國立臺灣大學生命科學院生命科學系

學士班學生論文

Department of Life Science
College of Life Science
National Taiwan University

Bachelor's Thesis

C9orf72 六核苷酸重複序列與 hnRNP H1 之間的交互作用

Exploring the structure of *C9orf72* hexanucleotide repeat expansion and its interactions with hnRNP H1

張翰彧

Han-Yu Chang

指導教授：温進德 博士

Advisor: Jin-Der Wen, Ph.D.

中華民國 114 年 1 月

January, 2025

國立臺灣大學學士班學生論文
口試委員會審定書

C9orf72 六核甘酸重複序列與 hnRNP H1 之間的結構互動

Exploring the structure of C9orf72 hexanucleotide repeat expansion and its interactions with hnRNP H1

本論文係張翰彧(B10B01011)在國立臺灣大學生命科學系完成之學士班學生論文，於民國 114 年 1 月 20 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

張壯德		（簽名）
朱雪津		（指導教授）

鄭貽生

系 主 任 ：_____（簽名）

（是否須簽章依各院系規定）

誌謝

自我加入實驗室已兩年，雖然大二到大三上學期都在修課沒什麼做實驗，但仍然學到很多，也過得很快樂。首先要我想感謝溫老師兩年來的教導，老師對我所提出的問題或是犯的錯誤也總是耐心的教導我應該要如何去做。不論是當初以導生的身分前來詢問實驗室時老師的建議，又或者是在實驗上、撰寫國科會計畫以及論文時對我提供的幫助都使我受益良多。也要感謝在我第一天進實驗室開始就帶我的世綸學長，不論是實驗上或是推甄都幫了我很多，幫助我快速釐清 Tweezers 上的各種問題，尤其是 G_4C_2 的 data 分析還有蛋白純化的實驗，對我後續能夠撰寫學士論文提供非常多的幫助。也非常感謝王碩跟恩慈提供我做 EMSA 實驗的各種建議。最後也要謝謝宜倫、聖元跟知宜為實驗室帶來許多歡樂。祝大家未來都順順利利。

中文摘要

人類第九號染色體第 72 開讀框六核苷酸(GGGGCC)重複擴增（*C9orf72* HRE）是導致肌萎縮性側索硬化症（漸凍症，ALS）和額顳葉癡呆症（FTD）的常見遺傳致病因素。過去的研究指出，由 *C9orf72* HRE 形成的核醣核酸凝集體（RNA foci）在疾病的多個方面具有重要關聯性。核醣核酸凝集體的一個已知特徵是能夠捕獲核醣核酸結合蛋白（RBPs），在 *C9orf72* 導致的漸凍症患者神經元細胞中發現有多種核醣核酸結合蛋白與核醣核酸凝集體共定位，其中，hnRNP H1 是一種剪接因子，能夠結合富含鳥嘌呤的序列，被認為在 *C9orf72* 導致的 ALS/FTD 患者中占主導地位。儘管疾病的其中一種常見致病原因已被發現，但其實際致病機制仍不清楚，更不用說 hnRNP H1 在疾病中的角色。

為了探究 $r(G_4C_2)_n$ 核醣核酸重複序列的結構特徵以及 hnRNP H1 扮演的角色，我們應用單分子光鉗技術以監測核醣核酸的即時構形變化及其與 hnRNP H1 的相互作用。通過這種獨特的單分子操控與測量方法，我們可以使結構反覆解開與摺疊來確定核醣核酸重複序列的構型。此外，我們將 hnRNP H1 的截短突變體 hnRNP H1-N1 添加到單分子實驗系統中，以研究其對 GGGGCC 串聯重複序列所形成之結構的影響。我們的結果表明，低重複次數的 GGGGCC 核醣核酸重複序列無法形成單分子的 G 四疊體，但仍然表現出結構變異性。

在 hnRNP H1-N1 存在的情況下，核醣核酸結構似乎被干擾，且形成新的結構。數據顯示，hnRNP H1-N1 的結合對特定結構具有專一性。此外，目標結構在 hnRNP H1-N1 的作用下傾向不穩定且易被解開。對蛋白誘導產生的新結構的分析提供了我們對於 hnRNP H1 的功能的新理解，在後續的研究中我們將做進一步的探討。

關鍵字：人類第九號染色體第 72 開讀框六核苷酸重複擴增、漸凍症、額顳葉癡呆症、核醣核酸凝集體、hnRNP H1、光鉗技術

ABSTRACT

The human *C9orf72* hexanucleotide (GGGGCC) repeat expansion (HRE) is a common genetic cause of both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Past studies indicated RNA foci formed by *C9orf72* HRE are strongly related to the diseases in different aspects. One of the known features of RNA foci was to sequester RNA-binding proteins (RBPs). Several RBPs were found to colocalize with the RNA foci found in neuron cells of c9ALS patients. Among them, hnRNP H1, a splicing factor with a preference for binding G-rich sequences, has been found to be predominant in ALS/FTD patients with *C9orf72* HRE. Although the common cause of these diseases has been identified, the underlying pathogenic mechanisms are yet unclear. Not to mention the function of hnRNP H1 in ALS/FTD.

To peek into the structural features of $r(G_4C_2)_n$ and the role of hnRNP H1 in the diseases, optical tweezers were applied in this research to monitor the real-time conformational change of RNA and its interaction with hnRNP H1. Through the unique single-molecule manipulations and measurements, we could determine the dynamic conformations of the RNA. Furthermore, hnRNP H1-N1, a truncation mutant of hnRNP H1, was added to determine its effect on the conformation of GGGGCC tandem repeats. Our results indicate that the short repeat of GGGGCC RNA was unable to form intramolecular G quadruplexes but still exhibited structural variations.

In the presence of the protein, the GGGGCC RNA could be remodeled to new conformations, suggesting that hnRNP H1-N1's binding has a structural specificity. Moreover, the protein-targeted structure tends to be destabilized by hnRNP H1-N1. The emergence of the protein-induced structure provides a great insight into the function of hnRNP H1. We will tackle these issues in future experiments.

Keywords: *C9orf72* HRE, amyotrophic lateral sclerosis, frontotemporal dementia, RNA foci, hnRNP H1, optical tweezers.

CONTENTS

FIGURE LIST

CHAPTER 1 INTRODUCTION

1.1 ALS and FTD

Amyotrophic lateral sclerosis (ALS) is a multisystem neurodegenerative disorder that causes the loss of function in motor neurons. ALS causes muscle weakness. The affected body parts expand through the progress of disease. Though the main symptom of ALS is motor neuron dysfunction, cognitive impairments were developed in around half of the patients. Some of the patients (≈13%) also have frontal temporal dementia (FTD) (Hardiman et al., 2017). Among >30 genes associated with ALS, the hexanucleotide repeat expansion (HRE) in chromosome 9 open reading frame 72 (*C9orf72*) is responsible for around 20% of familial ALS patients (DeJesus-Hernandez et al., 2011). ALS patients with *C9orf72* HRE are likely to develop cognitive impairments as well (Masrori & Van Damme, 2020). According to past genome-wide association study (GWAS) results, the GGGGCC HRE in *C9orf72* is likely to be the most common genetic cause of ALS-FTD (Jiang et al., 2016).

1.2 Introduction to *C9orf72* HRE

C9orf72 HRE, with expanded GGGGCC tandem repeats in the first intron of *C9orf72*, has been proven to be the most common genetic cause of ALS/FTD (DeJesus-Hernandez et al., 2011). Like other nucleotide repeat expansion disorders, ALS/FTD develops when the repeat number is beyond a threshold (DeJesus-Hernandez et al., 2011). This HRE has been found to form foci *in vitro* and *in vivo*; previous studies showed that *C9orf72* RNAs are able to form various secondary structures, leading to gel-like foci with different numbers of repeats (Figure 1)(Jain & Vale, 2017). Pathogenic mechanisms of *C9orf72* HRE can be classified into the following groups: loss-of-function of the gene,

RNA toxicity and toxicity of dipeptide repeat proteins. RNA foci account for both loss-of-function of the gene and RNA toxicity, as the foci down-regulate the transcription levels of *C9orf72*. Also, RNA binding proteins will be sequestered to the foci, leading to abnormality in function (Prudencio et al., 2015). Both sense and antisense of C9orf72 HRE may undergo repeat-associated non-AUG (RAN) translation and produce 5 dipeptide repeat (DPR) proteins: poly-GA, poly-GR, poly-GP, poly-PA and poly-PR. Poly-GR is found to correlate with neurodegeneration in ALS disease (Dong et al., 2024).

1.3 RNA structures formed by *C9orf72* HRE

The pathogenic mechanisms mentioned above involve the capability of $r(G_4C_2)_n$ in forming diverse secondary structures. According to previous studies, *C9orf72* HRE can form the following structures: RNA hairpin, G-quadruplex, I-motifs and R-loops.

RNA hairpins are found in various repeat numbers, in this research, the 4-repeat sense and antisense RNA we use was also predicted to form hairpins (Figure 2). Aside from canonically paired hairpin, guanine bases in this HRE are hydrogen-bonded through Hoogsteen pairing and form a planar square or tetrad structure. These structures together formed a G-quadruplex. G-quadruplex (GQ) is found to form in intermolecular or intramolecular ways, also, it may be parallel or antiparallel (Figure 3)(Kumar et al., 2016). RNA foci commonly seen in patients of ALS/FTD have been proved to be made by G-quadruplexes from $r(G_4C_2)_n$ both *in vitro* and *in vivo* (Figure 4) (Raguseo et al., 2023).

For antisense strand of *C9orf72* HRE, the C-rich RNA forms secondary structure similar to G-quadruplex named I-motif. It is composed of two parallel duplexes in head to tail orientation, the strands are held together by hemiprotonated C-C+ pairs (Kumar et al., 2016).

C9orf72 HRE is also able to form R-loops *in vitro*. R-loops are DNA/RNA hybrid structures composed of three strands: template DNA, complemented transcribed RNA, and single-stranded DNA that is replaced by RNA and can form G-quadruplexes. This structure will lead to disruption of transcription and cause the accumulation of aborted transcripts (Kumar et al., 2016).

1.4 hnRNP H1

As mentioned in 1.1, one of the pathogenic mechanisms of *C9orf72* HRE is to sequester RBPs; in past experiments, heterogeneous nuclear ribonucleoprotein (hnRNP) H1 is a splicing factor found to be predominant in patients of *C9orf72* caused ALS/FTD (Conlon et al., 2016). The known function of hnRNP H1 is to regulate alternative splicing of RNA (Gautrey et al., 2015), also it is likely a helicase that may change RNA conformation while binding.

As shown in Figure 5, hnRNP H1 is composed of 5 domains: RNA recognition motif (RRM)1, RRM2, low complexity (LC) 1, RRM3 and LC2. Unlike most hnRNPs, hnRNP H1 was found to bind consecutive guanine bases, which were capable of forming G-quadruplexes (Geuens et al., 2016). These RRMs are also known as quasi-RRM (qRRM)s and were found to interact with G rich sequence by 'encaging' the sequence in a structurally similar protein, hnRNP F (Dominguez et al., 2010). LC1 controls the alternative splicing function and is also essential for interaction with other RBPs, whereas LC2 may function as a transcriptional activation domain (Kim & Kwon, 2021). Among the three RRMs, RRM1 and RRM2 were found to majorly interact with G-rich structures in past research (Conlon et al., 2016; Kutluay et al., 2019).

1.5 Single-molecule techniques

Optical tweezers, a technique used to study the dynamic of single biomolecules, was

proposed by Arthur Ashikin in 1970 (Ashkin, 1970). After that, the first experiment utilizing optical tweezers to trap small biomolecules (bacterium and virus) was demonstrated in 1987 (Ashkin et al., 1987). By using a highly focused laser beam, scientists can manipulate small particles (usually a micro-sized polystyrene bead) in extremely small force (pico-Newton, pN). The principle of how optical tweezers trap the particle involves the balance of two types of forces produced by laser traps: scattering forces and gradient force. Scattering forces push the target along the direction of the propagation of laser, whereas gradient forces pull the target along the opposite direction (Moffitt et al., 2008). While these forces are matched at a specific position, the target will be trapped therein. Thus, we can freely control the target position through the movement of the laser trap.

Compared to traditional molecular biology approaches, optical tweezers provide access to measure a real-time change of a single molecule. While traditional methods usually combine a group of molecules and determine their property at the same time. Though the average properties of the whole group of molecules are obtained through experiments, the behavior of a single molecule at the right time is ignored. Through optical tweezers, the real-time conformation changes of RNA molecules can be determined.

1.6 Aims

Figuring out the mechanism of how *C9orf72* HRE causes ALS/FTD is crucial in the research in ALS/FTD. As mentioned in previous sections, we have now known that this HRE can cause various pathological effects in the cell through different ways, including DNA, RNA and proteins. RNA toxicity is one of these pathological effects, which is caused by the accumulation of the HRE transcript [$r(G_4C_2)_n$] and its interaction with

RNA-binding proteins (RBPs); the protein binding to $r(G_4C_2)_n$ will be sequestered and lose its function (Prudencio et al., 2015). Among more than a hundred RBPs found to interact with $r(G_4C_2)_n$, we have focused on hnRNP H1 and its interaction with RNA. As a potential splicing factor, hnRNP H1 may possess a helicase activity to bind to the target RNA and change its conformation. In this regard, optical tweezers provide a unique and convenient way to explore RNA-protein interactions and structural dynamics at the single-molecule level.

CHAPTER 2 EXPERIMENTAL PROCEDURES

2.1 RNA preparation

2.1.1 Plasmids

Plasmids are prepared by ligation of different inserts with vectors pVE60hp and pT7SP6vec for both tweezer and bulk experiments, respectively (Wen et al., 2008; 許鈺婕, 2023). The following table shows oligos used for plasmid constructs.

Name	Sequences (5' to 3')
G2C4-4x-F	TATGTAGGCCCCGGCCCCGGCCCCGGCCCCAT
G2C4-4x-R	GTACATGGGGCCGGGGCCGGGGCCGGGGCCTACA
G4C2-4x-F	TATGTAGGGGCCGGGGCCGGGGCCGGGGCCAT
G4C2-4x-R	GTACATGGCCCCGGCCCCGGCCCCGGCCCCTACA
GCswap-F	TATGTAGGGGCGGGCGCCGGGGCGGGCGCCAT
GCswap-R	GTACATGGCGCCCGCCCCGGCGCCCGCCCCTACA
T7SP6-4x-F	AATTCGGGGCCGGGGCCGGGGCCGGGGCCT
T7SP6-4x-R	GTACAGGCCCCGGCCCCGGCCCCGGCCCCG
GCswap-NH-F	AATTCGGGGCGGGCGCCGGGGCGGGCGCCT
GCswap-NH-R	GTACAGGCGCCCGCCCCGGCGCCCGCCCCG

Bacterial strain

XL-10-Gold® Ultracompetent Cells, Agilent (Stratagene, Cat. 200315): Tet^{r}Δ*(mcrA)183*Δ (*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB* lacI^{q}ZΔ*M15 Tn10* (Tet^{r}) Amy Cam^{r}]

Annealing of insert

By T4 PNK (NEB, #M0201S), the 5' end of DNA oligos are phosphorylated.

Reagent	Volume (μL)	Final conc.
10X T4 ligase buffer	5	1x
Forward oligomers	1	2 μM
Reverse oligomers	1	2 μM
T4 PNK (10 U/μL)	1	-
Total volume	Fill ddH_2O to 50 μL	

Cycle: 37℃, 30 min >> 80℃, 5 min>> 20℃ (slowly in an hour)

Ligation and transform

Make the following mixture and incubate at 25℃ for 5 min.

Reagent	Volume (μL)	Final conc.
2X Ligation Mix (Takara, #6023)	5	1x
Insert	1	8 nM
Vector	1	2.8 nM
Total volume	Fill ddH_2O to 10 μL	

The reaction (1.5 μL) was transformed to XL-10 competent cells. The transformed cells were streaked on a Lysogeny broth (LB) plate with 100 μg /mL of ampicillin and then cultured at 37℃ for 16 h. One single colony was selected and inoculated in 5 mL of TB broth with 100 μg/mL of ampicillin, which was incubated at 37℃, 250 rpm for 16 h. Vector (without the insert) was used as a negative control for transformation and bacterial culture.

Plasmid	Vector
p(G4C2)4-tweezers*	pVE60hp
P(G2C4)4-tweezers	pVE60hp
pGCswap-tweezers*	pVE60hp
pT7SP6-4x**	pT7SP6vec
pT7SP6-GCswap	pT7SP6vec

All plasmids are stored at -20℃ plasmid bank.

