplet in the event of lipolysis. Juxtaposition of triglyceride lipase structure to the trans-model suggested a possible interaction of

a conserved segment with the lipase by electrostatic interactions, opening the lipase lid to expose the catalytic center. To verify this binding model, the full-length protein was chopped into four segments without overlapping based on the bioinformatics analysis to study their individual binding affinity to the liposome by reconstitution. It was found that the 4 helices domain can be reconstituted into the liposome very efficiently, which agrees with the bioinformatics analysis and confirmed its lipid binding function. Interestingly, we found the PAT domain could also be reconstituted into the liposome, indicating its potential hydrophobic nature that could either be another lipid interaction domain or interact with the hydrophobic domain with the lipase.

Section 2 Introduction

Animals store most of excess energy in the form of neutral lipid triglycerides for later use as metabolic fuel. The hydrophobicity of triglycerides allows them to be densely packed into lipid droplets, providing an energy density 10 times that of hydrated proteins and carbohydrates [32]. The lipid droplets are composed of a triglyceride core surrounded by a monolayer of phospholipids and a variety of proteins [33] and can be readily visualized by light and electron microscopy [34-36]. The structure of lipid droplets is similar to that of lipoproteins but they still have very distinct features. Firstly, lipid droplets can be synthesized by all types of cells but lipoprotein synthesis are restricted to hepatocytes and enterocytes [37]. Secondly, lipid droplets can be considered as cell organelles as they mainly store lipids in place while lipoproteins are the intercellular carriers that transport lipid building blocks from one place to another. At last, lipoprotein only contains very limited types of proteins which enable them to execute the transportation function and reach target with limited interactions with the environment during the process. The proteome analysis of lipid droplet surface binding proteins reveals up to hundreds of

different proteins that has different functions in cell skeleton interaction, lipolysis and other metabolism processes [38-50]. More evidences [51-57] show that lipid droplet has direct contact with various organelles in the cell such as ER, mitochondria, peroxisomes, endosomes and other LDs. It is reported that LD originates from ER through a process called budding, where the DGAT on the ER surface synthesizes triglycerides into the intramembrane space between the two layers of ER membrane, resulting the outer membrane to bud out with the triglycerides inside [58, 59]. Utilization of the stored triglycerides requires enzymatic breakdown (lipolysis) by lipases, while the surface layer of the droplet controls the accessibility of lipases to the stored triglycerides. Among the proteins surrounding the lipid droplet surface, proteins in the PAT family (named after three earliest members) have raised great interest in recent years. The PAT family consists of the mammalian Perilipin, ADRP, TIP47, S3-12, and OXPAT, as well as insect lipid storage droplet protein 1 (Lsd1) and 2 (Lsd2) [60]. Perilipin, ADRP, and Lsd1 constitutively attach to the lipid droplets, and they maintain fat storage and regulation of lipolysis. TIP47, S3-12, OXPAT, and arguably Lsd2 bind reversibly to the droplets; hypothetically they are responsible for the packaging of newly synthesized triglycerides into lipid droplets [60]. A new nomenclature has been recently proposed for the mammalian PAT-family of proteins [61] as Perilipin to PLIN1, ADRP to PLIN2, TIP47 to PLIN3, S3-12 to PLIN4 and OXPAT to PLIN5. Accordingly, from now on we will refer Lsd1 and Lsd2 in *Drosophila* as PLIN1 and PLIN2, respectively. Mammals and insects share significant conservation in the molecular mechanism of lipid droplet metabolism, highlighting the tremendous potential of using genetic technical advantages of insects to discover novel features of lipid homeostasis [60, 62]. Studies of fruit fly models have established a correlation between triglyceride accumulation and the level of PLIN2 expression [63]. It is shown that the protein composition is related to the size of the lipid droplet.

For small lipid droplets in *Drosophila*, both PLIN1 and PLIN2 can be found but only PLIN1 shows up on large lipid droplets. PLIN1 is found exclusively associated with lipid droplets [64]. It dynamically interacts with lipid droplet to control access of lipase to triglycerides thus regulates the lipids homeostasis. In mammals, PLIN1 block the basal lipolysis but activates hormone sensitive lipase upon stimulation. In contrast to mouse perilipin, which protects triglycerides from hydrolysis [65, 66], depletion of *Drosophila* PLIN1 leads to adult-onset obesity [67] while overexpression of PLIN1 induces lipid droplet to shrink and aggregate [68]. PLIN1 serves as a lipolytic switch, which upon protein kinase A (PKA) mediated phosphorylation, promotes the activation of triglyceride lipolysis [69].

The association of these proteins on the surface of the lipid droplets is critical to their ability to properly regulate both storage and release of the triglycerides in the droplets. Despite the pressing need to understand the interaction between these proteins and the lipid droplets, progress has been hampered by the scarcity of three-dimensional structure information for these proteins. The first and so far the only structure determined was for the C-terminal TIP47 (residues 191-437) at 2.8 Å resolution using X-ray crystallography [70]. The structure consists of an α/β domain and a four-helix bundle that resembles the receptor-binding domain of apolipoprotein E. The deep hydrophobic cleft between the α/β domain and the four-helix bundle is consistent with binding to hydrophobic proteins and small molecules, rather than to the extended phospholipid membrane. This C-terminus construct was selected from one dozen truncations for soluble protein expression. Regrettably, it does not have the N-terminal 11-mer helical repeats that are probably responsible for reversible binding to lipid membranes.

Meanwhile, in vitro systems have recently become available for structural and functional studies. Recombinant PLIN1 has been purified and reconstituted in lipid droplet-like particles [71, 72]. Using an in vitro system, it was shown that phosphorylation of PLIN1 enhances the triglyceride lipase activity, demonstrating the direct connection between PLIN1 phosphorylation and the activation of lipolysis [71]. Hypothetically, PLIN1 phosphorylation causes changes on the droplet surface, making the internal triglycerides more accessible to the lipase [69]. The ability to reconstitute PLIN1 in lipoprotein particles opens the possibility to design structural studies to advance our understanding of the mechanisms of lipolysis regulation. Nevertheless, these lipoprotein particles are too large (~20 nm diameter) for solution NMR studies, and they are very difficult if not impossible to form diffraction quality single crystals for crystallography studies. Fortunately, several other structural techniques could be applied to this type of samples. For example, topologies of lipoprotein complexes have been determined using solid-state NMR [73, 74], fluorescence spectroscopy [75, 76] and electron paramagnetic resonance (EPR) [77]. Among these, solid-state NMR is particularly suitable to study these lipoprotein complexes because three-dimensional structure details could be obtained [78-81]. Here we report both NMR experimental data and structural models that could be useful to advance our understanding of protein targeting to lipid droplet and the function of PLIN1.

Section 3 Materials and Methods

Isotopically enriched (^{13}C , ^{15}N) ISOGRO, $^{15}\text{NH}_4\text{Cl}$ and uniformly labeled ^{13}C -glucose, were purchased from Sigma-Aldrich (St. Louis, MO). Lipid 1,2-dimyristoyl-sn-glycero-3-phosphoglycerol (DMPG) was purchased from Avanti Polar Lipids (Alabaster, AL). Benzonase was purchased from EMD Millipore (Billerica, MA). His-beads were purchased from Bio-Rad

(Hercules, CA). Lysozyme was purchased from MP Biomedicals (Santa Ana, CA). Thrombin was purchase from Bio Pharm Laboratories (Alpine, UT). All restriction cleavage enzymes and DNA polymerase as well as ligase were purchased from New England Biolabs (Ipswich, MA).

3.1 Protein expression and purification

The over-expression of PLIN1 (CG10374, NP_732904.2) as a fusion protein with thioredoxin-[His]₆-S tag was carried out as previously reported [71] with slight modifications to incorporate stable isotopes for NMR studies. In brief, *Drosophila* total mRNA was reverse transcribed using oligo-d(T)18-primer and the resulted cDNA was used as template to amplify the coding region of Lsd1 (CG10374, NP_732904.2) by PCR. The left and right primers were 5' - GACGACGACAAGATGGCAACTGCAACCAGCGGCAGTGGA and 5' - GAGGAGAAGCCCGGTCTAGACGCCGTTGATGTTATTGTG. The product was ligated into the vector pET-32 Ek/LIC that contains an N-terminal coding sequence for thioredoxin followed by His-Tag and S-Tag coding sequences. *E. coli* strain NovaBlue GigaSingles cells were transformed with the recombinant plasmid. Positive clones were confirmed by DNA sequencing. Plasmids extracted from these clones were transformed into *E. coli* Rosetta 2 cells for protein expression. Recombinant protein expression was induced with 1 mM IPTG when OD 600 reaches 0.8. After 4 hours induction, bacteria were collected and resuspended in buffer (50 mM Tris, pH 8, 1 mM EDTA, 100 mM NaCl, 1 mM PMSF) containing 0.3 mg/ml of lysozyme. After 30 min incubation, the preparation was centrifuged at 100,000g for 1 h. The fusion protein expressed as inclusion bodies as they can be found in the pellet. The inclusion bodies were resuspended in 0.1 M sodium acetate pH 5, 5 mM MgCl₂ and incubated with DNase I at 37 °C for 1 h. Sample was centrifuged (35,000g, 20 min) and the pellet was resuspended in 20 mM Tris, pH 7.5, 6 M urea,

0.3% octylglucoside, 0.5 M NaCl and vortexed at 40 °C for 15 min. After centrifugation (5000g, 20 min) the procedure was repeated twice. The final pellet was resuspended in 20 mM Tris, pH 8, 8 M urea, 10% Sarkosyl, 0.5 M NaCl, 1 mM DTT and sonicated in two steps for 1 min each time. The preparation was centrifuged (5000g, 20 min) and the pH adjusted to 7.5. Trx-Lsd1 stock solution (5 mg/ml) was kept in the freezer. Before usage the preparation was diluted to 0.5 mg/ml and dialyzed to a final concentration of 20 mM Tris, pH 7.5, 2 M urea, 0.10 M NaCl, 0.01% cholate. For ¹³C-¹⁵N double labeled protein expression, transformed *E. coli* Rosetta cells with the recombinant plasmid (pET32-CG10374) were grown in 200 mL Luria broth medium at 37 °C until optical density 0.8 at 600 nm. The bacteria pellet was collected by centrifugation, and cultured in 1 liter M9 minimal medium containing reagents enriched with NMR-active stable isotopes (¹³C-glucose and ¹⁵NH₄Cl). The medium was supplemented with properly labeled algae extracts (ISOGRO) to boost protein yield. When optical density reached 0.8, protein expression was induced by addition of 1 mM IPTG. After 6 hours, cells were harvested by centrifugation. Thioredoxin-Lsd1 fusion protein was purified essentially as previously described [71]. The final protein pellet was resuspended in 20 mM Tris, pH 8.0, 6 M Urea, 150 mM NaCl, 10 mM dithiothreitol and a solution of protein stock (1.7 mg/mL) was stored in the freezer.

3.2 Protein reconstitution

DPMG lipid powder was dissolved in organic solution Chloroform first. The solution was agitated in a round-bottom flask and let dry under N₂ gas flow at 37 °C dry bath (phase transition temperature for DMPG is 23°C). The coated DMPG film at the bottom of the flask were kept under N₂ gas flow for another 30 min before vacuum overnight. The thin DMPG film was dissolved in 20 mM Tris, pH 7.5, 2 M urea, 0.10 M NaCl to a target concentration of 6 mg/ml.

The dissolved lipids was vortexed vigorously for 30 s at RT and then applied to sonication at 30 °C for 1 hour. DMPG liposomes (400 nmol) were mixed with 8 nmol of fusion protein (Trx–Lsd1) in 20 mM Tris, pH 7.5, 2 M urea, 0.10 M NaCl, 0.01% cholate followed by exhaustive dialysis for 24 h against phosphate buffer (5 mM Na₃PO₄ pH 7.5, 20 mM NaCl) at 4 °C. After exhaustive dialysis, thioredoxin-PLIN1/DMPG complexes were brought to 60% (w/v) sucrose and subjected to ultracentrifugation in a sucrose density gradient (30 to 60 % (w/v)). The distinct white band floating at a density of 1.17g/ml was collected as a single fraction. Complexes were sedimented by ultracentrifugation in an aqueous buffer and resuspended in a buffer containing 5 mM Na₂HPO₄, 0.15 M NaCl, 0.1% octylglucoside at pH 7.4 and incubated with thrombin (1 unit/mg of protein) for 15 h at 4 °C to cleave the thioredoxin-[His]₆ tag. The cleaved tags were removed by passing the cleaved solution through a 10 kD cutoff membrane filter. After centrifugation, the pellet containing PLIN1/DMPG complex was washed with 5 mM phosphate buffer (pH 7.5) and excess of water was removed in the speed vac for 1 h. Based on proton NMR signal intensities, the sample contained about 35% wt of water. These complexes were previously reported to have apparent diameter of 20 nm [71], and they are likely small unilamellar vesicles (SUVs) [82]. The head groups of anionic lipids, such as DMPG, facilitate interaction with positively charged protein sidechain groups and they are important for membrane attachment of the protein. Neutral lipid DMPC was also tested, but it resulted in dramatic protein loss by sticking to the dialysis membranes and the centrifuge tubes. DMPG provided a cost effective way to obtain isotopically enriched PLIN1 lipoprotein complexes.

3.3 NMR spectroscopy

All Solid-state NMR experiments were carried out on a 600 MHz Varian INOVA spectrometer and a triple resonance magic-angle spinning (MAS) probe with a 1.6 mm spin module. All spectra were acquired with a MAS rate at 13.3 kHz. For ^{13}C 1D and ^{13}C - ^{13}C 2D experiments, the proton 90° pulse was 2.2 μs , cross polarization (CP) contact time 0.7 ms, locking fields of 73 kHz on ^1H and 80 kHz on ^{13}C channels, 100 kHz two pulse phase modulation (TPPM) decoupling [83], and dipolar-assisted rotational resonance (DARR) recoupling [78, 84]. For proton spin diffusion experiment, the proton-detected NHH pulse sequence was modified from the CHH sequence [79] with MISSISSIPPI solvent suppression [80] and an additional T2 filter (300 μs) [81] to suppress signals from the less mobile protein molecule as shown in Figure 16. The CP locking fields were 73 and 60 kHz on ^1H and ^{15}N channels, respectively. Contact time for the first CP (^1H to ^{15}N) was 1 ms, and 0.6 ms for the second CP (^{15}N to ^1H).

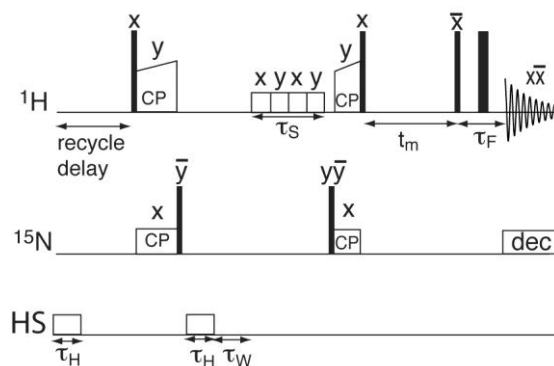


Figure 16 NHH spin diffusion pulse sequence modified from Luo et al. [79]. Proton polarization excited by the initial 90° pulse is transferred to ^{15}N by cross polarization and stored in the longitudinal direction by the first ^{15}N 90° pulse. The remaining proton signals are suppressed by MISSISSIPPI, which employs homospoil gradients and saturation pulses; $\tau_S=200$ μs , $\tau_H=20$, $\tau_W=50$ μs [80]. The polarization is brought back to the transverse plane by the second ^{15}N 90° pulse and transferred back to ^1H . Polarization propagates from protein protons to solvent (water and lipids) via spin-diffusion (t_m). A T_2 -filter (τ_F) is added to remove protein signals [81], leaving only polarization solvent protons to be detected. The π pulse is placed in the middle of τ_F . Narrow and wide filled rectangles represent $\pi/2$ and π pulses, respectively. This sequence can be used as CHH when the ^{15}N channel is replaced by ^{13}C .

3.4 MD simulation and spectral simulation

All MD simulations were performed on up to 600 processors on a linux cluster supercomputer using software GROMACS 4.5.5 and GROMOS96 54A7 force field combined with DMPG lipid interaction parameters with simple point charge-extended (SPCE) water model [85]. The non-bonded van der Waals interactions were estimated using Lennard-Jones potential with cutoff value of 1.2 nm and the bonds were constrained by linear constraint solver (LINCS) algorithm [86]. Electrostatic forces and energies were calculated using Particle-Mesh Ewald (PME) summation algorithm with cutoff value of 1.2 nm too [87].

Two equilibration phases, constant volume (NVT) and constant pressure (NPT) ensembles, were subsequently carried out, each with 1 fs time steps. In the first phase, the system was coupled to a strong temperature bath using V-rescale coupling [88] with temperature coupling constant of $\tau_T = 0.1$ ps to maintain system temperature at 300 K. In the second phase, Parrinello-Rahman pressure coupling [89] with coupling constant $\tau_P = 5.0$ ps to maintain the pressure semi-isotropically at 1 bar and a weak Nose-Hoover temperature coupling with a coupling constant $\tau_T = 0.5$ ps [90-92] was used to ensure a true NPT ensemble. A total of 15 ns equilibration was followed by a 50 ns production run of molecular dynamics (MD) in 2 fs step size, during which temperature and pressure were maintained using weak coupling methods (Nose-Hoover with $\tau_T = 0.5$ ps and Parrinello-Rahman with $\tau_P = 2$ ps).

For a given structure model, backbone and α -carbon chemical shifts were predicted by shiftX [93]. Based on these chemical shifts, ^{13}C - ^{13}C 2D spectra were simulated by program peaks2ucsf

in the Sparky package (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco) with assistance of a custom computer script.

3.5 Trypsin protection assay and Mass spectrometry

Trypsin was applied to the liposome with reconstituted protein in it to identify the binding domain of PLIN1 with the membrane. Trypsin was prepared by dissolving the powder to 1 mM HCl and stored at -20°C. The reconstituted protein on the liposome was incubated with trypsin at a ratio of 1 unit/mg protein at 37 °C overnight. The control experiment was done using buffer-exchanged protein in 2M urea, 0.05% cholate with trypsin at the same enzyme to protein ratio. The cleaved sample was run on SDS-PAGE first and the protein bands are excised for mass spectrometry. The sample was reduced, alkylated, and using another trypsin digestion before applied to HPLC-Mass spectrometry. Peptide mass fingerprints were used for database search using Mascot [94].

3.6 Truncated PLIN1 mutant clone and reconstitution

PLIN1 was truncated into 4 different pieces according to bioinformatics analysis without overlapping to study their individual lipid bilayer abilities. The full-length protein was separated into 4 different domains: the N-terminal PAT domain, the 4-Helix domain which shows lipid interactions, the R1 domain which is the segment in between of the first two, and the C-terminal domain R2. Since we don't have empty pET-32 Ek/LIC in hand, the recombinant pET-32-FL-

PLIN1 was used as both the PCR template and the final insertion plasmid. Primers designed to clone each domain are shown in Table 3. Since the full-length protein use LIC system as the recombination method, it does not disrupt the restriction cleavage sites on the plasmid. Double restriction cleavage was applied for cloning all the four segments into the vector. The vector with full-length PLIN1 was extracted from *E. coli* with Zyppy™ plasmid miniprep kit according to instructions. The extracted plasmid was used as template for the PCR amplification. The PCR was conducted in a 5 min 95 °C preparation followed by 45 cycles of 30 s 95 °C denature , 30 s 45 °C annealing, 60 s 60 °C extension. The PCR product and the vector with full-length PLIN1 were digested using Hind III and Kpn I as restriction cleavage enzymes, respectively. The ligation was performed with inserts: plasmids at the molar ratio of 3:1 at 37 °C for 2 hours. The ligated plasmid was transformed to *E. coli* DH5α for amplification. The extracted plasmids were later transformed to *E. coli* BL21 (DE3) for optimal protein expression. For soluble proteins, cell lysate was filtered against 0.2 μm filter to remove any remaining cell debris followed by passing through the Ni-NTA column (BioRad, Hercules, CA). For insoluble protein, cells were lysed followed by 36,000 g centrifugation. The pellet was tested on different buffers with 20 mM Tris, 50 mM NaCl buffers at pH 8 with different additions in the following order: 1) 1% Triton X-100, 2) 2M Urea, 3) 1% Triton X-100 + 2M urea, 4) 5% Sarkosyl, 5) 8M urea + 10% Sarkosyl. The reconstitution of these truncated mutant protein into the DMPG liposome was conducted the same way for the full-length protein.

Section 4 Results and Discussions

4.1 Spectra of ¹³C, ¹⁵N-PLIN1

Initially, ^{13}C 1D spectrum on ^{13}C , ^{15}N -uniformly labeled PLIN1 was performed to assess resolution and sensitivity, providing feedback for sample preparation process. The spectrum was acquired at several temperatures to assess cross polarization dynamics (Figure 17). Overall signal intensity increased by 30% when the temperature was lowered from -15 to -37 °C, due to the fact that cross polarization from proton to ^{13}C is more efficient when the dipolar coupling is strong in rigid molecules. The lipid methylene intensity increased by 50% instead, indicating that the mobility of acyl chains is reduced at lower temperatures than the protein, which presumably binds to the surface of the lipid membrane. From these 1D spectra, signals can be identified for carbonyl, aromatic ring, $\text{C}\alpha$, and side chain carbons.

