# 中国科学技术大学

# 博士学位论文


## TDP-43 串联 RRM 结构域的核磁片段筛选和 hADK2 的动态结构研究

作者姓名：　Gilbert Nshogoza

学科专业：生物化学与分子生物学

导师姓名：施蕴渝教授　阮科副教授

完成时间：二〇一九年九月二十八日

University of Science and Technology of China

A dissertation for doctor's degree


# NMR application in fragment-based screening against TDP-43 Tandem RRM and in dynamics of hADK2

**Author**：Gilbert Nshogoza

**Speciality**：Biochemistry and Molecular Biology

**Supervisors**：Prof. Yunyu Shi， Prof. Ke Ruan

**Finished time**: Sept 28th, 2019

# 中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文,是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：_____　　签字日期：_____

# 中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

□公开　□保密（____年）

作者签名：_____　　导师签名：_____

签字日期：_____　　签字日期：_____

# 摘要

本论文主要包括两个部分。第一章和第三章介绍了针对 TDP-43 的串联 RNA 识别模体（RRM）的基于 NMR 片段的筛选，第四章介绍了 NMR 弛豫色散和赝接触位移方法在获取人源 ADK2 蛋白质（hADK2）基态到激发态构象交换信息方面的应用。中间部分的第二章介绍了这项工作中使用的方法和材料。

TDP-43 最初是一种核蛋白，但在病理状态下易位至细胞质。 TDP-43 是一种 RNA 结合蛋白，由两个 RRM 结构域（RRM1 和 RRM2）组成。已知 RRM 结构域涉及蛋白质-核苷酸相互作用和蛋白质-蛋白质相互作用，并且会介导应激颗粒的形成。TDP-43 参与神经退行性疾病和癌症的发生发展，是潜在的治疗靶标。针对 TDP-43 串联 RRM 结构域，我们使用了蛋白信号观测和配体信号观测两种方法，来发现新型的小分子苗头化合物。首先对 TDP-43 串联 RRM 结构域使用了基于配体信号观测的核磁片段筛选，确定了三个结合较弱的苗头化合物。针对 $^{15}$N 标记的串联 RRM，以及单独的 RRM2，通过化学位移扰动（CSP）实验确定了这些苗头化合物的结合界面，并在 CSP 指导下通过分子对接（HADDOCK）建立了结构模型。这些苗头化合物主要与 RRM2 结构域结合，这表明 TDP-43 的 RRM2 结构域具有可成药性。这些苗头化合物为后续苗头化合物到先导化合物的演化提供了良好的出发点。

生物大分子具有高度的灵活性，并且会在很大尺度的时间范围内不断发生波动。结构生物学的重点主要是研究占有高比例的基态，而对于那些稀少的、瞬时的状态很难检测，因为它们对于大多数结构方法都是不可见的。但是，这些状态在诸如配体结合，酶催化和蛋白质折叠等生化过程中发挥重要作用。因此，对这些状态的结构和动力学描述对于理解蛋白质功能非常重要。核磁共振（NMR）溶液光谱学可以观察到细菌腺苷酸激酶动力学特性。但是，还缺乏对于人腺苷酸激酶的构象变化和动力学的研究。我们已经在微秒到毫秒的时间范围内，使用 NMR 弛豫弥散（CPMG）来检测 hADK2 激发态构象到基态构象

的转变。同时，我们还用镧系螯合肽标记了 hADK2，这为通过激发态和基态之间的慢交换来测定激发态赝接触位移（PCS）铺平了道路。

**关键词:** 核磁共振，片段筛选，RRM 结构域抑制剂，TDP-43，弛豫弥散，赝接触位移，蛋白质动力学，hADK2。

# ABSTRACT

This thesis contains two main parts. The first and third chapters present the NMR fragment-based screening against Tandem RNA Recognition Motifs (RRMs) of TDP-43, and the fourth chapter describes the application of NMR relaxation dispersion and Pseudocontact shifts methods for getting information about the ground-state to excited-states exchange in human ADK2 (hADK2). In the midst, the second chapter describes the methods and materials used in this whole work.

TDP-43 is originally a nuclear protein but translocates to cytoplasm in the pathological condition. TDP-43, as an RNA-binding protein, consists of two RRMs (RRM1 and RRM2). RRMs are known to involve both protein-nucleotide and protein-protein interactions and mediate the formation of stress granules. Thus, they assist the entire TDP-43 protein to participate in neurodegenerative and cancer diseases. Consequently, they are potential therapeutic targets. Protein-observed and ligand-observed nuclear magnetic resonance (NMR) spectroscopy were used to uncover the small molecule inhibitors against the tandem RRM of TDP-43. We identified three hits weakly binding the tandem RRMs by using the ligand-observed NMR fragment-based screening. The binding topology of these hits is then depicted by chemical shift perturbations (CSP) of the $^{15}$N-labeled tandem RRM and RRM2, respectively, and modeled by the CSP-guided High Ambiguity Driven biomolecular DOCKing (HADDOCK). These hits mainly bind to the RRM2 domain, which suggests the druggability of RRM2 domain of TDP-43. These hits also facilitate further studies regarding the hit-to-lead evolution against TDP-43 RRM domain.

Biological macromolecules are highly flexible and continually undergo conformational fluctuations on a broad spectrum of timescales. The focus of structural biology is mostly on studies of the highly populated, ground states of biological molecules. The states that are only sparsely and transiently populated are more difficult to probe because they are invisible to most structural methods. Yet, such states can play critical roles in biochemical processes such as ligand binding, enzyme catalysis, and protein folding. A description of these states in terms of structure and dynamics is, therefore, of great importance for understanding the protein functions.

Abstract

The advances in solution nuclear magnetic resonance (NMR) spectroscopy allowed the bacterial Adenylate Kinases dynamics to be observed with detail. However, nothing is known on conformational changes and dynamics of human Adenylate kinases. We have used the NMR relaxation dispersion (CPMG) to detect the hADK2 excited-state conformations exchanging with the ground state in the timescale of microsecond to millisecond. We also labeled hADK2 with lanthanide-chelating peptides, which paves way for the following experimental observation of Pseudocontact Shifts (PCSs) of excited-states in slow exchanging with the ground-state.

**Keywords:** Nuclear magnetic resonance; fragment-based screening; RRM domain inhibitor; TDP-43; Relaxation Dispersion, Pseudocontact Shifts, Protein dynamics, hADK2.

# ABBREVIATIONS

| | |
|---|---|
| µl | Micro liter |
| µM | Micro molar |
| AD | Alzheimer's disease |
| AD1 | aminoacridine derivative 1 |
| ADK/AK | Adenylate Kinases |
| ALS | Amyotrophic lateral sclerosis |
| ATG7 | Autophagy-related protein 7 |
| ATP | Adenosine Tri-phosphate |
| BMRB | Biological Magnetic Resonance Data Bank |
| CDK2/6 | cyclin-dependent kinase 2/6 |
| CF | Cystic Fibrosis |
| CM | Curcumin |
| CPEB1/4 | Cytoplasmic Polyadenylation Element-binding 1/4 |
| CPMG | Carr-Purcell-Meiboom-Gill |
| CPMG | Carr - Purcell - Meiboom - Gill |
| CSM | chemical shift mapping |
| CSPs | Chemical shift perturbations |
| DNA | Deoxyribonucleic acid |
| dNTP | deoxyribonucleoside triphosphate |
| eIF | early initiation factors |
| ELAV | Embryonic Lethal Abnormal Vision |
| FBLD | Fragment-based lead discovery |
| FBS | Fragment-based screening |
| FMRP | Fragile X Mental Retardation Protein |
| FTLD | Frontotemporal lobar degeneration |
| FUS | Fused in sarcoma |
| G3BP | Ras-GAP SH3 domain-binding protein |
| GLUTs | glucose transporters |

| | |
|---|---|
| GUK | Guanylate kinase |
| hADK | human Adenylate Kinase |
| hADK2 | human Adenylate Kinase 2 |
| HCC | hepatocellular carcinoma |
| HD | Huntington's disease |
| HDAC6 | Histone deacetylase 6 |
| hNFL | human low molecular weight neurofilament |
| hnRNPs | Heterogeneous nuclear ribonucleoproteins |
| HSQC | Heteronuclear single-quantum coherence |
| IPTG | Isopropyl β-d-1-thiogalactopyranoside |
| ITC | Isothermal titration calorimetry |
| LB | Luria-Bertani |
| LBT | Lanthanide binding tag |
| lncRNA | Long non-coding RNA |
| MALAT1 | Metastasis-associated-in-lung-adenocarcinoma-transcript-1 |
| mL | Milliliter |
| mM | Millimolar |
| NES | Nuclear export signal |
| NLS | Nuclear localization signal |
| NMD | nonsense-mediated mRNA decay |
| NMP | Nucleotide Mono phosphate |
| NMPK | Nucleoside monophosphate kinase |
| NMR | Nuclear Magnetic Resonance |
| NSCLC | non-small cell lung cancer |
| NTP | Ribonucleoside triphosphate |
| OD | Optical density |
| PABP | PolyA-Binding Protein |
| PCR | Polymerase chain reaction |
| PCS | Pseudocontact Shift |
| PD | Parkinson's disease |

| | |
|---|---|
| PMDs | protein-misfolding disorders |
| PRE | Paramagnetic Relaxation Enhancement |
| RBD | RNA binding domain |
| RNA | Ribonucleic acid |
| RNP | Ribonucleoproteins |
| RRM | RNA Recognition Motifs |
| SDS-PAGE | Sodium lauryl sulfate-polyacrylamide gel electrophoresis |
| SGs | Stress granules |
| SMN | Survival Motor Neuron |
| SNHG12 | Small nucleolar RNA host gene 12 |
| SPR | Surface Plasmon Resonance |
| SRSF3 | serine/arginine-rich splicing factor |
| STD | Saturation Transfer difference |
| STS | Staurosporine |
| *TARDBP* | Transactivation response DNA-binding Protein gene |
| TDP-43 | Transactive response DNA-binding Protein 43 |
| TNBC | Triple-negative breast cancer |
| TRIM16 | Tripartite motif-containing protein 16 |
| UMP-CMPK | uridylate-cytidylate kinase |
| UTR | Untranslated region |

# TABLE OF CONTENT

Table of content

Table of content

# Chapter 1 NMR fragment-based screening against Tandem RNA Recognition Motifs of TDP-43

## 1 Introduction

### 1.1 RNA recognition motifs (RRMs)

#### 1.1.1 RRM Function

RNA recognition motifs (RRMs) play diverse roles in post-transcriptional gene expression events such as RNA transport, localization, stability, and mRNA and rRNA processing. RRM is also known as the ribonucleoproteins (RNP) domain, as it contains the short and conserved elements RNP1 and RNP2, or RNA binding domain (RBD) are abundantly distributed in higher vertebrates [1]; and ubiquitously found in all kingdoms of life, including viruses and prokaryotes. In addition, they also participate in important functions such as microRNA biogenesis, apoptosis, and cell division [2,3]. RRMs are not only known to involve in protein-nucleotide interactions but also in protein-protein interactions[4].

#### 1.1.2 RRM Structure

In general, RRM domain is approximately composed of 90 amino-acids with β1α1β2β3α2β4(four β-strand packed together with two α-helix) as a canonical arrangement. There are five loops with different length, like special long Lys-rich loop2 (“KK” loop) in FUS RRM [5], between secondary structural elements. Thus, many structural variations have been discovered according to that scaffold, for example where either loop 5 forms the short a β-hairpin, or β-sheet extends the surface to five or six β-strands, in case of RRM1 of CPEB1 and CPEB4 [6]. In addition, the length of β-strands and α-helix can change as well, whereas the $U2AF^{35}$ RRM contains α1-helix three times longer than in canonical RRM [7]. Moreover, the additional of secondary structure elements at RRM extremities has been also revealed. In PTB

RRM2 and 3 for example, the β5-strand is added to extend the β-sheet surface at the C-terminus while β0-strand is added at N-terminus in CPEB1. Additional α-helix is also frequently present, for example α0-helix at N-terminus in Prp24 RRM or α3-helix in La at C-terminus [2] (**Figure 1.1**).


**Figure 1.1.Schematic representation of secondary structure elements of RRM and its numerous variations.** Canonical elements (blue), additional secondary stucture elements (Red)

Structurally, TDP-43 tandem RRMs are approximately 160 amino acids long and display β1α1β2β3α2β4 arrangement of secondary structure, with an additional β-hairpin named **β3'β3''**[8] or **β5**[9,10] which located between α2β4, and extends the β-sheet surface accessible to multiple RNA nucleotides binding **Figure 1.2**[2].

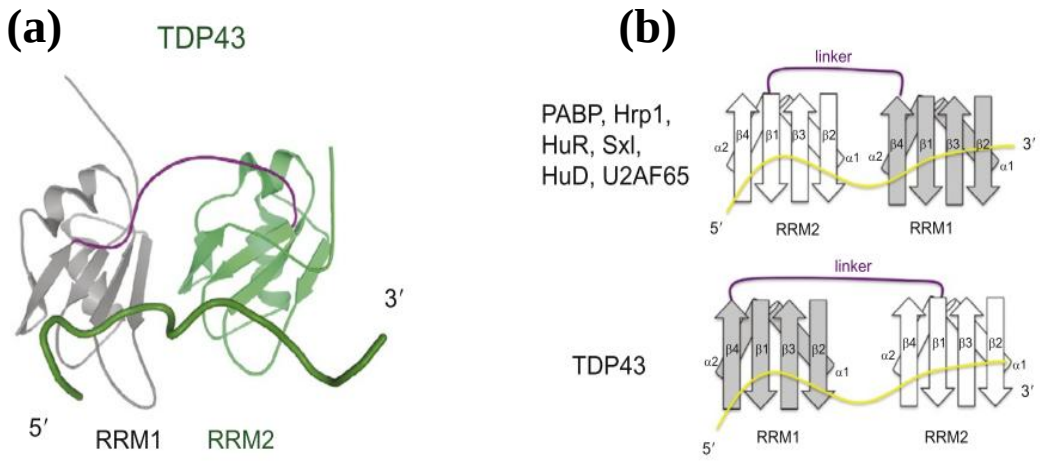


**Figure 1.2**. (**a**) Structure of TDP-43 tandem RRM in complex with RNA. (**b**) schematic representation to illustrate the two modes of RNA binding based on directionality of RNA binding by tandem RRM[8].

### 1.1.3 RNA Recognition mode by tandem RRM of TDP-43

Tandem RRMs and RNA complex structures disclosed enormous diversity in the the mode of RNA recognition. This is mainly due to the diversity in the arrangement of tandem RRM which directly impacts on the RNA-binding mode. As shown in **Figure 1.2**, the directionality of the RNA binding is conserved, inmost cases (Sxl, PABP, Hrp1, HuR, HuD, and Nucleolin) except TDP-43, whereas RRM2 recognizes the RNA 5'-end while the RNA 3'-end binds the RRM1. In their structures, RRMs lie side-by-side and make an extended surface on which RNA binds in linear and extended manner and their interdomain interactions make the complex structure more stable. However, the Nucleolin mode of binding is different from others as the two RRMs lie face-to-face for accomodating the RNA (**Figure1.3b**)[2,8].

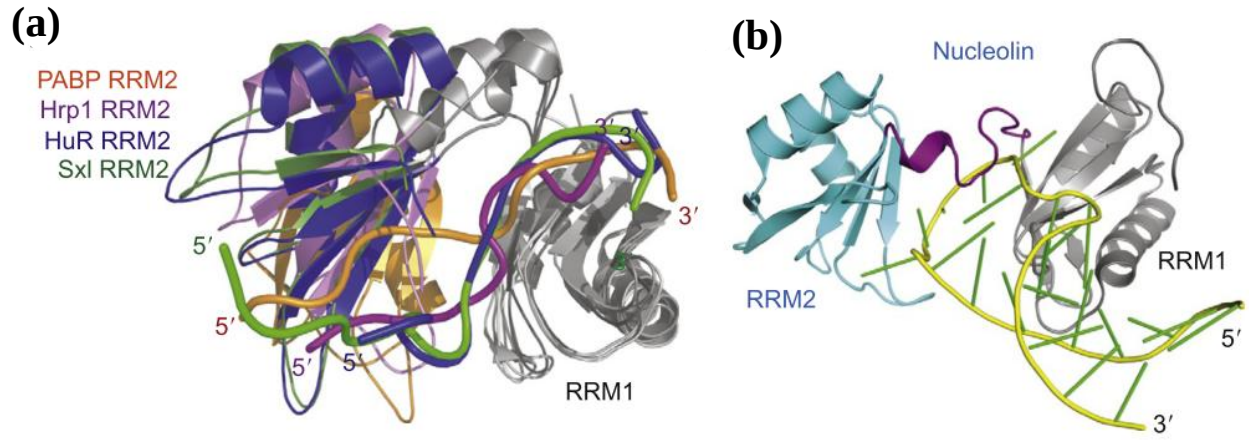


**Figure 1.3. RNA Recognition mode by tandem RRM-containing proteins.(a)**Overlay of PABP, Hrp1, HuR, and Sxl tandem RRMs in complex with RNA.RRM2 of PABP, Hrp1, HuR, and Sxl are shown in orange, magenta, blue, and green, respectively. RNA is shown in tube representation in the same color as RRM2 for individual structures. (**b**) Structure of Nucleolin RRMs bound to RNA[2].

Diverse studies revealed that TDP-43 tandem RRMs can interact with both short and long single-stranded nucleic acids rich in UG/TG either separately or collectively for achieving high affinity and specificity [8-10]. Given the RNA recognition mode by tandem RRMs, TDP-43 RRMs are independent of each other in unbound form but they establish a rigid structure upon RNA binding on flat surface β-sheet [8].

The AUG12 RNA binds on β-sheet surface in 5'to'3' direction from RRM1 to RRM2 (**Figure 1.2**). Spatially, in the common seen arrangement of other tandem RRMs, except in TDP-43, β2 strand of C-terminal RRM2 is closer to the β4 strand of the N-terminal RRM1 (β4β2 type), therefore the short linker of 10-aa brigdes only two β strands (**Figure 1.2**). On contrary, in TDP-43 the RNA binding direction and RRM orientation are done in reversed way (β2β4 type). Thus, the longer linker of 14-aa needs to bridge four β strands and embraces an extended conformation (**Figure 1.2**)[8].

In general, the RNA-recognition pocket is much shallower than the ATP-binding sites of kinases. It hence poses a grand challenge for conventional high throughput screening aimed at discovering strong binders. Conversly, fragment-based approach has been proven fruitful to uncover the initial hits, albeit at weak affinities.

## 1.2 TAR DNA-binding Protein 43kDa (TDP-43)

### 1.2.1 TDP-43 Structure

Human TDP-43 was first isolated during a search for novel transcriptional inactivators binding to the TAR DNA element of the HIV-1 virus, origin of the name: TAR DNA-binding Protein with molecular weight of 43 kDa approximately [11].

*TARDBP* is the gene encoding TDP-43, located on chromosome 1 and composed of six exons. The exon 1 is non-coding and its function is unknown; the rest four (2-6) exons are protein coding. *TARDBP* gene is evolutionarily highly preserved as proteins' sequences comparison of different species (human, mouse, *Drosophila melanogaster*, and *Caenorhabditis elegans*) showed a high degree of amino acids conservation. Human TDP-43 mRNA has been shown to be expressed in all analyzed tissues with different levels by using Northern blot and mostly present in pancreas, placenta, ovary, lung, testis, kidney, and spleen [12].

As illustrated in **Figure 1.4**, TDP-43 has five functional domains: two RNA recognition motifs (RRM1 and RRM2) with a considerable homology from human, mouse, worm, and fly species. These RRM domains have two highly conserved hexameric ribonucleoproteins 2 (RNP2), and octameric ribonucleoproteins 1 (RNP1) motifs, mainly involved in DNA and RNA binding [13]. In addition, It has a nuclear localization (NLS) domain and export signals (NES) facilitating its movement from nucleus to the cytoplasm or vice-versa [14], and C-terminal glycine rich region known for mediating the protein-protein interactions [15]

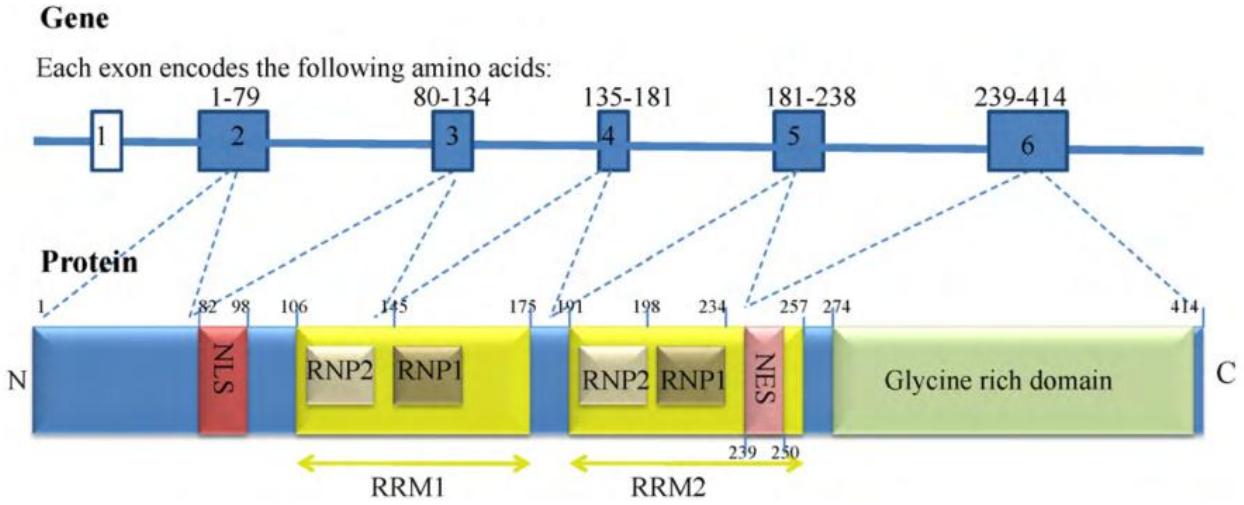


**Figure 1.4.** Schematic diagram of the *TARDBP* gene and TDP-43 protein.[16]

### 1.2.2 TDP-43 Function

Transactive response DNA-binding Protein 43kDa (TDP-43) is a RRM-containing protein, which plays important functions in mRNA metabolism regulation including transcription repression, exon skipping and RNA splicing [17,18]. TDP-43 functions as a transcription factor and it has been shown to bind to TAR DNA for repressing the HIV-1 transcription [11] and also to the SP-10 gene promoter required for spermatogenesis. The TDP-43 was shown to be involved in exon skipping and interact with other hnRNP family proteins for splicing inhibitory activity. In this regard, the TDP-43 has been reported to associate with SC35 and hnRNP A2 known as spicing regulator proteins [15,16].

TDP-43 binds specifically to UG-rich sequences at 3'-splice site of intron 8, thus it promotes CFTR exon 9 skipping andproduces an inactive CFTR protein in CF patients [12]. TDP-43 stabilizes the mRNA of human low-molecular-weight neurofilament (hNFL) by a direct interaction with the 3'UTR[19]. Depletion of TDP-43 has important consequences in essential metabolic processes in human cells like nuclear shape deformation, apoptosis and misregulation of the cell cycle through the control of cyclin-dependent kinase 6 (cdk6) expression[20].

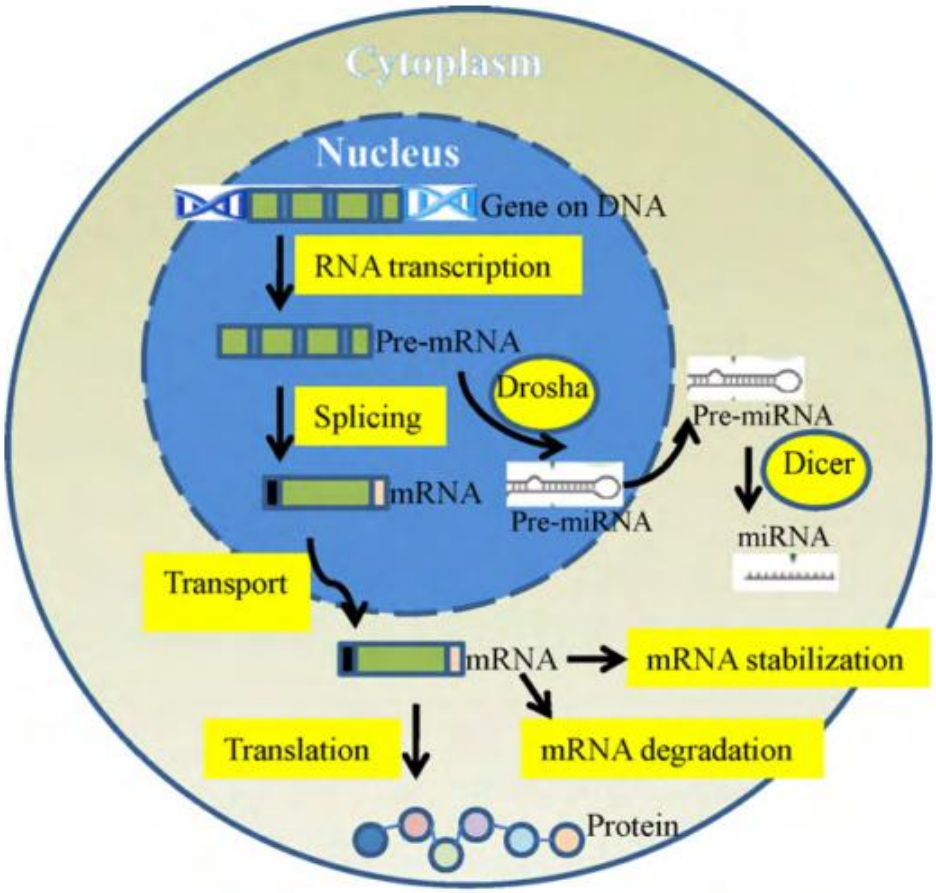


**Figure 1.5. Biological functions of TDP-43**. TDP-43 plays different roles in RNA processing (highlighted in yellow) in nucleus and cytoplasm [16]

The disruption of TDP-43 auto-regulation impacts both its localization and its level, which results in TDP-43 accumulation in cytoplasm. Based on its crucial roles in RNA processing, dysfunctional TDP-43 causes some abnormalities in alternative mRNA splicing, miRNA biogenesis, and RNA-rich granules formation (**Figure 1.6**)[21].

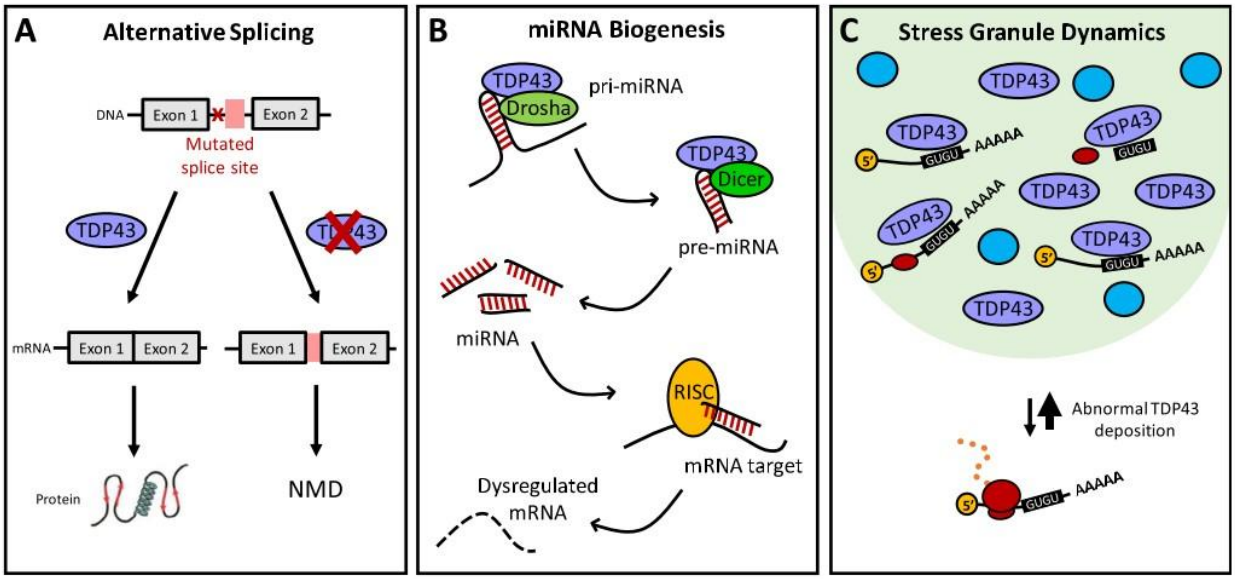


**Figure 1.6**.**TDP-43 deposition impacts RNA stability through several pathways. (A) Alternative splicing**. Mutations that introduce novel splice sites can lead to the inclusion of unannotated or ‘‘**cryptic**” exons (pink box). These faulty transcripts are often targeted by nonsense-mediated mRNA decay (NMD). Typically, TDP-43 is a strong repressor of these unannotated splicing events, but nuclear exclusion prevents TDP-43 from performing this function, and abnormal transcripts accumulate. **(B) miRNA biogenesis**. TDP-43 promotes several steps of miRNA biogenesis, and regulates the formation of key miRNAs that, in turn, control the stability and translation of mRNAs that are essential for neuronal survival, growth and development. **(C) Stress granule dynamics**. TDP-43 is one of several RNA-binding proteins (blue circles) that localize to SGs (light green) in response to various conditions. Because TDP-43 recognizes thousands of GU-rich transcripts, cytoplasmic TDP-43 deposition within SGs forces mRNA recruitment to SGs, shifting transcripts from actively translating polysomes to inert, though stable, SGs[21].

Furthermore, TDP-43 has been appeared to be enhanced in the embryonic and pupal stages of *Drosophila* and oocytes and first larval developmental stages of *C. elegans*[13]. Moreover, TDP-43 has a critical role in cell survival probably through maintaining genomic stability thereby demonstrated by its knockdown in human cells which caused the cell cycle disruption, dysmorphic nuclear shape and increased apoptosis [20].

### 1.2.3 TDP-43 and Stress Granules

Stress granules (SGs), processing bodies (P-bodies) and germ cell (or polar) granules are all considered as cytoplasmic RNA granules, cytoplasmic structures which represent physiological accumulations on mRNA and ribonucleoproteins that regulate gene expression through modulation of translation, trafficking and stability[22].

SGs, membrane-less organelles with a cytoprotective role, are in dynamic equilibrium between polysomes and P-bodies, both mRNA degradation and microRNA-mediated translational arrest take place in the stressful conditions [23,24]. The oxidative stress, osmotic and heat shocks, viral infection and proteasome inhibition are the types of the insults which have been elucidated to induce the SGs formation, as well as the molecular mechanism and signalling pathways which trigger these structures formation [25].

In cellular insults condition, several RNA-binding proteins (RBPs) with the mRNA are known to be the components of the SGs. Due to their prion-like domains which facilitate the aggregation of other RBPs/proteins in granules together with their target mRNAs, T cell-induces antigen 1 (TIA-1) and TIA-related (TIAR)are the main RBPs responsible for SG assembly. There other several components of SGs including PolyA-Binding Protein 1 (PABP-1), Ras-GAP SH3 domain-binding protein (G3BP), Staufen, Survival Motor Neuron (SMN), the Embryonic Lethal Abnormal Vision (ELAV) family member of HuR antigen (HuR), and Fragile X Mental Retardation Protein (FMRP) RBPs, together with ribosomal 48S pre-initiation complex, early translation initiation factors, microRNA-associated Argonaute proteins, p54/Rck helicase, XRN1 exonuclease and cytoskeletal proteins [24].