*Plasmids constructed by 楊世綸, **Plasmid constructed by 王渝暄

2.1.2 *In vitro* transcription

Presto™ Mini Plasmid Kit (Geneaid, PDH100) was used to isolate and purify plasmid from XL-10 cells. Plasmids were then digested by enzyme BssSI-v2 (NEB, #R0680S) (for optical tweezers experiments) or BsrGI-HF (NEB, #R3575S) (for smFRET experiments) at 37℃ for 4 h, and then purified by PCR Advanced PCR Clean Up System kit. Linearized plasmid was *in vitro* transcribed by HiScribe™ T7 Quick High Yield RNA Synthesis Kit (for optical tweezers experiments) or HiScribe™ SP6 RNA Synthesis Kit (for smFRET experiments) at 37℃ for 4 h. The RNA was then purified by MEGAclear™ Transcription Clean-Up Kit and stored at -20℃.

2.2 RNA preparation for optical tweezers experiments

2.2.1 Primers for handle

Name	Sequences (5’ to 3’)
5’handle-F	CTGAGGCTTAAGCGAGGGCGT
5’handle-R	CATATGTATATCTCCTTCTTAAAGTTAAACAAA
3’handle-F	GTACAGAACGCAATGCGTCTGGGCGCT

3'handle-R	Biotin-CACGAGGGAGCTTCCCAGGGGGAAACG

2.2.2 DNA handle preparation

PCR for handle

Handles were amplified by PCR with primers designed for the handle, the mixture for PCR reaction and setting of thermo cycler is listed below:

Reagent	Volume (μL)	Final conc.
10X Taq Buffer	50	1x
10nM dNTP	12.5	0.25 mM
5ng/μL template	5	5ng/100μL
100 μM forward primer	5	1 μM
100 μM forward primer	5	1 μM
2.5 U/μL Taq	5	2.5 U/100μL
Total volume	Fill ddH_2O to 500 μL	

Distribute the PCR mix to 7 PCR tubes for 70μL per tube.

Thermo cycle setting for 5' handle: 94℃ (30 s) → [94℃ (30 s) → 63℃ (30 s) → 68℃ (90 s)] repeated for 30 cycles → 68℃ (7 min) → 10℃ (hold).

Thermo cycle setting for 3' handle: 94℃ (30 s) → [94℃ (30 s) → 59℃ (30 s) →68℃ (90 s)] repeat for 30 cycles → 68℃ (7 min) → 10℃ (hold).

Purification of PCR product

The PCR products were purified by column and ethanol precipitation. Column purification using QIAquick PCR Purification Kit is carried out first. Before column purification, pool all the reactions in one tube and follow the protocols in the kit.

For ethanol precipitation, add 1/10 volume of 3 M NaOAc and 2.5 volume of 100%

EtOH, mix well by gently inverting the tube ~ 10 times. Incubate in -20°C freezer for > 1 h. After incubation, centrifuge at 13,000 g, 4°C for 20 min and carefully remove the supernatant. To wash the pellet, add 1mL of 70% EtOH, and centrifuge at 4°C for 5 min. Remove the supernatant carefully and air dry the pellet for 5-10 min. Add 30μL of ddH_2O to resuspend the pellet, measure UV with Nanodrop and stock at -20℃.

DIG labeling for 5'handle

Digoxigenin (DIG) was labeled to the 3' end of the 5' handle by an exchange reaction. Prepare a 25 μL reaction mixture in a PCR tube on ice:

Reagent	Volume (μL)	Final conc.
NEBuffer 2.1	2.5	1x
5' handle	1.5	1 μM
DIG-11-dUTP (1 mM)	2.5	0.1 mM
T4 DNA polymerase (3 U/μl)	1.7	0.2 U/μl
Total volume	Fill ddH_2O to 25 μL	

Incubate at 25℃ for 15 min. The final product was purified by PCR Advanced PCR Clean Up System kit and eluted with 30 μl elution buffer. Stored at -20℃.

2.2.3 Annealing of DNA handles and RNA

Annealing buffer

Reagent	Final conc.
Formamide	86.80%
EDTA (pH8.0)	1 mM
PIPES-KOH (pH 6.3)	43 mM
NaCl	0.43 M

The purified 5’ DIG-handle, 3’ handle and RNA were mixed and annealed in the following

thermos cycle setting:

Reagent	Final amount
RNA	4 pmol
3’handle	4 pmol
DIG-5’handle	2 pmol
Annealing buffer	80%

Incubation: 85℃ (1 min) → 62℃ (90 min) → 52℃ (90 min) → slowly cool down to 10℃ in 10 min → 10℃ (hold).

The annealed product was then diluted to indicated concentrations using 80% of annealing buffer as dilution buffer and stock at -20℃.

2.3 hnRNP H1 purification

2.3.1 Modification of hnRNP H1

Previously, our lab tried to extract full length hnRNP H1, however, the extracted protein aggregates and was not able to utilize in further experiments. To solve this problem, we have decided to present a truncation to hnRNP H1 and cut the two LC domains and RRM3, which is the C terminus of the protein. According to past research, LC1 mainly functions in liquid-liquid phase separation (LLPS) of the protein. Through LLPS, hnRNP H1 can interact with other proteins, and may cause aggregation of the protein. Past research has shown that there was no aggregation of protein after C terminus was cleaved (Kim & Kwon, 2021).

2.3.2 Plasmid of truncated mutant

The truncated mutant will be constructed by applying PCR to plasmid containing wild type sequence. By designing primers at both ends of N terminus (RRM1 and RRM2), new insert is made. To express protein, insert will be ligated to vector as same as plasmid (with His-tag), of wild type and further transformed to bacterial strands for protein expression. The truncated mutant will be named hnRNP H1-N1 in further description.

PCR for insert of hnRNP H1-N1

Plasmid used for mutant was constructed by 邱仕捷 in 2022 and stocked at -80℃. Primers for PCR reaction was designed using NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi).

Name	Sequences (5' to 3')
H1-N1-For	CAGGGATCCATGATGTTGGGCACGGAA
H1-N1-Rev	GGTGCTCGAGTTAAACTTCAGCTCTACTGCTCTTAAA

Q5 High-Fidelity DNA polymerase (NEB, #M0491S) was used for PCR of hnRNP H1-N1 insert for proofreading. The reagent of PCR reaction was added as the following chart:

Reagent	Volume (μL)	Final conc.
5X Q5 reaction buffer	10	1x
10nM dNTP	1	0.2 mM
16ng/μL template	1	16ng/50μL
100 μM forward primer	2.5	0.5 μM
100 μM forward primer	2.5	0.5 μM

Q5 polymerase	0.5
Total volume	Fill ddH_2O to 50 μL

Setting of thermo cycle: 94℃ (30 s) → [98℃ (10 s) → 63℃ (30 s) → 72℃ (30 s)] repeated for 30 cycles → 72℃ (2 min) → 4℃ (hold).

PCR products were purified by PCR Advanced PCR Clean Up System kit. Insert was further digested with enzyme BamHI-HF (NEB, #R3136S) and XhoI (NEB, #R0146S) at 37℃ for 4 h.

Bacterial culture

XL-10 will also be used in construct of hnRNP H1-N1 plasmid, the ligation step was same as mentioned in **2.1.1.** The plasmid containing insert sequences was further transformed into XL-10 competent cells. The transformed cells were streaked on a LB plate with 50 μg/mL of kanamycin and then cultured at 37℃ for 16 h. After overnight incubation, one single colony was selected and inoculated in 5 mL of LB broth with 50 μg /mL of kanamycin, which was incubated at 37℃, 225 rpm for 16 h. Vector will be used as negative control for transformation and bacterial culture. Presto™ Mini Plasmid Kit was used to isolate and purify plasmid from XL-10 cells for sequencing and further usage.

2.3.3 Protein purification

Buffer for protein extraction

All the buffers were degassed and β-Me (final 6 mM) was added before usage.

Buffer A (for binding, wash, and elution)

Reagent	Volume (mL)	Final conc.
1 M Tris-HCl, pH 7.2	2.5	10 mM

1 M imidazole-HCl, pH 7.5	1.5	25 mM
5 M NaCl	2.5	500 mM
Total volume	Fill ddH_2O to 1000 mL	

Buffer B (for elution)

Reagent	Volume (mL)	Final conc.
1 M Tris-HCl, pH 7.2	5	10 mM
1 M imidazole-HCl, pH 7.5	250	500 mM
5 M NaCl	50	500 mM
Total volume	Fill ddH_2O to 500 mL	

proTEV buffer (for proTEV reaction)

Reagent	Volume (mL)	Final conc.
1 M Tris-HCl, pH 7.2	15	50 mM
0.5M EDTA, pH 8.0	0.3	0.5 mM
50% glycerol	90	15%
1 M DTT	0.3	1 mM
Total volume	Fill ddH_2O to 300 mL	

Buffer S (for Superdex 200)

Reagent	Volume (mL)	Final conc.
1 M Tris-HCl, pH 7.2	5	10 mM
5 M NaCl	20	200 mM
Total volume	Fill ddH_2O to 500 mL	

Bacterial strain

BL21-CodonPlus (DE3)-RIPL Competent cells, Agilent (Stratagene, Cat. 230280): F– *ompT hsdS*(rB – mB –) *dcm*+ Tetr gal l(DE3) *endA* Hte [*argU proL* Cam^R] [*argU ileY leuW Step/*$Spec^R$]

Bacterial culture and harvest

Plasmid extract from XL-10 cell mentioned in last section will be transformed into BL21 competent cell for protein expression. BL-21 cells containing hnRNP H1-N1 were streaked on a LB-kanamycin plate (50 μg/mL kanamycin) and incubated at 37℃ for 16 h. Two colonies were picked and inoculated in 5 mL of LB-kanamycin and incubated at 37℃, 250 rpm for 16 h. After 16 h of culture, the bacteria were diluted by LB-kanamycin 1:500 and incubated at 37℃, 250 rpm. When OD_{600} reached 0.5, 0.1 mM IPTG was added to the cultured bacteria and kept incubated for 3 hours. The cultured bacteria were cooled down by swirling the flask in an ice-water bath. The bacteria were harvested by centrifugation at 5500 rpm, 4℃ for 15 min. The pellet was resuspended in buffer A and centrifuged at 5500 rpm, 4℃ for 15 min, the supernatant was removed. Measure the weight of the pellet and wrap the tube with aluminum foil, freeze in liquid nitrogen, and store the cell pellet at -80°C.

Cell lysis

Resuspend the pellet by 10X weight of buffer A and add 100 μl Protease Inhibitor Cocktail (Sigma, P8849-1ML). Bacteria were lysed by French press with 3 passes (4℃, 20 kpsi). The bacteria lysate was obtained by centrifugation at 4℃, 13000 rpm, 30 min for four times. After centrifugation, the supernatant was filtered through 0.22 μm filter, thawed the sample on ice.

Fast protein liquid chromatography (FPLC)

To separate hnRNP H1-N1, supernatant from cell lysis was further purified by FPLC

with 2 columns: Histrap FF Crude and Superdex 200 10/300 GL. Histrap FF Crude was the first column used in this purification, samples were flowed through this column for twice. For the first Histrap, samples were flowed through the column, meanwhile hnRNP H1-N1 (with His-tag) will bind to the column. The column was then washed by buffer A. After binding, protein was eluted by buffer B and collected into different fractions (1.5 mL/tube). The fractions were confirmed by 15% SDS-PAGE to make sure which fractions should be collected. The pooled elution fractions were concentrated and replace the buffer with proTEV buffer by 10K Amicon Ultra-15 device (Merck #UFC901024) by centrifugation at 4000 g, 4℃.

To cut the tag from protein, ProTEV Plus (Promega, # V610A) was added to the sample with 30μL per 1000μL of sample. Incubate at 30°C (water bath) for 3 hours and fill buffer A to 20mL for second Histrap (same procedure as the first time). The fractions were confirmed by 15% SDS-PAGE to make sure which fractions should be collected.

For improved protein quality, Superdex 200 is applied to separate proteins with molecular weight. Fractions collected from the second Histrap were pooled together and concentrated by 10K Amicon Ultra-15 device. The concentrated samples were then injected into FPLC, and buffer S were used for elution, add glycerol to the sample (final conc. 10%) for storage in -80℃. Before storage, liquid nitrogen was used for fast cooling.

Bradford assay

To determine the concentration of purified hnRNP H1-N1, different concentrations of BSAs were used as a standard. Diluted Bio-Rad protein assay dye was added to different dilution factors of sample. After 5 min of incubation, each sample was quantified by Nano Photometer by measuring its OD_{565}.

2.4 Optical tweezers

2.4.1 Experiments set up

Buffers for optical tweezers experiments

All buffers were filtered before usage. Note that DTT was added only when the experiments were done in the presence of protein.

T20

Reagent	Volume (μL)	Final conc.
1 M Tris-HCl, pH 8.0	1000	20 mM
Total volume	Fill ddH_2O to 50 mL	

TK100

Reagent	Volume (μL)	Final conc.
1 M Tris-HCl, pH 8.0	1000	20 mM
2 M KCl	2500	100 mM
1 M DTT	50	1 mM
Total volume	Fill ddH_2O to 50 mL	

TMg10

Reagent	Volume (μL)	Final conc.
1 M Tris-HCl, pH 8.0	1000	20 mM
1 M $MgCl_2$	500	10 mM
1 M DTT	50	1 mM
Total volume	Fill ddH_2O to 50 mL	

Sample preparation

For optical tweezers experiments, two beads were prepared: 2 μl of streptavidin (SA) beads (SPHERO^{TM}, SVP-20-5) was added to 1 mL of T20; 1μl of RNA sample (1 nM), and 2μl of anti-DIG (AD) beads (SPHERO^{TM}, DIGP-20-2) were mix with 17μl of T20 and incubate in 4℃. After 10 min, 980 μl of T20 was added to the mixture of AD beads and RNA.

minTweezers

The single molecule instrument used in this research is miniTweezers, single trap optical tweezers for single molecule experiments. For calibration, ddH_2O was added between objective lens and chamber, thus, buffers for experiments (TK100 or TMg10, depending on the purpose of experiments) were flowed into the chamber to wash each channel. The schematic diagram of the chamber and laser traps were shown in Figure 6 (Wen et al., 2008). Before the experiments, laser traps must be aligned and power for each beam was adjusted to a PSDs sum of nearly 12000. AD and SA beads were injected to both upper and lower channels respectively. Finally, the beads would be trapped to micropipette and laser trap respectively.

For ramping experiments, slowly move the AD bead near the SA bead and pull away until the biotin interacted with streptavidin, then the trace could be seen while the AD bead was being pulled. The program setting for force ramping: force: 2-X pN (depending on the testing subject); pull speed: 100 nm/s; and refolding time: 10s. As for constant force experiments, a sudden drop of force was applied once the unwinding trace was observed while ramping. The force of constant force was determined and adjusted considering the unwinding force. To determine if the protein binding effect the tether itself and caused bias to data, the force oscillation of AD bead was measured after the tether between two beads broke.

2.4.2 Data analysis

OT console written with Matlab was used for analyzing data recorded in miniTweezers experiments, unwinding trace, force, size of unwounded structure would be obtained through analyzing. Thus, the distribution of force and size will be fit with worm-like chain (WLC) model to calibrate the actual distance of the structure (Marantan & Mahadevan, 2018). Factors that determine the size of WLC curve are the offset value (nm), which is dependent on the end-to-end distance of the structure, and the size of structure (nt). The offset value normally used in this research is 2nm, the end-to-end distance of an RNA hairpin.

2.5 EMSA for hnRNP H1-N1

EMSA experiments were performed to determine the interaction between hnRNP H1-N1 and RNA besides single-molecule experiments. A sequence of hnRNP H1-N1 with different concentrations (1X, 2X, 4X…) were added to fixed concentration of RNAs and incubated for 30 minutes in 37°C water bath. Buffers used for EMSA were 10X buffers for optical tweezers experiments written in **2.4.1.** 8% native PAGE was applied for gel electrophoresis: 100mV, 55 min, with 100 mV 20 min pre-running. To be mentioned that while using Mg^{2+} in reaction, EDTA in gel running buffer was removed.

CHAPTER 3 RESULTS

3.1 RNA preparation for optical tweezers experiments

Four-repeat sense strand, $r(G_4C_2)_4$, of *C9orf72* HRE were mainly studied in this research. The RNA was synthesized by 楊世綸. For the control experiments, the antisense $r(G_2C_4)_4$ and GCswap mentioned in **2.1.1** were used. Figure 7 shows the electrophoresis images of RNA and DNA handles.