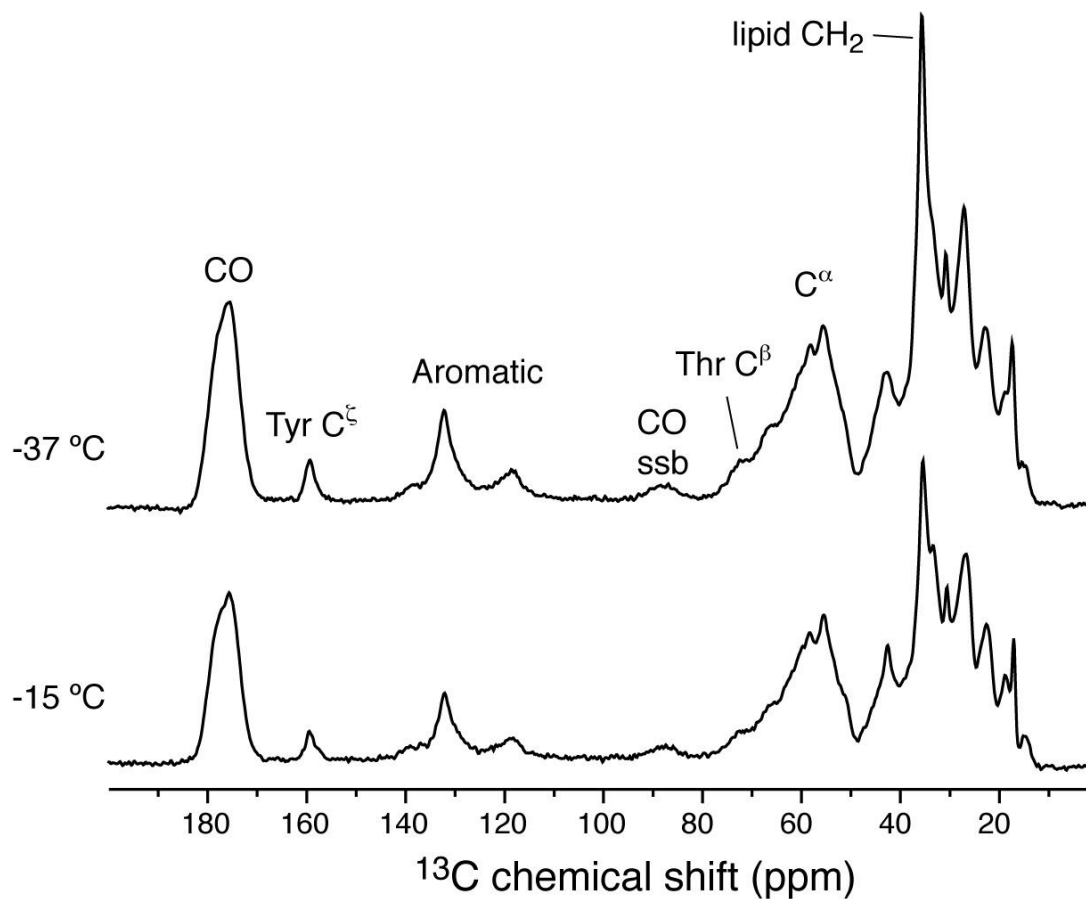


Figure 17 ^{13}C cross polarization spectra of ^{13}C , ^{15}N -PLIN1 lipoprotein complexes at different temperatures. Spectra were acquired on a 600 MHz instrument with the sample spinning at 13.3 kHz. For each spectrum 128 scans were accumulated. The data were apodized with 40 Hz line broadening. Calibrated sample temperature values are labeled in the figure.

Several ^{13}C - ^{13}C 2D correlation spectra were obtained to examine the chemical shift dispersion among amino acid types, and to analyze the secondary structures based on chemical shifts (the resonance position in a spectrum, in ppm unit) [95]. The ^{13}C - ^{13}C 2D spectrum (Figure 18), acquired with 10 ms DARR mixing [78], shows peaks in all the expected locations. For example, the serine, threonine, alanine, and isoleucine correlations have been labeled in the Figure.

Correlations observed in isolated regions of the spectrum (e.g., Ala $\text{C}\alpha$ - $\text{C}\beta$) provide a means to

performing analysis of secondary structure. Further analysis within the unique chemical shift ranges of various residue types showed that the $C\alpha$, $C\beta$ and CO (carbonyl carbon) signals were observed with chemical shifts indicative of their secondary structure. For example, 47% of the Ala $C\alpha$ - $C\beta$ correlations intensities (Figure 18 inset) are found in the area consistent with $C\alpha$ and $C\beta$ chemical shifts characteristic of α -helical structures; 31% of the intensities are found in the area characteristic of turn or random coil structures; 22% in the area characteristic of β -sheet. These numbers are consistent with the analysis based on circular dichroism spectrum: 34% α -helix, 16% β -sheets, 50% turns or random coil. Figure 18B shows the ^{13}C - ^{13}C 2D spectrum with 50 ms DARR mixing to create longer-range correlations. Among the many new peaks, the isoleucine $C\delta_1$ - $C\gamma_2$ and the inter-residue $C\alpha$ - $C\alpha$ correlations are also observed. Moreover, the individual outlying peaks allow estimation of linewidth to be about 0.5 ppm.

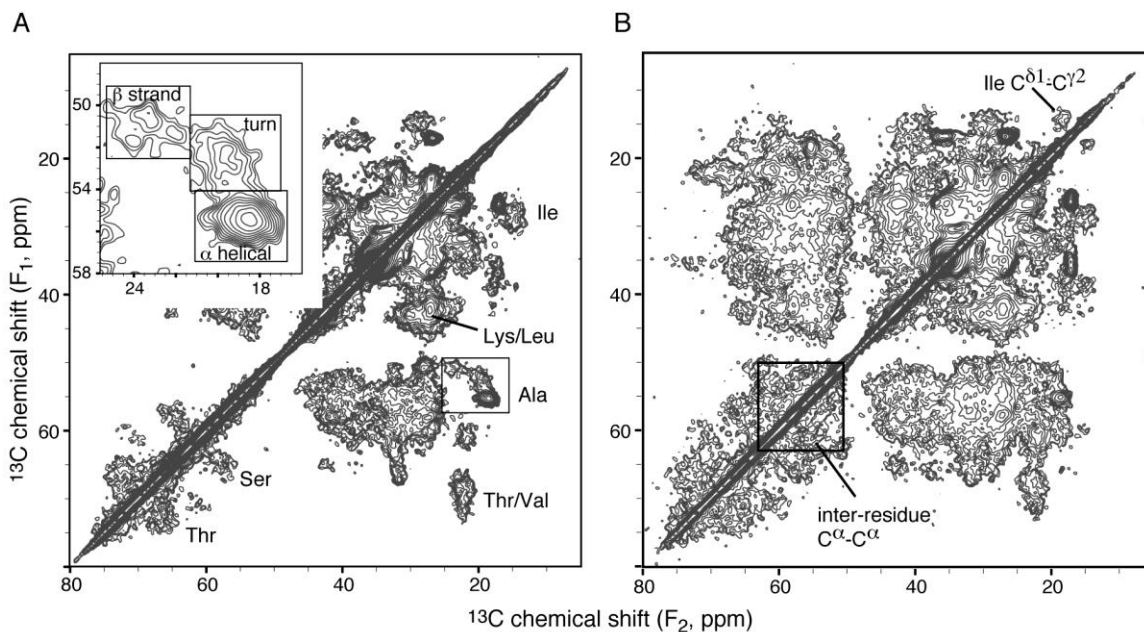


Figure 18 ^{13}C - ^{13}C 2D correlation NMR spectra of the ^{13}C , ^{15}N -PLIN1 sample with 10 ms (A) and 50 ms (B) DARR mixing. The inset is the expansion of the alanine $\text{C}\alpha$ - $\text{C}\beta$ region, with three boxes marking the characteristic chemical shift ranges for the three types of secondary structure. Data were acquired on a 600 MHz instrument with the sample spinning at 13.3 kHz. Each spectrum took 42 hours, with 96 scans per row. Sample temperature was -37°C .

4.2 Membrane insertion by proton spin diffusion

The attachment of PLIN1 on the surface of the lipid droplets is critical to its function in the activation of triglyceride hydrolysis. The four helices predicted in the central region have been hypothesized to be droplet targeting motifs based on their hydrophobicity (hydrophobic H6 and H8, amphipathic H7 and H9). To verify this hypothesis and to probe the depth of membrane insertion, NHH spin diffusion NMR experiment was performed at 45°C , higher than the gel to liquid-crystalline phase transition temperature (23°C) of DMPG (Figure 19A). This experimental temperature was chosen to give relatively sharp lipid methyl and methylene signals in proton spectra (Figure 19B). The proton resonances are assigned based on the published lipid NMR data.

Particularly, water resonates ca. 4.7 ppm, acyl chain methylene at ca. 1.3 ppm, and the terminal methyl groups of the acyl chains at ca. 0.8 ppm. In NHH spin diffusion experiments (Figure 16), proton polarization was first transferred to ^{15}N by cross polarization and all remaining proton signals were then wiped off. The polarization was transferred back to protein proton and propagated to water and lipids via spin-diffusion (t_m). The initial spectrum ($t_m = 0$ ms, Fig. 19C) shows weak residue amide signals not completely removed by 300 μs T_2 filter; longer T_2 delays hurt experimental sensitivity. The protein signal removal efficiency can be seen in Figure 20, where most of the signals were removed at $\tau_F = 600$ μs . Nevertheless, the aliphatic region (0-4 ppm) was free of interfering protein signals. The non-zero water signal likely originated from chemical exchange with lysine and arginine side chain amine protons, which could occur during ^{15}N to proton cross-polarization transfer. With increasing diffusion time (4 and 36 ms in Figure 19C), water and lipid signals grew stronger.

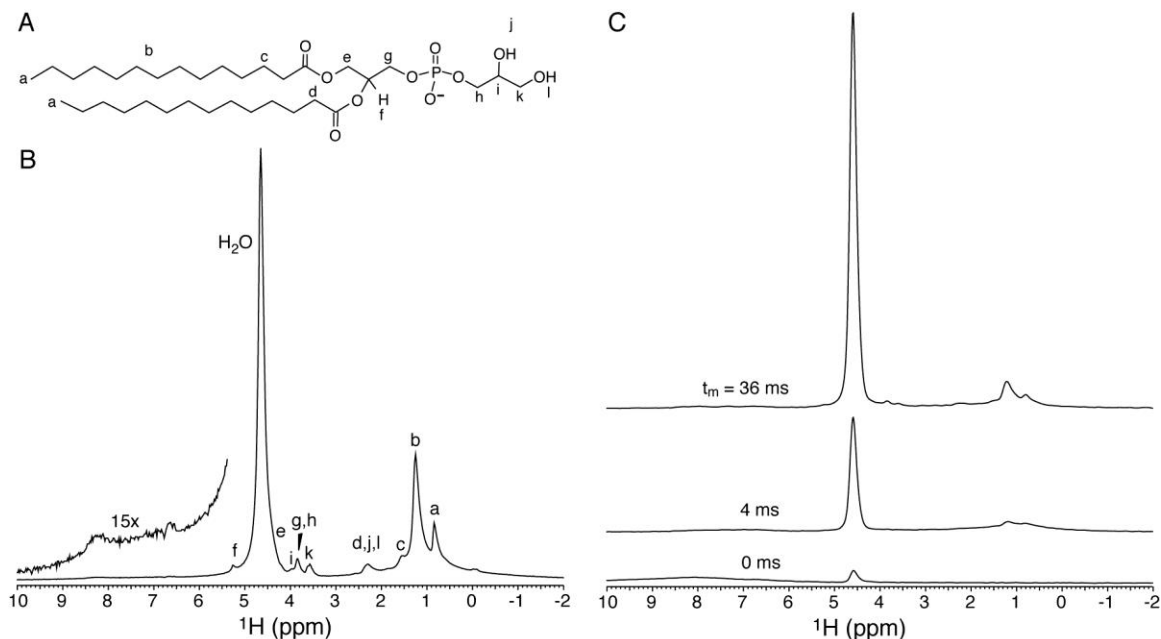


Figure 19 (A) DMPG molecule with proton designations. (B) Direct proton excitation spectrum on uniform ^{13}C , ^{15}N -PLIN1 reconstituted into DMPG vesicles at 45 °C. The amide region is magnified in the inset. (C) NHH spin diffusion spectra with diffusion time as indicated in the figure; 600 scans were acquired for each spectrum.

The dependence of peak intensities on serial diffusion times, namely, spin diffusion buildup curves, are shown in Figure 21. The buildup rate of how fast a buildup curve reaches its plateau relates to the distance between the protein and water or lipid groups. The curve of the terminal CH_3 groups of the acyl chains in Figure 21 exhibits a much faster buildup rate than water, indicating that a region of PLIN1 has close contact with the lipid acyl chains. This region of PLIN1 is buried in the middle of the bilayer in the native-like lipoprotein complexes; in the native lipid droplets, it very likely reaches the interface between the lipid monolayer and the triglyceride core. This deeply buried region should be critical to droplet attachment for PLIN1. Moreover, when PLIN1 is activated by phosphorylation, this region may be involved in opening the phospholipid monolayer to grant lipase access to the stored triglycerides.

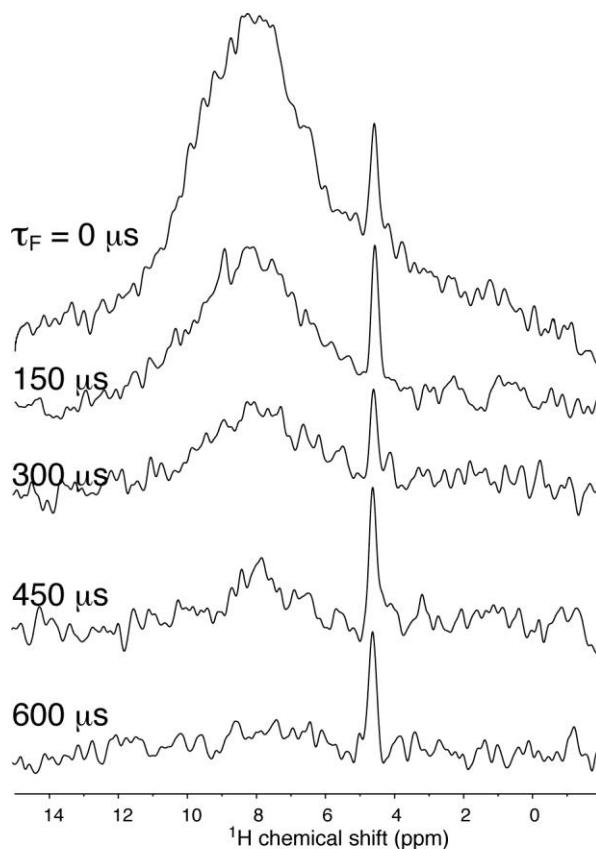


Figure 20 Suppression of protein signal by T_2 filter in the NHH spin diffusion sequence. Eight scans were acquired for each spectrum. The diffusion time (t_m) was zero.

The CH_2 buildup curve clearly shows a biphasic behavior, which can be decomposed into two exponential growth components. The fast component has a buildup rate similar to the CH_3 curve, very likely arising from a region of PLIN1 in close contact with the acyl chain. The slow component may be attributed to long-range diffusion from cytosolic regions of PLIN1. The buildup rate of water is also relatively fast, indicating close contact of water with the protein. The rate is slower than the rates of CH_2 and CH_3 in close contact with the protein, due to a smaller dipolar interaction resulted from fast motion of water molecules.

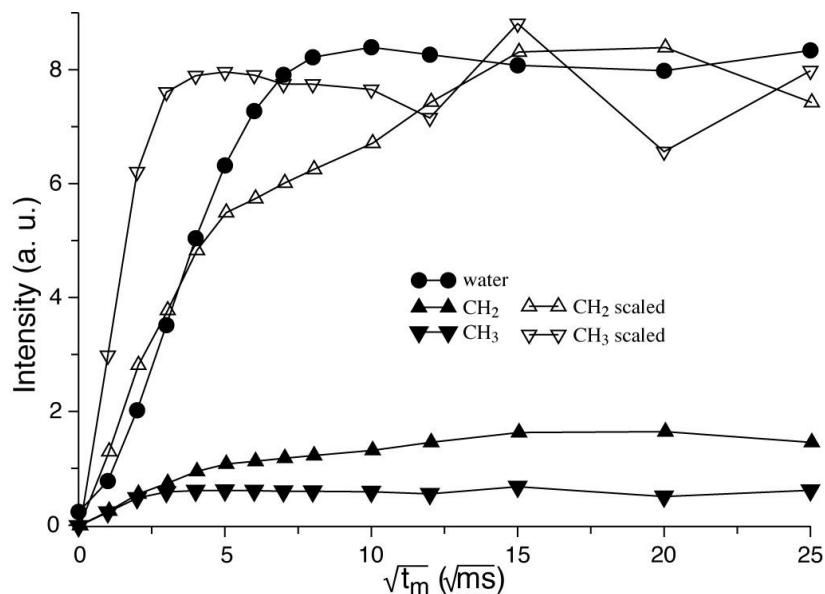


Figure 21 Proton 1D spin diffusion data buildup curves as a function of mixing time. The CH₂ and CH₃ curves are also rescaled for easy comparison of buildup rates. The fast growing CH₃ curve (down triangle) indicates a close contact between protein and the phospholipid acyl chain terminals, meaning that this part is deeply buried inside the lipid bilayer. While the CH₂ group (up triangle) shows a clear biphasic behavior indicating that there are two types of interactions with CH₂ groups from the protein, possibly the one buried inside the bilayer and the one in the cytosol part. The experiment was performed at 45 °C and data points were corrected for T₁ relaxation.

The proton spin diffusion data on water, CH₂ and CH₃ protons can provide a semi-quantitative evaluation of the protein regions located in different environments. The relative saturation intensities are roughly 85%, 8.5%, and 6.5% for water, CH₂, and CH₃ respectively, according to Figure 21 and taking into account that half of the CH₂ intensity arises from close contact (Figure 22). The sum of the latter two, 15%, represents percentage of the PLIN1 molecule anchored in the lipid membrane; this value is in very good agreement with the percentage (13%) of highly hydrophobic regions (residues 249 to 275 and 290 to 318) with regard to total number of amino acids (431) in the protein.

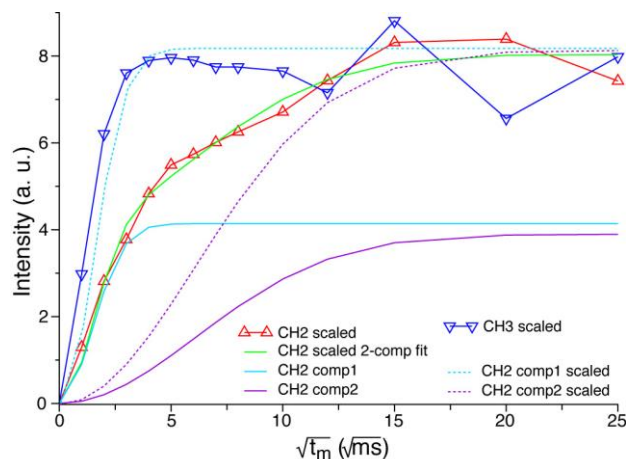


Figure 22 Decomposition of CH₂ buildup curve into two exponential growth components. The fast growth component was taken into account for calculating build up intensity from close contact with protein residues.

4.3 model building

So far, there is no structure or structure model for any PAT proteins bound to membrane. The conserved PAT domain of the PLIN1 is highly soluble and unlikely to participate in lipid interaction. No transmembrane domain was found in PLIN1 using HMMTOP server [96]. Bioinformatics analysis indicated that four predicted helices might participate in membrane targeting: hydrophobic helices H6 (249-261) and H8 (290-300) as well as amphipathic helices with high hydrophobic moment H7 (265-275) and H9 (301-318) [71]. The spin diffusion NMR experiments above corroborated this hypothesis. It is therefore interesting to build structural models for the membrane-binding domain, contained in the segment stretching from residue 249 to 318. This segment also includes a highly conserved hydrophilic motif ²⁸²EPENQARP²⁸⁹ that could act as a modulator of lipolysis.

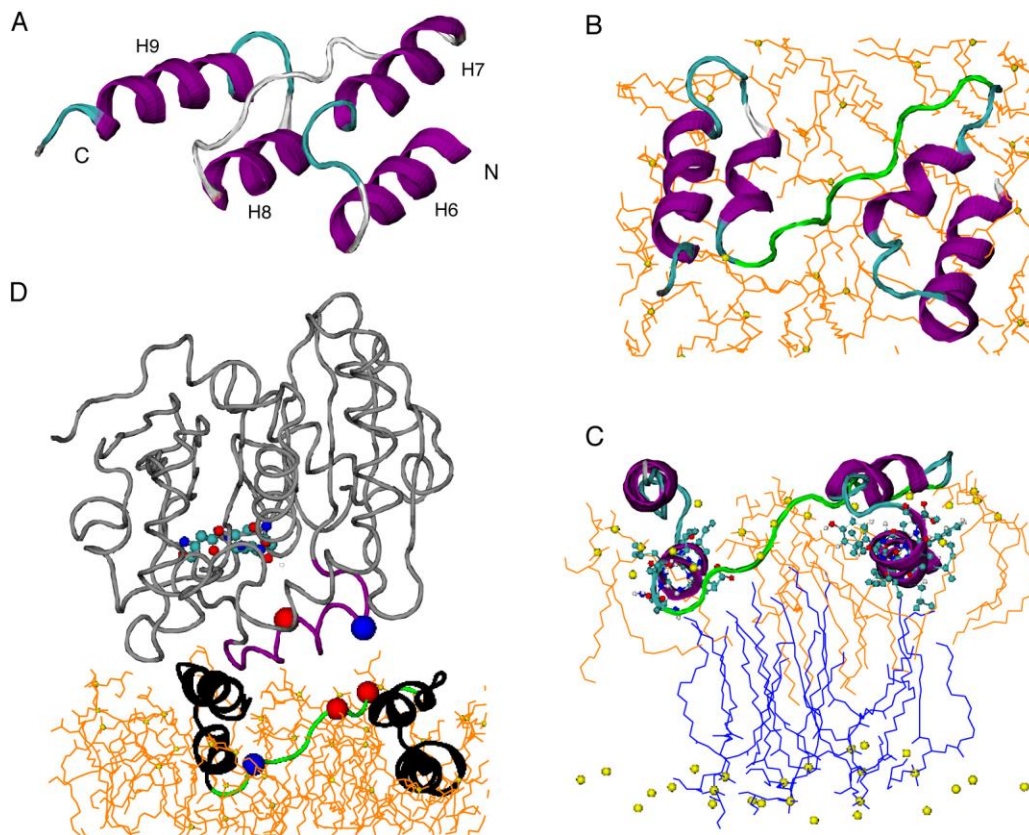


Figure 23 The trans-model of PLIN1 membrane-anchoring motif. (A) Starting model constructed based on hydrophobicity of predicted helices (hydrophobic H6 and H8, amphipathic H7 and H9) and NMR membrane insertion information. (B) Top view of the model after 50 ns MD simulation in DMPG lipid bilayer. Alpha helices are shown in purple, coils in gray, turns in cyan, and $^{282}\text{EPENQARP}^{289}$ motif in green. Lipids of the top leaflet are shown in orange, with the phosphorus shown in yellow sphere. (C) Side view of the model showing proximity between helices 6 and 8 with the lipid methyl groups. The lipids from top layer are shown in orange lines while those from bottom layer in blue. (D) Juxtaposition of the triglyceride lipase structure (PDB 4TGL) to the PLIN1 model. PLIN1 is shown in black, with the $^{282}\text{EPENQARP}^{289}$ motif shown in green, alpha carbon of E282 and E284 in red ball, alpha carbon of R288 in blue ball. TGL is shown in gray, with the opened lid (residues 82 to 96) in purple, the exposed catalytic center (S144, D203, H257) in CPK molecular models, alpha carbon of R86 in blue ball, and alpha carbon of D91 in red ball.