The fact that TDP-43 is RNA-binding protein, also shares so many features with other hnRNPs, and originally known as a nuclear protein but translocates to cytoplasm upon a pathological condition[14], it is reasonable to suspect that it may also play a role in cellular stress responses via SGs. Indeed, TDP-43 is able to assemble into stress granules (SGs), ribonucleoproteins (RNP) complexes, and cause the temporal arrest of protein synthesis, as protective response to oxidative stress and to environmental insults of different types. The deletion of the C-terminal region of amino acids residues 268-315 [26] or 274-414 [27] and the

RNA-recognition motif 1 domain proved to prevent the recruitment of TDP-43 into SGs; indicating TDP-43 implication in SGs assembly as a specific component [26,27]. Furthermore, the RNA binding domains (RRM1 and 2) in TDP-43 also contribute to the regulation of inclusions formation whereby the deletion of one or both RRMs leads to absence of inclusion formation under stress conditions [28]. Consistently, via a global proteomic approach to find out the TDP-43 interacting proteins, analysis of the TDP-43 interactors revealed an extensive interaction with the proteins which associate with RNA including the hnRNPs, RNA helicases, splicing factors, translation regulatory proteins as well as the proteins involved in mRNA transport and stability. Notably, TDP-43 was found to strongly co-localize with the two stress granule proteins (EIF4G and G3BP1)[29] and later with TIA-1, a primary SG protein and elF3 [28] in discrete cytoplasmic granules. In the same line, TDP-43 and hnRNP A2 are localized to stress granules due to oxidative stress, heat shock and exposure to thapsigargin[30].This clearly indicated that cytoplasmic TDP-43 associates with stress granules and participate in SG pathway.

Two different mechanisms how TDP-43 is recruited to SG can be envisaged: First, sinceTDP-43 protein has RNA binding motifs; it is conceivable that it is recruited into SG via bound mRNAs. Second, protein-protein interactions with other SG-associated proteins could be involved. RNA binding plays a crucial role for SG recruitment ofTDP-43, since deletion mutants lacking the principal RNA binding domain (RRM1) of TDP-43 showed poor recruitment to SG. This correlation between RNA binding and SG recruitment suggests that TDP-43 might be recruited into SG through binding to UG-rich RNA sequences (see model in **Figure 1.7**), although we cannot exclude that protein-protein interactions mediated by RRM1 domain are involved as well[27].

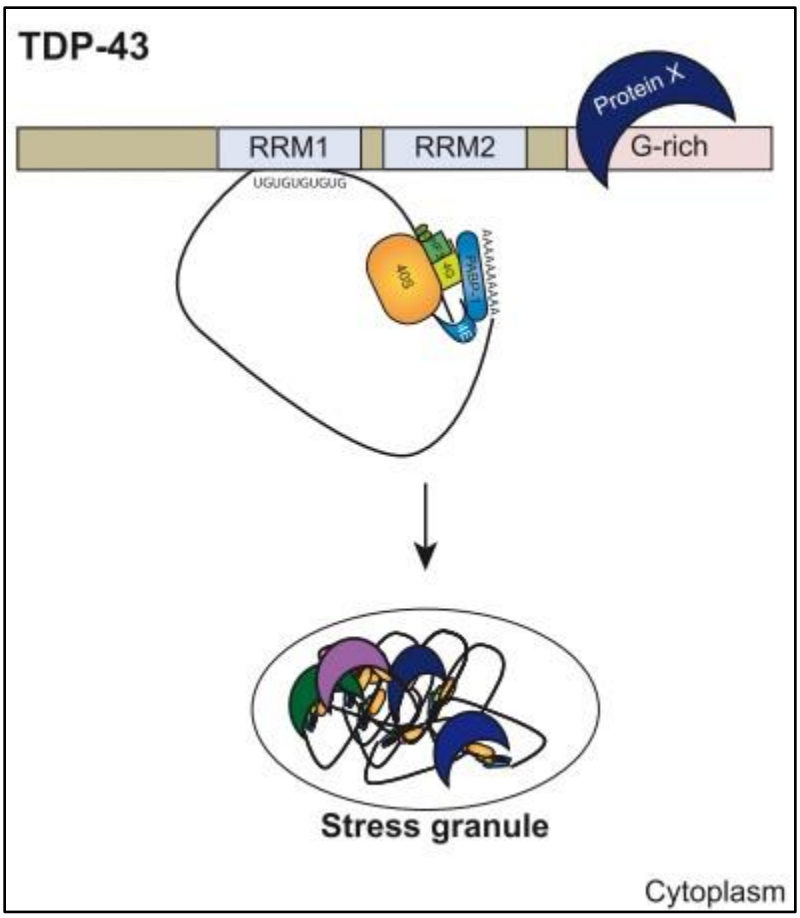


**Figure 1.7.** Upon cellular stress, translation of mRNAs is arrested and translationally silent pre-initiation complexes that contain mRNA, the small ribosomal subunit (40S), early initiation factors (e.g. eIF3, eIF4A, eIF4G) and PABP-1 are packaged into stress granules (SG). Bentmann group suggested that recruitment of TDP-43 into SG involves both protein-RNA and protein-protein interactions. TDP-43 binds to UG- rich mRNA sequences via its RNA binding domain (RRM1) and thus might be recruited into SG via its associated mRNAs. Since additional domains that did not show binding to UG-rich RNA, also contribute to SG recruitment of TDP-43, additional protein-protein interactions with proteins X are involved in SG recruitment of TDP-43[27].

After reporting the recruitment of TDP-43 to SGs, it was worthy to determine the functional role for TDP-43 in SGs and the cellular stress response. One group suggested that TDP-43 contributes to the cellular response to acute stress. Specifically, endogenous TDP-43 is recruited into SGs that are considered to be crucial elements in the protective response to cellular stress. Additionally, TDP-43 participates in regulating SGs whereby the reduced TDP-43 delays SG nucleation and secondary aggregation via the differential deregulation of important nucleating factors TIA-1 and G3BP at the mRNA level. Moreover, the number and size of TIA-1- and G3BP-positiveSGs are reduced in cells depleted of TDP-43 and subsequently treated with

oxidative stress (**Figure1.8**). Thus, TDP-43 contributes positively to SG assembly and their maintenance, as well as cellular survival following oxidative stress[30].

Additionally, cytoplasmic localization of TDP-43 protein may drive inappropriate interactions within SGs. Subsequent modification of SG homeostasis, like excessively promiscuous SG-mRNA binding or inappropriate persistence of SGs, may overwhelm regulatory processes and lead to intractable aggregation which represents a putative cytosolic gain-of-function toxicity pathway [31]. Recently, the study showed that cytosolic TDP-43 has multiple fates. Whereas a TDP-43 subpopulation is indeed recruited to SGs, mature aggregated TDP-43, produced with aggregate-prone TDP-43 variants or exposure to oxidative stress, generates distinct TDP-43 inclusions that are surprisingly devoid of SGs. In support of this observation, they found that SG components are predominantly excluded from TDP-43 pathology in ALS patient motor neurons. They suggested that pathological TDP-43 undergoes a conformational transition which disrupts its interaction with SGs, leading to the evolution of a distinct ALS inclusion pathology[32].

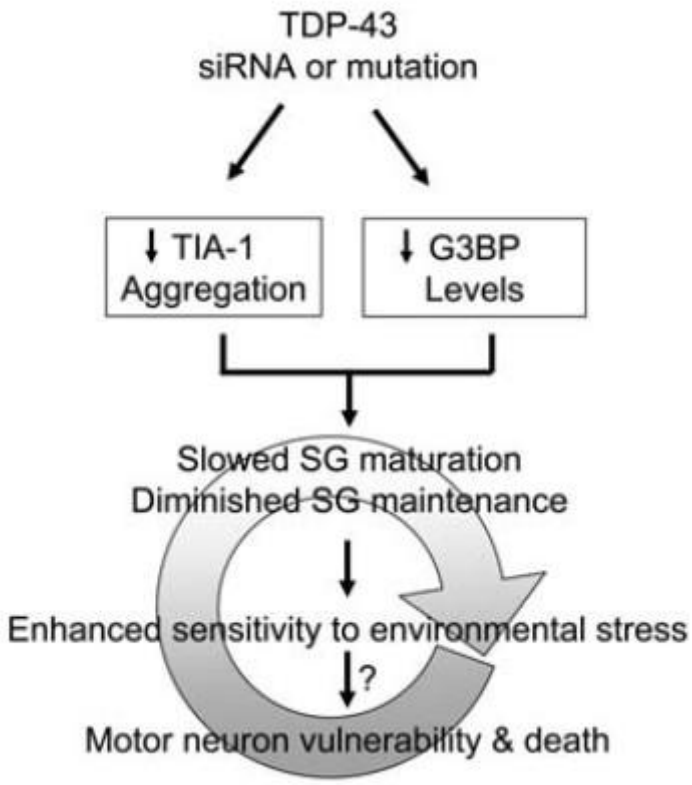


**Figure 1.8**.**Model of TDP-43 in regulation of SGs.** Reduced TDP-43 protein levels or TDP-43 mutations yield a reduction in G3BP levels and disrupt TIA-1 aggregation. These events yield slowed and diminished SG formation and poor maintenance. This may increase cellular susceptibility to acute stress stimuli and contribute to cellular death. This could set up a feed-forward amplification loop resulting in a maladaptive state in motor neurons, thereby contributing to an increased vulnerability over time [30].

### 1.2.4 TDP-43 involvement in diseases

#### a. TDP-43 and neurodegenerative diseases

The dysregulation of TDP-43 is hence associated with a variety of human diseases, especially neurodegenerative diseases, e.g., frontotemporal lobar degeneration (FTLD),amyotrophic lateral sclerosis (ALS), brain ischemia, aging and Alzheimer's disease (AD)[33-35]. ALS is an adult-onset disease characterized by the loss of both upper (from brain into the spinal cord) and lower (from spinal cord out to the lower muscles) motor neurons leading to progressive paralysis and death, within 3-5 years approximately of diagnosis. FTDL is the second most common form of dementia after AD, characterized by behavioral, language and emotions dysfunctions with manifestation of movement disorders in some individuals as early

symptoms[15]. For instance, in cases of FTLD and ALS, TDP-43 is the main constituent of their ubiquitin inclusions [36].

During the stress conditions, TDP-43 is localized in cytoplasm with mRNA binding to its RRM and glycine-rich domain, thus forms the isolated liquid compartment enriching mRNA and proteins. Such stress granules (SGs) in cells and pathological brain tissue play crucial roles in FTLD/ALS pathology [27,28]. Aggregate-prone TDP-43 variants or exposure to oxidative stress generates distinct TDP-43 inclusions devoid of SGs[32].

The TDP-43 immunoreactive histopathology, as seen in glia and neurons, has been shown to be involved in different related neurodegenerative diseases such as Alzheimer's disease (20-50%), hippocampal sclerosis (70%), Picks disease (33%), Parkinson's disease (7%), and in few cases of agryrophilic grain and Huntington's disease[14,37,38].The positive inclusions of TDP-43 are either independent or partially colocalize with the characteristic inclusions (β-amyloid, tau, α-synuclein, and expanded polyglutamines) that are found in these diseases (**Figure 1.9**)[16,38].

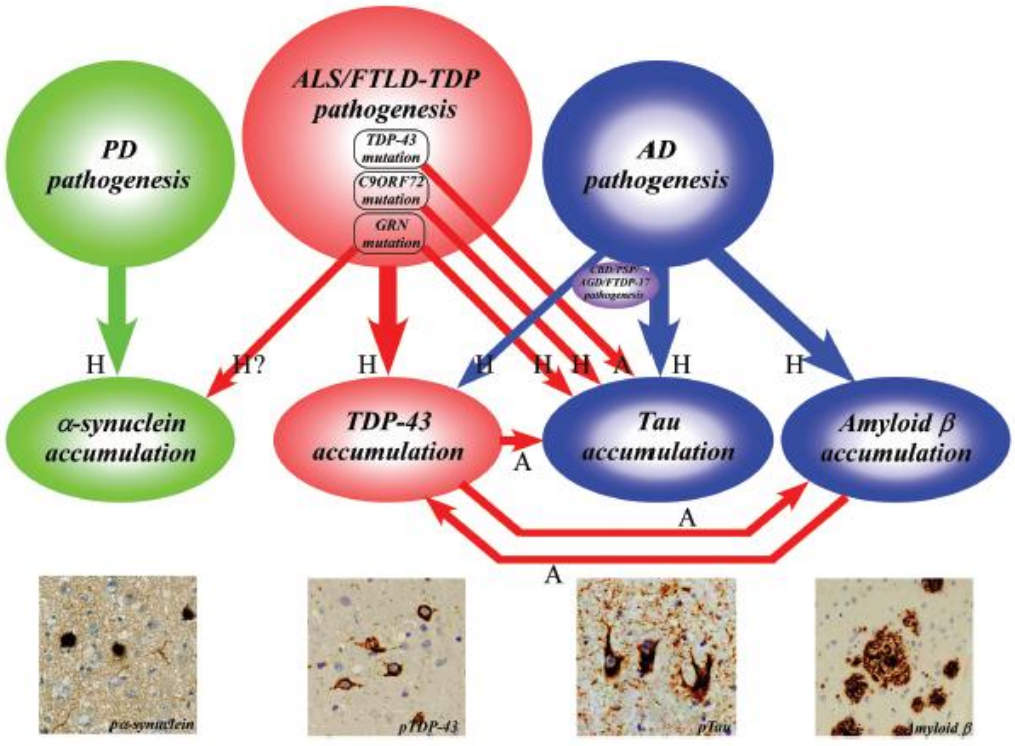


**Figure 1.9.**Description of the relationship between pathogenesis and the accumulation of proteins (including TDP-43) involved in degenerative diseases[39].

In the fundamental recommendations of the Strong criteria with respect to neuropathological diagnosis of ALS-FTSD, the neuronal cytoplasmic and nuclear inclusions within degenerating motor neurons in ALS can be composed of the different range of cytoskeletal proteins and RNA binding proteins. There are three major FTLD types depending on the hallmark pathological protein: FTLD-tau, FTLD-TDP and FTLD-FUS. The large majority of ALS cases with frontotemporal dysfunction belongs to the FTLD-TDP type and show TDP-43 immunoreactive inclusions within a range of neocortical and subcortical structures [40].

A plenty of evidences proved the involvement of TDP-43 in Alzheimer disease. The abnormal TDP-43 immunoreactivity has been commonly found in AD. The TDP-43 neuronal and glial inclusions estimation is nearly 25-30% of sporadic AD cases. After performing the immunohistochemistry with specific antibodies to pathologic forms of TDP-43, its higher frequency of pathology has been detected in AD. In addition, different studies suggested that presence of TDP-43 is associated with greater brain atrophy, particularly hippocampal atrophy, which continued to be greater in AD with TDP-43 pathology compared to AD without TDP-43 pathology [33].

Josephs team conducted a study to determine whether the TDP-43 independently has any clinical and neurological impact on the Alzheimer disease pathology. After accounting for age, apolipoprotein ε4, and other pathologies, TDP-43 showed a strong effect on cognition impairment, memory loss, and medial temporal atrophy in AD and hence an important contributor to the AD phenotype. By comparison, TDP-positive subjects were 10 times likely to be cognitively impaired at death than TDP-negative subjects. Additionally, the great cognitive impairment and medial temporal atrophy were correlated with great TDP-43 burden and more extensive TDP-43 distribution [35].

The toxicity of the TDP-43 overexpression requires the presence of functional RNA Recognition motifs [41-43]. Recently, the proteinopathy of both important mutations (D169G and K263E located at RRM1 and RRM2, respectively) was computationally explored and the mutants are more prone to aggregation and causing the neurological disorders [44]. The RNA-binding interruption may prevent RNA from being irreversibly sequestered in the aggregates.

Moreover, disruption of RNA-binding capacity may hamper further protein association by preventing RNA-nucleation of protein interactions, and hence may become one of therapeutic approaches for the neurodegenerative diseases.

### b. TDP-43 involvement in Cancer diseases

Although the TDP-43 involvement in neurodegenerative diseases is widely exploited, the study in cancer is still rare and recent. Some studies suggested that TDP-43 is a cancer responsive factor. Curcumin (CM) is a natural polyphenol compound isolated from *Curcuma longa* L which possesses different pharmacological effects including anticarcinogenic activities against a variety of tumors like human breast carcinoma [45,46].In human cells, the increased apoptosis upon TDP-43 loss is mediated via the retinoblastoma protein pRb pathway [20], while the CM is known to induce hyperphosphorylation of pRb[47]. Thus, TDP-43 down-regulation as response to CM treatment may activate pRb pathway and induce apoptosis in MCF-7 cells, and consequently TDP-43 positively contributes to the anticancer activity of curcumin [48]. Additionally, Staurosporine (STS) is used to induce apoptosis in U87 glioma cells and resulted in rapidly downregulating both mRNA and protein levels of TDP-43. However, blocking activity of Caspases significantly abolished the STS induced TDP-43 downregulation. More importantly, the STS cytotoxicity was dramatically enhanced by knockdown of TDP-43. Collectively, the normal levels of TDP-43 might be a crucial protective factor for cancer cells under apoptotic insult[49]. Moreover, the TDP-43 has been identified as novel protein binding partner of TRIM16 and prognostic indicator in both neuroblastoma and breast cancer patients. Therefore, TDP-43 binding to TRIM16 implicated as a tumor suppressor in inhibiting the viability and proliferation of neuroblastoma and breast cancer cells [50].

On the contrary, the TDP-43 inhibition suppressed the cervical cancer cell growth and induced the cell cycle arrest while its overexpression promote the cancer cell progression and drove the cell cycle [51].The cervical cancer, the third most common gynecological cancer, is classified as the fourth leading cause of cancer-related death in woman all over the world especially in developing countries [52,53]. Despite the modern and standard treatments including surgery, radiotherapy and chemotherapy for the cervical cancer, still the 5-year survival rate of

advanced patients remains very low [51,54]. The findings showed that the TDP-43 expression is up-regulated in cervical cancer tissues and mainly located in nucleus of cervical cells. This indicated that TDP-43 might play an important role in cervical cancer progression. Consistently, the knockdown of TDP-43 attenuated the cell growth and caused the G2/M cell cycle arrest. Furthermore, the effect of TDP-43 on cell cycle proteins (CCNA1 and CDK2 which are important cell cycle proteins in cancer progression) activity was investigated and found out that their activity decreased after knockdown of TDP-43 expression while it is enhanced upon the overexpression of TDP-43 [51].

Melanoma is the most malignant superficial tumor, and 80% of skin cancer leads to death [55]. The previous study showed that high expression of TDP-43 modulates glycolysis level through suppressing the miR-520 expression which, in turn, inhibited the PFKP expression in hepatocellular carcinoma (HCC) [56].It was worthy to find out the TDP-43 functional role in melanoma progression. Meanwhile, the glucose is known as the primary energy of tumor cells as the increased glucose uptake importantly sustains their survival and proliferation. The glucose transporters (GLUTs), the key players in rate limiting factors in glucose metabolism, were found to be strongly related to proliferation and metastasis of various tumors, including melanoma [57]. Hence, higher TDP-43 levels have been reported to be correlated with poor survival of melanoma patients. Thus, this indicated that TDP-43 is a potential molecular prognostic marker for melanoma. Moreover, TDP-43 knockdown reduced melanoma cells proliferation and metastasis, and impaired the glucose uptake through suppressing GLUT4 expression. Finally, the TDP-43 depletion affects tumorigenicity of melanoma cell lines whereby it significantly slowed down tumor growth and decrease the GLUT4 expression in *vivo*[58]. Together, this study found out that TDP-43 is a novel oncogene which regulates the melanoma proliferation and metastasis potentially through modulation of GLUT4 expression.

TDP-43 also plays an oncogenic role in glioma cells malignant progression by stabilizing small nucleolar RNA host gene 12 (SNHG12) [59]. Glioblastoma is the most common lethal primary brain tumor of adult people[60]. Long non-coding RNA (lncRNA) dysregulation is involved in tumorigenesis and regulation of different cellular processes in glioma. lncRNA SNHG12 is upregulated and promoted tumorigenesis in human osteosarcoma cells,

hepatocellular carcinoma cells, and gastric cancer cells. The inhibition of TDP-43 or SNHG12 significantly led to a decrease in proliferation of glioma cells, while the combined inhibition impeded glioma cell growth. Further, TDP-43 and SNHG12 knockdown reduced migration and invasion abilities of glioma cells. Regarding the correlation between TDP-43 and SNHG12, the results revealed that SNHG12 bound with TDP-43, hypothetically through RNA Recognition Motif (RRM). Additionally, SNHG12 expression depends on the TDP-43 presence in glioma cells whereby the TDP-43 knockdown decreased SNHG12 expression by reducing SNGH12's half-life. Taken together, this study indicated that TDP-43 exerts key roles in glioma cells malignant progression by stabilizing SNHG12. miR-195, downregulated in glioma tissues and cells, significantly impaired the malignant progression of glioma cells. TDP-43 upregulated miR-195 in an SNHG12-dependent manner [59].

Lung cancer is the main one detected in man and covers 17% of the new cancer cases worldwide and occupies approximately one fourth of the total cancer deaths. Almost 85% of the lung cancer cases belong to the non-small cell lung cancer (NSCLC), which include squamous cell carcinoma, large cell carcinoma and adenocarcinoma[61].MALAT1 (Metastasis-associated-in-lung-adenocarcinoma-transcript-1), a non-coding RNA overexpressed in non-small cell lung cancer (NSCLS), is overexpressed in the lung cancer tissues. The findings demonstrated that TDP-43 modulates the MALAT1activitiesthrough direct binding to UG repeats in MALAT1 RNA at 3' region by RRMs as they recognize UG/TG repeats in RNA/DNA, thus their participation is compulsory. This study demonstrated that TDP-43 controls the growth, invasion and migration of NSCLC cells through regulating the MALAT1 [62].

The alternative splicing has been reported as a potential hallmark of cancer. The TDP-43 is an important splicing regulator and responsible for the extensive alternative splicing events in triple-negative breast cancer (TNBC). TDP-43 acts together with another splicing factor, SRSF3 (serine/arginine-rich splicing factor), by forming a complex through RRM domain. Reduced tumor progression, including proliferation and metastasis, was observed upon the knockdown of TDP-43 in TNBC and RRM involvement is assured [63].

Collectively, all these studies evidenced that targeting the TDP-43 RRM domains may, therefore, be an effective therapeutic approach for neurodegenerative diseases and cancers.

### 1.2.5 TDP-43 and RRM inhibition

Although more is known about the TDP-43 biology and its association with neurodegenerative and cancer diseases, development of treatments toward the TDP-43 is mostly lag behind those targeting other proteins involved in such diseases [64]. RRM and RNA complexes have long been attractive targets for small molecules inhibition targeting the RNA not the protein. Firstly, aminoacridine derivative was discovered to interrupt the formation of RNA and U1A RRM1 complex[65]. The U1A protein, containing the RNA Recognition Motif, is a part of the spliceosome participating in the most eukaryotic pre-mRNA splicing. The U1A protein has been characterized for binding to the stem loop and internal loop target sites of RNA. Due to the RNA ability to form different and stable structures, the RNA targets in RNA-protein complexes change in both structure and sequence, consequently form distinct, potential target sites for small molecules[66].It has been investigated whether the aminoacridine derivative 1 (AD1) and 2,4,5,6-tetraaminoquinozaline are able to destabilize the complex of U1A RRM and stem loop 2 RNA, and this study demonstrated that AD1 is an effective inhibitor of the U1A-RNA complex whereby AD1 binds to and changes the conformation or dynamics of the RNA stem loop target site [65].

Additionally, a high-throughput screening assay based on AlphaScreen® technology was used to characterize DNA and RNA oligonucleotides (bt-TAR-32 and bt-TG6, respectively) binding to TDP-43 and their interaction inhibition was assessed. The novel, quantitative and non-radiometric assay was developed and validated for characterizing the bt-TAR-32 and bt-TG6 nucleic acids binding to TDP-43. Their binding to TDP-43 was saturable and of high affinity, with sub-nanomolar $K_D$ values and their association rates were similar, whereas the dissociation rate of bt-TAR-32 was slower than that of bt-TG6. Then, a screening of a small library of 7360 chemically diverse and drug-like compounds resulted in identifying a series of structurally similar compounds that disrupt bt-TAR-32 binding to TDP-43 with affinities ranging from 100 nanomolar to 10 micromolar [67].

Later, that series of 4-aminoquinoline derivatives were characterized for its capacity to modulate TDP-43 metabolism and function, whereby they bind to TDP-43, reduce its interaction with the oligonucleotide and stimulate caspase-mediated cleavage of TDP-43. The 4-aminoquinoline derivatives mediated nucleic acid inhibition from binding to TDP-43, where these compounds bound to both free and occupied TDP-43 at a site different from the oligonucleotide binding site, but with differing affinities. Although the 4-aminoquinoline and oligonucleotide binding sites are different, they could still bind to one of the RRM domains. Given their low molecular weight, it is unlikely that those compounds neither bind to both RRM domains simultaneously nor behave as competitive inhibitors compared to oligonucleotides. One possible scenario is that the aminoquinoline derivatives bind to either RRM1 or RRM2 and induce the conformational change that lower the binding affinity resulting in dissociation [68].Consistently, information is still lacking on their binding topology.

Furthermore, some medicinal treatments reduce the TDP-43 inclusions through autophagy pathway were discussed [69]. The protein misfolding and abnormal aggregation are common features of most age-related neurodegenerative diseases, including Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), Huntington's disease (HD), all classified as protein-misfolding disorders (PMDs)[70,71].The aggregation combines the highly diffusible small oligomers, fibrils, and protein inclusions considered as large aggregates with amyloid properties. The autophagy is the efficient mechanism and cellular catabolic route for the selective degradation of aggregation-prone proteins (including TDP-43) linked to neurodegeneration disease. Therefore, various compounds have been identified in high-throughput screening to enhance autophagy. Regarding the TDP-43, treatment of Parkinson's disease with rapamycin and its derivatives promoted the elimination of the most relevant aggregate-prone proteins involved in neurodegeneration, including TDP-43. Additionally, small molecules such as spermidine, carbamazepine, and tamoxifen have been shown to rescue motor dysfunction in mutant TDP-43 transgenic mice, in correlation with enhanced autophagy levels in ALS/FTD [69]. However, all those compounds have the limitations in enhancing the autophagy levels and may have detrimental consequences by exacerbating the disease progression despite their therapeutic potentials.

Although there is accumulative evidence about diseases treatment done through targeting the entire TDP-43, there is no compound directly targeting RRM domains of TDP-43 have been uncovered to our best knowledge.

## 1.3 NMR Fragment-based screening

Fragment-based lead discovery (FBLD) is a popular, effective and efficient approach for identifying hits for lead development. FBLD entails screening of small-molecule libraries against a target protein to identify weakly potent, bioactive molecules. The fragments are small, and less complex, this increases the probability of binding to a target protein, and consequently results in higher hit rates and efficient search of diverse chemical space. Additionally, the hits identified from fragment screening do not require changes and can be efficiently developed for specificity and inhibitory activity[72].

Nuclear magnetic resonance (NMR) spectroscopy is a unique tool to study molecular interactions in solution, and it became an essential technique to characterize events of molecular recognition and to obtain information about the interactions of small ligands with biologically relevant macromolecules including proteins and nucleic acids (DNA and RNA). It is widely used in drug discovery to identify hits in screening of compound libraries[73]. NMR spectroscopy is a powerful approach which has been extensively used by the pioneers in fragment-based drug discovery for detecting the molecular interactions between target and the fragment libraries [74-76] and to facilitate the structure-based drug design [77].

Binding events of ligands to receptors are the key for an understanding of biological processes. The new NMR spectroscopic techniques permitted to gain insight into protein-protein and protein-ligand interactions in solution on an atomic level. There are several approaches available, which can be used to study protein ligand interactions by solution NMR spectroscopy. The methods can be primarily divided into protein observed or ligand observed techniques. In a protein observed method, a spectrum of protein is acquired and the ligand is titrated. This provides information about the residues in the protein, which are involved in the direct

interaction with the ligand. These experiments identify binding events either by looking at the resonance signals of the ligand or the protein. Ideally, both techniques together deliver a complete picture of ligand binding to a receptor. Those two NMR methods are often used at the initial stage and at the optimization phase of drug-development as long as NMR can detect even a weak protein-inhibitor interaction [78,79].

### 1.3.1 NMR Ligand-based screening

Ligand-based NMR screening and the NMR determination of the bound conformation of a ligand are important tools in the rational drug-discovery process. Ligand-observed NMR experiments provide useful information about the protein-binding properties of small molecule ligands. Generally, these techniques rely on protons signals observation in the small molecule fragment and their behavioral changes in the presence/absence of a biomolecule of interest. As such, the techniques are relatively sensitive and are "label-free": the protein and small molecule are not required to undergo the modification in order to generate binding data. However, these techniques are limited to not often generate information about the binding site of the compound on the protein. None the less, these techniques have proven to be highly useful in the area of nucleic acid-ligand interactions [80]. The three ligand-observed NMR experiments (STD, WaterLOGSY, and CPMG) are often used against each other to carefully screen a library on a well-known tractable protein target.

#### Saturation-transfer difference (STD)

The saturation-transfer difference (STD) experiment, type of protein-ligand interaction study using NMR in solution, is performed using relatively small quantities of non-labeled of proteins, by recording a set of one-dimensional NMR spectra and it is based on the nuclear Overhauser effect. The success of this technique is a consequence of its robustness and the fact that it is focused on the signals of the ligand, without any need of processing NMR information about the receptor. STD identifies whether a ligand interacts with a protein even at the weak binding affinity, thus is suitable as an initial check of the protein-ligand interaction or for drug-screening purposes.

This method can detect a ligand-protein interaction even when the binding is weak. An NMR sample containing a small amount of protein and an excess amount of ligand is prepared in order to run this experiment for such a weak binding system. An optimal ligand: protein ratio for the STD experiment depends on the dissociation constant, i.e., on and off rates, and the solubility of the ligand. Then, two spectra, one with saturation of protein $^{1}$H signals (*on-resonance*) and the other without the saturation (*off-resonance*), are recorded for a set of STD experiments. When the saturated protein $^{1}$H magnetization is transferred to the ligand $^{1}$H magnetization, the ligand signal intensity of the two spectra differs (**Figure 1.10**). To make sure that the protein signals are saturated sufficiently without direct irradiation of ligand signals, it is worthy to repeat the experiments by selecting different irradiation frequencies. Also, the individual ligand and protein spectra must be recorded as controls to interpret the STD data [82, 83].