3.2 Purification of hnRNP H1-N1

To purify hnRNP H1-N1, transformed BL-21 cells were first cultured and lysed through French press. The lysates were first purified by Histrap FF Crude column. hnRNP H1-N1 was eluted through a concentration gradient of imidazole (25-500 mM). The TEV tag of the protein was cleaved, and the sample was passed through the Histrap column again. Elution fractions were verified by 15% SDS-PAGE. The size of hnRNP H1-N1 is approximately 21 KDa which can be seen on the electrophoresis result (Figure 8). Fractions 14~19 eluted from the second Histrap column were collected, concentrated, and dialyzed for the next step of purification.

To further improve the purity, the Superdex 200 column was used to separate proteins per size. Fractions from Superdex 200 were verified by 15% SDS-PAGE (Figure 9) The samples were stored at -80°C for further use.

3.3 Structures of $r(G_4C_2)_4$

The schematic setup of miniTweezers is as mentioned in **2.4.1** (Figure 2). The $r(G_4C_2)_4$ construct was initially bound with an AD bead and moved through laser beams. The AD bead was then moved to the SA bead until the other end of the tether binds to the SA bead. Once transitions are observed, ramping programs will begin, and data will be

recorded. The expected structures formed by $r(G_4C_2)_4$ are GQ and a hairpin with 22 nt (without first two Gs, Figure 2)

For $r(G_4C_2)_4$, two kinds of ions were used for miniTweezers experiments. TK100 and TMg10 as described in **2.4.1**.

3.3.1 $r(G_4C_2)_4$ in K^+

In past research, potassium ion was stated to promote and stabilize GQ structures in both RNA and DNA (Koirala et al., 2011). TK100, a buffer with 100mM K^+, was first applied in this study. The ramping obtained shows in Figure 10, RNAs tend to show a hopping state while the structure unfolds around 10 to 15 pN. From traces collected in experiments, similar patterns were observed. The structure size of $r(G_4C_2)_4$ was indeterminable due to the hopping with poor single-to-noise ratios.

3.3.2 $r(G_4C_2)_4$ in Mg^{2+}

Since experiments weren't successful in the monovalent cation, Mg^{2+} was applied. Although high concentrations of Mg^{2+} seem to destabilize GQ structures, a relatively low concentration of Mg^{2+} was found to promote GQ in single molecule fluorescence resonance energy transfer (FRET) experiments, thus TMg10 with 10mM Mg^{2+} was applied (Balaratnam & Basu, 2015; Lee et al., 2016).

Ramping experiments

Unlike in TK100, several stable transitions that fall around 18-24 pN were observed in TMg10. The transitions can be divided into three species: R, RH and HR (figure 11a). R stands for a clear and single rip; RH is a rip followed by a small hopping whereas HR is hopping happens first and then a rip. By exporting data to an F-X distribution plot, the force and extension of RNA seems to be separated into two groups. By using the extensible WLC model, the clusters fit to an 18 nucleotide (nt) structure and a 20 nt

structure (Figure 11b). In a total of 33 molecules and 944 traces, the structures with 18 nt and 20 nt account for 35.1% (331 traces) and 64.9% (613 traces) respectively. The predicted 22 nt hairpin and GQ did not fit into these two clusters.

In further predictions, these structures may be hairpin using 3 repeats of GGGGCC (18 nt) (Figure 12b). While the 20 nt structure was missing 4 nt from the full length, we tried to remove the first 4 guanines in the sequence and apply the prediction again, the most stable structure predicted by RNAfold was a 20 nt double hairpin (Figure 12a).

The cluster of 18 nt structures unwound between 15-25 pN with an average of 19.9 pN, whereas the cluster of 20 nt structures unwound between 18-26 pN with an average of 21.8 pN. These data show that $r(G_4C_2)_4$ prefers to form in the 20 nt structure with higher population and stability.

Constant force experiments

Constant force experiments were performed under 18-21 pN to measure structure size and folding/unfolding of $r(G_4C_2)_4$. The structure of $r(G_4C_2)_4$ seems to undergo repetitive folding and unfolding while the force is held at 19 or 20 pN. Therefore, the trace obtained was hopping (Figure 13). The traces then went through a Gaussian fitting to calculate the size of structure. Two different structure sizes were obtained through gaussian fitting: 18.25 and 19.49 nt (Figure 13), like the worm-like chain fitting to the FX plot shown in Figure 11b.

3.4 Structures of $r(G_2C_4)_4$ and GCswap

For control experiments, antisense RNA of *C9orf72* HRE was chosen as it can't form GQ and should only form in the predicted 22 nt hairpin structure. For control experiments, only TMg10 was used since the transition of wild type (WT) four repeat GGGGCC was indistinguishable in TK100.

3.4.1 r(G_2C_4)$_4$ in Mg^{2+}

The ramping data obtained in r(G_2C_4)$_4$ were very diverse in structure size and transition pattern (Figure 14). However, in many molecules recorded, the traces tend to disappear after a few cycles; the rest of the molecules form multimolecular structures.

3.4.2 GCswap in Mg^{2+}

As antisense seems to be unsuitable for control experiment, here a new construct of swapping two GC pairs in r(G_4C_2)$_4$ was made. As the GC swapped, this RNA sequence was predicted to form a 22 nt hairpin only (Figure 15a).

Ramping traces

Approximately all of transitions found in ramping data of GCswap in Mg^{2+} were a clear single rip, a few of which were a rip with slightly hopping (Figure 15b). While analyzing F-X plot, only one cluster was found, all the traces were fit to 22 nt worm-like chain (Figure 15c). In total of 12 molecules and 281 traces, the unfolding force was mainly found among 22-26 pN, with several points in 18-22 pN. The average unfolding force of GCswap RNA is 23.96 pN.

Constant force

Constant force experiments were performed under 19-23 pN to measure structure size and folding/unfolding of GCswap. The structure of GCswap seems to repeat folding and unfolding while the force is held at 20-23 pN depending on molecules (Figure 16). The traces obtained were then went through a Gaussian fitting to calculate the size of structure. The structure calculated through Gaussian fitting was 21.59 nt (Figure 16), slightly smaller than the worm-like chain fitting to FX plot.

3.5 Interactions between hnRNP H1-N1 and r(G_4C_2)$_4$

As transitions of WT four repeat *C9orf72* HRE were obtained, hnRNP H1-N1 was applied to miniTweezers experiments, 10nM of protein were diluted from stock using experimental buffer: TMg10. The concentration of protein was tested several times, while the concentration was high, traces observed often belong to multimolecular structures; while the concentration was below 10 nM, the efficiency of binding was not enough to collect sufficient data.

Ramping data

While protein was flowed into chamber of minTweezers, the tether of RNA seems to become unstable and break in relatively high force (around 30 pN). Thus, the high force cutoff in the program of ramping was changed to 27-30 pN depending on each molecule.

Similar to traces without protein, traces recorded in hnRNP H1-N1 experiments were also classified into the following groups: R, RH and HR (Figure 17a). Different from RNA only, some of R and HR seems to be slightly bigger than R recorded in $r(G_4C_2)_4$ without protein. By F-X plot, three groups of force and extensions were clustered (Figure 17b). After fitting to worm like chain model, 18 nt and 20 nt clusters in 3.3.2 were also presented in data with protein. Besides the two groups that were the same as RNA only, an additional group was found and fit with 22 nt worm like chain. In total of 15 molecules and 284 traces, 18 nt cluster accounts for 21.5% (104 traces); 20 nt cluster accounts for 28.5% (138 traces); 22 nt cluster accounts for 50% (242 traces). The average unfolding force in 18 nt cluster and 20 nt cluster were 19.78 pN and 20.98 pN, respectively. In the 22 nt group, the average unfolding force was 20.93 pN. Compared with data from RNA only, the percentage of different groups changed significantly (17c). However, the 20 nt cluster has a smaller p-value, suppose the protein binding may be targeting the 20 pN structure mainly and change its conformation to 22 nt structure.

By overlapping F-X plot in **3.3.2**, the two structures seen in r$(G_4C_2)_4$ without protein were likely the 18 nt cluster and 20 nt cluster here (Figure 18). In addition to that, the 22 nt cluster suggests that the binding of hnRNP H1-N1 may change the conformation of RNA to the predicted 22 nt hairpin mentioned in **3.3** (Figure 2)**.**

Constant force

Constant force experiments were performed under 17-21 pN to measure structure size and folding/ unfolding of r$(G_4C_2)_4$ while hnRNP H1-N1 was presented, also we tried to determine the actual conformational change caused by protein binding. Several traces under 17-21 pN were recorded and fit with gaussian to determine the structure size (Figure 19). The structure size found after fitting was mostly similar to RNA without protein (18.64 nt and 19.82 nt). The binding step of the protein and the 22 nt cluster were undeterminable through the analysis of constant force experiment.

3.6 Interactions between hnRNP H1-N1 and GCswap RNA

To see if the 22 nt structure presented in the presence of hnRNP H1-N1 can also be found in GCswap experiments while protein exists, 10 nM of hnRNP H1-N1 was also applied in GCswap experiments. Results in **3.4** implied that GCswap can only form 22 nt hairpin in ramping experiments and constant force experiments.

Ramping data

Overall, transitions observed were similar to GCswap without hnRNP H1-N1. R in **3.4** was also observed in the presence of protein, however, the percentages of trace with small hopping seem to rise a little (Figure 20a). A total of 15 molecules, 366 traces, the unfolding force of GCswap in protein experiments falls between 19-26 pN, with an average of 23.3 pN (Figure 20b).

Compared with data in **3.4**, the average unfolding force and the result of worm-like chain fitting seems to be mostly identical. By pooling data from **3.4** and here together in F-X plot, the traces with or without protein almost overlapped (Figure 20c).

Constant force

Same as constant force experiments in GCswap without hnRNP H1-N1, force was held at 19-23 pN to measure whether the structure size and folding/ unfolding of GCswap was affected by protein binding. Same with ramping experiments, the gaussian fitting result of molecules under 22 pN in GCswap with hnRNP H1-N1 was 21.12 (Figure 21), a little smaller than the result in **3.4**. This result seems to indicate that hnRNP H1-N1 either can't change the conformation of GCswap or unable to bind the hairpin formed by GCswap.

3.7 EMSA analysis of hnRNP H1 and r$(G_4C_2)_4$-NH

To further explore the binding of hnRNP H1-N1, EMSA was applied using RNA made for smFRET experiments by 王渝暄.

3.7.1 EMSA experiments in Mg^{2+}

EMSA experiments were performed in buffer containing Mg^{2+} as single-molecule experiments have spotted binding traces. RNA used in the experiments was r$(G_4C_2)_4$ without handles, the final concentration for RNA was 1μM. RNA was incubated with different concentrations of hnRNP H1-N1 in 37℃ for 30 minutes and run in 8% native PAGE. Protein concentration used in experiments were 1×, 2×, 4×, 8×, 12× and 16×. RNA was applied to check the pattern of RNA in absence of hnRNP H1-N1.

The result of the EMSA shown three main bands for r$(G_4C_2)_4$-NH, which were supposed to be multimer, dimer and monomer from upper to lower according to past research (Geng et al., 2024) (Figure 22a). By analysis of the result, it seems like hnRNP

H1-N1 binds to dimer band the most. The band of dimer declines while a small band appears near multimer band as the protein concentration rises. The multimer band remained unchanged and the monomer was likely to shift into a smear. Thus, the dimer band was chosen for the region of interest (ROI, red frame in Figure 22a) for analyzing with Image J. The result of Image J analysis of intensity stated that the RNA started to shift when protein concentration was 4 μM (64% remain) and declined to around 30% when protein concentration reached 12 μM.

3.7.2 EMSA experiments in K^+

EMSA experiments were performed in buffer containing K^+ to see if the pattern has changed since monovalent cations was said to enhance GQ (Koirala et al., 2011). For TK100, the concentration of the protein has changed because the binding of protein was undeterminable in 1 μM of hnRNP H1-N1. The result was mostly the same as result in TMg10. The only difference was the binding percentage of protein. The ROI intensity declined to 25% as the protein concentration reached 8 μM and remained at around 20% in 16 μM and 32 μM. These results seem to indicate that GQ since the lane of RNA only as the data in both ions were nearly the same.

CHAPTER 4 DISCUSSION

4.1 Do noises in trace affect the tether of RNA structure?

While the protein was applied to the single molecule system, noise in ramping traces were observed (Figure 17a). As a structure slightly larger than what we found in experiments of $r(G_4C_2)_4$ only was shown in **3.5**, this noise was suspected to cause bias and result in this 22 nt structure. To make sure that the noise didn't cause any bias, we recorded the force oscillation after the tether of RNA was broken to monitor if the bead itself was affected by protein sticking to the surface. By doing Gaussian fitting to the force recorded and compare the standard deviation of those with protein and without protein.

Through several comparisons (Figure 23), the standard deviation seems to be nearly the same: 0.3723 (no protein) and 0.3759 (with protein) for the samples shown. Through this measurement, we could exclude the effect of protein to bead and cause bias to data analysis, since the difference in the force while there is no tether was so small. A possible cause of the noise might be the binding protein to the DNA handle of RNA samples. Since the handles were both long compared to RNA constructs used, once protein stuck on the handle, there might be a small but continuous noise to the tether.

4.2 Structure preference of hnRNP H1-N1

As stated in **1.4**, hnRNP H1 should be able to bind G-rich sequence (G-tract) with the quasi-RRMs in protein (Dominguez et al., 2010). Though RRM3 was cleaved for hnRNP H1-N1 in this research, the main RRMs that interact with G-tract were supposed to be RRM1 and RRM2 (Conlon et al., 2016). Thus, binding patterns in experiments of both $r(G_4C_2)_4$ and GCswap should be observed since both constructs satisfied the "G rich"

criteria of hnRNP H1 to binds. However, the binding was only observed in $r(G_4C_2)_4$ in single molecule experiments. This indicated that the protein might be unable to bind to the hairpin formed by GCswap RNA.

Interestingly, the EMSA result of $r(G_4C_2)_4$ indicated that among structures formed by $r(G_4C_2)_4$, the binding of hnRNP H1-N1 still has some preference (Figure 22). Three main bands were observed in the naïve PAGE of $r(G_4C_2)_4$, however, only one of them seems to shift while the protein concentration rises. The upper band remains unchanged, the lower one seems to degrade to smear, only the middle band tends to shift with the concentration change, the phenomenon was same in both TMg10 and TK100. The binding affinity of hnRNP H1-N1 was higher in K^+ compared to same concentration under Mg^{2+}, supposed that hnRNP H1-N1 tends to bind GQs, as K^+ was found to stabilize GQs in the past research (Balaratnam & Basu, 2015).

The statistical test result of $r(G_4C_2)_4$ structure clusters changes in the presence of hnRNP H1-N1 shown that 20 nt double hairpin seems to decline the most after the protein was applied into the system. The p-value of the 18 nt and 20 nt group from the chi square were 1.76×10^{-7} and 1.53×10^{-38}, respectively (Figure 17c). From the statistical tests, we can state that the decline of 20 nt group was more than 18 nt group, which further implies that the conformation change caused by hnRNP H1-N1 mainly target the 20 nt double hairpin.

Interestingly, the unfolding trace of double hairpin usually appears as two rips, but the transition observed in our experiments was a single rip. According to past research, there might be kissing interaction between two loops of the double hairpin, which leads to the combination of two transitions into one (Li et al., 2006). The kissing effect may also be the reason why hnRNP H1-N1 target this double hairpin.

4.3 EMSA analysis state that hnRNP H1-N1 may exhibit non-specific binding

To further explore the binding preference of hnRNP H1-N1, GCswap without handle sequence was made for EMSA analysis. However, a mistake in the transcription step left DNA template left in final product. Thus, bands which were likely DNA fragments were detected in EMSA result (Figure 24). The bottom bands in EMSA analysis of GCswap were also found to shift toward smear as the bottom bands in both condition of EMSA analysis for $r(G_4C_2)_4$-NH. This phenomenon may be caused by non-specific binding of hnRNP H1-N1. Protein may exhibit non-specific binding depending on pi value, the relationship between pi value and reaction buffer decides the charge of protein, while the pi value was relatively high, the protein will be positively charged and interact with negatively charged RNA in a charge mediated manner (da Silva et al., 2018). By online tool: ExPASy Compute pI/Mw Tool (https://web.expasy.org/cgi-bin/compute_pi/pi_tool.cgi), we have acknowledged the pI value of hnRNP H1-N1 was 5.89. This indicated that the protein was negatively charged. Hence, there might be other reasons for the non-specific interaction between hnRNP H1-N1 and RNA. Other possible reasons might be the concentration of protein and the polarity of protein.