Two models of the most probable membrane-binding motif (residues 249 to 318) containing predicted helices 6 to 9 [71] were manually constructed based on hydrophobicity distribution: *trans*-model in Figure 23A and *cis*-model in Figure 25A, respectively. The amphipathic helices H7 and H9 most likely lie in the membrane-cytosol interface. The loop between H7 and H8 is so placed as to expose the hydrophilic conserved $^{282}\text{EPENQARP}^{289}$, which could be functionally

important. This requires the hydrophobic helices H6 and H8 to be placed parallel to the membrane surface. Had these two helices been aligned perpendicular to the membrane surface, this hydrophilic stretch would have been pulled into the membrane. First, the C α trace was designed with assistance of Visual Molecular Dynamics (VMD) [97] and then the all-atom molecule structure was generated by Structural Alphabet based protein Backbone Builder from alpha Carbon trace (SABBAC) 1.3 [98]. The two starting models (shown in Fig. 5A and Fig. S4) were energy minimized in vacuum using GROMACS 4.5.5 [99] and then embedded in a bilayer containing 512 DMPG molecules, with the amphipathic helices right above the lipid phosphate groups. The atomic level coordinates and interaction parameters for well-equilibrated DMPG lipid bilayer [100] were downloaded from Lipidbook [101]. These models were put in periodic triclinic boxes, solvated with adequate water and counterions were added to obtain electrically neutral system. These systems were then taken through steepest descent energy minimization and found to converge to physically realistic minimum energy value with maximum force less than 100 kJ/mol/nm. Then a total of 15 ns equilibration was performed, followed by a 50 ns production run of molecular dynamics (see details in Material and Methods).

The final structure of the *trans*-model is shown in Figure 23 B and C. All helices were able to maintain the helical structures, but helices 7 and 8 slightly unwound. In the final structure, H6 and H8 both have close contacts with the terminal methyl groups of the acyl chains from both leaflets, agreeing with the spin diffusion data (Figure 23C). The acyl chains of these lipids from the top leaflet wrapped around the helices to accommodate the perturbation caused by the protein. Such perturbation also caused the two leaflets to come close locally, allowing the terminal methyl groups from the bottom leaflet contact H6 and H8. Interestingly, lipids from the bottom leaflet

were getting very close to the surface in the region enclosed by the helices. This may imply that in real lipid droplet, triglyceride molecules might be brought close to the surface by the same mechanism, ready to leave the droplet in the event of lipolysis.

In Figure 23D, the crystal structure of a fungus triglyceride lipase [102] is juxtaposed to the PLIN1 model. In this structure, the lid (⁸²SSSIRNWIADLTFVP⁹⁶, with residues 83 to 84 and 91 to 95 as hinges and charged residues underlined) of the lipase is propped open by diethyl *p*-nitrophenyl phosphate to expose a patch of hydrophobic area of 800 Å² and the catalytic center residues S144, D203, H257. It is interesting that the footage of the lipase matches with PLIN1 *trans*-model, and that the lipase lid has a similar orientation to the ²⁸²EPENQARP²⁸⁹ (charged residues underlined) motif of PLIN1. More importantly, two possible electrostatic interactions, one between the two glutamates (²⁸²E and ²⁸⁴E) on PLIN1 and ⁸⁶R on the lipase and the other between ²⁸⁸R on PLIN1 and ⁹¹D on lipase, may be responsible to open the lipase lid in place of diethyl *p*-nitrophenyl phosphate. There is no structure of any *Drosophila* triglyceride lipase, however, sequence alignments show that several *Drosophila* fat body triglyceride lipases share conserved amino acids with the lid segment of the fungus lipase (Figure 24). The charged residues ⁸⁶R and ⁹¹D of the fungus lipase have their counterparts in the *Drosophila* lipases (Figure 24). Specifically, RLRNFTND of two isoforms of CG8552, which have been considered to be activated by PLIN1 in insects [60, 71], could facilitate electrostatic interaction for three pairs of amino acids with ²⁸²EPENQARP²⁸⁹ of PLIN1.

```

4TGL      69 GDSEKTIYIVFRGSSSIRNWIAD
CG8552B/C 260 GIDEKLSITLESIPRLRNFTND
          *  **   *           **   *

4TGL      81 GSSSIRNWIADLTFVPVSYP
CG11055   303 GSSIKVNRLIELPAEPLKLP
          ***   *   *   *   *   *

4TGL      72 EKTIIYIVFRGSSSIRNWIAD
CG8823    365 EKWAHLDFIWGTEARKYVYD
          **   *   *   *   *

```

Figure 24 Sequence alignments of fungus Rhizomucor miehei triglyceride lipase chain A (PDB ID: 4TGL) [102] with Drosophila melanogaster triglyceride lipases: two isoforms of CG8552 (NP_001188714.1 and NP_001188715.1) [103], CG11055 (NP_611463.1) [103], and CG8823 (NP_477331.1) [104]. The alignments were performed using program SIM [105]. ⁸⁶R and ⁹¹D of the fungus lipase and corresponding charged residues in other proteins are underlined.

The interaction between the two proteins may also cause reorganization of the PLIN1 structure and further perturbation to the local phospholipid molecules. Energy released from the electrostatic interactions could convert to mechanical energy, causing the lipase to push against H7 and H9 of PLIN1 (Figure 23D). This could result in an increased area for the region enclosed by the helices, creating a passage for the stored triglyceride molecules, which were already very close to the surface, to diffuse toward the catalytic center. The interaction of PLIN1 and triglyceride lipase is probably modulated by phosphorylation and Ca²⁺. Phosphorylation of PLIN1 promotes lipase activity in hydrolyzing triglycerides stored inside the lipid droplet [69, 71] and Ca²⁺ is also an activator of lipolysis [106]. These facts suggest that phosphorylation and binding of Ca²⁺ could promote conformational changes affecting certain protein regions. These changes would affect the interaction of the protein with lipid, affecting the accessibility of the lipases to the triglyceride molecules, or the interaction of PLIN1 with lipases and/or other proteins required in the activation process. Thus, future studies of the structures of PLIN1 with and without Ca²⁺

may identify structural changes that would explain the role of certain protein regions in the function of PLIN1.

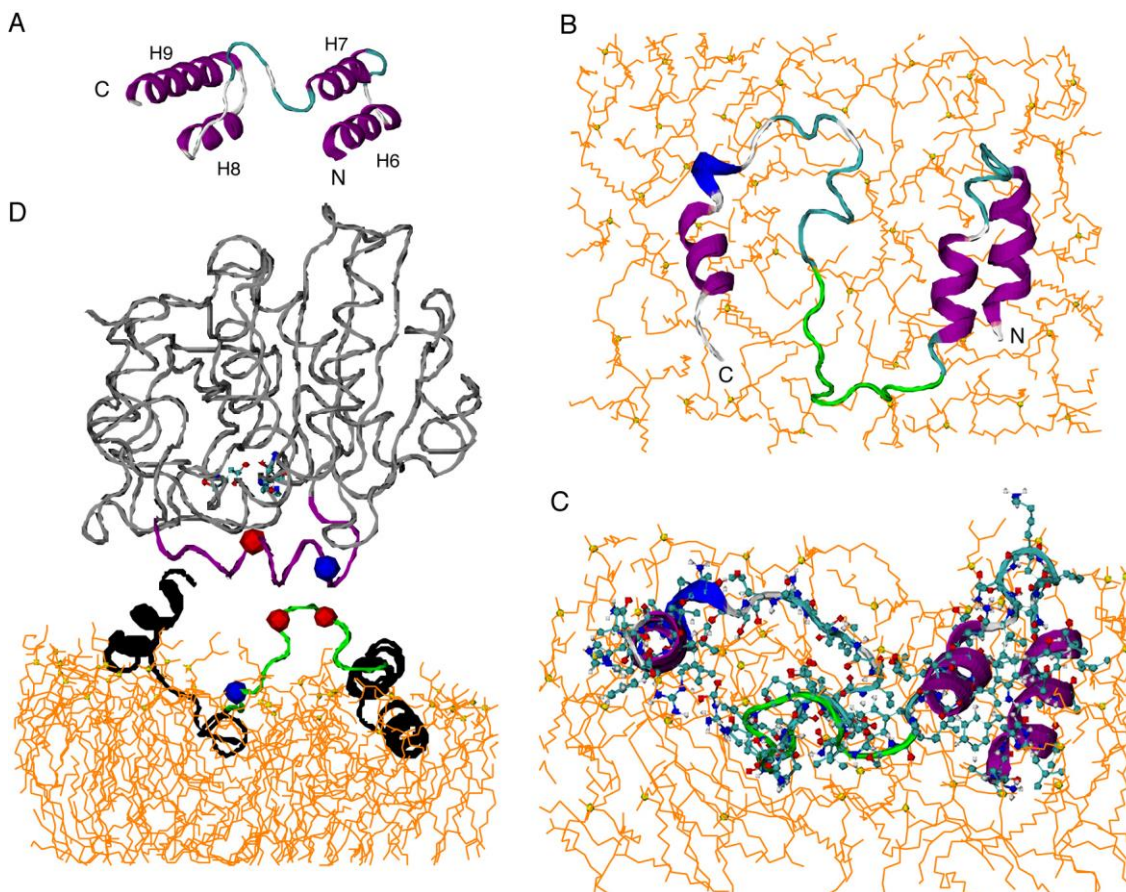


Figure 25 The cis-model of Lsd1 membrane-anchoring motif. (A) Starting model constructed based on hydrophobicity of predicted helices (hydrophobic H6 and H8, amphipathic H7 and H9) [71] and NMR membrane insertion information. (B) and (C) are two views of the model after 50 ns MD simulation in DMPG lipid bilayer. Alpha helices are shown in purple, 3-10 helices in blue, coils in gray, turns in cyan, and ²⁸²EPENQARP²⁸⁹ motif in green. Lipids are shown in orange, with the phosphorus shown in yellow sphere. Only one bilayer leaflet is shown since the protein does not penetrate into the other leaflet. (D) Juxtaposition of triglyceride lipase (PDB 4TGL) to Lsd1. Lsd1 is shown in black, with the ²⁸²EPENQARP²⁸⁹ motif shown in green, alpha carbon of E282 and E284 in red ball, alpha carbon of R288 in blue ball. TGL is shown in gray, with the opened lid (residues 82 to 96) in purple, the exposed catalytic center (S144, D203, H257) in CPK molecular models, alpha carbon of R86 in blue ball, and alpha carbon of D91 in red ball.

The starting and final structures of the *cis*-model are shown in Figure 24. Helix 8 is completely unwound and helix 9 also significantly changed its structure. Neither the *trans*-model nor *cis*-

model penetrates into the bottom layer of DMPG bilayer, in agreement with the proposed function of attaching the protein to the phospholipids monolayer covering the lipid droplet.

4.4 Model verification by NMR data

Backbone and β -carbon chemical shifts were predicted from structures by program shiftX [93]. Based on these chemical shifts, ^{13}C - ^{13}C 2D spectra were simulated and compared with experimental spectra (Figure 26 and 27). The simulated spectrum of *trans*-model agrees very well with experiments. The $\text{C}\beta$ -CO peaks of T262 and T291 (lower left corner of Figure 26A) do not have corresponding peaks in the experimental spectrum acquired with 10 ms DARR mixing, which is chosen for establishing short-range (1- and 2-bond) resonance correlations. Matching peaks are found for the experimental spectrum acquired with 50 ms DARR for long-range correlations (Figure 27B). These two threonine residues possibly undergo unfavorable dynamics, resulting in weaker dipolar coupling between $\text{C}\beta$ and CO. On the contrary, more unmatched peaks are found for the *cis*-model (Figure 26B and 27). The T291 and T 262 $\text{C}\beta$ -CO peak (lower left corner of Figure 26B) do not have matching peak even for the spectrum acquired with 50 ms DARR mixing (Figure 27B). A280, T300, and S303 are also mismatched. Therefore, the *cis*-model does not agree with NMR data.

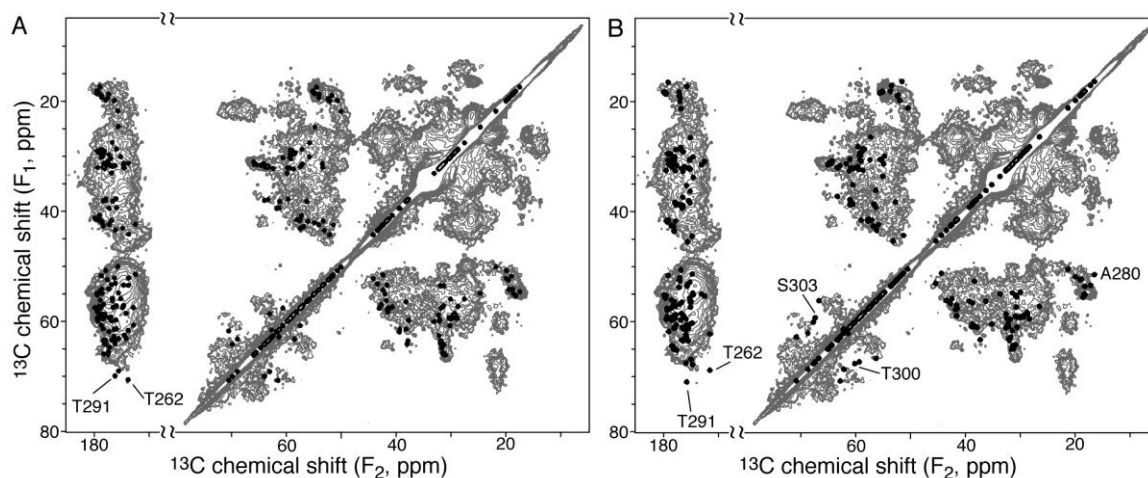


Figure 26 Verification of structure models by ^{13}C - ^{13}C 2D data. The experimental spectrum (gray) was obtained with 10 ms DARR mixing [78]. Simulated spectra (black) for the trans-model (A) and cis-model (B) only consist of $\text{C}\alpha$ - $\text{C}\beta$, $\text{C}\alpha$ - CO , and $\text{C}\beta$ - CO correlations, without other side chain carbons.

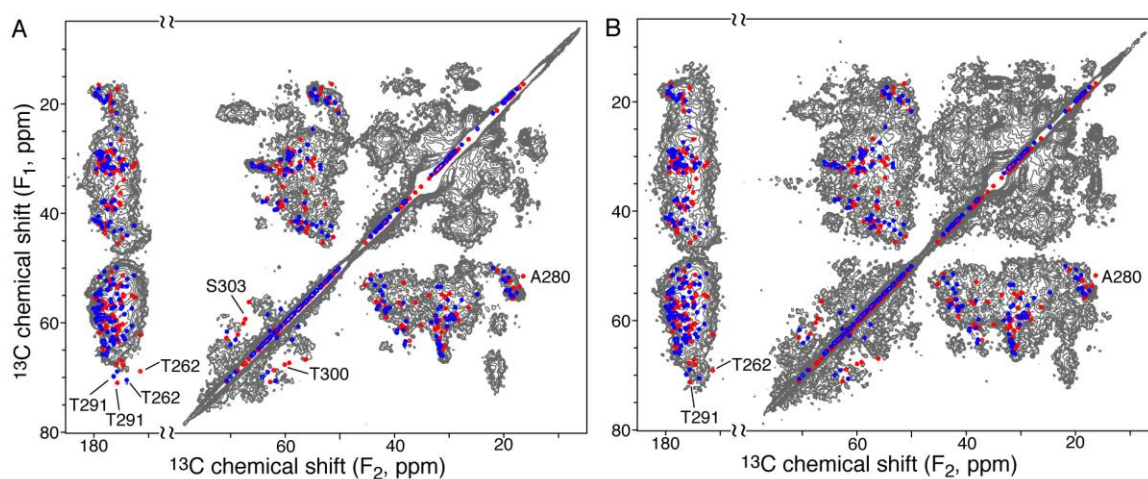


Figure 27 Verification of structure models by ^{13}C - ^{13}C 2D data with 10 ms (A) and 50 ms (B) DARR mixing. Simulated spectra are shown in blue and red for the trans-model and cis-model, respectively. The simulated spectra only consist of $\text{C}\alpha$ - $\text{C}\beta$, $\text{C}\alpha$ - CO , and $\text{C}\beta$ - CO correlations, without other side chain carbons. Outlying resonances are labeled.

4.5 Model verification by experimental data

The reconstituted PLIN1 protein in DMPG liposomes were subjected to protease digestion assay. The liposome were digested by trypsin at 37 °C overnight. High density Tricine gel was applied for separating these small peptides after digestion and revealed several small fragments around 1-3 kD as shown in Figure 28. The gel bands were cut off and sent for mass spec analysis.

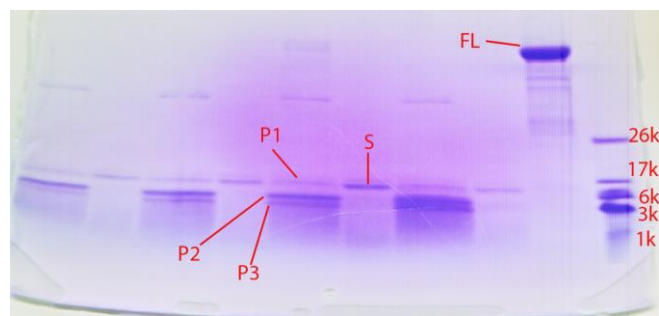


Figure 28 High density Tricine gel to separate small peptide after trypsin digestion of PLIN1-liposome. The very right lane is the marker with their molecular weights labeled. The second lane from the right is the full-length PLIN1 reconstituted into the liposome. Starting from the third lane from right to the very left, these eight lanes are the digested PLIN1-liposome supernatant and pellet fractions in alternative order. The best resolved bands (1 band from supernatant and 3 bands from pellet) as labeled in the figure were cut off for mass spec analysis.

According to the gel, there are two major bands in the pellet, consistent with our prediction that two helices are deeply buried inside the liposome, with corresponding molecular weights around 1.2 kD and 1.5 kD, respectively. The mass spec were performed on each band separated by the high density gel. A total of 36 unique peptides were identified in the full-length sample, 25 for P1, 27 for P2, 30 for P3 and 9 for supernatant.

According to the bioinformatics analysis, the full-length PLIN1 protein was truncated into 4 different segments to test the binding affinity to liposome, respectively. The four segments are named as the N-terminus PAT domain, R1 domain (the Region between PAT domain and 4-Helix domain), 4 Helix domain and R2 domain (Region after 4-Helix domain on the C terminus). All

segments were cloned into pET32 vector using primers shown in Table 3 and then transformed into *E. coli* for protein expression individually

Table 3 Primers designed to amplify each truncated mutant of *Lsd1*

domain	5' forward primer	3' backward primer	Restriction sites
PAT	5' GT ATT <u>GGT ACC</u> ATG GCA ACT GCA ACC AGC GGC AGT G	5' GTT AGA AGC TTA TTT CAG AAC CGG ACG CAC CAG GTG	HindIII, KpnI
R1	5' GT ATT <u>GGT ACC</u> CGC GCC GAT TCT GTC AAG CAA ATC	5' GTT AGA AGC TTA GAT GGT TCT TTG GGT AAG GCG GCG	HindIII, KpnI
4-Helix	5' GT ATT <u>GGT ACC</u> GCA GAG GCT CGC GCC CTC AAA AAG	5' GTT AGA AGC TTA AAT ATG GTG GGC AAC CTC TGT TG	HindIII, KpnI
R2	5' GT ATT <u>GGT ACC</u> ATC TAT ATC AAC CAC CGC ATC ATC	5' GTT AGA AGC TTA GTA GAC GCC GTT GAT GTT ATT GTG	HindIII, KpnI

Underline shows the complementary sequence to the template. Yellow means restriction cleavage sites. Red shows the added stop codon for the truncated recombination protein.