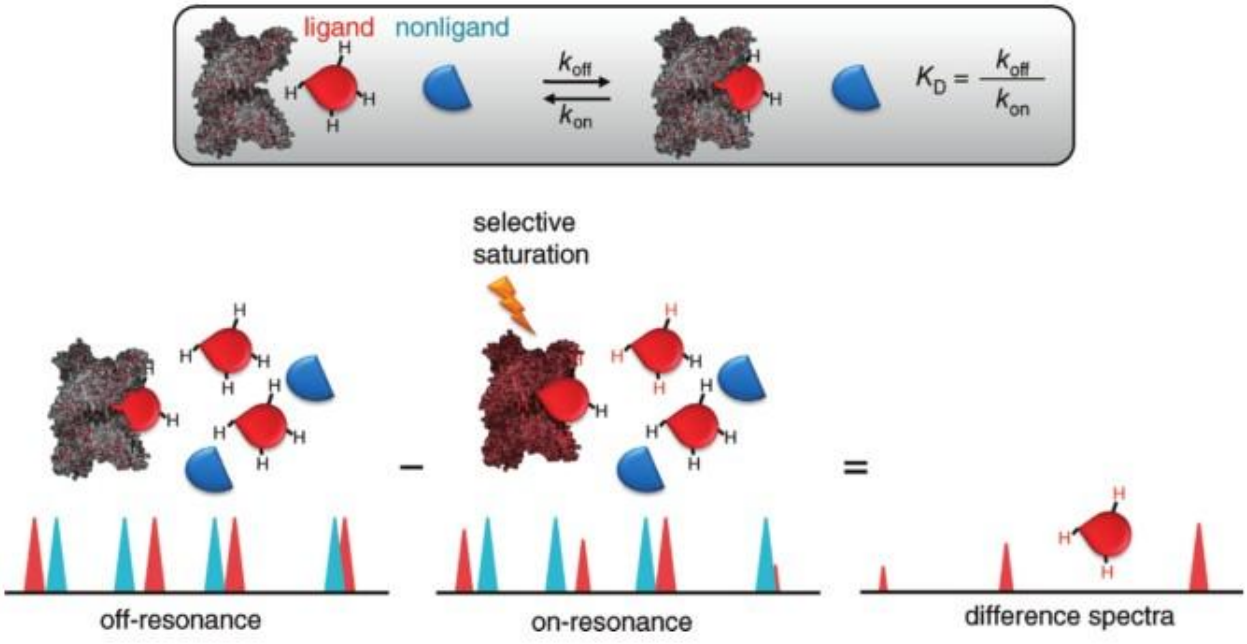


**Figure 1.10. Scheme of the STD-NMR experiment**. The exchange between free and bound ligand allows intermolecular transfer of magnetization from the receptor to the bound small molecule[81].

It can be used as a screening technique, for identification of lead structures or as a useful tool for identifying ligand moieties important for binding. The term binding epitope is frequently used in the STD-NMR literature to describe the hydrogens of the ligand that are closer to the protein upon binding [81]

## WaterLOGSY

The Water-Ligand Observed via Gradient SpectroscopY (Water-LOGSY) experiment is a NMR technique commonly uses 1D ligand-observed method for identifying the molecules that interact with target biomolecules (proteins, RNA and DNA fragments). Water-LOGSY is based on Nuclear Overhauser effect spectroscopy (NOESY) and on intermolecular magnetization transfer via the bulk water. The inverted water magnetization can be transferred to the bound ligand through different pathways, where the ligand interacts with water via water−ligand−protein or protein−ligand complexes, the rotational correlation times of which yield negative cross-relaxation rates and show negative NOEs with water. In contrast, small molecules that only interact with the bulk water (i.e., non-binders) undergo much more rapid tumbling, which translates into positive NOEs. Therefore, opposite signs for the signals from free versus protein-bound ligands are observed in the WaterLOGSY spectrum, which allows differentiating easily the binders from non-binders [82-84].

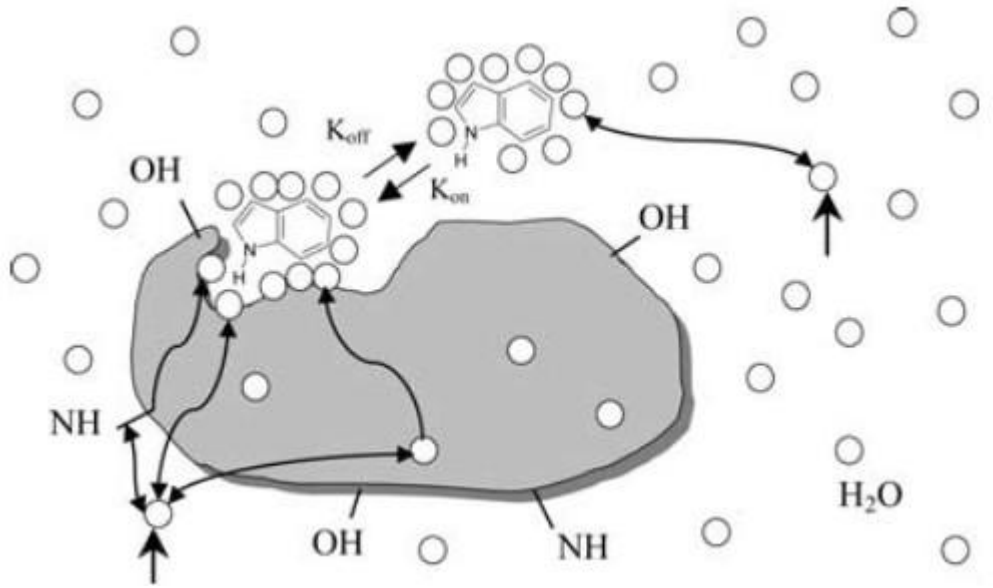


**Figure 1.11**.**The WaterLOGSY principle**. The protein is shown with the buried cavities and the active binding site. The ligand is shown in the bound and free states. Excitation of bulk water (circles) is shown with a solid arrow and some of the different magnetization transfer pathways are shown with dashed lines[84].

The WaterLOGSY represents a powerful NMR approach for primary screening of compounds for binding to the target of interest in the μM range. Those identified novel hits of

micromolar affinity can be then optimized using directed screening, combinatorial chemistry, medicinal chemistry and structure-based drug design. The method requires only limited amount of protein due to its high sensitivity [82-84].

**Carr-Purcell-Meiboom-Gill (CPMG)**

CPMG NMR is a straightforward and reliable approach for validating binding of a ligand to protein, and has been used more broadly in drug discovery to observe protein/ligand binding. CPMG is a relaxation-edited, ligand-observed $^{1}$H NMR technique that considers differences in relaxation properties between the nuclei in macromolecules and nuclei in small molecules to probe binding. More specifically, CPMG utilizes a pulse sequence that exploits the differences in transverse relaxation times ($T_2$), which are directly dependent on the molecular rotational correlation time ($\tau_c$) and thus on the molecular weight of a molecule or complex. Small molecules relax slowly (long $T_2$, second scale) exhibiting sharp and well defined peaks. In contrast, higher molecular weight species, such as proteins or nucleic acids, relax quickly (short $T_2$, millisecond scale) and tend to show broad NMR signals. Importantly, when small molecules are transiently bound to macromolecules, they exhibit changes in relaxation times that can be easily observed using CPMG pulse sequences [80,85,86].

CPMG provides an easy to interpret NMR spectrum, allowing for straightforward determination of binding of low molecular weight compounds. CPMG has similar sample and affinity requirements to STD and WaterLOGSY and can therefore be used at the same stage of lead discovery or ligand validation and is often used in parallel with these techniques. CPMG requires also only small amount of protein. In addition, CPMG does not require isotope labeling of either the ligand or protein, resulting in an accessible and low cost-effective NMR technique. Because CPMG monitors only the $^{1}$H NMR signals from the ligand it is highly sensitive, and the entire data acquisition can be completed in short time. CPMG is applicable to low-molecular weight compounds that bind to proteins/nucleic acids with on–off rates in the micro- to millisecond time scale, corresponding to affinities in the micromolar to millimolar range [80,85,86]

### 1.3.2 NMR Protein-based Screening and Chemical Shift Mapping

Upon formation of protein–ligand interactions several physical parameters of both the protein and the ligand change. First of all there will be changes in the local electron density due to for example, differences in the hydrophobicity at the interaction surface. Differences in the electron density have an influence on the most easily observable NMR parameter: the chemical shift. Large changes in chemical shifts are also induced by the spatial proximity of groups with magnetic susceptibility anisotropies, like aromatic rings. Importantly, the chemical shift is not only influenced by the change in the covalent molecular structure of a protein but also by the non-covalent interactions with ligands and solvent molecules. One of the most important protein observed methods for the investigation of protein–ligand interactions is the chemical shift mapping (CSM) also known as the chemical shift perturbation (CSP). Thereby, a series of NMR spectra of the protein are recorded in the absence and presence of varying amounts of the binding ligand. The most common experiment which is used for chemical shift mapping is the $^{15}$N-heteronuclear single quantum correlation (HSQC) experiment due to its superior signal dispersion.

Binding is most easily seen by overlaying all HSQCs recorded during the titration. If there is an interaction, the chemical shifts of the residues involved in the complex formation with ligand, seen as peaks in a $^{15}$N-HSQC, are displaced from their original position [79,87].

Protein-detected ligand titration experiments are often performed using isotope-labeled samples, so that many protein signals can be tracked at once. This is useful to determine a ligand-dissociation constant and estimate the ligand-binding site. Even in unfavorable conditions where other inter-nuclear interactions cannot be detected, the nuclear chemical shift is sensitive to local electronic changes. Chemical shift perturbations have long been used to indicate interaction sites. The observation of the protein signals depends on various factors, such as magnetic field strength, available machine time, protein size, buffer, and temperature. To separate many signals, $^{1}$H-$^{15}$N HSQC experiments, which mainly reflect $^{1}$H-$^{15}$N correlations of protein backbone, are employed[88].

As ligand concentration increases, protein NMR signals show one of two titration-spectral patterns: (1) changes in chemical shift positions (fast exchange regime) or (2) a decrease in the free-form signals and the bound-form signals increase (slow exchange regime). The fast exchange pattern is observed when the exchange rate is larger than the difference in chemical shifts between the free and the bound forms, while the slow exchange pattern is observed when the exchange rate is smaller than the shift difference. It is possible to observe both patterns in one protein NMR spectrum, as the differences in chemical shifts between the free and ligand-bound forms vary from site to site [79,87,89].

## 1.4 Significance of the study

Interactions of proteins with other molecules essentially define their functions. The interaction partner spans a wide range from ions, small molecules, lipids, peptides to other proteins or membranes. Nuclear magnetic resonance (NMR) spectroscopy is a very efficient technique in order to get information about protein–ligand interactions at atomic resolution. Besides providing structural information, it also allows for a fast screening, especially of weakly binding ligands. Here, we carried out the automated NMR fragment-based screening [90] to identify three hits of the tandem RRMs of TDP-43. Following chemical shift perturbations of the $^{15}$N labeled TDP-43 Tandem RRMs demonstrate that these hits bind to the same site, mainly on the RRM2 domain. It has also been validated by the chemical shift perturbation experiments for TDP-43 RRM2 alone. The CSP-driven HADDOCK was used to generate the protein-hits binding mode. Collectively, our work provides a class of compounds for further hit-to-lead evolution of TDP-43 RRM domain, and paves the path for targeting protein-nucleic acids interactions using the fragment-based approach.

# Chapter 2 Materials and methods

## 2.1 Cloning of TDP-43 Tandem RRM and RRM2

We selected the boundary containing tandem RRM (residues 101-269) including the linker according to the literature. We used Primer Premier 5.0 software to design the Primers by considering the avoidance oftheir mismatching and formation of dimers between them; and consider the restriction enzyme sites of the restriction enzymes used in the amplified gene fragments as well. In the construction of the expression vector of Tandem RRM, we selected the restriction enzymes Nde I and Xho I to digest the gene fragment and the vector pET22b (GE, Shanghai, China). The sequence of the primers is as follows:

Upstreamprimer: 5'-CCGGAACATATGCAGAAAACATCCGATTTAAATAGTGTTGGG-3’

Downstream primer: 5’-CCGCTCGAGCTGTCTATTGCTATTGTGCTTAGGTTC-3’.

The tandem RRM domain (191-262) of TDP-43gene synthesized by GENEWIZ (Suzhou, China) is used as a template in the PCR reaction and sub-cloned into pET22b vector with $His_6$ tag.While the RRM2 domain was amplified from the tandem RRM construct and then sub-cloned into pET22b vector (GE Healthcare, Shanghai, China) with $His_6$ tag. We used the Nde I and Xho I as restriction enzymes with the following primers:

Upstream primer: 5'-CCGGAACATATGAAAGTGTTTGTGGGGCGCTGTACAG -3’

Downstream primer: 5’-CCGCTCGAGCTGTCTATTGCTATTGTGCTTAGGTTC -3’.

All primers were synthesized by Sunny (Shanghai Sunny Biotechnology Co., Ltd). The DNA polymerase enzyme of PCR was Takara's PrimerSTAR Max. The PCR system and program settings are shown in Table 2.1 and Table 2.2, respectively.

## 2.2 Cloning of hADK2, hADK2-LBT1 and hADK2-LBT2

### 2.2.1 Cloning hADK2

The hADK 2 fragment of 717bp was amplified from huma cDNA by PCR with respective primers containing XhoI and NdeI digestion sites, for subsequent sub-cloning. Then they were sub-cloned into pET28a with His-tag and TEV protease cleavage site.

ADK2-foward 5' CCTTAACATATGGCTCCCAGCGTGCCAGCGGCAGAACC 3'
ADK2-reverse 5' CCGCTCGAGTTATGTGGCTTTGGAGAAGGCTGCTAGG 3'

All primers were synthesized by Sunny (Shanghai Sunny Biotechnology Co., Ltd). The DNA polymerase enzyme of PCR was Takara's PrimerSTAR Max. The PCR system and program settings are shown in Table 2.1 and Table 2.2, respectively.

### 2.2.2 Cloning hADK2-LBT1 and hADK2-LBT2

The insertion of Lanthanide binding Tags is done by site directed mutation method. We replaced the loop amino acids by the amino acids sequence of the tag and then follow the mutation creation method. The following primers were used to insert the LBTs into the protein.

**LBT1**: YIDTNNDGWYEGDELLT
ADK2-LBT1-R: 5’ CAACTATGGATGCTTACATCGACACGAACAACGACGGCTG 3’

ADK2-LBT1-F: 5’ GTACGAGGGCGACGAGTTGTTGACGAGTGATGAAATGGTAGTG 3’

**LBT2**: IDTNNDGWYEGDELLT
ADK2-LBT2-R: 5’ CAACTATGGATGCTATCGACACGAACAACGACGGCTGG 3’

ADK2-LBT2-F: 5’ GTACGAGGGCGACGAGTTGTTGACGAGTGATGAAATGGTAGTG 3’

Then all constructs were sub-cloned into pET22b vector (GE Healthcare, Shanghai, China) with $His_6$ tag.

### 2.2.3 Construction of Site-directed Mutants to insert the LBTs

For specific site mutation of amino acids, we selected the mutation kit purchased from TAKARA Company to construct a series of hADK2-LBTs. The experimental procedures are as follows:

1. Identify the mutant amino acid sites and introduce the corresponding nucleotide sequences of the LBTs amino acids into the designed primers. The upstream primers (including the LBTs nucleotides in the upstream primers) are designed along the vector coding direction from the beginning of the LBT sites, and the downstream primers are designed in the opposite direction next to the LBTs sites, with the aim of integrating the whole primers.

2. The wild-type hADK2 plasmid was used as template and LBTs primers were added to carry out PCR amplification. With the above-mentioned system, the extension time in the amplification procedure was extended to sufficiently amplify the complete vector sequence.

3. Blunting Kination reaction: after PCR reaction, the sample was linearized plasmid. After agarose gel identification and recovery, 8.5 μL recovery product was added to the PCR tube. 1 μL of 10 *Blunting Kination Buffer and 0.5 μL of Blunting Kination Enzyme Mix were added to the PCR tube. The enzyme was denatured and terminated by water bath at 37 $^{o}$C for 10 minutes.

4. Linking reaction: Blunting Kination product of 5 μL was mixed with 5 μL Solution I, and the reaction time was more than 1 hour in water bath at 16 $^{o}$C.

5. The conjugates were transformed into competent cells, cultured on a smear plate, and then sequenced to determine whether the LBTs were successfully inserted.

**Table 2.1 Reagents of PCR**

| Reaction System | Volume |
|---|---|
| PrimerSTAR Max (2X) | 25 μl |
| Upstream primer | 1 μl |
| Downstream primer | 1 μl |
| DNA Template | 1 μl |
| Sterilized double distilled water | 22 μl |

**Table 2.2 PCR Reaction procedure**

| PCR steps | Temperature | Time |
|---|---|---|
| 1. Initial denaturation | 98°C | 5 min |
| 2. Denaturation | 95°C | 30 s |
| 3. Annealing | 56°C | 30 s |
| 4. Elongation | 72°C | 10 s |
| | | (back tostep 2, 31 cycles) |
| 5. Final Elongation | 72°C | 10 min |
| 6. Final hold | 10°C | Forever |

## 2.4 Miniprep of target DNA fragments and double digestion

PCR products were identified and recovered by preparing a 1% Agarose gel (1×TAE + 1% Agarose). Specific steps are as follows:

1) Cut down gel pieces containing the targeted DNA fragments on the agarose gel with a clean blade under the UV light and weighing its quantity. 100mg of agarose gel, corresponding to 100ul volume. Put the pieces into a 1.5 ml micro-centrifuge tube. The same volume of Binding buffer (XP2) was added and placed in a 65 ° C water bath for about 10 min until the nucleic acid gel was completely melted.

2) Add the melted nucleic acid gel solution to the HisBind@DNA adsorption mini-column and incubate for 2 min at room temperature.

3) Centrifuge at 13,000 rpm for 1 min to discard the waste liquid from the collection tube.

4) Add 300 μL of Binding buffer (XP2) to the adsorption mini-column and centrifuge at 13,000 rpm for 1 min. Discard the waste.

5) Then centrifuge empty Mini-column tube again to dry for 2 min at 13,000 rpm and discard the waste.

6) Place the column into a new Eppendorf tube, add 30 μL of sterile double distilled water, and incubate for 2 min at room temperature. Centrifuge at 13,000 rpm for 1 min to elute the DNA. The target DNA fragment solution in the centrifuge tube is collected.

7) Take the 2 μL of collected gene and measure its concentration on One Drop (Spectrophotometer).

8) The gene fragment recovered by gel was double-digested with NdeI and XhoI restriction enzymes. The digestion reaction was carried out in a 37 °C water bath for about 3 hrs. The reaction system for digestion is shown in Table 2.3.

**Table 2.3 Components of the double digestion**

| **Reaction system** | **volume** |
|---|---|
| Target segment | 40 μl |
| NdeI (10U) | 2 μl |
| XhoI (10U) | 2 μl |
| 10x FastDigest Buffer | 6 μl |

## 2.5 Ligation and transformation of the target gene

The vector pET22b and pET28aweredigested with the same cleavage enzyme (NdeI and XhoI) for ligation, and then ligated with the double-digested gene in a 16 °C water bath overnight. The reaction system of the connection is shown in Table 2.4.

**Table 2.4 Composition of the target gene**

| Ligation reaction system | Volume |
|---|---|
| Double-digested gene fragment | 4μl |
| Solution **I** ligase | 5μl |
| Double-digested vector | 1μl |

The ligation product was transformed into *E. coli* competent cells (BL21). The reaction steps are as follows:

1) BL21 competent cells were removed from the -80 °C ultra-low temperature freezer and thaw them on ice for 10 min.

2) Add 10 μl of the ligation product to the competent cells and mix gently. Incubate the tubes on ice for 30 min.

3) Heat shock at 42 °C for 45 sec on water bath, quickly transfer back to ice and incubate them for 2 min.

4) Add 500 μl of LB medium to competent cells in a hood. The system was then incubated shaken vigorouslyat 37 ° C for 60 min to restore the growth of *E. coli*.

5) The grown bacterial solution was centrifuged at 5000 rpm for 2 min. Remove500 μL of the supernatant. The remaining bacterial solution was then resuspended and Plate out the suspension on a LB agar plate containing the appropriate antibioticand spread evenly with a pre-sterilized glass coating bar.

6) Place the plate into an incubator at 37 °C for 12 hrs, until the colonies are clearly visible, pick and carry out subsequent colony PCR for verification.

## 2.6 Bacterial liquid PCR to verify recombinant plasmid and sequencing

When the colony size on the plate is appropriate, pick up 3-5 single colonies with a 10 μl pipette tip and put the tip into 5 ml of LB medium containing the corresponding antibiotic. The culture tube was placed in a shaker at 37 ° C for 6 h. 1 μl of the bacterial liquid was taken out as a template to carry out bacterial liquid PCR for verification. The PCR experimental procedure was identical to the step of the previous PCR, except that the inexpensive rTaq enzyme was used as DNA polymerase enzyme in place of the PrimerSTAR Max. In the ultra-clean hood, the bacterial liquid positive for the bacterial liquid PCR was subjected to storage and sequencing, whereas a mixture of 800 μl bacterial solution and 200 μl of glycerol was done for stock.Finally, 500 μl was sent to the Tsingke Company for sequencing.

## 2.7 Expression of proteins

The expression strain is kanamycin resistant. The steps to express are as follows:

1) 5 μl of the stock solution was added to 100 ml of LB medium containing appropriate antibiotic and shaken overnight at 37 °C.

2) Transfer 100 ml of the bacterial solution to 1 L of LB medium, with the appropriate concentration of antibiotics, and continue to culture at 37 °C.

3) Monitor the OD value of the bacterial solution during the culture. When the $OD_{600}$ value reaches 0.8-1.0, take out the culture flask and place it on an ice-water mixture for 5 min. IPTG was added to a final concentration of 0.2 mM and induced at 16 °C for 20 hrs.

4) For the expression of $^{15}N$ labeled protein, it is necessary to pour the bacterial liquid into a pre-sterilized 750 ml Beckmann centrifuge bottle at 4 °C when the $OD_{600}$ value of the bacterial solution reaches 0.8-1.0. After centrifugation at 2800 g for 30 min, the cells were slowly resuspended in M9 medium and transferred to M9 medium containing $^{15}NH_4Cl$. After incubating the inorganic culture solution for 40 min at 37 °C, the flask was taken out; ice-bathed for 5 min, 0.2 mM IPTG was added and induced at 16 °C for 20 h.

## 2.8 Protein purification

The protein tag used for fusion expression in the experiment is 6xHis, which is capable of chelation of nickel ions on the nickel column beads, thereby binding the protein of interest to the gel. The gradient concentration of the imidazole solution can elute the non-specific bound proteins, and finally the high-concentration imidazole solution is used to elute the target protein, thereby preliminary purification of the target protein. The specific purification steps are as follows:

1) Put the induced bacterial solution into a centrifuge bottle and centrifuge at 6,000 rpm for 8 min at 4 °C.

2) After centrifugation, carefully pour out the supernatant. The cells were resuspended in 50 ml of binding buffer (20 mM Tris, 1 M NaCl, pH 7.5) thentransferred into a high pressure homogenizer, and *E. coli* cells were lysed at 6 ° C under a pressure of 800 bar.

3) Pour the lysate into a 40 ml tube and centrifuge at 14000 rpm for 25 min at 4 °C.

4) Put the centrifuged supernatant into a nickel column that was previously equilibrated with the binding buffer, and incubate the supernatant on the column for 5 min before collecting the flow through.

5) The unnecessary proteins were eluted with 20 ml of washing buffer containing 20 mM imidazole and further wash it with 10 ml of washing buffer containing 30 mM and 40 mM imidazole, respectively.

6) Elute the protein of interest from the nickel column with 10 ml of elusion buffer containing 500 mM imidazole.

7) Flow through and each step of the elute is sampled 10 μL to add 10 μL of 2x Coomassie blue loading buffer, the sample is boiled, centrifuged, and run the SDS-PAGE.

8) The eluted protein was transferred to a low salt buffer (20 mM Tris, 200 mM NaCl, pH 7.5) by dialysis or concentration buffer to remove the imidazole, and 5 mM β-mercaptoethanol was added.

9) Add an appropriate amount of TEV enzyme depending on the concentration of the protein of interest. The His-tag was removed by overnight digestion at 16 °C.

10) The digested protein solution removes β-mercaptoethanol from the system by concentrating with the buffer. The cut His-tag and the TEV enzyme were then removed against the nickel column.

11) The fully digested protein solution is further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 75 column (GE Healthcare, Shanghai, China) and the protein buffer contained 20 mMTris, 200 mM NaCl, pH 7.5.

12) The protein solution, collected from size exclusion chromatography, is concentrated to a suitable concentration and stored frozen at -80 °C.

## 2.9 SDS-PAGE

1) Prepare the 10% SDS-PAGE separating gel and 5% SDS-PAGE stacking gel according to the exact formula in table 2.5, plug in the comb until the gel dries completely, put it into the electrophoresis tank.

2) Add the running buffer in the electrophoresis tank, load the amount of proteins sample mixed with the same volume of the SDS loading buffer (100mM Tris, 20% Gan oil, SDS 4 %, bromophenol blue 0.2%, pH 6.8) and boiled for 5 min at 180 $^{o}$C temperature and centrifuged for 2 min at high speed. Then connect the electrophoresis tank to the power.

3) Start with low-voltage 40V, after the protein samples are separated we increased the voltage to 100V and when bromophenol blue indicator reached the bottom of the edge, stop the electrophoresis, remove the rubber plate and peel off therubber plate gently.

4) After electrophoresis, open a two-tier glass gently; get the gel, transfer he gel to a box.

5) Add the appropriate volume of dye solution (10% Acetic acid, 40% ethanol, 0.25% classic Brilliant Blue-250) to the box.

6) Test the dye by heating for 1 min, collect the dye for future use and decolorize the gel with the water and heat again for 6 min

7) Finally, observe the size and purity of the target and save the gel image.

**Table 2.5 SDS-PAGE gel components**

| **10% separating gel:** | **5% stacking gel:** |
|---|---|
| 40% Arc-Bis Mix1200µl | 40% Arc-Bis Mix200µl |
| 80% Glycerol 600 µl | - |
| Gel Buffer1200 µl | Gel Buffer400µl |
| dd$H_2O$ 600µl | dd$H_2O$ 1000µl |
| 10% APS 20 µl | 10% APS 20µl |
| TEMED 2 µl | TEMED 2µl |

## 2.10 NMR HSQC spectrum

The purified protein was exchanged by dialysis into phosphate buffer containing 25 mM $Na_2HPO_4$, 25 mM $NaH_2PO_4$, pH 7.5 and 200 mM NaCl.Then, the protein solution was concentrated to 0.2 mM and recorded $^{1}H$-$^{15}N$ HSQC spectra with a Varian 700 NMR spectrometerequipped with a 96 well auto-sampler and a 5 mm cryoprobe, at the temperature of 25 °C.

## 2.11 NMR Fragment-based screening of Tandem RRM

### 2.11.1 Preparation of initial screening samples

The scope of the screening is our pre-established library of small molecule fragments. The library contains 880 small molecules selected from the Chembridge compound library. These small molecules are mixed into a cocktail according to the physicochemical properties and the nuclear magnetic spectrum of the small molecule.

Fragment screening requires preparation of Tandem RRM protein, 2x phosphate buffer, and deuterium oxide.

Samples need to be prepared in batches with a protein concentration of 10 µM in each screening sample and a final concentration of 400 µM in small molecules. It is first necessary to calculate the volume to which each component of the sample is added.

The sample volume in each NMR tube is 500 μl, so the total volume is V1 = n * 500, where n is the number of samples prepared.

A half volume of $D_2O$ needs to be added to the screening sample, so the total volume of $D_2O$ is V2 = n * 500/2.

The protein concentration of each sample was 10 μM, the volume of the required total protein solution was V3 = n*500*10 / C, and C is the concentration of the protein.

The total volume of small molecules is V4 = n*10. Therefore, the volume of the added 2x buffer is: V5 = n*500 - V1 - V2 - V3. Next, each component except the small molecule was added to the V-shaped tube, and gently blown and mixed with a pipette.

Adjust the range of the eight-channel 1250 μl pipette to 490 μl, and take the protein solution from the V-tube and add it to the Thermomatrix 1.4 ml plastic tube. Then, the eight-channel 10 μl tip is used to sequentially take the small molecule mixture from the fragment library and add it to the tube. The small molecule and protein solution are thoroughly mixed using an oscillator at room temperature. Add them to the NMR tube, and place them on the injector in sequence.

### 2.11.2 Preparation of the secondary screening sample

By the results of the initial screening, a single small molecule that may bind to the protein can be directly selected. The sample preparation process for the secondary screening was similar to that of the primary screening except that a 10 μl mixture of small molecules was replaced with a single small molecule of 1 μl. For simplicity, the volume of small molecules can be ignored when calculating the volume of each component.

### 2.11.3 Processing of NMR Fragment screening results

The screening results are processed by the scripts provided by the ACD software and the files of .nd9 are obtained. When opened in ACD, the file can display $^1H$ spectra of small molecules, $^1H$ spectra of mixtures (or single small molecules), STD, WaterLOGSY, and CPMG spectra. If the small molecule binds to the target protein, a positive peak appears on the STD spectrum, and on the WaterLOGSY spectrum, the peak of the signal opposite to the solvent peak appears, and the signal intensity of the CPMG spectrum peak decreases. It should be noted that due to its nature, small molecules that bind to the target protein do not necessarily show

corresponding results in all three experiments. Therefore, when processing the selected data, if only one of the above cases occurs, small molecules may also be present. If it is combined, the small molecule should be selected for subsequent experimental determination.

### 2.11.4 HSQC titrationsfor small molecule binding validation and NMR chemical shift perturbations

We titrated the $^{15}$N-labeled protein with the compound obtained by primary screening, and judged whether the protein and the small molecule were bound by observing if there was a peak shift in the spectrum.

The $^{15}$N-labeled Tandem RRM protein was first expressed in M9 medium, followed by protein purification.The buffer made of 20 mM PBS, pH 7.5, 200 mM NaCl is used for concentrating the protein. The total volume 500 μl composed of 0.05-0.2 mM protein solution and 50 mM small molecule dissolved in DMSO were prepared for NMR titration. Then, NMR HSQC spectra of the protein were recorded using a series of hit/protein molar ratios of 0.0, 0.5, 1.0, 2.0, 4.0 and 8.0 for TDP-43 tandem RRMs and 0.0, 0.5, 1.0, 2.0, 4.0 for RRM2, respectively. Spectra were processed in NMRpipe[91], and analyzed with Sparky and overlapped. If some of the peaks on the NMR spectrum shifted, thus the small molecules bound to Tandem RRM and/or RRM2.

In addition, the chemical shift changes (Δδ) relative to the free form of protein were defined as

$$\Delta\delta = \sqrt{(\delta_{1_H})^2 + (0.2\delta_{15_N})^2} \qquad (2.1)$$

Where $\delta^1{}_H$ and $\delta^{15}{}_N$ are the chemical shift differences of the $^1$H and $^{15}$N dimensions, respectively. We referred to chemical shift assignments previously deposited in Biological Magnetic Resonance Data Bank: RRM1 (BMRB Entry 18765), RRM2 (BMRB Entry 19922), and tandem RRM (BMRB Entry 19290). All structures figures were prepared by Pymol (DeLano Scientific, LLC, Palo Alto, CA, USA).

## 2.12 Molecular Docking

HADDOCK is information-driven docking technique used for modeling the biomolecules structures by using the experimental or predictive restraints [92,93]. The CSPs, obtained from the NMR HSQC titration data, were used for both as HADDOCK restraints and defining the protein active residues. The tandem RRM structure (PDB: 4bs2) served as the starting structure, while the hit**1** PDB file was generated by the PRODRG [94]. The docking calculations were done by the HADDOCK web server and clustered 186 structures in 16 clusters according to the RMSD threshold of 2 Å.

## 2.13 Linewidth Analysis

The NMR HSQC spectra at molar ratios of 0:1 and 8:1 (hit/protein) were processed using the same NMRpipe script, e.g., 2-fold zero-filling, Fourier transformation, and phase corrections. The spectra analysis was done by using Sparky, with randomly selected peaks. After peak integration, the linewidth, i.e., the full width at half the peak height was automatically estimated by Sparky.

