Screening of Small Molecules and Lead Identification against CAG Repeat RNA for the Treatment of Huntington's Disease and Spinocerebellar Ataxia

M.Sc. Thesis

By
Kanav Gupta

DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
May 2024

Screening of Small Molecules and Lead Identification against CAG Repeat RNA for the Treatment of Huntington's Disease and Spinocerebellar Ataxia

A THESIS

Submitted in partial fulfillment of the requirements for the award of the degree
of
Master of Science

by
Kanav Gupta
2203171009

DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
May 2024

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Screening of Small Molecules and Lead Identification against CAG Repeat RNA for the Treatment of Huntington's Disease and Spinocerebellar Ataxia** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period July, 2022 to May, 2024 under the supervision of Prof. Amit Kumar.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

24/05/2024

Signature of student with date
Kanav Gupta

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Signature of the Supervisor of M.Sc. thesis (with date)
Prof. Amit Kumar

Kanav Gupta has successfully given his M.Sc. Oral Examination held on **9 May 2024.**

Signature of Supervisor of MSc thesis
Date: 24.05.24

P.V. Kodgire
Convener, DPGC
Date: 24/05/2024

ACKNOWLEDGEMENTS

My M.Sc. journey has been possible thanks to the support of countless individuals, and I would like to express my gratitude to them in this section. Special thanks to my supervisor, **Prof. Amit Kumar**, for his guidance and insights that have made this an inspiring experience. His enthusiasm and suggestions have shaped this research into its current form. I am also grateful to the Department of Biotechnology, Government of India, for providing the fellowship that funded my studies.

I extend my thanks to the head and faculty of the Department of Biosciences and Biomedical Engineering (BSBE) for their immense support. My lab members have been a constant source of help and encouragement. I am particularly thankful to **Mr. Krishna Singh Solanki** for his professional and personal guidance, serving as my troubleshooting team, and engaging in both scientific and non-scientific discussions. His support made my time and research journey at IIT Indore significantly easier**. Sakshi Shukla** provided critical assistance during the final phase of my thesis. Her meticulous attention to detail, innovative ideas, and unwavering commitment have been instrumental in advancing our research. I also acknowledge **Ashmad Nayak, Pronamika Chetia, Aritra Chakraborty, Priya Gupta, Aakriti Singh** for their thoughtful scientific discussions and emotional support.

Thanks to my dear friend Dhwani Thakkar for her encouragement, support and tolerating me and my baseless talks. I am also grateful to Vaishali Saini, and my seniors Rahul Chauhan, Kavita Singh, Kundan Solanki, and Priyanka Payal, who, although not members of my lab, provided invaluable assistance with my experiments and supported me personally.

I am grateful to Dr. Ravinder, Mr. Ghanshyam Bhavsar, and Mr. Kinny Pandey from SIC, IIT Indore, for their help with high-end techniques. I appreciate the assistance of departmental staff members Mr. Amit Kumar Mishra, Mr. Gaurav Kumar, and Mr. Arif Patel throughout my M.Sc.

program. My sincere gratitude goes to the Director of IIT Indore, and I also thank all the non-teaching staff, housekeeping staff, and security staff at IIT Indore.

On a personal level, I have been supported by many people who helped me face the challenges of my M.Sc. journey. I am especially grateful to my family. I deeply appreciate my parents, **Mrs. Ritu Gupta,** and **Mr. Sudhir Gupta**, for nurturing my curiosity and observational skills. My brother, **Himank Gupta**, has been a great support in my personal life.

While it would take many pages to mention everyone, I also want to thank all the individuals, named and unnamed, who have directly or indirectly contributed to my journey. I believe all these supportive people are part of the positive energy derived from the ultimate source of energy we call God, to whom I am deeply grateful.

Dedicated to my family.