The strategy for expressing the four segments was to use the pET32 with full-length PLIN1 gene as the template. The four segments were amplified by PCR as described in the Material and Methods. The amplified DNA were purified by agarose gel and subjected to digestion using Hind III and Kpn I as the restriction nucleases. The pET-32-fl-PLIN1 plasmids were also digested using the same nucleases. The digested product were purified again by gel and ligated by T4 ligase. The ligation products were transformed into *E. coli* DH5 α for amplification. The plasmids were confirmed correct by sequencing and transformed into *E. coli* BL21 (DE3) for protein expression. Protein expressions for all segments yield sufficient amount (> 10 mg/ml) for follow up tests. For R1 region, the recombinant proteins were mainly expressed as soluble form,

indicating they are probably not associated with membrane interaction. Interestingly, all segments have some expression in inclusion bodies, while 4-helix domain has some expression as soluble parts, indicating that solubility of recombinant proteins cannot directly reflect its actual natural hydrophobicity. All proteins were purified the same way as full-length PLIN1. It is worth to mention that all these segments can be dissolved in 2 M urea buffer, which make them easier to be solubilized comparing to the full-length protein, which requires 8 M urea + 10% Sarkosyl. The solubilized proteins were buffer-exchanged to 0.01% cholate buffer after passing through Ni-NTA column for purification. The purified proteins were subjected to DMPG liposome reconstitution assay as described before for full-length protein. The reconstituted liposomes were pelleted at 70,000 g centrifugation and washed with buffer before final centrifugation. Both the supernatant and pellet were subjected to SDS-PAGE analysis.

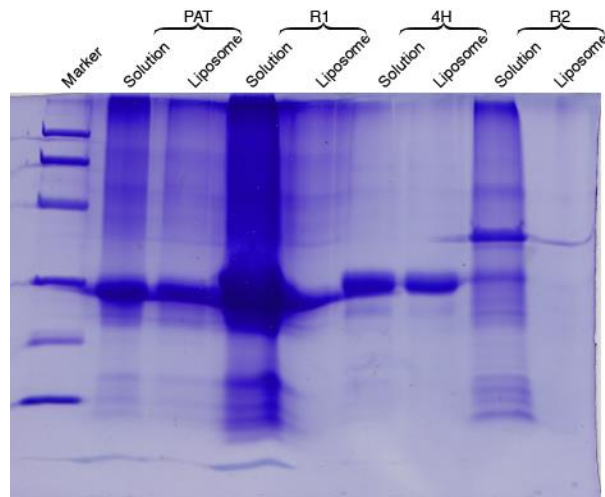


Figure 29 SDS-PAGE results of the reconstitution of each segment from PLIN 1 protein. The protein were reconstituted into the liposome through a detergent-mediated method as described in the Material and Methods. After reconstitution, the liposomes were centrifuged down. The solution and pellet are separated and checked by SDS-PAGE. As shown in the figure, the PAT domain and 4-Helices domain can be integrated into the DMPG liposome, indicating their hydrophobicity. R1 and R2 domains cannot be integrated into the liposome, they are most likely the soluble parts of the protein. For PAT and 4H domain, the protein bands in the solution part means there are excess amount of protein in the sample preparation and some are still left in the solution after reconstitution.

As shown in Figure 29, it is interesting that not only 4-helix domain can be reconstituted into the liposome, but the PAT domain also showed some ability to interact with the liposome. However, R1 and R2 domain have very low ability to be reconstituted into the liposome. It agrees with our NMR data and MD simulation model that 4-helix is the membrane interacting domain quite well. The PAT domain can also be reconstituted into the liposome. The bioinformatics analysis showed some hydrophobicity of this region, but no membrane interaction were predicted. It was reported that this domain was not required to bind to lipid droplet [60]. The GRAVY score of PAT domain gives -0.233 and PROSO (PROtein SOLubility) [107] shows very confused results because PROSO gives 0.557 which is just above the insolubility threshold 0.5, while PROSO II gives 0.667 as soluble protein, the scores is also just above the threshold 0.6. The MEMEX (MEMbrane EXperimentability prediction) [108] gives a high insolubility probability of 1.0. All these results together showed the complexity of this domain. The hydrophobicity of PAT domain might contribute to the hydrophobic interactions with lipase.

Section 5 Conclusion

NMR spin diffusion experiments were consistent with the predicted membrane attachment motif of PLIN1, and they indicated that some regions of PLIN1 have a deep contact with the phospholipid acyl chains near the bilayer center. For a native lipid storage droplet that is covered by a lipid monolayer, PLIN1 could penetrate to the interface of the monolayer and the triglyceride core. Two structure models for the membrane attachment motif were generated based on hydrophobicity analysis and NMR membrane insertion depth information, followed by optimization in lipid environment. Both models consist of four membrane interacting elements that are roughly parallel to the membrane surface. Two amphipathic elements stay on the

membrane surface, and two hydrophobic elements are buried deeper. Simulated NMR spectra for the *trans*-model agreed with experimental spectra. Juxtaposition of the triglyceride lipase structure to the PLIN1 *trans*-model suggests a possible interaction of the conserved sequence (EPENQARP), which is on a long loop between lipid binding elements, with the lipase. The long loop could bind to the lipase lid domain by electrostatic interactions and open the lid to expose the catalytic center. Interaction with the lipase could also cause reorganization of the membrane attachment elements of PLIN1, leading to an increased area for the region enclosed by the membrane attachment motif. A passage may be created by this process for the stored triglyceride molecules, which are already very close to the surface due to perturbation by PLN1 attachment, to diffuse toward the catalytic center. Therefore, this structural model could help design future experiments to elucidate the role of PLIN1 in lipolysis.

CHAPTER VI

INTERACTION BETWEEN MORTALIN AND SHETA2

Section 1 Abstract

Retinoids are a group of small molecules that have been clinically proven with anti-cancer activities. They specifically induce cell apoptosis in malignant cancer cells but not normal tissue cells. However, the toxicity of this class always hinders its application *in vivo*. Heteroatom substitution on the ring generates a new class called heteroarotinoids showing dramatically reduced toxicity. Recent modification leads to a new novel class flexible heteroarotinoids (Flex-Het) with a flexible linker between the heteroatom ring and aryl ring. Among these Flex-Hets, SHetA2 with thiourea linker is the most promising one, showing the greatest anti-cancer activity. Mass spec studies indicates SHetA2 binds to Mortalin, a heat shock protein 70 family member located in the mitochondria. It has been shown that in cancer cells, this protein is released to cytoplasm, sequestering p53 to block it from inducing apoptosis. In this study, we study the structure of the Mortalin substrate binding domain by solution NMR and show that it can interact with SHetA2. Chemical shift perturbation experiment demonstrates that the binding sites of Mortalin to SHetA2 are the substrate binding pocket. Docking of SHetA2 onto Mortalin by AutoDock result at least two favorable configurations, both are stabilized by 1) hydrophobic interactions of the hydrophobic heteroatom ring of SHetA2 with the hydrophobic substrate binding pocket of Mortalin; 2) the hydrogen bonds formed between the aryl ring NO₂ group of SHetA2

and side chains amide groups from either R513 or K576. According to this binding model, we design several new molecules that can occupy both configurations of SHetA2 and test their binding affinity with AutoDock. We find one of the candidate shows a much higher binding affinity to the protein. These findings will guide future generations of anti-cancer Flex-Het design.

Section 2 Introduction

The study on retinoids can be originated back to the exploration of metabolism of Vitamin A [109]. It was found that Vitamin A was transformed to an activator, retinoic acid, which regulates RNA and protein expression *in vivo* [110-112]. Retinoids are natural and synthetic vitamin A analogs with at least one aromatic ring that function similar to steroid/thyroid hormone [113]. These retinoids activate the dimer retinoid acid receptors (RARs and RXRs) to regulate specific gene expressions [114]. The retinoids (especially all-*trans* retinoic acid and 13-*cis*-retinoic acid) have shown potential in the chemoprevention of a variety of different cancers in animal models and human cell lines [115-126]. However, the clinical application of retinoids are limited by their significant toxicity [127]. Lots of efforts have been expended to develop mimics of all *trans*-retinoic acid with expectation of similar clinical effect but less toxicity. The search led to a novel group of compounds called heteroarotinoids (Hets) [128]. Heteroarotinoids are retinoid mimics containing an aryl ring and a heteroatom in the partially saturated ring. It has similar anti-cancer activity by activating RARs and induce gene regulation [129, 130]. The structure activity relationship studies found that increasing the specificity of heteroarotinoids for individual nuclear retinoic acid receptors could not separate the efficacy from the toxicity. The strategically insertion of heteroatom into the ring structure showed reduced toxicity in analogous of several retinoids

[129, 131, 132]. In particular, the toxicity of all-*trans* retinoic acid was reduced by 1,000-fold. Various Hets have been synthesized and tested on various cancer cell lines and several of them showed promising anti-cancer activities with reduced toxicity. Further exploration led to a group of compounds called Flexible Heteroarotinoids (Flex-Hets) where the linker between the heterocyclic ring and the aryl ring varied [132]. It was found that by introducing more flexible urea and thiourea linkers, the anti-cancer activity increased dramatically. Among all the Hets examined, the most potent one is called SHetA2 [133], the structure of which contains a thiourea group as the linker and Sulphur as the heteroatom in the heterocyclic ring. The official name is *N*-(3,4-dihydro-2,2,4,4-tetramethyl-2*H*-1-benzothiopyran-6-yl)-*N'*-(4-nitrophenyl)thiourea. It showed the greatest level of efficacy and potency in the inhibition of a variety of different cancer cell lines while retaining low if not none level of toxicity on normal cells [132, 134-139]. SHetA2 selectively inhibits cancer cell growth and induces cell apoptosis independent of the retinoic acid receptor pathway [140]. It was reported to suppress mitochondria permeability transition and enhance the cytochrome c release from mitochondria [134].

SHetA2 showed strong activity against human head and neck cell carcinoma [130], ovarian cancer [141], lung cancer [142] and kidney cancer cells [143] by inducing apoptosis while exhibited very low toxicity and excellent discrimination between malignant and benign cells. Oral administration of SHetA2 in rats [144] showed inhibition of ovarian and kidney cancerous xenograft tumors growth at 10 mg/kg/day. It showed no genotoxicity by bacterial-Ames test, both *in vivo* and *in vitro* chromosomal aberration on mouse micronucleus [145]. The pre-clinic tests showed no significant toxicity at the highest given doses to rats and dogs for a 14 day period. In a 28-day subchronic toxicity test, No Observed Adverse Effect Level (NOAEL) was achieved with

a dose of 1500 mg/kg/day on dogs [144]. Thus the therapeutic window for administrative safety with SHetA2 can be up to 150-fold above the effective dose *in vivo*. Pharmacokinetics studies [146] in mice revealed 10 mM concentration in plasma were achieved within 5 min. Oral bioavailability values were found to be 15% at 20 mg/kg and 19% at 60 mg/kg. The initial $t_{1/2}$ of 40 min and terminal $t_{1/2}$ of 11.4 hours were obtained. Total body clearance was 1.81 $\mu\text{L}/\text{h}/\text{kg}$. All these data demonstrated the behaviors of SHetA2 is a favorable candidate for future development. Recently, a second generation of Flex-Hets were designed, synthesized and evaluated for biological activity on ovarian cancer cells [147]. But the most active candidate is still only at comparable level to SHetA2. SHetA2 so far is still the best anti-cancer Flex-Het. It effectively inhibits all the 60 human cancer cells lines in the National Cancer Institute (NCI) and was promoted into the RAPID (Rapid Access to Preventive Intervention Development, NSC 721689) program for preclinical studies as therapeutic and chemopreventive agent for cancer.

It is well accepted that SHetA2 induces apoptosis in cancer cells. However, the detailed mechanism of the SHetA2 still remains elusive. One *in vitro* study showed that SHetA2 could reduce the level of Bcl-2 protein but not Bax and induce PARP-1 and caspase 3 cleavage to induce kidney cancer cell Caki-1 apoptosis [143]. It was also found that SHetA2 could repress the nuclear factor κB which regulates Bcl-2, cell differentiation as well as apoptosis. In both cancer and normal cells, SHetA2 induces G1 cell cycle arrest through reduction of cyclin D1. However, no directly evidence was found that SHetA2 directly interacts with these proteins. In order to search for direct interaction of SHetA2 with the protein, the small molecule was conjugated onto a magnetic sphere with slightly modification on the methyl group on the heteroatom ring [148]. The methyl group was converted to hydroxyl group so it can be linked to a long linker group on

the magnetic sphere. Both SHetA2 conjugated and non-conjugated spheres were incubated with human ovarian cancer cell A2780 whole cell protein extracts followed by washing with protein extraction buffer as well as excess amount of free SHetA2 molecules. The protein that was eluted from SHetA2-conjugated but not from non-conjugated spheres was subjected to mass spectrometry analysis. The protein Mortalin (HSPA9) was identified in both gel band excision QStar experiment and straight eluents analysis by Orbitrap. This is the first direct evidence that shows direct interaction of SHetA2 with protein in cell lysate. Further co-immunoprecipitation experiments confirmed the interaction between SHetA2 with Mortalin as SHetA2 disrupted the binding of Mortalin to p53 and p66shc inside cancer cells, where both mediate the mitochondria apoptosis pathway.

Mortalin was first identified as a cellular mortality factor as comparing the mortal and immortal cells as it segregated with loss of immortality of serially passaged mouse embryonic fibroblasts cybrids [149, 150]. Immunostaining found this protein to be a pancytoplasmic protein in mortal cells. However, the antibody against Mortalin showed a distinct pattern which is perinuclear in immortal MEFs. It was found later that mouse possesses two Mortalins, MOT-1 (pancytoplasmic) and MOT-2 (perinuclear), with contrast functions but only two amino acids differences [151, 152]. Overexpression of MOT-1 induces cellular senescence, while MOT-2 induces malignant transformation [153, 154]. Human only has one Mortalin homologue which functions as MOT-2 in mouse [154, 155]. The encoding gene of human Mortalin is located on chromosome 5q31.1 (gene name HSPA9B) [155, 156]. It was first called HMOT-2 until Mortalin was accepted. It is also known as p66mot-1, mtHsp70 (mitochondria stress-70 protein), GRP75 (glucose-regulating protein 75), HspA9 and PBP74 (peptide-binding protein 74). Human Mortalin is the mitochondria

heat shock protein 70 family member, but not induced by heat. It is a highly conserved molecular chaperone with high degree of identity with other Hsp70 family members as well as *E. coli* DnaK [157], *S. cerevisiae* SSC1p [158], rat cytosolic Hsc70 [159] and endoplasmic reticulum (ER) Bip [160]. As the major mitochondria chaperone, the primary function of Mortalin is assisting mitochondria targeting protein folding and transportation through the mitochondria membrane with the help of Hsp60 [161, 162]. Although Mortalin mainly distributed in the mitochondria, it can also be found in cytoplasmic regions. With different binding partners, it possesses diverse functions such as intracellular trafficking (Fibroblast Growth Factor-1, FGF-1) [163, 164], internalization of receptors (Interleukin-1 receptor type I) [165], protein modification (Mevalonate pyrophosphate decarboxylase MPD or MVD1 [166], voltage-dependent anion-selective channel 1 VDAC-1 [167]), iron-sulfur cluster biogenesis (Fe-S cluster [168] and J protein cochaperone), apoptosis suppression under mild stress response (p53, p66shc [169-172]) and immune responses (complement C8 and C9 [173, 174]). In particular, all Hsp70 family members have elevated expression levels in tumor cells, associating pre-malignant cell response to the selection process during tumorigenesis. These molecular chaperones contribute to the cellular immortalization due to the fact they buffer the conformational consequences of mutant proteins and help the cells survive through various stresses [175]. The mitochondria Hsp70 member Mortalin was also found overexpressed in several tumor cell lines [176-178]. The interaction between Mortalin and p53 was first identified in the cytoplasm of tumor cells. Mortalin sequesters p53 in the cytoplasm fraction through physical contact and inhibits its normal transcriptional activation function which leads to uncontrolled cell proliferation, a hallmark of cancer cells [179-181]. And this interaction can be abrogated by MKT-077, a well-known inhibitor of Mortalin, which reactivates the wild-type p53 function. Overexpression of Mortalin

can reverse the p53-dependent suppression of centrosome duplication indicating the interaction was localized to centrosomes [182]. The physical contact of Mortalin and p53 happens during late G1, S and G2 phases of cell cycle while disassociation from them occurs during mitosis. This interaction has not been observed in normal cells, indicating that normal cell life expansion by Mortalin is p53-independent [161, 183].

Human Mortalin is a 74 kD protein with 679 amino acids. It shows higher similarity to bacterial homologue comparing to other Hsp70 family members, indicating it might descend from the precursor mitochondria which is an endosymbiont bacteria trapped inside of the cell. Due to the insolubility, no full-length Mortalin structure is available so far. A very recent study successfully expressed the full length protein as soluble form with the help of coexpression with the Hsp70-escort protein1 (hHep1) [184]. However, only small angle X-ray scattering data was assessed and the low resolution surface shape was acquired. The *ab initio* model was superpositioned using two *E.coli* DnaK domains (NBD and SBD) as well as full-length DnaK. The superposition suggested an elongated shape of monomeric Mortalin comparing to DnaK. No detailed structure of the full-length protein have been revealed yet. Based on the evolutionary conservation comparing to its Hsp70 family members, it is proposed that this protein should contain two domains just like the other Hsp70s. The two domains are ~42 kD Nucleotide-binding domain (NBD) and ~25 kD substrate-binding domain (SBD, or peptide-binding domain as PBD) with protease sensitive linker connecting them. The SBD can be further divided into two subdomains, the 13 kD β -sandwich domain (SBD β) and the ~12 kD C-terminal helical domain (SBD α). They formed the so-called kettle model where NBD is the handle, SBD is the pot and the C-terminus is the lid. When NBD is bound to ADP, the two domains do not interact and it is a favorable

configuration for substrate to get into the substrate binding pocket inside the 'pot'. Upon ADP-ATP exchange, the SBD β and linker region dock onto the NBD and induce a dramatic conformational change on the SBD through allosteric regulation, leading to the openness of the C-terminal 'lid' and release of substrate. So far, SBD β together with the first two α -helices from SBD α and the whole NBD has been crystalized and their structure determined [185]. Using different deletion mutant recombinant protein, the co-immunoprecipitation assay presumed that Mortalin binds to p53 on the NBD, encompassing residues 253-282 although 105-538 segment didn't show binding [186]. These data are still ambiguous if not wrong and deletion might cause partially structure unfold and result non-specific interactions. While in another study, it was shown that the purified SBD of Mortalin, instead of NBD, associated with p53 in a concentration-dependent manor [187]. A mutation of V482F could abrogate the interaction which is also observed in DnaK, the bacteria homologue of Mortalin [188]. More interestingly, pre-incubation of Mortalin with a short peptide could greatly reduce its binding ability to p53. And the interaction between Mortalin and p53 can be abrogated by adding ATP if full-length protein was used [187]. 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 showed the exact interacting domains between Mortalin and p53 yet.

In this study, the structure of recombinant Mortalin SBD was accessed using solution NMR and it is found to be very similar to its crystal structure. We also used solution NMR to trace the chemical shifts of the purified recombinant Mortalin SBD protein titrated by SHetA2 with different molar ratios. Our results suggest that SHetA2 could physically bind to the substrate binding pocket of Mortalin. This could potentially disrupt the interaction of Mortalin with p53

due to the competitive inhibition thus leading to P53 translocation to nucleus and induce apoptosis. The docking experiments showed two main configurations of the interaction between SHetA2 and Mortalin which agreed with NMR data. According to these results, different SHetA2 analogs were designed and tested with docking. It was shown that the small molecule that can occupy the space of both configurations of SHetA2 gave the most binding potential. This result would give further guide lines for improving the Flex-Het design and cancer drug development.

Section 3 Materials and Methods

Purified SHetA2 powder is kindly provided by Dr. Benbrook. The isotope (^{13}C , ^{15}N) enriched ISOGRO, ^{13}C -glucose, $^{15}\text{NH}_4\text{Cl}$ are purchased from Sigma-Aldrich (St. Louis, MO). Ni-NTA purification kit is purchased from Bio-Rad (Hercules, CA). Lysozyme is purchased from MP Biomedicals (Santa Ana, CA). Thrombin is purchase from Bio Pharm Laboratories (Alpine, UT).

3.1 Protein expression and purification

Plasmid pET-52 with Mortalin SBD (AA#439-597) encoding gene was kindly provided by Dr. Benbrook. The plasmid was transformed into *E. coli* BL21 (DE3) for optimal protein expression. The stock cells were inoculated into LB medium and incubated at 37 °C until OD600 reaches 0.8. The cells were pelleted at 4000 g for 10 min and resuspended in 5 times volume of ^{13}C -glucose, $^{15}\text{NH}_4\text{Cl}$ minimum medium. The culture was incubated at 37 °C until OD600 reaches 0.8 again and then transferred to 20°C for induction. Final concentration of 1 mM IPTG was added to the culture and the cells are harvested after 20 hours induction by centrifugation. The cells are lysed by incubation with lysozyme at room temperature followed by sonication. The protein was found in the soluble part. The lysates were filtered against 0.2 μm pore size filter to remove any

remaining bacteria and then transferred to the Ni-NTA column for protein purification. The column was washed with 10 mM imidazole buffer and eluted with 1 M imidazole. The eluents were buffer exchanged to Tris-buffer and subsequently digested by HRV-3C protease to remove His-tag. The cleaved product were passed through Ni-NTA column again to remove the cleaved His-tag. The purity of the protein were examined by SDS-PAGE. Final concentration of the protein is 20 mg/ml and stored at 4 °C before use.

3.2 NMR spectroscopy

All NMR experiments were carried out on the Agilent INOVA 600 MHz spectrometer with the Nalorac ^1H , ^{13}C , ^{15}N -labeled PFG triple resonance probe. VnmrJ software with Biopack suite of pulse programs was used to collect the data. All data were acquired at room temperature (25 °C). Backbone chemical shifts were assigned in a sequential manner using the following experiments: NHSQC, HNCA, HNcoCA, HNCACB, HNcoCACB, HNCOC and HNcaCO. Data were processed using NMRPipe and analyzed using SPARKY. TALOS+ was used to calculate the secondary structure constraints and predict the secondary structure according to the backbone chemical shifts.