4.4 Helicase activity of hnRNP H1-N1

As mentioned in **1.4**, hnRNP H1 may possess the activity of helicase. Usually, the binding step of a helicase could be visualized through optical tweezers (Spies, 2014). Unfortunately, the binding of hnRNP H1-N1 seems to happen in a sudden and disassociated away after changing the conformation of RNA. In this case, analyzing the

mean unfolding force of structures in different conditions can help define the helicase activity of hnRNP H1.

After the t-test, the calculated p-value for the two clusters (18 nt and 20 nt) were 0.188 and 3.05×10^{-8}, respectively (Figure 25). The result of the t-test indicated that only the decline of unfolding force in 20 nt structure was significant (Figure 25b), echoing to the inference in **4.2**. To conclude, the binding of hnRNP H1-N1 destabilizes the 20 nt double hairpin, and this might be the evidence that hnRNP H1 possess the activity of helicase.

4.5 Possible candidates for the additional group found in experiments of $r(G_4C_2)_4$ in the presence of hnRNP H1-N1

4.5.1 22 nt hairpin

GCswap was used for control and the result indicated that it only forms one 22 nt hairpin, however, the F-X plot of GCswap did not perfectly match the additional group appears in presence of hnRNP H1-N1 (Figure 26). While analyzing the F-X plot and processing the fitting of WLC model, the new structure appears in the presence of protein cannot perfectly fit to 22 nt WLC curve (Figure 17b). Despite the fitting result, the overlapping of two clusters in Figure 26 and the indicates that these two clusters were not really that different, the reason why WLC curve did perfectly fit to the cluster may be the noise in experiments. On the other hand, considering the unfolding force of two structure, 22 nt hairpin predicted to form by $r(G_4C_2)_4$ was more stable than GCswap; initial ΔG = -13.20 kcal/mol ($r(G_4C_2)_4$) comparing to Initial ΔG = -11.40 kcal/mol (GCswap). By contrast to the result where GCswap seems to be relatively stable according to unfolding force.

4.5.2 G-quadruplex

Since the fitting result suggested that the cluster found in experiments of hnRNP H1-N1 might not be the 22 nt hairpin predicted. As the considering the fitting of WLC model requires an offset value depending on the end-to-end distance of the structure, we tried to find out if there's any suitable value. Also, though GQs weren't observed in single-molecule experiments without protein, we thought it still got a chance to form in protein experiments as hnRNP H1-N1 have a chance to unwind the structure and cause rearrangement to fold into GQ since the fitting result indicates that the structure was slightly bigger than a 22 nt hairpin.

By consulting past research, offset value and structure size used for WLC fitting for RNA GQ was 1.0nm and 22 nt (parallel), calculated from 3D structure on Protein Data Bank (PDB, accession number 8X0S) (Geng et al., 2024). This value was calculated considering the distance of first G to last G. After applying this data to the fitting, we found that the curve was totally apart from the cluster (Figure 27). To mention that the structure for the reference was under the condition of K^+, the offset value was not the actual value of a GQ in Mg^{2+}. Moreover, to check the offset value we obtained by measuring on PDB, other hairpins on PDB were also measured (accession number 1HS2, 1JWC), none of which matched the offset value that were usually used in optical tweezer experiments when structures were hairpin. The offset value used in regular fitting were derived from the Watson–Crick base pairing with X-ray structure, measuring conducted on PDB were all smaller, suppose the actual value should be larger than what we found through PDB. Thus, there were still chances of the existence of GQ.

However, past research conducted on Ewing sarcomas has stated that the binding of hnRNP H1 to *EWSR1*-exon 8 G-rich sequences destabilizes GQs, moreover, qRRM1 and

2 were found to disturb the stability of G quadruplex (Vo et al., 2022). The fitting result and finding in another disease made the chance of this new structure to be GQ low.

4.6 Future perspectives

In this study, several targets were found, we found evidence of hnRNP H1-N1 as a helicase and the binding preference of hnRNP H1-N1. Also, a structure that has a chance to be GQ monomer was found. For future perspective, how to verify these findings and further study into the relationship between hnRNP H1 and ALS/FTD should be the main topic.

First, to determine the helicase activity of hnRNP H1, suppose an RNA with longer repeat number can provide a trace that is more significant while protein binds. As the RNA foci in the neuron cell of ALS patients are also repeat number dependent, increase the number of GGGGCC repeat may also help in finding the relationship of protein and ALS/FTD (Lee et al., 2013). Since high repeat number may be hard to manage in optical tweezer and 10 repeats of GGGGCC was found to form aggregate, increase the repeat number to 10 seems reasonable (Raguseo et al., 2023).

To determine if the additional group of structure in protein experiments, several controls might be useful. TERRA has been discovered to form RNA GQs in past optical experiments, and the G-tract in TERRA may also allow hnRNP H1 to bind, it seems to be a suitable control for future experiments (Yangyuoru et al., 2013). In addition to TERRA, modifying the sequence to $r(G_4T_2)_n$ may also help, as past research had proved it was capable of forming unimolecular GQs by NMM staining and gel electrophoresis (Raguseo et al., 2023).

References

Ashkin, A. (1970). Acceleration and trapping of particles by radiation pressure. *Physical review letters*, *24*(4), 156.

Ashkin, A., Dziedzic, J. M., & Yamane, T. (1987). Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, *330*(6150), 769-771.

Balaratnam, S., & Basu, S. (2015). Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA g-quadruplexes. *Biopolymers*, *103*(7), 376-386.

Conlon, E. G., Lu, L., Sharma, A., Yamazaki, T., Tang, T., Shneider, N. A., & Manley, J. L. (2016). The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *elife*, *5*, e17820.

da Silva, F. L. B., Derreumaux, P., & Pasquali, S. (2018). Protein-RNA complexation driven by the charge regulation mechanism. *Biochemical and biophysical research communications*, *498*(2), 264-273.

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., & Adamson, J. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, *72*(2), 245-256.

Dominguez, C., Fisette, J.-F., Chabot, B., & Allain, F. H. (2010). Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nature structural & molecular biology*, *17*(7), 853-861.

Dong, D., Zhang, Z., Li, Y., Latallo, M. J., Wang, S., Nelson, B., Wu, R., Krishnan, G.,

Gao, F.-B., & Wu, B. (2024). Poly-GR repeats associated with ALS/FTD gene C9ORF72 impair translation elongation and induce a ribotoxic stress response in neurons. *Science signaling*, *17*(848), eadl1030.

Gautrey, H., Jackson, C., Dittrich, A.-L., Browell, D., Lennard, T., & Tyson-Capper, A. (2015). SRSF3 and hnRNP H1 regulate a splicing hotspot of HER2 in breast cancer cells. *RNA biology*, *12*(10), 1139-1151.

Geng, Y., Liu, C., Xu, N., Suen, M. C., Miao, H., Xie, Y., Zhang, B., Chen, X., Song, Y., & Wang, Z. (2024). Crystal structure of a tetrameric RNA G-quadruplex formed by hexanucleotide repeat expansions of C9orf72 in ALS/FTD. *Nucleic Acids Research*, gkae473.

Geuens, T., Bouhy, D., & Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. *Human genetics*, *135*, 851-867.

Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E. M., Logroscino, G., Robberecht, W., Shaw, P. J., Simmons, Z., & Van Den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *Nature reviews Disease primers*, *3*(1), 1-19.

Jain, A., & Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature*, *546*(7657), 243-247.

Jiang, J., Zhu, Q., Gendron, T. F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J. E., Jafar-Nejad, P., Drenner, K., & Schulte, D. (2016). Gain of toxicity from ALS/FTD-linked repeat expansions in C9ORF72 is alleviated by antisense oligonucleotides targeting GGGGCC-containing RNAs. *Neuron*, *90*(3), 535-550.

Kim, G. H., & Kwon, I. (2021). Distinct roles of hnRNPH1 low-complexity domains in splicing and transcription. *Proceedings of the National Academy of Sciences*,

118(50), e2109668118.

Koirala, D., Dhakal, S., Ashbridge, B., Sannohe, Y., Rodriguez, R., Sugiyama, H., Balasubramanian, S., & Mao, H. (2011). A single-molecule platform for investigation of interactions between G-quadruplexes and small-molecule ligands. *Nature chemistry*, *3*(10), 782-787.

Kumar, V., Islam, A., Ahmad, F., & Hassan, M. I. (2016). Structural insight into C9orf72 hexanucleotide repeat expansions: Towards new therapeutic targets in FTD-ALS. *Neurochemistry International*, *100*, 11-20.

Kutluay, S. B., Emery, A., Penumutchu, S. R., Townsend, D., Tenneti, K., Madison, M. K., Stukenbroeker, A. M., Powell, C., Jannain, D., & Tolbert, B. S. (2019). Genome-wide analysis of heterogeneous nuclear ribonucleoprotein (hnRNP) binding to HIV-1 RNA reveals a key role for hnRNP H1 in alternative viral mRNA splicing. *Journal of virology*, *93*(21), 10.1128/jvi. 01048-01019.

Lee, I.-R., Hsu, H.-Y., & Wu, J.-Y. (2016). The Effect Magnesium Cations to the Formation of G-Quadruplex Studied by Single-Molecule Spectroscopy. *Biophysical Journal*, *110*(3), 635a-636a.

Lee, Y.-B., Chen, H.-J., Peres, J. N., Gomez-Deza, J., Attig, J., Štalekar, M., Troakes, C., Nishimura, A. L., Scotter, E. L., & Vance, C. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell reports*, *5*(5), 1178-1186.

Li, P. T., Bustamante, C., & Tinoco Jr, I. (2006). Unusual mechanical stability of a minimal RNA kissing complex. *Proceedings of the National Academy of Sciences*, *103*(43), 15847-15852.

Marantan, A., & Mahadevan, L. (2018). Mechanics and statistics of the worm-like

chain. *American Journal of Physics*, *86*(2), 86-94.

Masrori, P., & Van Damme, P. (2020). Amyotrophic lateral sclerosis: a clinical review. *European journal of neurology*, *27*(10), 1918-1929.

Moffitt, J. R., Chemla, Y. R., Smith, S. B., & Bustamante, C. (2008). Recent advances in optical tweezers. *Annu. Rev. Biochem.*, *77*(1), 205-228.

Prudencio, M., Belzil, V. V., Batra, R., Ross, C. A., Gendron, T. F., Pregent, L. J., Murray, M. E., Overstreet, K. K., Piazza-Johnston, A. E., & Desaro, P. (2015). Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nature neuroscience*, *18*(8), 1175-1182.

Raguseo, F., Wang, Y., Li, J., Petrić Howe, M., Balendra, R., Huyghebaert, A., Vadukul, D. M., Tanase, D. A., Maher, T. E., & Malouf, L. (2023). The ALS/FTD-related C9orf72 hexanucleotide repeat expansion forms RNA condensates through multimolecular G-quadruplexes. *Nature Communications*, *14*(1), 8272.

Spies, M. (2014). Two steps forward, one step back: determining XPD helicase mechanism by single-molecule fluorescence and high-resolution optical tweezers. *DNA repair*, *20*, 58-70.

Vo, T., Brownmiller, T., Hall, K., Jones, T. L., Choudhari, S., Grammatikakis, I., Ludwig, K. R., & Caplen, N. J. (2022). HNRNPH1 destabilizes the G-quadruplex structures formed by G-rich RNA sequences that regulate the alternative splicing of an oncogenic fusion transcript. *Nucleic Acids Research*, *50*(11), 6474-6496.

Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S. H., Noller, H. F., Bustamante, C., & Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. *Nature*, *452*(7187), 598-603.

Yangyuoru, P. M., Zhang, A. Y., Shi, Z., Koirala, D., Balasubramanian, S., & Mao, H. (2013). Mechanochemical properties of individual human telomeric RNA (TERRA) G-quadruplexes. *ChemBioChem*, *14*(15), 1931-1935.

許鈺婕. (2023). *以單分子螢光共振能量轉移觀測 mRNA 的長度對反義寡核苷酸以及 30S 核醣體次單元的交互作用影響* (國立臺灣大學碩士學位論文).

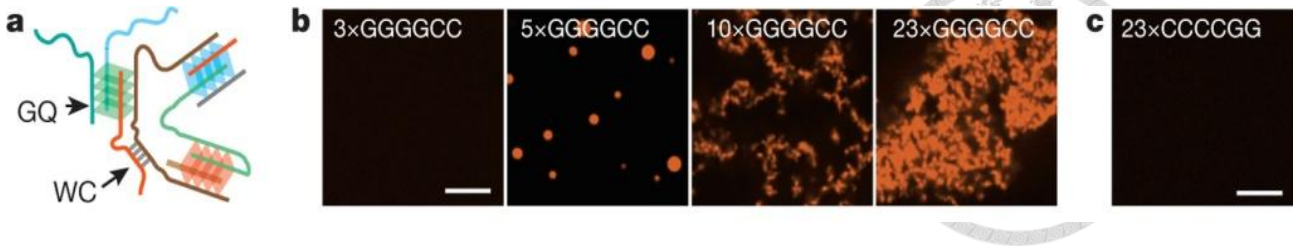

Figure 1. RNA GGGGCC repeats form foci

RNA G_4C_2 tandem repeats may form foci, depending on the number of repeats. Foci can be detected from 5× repeats or longer. By contrast, 23× of CCCCGG cannot form any detectable foci. Adapted from Jain & Vale, 2017, permission acquired from Springer Nature.

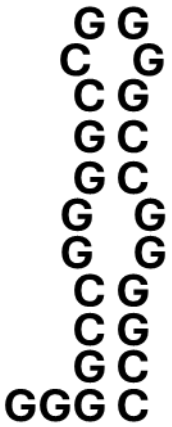

Figure 2. Hairpin formed by r(G4C2)4

22 nt hairpin formed by r(G4C2)4, predicted using: https://www.unafold.org/

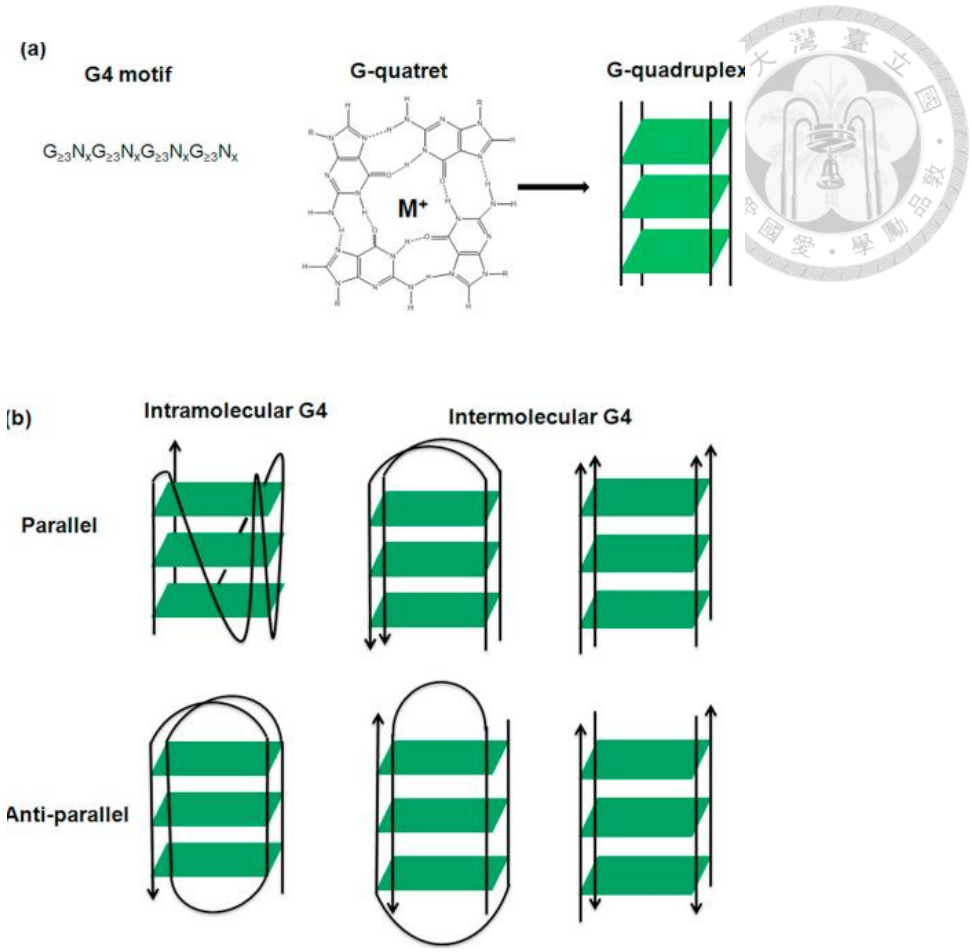

Figure 3. G quadruplex

(a) G-quartet formed by 4 Gs, G-quartet can stack together and form G quadruplex. (b) Different GQs. Adapted from Kumar et al., 2016, permission acquired from Springer Nature.