Abstract

Huntington's disease (HD) is a severe neurodegenerative condition caused by the expansion of CAG repeats in the HTT gene. This expansion leads to the formation of mutant huntingtin proteins, which aggregate and cause cellular toxicity. HD predominantly affects the striatum, particularly the caudate nucleus and putamen, resulting in the degeneration of GABAergic medium spiny neurons. These pathological changes are accompanied by a decrease in GABA and substance P levels, as well as a reduction in dopamine receptor density. Recent therapeutic research has shifted towards targeting RNA to mitigate the pathogenesis of HD. This study focuses on the screening of FDA-approved small molecules for their potential to bind with CAG repeat sequences and alleviate toxicity caused by CAG repeat sequences. Biophysical analyses identified molecule D25 as having specificity and selectivity towards RNA motifs containing CAG repeats. Further experimentation confirmed D25's efficacy in reducing polyglutamine protein formation through cell-based studies and neurobehavioral improvement in *in-vivo* models. Thus, D25 can be used as a potential therapeutic drug against neurological diseases caused due to CAG repeat expansion such as HD and SCA. This approach could pave the way for developing targeted and effective treatments for HD and similar trinucleotide repeat disorders.

Keywords: **Huntington’s Disease, CAG, Small Molecules, HTT gene, PolyQ**

LIST OF PUBLICATIONS

(a) Publication from thesis

Kanav Gupta, Amit Kumar*. Oseltamivir Alleviates Toxicity from CAG Repeat RNA in Huntington's Disease and Spinocerebellar Ataxia through Interaction with Expanded Repeats. (Manuscript to be communicated)

(b) Publication apart from thesis:

Sakshi Shukla, **Kanav Gupta**, Krishna Singh, Amit Kumar*. An Updated Canvas of the RFC1-mediated CANVAS (Cerebellar Ataxia, Neuropathy and Vestibular Areflexia Syndrome**), Molecular Neurobiology**. (Submission ID: 5b270c26-12b5-4523-9be4-943aec0409fe) (Under peer review)

TABLE OF CONTENTS

LIST OF FIGURES

Nomenclature

Λ	Wavelength
ε	Extinction coefficient
°C	Degree Centigrade
Δ	Delta
µL	Microliter
mL	Milliliter
nm	Nanometer
nM	Nano molar
µM	Micro molar
M	Molar
s	Seconds
min	Minutes
hr	Hour

Acronyms

CD	Circular Dichroism
FBA	Fluorescence Binding assay
FID	Fluorescence Intercalator Displacement Assay
D/N	Drug/Nucleic Acid
DMSO	Di methyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
LB	Luria Broth
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
RNA	Ribonucleic acid
Kd	Dissociation constant
Ksv	Stern-Volmer Constant
DTT	Dithiothreitol
PMSF	Phenylmethylsulphonyl fluoride
SDS	Sodium dodecyl-sulfate
Ni-NTA	Nickel-Nitrilotriacetic Acid
IPTG	Isopropylthio-β-galactoside
CV	Column Volume
OD	Optical Density
TNR	Trinucleotide repeat
UTR	Untranslated region
AONs	Antisense Oligonucleotides
APS	Ammonium Persulphate
TEMED	Tetramethylethylenediamine

Chapter 1
Introduction

Neurological disorders, as medically described, are disorders of the nervous system involving all autonomous CNS and PNS. In the upcoming decades, the number of fatalities and disabilities brought on by neurological disorders are predicted to increase and are now becoming more widely acknowledged as a worldwide public health concern. Even while communicable neurological disorders such as Acute meningitis, caused by *N. meningitidis*, have decreased overthe past 30 years, owing to medical amelioration, but the absolute number of fatalities has climbed by 39%, and an increase of 15 % is observed in disability-adjusted life years (DALYs), which are the sum of years of life lost and years lived with impairment caused due to neurological disorders (Mahajan and Patil 2021). Neurological disorders implicate both benign and malignant tumors, seizures, disorders of brain development, cerebrovascular accidents, neurological infections, headaches, degenerative disorders, and facial pain (Feigin et al. 2020).

Neurodegenerative disorders are neurological disorders involving cumulative damage to nerve cells cardinally in brain and involve disorders such as Creutzfeldt-Jakob disease, Vascular dementia, Progressive supranuclear palsy, Frontotemporal dementia, Huntington's disease, Dementia with Lewy bodies, Parkinson's disease, Alzheimer's disease (AD) (‘Focus on Neurodegenerative Disease’ 2018). Globally, 50 million people were affected by neurodegenerative disorders in 2019, a number projected to reach 160 million by 2060, often leading to dementia (Armstrong 2020).