3.3 SHetA2 titration

SHetA2 was found to be water insoluble. DMSO was recruited as the solvent to dissolve SHetA2 and the highest concentration we can have is 170 mg/ml. The DMSO dissolved SHetA2 was added into the ^{15}N labeled protein solution sample at a molar ratio of protein to SHetA2 at 1:0.2, 1:0.4, 1:0.6, 1:0.8, 1:1, 1:2, 1:4 and 1:8. A total of 10 mg protein in 400 μl buffer with 10% D₂O was used for the titration experiment. For the first five titrations (1: 0.2 to 1: 1), 3.3 μl of 8.5

mg/ml SHetA2 dissolved in DMSO was added into the sample. The same volume of SHetA2 solution with different concentrations added into the sample for the last three titrations (3.3 μ l 42.5 mg/ml to achieve 1: 2, 3.3 μ l 85 mg/ml to achieve 1: 4 and 3.3 μ l 170mg/ml to get to 1: 8). This would have the minimum effect on the protein concentration to ensure spectrum quality. As a control, the same amount of DMSO (solvent only without SHetA2) was tested on another batch of ^{15}N labeled protein in parallel to the experiment and the same spectra with the same parameters were acquired. A quick proton 1D experiment was always checked before the NHSQC 2D experiment to ensure the sample is still in good condition and no significant loss of signals. The chemical shifts were assigned to each individual amino acids on both spectra. The chemical perturbation was calculated by using the difference between the SHetA2 titrated chemical shift minus the *apo*-state and the DMSO titrated chemical shifts minus the *apo*-state.

$$\delta = \sqrt{(\delta_{N_l} - \delta_{N_s})^2 + [10 \times (\delta_{H_l} - \delta_{H_s})]^2}, \quad (117)$$

where the δ_{N_l} and δ_{H_l} are the ^{15}N and ^1H chemical shift difference between the final SHetA2 titration (protein: SHetA2 = 1:8) and *apo*-state. The δ_{N_s} and δ_{H_s} are the ^{15}N and ^1H chemical shift difference between the corresponding DMSO titration and *apo*-state. Two *apo*-state spectra were used to eliminate any variations between the two samples titrated in different experiments so that more accuracy would be expected.

3.4 Protein assignments

The peaks were first selected in NHSQC spectrum with a contour level that is sufficient to pick around 200 peaks (~168 residues after HRV-3C protease digestion). Using these 200 peaks as the reference, the peaks were selected from other 3D spectra (~400 peaks were selected from HNCA,

HNCO and HNcoCACB, ~ 200 peaks were selected from HNcoCA, HNcaCO and ~800 peaks were selected from HNCACB). The peak coordinates data along with intensity data were uploaded to the PINE server for auto-assignments followed by extensive labor for manual corrections. The finalized labeled spectra will be uploaded onto the BMRB.

3.5 AutoDock docking simulation

AutoDock was applied for predicting the interaction sites of small molecules (ligands) with Mortalin (macromolecule receptor). AutoDock 4.2 [189] suite was used for all the docking experiments and the results are first analyzed using AutoDock Tool (ADT) and then in VMD. For all small molecule ligand, the structure were generated using ChemSketchTM (Advanced Chemistry Development, Inc. ADC/Labs, Toronto, Canada) and converted to PDB file using OpenBabelGUI [190]. The ligand were set to rotate freely though single bonds to accommodate best conformation in space after docking. For protein, the crystal structure of Mortalin (3N8E) was downloaded from Protein Data Bank. The protein structure were prepared by ADT to obtain correct ionization and tautomeric states of amino acid residues. Only polar hydrogens were added back to the protein structure. Kollman united atom partial charges and salvation parameter were assigned. The docking grid box was selected to not only cover all the NMR perturbation sites to ensure the correct guide lines but also the surface of the protein to ensure binding specificity. AutoGrid was run first to prepare the coordinates system and then the Lamarckian Genetic Algorithm was sued with a population size of 150 and 25 million maximum evaluations for 10 runs for AutoDock. All dockings were run on a personal computer with a 2.3 GHz Intel core i7 CPU.

Section 4 Results and Discussion

4.1 Recombinant Mortalin SBD expression and sample preparation

The Mortalin SBD (AA 439-597) encoding gene was introduced into pET-52 vector through Nco I and BamH I 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 followed by a HRV-3C protease cleavage site (LEVLFG/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. The plasmid was transformed into *E. coli* BL21 (DE3) for optimal expression. The cell was cultured in M9 minimum medium and induced by IPTG at 37 °C first, but it was found that no soluble protein was expressed at this temperature. This is probably caused by the fast expression of protein in high temperature and the folding process cannot complete in such a short time. The cell was then induced at a lower temperature (20 °C) overnight (20 hours). The recombinant protein was found in the soluble part in the cell lysates. The lysates were filtered to remove any remaining cells before loading onto the Ni-NTA column. The column was washed with three column volume of lysate buffer with 10 mM imidazole followed by 1 M imidazole elution. The eluent was buffer exchanged to Tris-buffer by passing through a 10 kD cutoff membrane. The protein was digested by HRV-3C protease to remove the His tag on the N terminus. The cleaved 10 His tag was further removed by passing the protease-treated solution through Ni-NTA column again. The final sample was concentrated using the 10 kD cutoff membrane. The purity of the sample was over 95% checked by SDS-PAGE as shown in Figure 30.

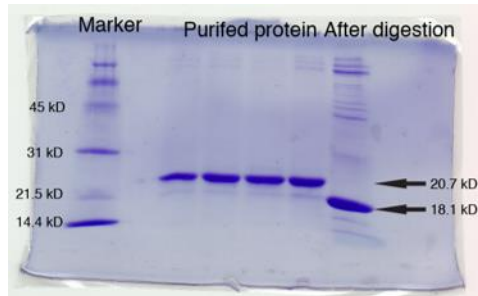


Figure 30 SDS-PAGE of purified Mortalin and digestion by HRV 3C protease. The lane on the left is the marker proteins with molecular weight labeled. The four lanes in the middle are the purified protein through Ni-NTA. The very right lane shows the protein after treated with HRV-3C protease. After removal the His-tag, the molecular weight of the protein goes from 20.7 kD to 18.1 kD.

4.2 Solution NMR of Mortalin SBD and backbone assignments

The 450 μL sample was mixed with 50 μL D_2O for locking signal before loading into the NMR tube. A standard 5 mm solution NMR tube was used for the experiment. First, an NSHQC spectrum was acquired to assess the quality of the sample. It was shown that the spectrum was well dispersed and proton resonance ranged from 7 ppm to 9.7 ppm, indicating the protein was well-folded as shown in Figure 31.

were selected with such a contour level that approximately 200 peaks can be selected for the 168 AA recombinant protein after cleavage. The resonances of the selected peaks in NHSQC were used as the guiding position in the 3D data set for peak selection. With proper contour levels, around 400 peaks were selected from HNCA, HNCO and HNcoCACB, ~ 200 peaks were selected from HNcoCA, HNcaCO and ~800 peaks were selected from HNCACB spectrum. The HNCA, HNcaCO and HNCACB provide the magnetization transfer information for both inter- and intra-residue, while HNcoCACB, HNcoCA and HNCO give only the intra-residue correlation. All these spectra are connected in such a way that each peak can find their partner in the other spectra to confirm the correct assignment. For example, normally two peaks from different carbon resonance frequencies but sharing the same proton and nitrogen frequency can be observed in the HNCA spectrum of residue i , one from residue i and the other from residue $i-1$. The peak from residue i can be found has a partner at the exact position in the HNcoCA spectrum of residue $i+1$ with a different nitrogen frequency, while the peak from residue $i-1$ has a partner in the HNcoCA spectrum of residue i itself. In this way, all resonance peaks from different spectra can be correlated through this process called backbone walk as shown in Figure 32.

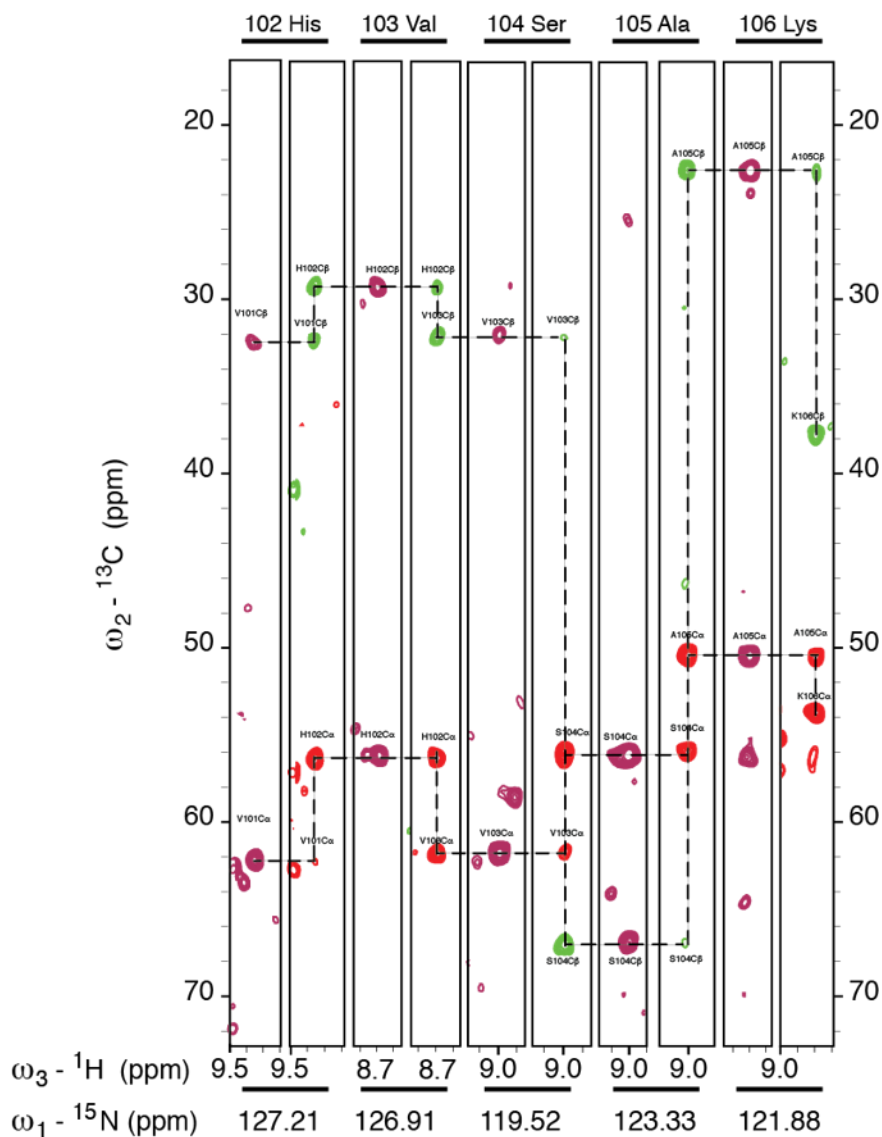


Figure 32 The backbone assignments example. The HNCACB and HNcoCACB spectra for residues 102 to 106 were displayed in each slices. Each HNCACB slice contains four peaks: the C and C from residue i and the C and C from residue $i-1$. Each HNcoCACB spectrum of residue i contains two peaks both from residue $i-1$. All these peaks can be correlated to form a backbone walk trace that connects all chemical shifts in a sequential manner.

The peak positions along with intensities data were exported and uploaded onto the PINE server 2.0 [191] for auto assignments. The assigned data returned in high confidence for over 90% of the amino acids (Figure 33). The predicted structure directly from the assignments agrees very well

with the crystal structure (PDB ID: 3N8E), where multiple β -sheets structure distributed in the N-terminus while two helices with a turn in between were found close to the C-terminus (Figure 35). The structure has not been finalized since some adjustments and corrections would be made on the assignments. Part of second helix toward the C-terminus was unstructured potentially because the missing of C-terminus 3 helices bundle and the folding of this region is not complete.

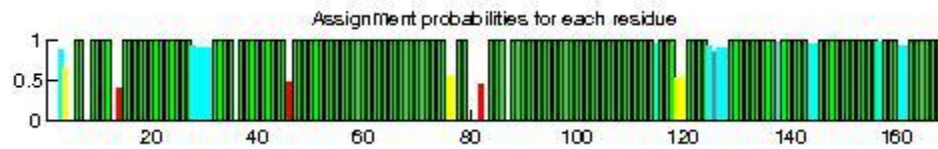


Figure 33 Assignments possibilities for each amino acids in Mortalin according to all 7 spectra uploaded onto the PINE server 2.0 including NHSQC, HNCA, HNcoCA, HNCOC, HNcaCO, HNCACB, HNcoCACB. Green bars indicates 100% confidence. Yellow-green indicates over 90% confidence. Cyan means over 80%, yellow means ambiguous around 50%, while red means lower than 50% confidence.

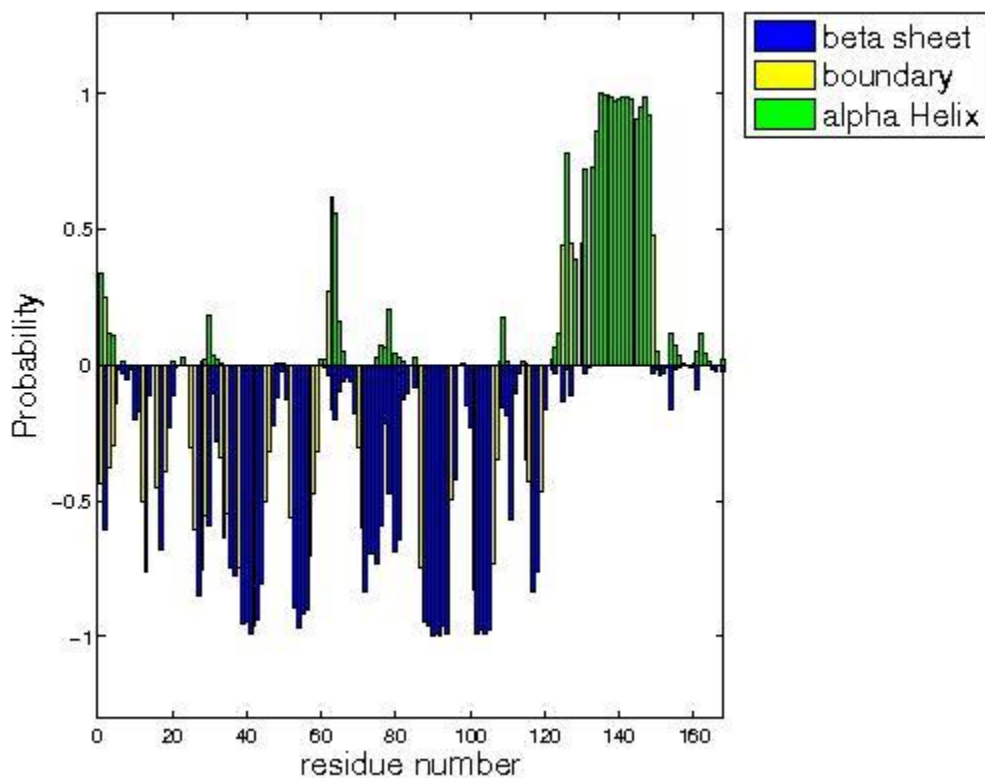


Figure 34 Secondary structure probabilities predicted based on the chemical shifts directly assigned to each amino acids backbone. Green bar means α -helical while blue bar indicates β -sheet structure and the length of each bar represents possibility.

The auto assigned spectra were examined in SPARKY program and manual adjustments and corrections were performed. The final corrected peak data were analyzed using TALOS+, a software which predicts the secondary structure based on ϕ/ψ location according to the NMR chemical shifts and amino acid type of the residue and its neighbors. The TALOS+ predicted secondary structure was illustrated in Figure 35. Again, it agrees very nicely with the crystal structure of the protein, with an eight β -sheets structure on the N-terminus and two connected α -helices on the C-terminus. The second helix was partially unstructured on the carboxyl end and

this is because of the flexible C-terminus of the protein in the solution. While in the crystal environment, the C-termini of the two asymmetric units are in close contact, keeping the helices more stable comparing to the solution states.

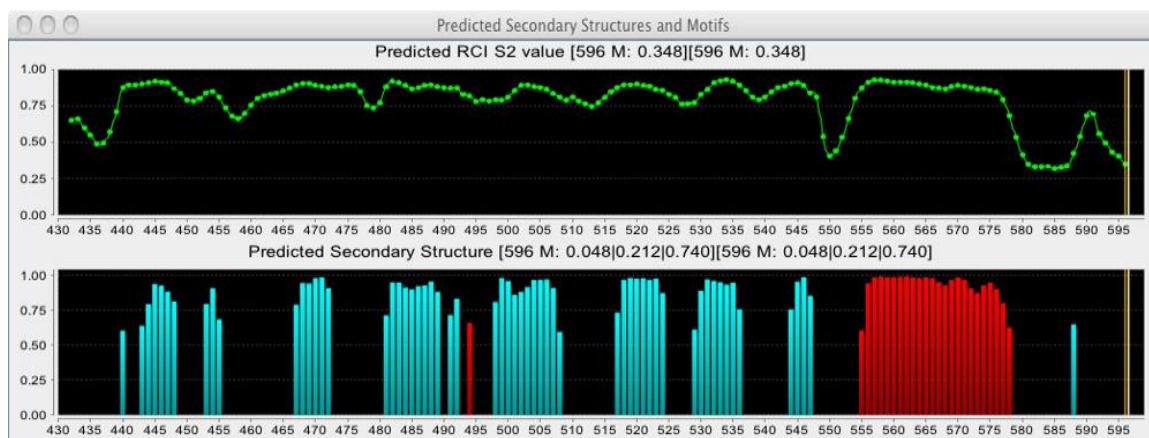


Figure 35 Secondary structure predicted by TALOS+ using the corrected Mortalin backbone chemical shifts. The structure predicted has a high overall confidence, especially in those structured regions. Both ends has a rather low confidence due to the fact that they are flexible regions in solution. Blue bars represent β -sheets structures while red bars represent α -helices and the lengths represent the probability.

4.3 SHetA2 titration and chemical shift perturbation

The interaction of protein with other agents such as ligand or other proteins can cause a local structure change or at least local electron density change. These changes can be readily detected by NMR spectroscopy. To investigate the interaction sites of Mortalin with SHetA2, chemical shift perturbation experiments were conducted. The SHetA2 was found to be water insoluble, so DMSO was used as the solvent for SHetA2. The highest concentration we got was 170 mg/ml. For the titration experiment, 500 μ L of mg/ml protein sample was prepared and loaded into the NMR tube. NQSQC spectrum was acquired for the *apo*-state protein (without SHetA2). Then 33

μL SHetA2 solution was added into the sample tube to achieve the molar ratio of protein: SHetA2 at 1:0.2. The NHCQC spectrum was acquired. The same procedure was repeated until the spectrum of the protein: SHetA2 ratio of 1:8 ratio was done. It was clear that some peaks are shifted as ligand added into the sample, as shown in Figure 36.

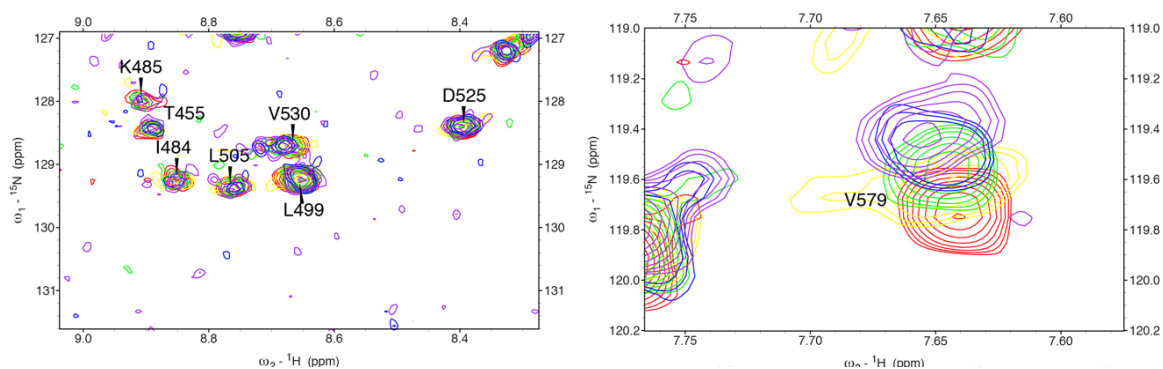


Figure 36 Protein peak shifts caused by adding SHetA2. The protein backbone resonance peaks in the NHCQC spectrum were affected by SHetA2 titration. On the left, titration has little effect on these peaks. However, on the right, this V579 peak has a relative large position shift. Different colors indicating different molecular ratio of protein to SHetA2. (Red, apo-state 1:0; yellow, 1:0.4; green, 1:0.8, blue 1:2, purple, 1:8)

In order to eliminate the disturbance of the solvent to the protein, a control experiment where the same amount of pure DMSO was added to the protein sample was performed and the NHCQC spectra on each step were recorded. All spectra from each titration steps were assigned using the apo-state as a reference. The chemical shift perturbations on each protein to SHetA2 ratio were calculated as described in the Material and Methods. The chemical shift perturbation of each amino acid were shown in Figure 37. The top 10 most shifted peaks are L457, V471, S473, A475, G514, V515, I518, D572, K576, and E586. All these peaks together form a nice circle around the putative substrate binding pocket (as compared to the structure of HSP70 and DnaK) of the Mortalin SBD, indicating SHetA2 binding to this substrate binding cavity. This substrate binding pocket might be also the interaction domain of Mortalin with p53 protein. Due to competitive

inhibition, binding of SHetA2 disrupt the interaction between Mortalin and p53, releasing p53 from the complex and the later translocate to the nucleus to induce apoptosis of the cancer cell.