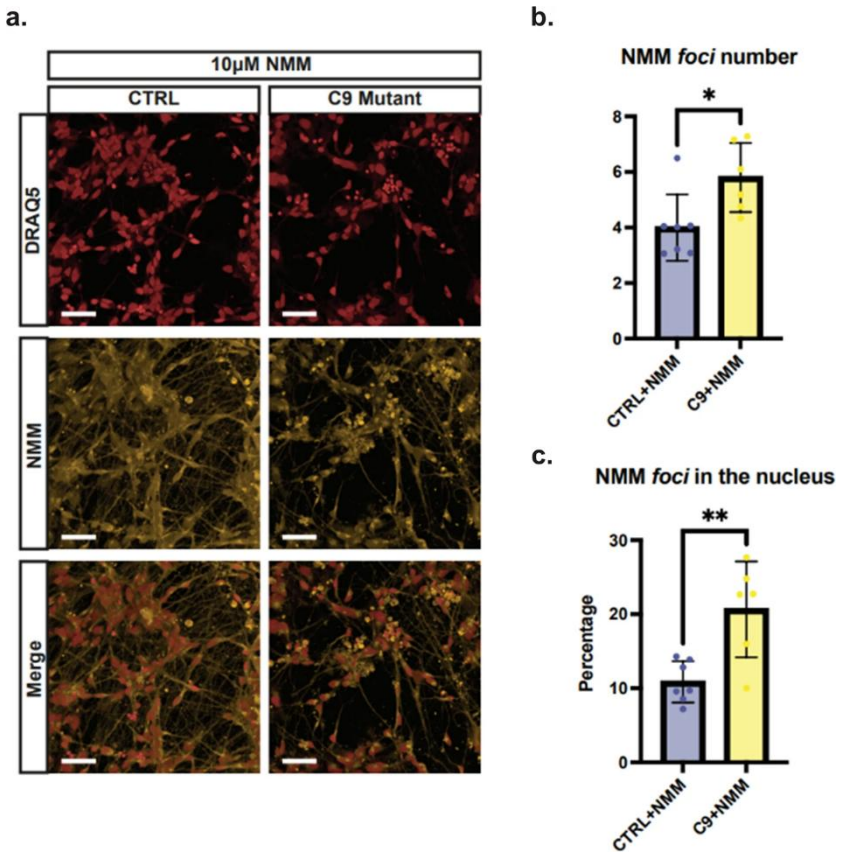

Figure 4. Foci of RNA G-quadruplex detected in iPSC

From iPSC generated using C9ALS patient's neuron cells, RNA foci were detected and NMM (GQ specific dye) was used to detect GQ. The result indicated that RNA foci in the neuron cell of a ALS patients were composed of GQs. (a) Image of iPSC stained with different dyes. Statistics of NMM foci in the cell (b) and in the nucleus (c). Adapted from Raguseo et al., 2023, open access.

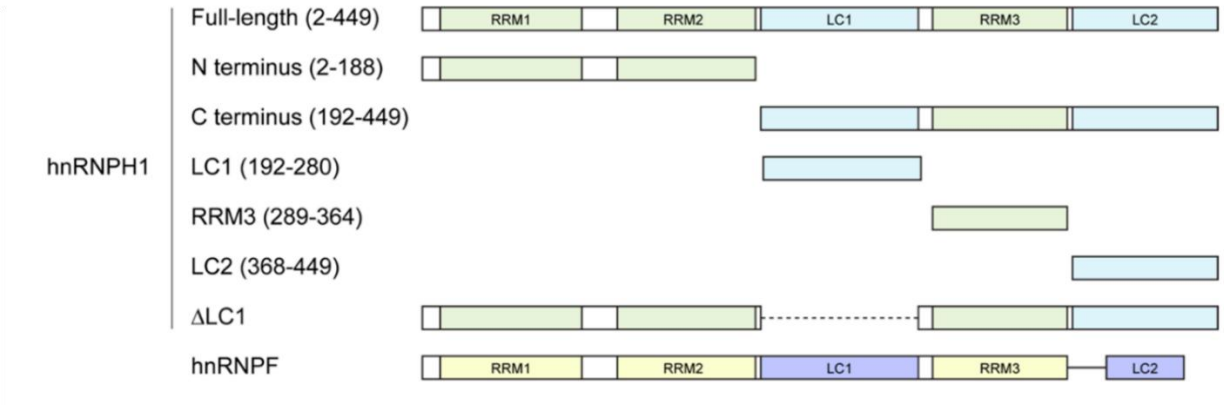

Figure 5. Domains of hnRNP H1

Different domains are colored and labeled. The white box between domains are spacer regions. Adapted from Conlon et al., 2016, open access.

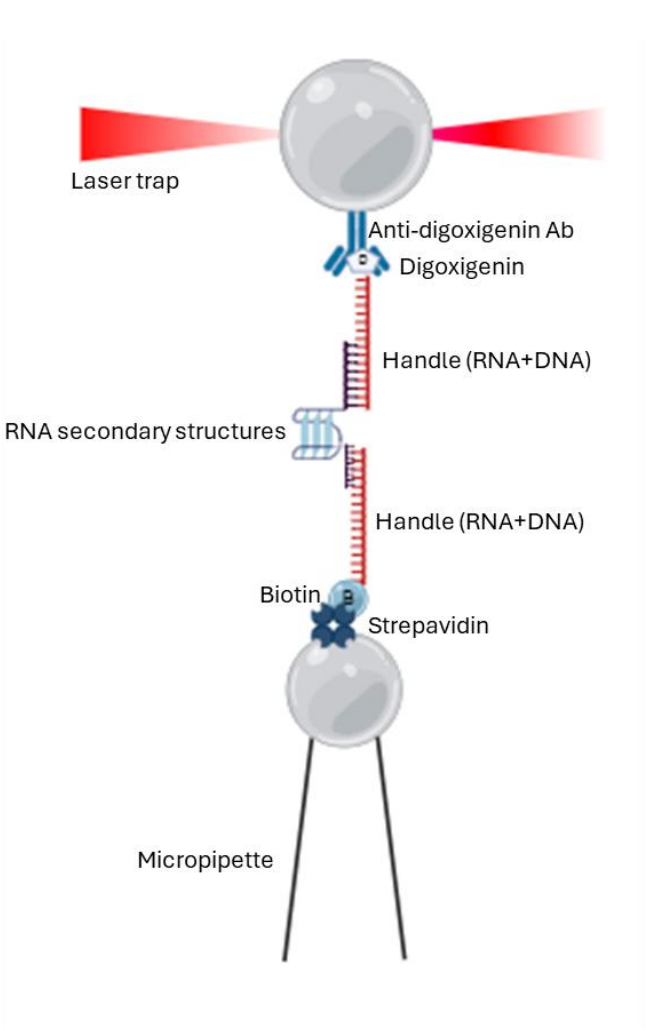

Figure 6. Schematic diagram of the miniTweezers setup

MiniTweezers, the instrument of optical tweezers used in this research, contain a pair of 845 nm lasers that form a laser trap in the chamber. This trap allows us to control a polystyrene bead that is coated with anti-digoxigenin antibody. The RNA sample containing the 5' and 3' flanking regions are annealed with complementary DNA sequences (termed handles), which are labeled with a digoxigenin and a biotin tags on each end, respectively. The digoxigenin end could bind with the anti-digoxigenin antibody-coated bead. On the other hand, the biotin end could bind with another bead that is coated with streptavidin. The bead with streptavidin is held at the tip of a micropipette by negative pressure (Wen et al., 2008)

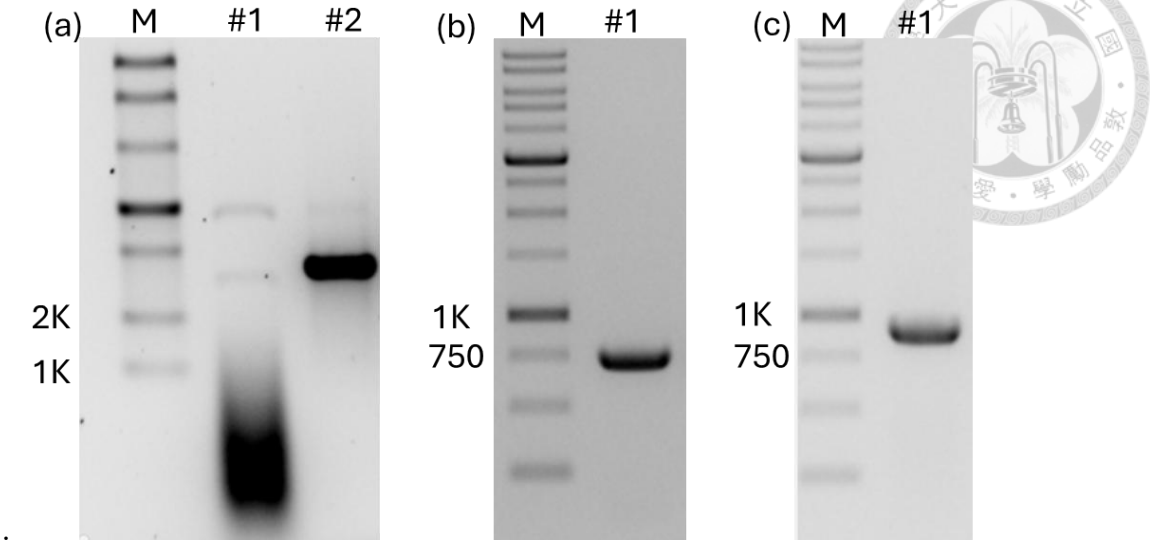

Figure 7. Gel electrophoresis of r(G_2C_4)$_4$ and handles

(a) *In vitro* transcription of r(G_2C_4)$_4$. The length of RNA was 1720 nt. Samples were run in 1% agarose gel stained with SYBR green dye. #1: without purification, the lower smear may be DNA templates. #2: Elution of of the transcription product. Data from 2023/03/16. (b)&(c) Purified 5' (732bp) and 3' (942bp) handle. Samples were run in 1% agarose gel. Data from 2023/03/20& 2023/04/20 repectively.

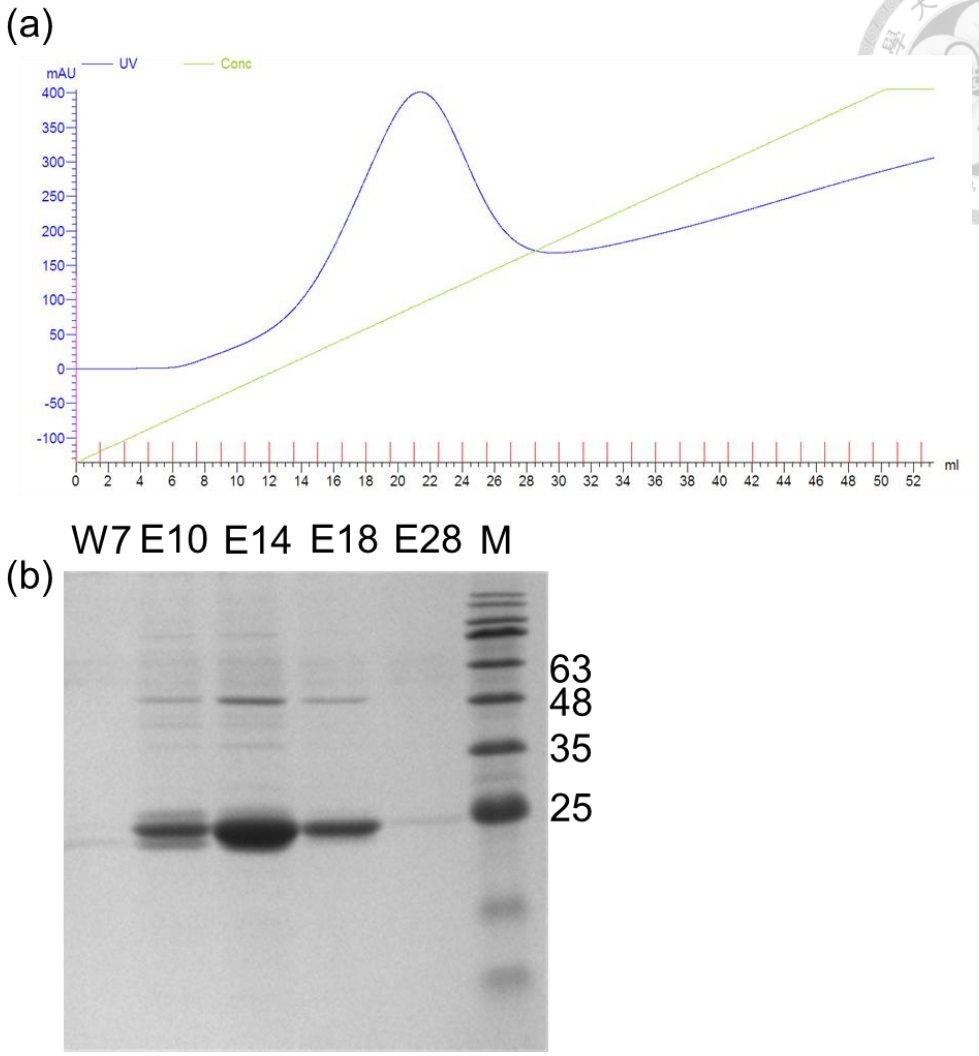

Figure 8. Purification hnRNP H1-N1 through Histrap FF Crude column

(a) Samples after ProTEV cleavage were purified through the Histrap column again. Fractions around the peak were applied to 15% SDS-PAGE to check if the protein was cleaved. (b) The eluted fractions were examined by 15% SDS-PAGE. The cleaved protein should be 21 kDa. The 7th fraction from the wash step was also applied to check. Data from 2024/08/07.

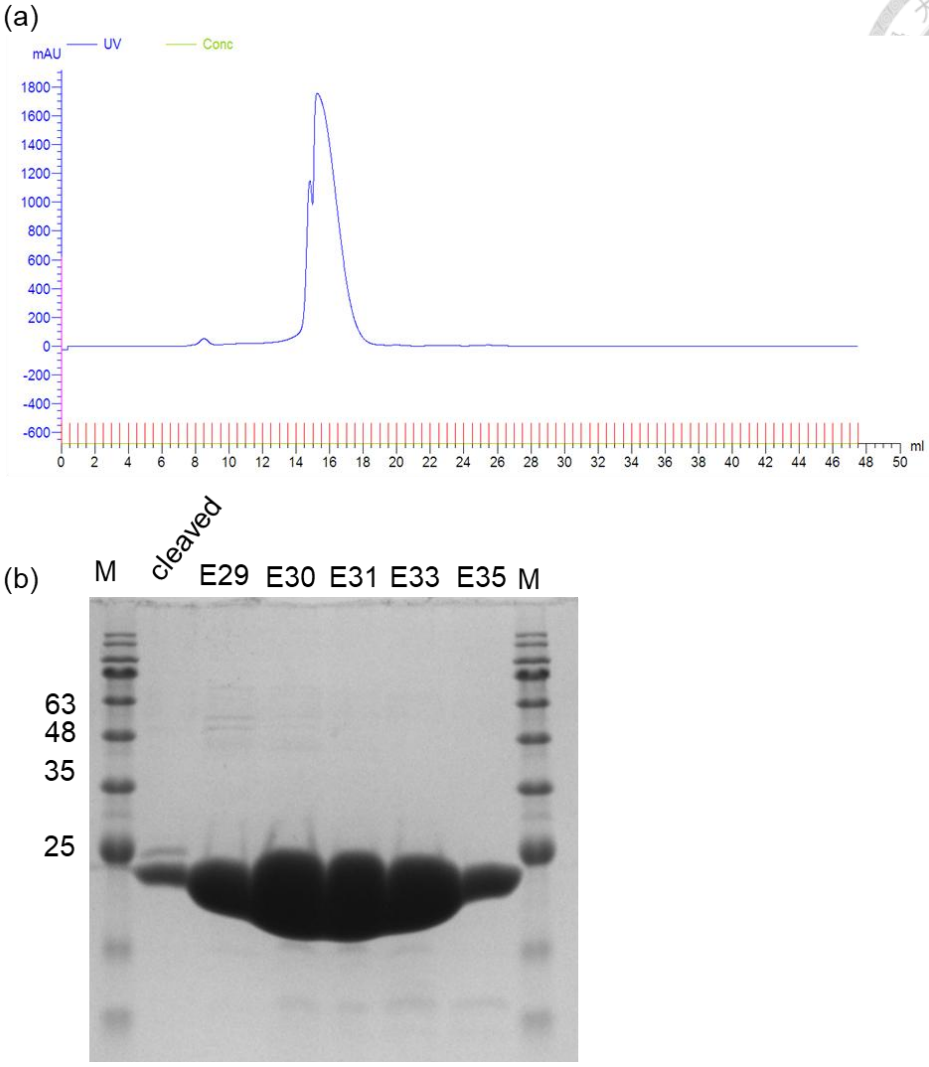

Figure 9. Purification hnRNP H1-N1 through Superdex 200 column

(a) Samples after second Histrap were purified through Superdex 200 column, fractions among the peak were applied to 15% SDS-PAGE to check if the protein was cleaved. (b) The eluted fractions were examined by 15% SDS-PAGE. The first lane was samples after second Histrap, it was also applied to check the protein. Data from 2024/08/08.