## 2.14 NMR Dynamics

All dynamics NMR experiments were acquired either on Varian 700MHz spectrometer equipped with a 5-mm cryoprobe or an Agilent 500MHz spectrometer equipped with a room temperature probe at 25 $^{o}$C. Data were processed using the NMRPipe[91] software and visualized using Sparky. The $^{15}$N CPMGrelaxation dispersion NMR experiments were performed on $^{2}$H-$^{15}$N hADK2 by bound Ap5a substrate purchased from Sigma-Aldrich. Relaxation data were fit to a two-site model describing the open and closing of the two lids using the full Carver-Richards equation[95] as described below:

$$R_2^{\text{eff}} = \frac{1}{2}\left(R_{2A}^0 + R_{2B}^0 + k_{ex} - \frac{1}{\tau_{CP}}\cosh^{-1}\left(D_+ \cosh\eta_+ - D_- \cos\eta_-\right)\right) \quad (2.2)$$

$$\text{where } D_\pm = \frac{1}{2}\left(\frac{\psi + 2\Delta\omega_N^2}{\sqrt{\psi^2+\zeta^2}} \pm 1\right), \eta_\pm = \frac{\tau_{CP}}{\sqrt{2}}\sqrt{\sqrt{\psi^2+\zeta^2} \pm \psi},$$

$$\psi = \left(R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex}\right)^2 - \Delta\omega_N^2 + 4p_A p_B k_{ex}^2,$$

$$\zeta = 2\Delta\omega_N\left(R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex}\right) \text{ and } \nu_{CPMG} = \frac{1}{2\tau_{CP}}$$

Here, $k_{ex}$ is the rate of exchange between states A and B, $R_{2A}^0$ and $R_{2B}^0$ are the transverse exchange rates (assumed to be the same) for the states A and B, respectively. $p_A$ and $p_B$ are the populations of the respective states, $\Delta\omega$ is the chemical shift difference between the two exchanging states and $\tau_{CP}$ is considered as the time between 180°C pulses in the NMR experiments, and more detailed parameters settings were used in our previous study [96] and [97]. Notably, the two-state exchange is a model which is certainly oversimplified as two domains (LID and AMP) can move during the Ap5a binding. The NMR experiments are performed under saturating concentrations of Ap5a. The $^{15}$N-labeled hADK2 or hADK2-LBT-tagged proteins were first expressed in M9 medium, followed by protein purification. The buffer made of 20 mM PBS, pH 7.5, 200 mM NaCl is used for concentrating the protein. The total volume 500 µl composed of 0.05-0.2 mM protein solution was prepared for NMR titration. Then, NMR HSQC spectra of the protein were recorded using a series of $Tm^{+3}$or $Tb^{+3}$/protein molar ratios of 0.0, 0.5, 0.75, and 1.0, for hADK2 or hADK2-LBT-tagged.The paramagnetic ions $Tm^{+3}$ and $Tb^{+3}$(at saturating concentrations of 0.5 molar ratio of the protein concentration) have been used to titrate hADK2-LBT-tagged in order to observe the paramagnetic effect.

# Chapter 3 Results and Discussion

## 3.1 Results

### 3.1.1 Protein expression and purification of tandem RRM and RRM2 domain of TDP-43

Both the tandem RRM domain (102-265aa) and RRM2 domain (192-265aa) cloned in the pET22b were expressed into the *E.coli* BL21. The tandem RRM protein (20.2 kDa)and RRM2 werepurified and eluted on Nickel column, then confirmed by SDS-PAGE (**Figure 3.1a, c**). The TDP-43 tandem RRM and RRM2 protein solutionswere further purified by size exclusion chromatography using a HiLoad 16/600 Superdex 75 column and collect the purified proteins indicated by the peaks (**Figure 3.1b,d**) for further steps.

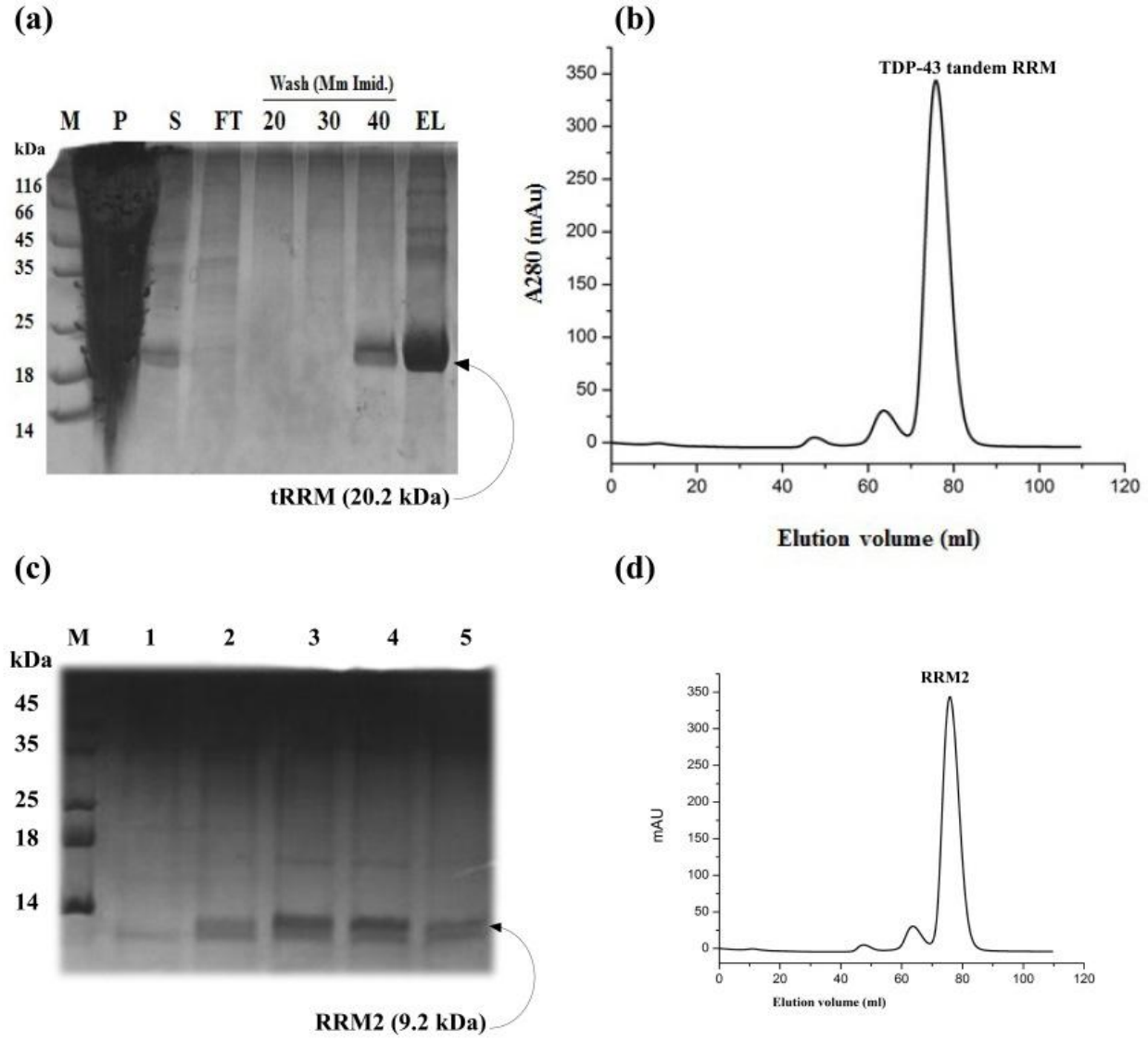


**Figure 3.1.TDP-43 tandem RRM expression and purification.**(**a, c**) The expression of the TDP-43 tandem RRM(20. 2 kDa) and RRM2 (9. 2 kDa) domainsconfirmed by SDS-PAGE. M: Molecular weight; P: Pellet; S: Supernatant; FT: Flow through; 20, 30 and 40: Imidazole concentrations of washing buffer; EL: Eluted protein (Lanes in **a**). M: Molecular weight; Numbers: 1-5, fractions collected from the peak of RRM2 (Lanes in **b**). (**b, d**) Purification of the tandem RRM and RRM2 domains by HiLoad 16/600 Superdex 75 column.

### 3.1.2 The primary and the secondary screening

Since TDP-43 tandem RRMs lack a well-defined small molecule binding pocket with respect to the kinases, we carried out the NMR fragment-based screening to identify weak binders. NMR ligand-observed methods detect the weak protein-ligand binding by detecting the

changes in the characteristics of the ligand spectrum that occur upon binding to the protein. Using the ligand-based experiments, i.e., Saturation Transfer difference (STD) [81], WaterLOGSY [84], CPMG [85] and ligand-based 1D proton, we found 17hits from the primary screening of 89 cocktails containing 10 compounds each (**Figure 3.2a**).The binders present signals while the non-binders present no signals in the STD spectra. Accordingly, the binders show inverted or fast decay of signals in the WaterLOGSY and CPMG experiments, respectively[98].

The combined output of these spectra enabled the identification of primary screening hits from cocktails. It is worth noting that the reference 1D proton spectra of each individual compound might be slightly different from the screening spectra as a different buffer was used to be better compatible with TDP-43 tandem RRMs. The primary screening hits were further validated by the secondary screening for individual hit using the same set of NMR experiments (**Figure 3.2b**). The aromatic peaks of the hit are depicted as they suffer less from the interference of buffer signals. The secondary screening eliminated 13 primary hits, probably due to sample aggregation in cocktails, ambiguous selection of hits with degenerated chemical shifts, and/or spectrometer instability. Among the remaining 4 hits, hit **2** demonstrates a distinct topology relative to hit **1** and **3** (**Figure 3.2c**)[98].

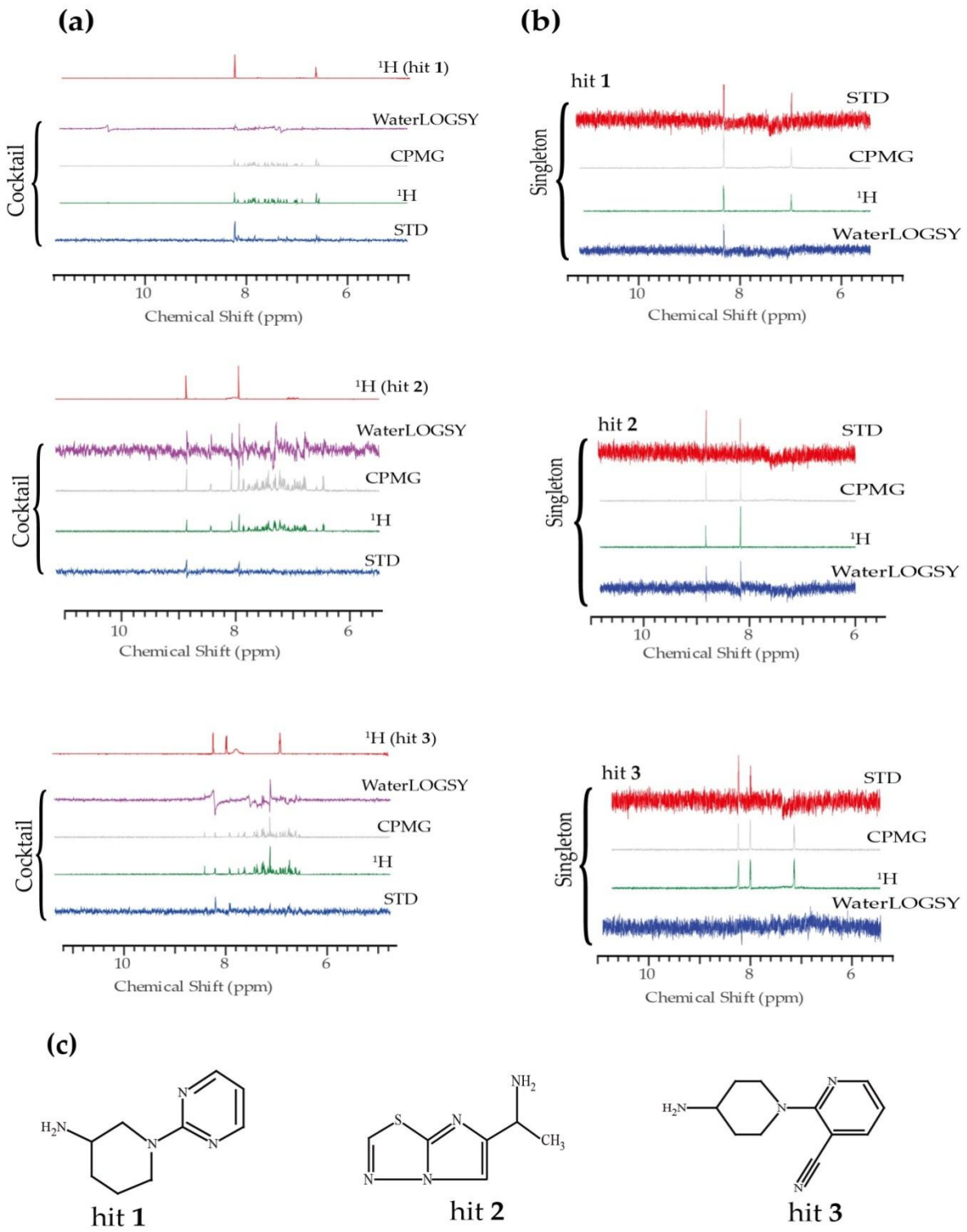


**Figure 3.2.NMR fragment-based screening against the tandem RRM domain of TDP-43**. (**a**) The primary screening WaterLOGSY, CPMG, $^{1}$H and STD spectra for three respresentative cocktails. The $^{1}$H reference spectrum of respective hit is shown for comparison. (**b**) The secondary screening spectra for individual hit **1**, **2**, and **3**, respectively. (**c**) The chemical structures of hits **1**, **2** and **3**.

The foursecondary screening hits were then cross validated using the chemical shift perturbations (CSPs) of the $^{15}$N-labeled tandem RRMs of TDP-43 (**Figure 3.4, 3.5**)and 3 of them induced significant chemical shift changes of the tandem RRM[98].

### 3.1.3 Chemical shift perturbations (CSPs) of Tandem RRM

This approach has been extensively applied in interrogation of protein-ligand interactions in an affinity ranging from nM to mM. As CSP is a sensitive indicator of chemical environment changes induced by ligand titration, it is particularly powerful in detection of weak bindings. The linewidths of the amide signals of TDP-43 tandem RRM show almost no changes upon titration of hit **1** (**Table 3.1**), which suggests that hit **1** induces no protein aggregation[98]. This is a useful approach to remove false positives, which are commonly found in drug screening because of protein aggregation [99].

**Table 3.1 NMR linewidth (lw) analysis of $^{15}$N labeled TDP-43 tandem RRM upon addition of hit 1.**

| Residue | Free | | + hit 1* | |
|---|---|---|---|---|
| | lw$^{15}$N (hz) | lw$^{1}$H (hz) | lw$^{15}$N (hz) | lw$^{1}$H (hz) |
| D236 | 12.9 | 29.0 | 12.6 | 29.9 |
| F124 | 11.4 | 27.4 | 11.2 | 27.5 |
| E156 | 12.1 | 24.5 | 12.5 | 23.8 |
| F211 | 13.0 | 28.1 | 12.9 | 28.1 |
| G245 | 12.3 | 30.0 | 12.0 | 30.3 |
| G196 | 11.5 | 29.9 | 11.4 | 30.8 |

* At the hit/protein molar ratio of 8:1.

Titration of hit **1** induces dose-dependent CSPs of residues G245, E246, H256, I257, S258 (**Figure 3.4b**). However, the curve does not reach the saturation point (**Figure 3.3**), as limited by the weak binding affinity and the low aqueous solubility of the hit**1,** thus further titration of the small molecule is not feasible. Hence the binding affinity of those weak binders cannot be robustly estimated from CSPs.

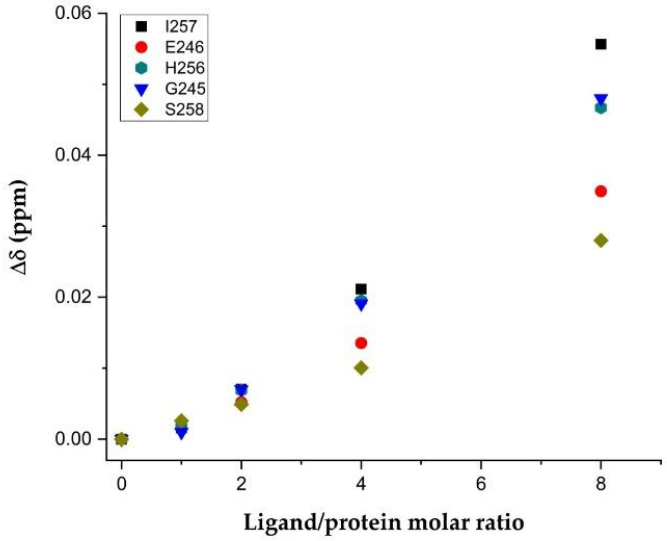


**Figure 3.3.The dose-dependent titration curve for the tandem RRM disturbed residues upon hit 1 titration** at different ligand/protein molar ratios: 0:1, 1:1, 2:1, 4:1, 8:1. Annoted are the disturbed residues.

The disturbed residues were then mapped on the surface representation of the solution structure of TDP-43 tandem RRMs (PDB code: 4BS2) [8]. Residues H256, I257, S258 locate on the β4 strand, while residues Gly245 and Glu246 bridges the α2 and β3 (**Figure 3.4c**)[98].

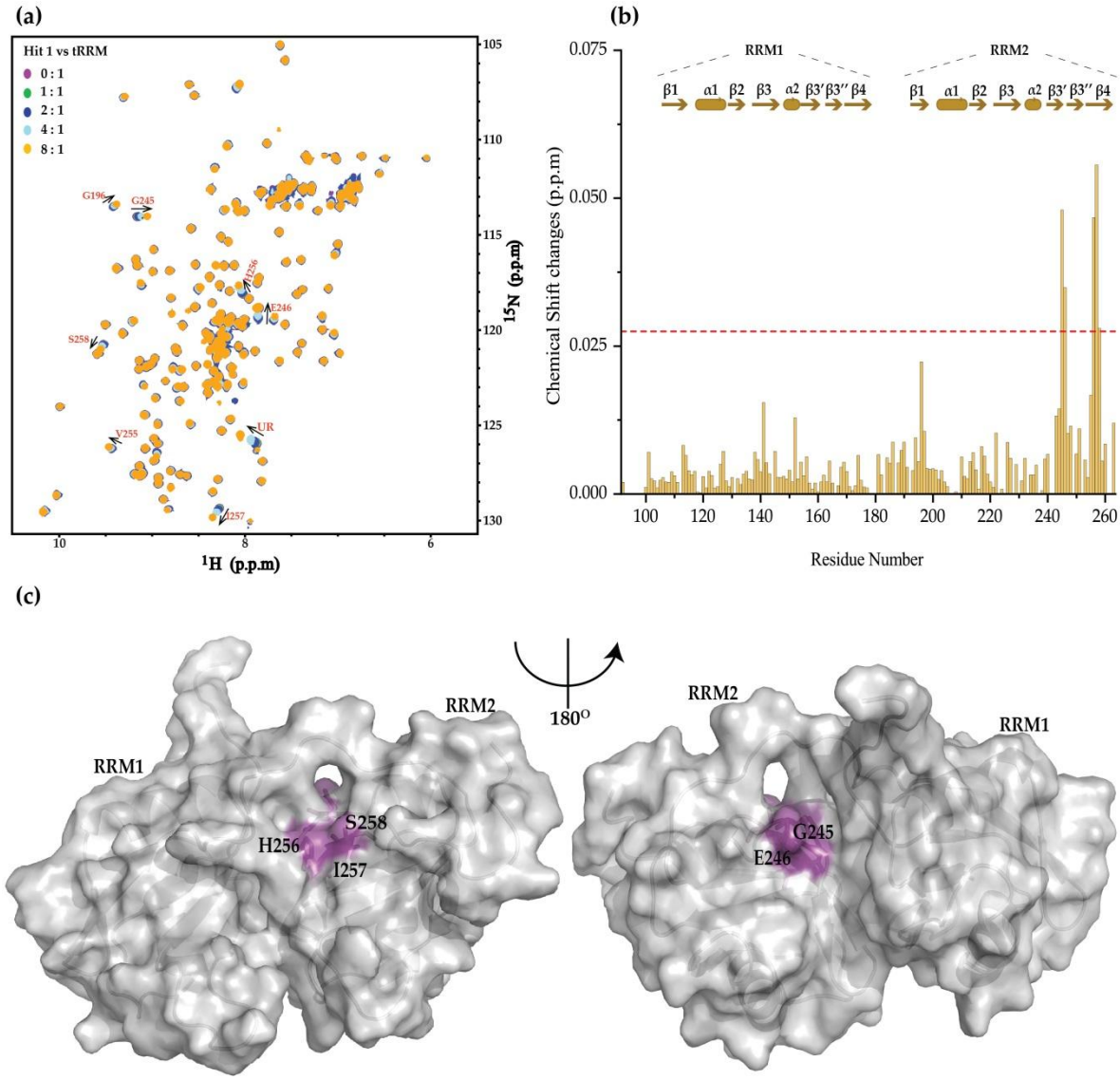


**Figure 3.4.Binding topology of hit 1 on the tandem RRMs of TDP-43 using NMR chemical shift perturbations.** (**a**) The chemical shift perturbations of $^{15}$N-labeled tandem RRM domain of TDP-43 upon titration of hit **1**. Theligand/protein molar ratios are annotated. The perturbed residues are labeled and the arrows indicate the direction of chemical shift changes. UR stands for unassigned residue. (**b**) Chemical shift changes of the TDP-43-tandem RRM domain at the ligand protein molar ratio of 8 : 1. The red horizontal dashed line represents two standard deviations above the averaged chemical shift changes of residues. (**c**) Surface representation of TDP-43tandem RRM domain (PDB code: 4BS2)showing the purple-colored residues with significant chemical shift changes.

Consistently, hits **2** and **3**titrationsalso point to the same binding topology in the tandem RRM of TDP-43 (**Figure 3.5**). For example, hit **2**perturbed residues G245, H256 and I257(**Figure 3.5a, 3.5c**), while hit **3**induces significant CSPs for residues G245, E246, H256, and

I257(**Figure 3.5b, 3.5d**). The similarity of the binding pattern of the three hits suggests that weak but specific binders were successfully identified using the NMR fragment-based screening.

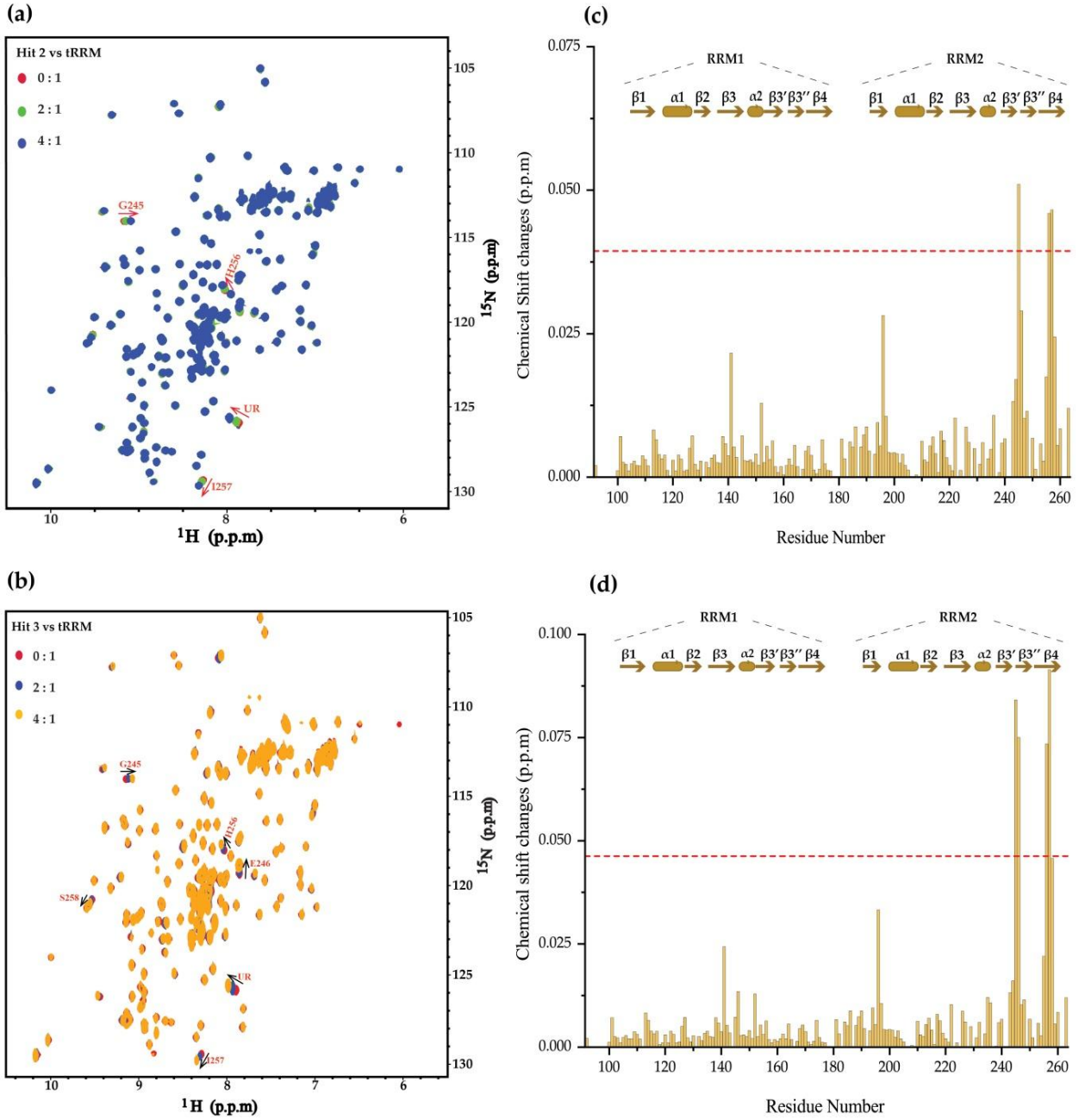


**Figure 3.5.Chemical shift perturbations of tandem RRM upon binding of hit 2 and 3.** (**a, b**) The chemical shift perturbations of TDP-43 tandem RRM domain induced by titration of hit **1** and **2**, respectively. Annotated are the hits: protein molar ratios. UR stands for unassigned residue. (**c, d**) Residue-by-residue chemical shift changes of tandem RRM at the hit/protein molar ratio of 8:1 for compound**2** and **3**, respectively. The red dashed lines represent two standard deviations above the averaged chemical shift changes of residues.

### 3.1.4 TDP-43 RRM2 Chemical shift perturbations (CSPs)

Having confirmed that 3 different hits bind on the same site of TDP-43 RRM2 domain, we further investigated whether RRM2 alone is sufficient for ligand binding. Hit 2 was thus titrated to the $^{15}$N-labeled RRM2 domain of TDP-43 (**Figure 3.6a**). Consequently, the residues G245 on loop bridging the α2 and β3' and H256, I257 located on β4-strand were perturbed(**Figure 3.6b**). Those residues were mapped on the surface representation of TDP-43 RRM2 [10] domain in complex with a single-stranded DNA (**Figure 3.6c**). The hit binds to the same sites of either TDP-43 tandem RRM or RRM2 alone. That is to say, TDP-43 RRM2 is the main contributor for ligand binding, and should be considered as the target for follow-up hit-to-lead evolutions[98].

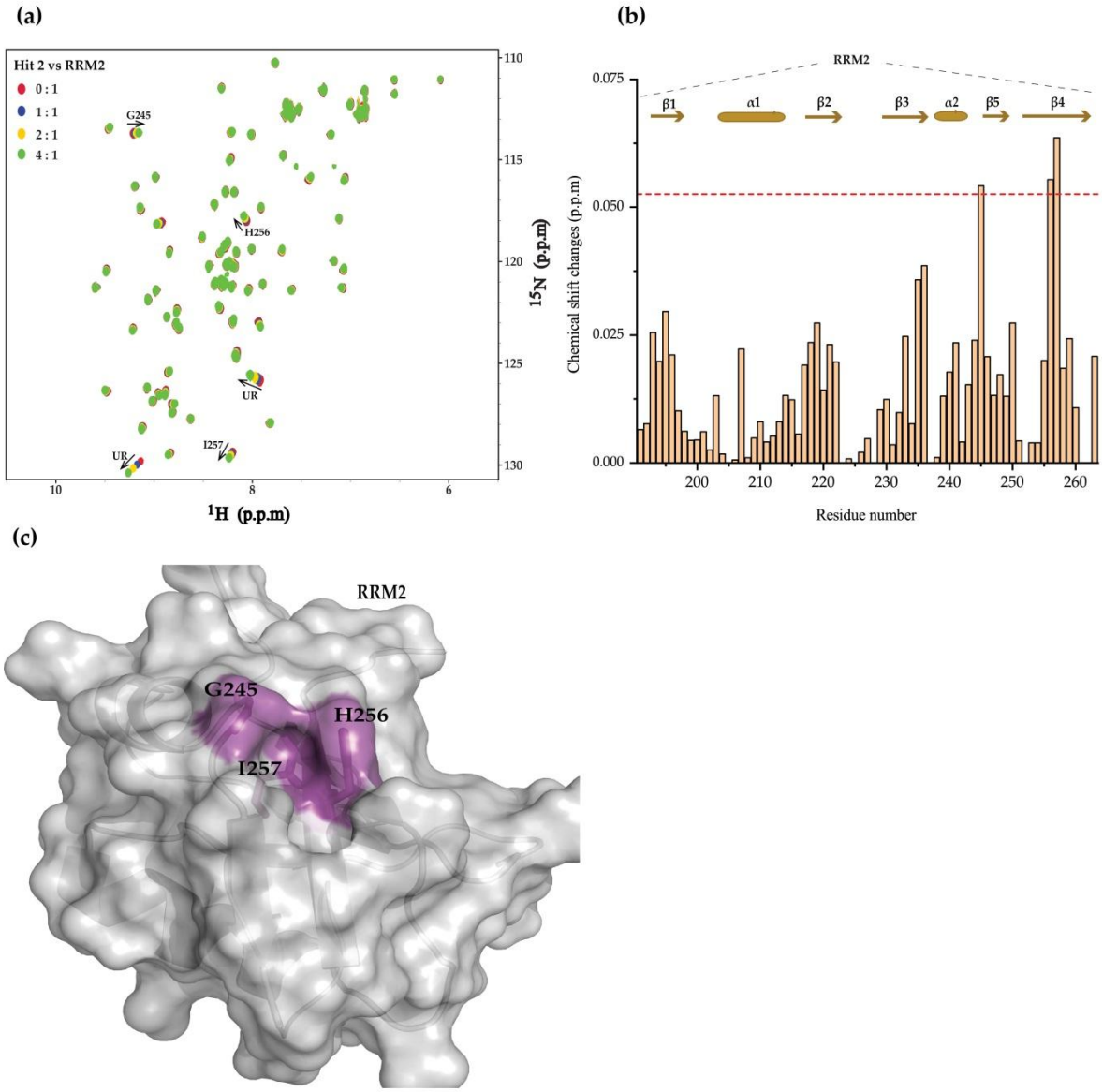


**Figure 3.6.Chemical shift perturbations of TDP-43-RRM2 domain upon hit 2 titration.** (**a**) The chemical shift perturbations of RRM2 domain of TDP-43 by hit **2** titration. The hit **2**: protein molar ratios and corresponding hit **2** chemical structure are denoted. The shifted residues are labeled and the arrows pointing out the trend of chemical shift changes. UR stands for unidentified residue; (**b**) Chemical shift changes of the TDP-43 RRM2 domain residues at a hit **2**: protein molar ratio of 4 : 1. The blue dashed lines represent two standard deviations above the averaged chemical shift changes of residues. (**c**) Residues (colored in purple), undergo significant chemical shift changes, are mapped on the surface representation of TDP-43-RRM2 domain (PDB code: 1WF0) [10].