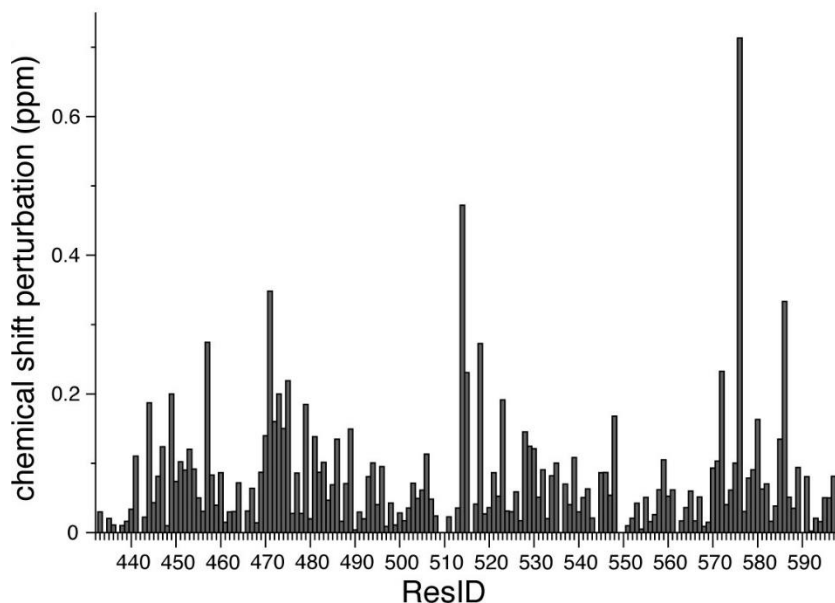


Figure 37 Chemical shift perturbation of SHetA2 on each amino acid residue of Mortalin. The height of the bar represents the relative chemical shift changes of SHetA2 titrated spectrum from DMSO titrated spectrum. The peaks from the final titration (protein: ligand=1:8) are used for the calculation.

4.4 Multi-configurations of interactions between SHetA2 and Mortalin by AutoDock

Mortalin has been shown to have the ability to sequester tumor suppress protein p53 in the cytoplasm region, inhibiting its ability to translocate into the nucleus and perform transcriptional regulation. In order to prove the SHetA2 can effectively bind to Mortalin and subsequently disrupt the interaction between Mortalin and p53, AutoDock was applied to study the interactions between SHetA2 and Mortalin. AutoDock is a suite of automatic docking tools that predict the interactions of ligand with biological macromolecules with known structure. It use Lamarckian

Generic Algorithm as well as empirical free energy scoring function to find the global minimum of the interaction energy between the ligand and receptor and explore all available degrees of freedom. The docking of SHetA2 was performed using the Mortalin crystal structure (PDB ID: 3N8E) as the macromolecule. The resulted lowest energy states were shown in Figure. As guided by the titration data, the two lowest energy states were found with direct interactions with the NMR titration sites on Mortalin, with binding energy of -8.45 kcal/mol and -7.7 kcal/mol, respectively. For configuration 1 (-8.45 kcal/mol) as shown in Figure 38, the hydrophobic *gem*-dimethyl group of SHetA2 was deeply inserted inside the substrate binding cavity of Mortalin, forming a strong hydrophobic interactions with methyl group on threonine T449. Additional hydrophobic interactions can also be found involving the hydrophobic end of SHetA2 and I447, V482, I484 and I518 deeply buried inside the substrate binding pocket of Mortalin. The two oxygen atoms on the nitro group of SHetA2 form hydrogen bonds with the two amidine protons from arginine (R513) to stabilize the hydrophilic end of the molecule (Table 4). The C-terminal E586 potentially interacts with R513 in the *apo*-state, adding of SHetA2 disrupts this interaction and E586 showed a large chemical shift change in the perturbation chart. In addition, another hydrogen bond can be found between carboxyl group of the serine (S473) and the NH group on the thiourea linker, and S473 also showed a large chemical shift change in the perturbation plot. For configuration 2 (- 7.7 kcal/mol) as shown in Figure 39, one of the oxygen atom from the nitro group on SHetA2 forms hydrogen bond with the amine group on K576, which is the most perturbed residue in the protein (Table 4 and Figure 37). Additional hydrogen bonds can be found between the NH groups on the SHetA2 thiourea linker with glutamine (Q470). The hydrophobic ring structure is barely inside the substrate binding pocket, in close proximity with S473 and V471. Some of these sites can be directly seen from the chemical shift perturbation chart such as

S473, V471, R576 and I518. However, some of them are not directly reflected by the perturbation chart potentially because the local secondary structure are not affected much and their backbone chemical shifts are not as sensitive. For these amino acids, side chain NMR experiments like NOESY could give more accurate perturbation effects.

Table 4 The hydrogen bond interactions between the Mortalin (PDB 3N8E) and two configurations of SHetA2

Protein	configuration	Hydrogen bond	Bond length	Binding energy
Mortalin (3N8E)	1	R(513) N-H ... O = N	2.84 Å	-8.45 kcal/mol
		S(473) C = O ... H-N	3.13 Å	
	2	K(576) N-H ... O = N	2.64 Å	-7.7 kcal/mol
		Q(470) C = O ... H-N	2.80 Å	
		Q(470) C = O ... H-N'	2.92 Å	

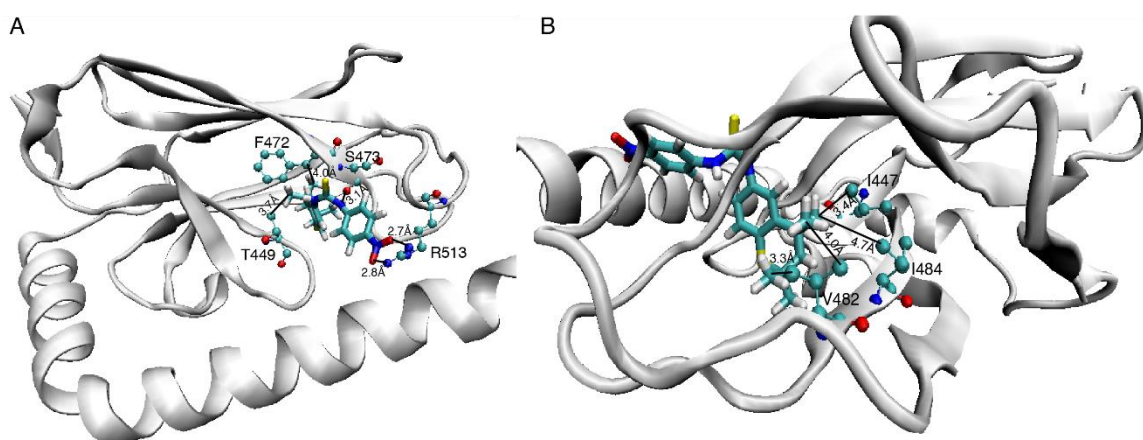


Figure 38 The first configuration of SHetA2 binding to Mortalin with binding energy of -8.45 kcal/mol. A) The nitro group on SHetA2 form hydrogen bonds with R513 from Mortalin. One of the NH group on the thiourea linker also forms hydrogen bond with CO group from S473. The hydrophobic tail of SHetA2 is deeply buried inside the substrate

binding cavity with strong hydrophobic interactions with F472 and T449. B) View from the other side of the substrate binding pocket of Mortalin. Additional hydrophobic interactions can be found between SHetA2 heteroatom ring with surrounding hydrophobic residues I447, I484, V482 and I518.

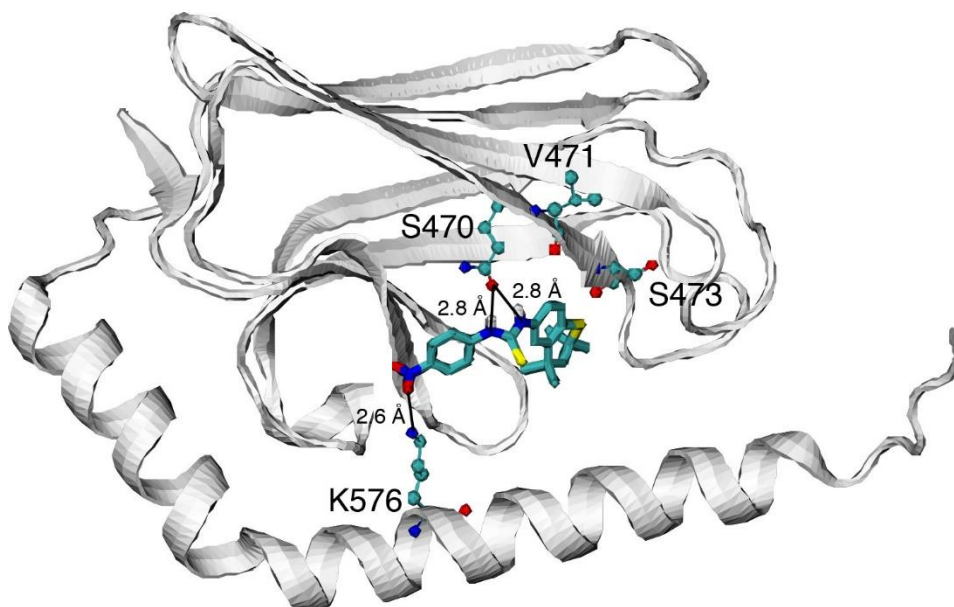


Figure 39 The second configuration of SHetA2 binding to Mortalin with binding energy of -7.7 kcal/mol. The three hydrogen bonds between SHetA2 and Mortalin K576 and S470 stabilize the structure in this position. The hydrophobic heteroatom ring is shallowly buried inside the substrate binding cavity, in close proximity with residues V471 and S473.

4.5 Improvements on small molecule structure to enhance binding

Clinical studies showed that SHetA2 disrupts the interaction between Mortalin and p53, releasing p53 from the complex to translocate into the nucleus to regulate the apoptosis. We have proved SHetA2 can interact with Mortalin and this interaction could be the mechanism of how SHetA2 competitively binds to the Mortalin substrate binding pocket, which most likely to be the p53 interaction domain. However, multi-configurations of SHetA2 binding indicate this is not a stable

interaction. The fact that the chemical shift perturbation in the SHetA2 titration experiments didn't not reach a saturation level even at 1: 8 molar ratio of protein: ligand further supports this conclusion. In order to investigate the binding preference of Mortalin to the small ligands, first, different SHetA2 analogs we have in hand were tested in the AutoDock to check their binding energy. A total of 31 different SHetA2 analogs were tested as listed in Table 5.

Table 5 Binding Energy of SHetA2 analogs

Analog No. or Name	Structure	Binding Energy (kcal/mol)
1) SHetA3-16		-8.54
2) SHetA4-17		-8.46
3) SHetC2		-8.31
4) 4,4 - HPR		-8.43

5) SL-1-50	-7.98
6) OHet72	-8.28
7) All- <i>trans</i> -RA	-7.56
8)	-7.51
9)	-7.69

10)

-8.27

11)

-8.09

12)

-7.53

13)

-7.86

14)

-8.65

15)

-7.79

16)	-8.23
17)	-7.81
18)	-9.65
19)	-9.15
20)	-7.57
21)	-8.05
22)	-8.67

23)

-8.3

24)

-8.5

25)

-8.76

26)

-9.58

27)

-8.41

28)

-8.44

29)		-8.63
30)		-7.97
31)	32)	-8.06

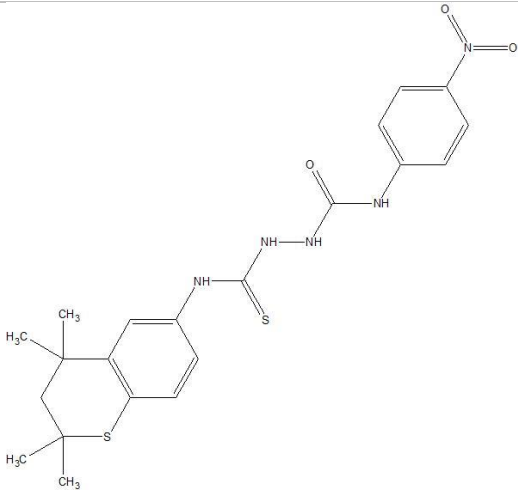
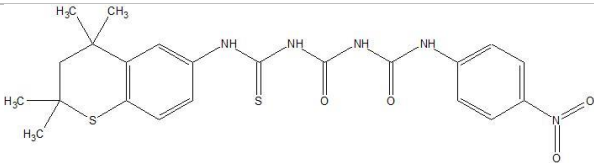
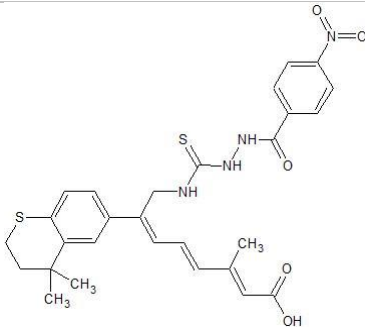
Comparing the binding energies of difference analogs with different functioning groups, we find that the NO₂ group on the aryl ring is more favorable comparing to the CO₂ group, while COEt group would achieve the same binding energy level as the NO₂ group. The Sulphur or Oxygen atom in the linker region does not make much difference to the binding energy considering they are chemically very similar. However, in the heteroatom ring, Sulphur and Oxygen substituted rings have very different binding energies, indicating Sulphur in the heteroatom ring instead of Oxygen is very important. While the heteroatom ring on one end, the other end has a better binding when it is a ring structure comparing to chain structure. When both ends of the molecule are hydrophobic heteroatom rings (No. 18), it yields the highest binding energy at -9.65 kcal/mol. In this molecule, one of the heteroatom ring was deeply buried inside the substrate binding pocket just like SHetA2, while the other end has close contact with Methionine (M584). The binding is different from SHetA2 where the Hydrogen bonding contributes more to the structure. We also

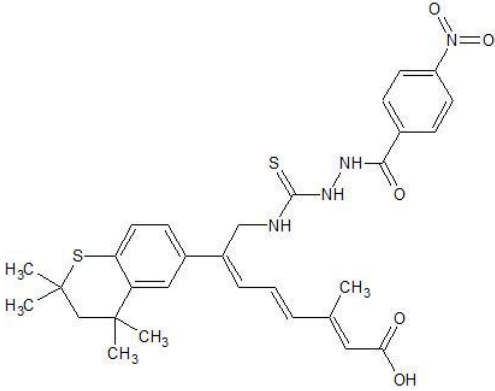
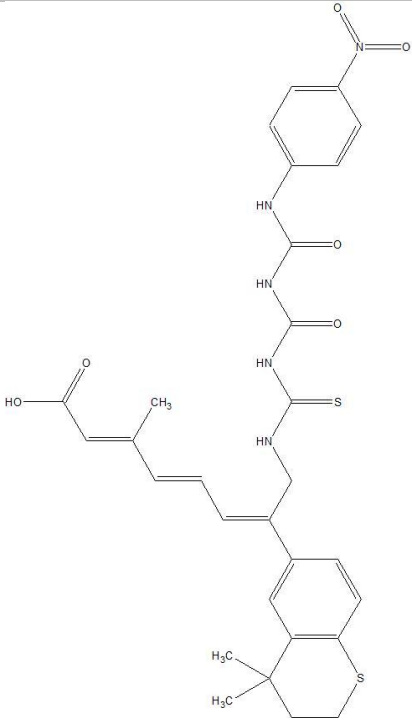
find that the molecules that are longer in structure bind better, which supports our findings in the NMR titration study. So if the molecule is long enough to occupy both configurations of SHetA2, it would have the highest binding affinity to Mortalin.

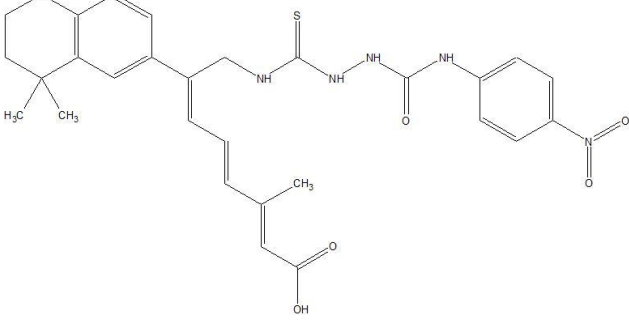
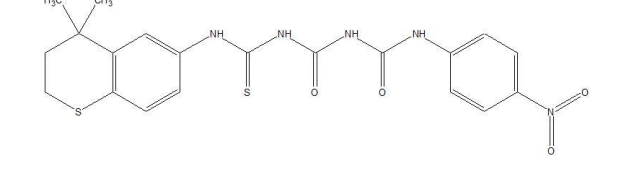
It has been shown so far that SHetA2 could bind to Mortalin in the substrate binding pocket. However, there could potentially be multiple configurations according to the NMR titration data and AutoDock results. The fact that the chemical shift changes in the NMR titration experiments didn't reach a plateau even at the 1:8, which is the highest concentration we can get for SHetA2, also indicates that there could be multiple configurations. Both the two lowest binding energy configuration of SHetA2 showed strong interactions with the sites that has significant chemical shift changes. However, configuration 1 was stabilized by the hydrogen bonds between the nitro groups on the SHetA2 and R513 on the protein. However, R513 was not significantly perturbed by the SHetA2 titration according to the NMR data. But the hydrophobic heteroatom ring of SHetA2 is deeply inside the substrate binding pocket of Mortalin, in close contact with I518 and V515, both of which show significant chemical shift changes on the perturbation experiments. In contrast, the configuration 2 was stabilized by the hydrogen bond between one of the nitro oxygen on SHetA2 and the side chain amide group of K576, which shows significant perturbation. However, the hydrophobic heteroatom ring is barely inside the cavity, resulting an unfavorable binding state (binding energy is higher comparing to configuration 1). All these evidences indicate that the binding of SHetA2 to Mortalin is a dynamic process. According to both the titration data and the small analogs we have been analyzed by AutoDock, we predict that if the linker region can be longer, and the molecule could reach to the K576 as well as keep the hydrophobic heteroatom ring inside the hydrophobic substrate binding pocket of Mortalin, it

should be the most favorable binding position. To test this idea, we designed several novel molecules to be tested on AutoDock as shown in Table 6.

Table 6 AutoDock results of small molecules designed for better binding affinity.

Name	Structure	Binding Energy (kcal/mol)
Analog 1		-9.56
Analog 2		-10.27
Analog 3		-11.09
Analog 4		-11.41

<p>Analog 5</p>	 <p>The structure of Analog 5 features a central 1,3-diene system. The left end of the diene is substituted with a 2,2,3-trimethyl-1,3-dithiane ring. The right end is substituted with a propenoic acid derivative, specifically a 2-methyl-3-(4-nitrophenylhydrazono)propenoic acid moiety. The diene is in a s-cis conformation.</p>	<p>-10.02</p>
<p>Analog 6</p>	 <p>The structure of Analog 6 is a long-chain molecule. It begins with a 2,2,3-trimethyl-1,3-dithiane ring at the bottom. This is connected via a diene system to a propenoic acid derivative. The diene is in a s-cis conformation. The propenoic acid part is substituted with a 2-methyl-3-(4-nitrophenylhydrazono)propenoic acid moiety. The diene is in a s-cis conformation.</p>	<p>-11.93</p>

Analog 7		-10.36
Analog 8		-9.14

First, we elongated the linker of SHetA2 from four bond thiourea chain to a six bond chain. In agreement with our previous predictions, the new structure analog 1 binds to Mortalin very well, and the binding energy is -9.56 kcal/mol. If we further elongate the linker to seven and eight bonds, the binding affinity further increases, with binding energy of -10.27 and -11.09 kcal/mol, respectively. Taking a look at their structures after docking in Figure 40A, all three hydrophobic heteroatom rings are almost at the same position inside the hydrophobic substrate binding pocket of Mortalin. The long arm is reaching out in contact with K576 on the helix of Mortalin. The longer linker between the two interacting region is more favorable so that the tension on the linker would be smaller, resulting a better binding affinity. The short arm has less freedom of rotation thus more rigid to constrain the structure. To further test our conclusion on the multi-configurations of SHetA2, another set of small molecules (analog 4 to 7) was designed with the purpose to occupy all interaction sites from both configurations. All analogs 4 to 7 have two arms with different lengths. Analog 4 was designed with one arm the same as analog 1 and analog 6

was designed to with one arm the same as analog 3, analog 7 with an arm the same as analog 3, while analog 5 was designed to test the preference of double-double methyl group or single-double methyl group on the heteroatom ring structure. According to the AutoDock results, with no surprise, analog 4 and 6 have significant lower binding energies to Mortalin comparing to the previous tested molecules. Checking the binding structure as shown in Figure 40B, both arms are occupying the positions exactly they are designed to bind. With the long arm reaching to the K576 forming hydrogen bonds, while the short arm forming hydrogen bonds with R513. And the longer arm analog 6 has an even lower binding energy comparing to the analog 4, in agreements with the differences between analog 3 and 1. Analog 5 shows a higher binding energy due to the double-double methyl group on the heteroatom ring that might affect the hydrophobic interactions inside the substrate binding pocket. To test the effect of these methyl groups, analog 8 was designed with only single-double methyl groups comparing to analog 3, however, this does not increase its binding ability, potentially because in the single arm structure, the heteroatom ring has more space to move around inside the pocket to find a better binding sites, while in the double armed structure, this flexibility is more limited due to the tension created by the second arm.