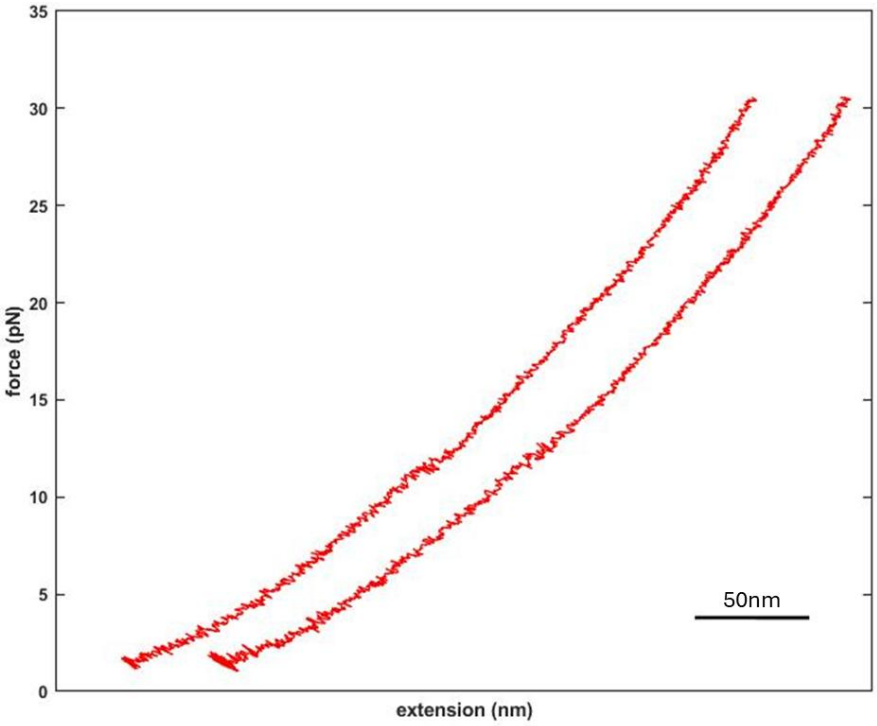

Figure 10. Ramping traces of r(G4C2)4 inTK100

Traces in TK100 often show small hopping between 10-15 pN. Two traces were chose to represent the data obtained in TK100 experiments. Data from 2024/11/19.

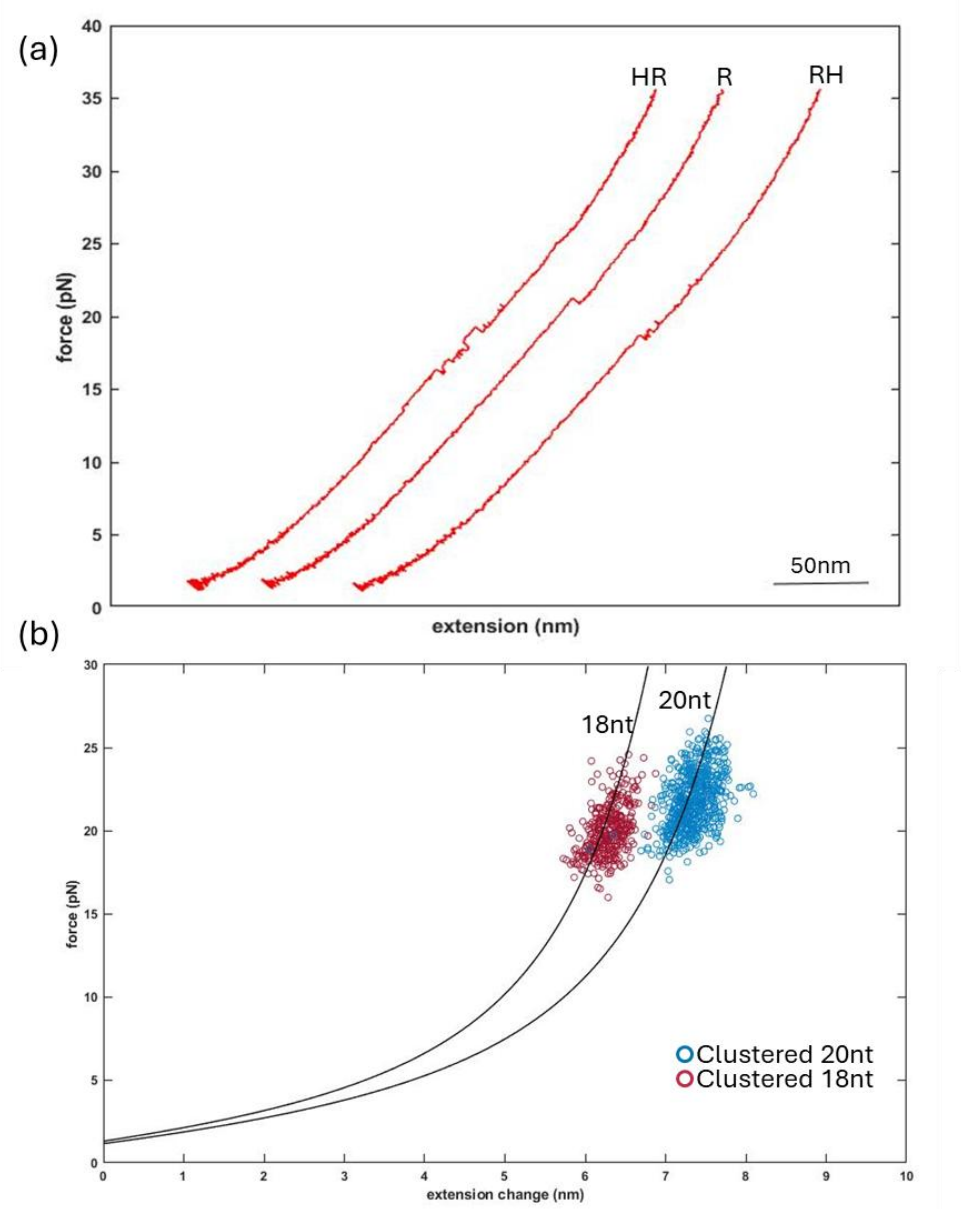

Figure 11. Ramping traces and F-X plot of r$(G_4C_2)_4$ inTMg10

(a) Different traces observed in ramping experiments. Both traces with hopping (HR and RH) appera in 15-20 pN, while R appears in 20-25 pN. The traces were from experiments in 2024/11/26. (b) F-X plot of all of ramping data in TMg10, after clustering, the traces were devided into 18 nt and 20 nt groups. These groups are respectively fitted to extensible WLC model.

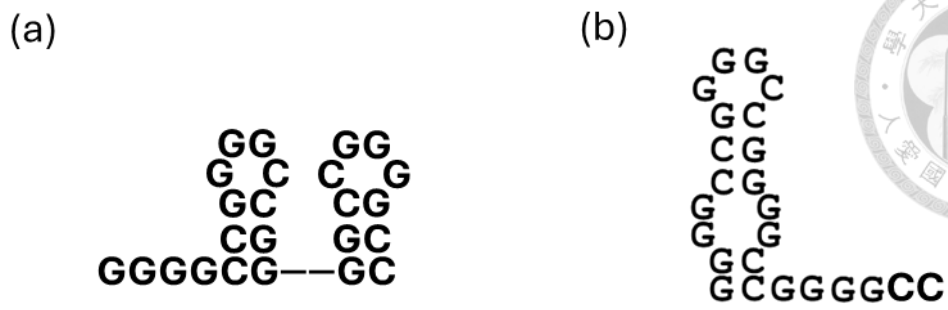

Figure 12. Possible structures of r(G$_4$C$_2$)$_4$ inTMg10

(a) 20 nt double hairpin; (b) 18 nt hairpin using 3 of the repeats.

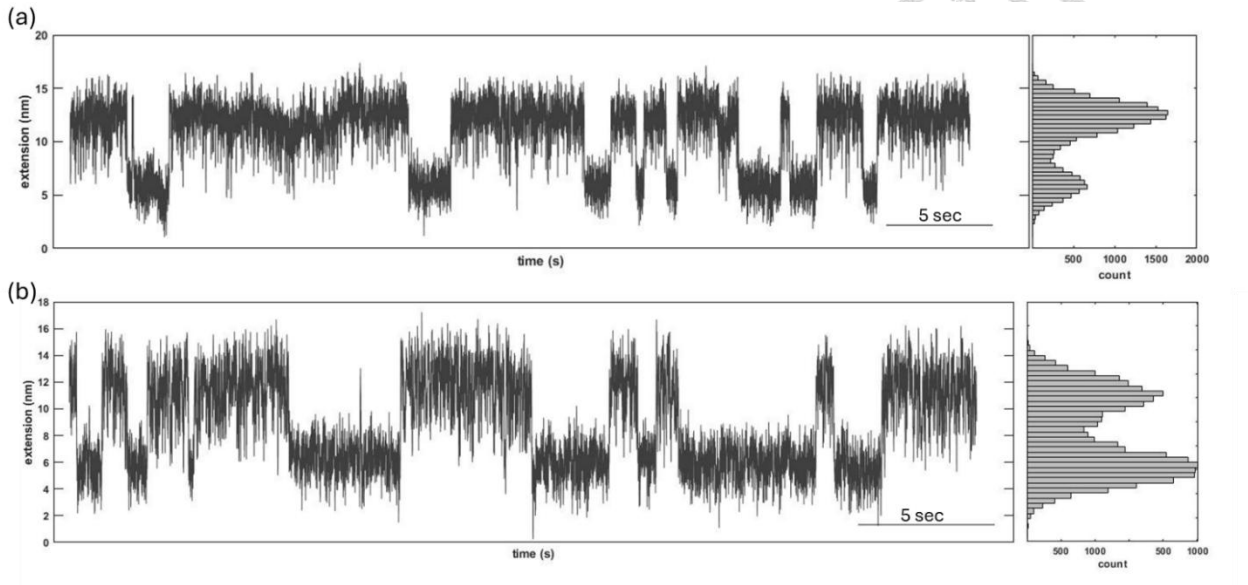

Figure 13. Constant force traces of $r(G_4C_2)_4$ inTMg10

For the trace of constant force, the upper state is the unfolded state, while the lower is the folded state. (a) Trace under 20 pN. The transition size between two states can be estimated by fitting the trace with two Gaussians. (b) Trace under 19 pN. Both traces were from 2024/11/12.

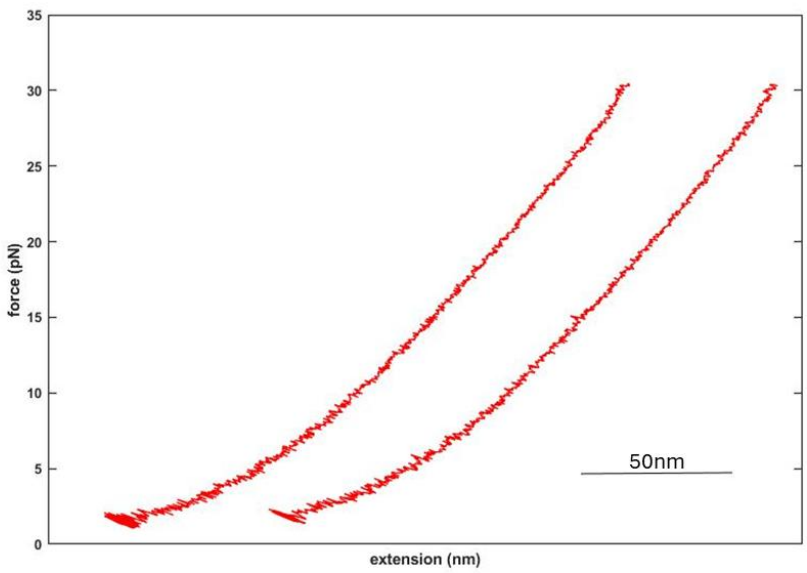

Figure 14. Ramping traces of r(G_2C_4)$_4$ inTMg10

Traces observed in ramping experiments of r(G_2C_4)$_4$ in TMg10. Most of the traces show either no transitions (left) or small hoppings (right). Data from 2024/11/22.

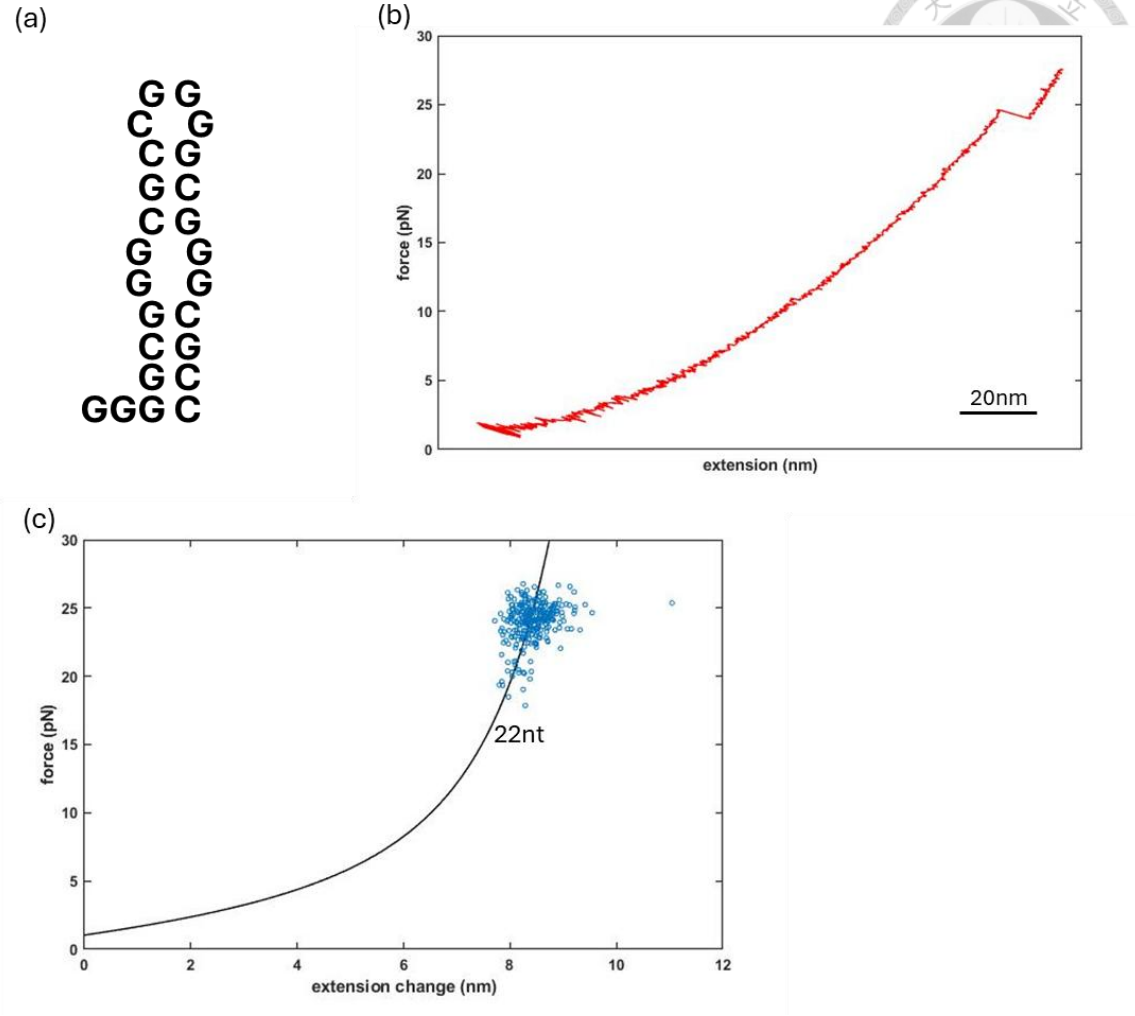

Figure 15. Structures of GCswap RNA, ramping traces and F-X plot of GCswap inTMg10

(a) 22 nt hairpin, predicted by RNAfold. (b) A representative trace observed in ramping experiments of GCswap in TMg10. Most traces show a single rip. (c) F-X plot of GCswap. Only one cluster was identified. All the traces were fit to a 22 nt worm-like chain model. Data of (b) was from 2024/12/31.