### 3.1.5 Comparison of hits and nucleotides binding sites

We further compared the small molecule binding topology with the nucleic acid recogition sites of TDP-43 RRM domain. In TDP-43 tandem RRMs, 10 out of 12 nucleotides of the AUG12 RNA (GUGUGAAUGAAU) interact with RRM1 and RRM2 (PDB code: 4BS2) [8].Among them, the first five ($G_1U_2G_3U_4G_5$) nucleotides are accommodated on RRM1 β-sheet and the following two nucleotides ($A_6A_7$) act as a connector between two RRMs, while the next three nucleotides ($U_8G_9A_{10}$) lie on the RRM2. The $U_8$ nt of RNA is recognized on Ser258 (β4) through hydrogen bonds, on the backbone carbonyl oxygen of Asn259 (β4) and the backbone amide of Glu261 from the C-terminus [8]. Comparatively, all three hits have perturbed some residues located on the β4-strand, specifically hits **1** and **3** disturbed Ser258 (β4), this also interacts with $U_8$ nucleotide in tandem RRM (**Figure 3.7a**). Furthermore, the RRM2 residues D247 (loop α2-β3') and I249 (β3') are involved in inter-RRM interactions upon RNA binding on tandem RRM of TDP-43. While, this study revealed that their nearby residues G245 and E246 (loop α2-β3') display higher chemical shift perturbations induced by the hits binding (**Figure 3.4b, 3.5c, 3.5d**)[98].

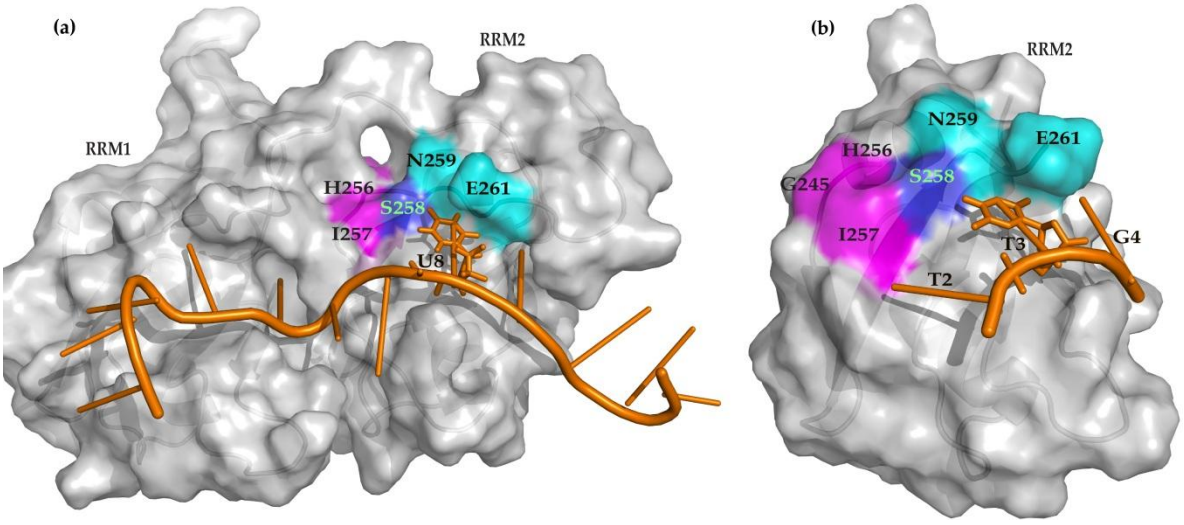


**Figure 3.7.Comparison of binding sites of nucleic acids and hits on TDP-43.** (**a**) Surface representaion of TDP-43 tandem RRMs in complex with AUG12 RNA (orange cartoon), where residues interacting with U8 nucleotide (in stick) and hitsare highlighed in cyan and magenta, respectively. Residue S258 (blue) is interacting with both U8 and hit. (**b**)Surface representation of TDP-43 RRM2 in complex with ssDNA (PDB code: 3D2W) using the same coloring scheme.

Accordingly, the crystal structure of TDP-43 RRM2 in complex with ss-DNA 5'-GTTGAGCGTT-3' (PDB entry: 3D2W) reveals that only three 5' end nucleotides (T2, T3, G4) make extensive contacts with β-sheet residues of RRM2, whereby T3 particularly contacts with Ser258, Asn259 and Glu261 through hydrogen bonds [10], while in our study the residues His256, Ile257 nearby the Ser258 (β4) have been perturbed upon hit binding on single RRM2(**Figure 3.7b**). This suggests that the fragment screening hits bind to a proximal site for RNA/DNA recognition, thus new hits can be designed using fragment grow strategy to block the DNA/RNA recognition capability of TDP-43 RRM2[98].

To further characterize the binding mode, a data-driven approach "Haddock" [92] was used to model the tandem RRM-hit 1 complex structure. Residues G245, E246, H256, I257, and S258 were defined as active ones in binding site. Among the docking poses generated by HADDOCK, the best-fit ones were filtered out based on CSP and STD restraints [100-102]. One representative docking pose (**Figure 3.8**) indicates that hit 1 forms hydrogen bond with the side chain of S258, and the aromatic ring of hit 1 is proximal to residues G245, E246, H256, and I257. These docking poses pave the path for following structure-guided hit-to-lead evolution[98].

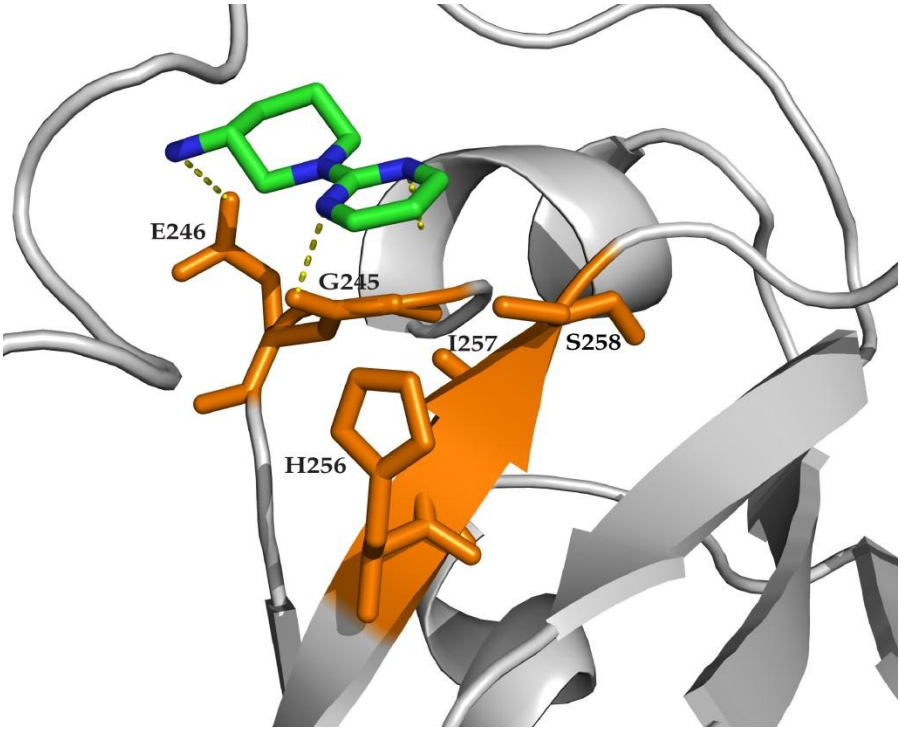


**Figure 3.8.The representative docking model of hit 1 in consistent with experimental CSP and STD restraints.** Hit 1 (green color) in binding site of tRRM (PDB: 4bs2) where the carbonyl hydrogen is oriented toward G245 while the side chain hydrogen interacts with E246 residue of tandem RRM. Other active residues (orange sticks), H256, I257 and S258 are located in proximal of the hit **1.**

## 3.2 Discussion and Conclusion

Proteins containing RRM domains function in important aspects of the posttranscriptional regulation of gene expression, mRNA maturation, and other RNA processing machinery. These proteins perform their diverse roles depending on dual ability to recognize RNA and to interact with other proteins by using their RRM domain [66]. As TDP-43 is closely correlated with neurodegenerative and cancers diseases [63,103], the RRM domain of TDP-43 becomes an attractive therapeutic target. However, there is no direct inhibitor targeting the RRM discovered to date.

Weuncovered three small molecules binding to tandem RRM domain of TDP-43by using NMR fragment-based screening techniques[98].NMR spectroscopy is one of a plethora of biophysical methods that is particularly powerful to detect even ultra-weak protein-ligand interactions. Accordingly, chemical shift perturbations observed in HSQC spectra or the linewidth analysis of the small molecules allow the determination of binding affinity [104,105]. This is sometimes recalcitrant, as the titration to saturation point may be infeasible in case of weak binding affinities and low aqueous solubility of compounds[98].

NMR is extensively applied in fragment-based lead discovery [72].The central idea is to screen a small library (500-2000 molecules) of low-molecular-weight compound (110-250 Da), as their low complexity enhances the probability of matched interactions between the target and these fragment compounds. The reasonable hit rate indicates the druggability of TDP-43 tandem RRM domain.

Although the 4-aminoquinolines molecules have been discovered through high throughput screening against the full-length TDP-43 [34], the enlightenment on binding site is still lacking. TDP-43 contains two RNA-binding RRM domains and the C-terminal low complexity domain, which may form liquid-liquid phase separation as a reservoir of mRNAs. It is hence essential to determine the small molecule binding topology on TDP-43. The tandem RRM of TDP-43 is composed of canonical RRM arrangement (β1α1β2β3α2β4), with an additional β-hairpin (β3'β3" or β5) found between α2 and β4 which extends the β-sheet surface for RNA recognition [2,8,10].

The binding topology of our fragment screening hits and CSP-guided HADDOCK modelingreveal a ligand-binding “hot spot” of TDP-43 RRM2, proximal to H256, I257 and S258[98]. Interestingly, these residues are also close to the RRM1 and RRM2 interface. Previous study proposed that both RRM domains are indispensable, for achieving the greater binding affinity between the TDP-43 and nucleic acids [10]. Since this “hot spot” is partially overlapped with the RNA/DNA recognition site, it directs the following structure-guided hit-to-lead evolution against TDP-43 tandem RRM domains.

# Chapter 4 Conformational dynamics of human Adenylate Kinase 2 (hADK2)

## 4.1 Introduction

Phosphate esters and anhydrides are high-energy linkages that are extremely resistant to nucleophilic attack and are therefore fundamental to storage of biochemical energy, genomic stability,and long signaling state lifetimes[106]. While stability is critical, organisms must rapidly find response and effectively to their environment, reason why they have evolved enzymes to catalyze transfer of phosphoryl groups with exquisite specificity and enormous rate accelerations compared to the uncatalyzed reaction in solution. Phosphoryl transfer plays key roles in signaling, energy transduction, protein synthesis, and maintaining the integrity of the genetic material[107].

To fulfill their functions the proteins/enzymes spontaneously search for an energy and amino acid sequences determine the unique order structure of proteins, and the folding process. Consequently, proteins become inherently dynamic and are often described as an ensemble of interconverting conformations.The state with the lowest energy forms the most stable structure, which is also called “ground state”. The structure determines function, reason why in the past decade people focused onthe analysis of the ground state structure of proteins in the field of structural biology, in order toexplain why proteins have a specific function. However, growing evidences have shown that proteins in solution do not exist as static structures in the form of building blocks, on contrary in the ground and excited states whereby the structural change is the origin of protein dynamic existence [108].

Besides understanding the static structure of proteins, it is necessary to understand the mechanism and process of their functions. It is necessary to analyze the structure of the excited states in different exquisite time and spatial resolution. As long as excited conformations play an important role in the function of proteins, they help to explain more comprehensively the relationship between protein structure and function, thereforethe way proteins fulfill their roles [109]. Excited states of proteins are essential for their functions, such as ligand binding, molecular

recognition, catalytic reactions and so on. Structural effects play an important role in the process[110].

The ground state of a protein is the most energetic low state, while the excited state also known as transient is high energy state, so its population is also low[111]. Excited states cannot be observed directly on NMR spectra due to their small population and short existence time. Therefore, the study of the excited state is mainly based on the signal transformation of the ground state by means of dynamic experiments to deduce the properties of the excited state[109,112].

NMR Dynamics describes proteins in different states.In general, the information acquired during the conversion between states at a certain rate cannot directly provide structural constraints. Therefore, it is necessary to obtain the structure information of excited states by dynamic experiments in order to find out the solution to this problem.
In this thesis part, we briefly summarize some of nuclear magnetic resonance (NMR) methods appliedin this research to study the existence of human ADK2 (hADK2) conformational changes and dynamics during its open- and closed-like conformational states.Thus, this work will yield novel insight into the behavior of human ADK2 by using different NMR techniques to determine their conformational dynamics.

## 4.2 Nuclear magnetic resonance techniques used in this research

Recent methodological advancements in Nuclear magnetic resonance(NMR) have extended the ability to characterize protein dynamics and promise to shed new light on the mechanisms by which the molecules function. However, X-ray crystal diffraction is still the most important method to obtain accurate atomic structure of proteins. The static and homogeneous three-dimensional atomic structure model is the basis of proteinsfunction understanding.

On another hand, NMR has a unique advantage in measuring the dynamic effects of proteins in solution. It can observe the dynamic characteristics of proteins at the atomic level in the time scale from picosecond to seconds, which plays an important role in comprehensively understanding the conformation of proteins in performing their functions[113]. Protein motions at

different time scales correspond to different types of functions, such as side chain motions from picosecond to nanosecond, loop motions from nanosecond to microsecond, and microsecond to millisecond. The second range corresponds to the movement of protein domains/subunits [112]. NMR is uniquely suited to characterize dynamics of a protein in solution and a variety of different NMR techniques have been developed to characterize the conformational/chemical exchange processes, including line shape analysis, ZZ exchange, R1ρ, Carr-Purcell-Meiboom-Gill relaxation dispersion (CPMG RD) and chemical exchange saturation transfer (CEST) experiments, and each experiment is sensitive to exchange processes on a specific timescale as shown in **figure 4.1**[112,114,115].

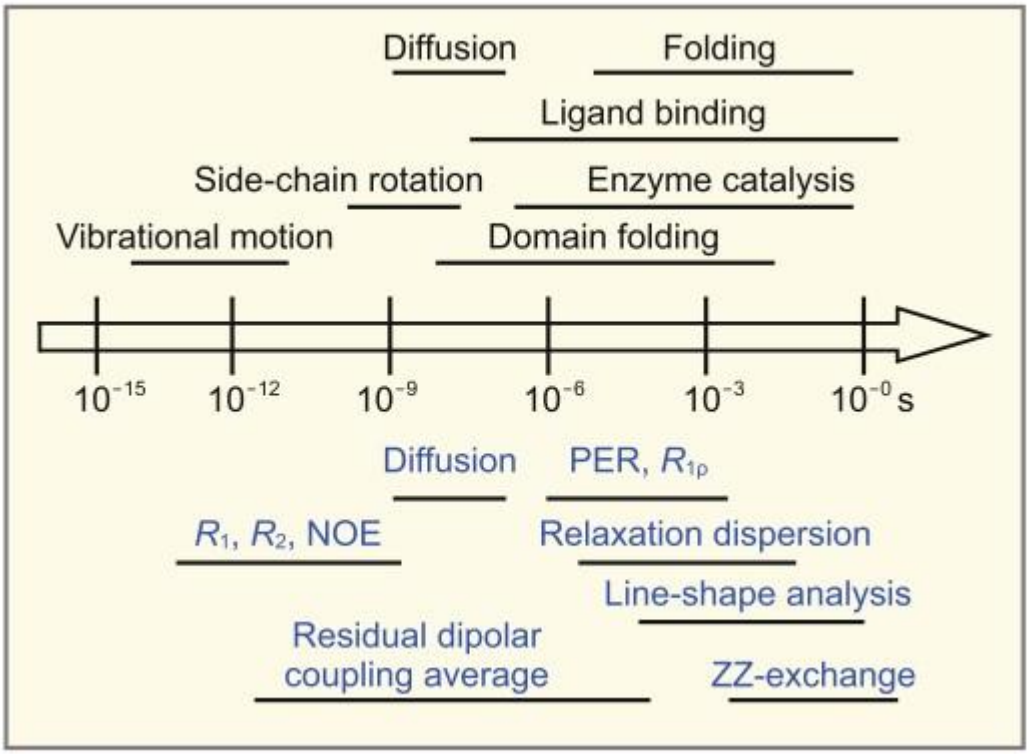


**Figure 4.1**: Major NMR methods used to study the protein dynamics at different time scales[112].

### 4.2.1 Description of Chemical Exchange

The ground state of proteins coexists in a dynamic equilibrium with other conformations, whose energies are only slightly higher than that of the ground state. Several studies have suggested that these alternative conformations are likely to be functional, e.g. in enzyme catalysis [116]. These higher energy states, transiently formed and populated at only very low levelsdesignated in what follows as "invisible" or excited conformational states (excited states for short), have remained recalcitrant to detailed quantitative analysisbecause they are often not directly observable in the NMR spectrum. Yet, they may manifest themselves through line broadening of the NMR signals, an effect commonly known as "chemical exchange".

If an atom exchanges in two or more environments, its chemical shift or relaxation properties vary in different environments, and its NMR spectra change accordingly. In most cases, chemical exchange refers to non-covalent bond changes, such as protein folding, protein-ligand binding, and in a few cases, molecular reactions to form covalent bonds. There is a common formula used to express the exchange of nuclear spins between states A and B, in order to illustrate the mechanism of chemical exchange[115]:

$$\mathrm{A} \underset{\mathrm{k_2}}{\overset{\mathrm{k_1}}{\rightleftarrows}} \mathrm{B} \quad (4.1)$$

$k_1$ is conversion rate constantfrom A to B while $k_2$ is from B to A, hence we can defineglobal equilibrium constants as:

$$K_{eq} = \frac{k_1}{k_2} \quad (4.2)$$

In the equilibrium state, the proportions of the two states are respectively:

$$\rho_A = \frac{k_2}{k_1 + k_2} = \frac{1}{1 + K_{eq}} \quad \text{and} \quad \rho_B = \frac{k_1}{k_1 + k_2} = \frac{K_{eq}}{1 + K_{eq}} \quad (4.3)$$

For the convenience of describing the exchange at different time scales, we first define the difference of chemical shifts (Δω) between the apparent exchange rate $k_{ex}$ and states A and B:

$$k_{\mathrm{ex}} = k_1 + k_2 \quad (4.4)$$

$$\Delta_\omega = \omega_A + \omega_B \quad (4.5)$$

According to the relationship between $k_{ex}$ and Δω, chemical exchange can be divided into three types: fast exchange, slow exchange and medium exchange. In fast exchange regime, the exchange rate is much faster than the chemical shift difference (in Hz) for fast exchange ($k_{ex} \gg$ Δω), here a peak is observed in NMR spectra because there is not enough time to establish the related resonance frequencies in both states A and B.

In slow exchange condition, the apparent exchange rate is much slower than the chemical shift difference (in Hz), which is a slow exchange ($k_{ex} \ll \Delta\omega$).At this point, two discrete peaks can be observed on the NMR spectrum, each of which corresponds to two different states, A and B.

The intermediate exchange ($k_{ex} \approx \Delta\omega$) causes the line broadening to be very severe and even causes the peak to disappear.

### 4.2.2 Relaxation Dispersion

Chemical exchange in proteins can be monitored by NMR relaxation dispersion, whereby information about the exchanging states is acquired.

For moderately slow exchange (100μs~10 ms), the presence of a small number of excited states (≥0.5%) leads to a significant increase in the apparent lateral relaxation rate and a significant broadening of ground state NMR signals. Because of the low population, many experimental methods cannot see this state, and the exchange rate can contribute to the apparent lateral relaxation rate, so Relaxation dispersion (RD) experiments can study this "invisible" state. Common RD experiments include CPMG (Carr-Purcell-Meiboom-Gill) and $R_{1\rho}$Experiment [115]. CPMG experiments add a series of repetitive $180^{o}$pulses to the xy plane during the z-axis precession by heteronuclear magnetization, and reduce the contribution of the exchange relaxation rate to the transverse relaxation rate by adjusting the repetition frequency of these pulses. The experiment of $R_{1\rho}$ is similar to that of spin locking field $B_1$ by adjusting the intensity of spin locking field $B_1$.

The attenuation effect of RD experiments on the exchange rate contribution depends on the relationship between the exchange rate and the applied magnetic field strength $B_1$. In CPMG, the added magnetic field intensity $B_1$ is realized by a series of 180$^o$C. When the repetition frequency of 180 $^o$C in CPMG is so high that the pulse interval is reduced to zero, CPMG is equivalent to $R_{1\rho}$. Therefore, the magnetic field intensity $B_1$ added in $R_{1\rho}$ experiment is higher than that of CMPG[115].

The **Figure 4.2** describes the CPMG basic principle. As illustrated by **Figure 4.2a**, The CPMG experiment is suitable for the case where the exchange rate is200 $s^{-1} \leq k_{ex} \leq 2000\ s^{-1}$ and $p_E \geqslant 0.5\%$. The effective transverse relaxation rates ($R_{2,eff}$) of the visible peaks are quantified as a function of the frequency $\nu_{CPMG} = 1/(4\tau_{CPMG})$ at which refocusing 180° pulses are applied during a relaxation delay $T_{relax}$. The evolution of magnetization is illustrated in **Figure 4.2b**, where without any loss in generality it has been assumed that the carrier is placed on resonance for the major-state correlation being considered. After the 90°pulse at the start of the echo train, magnetization from this stateremains aligned along the *y* axis in the rotating frame while themagnetization from the minor state precesses in the *xy* plane, losing phase with the ground state. On average, the excited-state spins accumulate a phase equals to $\Delta\omega/k_{EG}$ between exchange events, so exchange leads to a reduction in the major-state signal and hence a nonzero exchange-induced relaxation rate $R_{ex}$. Each 180° pulse of the CPMG train inverts the sense in which the spins precess around the z axis, reducing the phase accumulation and decreasing $R_{ex}$. Exchange parameters and minor-state chemical shifts can be obtained from fits of the $R_{2,eff}(\nu_{CPMG})$ profile so long as the exchange contribution to $R_{2,eff}$(i.e., $R_{ex}$) is quenched as $\nu_{CPMG}$ increases (**Figure 4.2c**). In the slow-exchange limit, $R_{ex} = k_{GE}$ when $\nu_{CPMG} = 0$, so the maximum observable change in the $R_{2,eff}(\nu_{CPMG})$ profile is $k_{GE}$[117].

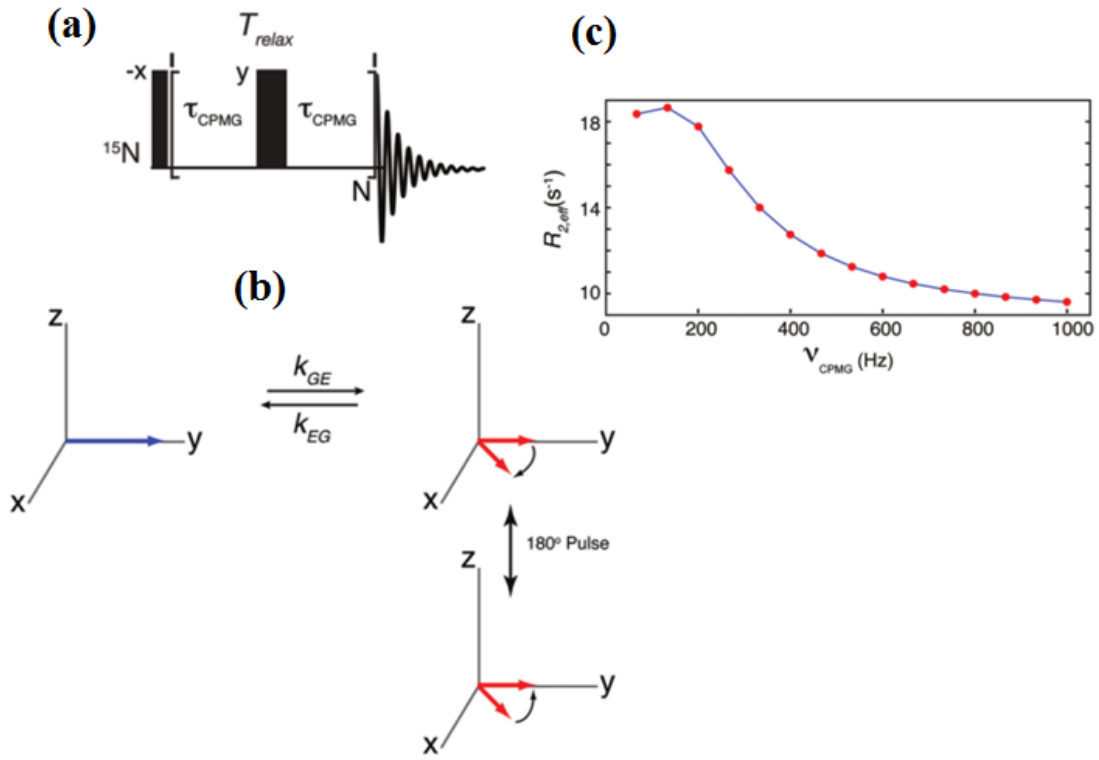


**Figure 4.2:** Basic Principle of CMPG Experiment [117]

From the dependence of $R_{ex}$ on $\nu_{CPMG}$explored in the relaxationdispersion experiment, three physical parameters for a two-site dynamic process can, in principle, be obtained: rates of interconversion ($k_{ex}$), relative populations of the exchanging species ($p_A$ and $p_B$) and chemical shifts between the exchanging species ($\Delta\omega$).

### 4.2.3 Pseudocontact shift (PCS)

The paramagnetic NMR is a toolbox useful to obtain structure and dynamic information evenfor systems hardly affordable with other techniques. It can assist, for instance, in the assignment of the NMRspectra, in drug screening, in the definition of the interaction surfaces in protein complexes, and in the characterization of the structural heterogeneity of biomolecular systems, including the detection of low populated conformational states[118].

**The principle of Pseudocontact Shift**

The magnetic anisotropic lanthanide metal ions can be rigidly linked to the specific sites of proteins by lanthanide binding tags (LBT) on proteins. This will cause a large shift in the NMR peaks of proteins. Except for the residues near the label, which will be affected by the local chemical environment changes caused by direct contact, the large-scale peak shifts occurring elsewhere are not caused by the binding. Therefore, this changein chemical shifts result from both through-bond (contact shifts) and through-space and is referred to as pseudocontact shifts (PCSs) interactions. PCS is produced by the dipole-dipole interaction between the unpaired electrons of lanthanide metal ions and the nuclear spins of proteins. It is informative to depict the PCSs as isosurfaces superimposed onto the 3D structure of the protein. The isosurfaces present the coordinates for which Equation 4.6 predicts the same PCS values,[119]. **Figure 4.3** shows the specific mechanism of PCS.

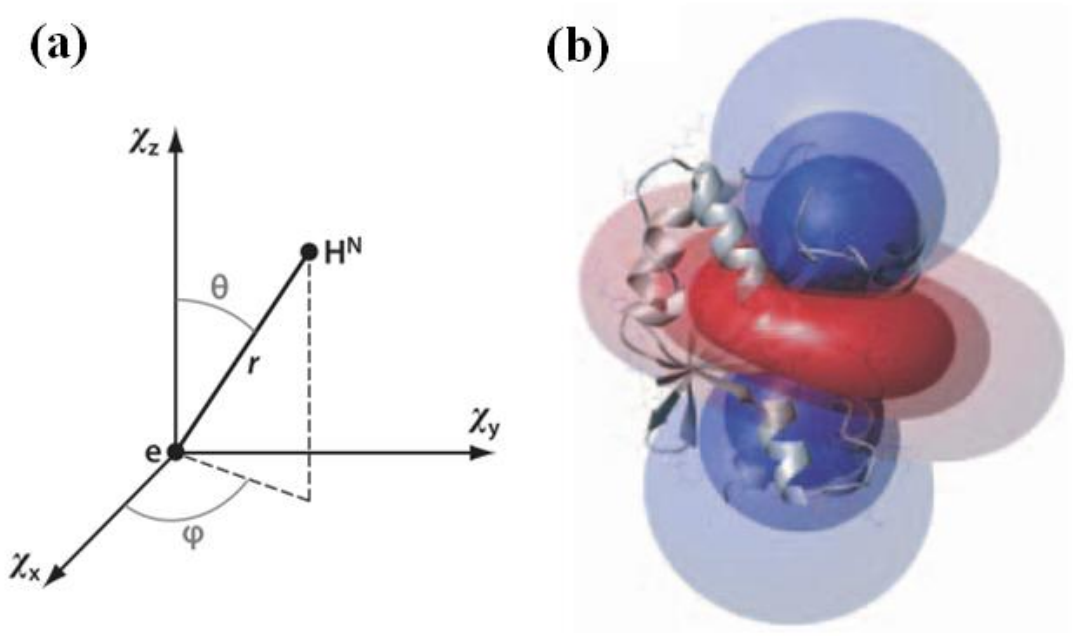


**Figure 4.3**: Lanthanide Metal Ions Induce Pseudocontact shift on Proteins. **(a)** Geometric information of the pseudocontact shift (PCS) effect describing angle and distance of H atom relative to Δχ tensor principal axis coordinate system by equation 4.6 **(b)**The isosurface depicting the PCS effect induced by lanthanide metal ions on proteins. Blue and red surfaces identify the spatial locations of positive and negative PCSs relative to lanthanide metal ions, and the color ranges from dark to light representing (±3, ±1.5, and ±0.5 ppm), respectively[119,120].

The magnetization tensor of lanthanide metal ions consists of two parts: the magnetic isotropic part and the magnetic anisotropic part $\Delta\chi$. The magneto-isotropic part only produces PRE effect, while the magneto-anisotropic part $\Delta\chi$ causes PCS effect. PCS is related to the direction and distance of nuclei relative to lanthanide metal ions. This can be described by the formula 4.6 [121].

$$\Delta\delta^{PCS} = 1/(12\pi r^3)[\Delta\chi_{ax}(3\cos^2\theta-1)+1.5\,\Delta\chi_{rh}\sin^2\theta\cos 2\varphi] \qquad (4.6)$$

Where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ denote, respectively, the axial and rhombic components of the $\Delta\chi$ tensor which is the anisotropy component of the magnetic susceptibility tensor $\chi$describing the magnetic moment of the paramagnetic center. Pseudocontact shifts $\Delta\delta^{PCS}$ (measured inparts per million) present some of the most useful paramagnetic data. They depend on the polar coordinates $r$, $\theta$, and $\varphi$ of the nuclear spin with respect to the $\Delta\chi$ tensor of the metal ion and leads to measurable effects at distances of up to 40Å.

$$\Delta\chi_{ax} = \chi_z - (\chi_x + \chi_y)/2 \qquad (4.7)$$

$$\Delta\chi_{rh}) = (\chi_x - \chi_y) \qquad (4.8)$$

The $\chi$ and $\Delta\chi$ tensors share the same principalaxes. The $\Delta\chi$ tensor vanishes for $\chi_x = \chi_y = \chi_z$. The anisotropy of the $\chi$ tensor is governed by the coordination environment.

**Lanthanide ion tags**

Paramagnetic metals can occur naturally in a macromolecule [122]or be introduced by replacing an existing metal[123], covalently or noncovalently added as a chelate [124,125], bound by an unnatural amino acid [126,127], or lanthanide-binding peptide sequence tag (LBT) engineered into the protein sequence [128-130] as shown in **Figure 4.3**. However, there is one common limitation to artificial tags which is the presence of metal ion motion, resulting in a removal of PCS on average due to the rapid movement of the label, so the PCS value obtained will be small.