In 2020, a hypothesis was put forward to explain the cause of variable neurodegenerative disorders, which entails the primary cause as accelerated aging, governed by allostatic load, resulting in differential aging of neuroanatomical pathways generally long-range synaptic connections and release of several reactive breakdown products such as TDP- 43, FUS, Ab, tau and α-synuclein (Armstrong 2020). Age significantly increases the risk of acquiring a neurodegenerative illness. Given the rise in life expectancy, more individuals may develop neurodegenerative disorders in future decades. Genetic mutations can lead to accelerated genesis of similar misfolded and aggregated proteins, which can overwhelm ubiquitin and phagosome- lysosome system (Ross and Poirier 2004).

1.1 Repeat Expansion disorders

Repeat instability is one of the major causes of neurodegenerative disorders (Pearson, Edamura, and Cleary 2005). Microsatellites are short tandem repeats (STR) of (1-6) base pairs polymorphic sequences dispersed throughout the human genome. Short tandem repeats (STRs) are highly variable genetic elements, comprising at least 6.77% of the human genome. The instability of these sequences increases with their length, making larger STR expansions more prone to further expansion during the process of DNA replication. These microsatellites, over the past several decades, have been used for linkage studies; however, in the early 90s, two neurological disorders were found to be caused by repeat instability in these microsatellite regions (Fu et al. 1991). The size of the repeat instability can range from trinucleotide (FXTAS.FXS), tetranucleotides (DM2), pentanucleotides (SCA10), hexanucleotides (ALS), dodecanucleotides (EPM1) (Paulson 2018). Characteristically these expansions are dynamic in nature, unlike static mutation. These repeat expansions can be found in 5′ UTR, coding exons, introns, and 3′ UTR region of a gene.

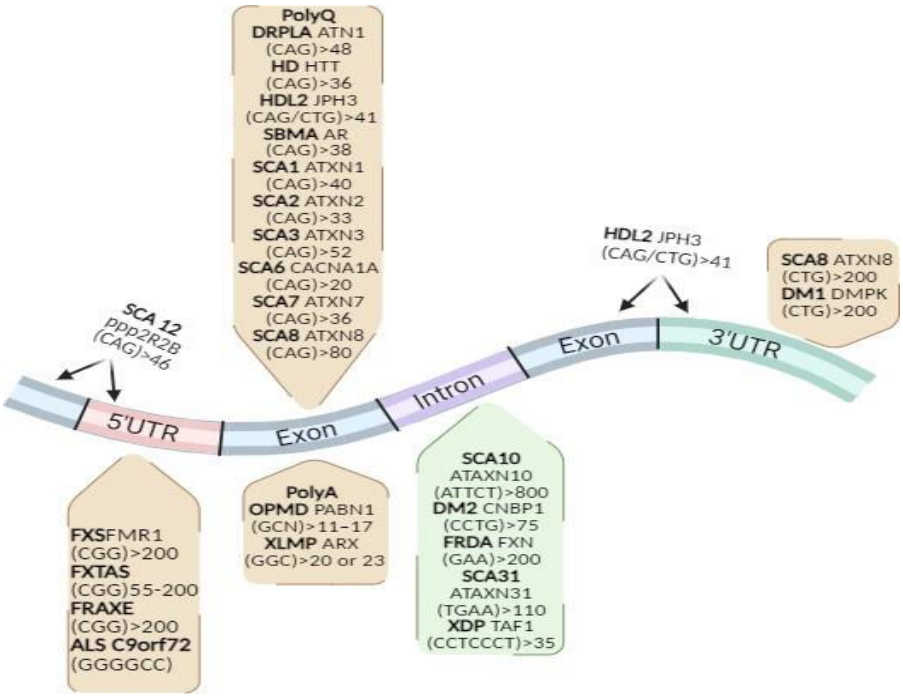


Figure 1.1. Representing different types of repeat expansion disorders

1.2 Pathogenic mechanism

The severity of the disease shows a robust negative correlation with the length of repeats (Budworth and McMurray 2013). There are mainly 4 different pathogenic mechanisms of repeat expansion disorders a) loss of function of repeats (FXS) through transcription repression b) gain of function of RNA (protein sequestration) (DM1) c) gain of function of toxic protein (Fiszer and Krzyzosiak 2013) (HD) d) RAN translation (Malik et al. 2021).