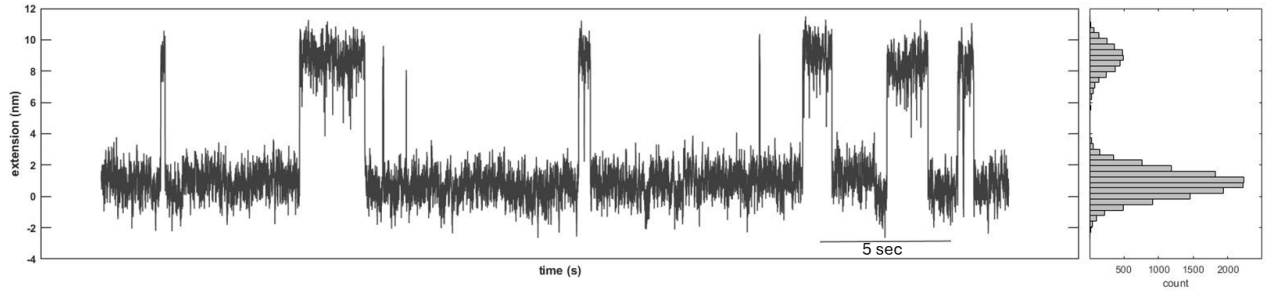

Figure 16. Constant force traces of GCswap inTMg10

For the trace of constant force, the upper state is the unfolded state, while the lower is the folded state. Trace was obtained by holding the structure under 22 pN, the Gaussian fitting for the two states of trace indicated the size of the transition between states. Data from 2024/12/20.

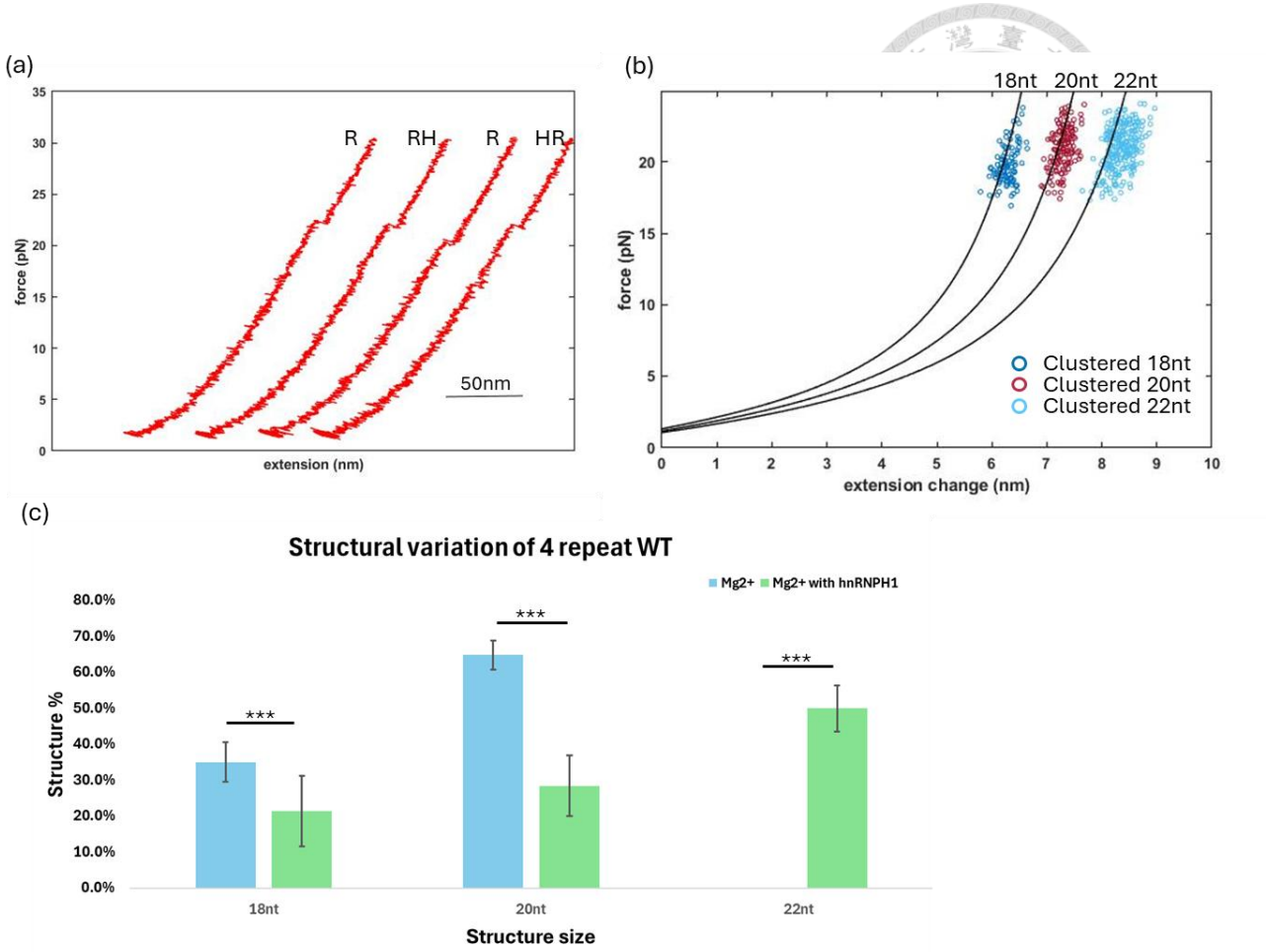

Figure 17. Ramping traces and F-X plot of r(G4C2)4 inTMg10 with hnRNP H1-N1

(a) Different traces observed in ramping experiments, all traces were found in 18-24 pN. Two R were shown because of the size difference. The traces were from experiments in 2024/11/12; (b) F-X plot of all of ramping data, after clustering, the traces were devided into 18 nt, 20 nt and 22 nt groups. These groups are further fitted to extensible WLC model; (c) Statistical analyse of changes in structure size after normalization, all of the data has $p<0.001$ after chi square test, while 20 nt has a smaller p value than 18 nt. The error bar shown in figure indicated the standard error of all the data were between ±10%.

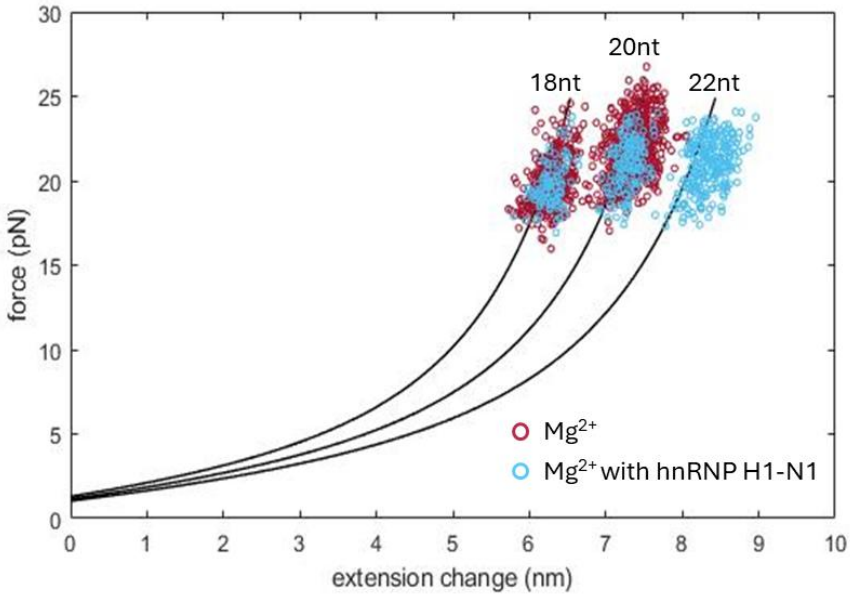

Figure 18. F-X plot of r(G4C2)4 from with/without hnRNP H1-N1

Data from both conditions were pooled together, as 18 nt and 20 nt mostly overlaps.

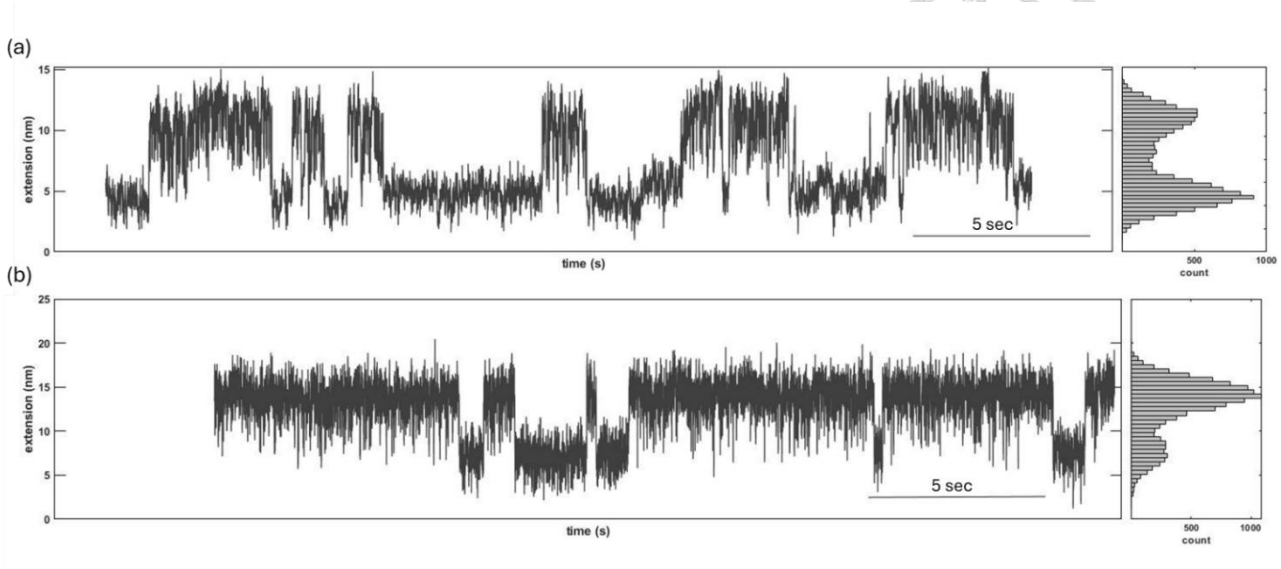

Figure 19. Constant force traces of r(G_4C_2)$_4$ inTMg10

For the trace of constant force, the upper state is unfolding state, while the lower is folded state. (a) Trace under 19 pN, the Gaussian fitting for the to states of trace indicated the size of the transition between states. Data from 2024/11/26. (b) Trace under 20 pN. from 2024/11/12.

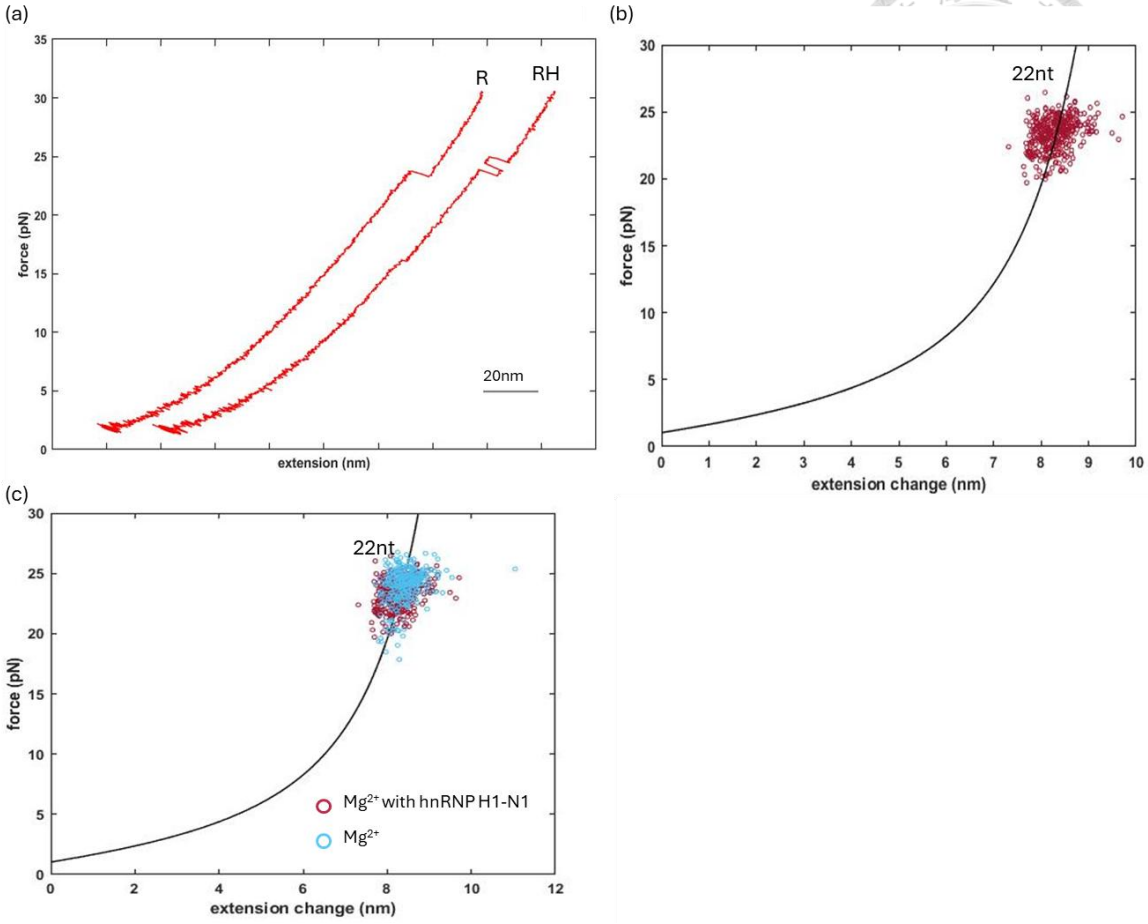

Figure 20. Ramping traces and F-X plot of GCswap inTMg10 with/without hnRNP H1-N1

(a) Different traces observed in ramping experiments, all traces were found in 18-27 pN. The traces were from experiments in 2024/12/20. (b) F-X plot of all of ramping data. (c) Data from both conditions were pooled together, the distribution of force and extension mostly overlaps.

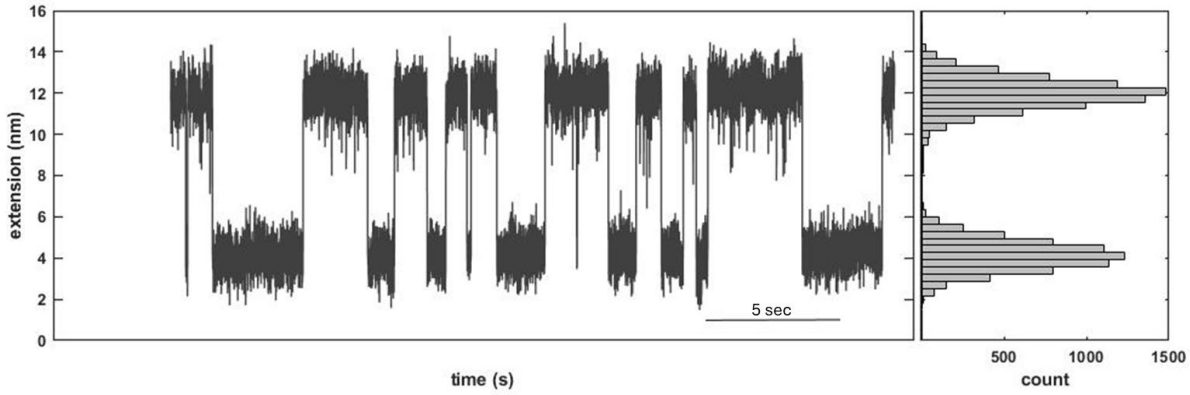

Figure 21. Constant force traces of r(G_4C_2)$_4$ inTMg10

For the trace of constant force, the upper state is unfolding state, while the lower is folded state. Trace was obtained by holding under 22 pN, the Gaussian fitting for the to states of trace indicated the size of the transition between states. Data from 2024/12/31.