Although polypeptide-basedtags have the advantage of ease-of-use by simple encoding into a protein expression vector for inserting metals,predicting tag motion for structure calculations is challenging due to an increase in tag size (~1800 Da) and number of rotatable bonds when compared to the nitroxide labels (~170 Da). Additionally, engineering proteins to incorporate the LBTs within the protein rather than at the N or C termini has proven successful; this can be done either through inserting into a loop connecting two defined secondary structural elements or through secondary attachment to a cysteine residue [128,131,132]. Such alterations likely will reduce, but not eliminate, peptide-based tag motion. However, the eliminationof unnecessary amino acidsto improve existing LBT sequences did contribute to neither conformational stability nor lanthanide binding[133].Up to now, there are many kinds and quantities of lanthanide metal chelating agents, the development of labels is continuing, and the application of PCS is gradually becoming mature.

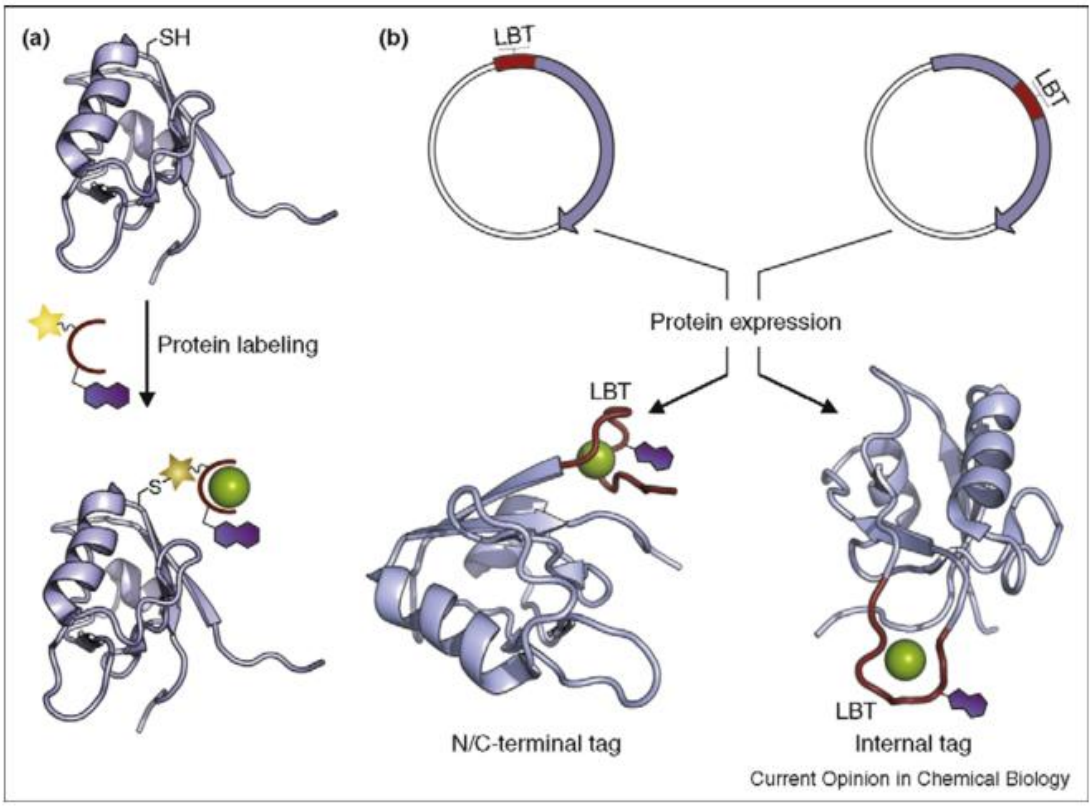


**Figure 4.3**. Methods for the site-specific incorporation of lanthanide-binding ion tags into target proteins. **(a)** Attachment of synthetic lanthanide-chelatingprosthetic groups (red) via chemical modification between a reactive group (star) and cysteines (shown here) or other amino acids. LBT peptides mayalso be linked to the target proteinby chemical modification of target cysteines. **(b)** Peptide-based lanthanide-binding ion tags (red) are encoded at the DNA level and fused to either the N-terminus or C-terminus of the target protein or into a loop[134].

Paramagnetic lanthanide ions are becoming increasingly utilized as structural probes because of their limited chemical reactivity, bioorthogonal nature, and similar coordination chemistry across the group, and strong magnetic properties. The magnetic susceptibility ($\chi$)-tensor for most paramagnetic lanthanide ions is characterized by significant anisotropy that leads to partial alignment of lanthanide chelates in a magnetic field and, as an independent phenomenon, PCSs of resonances. Anisotropy is greatest for the $Tb^{3+}$, $Dy^{3+}$and $Tm^{3+}$ ions. Though paramagnetic, $Gd^{3+}$ has an isotropic $\chi$-tensor and does not cause alignment or PCSs. The diamagnetic lanthanide ions $La^{3+}$, $Y^{3+}$ and $Lu^{3+}$ are useful in control experiments because they neither align nor cause PCSs and PREs[135].

**Experimental Measurement of PCSs**

The measurement of PCS is essentially the measurement of chemical displacement, as PCS is the difference of chemical shift between paramagnetic and diamagnetic states. Any experiment that can obtain chemical shift can be used to measure PCS. However, practically for the sake of simplicity, two-dimensional spectrum is generally used. The diamagnetic spectra use one kind of diamagnetic metals (such as $Y^{3+}$ or $Lu^{3+}$), while paramagnetic spectra use two or more kinds of paramagnetic metals (such as $Tm^{3+}$, $Tb^{3+}$). After the overlapping of these spectra, the positions of the paramagnetic and diamagnetic belonging to the same residue are approximately in a straight line, which can help us to identify the peaks. If there are many peaks, and the peaks of many residues are in the same straight line, ZZ-exchange spectra can be used to assist the identification [119].

Another way to measure the PCS depends on rapid chemical exchange between the diamagnetic and paramagnetic states of the protein. This method can be applied if the metal ion or metal ion complex binds sufficiently weakly to exchange rapidly between different protein molecules. In the fast metal exchange process, titration of the protein with the paramagnetic metal ion leads to continuous changes in chemical shifts from the position of the diamagnetic peak. In the slow metal exchange regime, finite exchange rates can be used to generate cross-peaks in exchange experiments performed with an equimolar mixture of paramagnetic and diamagnetic metal ions, linking each paramagnetic peak with its diamagnetic partner[119,136].

All above described methods have been applied to study conformational changes and dynamics during the open- and closed-like conformational states of hADK2.

## 4.3 Human Adenylate Kinases

Protein kinases are enzymes with critical roles in cellular processes and pathology. As a result, researchers have studied their activity and regulatory mechanisms extensively .
Adenylate kinases (ADK/AK) are phosphotransferases that catalyze the reversible conversion (by transferring of a γ-phosphate group) of ATP and AMP to two molecules of ADP, an essential reaction for many processes in living cells, importantly in synthesisof nucleic acid and cellular energy metabolism (**Figure 4.4**).The AMPis the preferred substrate of all ADKs while their main phosphate donor is ATP, even though some can phosphorylate other substrates like dAMP, CMP, dCMP, and use other phosphate donors such as GTP, NTPs and dNTPs.Nine different adenylate kinase isoenzymes have beenidentified and characterized so far in human tissues, named ADK1 to ADK9 according to their order of discovery.ADKs belong to the nucleoside monophosphate kinase (NMPK) family, which also includes other members such as guanylate kinases, thymidylate kinases, and UMP/CMP kinases[137].

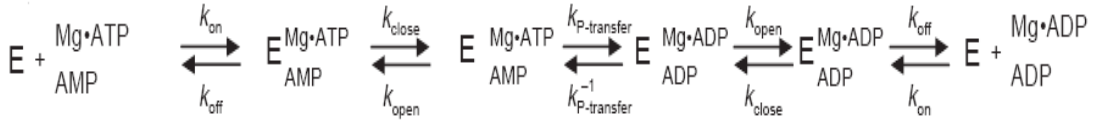


**Figure 4.4**. Reaction of reversible conversion of ATP and AMP into two ADP molecules.

Adenylate kinase is a 3-domain protein consisting of a large central CORE domain that contains the phosphate donor binding glycine-rich region or p-loop. On one side The Core domain is flanked by a LID domain (ATP-binding)and the NMP domain (AMP-binding) on the other side. In the absence of ligands, the enzyme is in the free "open" conformation, in which both the AMPbd and the LID are far away from the CORE. Upon phosphate donor (ATP or other) binding, the LID domain closes, enabling the substrate to bind. Then the phosphoryl transfer occurs. This results in two bound ADPs, of which the one bound at the phosphate donor binding site is released with a subsequent opening ofthe LID domain. Subsequently, this opening signals to the substrate binding domain to open and release the second ADP (**Figure4.5**) [138].

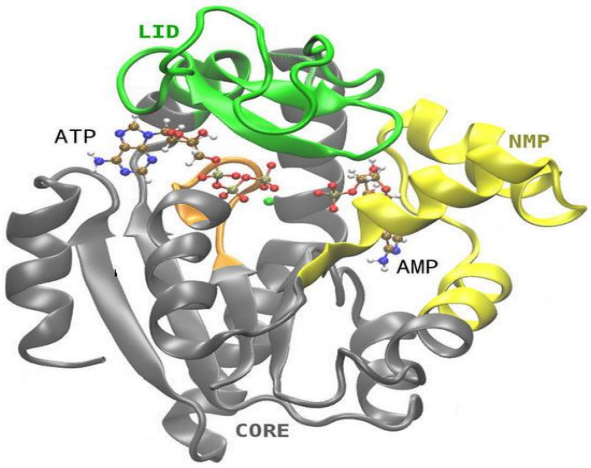


**Figure 4.5.** Structure of Adenylate Kinase (AKE). ATP, that is complexed with $Mg^{+2}$, is bound between the CORE (grey) and LID domains (green), in the so called ATP binding site, while AMP is sandwiched between CORE and NMP (yellow), in the AMP binding site. ATP and AMP ligands are represented as ball and stick [138].

### 4.3.1 Human adenylate kinase 2 (hADK2)

The ADK2 is present in the mitochondrial intermembrane space and its mRNA is strongly expressed in liver, heart, skeletal muscle and pancreas, moderately expressed in kidney, placenta and brain and weakly expressed in lung tissues. However, at the protein level, ADK2 is detected abundantly in liver, kidney and heart and faintly in lung and skeletal muscle. ADK2 can use all nucleoside triphosphates to phosphorylate AMP[139,140].

Deficiencies of the mitochondrial ADK2 have also been implicated in human disease. The reticular dysgenesis, the most severe form of inborn human combined immunodeficiency, has beenfound in individuals with the mutations in gene encoding ADK2. Complete absence of granulocytes and lymphocytes, hypoplasia of the thymus and secondary lymphoid organs, and lack of innate and adaptive humoral and cellular immunity are the symptoms of the disease[137]. Additionally,apart the fatal septicemia caused by reticular dysgenesis within days after birth, sensorneural deafness is also associated with the disease[141]. The ADK2 also may be implicated in the proliferation of hematopoietic precursors and in controlling cell apoptosis due to its mitochondrial location, where it plays a central role in providing the required energy[142].

Evidence also supported the role played by ADK2 in apoptosis, whereby it has been suggested that ADK2 activates a novel apoptotic pathway through formation of a complex with FADD (Fas-Associated protein with Death Domain) and Caspase-10 (AFAC10 complex) that may be involved in tumorigenesis. This complex then mediates an amplification loop that ensures the execution of apoptosis, although the decision as to whether cells live or die is made by the apoptosome[143].

### 4.3.2 Why study hADK2 conformational dynamics

In large variants of adenylate kinase, the AMP and ATP substrates are buried in a domain that undergoes conformational changes from an open to a closed state when bound to substrate; the ligand is then contained within a highly specific environment required for catalysis. Enzymes dynamically access structures with higher free energies than those of the stable ground states. These high-energy states are transient (**Figure 4.6**) and not sufficiently populated to permit direct spectroscopic observation. However, the extent and manner by which protein motions and conformational dynamics (especially during those transient high-energy states) underlie the function and regulation of these important enzymes are still not well understood[144].


**Figure 4.6.** A minimal kinetic mechanism of ADK catalysis indicates the presence of a transient high-energy state (highlighted in red) that is simultaneously open and substrate bound [144].

ADKs are generally considered as an excellent model to study protein dynamics due to conformational changes fromdomain movements. We mainly focused on human ADK2 enzyme based on the following criteria: (i) Generally, ADK catalyzes the reversible conversion of ATP and AMP into two ADP molecules (**Figure 4.4**); this allows measurements of enzyme dynamics during stable ground states and high-energy states including the transients ones. (ii) Different crystal structures of human ADKs (including hADK2) and several other organisms are known

either free or in complex with substrates and inhibitors. (iii) The hADK2 isoform is monomeric (26.4kDa) and its high-quality NMR spectra enable the analysis of dynamics. (iv) ADKs have two substrate lids that close upon substrate binding as a requirement for phosphoryl transfer.

Some experimental studies have performed on the conformational transitions of ADK from different bacteria.NMR approach was used to examine the native energy landscape of adenylate kinase in its different forms (free form, complex with natural substrates, and with a tightly binding inhibitor). Binding of ATP induces a dynamic equilibrium, before the NMP domain,in which the ATP binding motif populates both the open and the closed conformations with almost equal populations. The same event happens for AMP binding, which induces equilibrium between open and closed conformations of the AMP binding motif. These ATP- and AMP-bound structural ensembles represent complexes that exist transiently during catalysis. The binding of AMP and ATP is simultaneously required to force both substrate binding motifs to close cooperatively, which is highly cooperative in nature. Dynamical correlation map shows interesting cross-peak between LID and NMP domain which suggests the presence of correlated motion between them. Briefly, ADK substrates act to shift pre-existing equilibria toward catalytically active states[138,145].

Consistently, another team found that the additional of $Mg^{2+}$ cofactor activates two distinct molecular events, phosphoryl transfer and lid-opening [146]. On contrary, some studies showed that free ADK without any substrate can also form the ‘‘close” conformation, at the microsecond to millisecond time scale, along the enzymatic reaction trajectory[147,148].

Later, Crystal structure of human adenylate kinase 4 (ADK4) L171P mutant suggested the role of hinge IV region in protein domain motionby dramatically changing the orientation of the LID domain, which has been described as a novel twisted and-closed conformation in contrast to the open and closed conformations in other ADKs[149].

The direct NMR-based detection of the conformational dynamics of a catalytically indispensable high-energy state of an adenylate kinase 2 (ADK2) variants from bacteria (*Aquifex aeolicus*) have been reported. Kovermann team found that the *Aquifex*’s ADK2 catalytic activity

is characterized by its dynamic interconversion between the high-energy state and a ground state structure as characterized via NMR relaxation dispersion experiments[144]. The same team developed a chemical approach whereby the introduced covalent disulfide bond allowed a high-energy state of apo-ADK to be arrested. This enabled the comprehensive biophysical characterization of the structure, dynamics, and activity of the crucial conformation. They discovered that the structural sampling of the substrate free enzyme corresponds to the complete amplitude that is associated with formation of the closed and catalytically active state. Furthermore, they found that the ligand binding affinity is improved in the trapped high-energy state, compared with the wild-type enzyme. This demonstrates that substrate binding to the high-energy state is not occluded by steric hindrance[150].

Notably, many details of the ADK functional mechanism remain unclear and a complete picture of the whole process is still lacking. For instance, even though a random binding mechanism is commonly accepted, the order in which the ligands bind is not entirely clarified[149,151,152]. Furthermore, there is a contradiction found in models describing the conformational changes happening in the enzyme. Some authors suggest an independent motion of the domains[147,153], while others propose a concerted one [138,154].

In this context, the coordination information for the domain motions or the existence of transition and intermediate states during the hADK2 action in presence of Ap5a would represent an important breakthrough by using the NMR techniques such as relaxation dispersion measurements and paramagnetic approaches through insertion of LBT into the protein loop. Although, proteins with similar functions have similar active sites and exhibit similar ligand binding interactions, despite global differences in sequence and structure; we aimed atproving that the patterns of regulated dynamics and motions are functionally important for human Adenylate Kinases activity.

## 4.4 Results

### 4.4.1 Protein expression and purification of hADK2

The hADK2 was obtained by PCR using its respective primers and digested with XhoI and NdeI and then was cloned into pET28a vector digested by the same restriction enzymes (XhoI and NdeI). After the protein expression, the target protein was confirmed by SDS-PAGE (**Figure 4.6a**). We have further purified hADK2 protein by the 16/60 Superdex 75 size exclusion column to separate it from other unspecified proteins, and was collected and then concentrated to 15mg/ml for further experiments (**Figure 4.6b**).

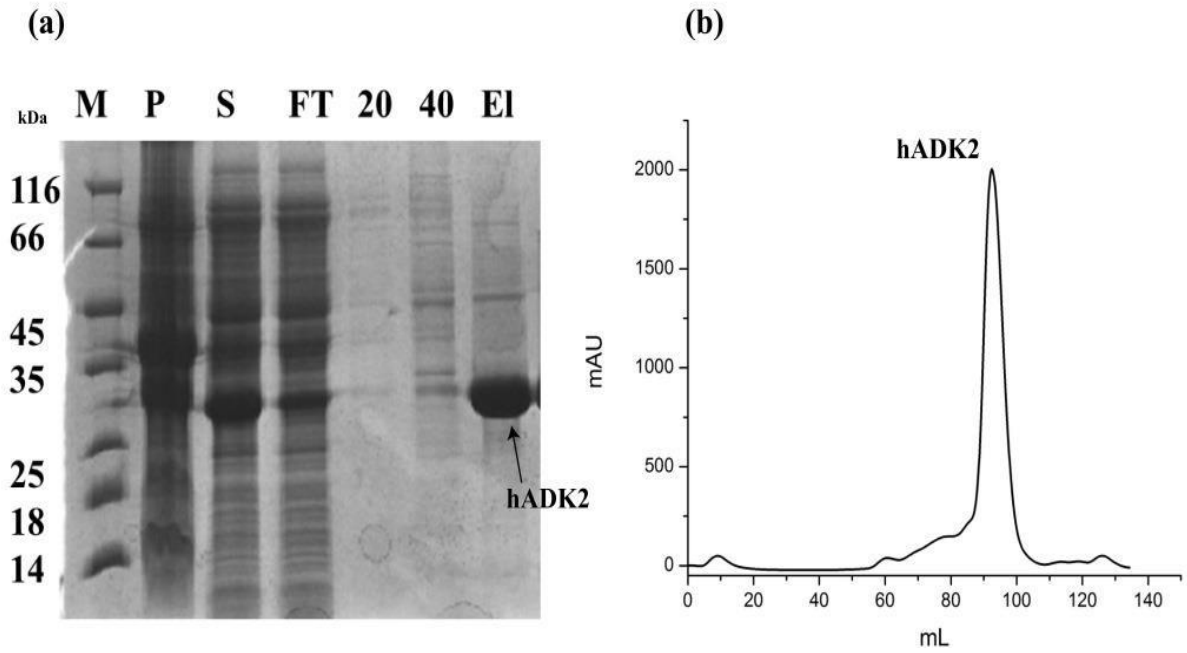


**Figure 4.6. The hADK2 expression and purification. (a)** The expression of the hADK2 (26.4kDa) confirmed by SDS-PAGE. Lanes M: Molecular weight; P: Pellet; S: Supernatant; FT: Flow through; 20 and 40: Imidazole concentrations (Mm) of washing buffer; EL: Eluted protein. **(b)** Purification of the hADK2 by Size exclusion chromatography.

### 4.4.2 Free and bound states of hADK2

We compared the HSQC spectra of the free hADK2 (open state) and hADK2 bound (closed state) by the Ap5a substrate. Notably, P1, P5-di(adenosine-50) pentaphosphate (Ap5a) consists of one ATP and one AMP molecule connected by a phosphoryl group, and is a good structural mimic of the enzyme's two native substrates.

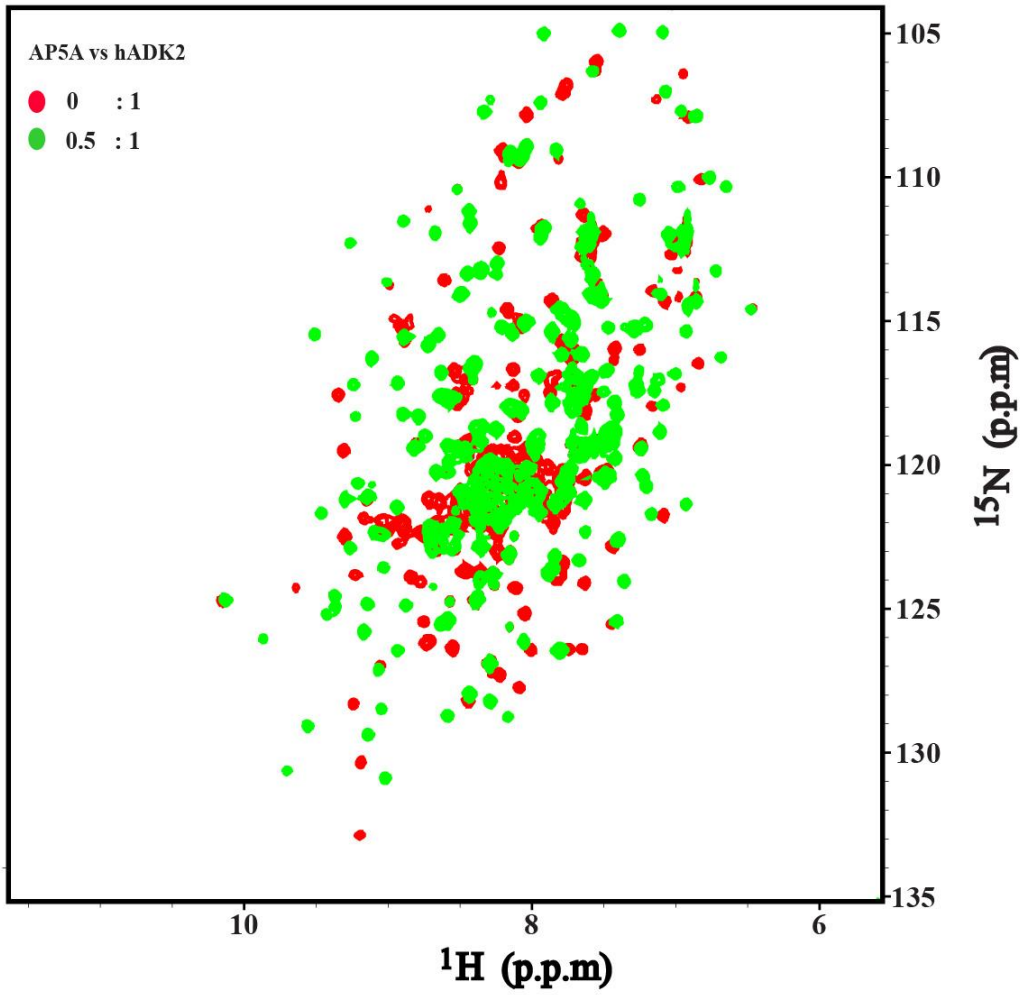


**Figure 4.7**: Free and bound states of hADK2. $^{1}H$-$^{15}N$-HSQC spectrum overlaying hADK2-free (red) and hADK2-bound by saturating concentrations of Ap5a substrate inhibitor (green) acquired at 25$^{o}$C, demonstrating that opening and closing dynamic is detected.

The **Figure 4.7** shows that there are significant spectral differences between Ap5a-saturated conditions to the apo-hADK2 and this implies that the enzyme in apo-state fluctuates between the open and closed conformations. Furthermore, in the presence of the Ap5a substrate,

most of the expected resonances are visible in NMR spectrum, but some of these resonances are missing in the NMR spectrum of the unbound hADK2, this may due to the conformational interchange regime of the apo-hADK2 upon ligand binding. These results are consistent with the previous study which suggested that the ADKE experiences transient high-energy state that is simultaneously open and substrate bound [144]. This spectra superimposition also reveals the strong ability of Ap5a to interact with hADK2 and make it more compact.

### 4.4.3 Relaxation dispersions for Ap5a-bound hADK2

We have used the $^{15}$N-CPMG NMR relaxation dispersion experiments to quantify the physical parameters for a two-site dynamic process in the presence of inhibitor Ap5a. The overall transverse relaxation rate ($R_2^{eff}$) is measured from the increased dephasing of coherence caused by the conformation exchange in the microsecond-to-millisecond timescale, resulting in additional line-broadening of NMR signals by an amount ($R_{ex}$), which can be refocused by increasing the power of an applied radio-frequency field ($\nu_{CPMG}$).

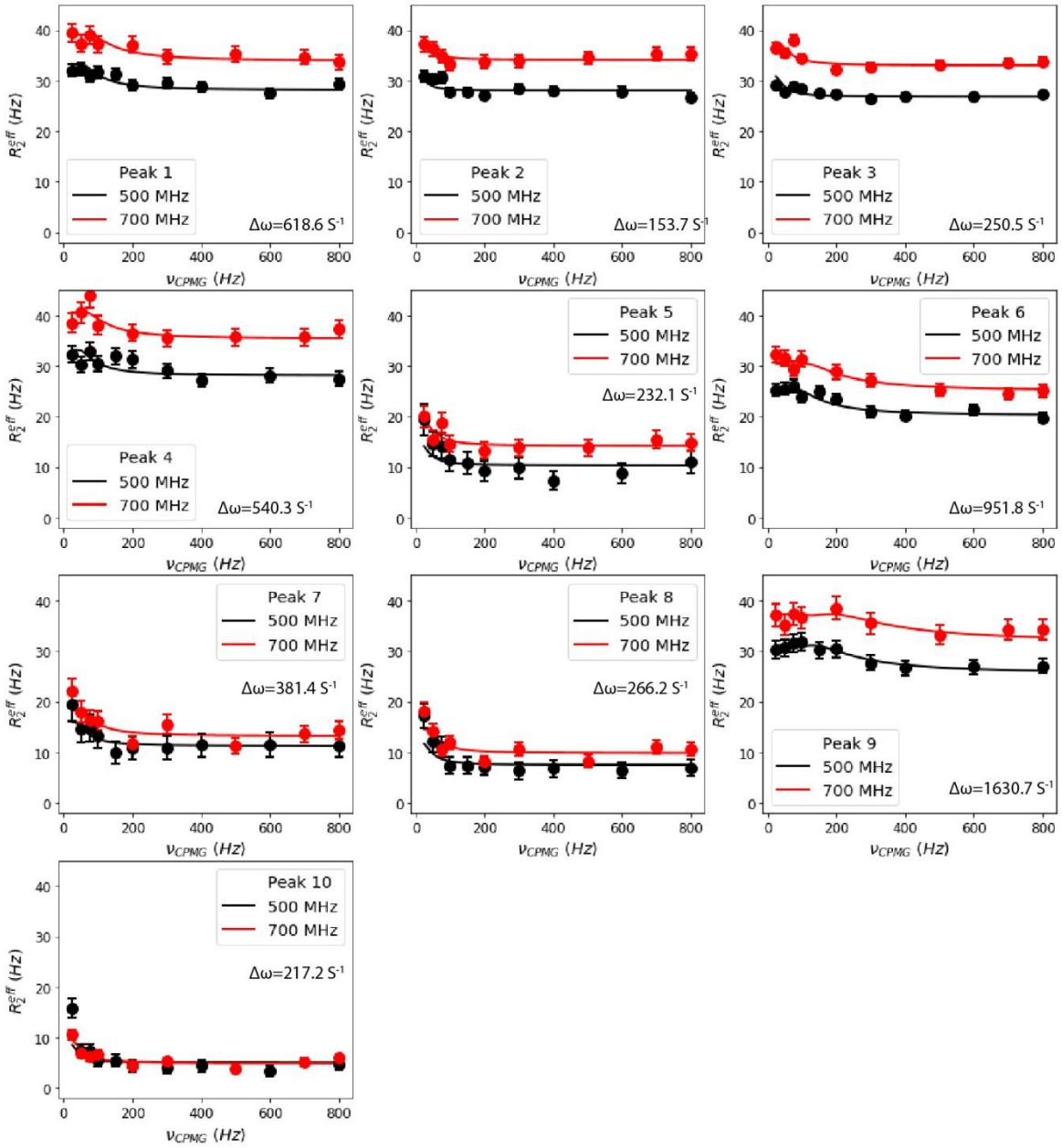


**Figure 4.8.** NMR Relaxation dispersions profiles of representative exchanging residues sensitive to opening-closing state in the presence of Ap5a ligand on $^{2}H$-$^{15}N$ hADK2. Data were acquired with 700 MHz (red) and 500 MHz (black) spectrometers, respectively at 25 °C.

$^{15}N$ NMR relaxation dispersion experiments were performed on Ap5a-bound hADK2, these experiments allowed to find out the residues which undergo the fluctuations in their local environments and kinetics characterization, and consequently detectable chemical exchanges [155]. Dispersion profiles were analyzed simultaneously to extract common exchange rate ($k_{ex}$) of

108.4 ±0.18 $S^{-1}$ and low excited state population $p_{E=}$ 0.05, along with values of chemical shift exchange (Δω) for each residue as indicated on **figure 4.8**.

These results suggested that the enzyme fluctuates between the closed and open conformation in slow exchange regime as the $k_{ex}$<<Δω with the equilibrium shifted towards the closed form by considering the representative exchanging residues of $^{2}H/^{15}N$ hADK2, in the presence of the inhibitor Ap5a whose both nucleotides (ATP and AMP) are covalently connected.

### 4.4.4 Design and insertion of Lanthanide Binding Tags into hADK2

Lanthanide-tagged proteins are valuable for exploiting the unique properties of Ln ions for investigating protein structure, function, and dynamics. Introduction of the Ln into the target is accomplished via chemical modification with synthetic lanthanide-chelating prosthetic groups or by coexpression with peptide-based binding tags known as Lanthanide Binding Tags (LBTs). The paramagnetic lanthanide ions induce pseudocontact shifts that yield valuable distance constraints for investigating the functions and dynamics for both proteins and protein complexes and have been applied in vivo.

We insertedthe Lanthanide Binding Tags (LBTs) into hADK2, such that we may determine the chemical shifts (so called pseudocontact shifts) of the excited states measured under the paramagnetic conditions. This allows having valuable structural information of the excited states. To utilize lanthanide-binding coexpression tags in NMR spectroscopy, the lanthanide tag must be well ordered with respect to the protein. One method to reduce the conformational dynamics is to restrain both termini of the LBT to decrease interdomain motion, which could be achieved by making the LBT integral to the protein sequence. Therefore, it may be possible to predict appropriate placement of LBTs in structures on the basis of analyses of protein structures or directly from the protein sequence [128].

The X-ray crystal structure of hADK2 (PDB entry 2c9y) was used as the starting structure. We have checked the structure to see which region is appropriate and suitable to insert LBT in order to avoid potential steric repulsions. We then chose the loop region GKLVS (71-75) of hADK2 denoted "γ" as a tentative insertion point of LBT tags with different lengths: long

(**YIDTNNDGWYEGDELLT**) denoted "**1**" and short (**IDTNNDGWYEGDELLT**) denoted "**2**" resulting in two different loop-LBT constructs: hADK2-LBT1γ and hADK2-LBT2γ.

All LBTs were inserted by using QuickChange II Site-Directed Mutagenesis kit (Agilent).For all NMR spectroscopic applications, $^{15}$N-labeled proteins were expressed in M9media and induced by IPTG for 20 hours at 16$^{o}$C. Expression of the hADK2-LBT1γand hADK2-LBT2γ resulted in very good yields of the soluble fusion protein (40-60 mg/L). The fractions, containing hADK2-LBT1γ and hADK2-LBT2γ from expressed, harvested and purified, were subjected to size-exclusion chromatography to yield the desired LBT-tagged hADK2 products (**Figure 4.9**).