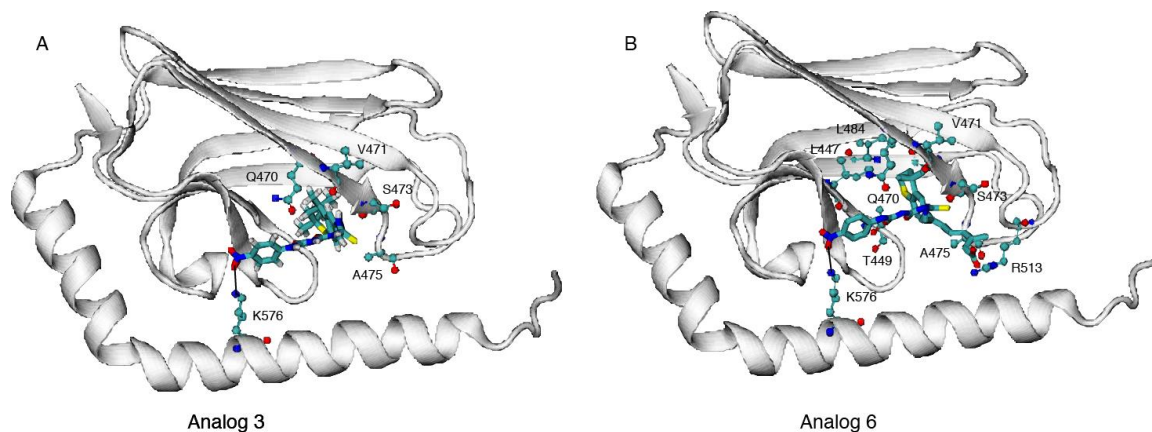


Figure 40 Docking of Analog 3 and 6 into Mortalin. A) Analog 3, which possess a longer linker (8 bonds) comparing to SHetA2, is docked into the hydrophobic substrate binding pocket of Mortalin on the heteroatom ring end while the nitro group on the other end can reach to K576, forming a hydrogen bond. These are the strongest interactions shown in the NMR titration experiments. B) Analog 6, which were designed with two one extra arm can occupy both configurations of SHetA2, with hydrophobic interactions stabilizing the heteroatom ring in the substrate binding pocket, while the other two arms form hydrogen bonds with K576 and R513, respectively.

Section 4 Conclusion

SHeta2 has been clinically shown anti-apoptosis activity in cancer cells through disruption of Mortalin and p53 complex, releasing p53 to translocate into the nucleus to function as an apoptosis regulator. In this study, we investigate the structure of Mortalin using solution NMR, which indicates it has a high similarity to the crystal structure expect the more flexible unstructured C-terminus. We also demonstrate that SHetA2 interacts with Mortalin through binding to its substrate binding pocket with multiple configurations. Through this competitive binding, it disrupts the interactions between p53 and Mortalin. This mechanism can be further enhanced by designing new small molecule that occupy the interaction sites for both configurations of SHetA2. Our findings will definitely guide the chemical design of Flex-Hex SHetA2 analogs in the future studies.

REFERENCES

1. Bloch, F., *Nuclear Induction*. Physical Review, 1946. **70**(7-8): p. 460-474.
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. Mittermaier, A.K. and L.E. Kay, *Observing biological dynamics at atomic resolution using NMR*. Trends in Biochemical Sciences, 2009. **34**(12): p. 601-611.
4. Glaubitz, C. and A. Watts, *Magic angle-oriented sample spinning (MAOSS): A new approach toward biomembrane studies*. J Magn Reson, 1998. **130**(2): p. 305-16.
5. Andronesi, O.C., et al., *Probing membrane protein orientation and structure using fast magic-angle-spinning solid-state NMR*. J Biomol NMR, 2004. **30**(3): p. 253-65.
6. Sizun, C. and B. Bechinger, *Bilayer sample for fast or slow magic angle oriented sample spinning solid-state NMR spectroscopy*. J Am Chem Soc, 2002. **124**(7): p. 1146-7.
7. Wallin, E. and G. von Heijne, *Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms*. Protein Sci, 1998. **7**(4): p. 1029-38.
8. Hopkins, A.L. and C.R. Groom, *The druggable genome*. Nat Rev Drug Discov, 2002. **1**(9): p. 727-30.
9. Andrew, E.R., A. Bradbury, and R.G. Eades, *Nuclear Magnetic Resonance Spectra from a Crystal Rotated at High Speed*. Nature, 1958. **182**(4650): p. 1659-1659.
10. Wishart, D.S. and B.D. Sykes, *Chemical shifts as a tool for structure determination*. Methods Enzymol, 1994. **239**: p. 363-92.
11. Cross, T.A., M.H. Frey, and S.J. Opella, *N-15 Spin Exchange in a Protein*. J Am Chem Soc, 1983. **105**(25): p. 7471-7473.
12. Gerlach, W. and O. Stern, *Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld*. Zeitschrift für Physik, 1922. **9**(1): p. 349-352.
13. Kellogg, J.M.B., et al., *The Magnetic Moments of the Proton and the Deuteron. The Radiofrequency Spectrum of H₂ in Various Magnetic Fields*. Physical Review, 1939. **56**(8): p. 728-743.
14. Gorter, C.J. and L.J.F. Broer, *Negative result of an attempt to observe nuclear magnetic resonance in solids*. Physica, 1942. **9**(6): p. 591-596.
15. Bloch, F., W.W. Hansen, and M. Packard, *The Nuclear Induction Experiment*. Physical Review, 1946. **70**(7-8): p. 474-485.
16. Bloembergen, N., E.M. Purcell, and R.V. Pound, *Relaxation Effects in Nuclear Magnetic Resonance Absorption*. Physical Review, 1948. **73**(7): p. 679-712.
17. Arnold, J.T., S.S. Dharmatti, and M.E. Packard, *Chemical Effects on Nuclear Induction Signals from Organic Compounds*. The Journal of Chemical Physics, 1951. **19**(4): p. 507-507.
18. Gutowsky, H.S., *Nuclear Magnetic Resonance in Metals: Temperature Effects for Na²³*. Physical Review, 1951. **83**(5): p. 1073-1074.

19. Ramsey, N.F. and E.M. Purcell, *Interactions between Nuclear Spins in Molecules*. Physical Review, 1952. **85**(1): p. 143-144.
20. Lowe, I.J. and R.E. Norberg, *Free-Induction Decays in Solids*. Physical Review, 1957. **107**(1): p. 46-61.
21. Ernst, R.R. and W.A. Anderson, *Application of Fourier Transform Spectroscopy to Magnetic Resonance*. Review of Scientific Instruments, 1966. **37**(1): p. 93-102.
22. Aue, W.P., et al., *Sensitivity of two-dimensional NMR spectroscopy*. Journal of Magnetic Resonance (1969), 1978. **29**(3): p. 523-533.
23. Bax, A. and L. Lerner, *Two-dimensional nuclear magnetic resonance spectroscopy*. Science, 1986. **232**(4753): p. 960-7.
24. Wüthrich, K., K. Nagayama, and R.R. Ernst, *Emerging techniques Two-dimensional NMR spectroscopy*. Trends in Biochemical Sciences, 1979. **4**(8): p. N178-N181.
25. Clore, G.M. and A.M. Gronenborn, *Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy*. Crit Rev Biochem Mol Biol, 1989. **24**(5): p. 479-564.
26. 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.
27. Moon, R.B. and J.H. Richards, *Determination of intracellular pH by 31P magnetic resonance*. J Biol Chem, 1973. **248**(20): p. 7276-8.
28. Navon, G., et al., *High-resolution 31P nuclear magnetic resonance studies of metabolism in aerobic Escherichia coli cells*. Proc Natl Acad Sci U S A, 1977. **74**(3): p. 888-91.
29. Hoult, D.I., et al., *Observation of tissue metabolites using 31P nuclear magnetic resonance*. Nature, 1974. **252**(5481): p. 285-287.
30. Lauterbur, P.C., *Image Formation by Induced Local Interactions - Examples Employing Nuclear Magnetic-Resonance*. Nature, 1973. **242**(5394): p. 190-191.
31. Kumar, A., D. Welti, and R.R. Ernst, *Imaging of Macroscopic Objects by Nmr Fourier Zeugmatography*. Naturwissenschaften, 1975. **62**(1): p. 34-34.
32. Wolins, N.E., D.L. Brasaemle, and P.E. Bickel, *A proposed model of fat packaging by exchangeable lipid droplet proteins*. FEBS Lett., 2006. **580**(23): p. 5484-5491.
33. Tauchi-Sato, K., et al., *The surface of lipid droplets is a phospholipid monolayer with a unique fatty acid composition*. Journal of Biological Chemistry, 2002. **277**(46): p. 44507-44512.
34. Digel, M., R. Ehehalt, and J. Fullekrug, *Lipid droplets lighting up: insights from live microscopy*. FEBS Lett, 2010. **584**(11): p. 2168-75.
35. Fujimoto, T., et al., *Imaging lipid droplets by electron microscopy*. Methods Cell Biol, 2013. **116**: p. 227-51.
36. Ohsaki, Y., M. Suzuki, and T. Fujimoto, *Open questions in lipid droplet biology*. Chem Biol, 2014. **21**(1): p. 86-96.
37. Vance, J.E. and D.E. Vance, *The assembly of lipids into lipoproteins during secretion*. Experientia, 1990. **46**(6): p. 560-9.

38. Brasaemle, D.L., et al., *Proteomic analysis of proteins associated with lipid droplets of basal and lipolytically stimulated 3T3-L1 adipocytes*. J Biol Chem, 2004. **279**(45): p. 46835-42.
39. Cermelli, S., et al., *The lipid-droplet proteome reveals that droplets are a protein-storage depot*. Curr Biol, 2006. **16**(18): p. 1783-95.
40. Soulages, J.L., et al., *Developmental changes in the protein composition of Manduca sexta lipid droplets*. Insect Biochem Mol Biol, 2012. **42**(5): p. 305-20.
41. Beilstein, F., et al., *Proteomic analysis of lipid droplets from Caco-2/TC7 enterocytes identifies novel modulators of lipid secretion*. PLoS One, 2013. **8**(1): p. e53017.
42. Na, H., et al., *Proteomic studies of isolated lipid droplets from bacteria, C. elegans, and mammals*. Methods Cell Biol, 2013. **116**: p. 1-14.
43. Zhang, P., et al., *Proteomic study and marker protein identification of Caenorhabditis elegans lipid droplets*. Mol Cell Proteomics, 2012. **11**(8): p. 317-28.
44. Hodges, B.D. and C.C. Wu, *Proteomic insights into an expanded cellular role for cytoplasmic lipid droplets*. J Lipid Res, 2010. **51**(2): p. 262-73.
45. Wang, H., D. Gilham, and R. Lehner, *Proteomic and lipid characterization of apolipoprotein B-free luminal lipid droplets from mouse liver microsomes: implications for very low density lipoprotein assembly*. J Biol Chem, 2007. **282**(45): p. 33218-26.
46. Ding, Y., et al., *Proteomic profiling of lipid droplet-associated proteins in primary adipocytes of normal and obese mouse*. Acta Biochim Biophys Sin (Shanghai), 2012. **44**(5): p. 394-406.
47. Sato, S., et al., *Proteomic profiling of lipid droplet proteins in hepatoma cell lines expressing hepatitis C virus core protein*. J Biochem, 2006. **139**(5): p. 921-30.
48. Zhu, Z., et al., *Dynamics of the lipid droplet proteome of the Oleaginous yeast rhodosporidium toruloides*. Eukaryot Cell, 2015. **14**(3): p. 252-64.
49. Baumeier, C., et al., *Caloric restriction and intermittent fasting alter hepatic lipid droplet proteome and diacylglycerol species and prevent diabetes in NZO mice*. Biochim Biophys Acta, 2015. **1851**(5): p. 566-76.
50. Dahlhoff, M., et al., *Characterization of the sebocyte lipid droplet proteome reveals novel potential regulators of sebaceous lipogenesis*. Exp Cell Res, 2015. **332**(1): p. 146-55.
51. Lee, J.S., et al., *Pharmacological ER stress promotes hepatic lipogenesis and lipid droplet formation*. Am J Transl Res, 2012. **4**(1): p. 102-13.
52. Robenek, M.J., et al., *Lipids partition caveolin-1 from ER membranes into lipid droplets: updating the model of lipid droplet biogenesis*. FASEB J, 2004. **18**(7): p. 866-8.
53. Robenek, H., et al., *Adipophilin-enriched domains in the ER membrane are sites of lipid droplet biogenesis*. J Cell Sci, 2006. **119**(Pt 20): p. 4215-24.
54. Roingeard, P., et al., *Hepatitis C virus budding at lipid droplet-associated ER membrane visualized by 3D electron microscopy*. Histochem Cell Biol, 2008. **130**(3): p. 561-6.
55. Xu, N., et al., *The FATP1-DGAT2 complex facilitates lipid droplet expansion at the ER-lipid droplet interface*. J Cell Biol, 2012. **198**(5): p. 895-911.
56. Wilfling, F., et al., *Triacylglycerol synthesis enzymes mediate lipid droplet growth by relocalizing from the ER to lipid droplets*. Dev Cell, 2013. **24**(4): p. 384-99.

57. Barbosa, A.D., D.B. Savage, and S. Siniouoglou, *Lipid droplet-organelle interactions: emerging roles in lipid metabolism*. *Curr Opin Cell Biol*, 2015. **35**: p. 91-97.
58. Popescu, C.I., et al., *Hepatitis C virus life cycle and lipid metabolism*. *Biology (Basel)*, 2014. **3**(4): p. 892-921.
59. Martin, S. and R.G. Parton, *Lipid droplets: a unified view of a dynamic organelle*. *Nat Rev Mol Cell Biol*, 2006. **7**(5): p. 373-8.
60. Bickel, P.E., J.T. Tansey, and M.A. Welte, *PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores*. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids*, 2009. **1791**(6): p. 419-440.
61. Kimmel, A.R., et al., *Adoption of PERILIPIN as a unifying nomenclature for the mammalian PAT-family of intracellular lipid storage droplet proteins*. *J Lipid Res*, 2010. **51**(3): p. 468-71.
62. Namigai, E.K. and Y. Suzuki, *Functional conservation and divergence of BMP ligands in limb development and lipid homeostasis of holometabolous insects*. *Evol Dev*, 2012. **14**(3): p. 296-310.
63. Gronke, S., et al., *Control of fat storage by a Drosophila PAT domain protein*. *Curr. Biol.*, 2003. **13**: p. 603-606.
64. Arrese, E.L., et al., *Expression of lipid storage droplet protein-1 may define the role of AKH as a lipid mobilizing hormone in Manduca sexta*. *Insect Biochem. Mol. Biol.*, 2008. **38**(11): p. 993-1000.
65. Martinez-Botas, J., et al., *Absence of perilipin results in leanness and reverses obesity in Lepr db db mice*. *Nature Genetics*, 2000. **26**(4): p. 474-479.
66. Tansey, J.T., et al., *Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity*. *Proc Natl Acad Sci U S A*, 2001. **98**(11): p. 6494-6499.
67. Beller, M., et al., *PERILIPIN-Dependent Control of Lipid Droplet Structure and Fat Storage in Drosophila*. *Cell Metab*, 2010. **12**(5): p. 521-532.
68. Marcinkiewicz, A., et al., *The phosphorylation of serine 492 of perilipin A directs lipid droplet fragmentation and dispersion*. *Journal of Biological Chemistry*, 2006. **281**(17): p. 11901-11909.
69. Patel, R.T., et al., *Activation of the lipid droplet controls the rate of lipolysis of triglycerides in the insect fat body*. *J. Biol. Chem.*, 2005. **280**: p. 22624-22631.
70. Hickenbottom, S.J., et al., *Structure of a lipid droplet protein: the PAT family member TIP47*. *Structure*, 2004. **12**(7): p. 1199-1207.
71. Arrese, E.L., et al., *Function and structure of lipid storage droplet protein 1 studied in lipoprotein complexes*. *Arch. Biochem. Biophys.*, 2008. **473**(1): p. 42-47.
72. Arrese, E.L., et al., *Purification and characterization of recombinant lipid storage protein-2 from Drosophila melanogaster*. *Prot. Pept. Lett.*, 2008. **15**(9): p. 1027-1032.
73. Gonzalez, M.C., et al., *The central type Y amphipathic alpha-helices of apolipoprotein AI are involved in the mobilization of intracellular cholesterol depots*. *Arch Biochem Biophys*, 2008. **473**(1): p. 34-41.

74. Prieto, E.D. and H.A. Garda, *Membrane insertion topology of the central apolipoprotein A-I region. Fluorescence studies using single tryptophan mutants.* Biochemistry, 2011. **50**(4): p. 466-79.
75. Garda, H.A., E.L. Arrese, and J.L. Soulages, *Structure of apolipoprotein-III in discoidal lipoproteins - Interhelical distances in the lipid-bound state and conformational change upon binding to lipid.* Journal of Biological Chemistry, 2002. **277**(22): p. 19773-19782.
76. Soulages, J.L. and E.L. Arrese, *Fluorescence spectroscopy of single tryptophan mutants of apolipoprotein-III in discoidal lipoproteins of dimyristoylphosphatidylcholine.* Biochemistry, 2000. **39**(34): p. 10574-80.
77. Lai, A.L., et al., *Fusion activity of HIV gp41 fusion domain is related to its secondary structure and depth of membrane insertion in a cholesterol-dependent fashion.* J Mol Biol, 2012. **418**(1-2): p. 3-15.
78. Takegoshi, K., S. Nakamura, and T. Terao, *^{13}C - ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR.* Chem. Phys. Lett., 2001. **344**(5-6): p. 631-637.
79. Luo, W. and M. Hong, *A 1D sensitivity-enhanced ^1H spin diffusion experiment for determining membrane protein topology.* Sol. St. Nucl. Magn. Reson., 2006. **29**(1-3): p. 163-169.
80. Zhou, D.H. and C.M. Rienstra, *High-performance solvent suppression for proton-detected solid-state NMR.* J. Magn. Reson., 2008. **192**: p. 167-172.
81. Kumashiro, K.K., et al., *A Novel Tool for Probing Membrane Protein Structure: Solid-State NMR with Proton Spin Diffusion and X-Nucleus Detection.* J. Am. Chem. Soc., 1998. **120**(20): p. 5043-5051.
82. Cruciani, O., et al., *An Improved NMR Study of Liposomes Using 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine as Model.* Molecules, 2006. **11**: p. 334-344.
83. Bennett, A.E., et al., *Heteronuclear decoupling in rotating solids.* J. Chem. Phys., 1995. **103**(16): p. 6951-6958.
84. Takegoshi, K., S. Nakamura, and T. Terao, *C -13- H -1 dipolar-driven C -13- C -13 recoupling without C -13 rf irradiation in nuclear magnetic resonance of rotating solids.* Journal of Chemical Physics, 2003. **118**(5): p. 2325-2341.
85. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit.* Bioinformatics, 2013. **29**(7): p. 845-854.
86. Hess, B., et al., *LINCS: A linear constraint solver for molecular simulations.* J. Comput. Chem., 1997. **18**(12): p. 1463-1472.
87. Darden, T., D. York, and L. Pedersen, *Particle Mesh Ewald - an $N \cdot \log(N)$ Method for Ewald Sums in Large Systems.* J. Chem. Phys., 1993. **98**(12): p. 10089-10092.
88. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* J. Chem. Phys., 2007. **126**(1): p. 014101:1-7.
89. Parrinello, M. and A. Rahman, *Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method.* J. Appl. Phys., 1981. **52**(12): p. 7182-7190.
90. Nose, S. and M.L. Klein, *Constant Pressure Molecular-Dynamics for Molecular-Systems.* Mol. Phys., 1983. **50**(5): p. 1055-1076.

91. Nose, S., *A Molecular-Dynamics Method for Simulations in the Canonical Ensemble*. Mol. Phys., 1984. **52**(2): p. 255-268.
92. Hoover, W.G., *Canonical Dynamics - Equilibrium Phase-Space Distributions*. Phys. Rev. A, 1985. **31**(3): p. 1695-1697.
93. Neal, S., et al., *Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts*. J. Biomol. NMR, 2003. **26**: p. 215-240.
94. Hirose, M., et al., *MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming*. Comput Appl Biosci, 1993. **9**(2): p. 161-7.
95. Frericks, H., et al., *Magic-angle spinning solid-state NMR of a 144 kDa membrane protein complex: E. coli cytochrome bo3 oxidase*. J. Biomol. NMR, 2006. **36**(1): p. 55-71.
96. Tusnády, G.E. and I. Simon, *The HMMTOP transmembrane topology prediction server*. Bioinformatics, 2011. **17**: p. 849-850.
97. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. J. Mol. Graph., 1996. **14**(1): p. 33-8, 27-8.
98. Maupetit, J., R. Gautier, and P. Tuffery, *SABBAC: online Structural Alphabet-based protein Backbone reconstruction from Alpha-Carbon trace*. Nucleic Acids Res. , 2006. **34**: p. W147-151.
99. Pronk, S., et al., *GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit*. Bioinformatics, 2013. **29**(7): p. 845-854.
100. Piggot, T.J., D.A. Holdbrook, and S. Khalid, *Electroporation of the E. coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes*. J. Phys. Chem. B, 2011. **115**(45): p. 13381-13388.
101. Domanski, J., et al., *Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations*. J. Membrane Biol., 2010. **236**(3): p. 255-258.
102. Derewenda, U., et al., *Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase*. Biochemistry, 1992. **31**(5): p. 1532-1541.
103. Hoskins, R.A., et al., *Sequence Finishing and Mapping of Drosophila melanogaster Heterochromatin*. Science, 2007. **316**(5831): p. 1625-1628.
104. Zinke, I., et al., *Suppression of food intake and growth by amino acids in Drosophila: the role of pumppless, a fat body expressed gene with homology to vertebrate glycine cleavage system*. Development, 1999. **126**(23): p. 5275-5284.
105. Huang, X. and W. Miller, *A time-efficient, linear-space local similarity algorithm*. Adv. Appl. Math., 1991. **12**(3): p. 337-357.
106. Arrese, E.L., et al., *Calcium and cAMP are second messengers in the adipokinetic hormone-induced lipolysis of triacylglycerols in Manduca sexta fat body*. J. Lipid Res., 1999. **40**(3): p. 556-564.
107. Smialowski, P., et al., *Protein solubility: sequence based prediction and experimental verification*. Bioinformatics, 2007. **23**(19): p. 2536-42.
108. Martin-Galiano, A.J., P. Smialowski, and D. Frishman, *Predicting experimental properties of integral membrane proteins by a naive Bayes approach*. Proteins, 2008. **70**(4): p. 1243-56.