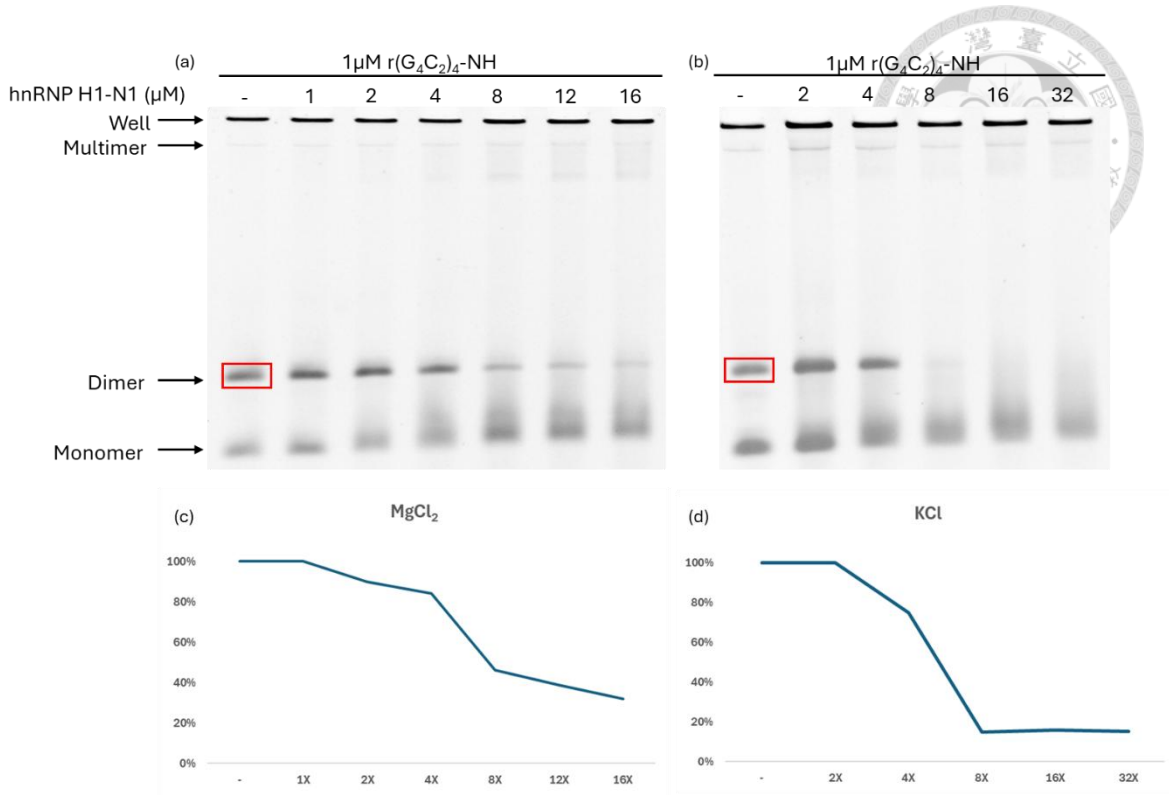

Figure 22. EMSA analysis of r(G4C2)4-NH in different ion conditions

(a) EMSA of r(G4C2)4-NH in TMg10 with 8% native PAGE. (b) EMSA of r(G4C2)4-NH in TK100 with 8% native PAGE. The red box indicates the ROI that was analyzed by Image J. Data were from (a) 2024/12/15 and (b) 2025/01/06. (c) and (d), line charts indicated the change of ROIs.

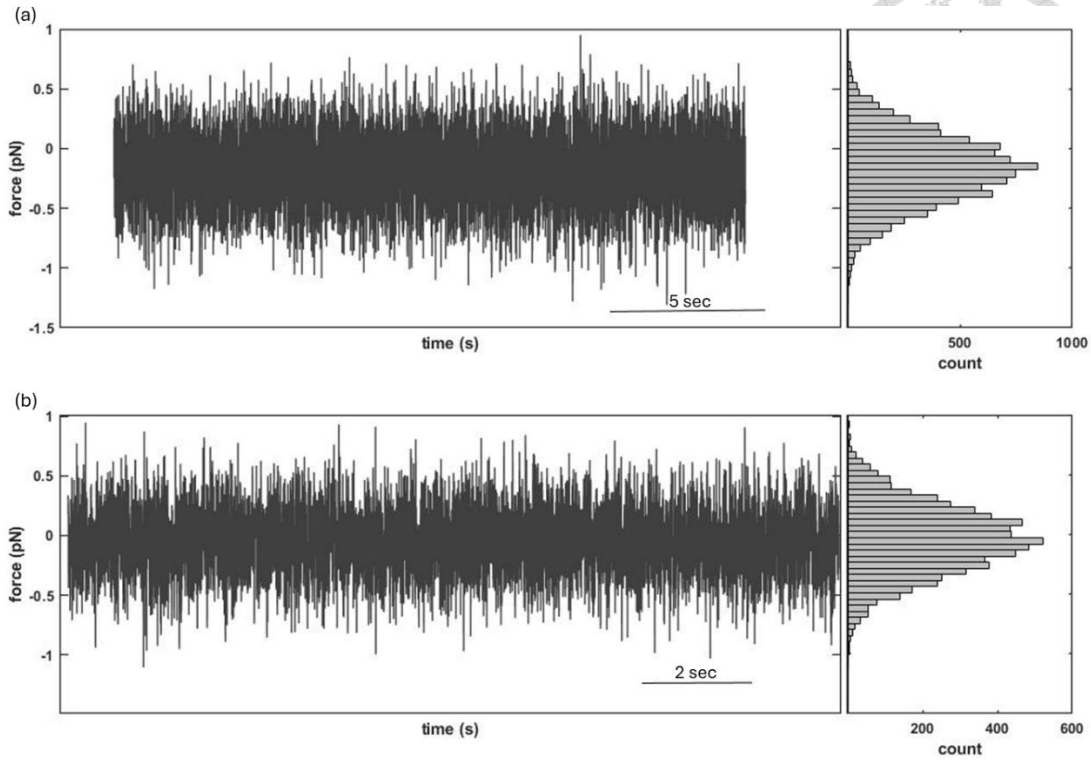

Figure 23. Gaussian fitting for the determination for the effect of noise

(a) Fluctuations of the AD bead recorded while potein was not in the system. The trace was fit with one Gaussian to calculate standard deviation. (b) Fluctuations of AD bead recorded while potein was in the system, the trace was then fit with one Gaussian to calculate standard deviation. Both traces were from 2024/12/03.

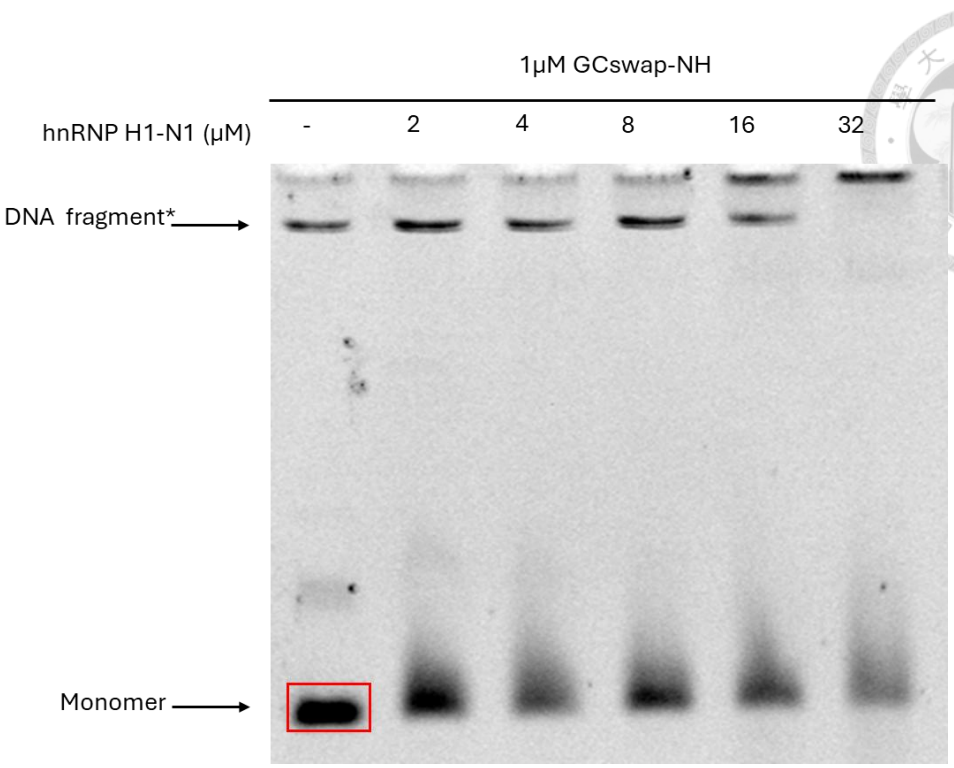

Figure 24. EMSA analysis of GCswap in Mg^{2+}

EMSA of GCswap-NH in TMg10 with 8% native PAGE. The red box indicates the ROI that was analyzed by Image J. Data was from 2025/01/11. * DNA fragments appear because DNase I was not added to the reaction mix after transcription.

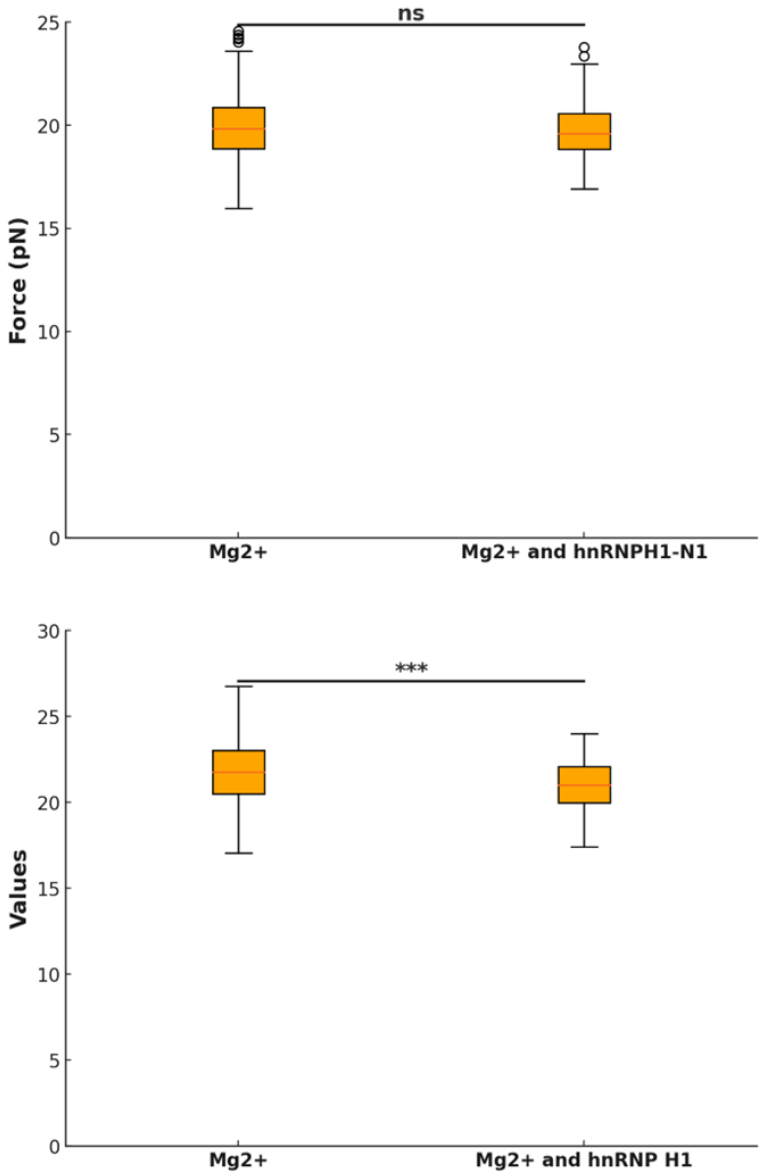

Figure 25. Box plot of unfolding force in 18 nt and 20 nt groups

(a) The unfolding force of 18 nt groups in different conditions. P-value obtained through the t-test was 0.188, indicate that thechange was none-significant (ns). (b) The unfolding force of 20 nt groups in different conditions. P-value obtained through the t-test was 3.05×10^{-8}, indicate that thechange was very significant (***).

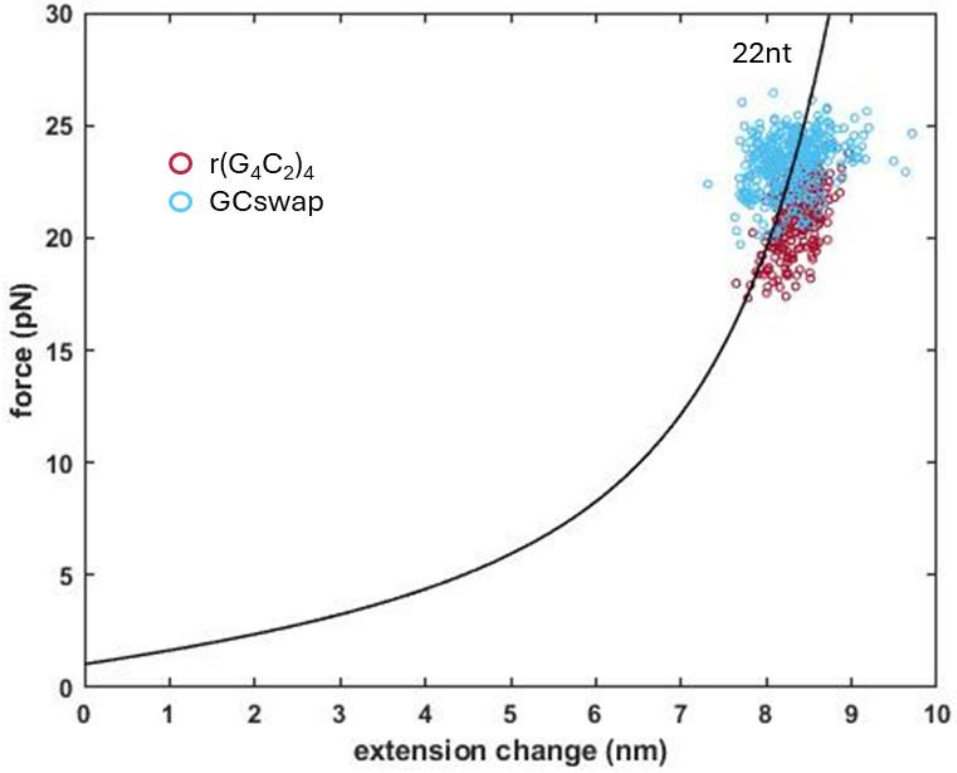

Figure 26. F-X plot of GCswap and r(G4C2)4 inTMg10 with hnRNP H1-N1

F-X plot of GCswap and r(G4C2)4, only the 22 nt cluster from r(G4C2)4 was shown. Both data were from protein experiments, red dots stands for r(G4C2)4, whereas cyan dot stands for GCswap.

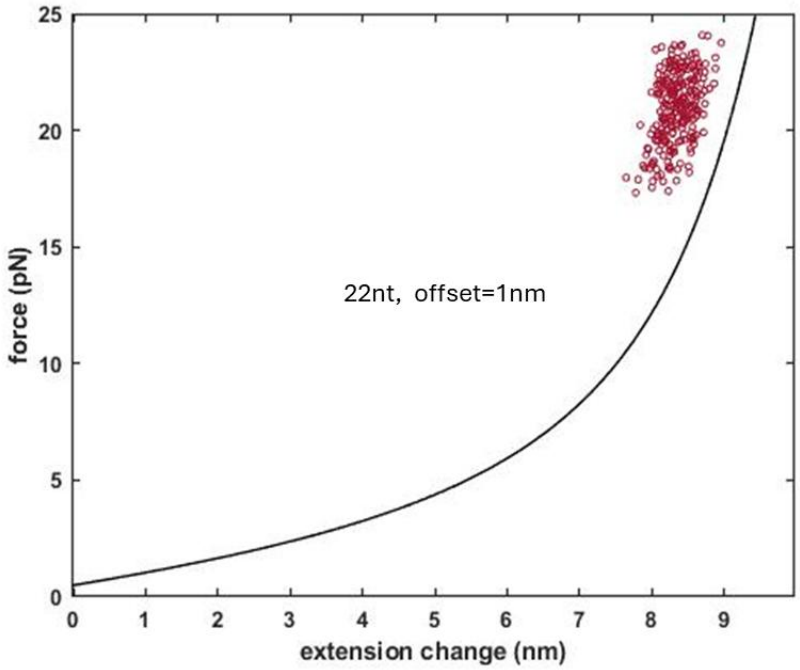

Figure 27 Different fitting of WLC model to the new structure found in the presence of hnRNP H1-N1

F-X plot of $r(G_4C_2)_4$ with WLC applying statistics of parallel GQ fitted.