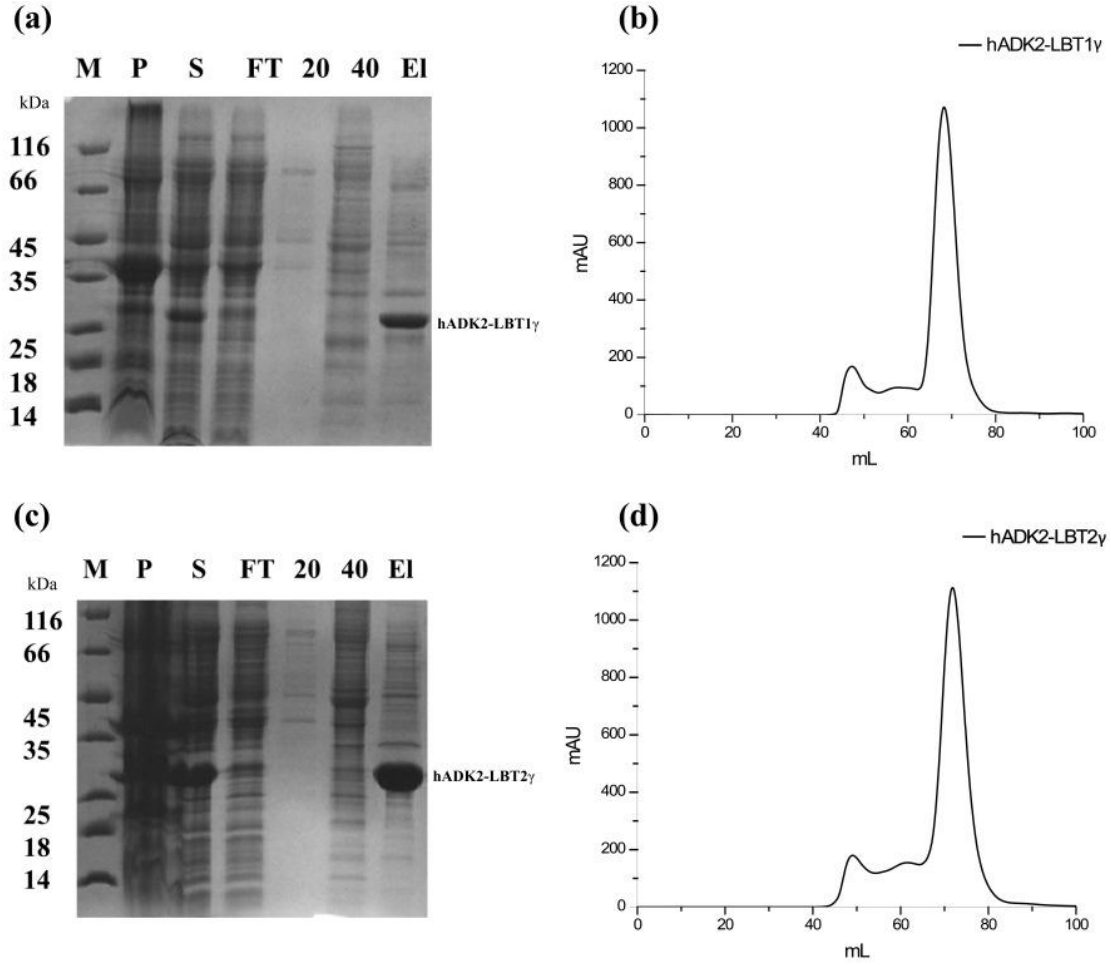


**Figure 4.9.The expression and purification of hADK2-LBT1γ and hADK2-LBT2γ.(a, c)** The expression of the hADK2-LBT1γ and hADK2-LBT2γ were confirmed by SDS-PAGE, respectively. Lanes M: Molecular weight; P: Pellet; S: Supernatant; FT: Flow through; 20 and 40: Imidazole concentrations (Mm) of washing buffer; EL: Eluted protein. **(b, d)**Size-exclusion of chromatograms for the hADK2-LBT1γ and hADK2-LBT2γ, respectively.

We then checked if the inserted LBT tags did not alter significantly the normal structure of the hADK2 protein, fortunately we realized that the whole spectrum remains largely unchanged by considering the uniformity of the superimposed spectra of with and without LBT inserted (**Figure 4.10**). However, there is the minimal structural perturbation probably induced by the LBTs onto the protein structure. We may consider the surplus of the residues (green) as the additional residues from the inserted LBT tags.

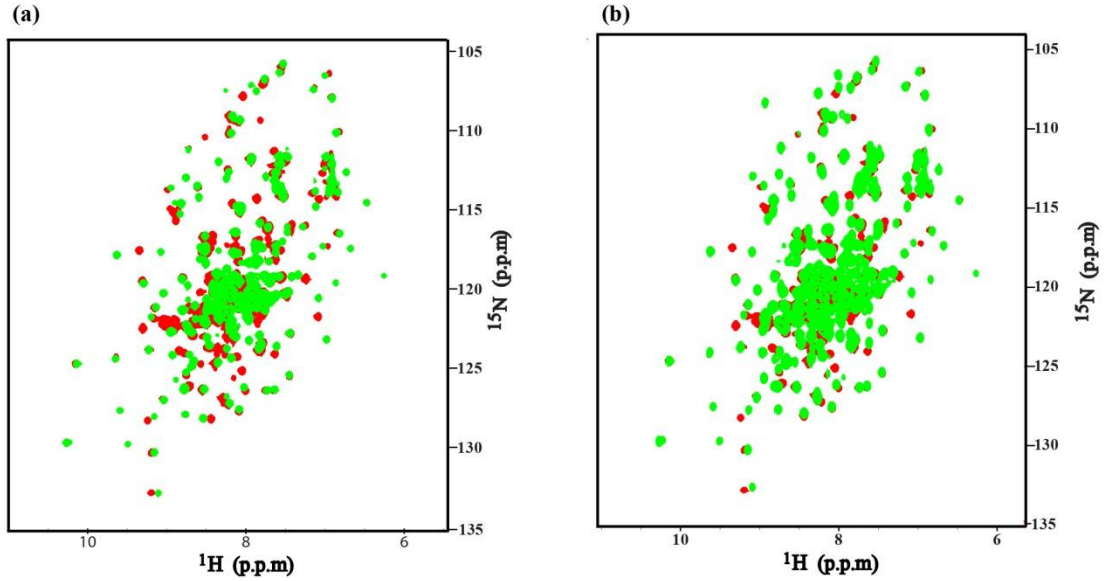


**Figure 4.10**. **The inserted loop-LBTs into hADK2.** No significant change in the fusion protein occurs from insertion of the LBTs, as shown by $^{15}$N-HSQC spectra overlaying hADK2 without LBT tag (red) and **(a)**hADK2-LBT1γ (green); **(b)** hADK2-LBT2γ (green). Only some minor, local deviations are apparent probably at the insertion site itself.

### 4.4.5 The hADK2-LBT-tagged titrated by Ap5a

The **Figure 4.11** displays a comparison of Ap5a bound and unbound states which show a high degree of similarity for both hADK2-LBT tagged types after titration. Through deep observation, we noted that the perturbed residues are almost the same in both cases.

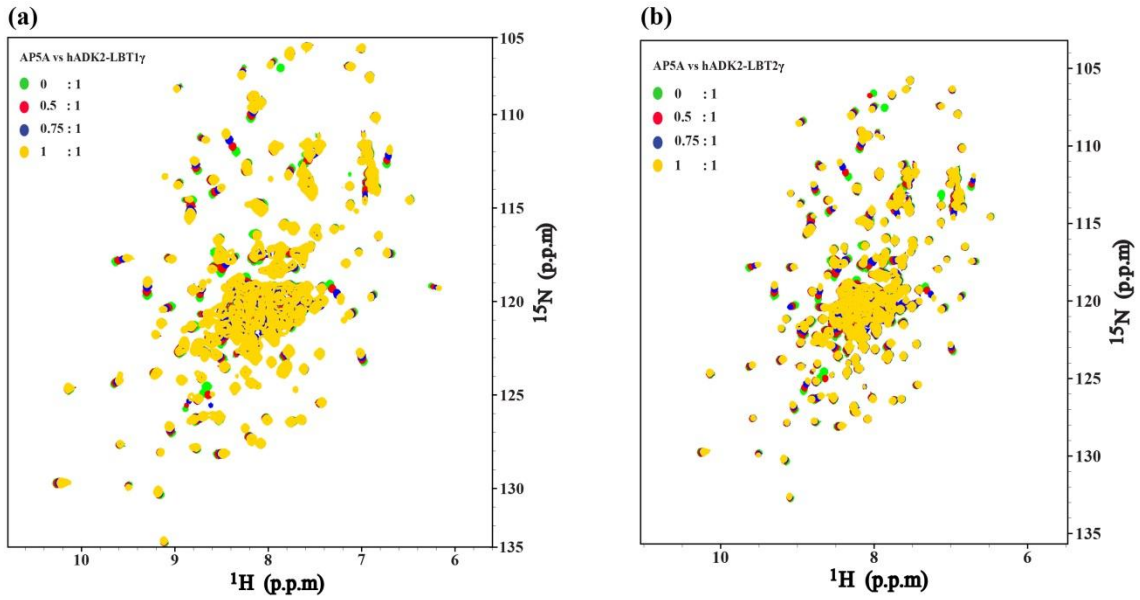


**Figure 4.11. NMR-HSQC spectral perturbation study of closed-states of ADK2-LBT1γ and ADK2-LBT2γ.** Overlay of the $^{1}$H-$^{15}$N-HSQC spectra of the hADK2-LBT1γ **(a)** and hADK2-LBT2γ **(b)** titrated with increasing molar-ratios of Ap5a. The Ap5a/protein molar ratios are annotated.

A comparison, between apo and the Ap5a-bound states of hADK2-LBTs-tagged, is made based on $^{1}$H-$^{15}$N NMR HSQC spectra that reveal profound spectral differences shown by the chemical shift changes for some residues explaining an average of the shifts from the open to closed Ap5a-bound states regarding some residues, which is in consistency with the previous study [145].

Apparently, both hADK2-LBTs-tagged variants behave similarly at closed Ap5a-bound state. However, we may suggest that these chemical shifts observed for some residues during open-to-closed transition may depend on substantial changes to the electronic environment from

the highly charged Ap5a molecule because the Ap5a changes both the overall protein conformation and the local chemical environment close to the interaction site. Thus, those disturbed residues may be located on both NMP-binding and ATP-binding domains where the Ap5a substrate binds or on the Helix 9 which actually plays an important role in in the enzyme's opening-to-closing dynamics and so facilitates the two domains (LID and NMP) to move closer and be involved in closed state arrangement [144,156].

### 4.4.6 The hADK2-LBTs-tagged titrated by paramagnetic ions $Tb^{3+}$ and $Tm^{3+}$

To observe the PCS effect, we then titrated the hADK2-LBT-tagged with $Tm^{3+}$ and $Tb^{3+}$ (**Figure 4.12**). In our case study, the metal-free protein (hADK2-LBT1γ) serves as an acceptable diamagnetic reference[157].

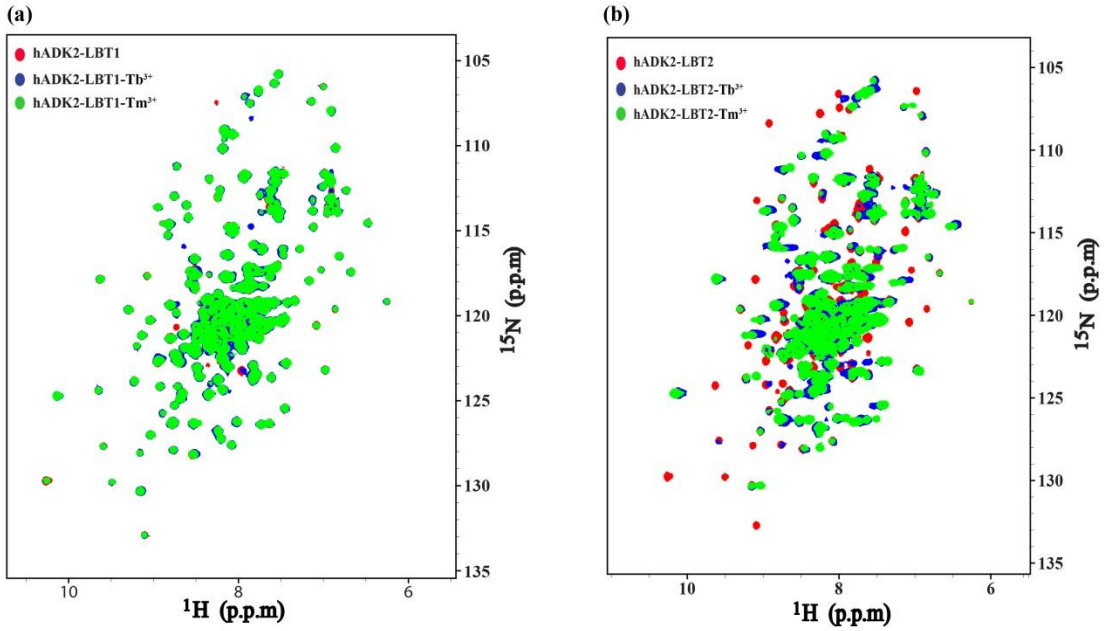


**Figure 4.12.$^{15}$N HSQC spectra of the variants. (a)** hADK2-LBT1γ (red) and **(b)** hADK2-LBT2γ (red), respectively, in complex with $Tb^{3+}$(blue) and $Tm^{3+}$ (green).

The resonances in the $^{1}H$-$^{15}N$ HSQC spectra for $Tb^{3+}$, $Tm^{3+}$, and the hADK2-LBT1γ wild type are not displaced on diagonal lines due to the similarity of the superimposed 2D spectra in which the paramagnetic shifts are not observed. The lanthanide binding did not induce noticeable chemical shift changes in a majority of the protein as observed (**Figure 4.12a**). However, in

**Figure 4.12b** with the short LBT tag lacking the first Y amino acid, we may observe some residues experiencing ambiguous small changes in presence of $Tb^{3+}$and $Tm^{3+}$ paramagnetic lanthanides ions on hADK2-LBT2γ $^{1}H$-$^{15}N$ HSQC spectrum. However, these shifts may due to flexibility of LBTs which exhibitsignificant mobility relative to the target molecule, the reason for which remains unclear.

Paramagnetic shifts and signal broadening could be observed in the $^{1}H$-$^{15}N$ HSQC spectra of the paramagnetic protein complexes, demonstrating the utility of loop-LBTs for the measurement of paramagnetic effects. The paramagnetic NMR spectra of loop-LBT wild type and mutants were not assigned due to limited time even though we tried our best to do it and the $\Delta\chi$ tensor components were not calculated because of lack of the pseudocontact shifts.

## 4.5 Discussion and Future perspectives

In biology many proteins exert their activity assuming different conformations and passing from the active to inactive state. Enzyme function depends on transitions from the ground state to higher energy states. Deviations from the ground-state structure result from chemical reactivity and conformational flexibility. Therefore, an understanding of biological function must include knowledge of the kinetics and energetics of conformational and chemical dynamical processes experienced by proteins[158]. The ground states include substrate-free and substrate-bound states, but high-resolution ground state structures alone cannot fully explain enzymes' catalytic power. Their interdomain interactions, ligand binding, and covalent modifications involve changes in motion and conformational equilibrium in a manner that can be correlated with function [159].

Consequently, Nuclear magnetic resonance (NMR) methods provide complementary information about protein conformation and dynamics in solution. NMR has the unique capacity to investigate dynamic properties of molecules over a range of time scales at atomic resolution insolution[115]. In particular, high resolution solution heteronuclear ($^{2}H$, $^{13}C$, and $^{15}N$) spin relaxation NMR techniques have been developed and applied extensively for investigation of dynamic processes in proteins on picosecond-nanosecond (ps-ns) and on microsecond-millisecond (μs-ms) time scales.

Although wild type hADK2 is saturated with Ap5a (**Figure 4.7**), a significant proportion of hADK2 may remain in a partially open conformation. Our results may be consistent with different previous studies. Firstly, this model was compared to an induced-fit model where the transition of ADKe from an open state to a closed state is initiated by ATP binding, while with Ap5A, there is a 40:60 split between partially open and fully closed conformations[160]. Also this event, in which closed and open substrate-bound states have been sampled simultaneously in substrate-free apo-enzyme, has been previously demonstrated [144]. On another hand, the previous study showed that motions in apo *Aquifex* Adk occur preferentially in the direction of the catalytically competent closed conformation. The NMR dynamic experiments revealed that in the absence of substrate, *Aquifex* Adk sampled a minor state comparable to the closed state observed during catalysis. Its conformational transition was relatively slow (microsecond-to-

millisecond) occurring on the timescale of catalytic turnover and thought to be facilitated by individual high-frequency local fluctuations[147].

We note that in the slow-exchange regime, we expected to see signals for both states in the HSQC spectrum, which we did observe here for Ap5a-bound state of hADK2-LBT tagged (**Figure 4.10**). A plausible explanation for this phenomenon is a high population of the open state that give rise to strong signals. In addition, there is no line broadening of signals corresponding to the open state, which is considerably faster than the closing rate, and did not broaden the already strong signals during the detection. Furthermore, the presence of NMR dispersions is also caused by motions during the open-to-close transition upon Ap5a binding[146].

The similarity of the $^{1}H$-$^{15}N$ HSQC spectra of wild type hADK2 and the LBT insertion mutants shows the minimal structural perturbation induced by the LBTs onto the protein structure. Although, a small degree of heterogeneity is observed with some peaks in hADK2-LBT2γ $^{1}H$-$^{15}N$ HSQC spectra,it is unclear whether the LBT or lanthanides ions contributed to that heterogeneity. Reason why it is worthy to pursue this studies by adding the ligand to the titrated hADK2-LBT2γ for investigating the conformational exchange using $^{15}N/^{1}H$ CPMG or $^{15}N$ CEST [96].

Furthermore, the use of two-loop-LBT approach may also be useful in future work due to its advantage as it only needs to design and prepare a single protein construct. This approach is also suitable to augment and cross validate studies on dynamics of multi-domain complexes [161]. In the same line, large residual dipolar couplings (RDCs) could be determined by NMR spectroscopy for all LBT-loop constructs and reveal if the LBT2 series was rigidly incorporated into the hADK2 structure.

# References


1. Venter, J.C. The sequence of the human genome. *Science* **2001**, *291*, 1304-1351, doi:10.1126/science.1058040.

2. Afroz, T.; Cienikova, Z.; Clery, A.; Allain, F.H. One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. *Methods Enzymol* **2015**, *558*, 235-278, doi:10.1016/bs.mie.2015.01.015.

3. Clery, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **2008**, *18*, 290-298, doi:10.1016/j.sbi.2008.04.002.

4. Kielkopf, C.L.; Lucke, S.; Green, M.R. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **2004**, *18*, 1513-1526, doi:10.1101/gad.1206204.

5. Liu, X.; Niu, C.; Ren, J.; Zhang, J.; Xie, X.; Zhu, H.; Feng, W.; Gong, W. The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site. *Biochim Biophys Acta* **2013**, *1832*, 375-385, doi:10.1016/j.bbadis.2012.11.012.

6. Afroz, T.; Skrisovska, L.; Belloc, E.; Guillen-Boixet, J.; Mendez, R.; Allain, F.H. A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes Dev* **2014**, *28*, 1498-1514, doi:10.1101/gad.241133.114.

7. Kielkopf, C.L.; Rodionova, N.A.; Green, M.R.; Burley, S.K. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **2001**, *106*, 595-605.

8. Lukavsky, P.J.; Daujotyte, D.; Tollervey, J.R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F.E.; Damberger, F.F.; Allain, F.H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol* **2013**, *20*, 1443-1449, doi:10.1038/nsmb.2698.

9. Kuo, P.H.; Chiang, C.H.; Wang, Y.T.; Doudeva, L.G.; Yuan, H.S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res* **2014**, *42*, 4712-4722, doi:10.1093/nar/gkt1407.

10. Kuo, P.H.; Doudeva, L.G.; Wang, Y.T.; Shen, C.K.; Yuan, H.S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res* **2009**, *37*, 1799-1808, doi:10.1093/nar/gkp013.

11. Ou, S.; Wu, F.; Harrich, D.; Garcia-Martinez, L.; Gaynor, R. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Virol* **1995**, *69*, 3584-3596.

12. Buratti, E.; Dork, T.; Zuccato, E.; Pagani, F.; Romano, M.; Baralle, F.E. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* **2001**, *20*, 1774-1784, doi:10.1093/emboj/20.7.1774.

13. Ayala, Y.M.; Pantano, S.; D'Ambrogio, A.; Buratti, E.; Brindisi, A.; Marchetti, C.; Romano, M.; Baralle, F.E. Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J Mol Biol* **2005**, *348*, 575-588, doi:10.1016/j.jmb.2005.02.038.

14. Banks, G.T.; Kuta, A.; Isaacs, A.M.; Fisher, E.M. TDP-43 is a culprit in human neurodegeneration, and not just an innocent bystander. *Mamm Genome* **2008**, *19*, 299-305, doi:10.1007/s00335-008-9117-x.

15. Lagier-Tourenne, C.; Polymenidou, M.; Cleveland, D.W. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum Mol Genet* **2010**, *19*, R46-64, doi:10.1093/hmg/ddq137.

16. Warraich, S.T.; Yang, S.; Nicholson, G.A.; Blair, I.P. TDP-43: a DNA and RNA binding protein with roles in neurodegenerative diseases. *Int J Biochem Cell Biol* **2010**, *42*, 1606-1609, doi:10.1016/j.biocel.2010.06.016.

17. Buratti, E.; Brindisi, A.; Giombi, M.; Tisminetzky, S.; Ayala, Y.M.; Baralle, F.E. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J Biol Chem* **2005**, *280*, 37572-37584, doi:10.1074/jbc.M505557200.

18. Mercado, P.A.; Ayala, Y.M.; Romano, M.; Buratti, E.; Baralle, F.E. Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Res* **2005**, *33*, 6000-6010, doi:10.1093/nar/gki897.

19. Strong, M.J.; Volkening, K.; Hammond, R.; Yang, W.C.; Strong, W.; Leystra-Lantz, C.; Shoesmith, C. TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol Cell Neurosci* **2007**, *35*, 320-327, doi:10.1016/j.mcn.2007.03.007.

20. Ayala, Y.M.; Misteli, T.; Baralle, F.E. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *P Natl Acad Sci USA* **2008**, *105*, 3785-3789, doi:10.1073/pnas.0800546105.

21. Weskamp, K.; Barmada, S.J. TDP43 and RNA instability in amyotrophic lateral sclerosis. *Brain Res* **2018**, *1693*, 67-74, doi:10.1016/j.brainres.2018.01.015.

22. Anderson, P.; Kedersha, N. RNA granules: post-transcriptional and epigenetic modulators of gene expression. *Nat Rev Mol Cell Biol* **2009**, *10*, 430-436, doi:10.1038/nrm2694.

23. Kedersha, N.; Stoecklin, G.; Ayodele, M.; Yacono, P.; Lykke-Andersen, J.; Fritzler, M.J.; Scheuner, D.; Kaufman, R.J.; Golan, D.E.; Anderson, P. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J Cell Biol* **2005**, *169*, 871-884, doi:10.1083/jcb.200502088.

24. Kedersha, N.; Tisdale, S.; Hickman, T.; Anderson, P. Real-time and quantitative imaging of mammalian stress granules and processing bodies. *Methods Enzymol* **2008**, *448*, 521-552, doi:10.1016/S0076-6879(08)02626-8.

25. Kedersha, N.; Anderson, P. Mammalian stress granules and processing bodies. *Methods Enzymol* **2007**, *431*, 61-81, doi:10.1016/S0076-6879(07)31005-7.

26. Colombrita, C.; Zennaro, E.; Fallini, C.; Weber, M.; Sommacal, A.; Buratti, E.; Silani, V.; Ratti, A. TDP-43 is recruited to stress granules in conditions of oxidative insult. *J Neurochem* **2009**, *111*, 1051-1061, doi:10.1111/j.1471-4159.2009.06383.x.

27. Bentmann, E.; Neumann, M.; Tahirovic, S.; Rodde, R.; Dormann, D.; Haass, C. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J Biol Chem* **2012**, *287*, 23079-23094, doi:10.1074/jbc.M111.328757.

28. Liu-Yesucevitz, L.; Bilgutay, A.; Zhang, Y.J.; Vanderweyde, T.; Citro, A.; Mehta, T.; Zaarur, N.; McKee, A.; Bowser, R.; Sherman, M., et al. Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. *PLoS One* **2010**, *5*, e13250, doi:10.1371/journal.pone.0013250.

29. Freibaum, B.D.; Chitta, R.K.; High, A.A.; Taylor, J.P. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *J Proteome Res* **2010**, *9*, 1104-1120, doi:10.1021/pr901076y.

30. McDonald, K.K.; Aulas, A.; Destroismaisons, L.; Pickles, S.; Beleac, E.; Camu, W.; Rouleau, G.A.; Vande Velde, C. TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum Mol Genet* **2011**, *20*, 1400-1410, doi:10.1093/hmg/ddr021.

31. Monahan, Z.; Shewmaker, F.; Pandey, U.B. Stress granules at the intersection of autophagy and ALS. *Brain Research* **2016**, *1649*, 189-200, doi:10.1016/j.brainres.2016.05.022.

32. Chen, Y.C., T.J. Aggregation of the nucleic acid–binding protein TDP-43 occurs via distinct routes that are coordinated with stress granule formation. *J. Biol. Chem.* **2019**.

33. Wilson, C.A., Dugger, B. Dickson, D. W., Wang, D. S. TDP-43 in aging and Alzheimer's disease. *Int J Clin Exp Pathol* **2011**, *4*, 147-155.

34. Sun, M.; Yamashita, T.; Shang, J.; Liu, N.; Deguchi, K.; Liu, W.; Ikeda, Y.; Feng, J.; Abe, K. Acceleration of TDP43 and FUS/TLS protein expressions in the preconditioned hippocampus following repeated transient ischemia. *J Neurosci Res* **2014**, *92*, 54-63, doi:10.1002/jnr.23301.

35. Josephs, K.A.; Whitwell, J.L.; Weigand, S.D.; Murray, M.E.; Tosakulwong, N.; Liesinger, A.M.; Petrucelli, L.; Senjem, M.L.; Knopman, D.S.; Boeve, B.F., et al. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathol* **2014**, *127*, 811-824, doi:10.1007/s00401-014-1269-z.

36. Neumann, M.; Sampathu, D.M.; Kwong, L.K.; Truax, A.C.; Micsenyi, M.C.; Chou, T.T.; Bruce, J.; Schuck, T.; Grossman, M.; Clark, C.M., et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **2006**, *314*, 130-133, doi:10.1126/science.1134108.

37. Kraemer, B.C.; Schuck, T.; Wheeler, J.M.; Robinson, L.C.; Trojanowski, J.Q.; Lee, V.M.; Schellenberg, G.D. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol* **2010**, *119*, 409-419, doi:10.1007/s00401-010-0659-0.

38. Neumann, M. Molecular neuropathology of TDP-43 proteinopathies. *Int J Mol Sci* **2009**, *10*, 232-246, doi:10.3390/ijms10010232.

39. Takeda, T. Possible concurrence of TDP-43, tau and other proteins in amyotrophic lateral sclerosis/frontotemporal lobar degeneration. *Neuropathology* **2018**, *38*, 72-81, doi:10.1111/neup.12428.

40. Strong, M.J.; Abrahams, S.; Goldstein, L.H.; Woolley, S.; McLaughlin, P.; Snowden, J.; Mioshi, E.; Roberts-South, A.; Benatar, M.; HortobaGyi, T., et al. Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria.

*Amyotroph Lateral Scler Frontotemporal Degener* **2017**, *18*, 153-174, doi:10.1080/21678421.2016.1267768.

41. Ash, P.E.; Zhang, Y.J.; Roberts, C.M.; Saldi, T.; Hutter, H.; Buratti, E.; Petrucelli, L.; Link, C.D. Neurotoxic effects of TDP-43 overexpression in C. elegans. *Hum Mol Genet* **2010**, *19*, 3206-3218, doi:10.1093/hmg/ddq230.

42. Li, Y.; Ray, P.; Rao, E.J.; Shi, C.; Guo, W.; Chen, X.; Woodruff, E.A., 3rd; Fushimi, K.; Wu, J.Y. A Drosophila model for TDP-43 proteinopathy. *Proc Natl Acad Sci U S A* **2010**, *107*, 3169-3174, doi:10.1073/pnas.0913602107.

43. Voigt, A.; Herholz, D.; Fiesel, F.C.; Kaur, K.; Muller, D.; Karsten, P.; Weber, S.S.; Kahle, P.J.; Marquardt, T.; Schulz, J.B. TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS One* **2010**, *5*, e12247, doi:10.1371/journal.pone.0012247.

44. Bhandare, V.V.; Ramaswamy, A. The proteinopathy of D169G and K263E mutants at the RNA Recognition Motif (RRM) domain of tar DNA-binding protein (tdp43) causing neurological disorders: A computational study. *J Biomol Struct Dyn* **2018**, *36*, 1075-1093, doi:10.1080/07391102.2017.1310670.

45. Anand, P.; Kunnumakkara, A.B.; Newman, R.A.; Aggarwal, B.B. Bioavailability of curcumin: problems and promises. *Mol Pharm* **2007**, *4*, 807-818, doi:10.1021/mp700113r.

46. Duvoix, A.; Blasius, R.; Delhalle, S.; Schnekenburger, M.; Morceau, F.; Henry, E.; Dicato, M.; Diederich, M. Chemopreventive and therapeutic effects of curcumin. *Cancer Lett* **2005**, *223*, 181-190, doi:10.1016/j.canlet.2004.09.041.

47. Srivastava, R.K.; Chen, Q.; Siddiqui, I.; Sarva, K.; Shankar, S. Linkage of curcumin-induced cell cycle arrest and apoptosis by cyclin-dependent kinase inhibitor p21(/WAF1/CIP1). *Cell Cycle* **2007**, *6*, 2953-2961, doi:10.4161/cc.6.23.4951.

48. Fang, H.Y.; Chen, S.B.; Guo, D.J.; Pan, S.Y.; Yu, Z.L. Proteomic identification of differentially expressed proteins in curcumin-treated MCF-7 cells. *Phytomedicine* **2011**, *18*, 697-703, doi:10.1016/j.phymed.2010.11.012.

49. Nan, Y.N.; Zhu, J.Y.; Tan, Y.; Zhang, Q.; Jia, W.; Hua, Q. Staurosporine induced apoptosis rapidly downregulates TDP- 43 in glioma cells. *Asian Pac J Cancer Prev* **2014**, *15*, 3575-3579.

50. Kim, P.Y.; Tan, O.; Liu, B.; Trahair, T.; Liu, T.; Haber, M.; Norris, M.D.; Marshall, G.M.; Cheung, B.B. High TDP43 expression is required for TRIM16-induced inhibition of cancer cell growth and correlated with good prognosis of neuroblastoma and breast cancer patients. *Cancer Lett* **2016**, *374*, 315-323, doi:10.1016/j.canlet.2016.02.021.

51. Zhou, Y.L.; Li, Y.D.; Guo, F.J. Expression of TDP43 in cervical cancer. *Int J Clin Exp Patho* **2016**, *9*, 1467-1473.

52. Fan, Z.; Cui, H.; Xu, X.; Lin, Z.; Zhang, X.; Kang, L.; Han, B.; Meng, J.; Yan, Z.; Yan, X., et al. MiR-125a suppresses tumor growth, invasion and metastasis in cervical cancer by targeting STAT3. *Oncotarget* **2015**, *6*, 25266-25280, doi:10.18632/oncotarget.4457.

53. Kamangar, F.; Dores, G.M.; Anderson, W.F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol* **2006**, *24*, 2137-2150, doi:10.1200/JCO.2005.05.2308.

54. Meijer, C.J.; Snijders, P.J. Cervical cancer in 2013: Screening comes of age and treatment progress continues. *Nat Rev Clin Oncol* **2014**, *11*, 77-78, doi:10.1038/nrclinonc.2013.252.