109. Blomhoff, R., M.H. Green, and K.R. Norum, *Vitamin A: physiological and biochemical processing*. Annu Rev Nutr, 1992. **12**: p. 37-57.
110. DeLuca, H.F. and A.B. Roberts, *Pathways of retinoic acid and retinol metabolism*. Am J Clin Nutr, 1969. **22**(7): p. 945-52.
111. Trechsel, U. and H. Fleisch, *Retinol and retinoic acid modulate the metabolism of 25-hydroxyvitamin D3 in kidney cell culture*. FEBS Lett, 1981. **135**(1): p. 115-8.
112. Bhat, P.V. and A.M. Jetten, *Metabolism of all-trans-retinol and all-trans-retinoic acid in rabbit tracheal epithelial cells in culture*. Biochim Biophys Acta, 1987. **922**(1): p. 18-27.
113. De Luca, L.M., *Retinoids and Their Receptors in Differentiation, Embryogenesis, and Neoplasia*. FASEB Journal, 1991. **5**(14): p. 2924-2933.
114. Chambon, P., *A decade of molecular biology of retinoic acid receptors*. FASEB J, 1996. **10**(9): p. 940-54.
115. Bonanni, B. and M. Lazzeroni, *Retinoids and breast cancer prevention*. Recent Results Cancer Res, 2009. **181**: p. 77-82.
116. Hill, D.L. and C.J. Grubbs, *Retinoids and cancer prevention*. Annu Rev Nutr, 1992. **12**: p. 161-81.
117. Bollag, W. and E.E. Holdener, *Retinoids in cancer prevention and therapy*. Ann Oncol, 1992. **3**(7): p. 513-26.
118. Sporn, M.B., *Vitamin A and its analogs (retinoids) in cancer prevention*. Curr Concepts Nutr, 1977. **6**: p. 119-30.
119. Moon, T.E., et al., *Retinoids in prevention of skin cancer*. Cancer Lett, 1997. **114**(1-2): p. 203-5.
120. Kurie, J.M., *The biologic basis for the use of retinoids in cancer prevention and treatment*. Curr Opin Oncol, 1999. **11**(6): p. 497-502.
121. Baj, G., et al., *Retinoids in breast cancer prevention and treatment. A review of the literature*. Eur J Gynaecol Oncol, 2000. **21**(4): p. 411-5.
122. Nutting, C.M. and R.A. Huddart, *Retinoids in the prevention of bladder cancer*. Expert Rev Anticancer Ther, 2001. **1**(4): p. 541-5.
123. Veronesi, U. and A. Decensi, *Retinoids for ovarian cancer prevention: laboratory data set the stage for thoughtful clinical trials*. J Natl Cancer Inst, 2001. **93**(7): p. 486-8.
124. Averbuch, S.D., *Lung cancer prevention: retinoids and the epidermal growth factor receptor-a phoenix rising?* Clin Cancer Res, 2002. **8**(1): p. 1-3.
125. Niles, R.M., *The use of retinoids in the prevention and treatment of skin cancer*. Expert Opin Pharmacother, 2002. **3**(3): p. 299-303.
126. Decensi, A., et al., *Breast cancer prevention trials using retinoids*. J Mammary Gland Biol Neoplasia, 2003. **8**(1): p. 19-30.
127. Mariani, L., et al., *Chemoprevention of breast cancer with fenretinide (4-HPR): study of long-term visual and ophthalmologic tolerability*. Tumori, 1996. **82**(5): p. 444-9.
128. 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*. Journal of Medicinal Chemistry, 2004. **47**(4): p. 999-1007.

129. Benbrook, D.M., et al., *Biologically active heteroarotinoids exhibiting anticancer activity and decreased toxicity*. Journal of Medicinal Chemistry, 1997. **40**(22): p. 3567-3583.
130. 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.
131. Brown, C.W., et al., *Novel heteroarotinoids as potential antagonists of Mycobacterium bovis BCG*. Journal of Medicinal Chemistry, 2004. **47**(4): p. 1008-1017.
132. 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.
133. Liu, Z.F., et al., *In vitro and in vivo metabolism of a sulfur-containing heteroarotinoid SHetA2*. Clinical Cancer Research, 2003. **9**(16): p. 6248s-6248s.
134. 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.
135. Chengeclza, S., T. Lui, and D.M. Benbrook, *Effects of SHetA2 on thymidine phosphorylase expression and NF-kappa B activity*. Molecular Cancer Therapeutics, 2007. **6**(12): p. 3442s-3443s.
136. Liu, T., et al., *Regulation of apoptosis, cell cycle progression and differentiation by SHetA2 in renal cancer cells is associated with alterations in Bcl-2, cyclin D1 and E-cadherin proteins*. Molecular Cancer Therapeutics, 2007. **6**(12): p. 3336s-3336s.
137. 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.
138. Chengeclza, 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.
139. 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.
140. Chun, K.H., et al., *The synthetic heteroarotinoid SHetA2 induces apoptosis in squamous carcinoma cells through a receptor-independent and mitochondria-dependent pathway*. Cancer Res, 2003. **63**(13): p. 3826-32.
141. Le, T.C., et al., *Heteroarotinoids with anti-cancer activity against ovarian cancer cells*. Open Med Chem J, 2007. **1**: p. 11-23.
142. Lin, Y.D., 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*. Molecular Cancer Therapeutics, 2008. **7**(11): p. 3556-3565.
143. Liu, T., et al., *Development of flexible-heteroarotinoids for kidney cancer*. Mol Cancer Ther, 2009. **8**(5): p. 1227-38.
144. Kabirov, K.K., et al., *Oral toxicity and pharmacokinetic studies of SHetA2, a new chemopreventive agent, in rats and dogs*. Drug and Chemical Toxicology, 2013. **36**(3): p. 284-295.

145. 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.
146. Zhang, Y., et al., *High performance liquid chromatographic analysis and preclinical pharmacokinetics of the heteroarotinoid antitumor agent, SHetA2*. Cancer Chemother Pharmacol, 2006. **58**(5): p. 561-9.
147. 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. 209-17.
148. 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. 412-23.
149. 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.
150. Wadhwa, R., et al., *Spontaneous immortalization of mouse fibroblasts involves structural changes in senescence inducing protein, mortalin*. Biochem Biophys Res Commun, 1993. **197**(1): p. 202-6.
151. Wadhwa, R., et al., *Differential subcellular distribution of mortalin in mortal and immortal mouse and human fibroblasts*. Exp Cell Res, 1993. **207**(2): p. 442-8.
152. 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 Res, 2000. **7**(3): p. 229-31.
153. 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. 22239-42.
154. Kaul, S.C., et al., *Malignant transformation of NIH3T3 cells by overexpression of mot-2 protein*. Oncogene, 1998. **17**(7): p. 907-11.
155. Kaul, S.C., et al., *Mouse and human chromosomal assignments of mortalin, a novel member of the murine hsp70 family of proteins*. FEBS Lett, 1995. **361**(2-3): p. 269-72.
156. Ohashi, M., et al., *The gene encoding PBP74/CSA/motalin-1, a novel mouse hsp70, maps to mouse chromosome 18*. Genomics, 1995. **30**(2): p. 406-7.
157. Calloni, G., et al., *DnaK Functions as a Central Hub in the E. coli Chaperone Network*. Cell Reports, 2012. **1**(3): p. 251-264.
158. Westermann, B., et al., *The role of the GrpE homologue, Mge1p, in mediating protein import and protein folding in mitochondria*. The EMBO Journal, 1995. **14**(14): p. 3452-3460.
159. Yehiely, F. and M. Oren, *The Gene for the Rat Heat-Shock Cognate, Hsc70, Can Suppress Oncogene-Mediated Transformation*. Cell Growth & Differentiation, 1992. **3**(11): p. 803-809.
160. Motojima, K. and S. Goto, *Rat-Liver Bip/Grp78 Is down-Regulated by a Peroxisome-Proliferator, Clofibrate*. Febs Letters, 1992. **308**(2): p. 207-210.
161. Wadhwa, R., et al., *Identification and characterization of molecular interactions between mortalin/mtHsp70 and HSP60*. Biochemical Journal, 2005. **391**: p. 185-190.

162. Deocaris, C.C., S.C. Kaul, and R. Wadhwa, *On the brotherhood of the mitochondrial chaperones mortalin and heat shock protein 60*. *Cell Stress & Chaperones*, 2006. **11**(2): p. 116-128.
163. Mizukoshi, E., et al., *Fibroblast growth factor-1 interacts with the glucose-regulated protein GRP75/mortalin*. *Biochem J*, 1999. **343 Pt 2**: p. 461-6.
164. Mizukoshi, E., et al., *Cell-cycle dependent tyrosine phosphorylation on mortalin regulates its interaction with fibroblast growth factor-1*. *Biochem Biophys Res Commun*, 2001. **280**(4): p. 1203-9.
165. Sacht, G., et al., *ATP-sensitive association of mortalin with the IL-1 receptor type I*. *Biofactors*, 1999. **9**(1): p. 49-60.
166. Wadhwa, R., et al., *Mortalin-MPD (mevalonate pyrophosphate decarboxylase) interactions and their role in control of cellular proliferation*. *Biochem Biophys Res Commun*, 2003. **302**(4): p. 735-42.
167. Schwarzer, C., et al., *Voltage-dependent anion-selective channel (VDAC) interacts with the dynein light chain Tctex1 and the heat-shock protein PBP74*. *Int J Biochem Cell Biol*, 2002. **34**(9): p. 1059-70.
168. Schilke, B., et al., *Evolution of mitochondrial chaperones utilized in Fe-S cluster biogenesis*. *Curr Biol*, 2006. **16**(16): p. 1660-5.
169. 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. 1080-4.
170. Wadhwa, R., et al., *Inactivation of tumor suppressor p53 by mot-2, a hsp70 family member*. *J Biol Chem*, 1998. **273**(45): p. 29586-91.
171. Orsini, F., et al., *The life span determinant p66Shc localizes to mitochondria where it associates with mitochondrial heat shock protein 70 and regulates trans-membrane potential*. *J Biol Chem*, 2004. **279**(24): p. 25689-95.
172. Pellegrini, M., S. Pacini, and C.T. Baldari, *p66SHC: the apoptotic side of Shc proteins*. *Apoptosis*, 2005. **10**(1): p. 13-8.
173. Saar Ray, M., et al., *Mortalin/GRP75 binds to complement C9 and plays a role in resistance to complement-dependent cytotoxicity*. *J Biol Chem*, 2014. **289**(21): p. 15014-22.
174. Pilzer, D., et al., *Mortalin inhibitors sensitize K562 leukemia cells to complement-dependent cytotoxicity*. *Int J Cancer*, 2010. **126**(6): p. 1428-35.
175. Mosser, D.D. and R.I. Morimoto, *Molecular chaperones and the stress of oncogenesis*. *Oncogene*, 2004. **23**(16): p. 2907-18.
176. Wadhwa, R., et al., *Upregulation of mortalin/mthsp70/Grp75 contributes to human carcinogenesis*. *Int J Cancer*, 2006. **118**(12): p. 2973-80.
177. Starenki, D., et al., *Mortalin (GRP75/HSPA9) upregulation promotes survival and proliferation of medullary thyroid carcinoma cells*. *Oncogene*, 2014.
178. Deocaris, C.C., et al., *Mortalin sensitizes human cancer cells to MKT-077-induced senescence*. *Cancer Lett*, 2007. **252**(2): p. 259-69.

179. 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. 6818-21.
180. 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. 246-53.
181. 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.
182. Ma, Z., et al., *Mortalin controls centrosome duplication via modulating centrosomal localization of p53*. *Oncogene*, 2006. **25**(39): p. 5377-90.
183. Kaula, S.C., et al., *Inactivation of p53 and life span extension of human diploid fibroblasts by mot-2*. *FEBS Lett*, 2000. **474**(2-3): p. 159-64.
184. Dores-Silva, P.R., et al., *Human Mitochondrial Hsp70 (Mortalin): Shedding Light on ATPase Activity, Interaction with Adenosine Nucleotides, Solution Structure and Domain Organization*. *PLoS ONE*, 2015. **10**(1): p. e0117170.
185. 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.
186. Kaul, S.C., et al., *An N-terminal region of mot-2 binds to p53 in vitro*. *Neoplasia*, 2001. **3**(2): p. 110-4.
187. 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.
188. Laufen, T., et al., *Mechanism of regulation of hsp70 chaperones by DnaJ cochaperones*. *Proc Natl Acad Sci U S A*, 1999. **96**(10): p. 5452-7.
189. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility*. *J Comput Chem*, 2009. **30**(16): p. 2785-91.
190. O'Boyle, N.M., et al., *Open Babel: An open chemical toolbox*. *J Cheminform*, 2011. **3**: p. 33.
191. 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.

APPENDICES

Section 1: The SNARE complex

The SNARE complex is a system that mediates the fusion of vesicles inside the neuron cell with the cell membrane, releasing the neurotransmitter inside of the vesicle to activate sub-stream receptors. The name of SNARE originated from a protein called N-ethylmaleimide-Sensitive Factor (NSF), which is sensitive to the small molecule N-ethylmaleimide. The NSF can bind to a protein called Soluble NSF Attachment Protein (SNAP). And its receptor is named SNAP Receptor (SNARE). The SNAP 25 comes from SynNaptosome Associated Protein of 25 kD. In mouse, the SNARE complex consists of three proteins, Syntaxin 1A, SNAP 25 and Synaptobrevin, all of which contain the SNARE motif. These motifs together form the four helix bundle which pulls close of the vesicle and cell membrane to mediate membrane fusion. This complex is resistant to SDS but can be disassociated by heat (over 95 °C).

All four SNARE motifs (Syntaxin 1A, both motifs of SNAP25, Synaptobrevin 2) from rat were cloned, expressed and purified, respectively. The complex were formed by incubating all purified components at a 1:1:1:1 ratio at 4 °C for 24 hours. We found that the SNARE complex can be formed after the incubation, however, not only the 1st order complex (which consists one component each), the higher order complex (oligomers of the complex) can also be found as indicated in the SDS-PAGE gel. The SNARE complex formation also showed a dynamic balance between the complex and the components. The higher concentration of the components, the more complex formed. However, there were always components left in the sample, no matter reconstitute first or forming complex first, making it hard to get a pure sample of the SNARE

complex to study its structure.

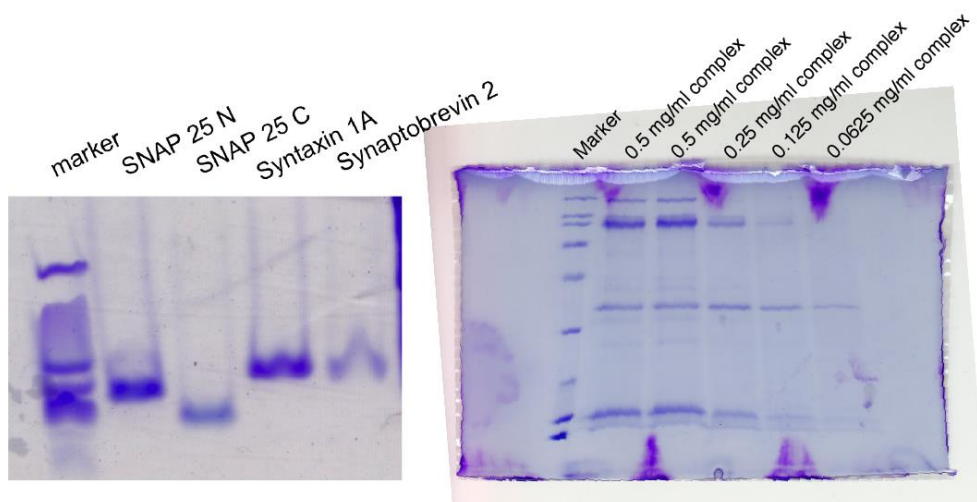


Figure 41 Purification of all four SNARE motifs and the dynamic formation of SNARE complex. On the left, all four SNARE motifs are purified and checked by the SDS-PAGE. On the right, we show that the SNARE complex is in a dynamic balance between the complex and components. With decreasing concentration from left to right, more high-order complex were disassociated. The first order complex remains almost the same concentration (the band in the middle).

Section 2: The aligned NMR

Application of NMR on the solid sample results broad spectra due to the chemical shift anisotropy, which is averaged out by Brownian motion in the liquid. However, in solid, chemical shift anisotropy is dominant. One way to overcome is the Magic Angle Spinning (MAS), in which the sample was placed 54.74° with respect to the external magnetic field. So the anisotropy term was effectively eliminated due to the $(3\cos^2\theta-1)$ will become 0 for the anisotropy. Another way to go around is to align the samples in a certain direction so that the chemical shift anisotropy aligned into one direction and the orientation of the membrane protein can be accessed by a specifically designed pulse sequence. The lipids can aligned automatically under eternal

magnetic field, which makes a perfect system to study membrane proteins. Here, we tried to align the DMPC/DHPC bicelle (at a ratio of $q = \text{DMPC} : \text{DHPC} = 3.5 : 1$) and the sample aligned very well in the magnet after 2.5 hours equilibration. The alignments can be probed by the ^{31}P NMR. The two distinct peaks are coming from the phosphorus belonging to the head group of each lipid molecule, respectively, with about 8 ppm separation (Figure 42). The small peptide Fowlicildin was mixed with the aligned bicelle sample and the same alignments of lipids can be achieved telling by ^{31}P spectrum. However, the ^{13}C signals were relatively low and ^{15}N cannot be found. With either 2 ms or 0.5 ms contact time with 1 k scans. It was reported longer contact time results more signals from mobile region of the peptide and shorter contact time results more signals from rigid regions. However, we haven't been successful on acquiring the lower sensitive spectra so far.

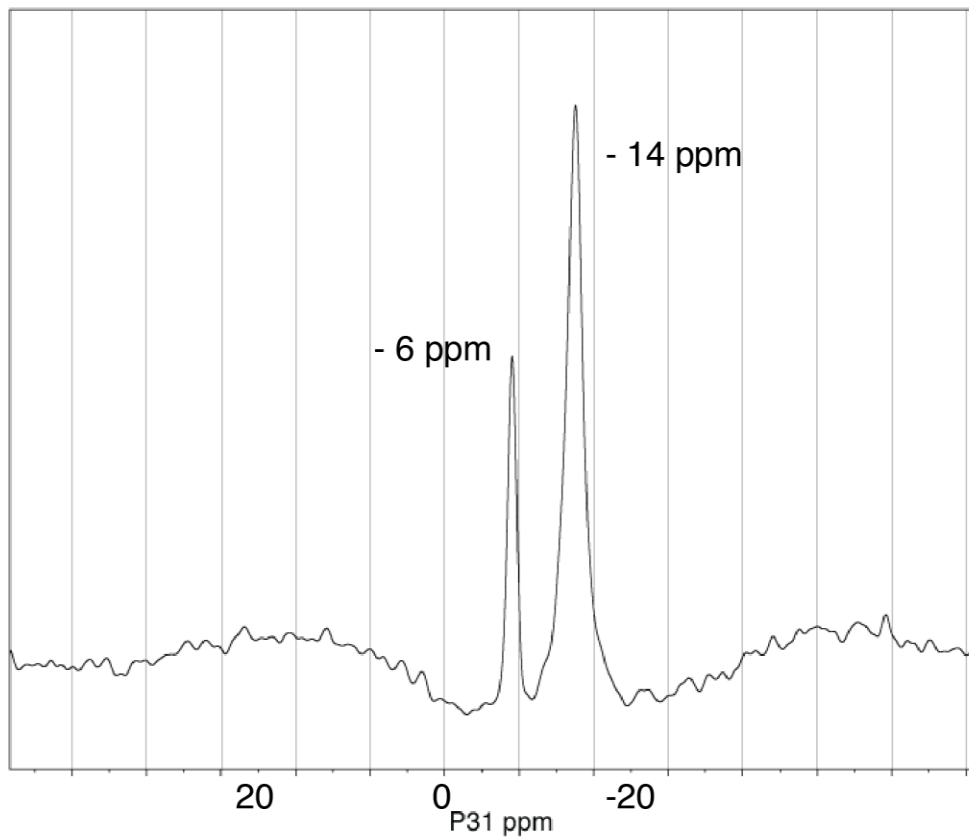


Figure 42 31P 1D spectrum of aligned DMPC/DHPC bicelle sample. The DMPC/DHPC lipids are mixed with $q=3.5$ and aligned under a 400 MHz magnet for 2.5 hours. The distinct two peaks shown in the 31P spectrum are coming from the head group phosphorus from the head group of each lipid, respectively. The 6 ppm peak is from DHPC, which is parallel to the external magnetic field after alignments and 14 ppm peak is from DMPC which is perpendicular to the external magnetic field after alignments.

VITA

Penghui Lin

Candidate for the Degree of

Doctor of Philosophy

Thesis: APPLICATION OF NUCLEAR MAGNETIC RESONANCE TO PROTEIN
STRUCTURE AND PROTEIN-LIGAND INTERACTION STUDIES

Major Field: Photonics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy/Education in
Photonics at Oklahoma State University, Stillwater, Oklahoma in July, 2015.

Completed the requirements for the Master of Science in Pathogen Biology at
Fudan University, Shanghai, China in 2009.

Completed the requirements for the Bachelor of Science in Bioengineering at
Harbin Institute of Technology, Harbin, China in 2006.

Experience:

Teaching/Researching Assistant at Oklahoma State University

Professional Memberships:

American Physical Society