55. Olszanski, A.J. Current and future roles of targeted therapy and immunotherapy in advanced melanoma. *J Manag Care Spec Pharm* **2014**, *20*, 346-356, doi:10.18553/jmcp.2014.20.4.346.

56. Park, Y.Y.; Kim, S.B.; Han, H.D.; Sohn, B.H.; Kim, J.H.; Liang, J.Y.; Lu, Y.L.; Rodriguez-Aguayo, C.; Lopez-Berestein, G.; Mills, G.B., et al. Tat-Activating Regulatory DNA-Binding Protein Regulates Glycolysis in Hepatocellular Carcinoma by Regulating

the Platelet Isoform of Phosphofructokinase Through MicroRNA 520. *Hepatology* **2013**, *58*, 182-191, doi:10.1002/hep.26310.

57. Slominski, A.; Kim, T.K.; Brozyna, A.A.; Janjetovic, Z.; Brooks, D.L.; Schwab, L.P.; Skobowiat, C.; Jozwicki, W.; Seagroves, T.N. The role of melanogenesis in regulation of melanoma behavior: melanogenesis leads to stimulation of HIF-1alpha expression and HIF-dependent attendant pathways. *Arch Biochem Biophys* **2014**, *563*, 79-93, doi:10.1016/j.abb.2014.06.030.

58. Zeng, Q.; Cao, K.; Liu, R.; Huang, J.; Xia, K.; Tang, J.; Chen, X.; Zhou, M.; Xie, H.; Zhou, J. Identification of TDP-43 as an oncogene in melanoma and its function during melanoma pathogenesis. *Cancer Biol Ther* **2017**, *18*, 8-15, doi:10.1080/15384047.2016.1250984.

59. Liu, X.; Zheng, J.; Xue, Y.; Qu, C.; Chen, J.; Wang, Z.; Li, Z.; Zhang, L.; Liu, Y. Inhibition of TDP43-Mediated SNHG12-miR-195-SOX5 Feedback Loop Impeded Malignant Biological Behaviors of Glioma Cells. *Mol Ther Nucleic Acids* **2018**, *10*, 142-158, doi:10.1016/j.omtn.2017.12.001.

60. Yuan, X.; Curtin, J.; Xiong, Y.; Liu, G.; Waschsmann-Hogiu, S.; Farkas, D.L.; Black, K.L.; Yu, J.S. Isolation of cancer stem cells from adult glioblastoma multiforme. *Oncogene* **2004**, *23*, 9392-9400, doi:10.1038/sj.onc.1208311.

61. Moran, C. Importance of molecular features of NSCLC for choice of treatment. *Am. J. Pathol.* **2011**, *178*, 1940-2198.

62. Guo, F.; Jiao, F.; Song, Z.; Li, S.; Liu, B.; Yang, H.; Zhou, Q.; Li, Z. Regulation of MALAT1 expression by TDP43 controls the migration and invasion of non-small cell lung cancer cells in vitro. *Biochem Biophys Res Commun* **2015**, *465*, 293-298, doi:10.1016/j.bbrc.2015.08.027.

63. Ke, H.; Zhao, L.; Zhang, H.; Feng, X.; Xu, H.; Hao, J.; Wang, S.; Yang, Q.; Zou, L.; Su, X., et al. Loss of TDP43 inhibits progression of triple-negative breast cancer in

coordination with SRSF3. *Proc Natl Acad Sci U S A* **2018**, *115*, E3426-E3435, doi:10.1073/pnas.1714573115.

64. Boxer, A.L.; Gold, M.; Huey, E.; Gao, F.B.; Burton, E.A.; Chow, T.; Kao, A.; Leavitt, B.R.; Lamb, B.; Grether, M., et al. Frontotemporal degeneration, the next therapeutic frontier: molecules and animal models for frontotemporal degeneration drug development. *Alzheimers Dement* **2013**, *9*, 176-188, doi:10.1016/j.jalz.2012.03.002.

65. Gayle, A.Y.; Baranger, A.M. Inhibition of the U1A–RNA Complex by an Aminoacridine Derivative. *Bioorg.Med. Chem.Lett.* **2002**, *12*, 2839–2842.

66. Varani, G.; Nagai, K. RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **1998**, *27*, 407-445, doi:10.1146/annurev.biophys.27.1.407.

67. Cassel, J.A.; Blass, B.E.; Reitz, A.B.; Pawlyk, A.C. Development of a novel nonradiometric assay for nucleic acid binding to TDP-43 suitable for high-throughput screening using AlphaScreen technology. *J Biomol Screen* **2010**, *15*, 1099-1106, doi:10.1177/1087057110382778.

68. Cassel, J.A.; McDonnell, M.E.; Velvadapu, V.; Andrianov, V.; Reitz, A.B. Characterization of a series of 4-aminoquinolines that stimulate caspase-7 mediated cleavage of TDP-43 and inhibit its function. *Biochimie* **2012**, *94*, 1974-1981, doi:10.1016/j.biochi.2012.05.020.

69. Vidal, R.L.; Matus, S.; Bargsted, L.; Hetz, C. Targeting autophagy in neurodegenerative diseases. *Trends Pharmacol Sci* **2014**, *35*, 583-591, doi:10.1016/j.tips.2014.09.002.

70. Aguzzi, A.; O'Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov* **2010**, *9*, 237-248, doi:10.1038/nrd3050.

71. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* **2003**, *4*, 49-60, doi:10.1038/nrn1007.

72. Gossert, A.D.; Jahnke, W. NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. *Prog Nucl Magn Reson Spectrosc* **2016**, *97*, 82-125, doi:10.1016/j.pnmrs.2016.09.001.

73. Meyer, B.; Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* **2003**, *42*, 864-890, doi:10.1002/anie.200390233.

74. Hajduk, P.J.; Bures, M.; Praestgaard, J.; Fesik, S.W. Privileged molecules for protein binding identified from NMR-based screening. *J Med Chem* **2000**, *43*, 3443-3447.

75. Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. NMR-based screening in drug discovery. *Q Rev Biophys* **1999**, *32*, 211-240.

76. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531-1534.

77. Fesik, S.W. NMR structure-based drug design. *J Biomol Nmr* **1993**, *3*, 261-269.

78. Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T.L.; Homans, S.W.; Kessler, H.; Luchinat, C., et al. Perspectives on NMR in drug discovery: a technique comes of age. *Nat Rev Drug Discov* **2008**, *7*, 738-745, doi:10.1038/nrd2606.

79. Becker, W.; Bhattiprolu, K.C.; Gubensak, N.; Zangger, K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem* **2018**, *19*, 895-906, doi:10.1002/cphc.201701253.

80. Li, Y.; Kang, C. Solution NMR Spectroscopy in Target-Based Drug Discovery. *Molecules* **2017**, *22*, doi:10.3390/molecules22091399.

81. Viegas, A.; Manso, J.o.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *Journal of Chemical Education* **2011**, *88*, 990-994, doi:10.1021/ed101169t.

82. Dalvit, C.; Caronni, D.; Mongelli, N.; Veronesi, M.; Vulpetti, A. NMR-based quality control approach for the identification of false positives and false negatives in high throughput screen. *Curr. Drug Discovery Technol.* **2006**, *3*, 115−124.

83. Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J Biomol Nmr* **2000**, *18*, 65-68.

84. Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J Biomol Nmr* **2001**, *21*, 349-359, doi:Doi 10.1023/A:1013302231549.

85. Hajduk, P.J.; Olejniczak, E.T.; Fesik, S.W. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. *Journal of the American Chemical Society* **1997**, *119*, 12257-12261, doi:DOI 10.1021/ja9715962.

86. Moschen, T.; Grutsch, S.; Juen, M.A.; Wunderlich, C.H.; Kreutz, C.; Tollinger, M. Measurement of Ligand-Target Residence Times by (1)H Relaxation Dispersion NMR Spectroscopy. *J Med Chem* **2016**, *59*, 10788-10793, doi:10.1021/acs.jmedchem.6b01110.

87. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Mag Res Sp* **2013**, *73*, 1-16, doi:10.1016/j.pnmrs.2013.02.001.

88. Ishima, R. Protein-Inhibitor Interaction Studies Using NMR. *Appl NMR Spectrosc* **2015**, *1*, 143-181, doi:10.2174/978160805962111501007.

89. Kim, H.Y.; Wyss, D.F. NMR screening in fragment-based drug design: a practical guide. *Methods Mol Biol* **2015**, *1263*, 197-208, doi:10.1007/978-1-4939-2269-7_16.

90. Gao, J.; Ma, R.; Wang, W.; Wang, N.; Sasaki, R.; Snyderman, D.; Wu, J.; Ruan, K. Automated NMR fragment based screening identified a novel interface blocker to the LARG/RhoA complex. *PLoS One* **2014**, *9*, e88098, doi:10.1371/journal.pone.0088098.

91. Delaglio, F.; Grzesiek, S.; Vuister, G.W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol Nmr* **1995**, *6*, 277-293, doi:10.1007/bf00197809.

92. G.C.P., v.Z.; Rodrigues, J.P.G.L.M.; Trellet, M.; Schmitz, C.; Kastritis, P.L.; Karaca, E.; Melquiond, A.S.J.; van Dijk, M.; de Vries, S.J.; Bonvin, A.M.J.J. The HADDOCK2.2 webserver: User-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **2015**, *428*, 720-725

93. Wassenaar, T.A., et al. WeNMR: Structural Biology on the Grid. *J Grid Comput* **2012**, *10*, 743-767, doi:10.1007/s10723-012-9246-z.

94. Schuttelkopf, A.W.; van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 1355-1363, doi:10.1107/S0907444904011679.

95. Carver, J.P.; Richards, R.E. A general two-site solution for the chemical exchange produced dependence of T2 upon the Carr-Purcell pulse separation. . *J. Magn. Reson.* **1972**, *6*, 89-105.

96. Xu, D.F.; Li, B.; Gao, J.; Liu, Z.J.; Niu, X.G.; Nshogoza, G.; Zhang, J.H.; Wu, J.H.; Su, X.C.; He, W., et al. Ligand Proton Pseudocontact Shifts Determined from Paramagnetic Relaxation Dispersion in the Limit of NMR Intermediate Exchange. *J Phys Chem Lett* **2018**, *9*, 3361-3367, doi:10.1021/acs.jpclett.8b01443.

97. Korzhnev, D.M.; Kay, L.E. Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc Chem Res* **2008**, *41*, 442-451, doi:10.1021/ar700189y.

98. Nshogoza, G.; Liu, Y.; Gao, J.; Liu, M.; Moududee, S.A.; Ma, R.; Li, F.; Zhang, J.; Wu, J.; Shi, Y., et al. NMR Fragment-Based Screening against Tandem RNA Recognition Motifs of TDP-43. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20133230.

99. Zega, A. NMR Methods for Identification of False Positives in Biochemical Screens. *J Med Chem* **2017**, *60*, 9437-9447, doi:10.1021/acs.jmedchem.6b01520.

100. Liu, J.; Gao, J.; Li, F.; Ma, R.; Wei, Q.; Wang, A.; Wu, J.; Ruan, K. NMR characterization of weak interactions between RhoGDI2 and fragment screening hits. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2017**, *1861*, 3061-3070, doi:10.1016/j.bbagen.2016.10.003.

101. Aguirre, C.; ten Brink, T.; Cala, O.; Guichou, J.F.; Krimm, I. Protein–ligand structure guided by backbone and side-chain proton chemical shift perturbations. *J Biomol Nmr* **2014**, *60*, 147–156.

102. Aguirre, C.; ten Brink, T.; Guichou, J.F.; Cala, O.; Krimm, I. Comparing Binding Modes of Analogous Fragments Using NMR in Fragment-Based Drug Design: Application to PRDX5. *Plos One* **2014**, *9*, doi:ARTN e10230010.1371/journal.pone.0102300.

103. Harrison, A.F.; Shorter, J. RNA-binding proteins with prion-like domains in health and disease. *Biochem J* **2017**, *474*, 1417-1438, doi:10.1042/BCJ20160499.

104. Shortridge, M.D.; Hage, S.H.; Harbison, G.S.; Powers, R. Estimating Protein-Ligand Binding Affinity using High-Throughput Screening by NMR. *J Comb Chem.* **2008**, *10*, 948–958.

105. Mashalidis, E.H.; Sledz, P.; Lang, S.; Abell, C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc* **2013**, *8*, 2309-2324, doi:10.1038/nprot.2013.130.

106. Kamerlin, S.C.; Sharma, P.K.; Prasad, R.B.; Warshel, A. Why nature really chose phosphate. *Q Rev Biophys* **2013**, *46*, 1-132, doi:10.1017/S0033583512000157.

107. Lassila, J.K.; Zalatan, J.G.; Herschlag, D. Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem* **2011**, *80*, 669-702, doi:10.1146/annurev-biochem-060409-092741.

108. Tzeng, S.R.; Kalodimos, C.G. Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* **2011**, *21*, 62-67, doi:10.1016/j.sbi.2010.10.007.

109. Mittermaier, A.; Kay, L.E. New tools provide new insights in NMR studies of protein dynamics. *Science* **2006**, *312*, 224-228, doi:10.1126/science.1124964.

110. Ishima, R.; Torchia, D.A. Protein dynamics from NMR. *Nat Struct Biol* **2000**, *7*, 740-743, doi:10.1038/78963.

111. Clore, G.M.; Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **2009**, *109*, 4108-4139, doi:10.1021/cr900033p.

112. Li, C.; Tang, C.; Liu, M. Protein dynamics elucidated by NMR technique. *Protein Cell* **2013**, *4*, 726-730, doi:10.1007/s13238-013-3912-1.

113. Gaspari, Z.; Perczel, A. Protein Dynamics as Reported by NMR. *Annu Rep Nmr Spectro* **2010**, *71*, 35-75, doi:10.1016/S0066-4103(10)71002-4.

114. Ziarek, J.J.; Peterson, F.C.; Lytle, B.L.; Volkman, B.F. Binding site identification and structure determination of protein-ligand complexes by NMR a semiautomated approach. *Methods Enzymol* **2011**, *493*, 241-275, doi:10.1016/B978-0-12-381274-2.00010-8.

115. Palmer, A.G.; Kroenke, C.D.; Patrick Loria, J. Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. **2001**, *339*, 204-238, doi:10.1016/s0076-6879(01)39315-1.

116. Bouvignies, G.; Vallurupalli, P.; Kay, L.E. Visualizing Side Chains of Invisible Protein Conformers by Solution NMR. *Journal of Molecular Biology* **2014**, *426*, 763-774, doi:10.1016/j.jmb.2013.10.041.

117. Vallurupalli, P.; Bouvignies, G.; Kay, L.E. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* **2012**, *134*, 8148-8161, doi:10.1021/ja3001419.

118. Ravera, E.; Carlon, A.; Fragai, M.; Parigi, G.; Luchinat, C. Paramagnetic NMR as a new tool in structural biology. *Emerging Topics in Life Sciences* **2018**, *2*, 19-28, doi:10.1042/etls20170084.

119. Otting, G. Protein NMR using paramagnetic ions. *Annu Rev Biophys* **2010**, *39*, 387-405, doi:10.1146/annurev.biophys.093008.131321.

120. Pintacuda, G.; Keniry, M.A.; Huber, T.; Park, A.Y.; Dixon, N.E.; Otting, G. Fast structure-based assignment of 15N HSQC spectra of selectively 15N-labeled paramagnetic proteins. *J Am Chem Soc* **2004**, *126*, 2963-2970, doi:10.1021/ja039339m.

121. Bertini, I.; Luchinat, C.; Parigi, G. Magnetic susceptibility in paramagnetic NMR. *Prog Nucl Mag Res Sp* **2002**, *40*, 249-273, doi:Doi 10.1016/S0079-6565(02)00002-X.

122. Sjodt, M.; Macdonald, R.; Spirig, T.; Chan, A.H.; Dickson, C.F.; Fabian, M.; Olson, J.S.; Gell, D.A.; Clubb, R.T. The PRE-Derived NMR Model of the 38.8-kDa Tri-Domain IsdH Protein from Staphylococcus aureus Suggests That It Adaptively Recognizes Human Hemoglobin. *J Mol Biol* **2016**, *428*, 1107-1129, doi:10.1016/j.jmb.2015.02.008.

123. Keniry, M.A.; Park, A.Y.; Owen, E.A.; Hamdan, S.M.; Pintacuda, G.; Otting, G.; Dixon, N.E. Structure of the theta subunit of Escherichia coli DNA polymerase III in complex with the epsilon subunit. *J Bacteriol* **2006**, *188*, 4464-4473, doi:10.1128/JB.01992-05.

124. Loh, C.T.; Graham, B.; Abdelkader, E.H.; Tuck, K.L.; Otting, G. Generation of pseudocontact shifts in proteins with lanthanides using small "clickable" nitrilotriacetic acid and iminodiacetic acid tags. *Chemistry* **2015**, *21*, 5084-5092, doi:10.1002/chem.201406274.

125. Yang, Y.; Wang, J.T.; Pei, Y.Y.; Su, X.C. Site-specific tagging proteins via a rigid, stable and short thiolether tether for paramagnetic spectroscopic analysis. *Chem Commun (Camb)* **2015**, *51*, 2824-2827, doi:10.1039/c4cc08493d.

126. Loh, C.T.; Ozawa, K.; Tuck, K.L.; Barlow, N.; Huber, T.; Otting, G.; Graham, B. Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. *Bioconjug Chem* **2013**, *24*, 260-268, doi:10.1021/bc300631z.

127. Park, S.H.; Wang, V.S.; Radoicic, J.; De Angelis, A.A.; Berkamp, S.; Opella, S.J. Paramagnetic relaxation enhancement of membrane proteins by incorporation of the metal-chelating unnatural amino acid 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid (HQA). *J Biomol Nmr* **2015**, *61*, 185-196, doi:10.1007/s10858-014-9884-5.

128. Barthelmes, K.; Reynolds, A.M.; Peisach, E.; Jonker, H.R.; DeNunzio, N.J.; Allen, K.N.; Imperiali, B.; Schwalbe, H. Engineering encodable lanthanide-binding tags into loop regions of proteins. *J Am Chem Soc* **2011**, *133*, 808-819, doi:10.1021/ja104983t.

129. Martin, L.J.; Imperiali, B. The best and the brightest: exploiting tryptophan-sensitized Tb(3+) luminescence to engineer lanthanide-binding tags. *Methods Mol Biol* **2015**, *1248*, 201-220, doi:10.1007/978-1-4939-2020-4_14.

130. Su, X.C.; McAndrew, K.; Huber, T.; Otting, G. Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. *J Am Chem Soc* **2008**, *130*, 1681-1687, doi:10.1021/ja076564l.

131. Barb, A.W.; Ho, T.G.; Flanagan-Steet, H.; Prestegard, J.H. Lanthanide binding and IgG affinity construct: potential applications in solution NMR, MRI, and luminescence microscopy. *Protein Sci* **2012**, *21*, 1456-1466, doi:10.1002/pro.2133.

132. Saio, T.; Ogura, K.; Yokochi, M.; Kobashigawa, Y.; Inagaki, F. Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect. *J Biomol Nmr* **2009**, *44*, 157-166, doi:10.1007/s10858-009-9325-z.

133. Barb, A.W.; Subedi, G.P. An encodable lanthanide binding tag with reduced size and flexibility for measuring residual dipolar couplings and pseudocontact shifts in large proteins. *J Biomol Nmr* **2016**, *64*, 75-85, doi:10.1007/s10858-015-0009-6.

134. Allen, K.N.; Imperiali, B. Lanthanide-tagged proteins--an illuminating partnership. *Curr Opin Chem Biol* **2010**, *14*, 247-254, doi:10.1016/j.cbpa.2010.01.004.

135. Otting, G. Prospects for lanthanides in structural biology by NMR. *J Biomol Nmr* **2008**, *42*, 1-9, doi:10.1007/s10858-008-9256-0.

136. John, M.; Otting, G. Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. *Chemphyschem* **2007**, *8*, 2309-2313, doi:10.1002/cphc.200700510.

137. Panayiotou, C.; Solaroli, N.; Karlsson, A. The many isoforms of human adenylate kinases. *Int J Biochem Cell Biol* **2014**, *49*, 75-83, doi:10.1016/j.biocel.2014.01.014.

138. Jana, B.; Adkar, B.V.; Biswas, R.; Bagchi, B. Dynamic coupling between the LID and NMP domain motions in the catalytic conversion of ATP and AMP to ADP by adenylate kinase. *Journal of Chemical Physics* **2011**, *134*, doi:Artn 03510110.1063/1.3516588.

139. Noma, T.; Song, S.; Yoon, Y.S.; Tanaka, S.; Nakazawa, A. cDNA cloning and tissue-specific expression of the gene encoding human adenylate kinase isozyme 2. *Biochim Biophys Acta* **1998**, *1395*, 34-39, doi:10.1016/s0167-4781(97)00193-0.

140. Noma, T.; Fujisawa, K.; Yamashiro, Y.; Shinohara, M.; Nakazawa, A.; Gondo, T.; Ishihara, T.; Yoshinobu, K. Structure and expression of human mitochondrial adenylate kinase targeted to the mitochondrial matrix. *Biochem J* **2001**, *358*, 225-232, doi:10.1042/0264-6021:3580225.

141. Lagresle-Peyrou, C.; Six, E.M.; Picard, C.; Rieux-Laucat, F.; Michel, V.; Ditadi, A.; Demerens-de Chappedelaine, C.; Morillon, E.; Valensi, F.; Simon-Stoos, K.L., et al. Human adenylate kinase 2 deficiency causes a profound hematopoietic defect associated with sensorineural deafness. *Nat Genet* **2009**, *41*, 106-111, doi:10.1038/ng.278.

142. Pannicke, U.; Honig, M.; Hess, I.; Friesen, C.; Holzmann, K.; Rump, E.M.; Barth, T.F.; Rojewski, M.T.; Schulz, A.; Boehm, T., et al. Reticular dysgenesis (aleukocytosis) is caused by mutations in the gene encoding mitochondrial adenylate kinase 2. *Nature Genetics* **2009**, *41*, 101-105, doi:10.1038/ng.265.

143. Lee, H.J.; Pyo, J.O.; Oh, Y.; Kim, H.J.; Hong, S.H.; Jeon, Y.J.; Kim, H.; Cho, D.H.; Woo, H.N.; Song, S., et al. AK2 activates a novel apoptotic pathway through formation of a complex with FADD and caspase-10. *Nat Cell Biol* **2007**, *9*, 1303-U1176, doi:10.1038/ncb1650.

144. Kovermann, M.; Aden, J.; Grundstrom, C.; Sauer-Eriksson, A.E.; Sauer, U.H.; Wolf-Watz, M. Structural basis for catalytically restrictive dynamics of a high-energy enzyme state. *Nat Commun* **2015**, *6*, 7644, doi:10.1038/ncomms8644.

145. Aden, J.; Wolf-Watz, M. NMR identification of transient complexes critical to adenylate kinase catalysis. *J Am Chem Soc* **2007**, *129*, 14003-14012, doi:10.1021/ja075055g.

146. Kerns, S.J.; Agafonov, R.V.; Cho, Y.J.; Pontiggia, F.; Otten, R.; Pachov, D.V.; Kutter, S.; Phung, L.A.; Murphy, P.N.; Thai, V., et al. The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol* **2015**, *22*, 124-131, doi:10.1038/nsmb.2941.

147. Henzler-Wildman, K.A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M.A.; Petsko, G.A.; Karplus, M., et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838-844, doi:10.1038/nature06410.

148. Henzler-Wildman, K.A.; Lei, M.; Thai, V.; Kerns, S.J.; Karplus, M.; Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **2007**, *450*, 913-U927, doi:10.1038/nature06407.

149. Liu, R.; Xu, H.; Wei, Z.; Wang, Y.; Lin, Y.; Gong, W. Crystal structure of human adenylate kinase 4 (L171P) suggests the role of hinge region in protein domain motion. *Biochemical and Biophysical Research Communications* **2009**, *379*, 92-97, doi:10.1016/j.bbrc.2008.12.012.

150. Kovermann, M.; Grundstrom, C.; Sauer-Eriksson, A.E.; Sauer, U.H.; Wolf-Watz, M. Structural basis for ligand binding to an enzyme by a conformational selection pathway. *Proc Natl Acad Sci U S A* **2017**, *114*, 6298-6303, doi:10.1073/pnas.1700919114.

151. Halder, R.; Manna, R.N.; Chakraborty, S.; Jana, B. Modulation of the Conformational Dynamics of Apo-Adenylate Kinase through a pi-Cation Interaction. *J Phys Chem B* **2017**, *121*, 5699-5708, doi:10.1021/acs.jpcb.7b01736.

152. Song, H.D.; Zhu, F. Conformational dynamics of a ligand-free adenylate kinase. *PLoS One* **2013**, *8*, e68023, doi:10.1371/journal.pone.0068023.

153. Pontiggia, F.; Zen, A.; Micheletti, C. Small- and Large-Scale Conformational Changes of Adenylate Kinase: A Molecular Dynamics Study of the Subdomain Motion and Mechanics. *Biophys. J.* **2008**, *95*, 5901–5912.

154. Whitford, P.C.; Miyashita, O.; Levy, Y.; Onuchic, J.N. Conformational transitions of adenylate kinase: switching by cracking. *J Mol Biol* **2007**, *366*, 1661-1671, doi:10.1016/j.jmb.2006.11.085.

155. Palmer, A.G. Enzyme Dynamics from NMR Spectroscopy. *Accounts of Chemical Research* **2015**, *48*, 457-465, doi:10.1021/ar500340a.

156. Brokaw, J.B.; Chu, J.W. On the roles of substrate binding and hinge unfolding in conformational changes of adenylate kinase. *Biophys J* **2010**, *99*, 3420-3429, doi:10.1016/j.bpj.2010.09.040.

157. Jensen, M.R.; Led, J.J. Metal-protein interactions: structure information from Ni2+-induced pseudocontact shifts in a native nonmetalloprotein. *Biochemistry-Us* **2006**, *45*, 8782–8787.

158. Bilderback, T.; Fulmer, T.; Mantulin, W.W.; Glaser, M. Substrate binding causes movement in the ATP binding domain of Escherichia coli adenylate kinase. *Biochemistry-Us* **1996**, *35*, 6100-6106, doi:10.1021/bi951833i.

159. Schulz, G.E. Induced-fit movements in adenylate kinases. *Faraday Discuss* **1992**, 10.1039/fd9929300085, 85-93, doi:10.1039/fd9929300085.

160. Onuk, E.; Badger, J.; Wang, Y.J.; Bardhan, J.; Chishti, Y.; Akcakaya, M.; Brooks, D.H.; Erdogmus, D.; Minh, D.D.L.; Makowski, L. Effects of Catalytic Action and Ligand Binding on Conformational Ensembles of Adenylate Kinase. *Biochemistry-Us* **2017**, *56*, 4559-4567, doi:10.1021/acs.biochem.7b00351.

161. Barthelmes, D.; Granz, M.; Barthelmes, K.; Allen, K.N.; Imperiali, B.; Prisner, T.; Schwalbe, H. Encoded loop-lanthanide-binding tags for long-range distance measurements in proteins by NMR and EPR spectroscopy. *J Biomol Nmr* **2015**, *63*, 275-282, doi:10.1007/s10858-015-9984-x.

# Acknowledgement

I would like to express my deepest gratitude to my supervisor Professor Yunyu Shi, for her kindness, encouragement, advices and suggestions during my PhD work. It is trully an honor to be admitted and work in her special and hard-working laboratory.

I am sincerely grateful and highly acknowledge this work to Professor Ke Ruan for his guidance, advices, trust, suggestions and encourgements during my PhD research. I'm greatly thankful for the different and huge invaluable things he taught me over the past four years. He is a real inspiring supervisor.

My grateful thanks also go to Professor Wu Jihui, Professor Zhiyong Zhang, Professor Qingguo Gong, Professor Tang Yajun, Dr. Li Fudong for their academic and technical advices to our better knowledge in Bio-NMR lab.

I thank Dr Jiuyang Liu, Dr Na Wang, Dr Ma Rongsheng and Dr Jiahai Zhang for their technical and experimental support provided to me during this PhD period.
I owe my sincere thanks to all my former and current labmates Dr Gao Jia, Dr Lv Mengqi, Dr Liu Xiaodan, Dr Chen Jiajing, Liu Yaqian, Liu Mingqi, Fan Weiwei, Xie Guodong, Xu Zheng, Zhao Shujie, Yiyang Jiang, Yu Hailong, Bao Hongyu, Xu Ling, Cheng Ling, Shen Siyuan, Liu Yongrui, Wang Chongyuan, Xie Changli, Wang Sun Man, Zhang Beibei, and Anam Nayab for their friendly help and company.
I want to thank all my friends, and countrymates for making my studies enjoyable experience in University of Science and Technology of China .

My most heartfelt thanks belong to my wife Odette Mujawamariya and our daughter Nshogoza Ihirwe Levana for their love, sacrifice and encourgements. Furthemore, I thank SEMATETU J. Damascène family and Spéciose UWAMARIYA for their understanding, endless support, motivation, and prayers.

**Acknowledgement**

Finally, my deep thanks go to the University of Science and Technology of China for the admission, to Chinese Scholarship Council for the financial support and to the government of Rwanda through the University of Rwanda for the study leave.

# List of Publications

1. **Nshogoza, G**.; Liu, Y.; Gao, J.; Liu, M.; Moududee, S.A.; Ma, R.; Li, F.; Zhang, J.; Wu, J.; Shi, Y.; Ruan, K.; NMR Fragment-Based Screening against Tandem RNA Recognition Motifs of TDP-43, *Int. J. Mol. Sci.* 2019, 20, 3230. Doi:10.3390/ijms20133230.

2. Tang, H.; **Nshogoza, G**.; Liu, M.; Liu, Y.; Ruan, K.; Ma, R.; Gao, J. Identification of Novel Hits of the NSD1 SET Domain by NMR Fragment-Based Screening. *Chin. J. Magn Reson* 2019, 36, 148–154. Doi:10.11938/cjmr20182696.

3. Liu, J.; Zhang, S.; Liu, M.; Liu, Y.; **Nshogoza, G**.; Gao, J.; Ma, R.; Yang, Y.; Wu, J.; Zhang, J., et al. Structural plasticity of the TDRD3 Tudor domain probed by a fragment screening hit. *Febs J.* 2018, 285, 2091–2103. Doi:10.1111/febs.14469.

4. Xu, D.; Li, Bin.; Gao, J.; Liu, Z.; Niu, X.; **Nshogoza, G**.; Zhang, J.; Wu, J.; Su, X. C.; He, W.; Ma, R.; Yang, D.; Ruan, K.; Ligand Proton Pseudocontact Shifts Determined from Paramagnetic Relaxation Dispersion in the Limit of NMR Intermediate Exchange, *J. Phys. Chem. Lett.* 2018, 9, 3361−3367. Doi:10.1021/acs.jpclett.8b01443.

5. Moududee, S.A.; Jiang, Y.; **Nshogoza, G**.; Xie, G.; Xu, Z.; Wu, J.; Gong, Q.; Tang, Y.; Shi, Y.; Structural and functional characterization of hMEX-3C Ring finger domain as an E3 ubiquitin ligase, *protein science,*2018, 27, 1661–1669. Doi: 10.1002/pro.3473.

6. Xie, C.; He, C.; Jiang,Y.; Yu, H.; Cheng, L.; **Nshogoza, G.**; Ala, M.S.; Tian, C.; Wu, J.; Shi, Y.; and Li, F., Structural insights into the recognition of phosphorylated Hop1 by Mek1, *Acta Crysta D: Struct Biol.*, 2018, 74, 1027-1038.Doi:10.1107/S2059798318011993.

## Academic Conference attended

- RNA structural biology symposium, 2017. Hefei, P.R. China