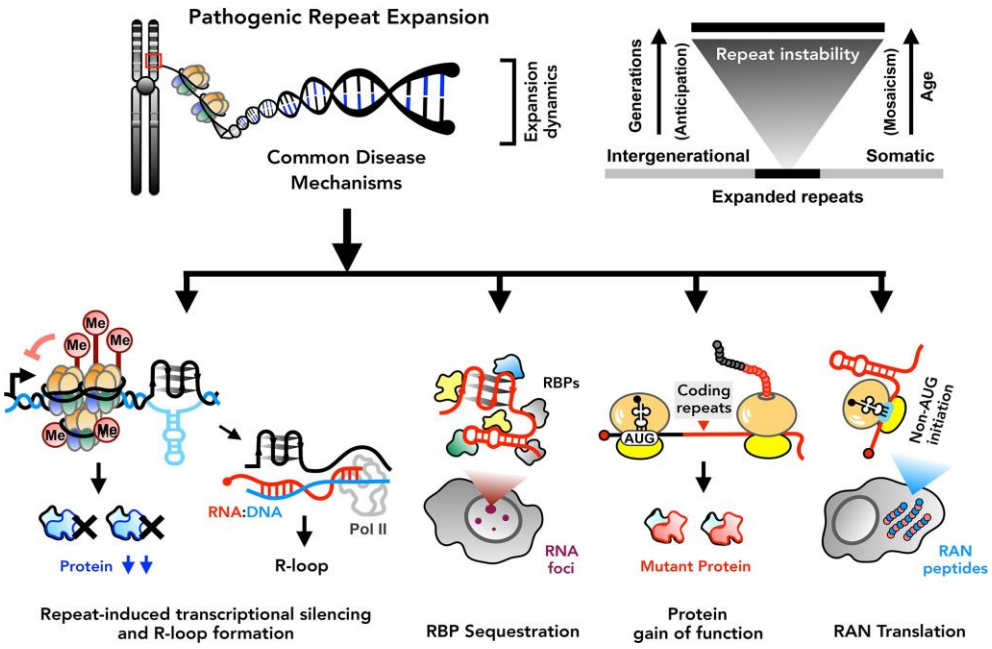


Figure 1.2. Represents different types of repeat expansion disorders

Diseases such as Fragile X syndrome and Friedreich's ataxia are caused by the expansion of (CGG) repeats and (GAA). (TTC) which leads to the loss of function of repeats. The loss of function of repeats in these diseases is due to transcriptional silencing caused by DNA methylation.

As a result of transcriptional silencing, reduced levels of FMRP and FRATXAN proteins are observed in FXS and Friedreich's ataxia, respectively. A decrease in the transcription of FXS leads to hyper-methylation of histone H3 lysine 9, triple-stranded DNA (triplex DNA), and formation of RNA: DNA hybrid.

The gain of function of RNA is another mechanism by which repeat expansion causes diseases. Myotonic Dystrophy is a well-renowned disease that is caused by the expansion of CAG repeat in 3′ UTR. RNA gain of function leads to the formation of RNA foci, which sequesters several important proteins involved in splicing, which leads to aberrant splicing defects. Several proteins, such as MBNL, are sequestered, leading to ribonuclear foci. RNA-mediated toxicity is also observed in HD and other ataxia,

Gain of protein function encompasses most repeat expansion disorders and are characterized by Poly-Q and Poly-A protein formation, which contains polyglutamine in their amino-terminal domain of proteins. SBMA was the first Poly-Q disease caused by the expansion of the gene of the androgen receptor in the X-chromosome. The expansion in the exon of the IT15 gene causes HD. These mutant proteins have aggregation ability and cause several cellular dysregulations, such as mitochondrial dysfunction, impairment of axonal transport, and transcriptional dysregulation.

RAN translation is a type of translation that does not require a start codon. Due to expansion in mRNA, it attains secondary structures such as hairpins, which can sequester translation factors and stall ribosomes during translation initiation. DM1 and SCA8 both showed non-AUG translation.

1.3 CAG repeat trinucleotide expansion disorder

Trinucleotide expansion disorder involving expansion of (CAG) repeats in coding exons are involved with polyglutamine toxic protein which form aggregates inside the cell. One of the major disease involving (CAG) repeat expansion is Huntington’s disease. In 1993, it was found that HD is caused due to repeat expansion in IT15 gene of chromosome 4 (4p13) (MacDonald 1993). Huntington’s disease occurs when this expansion increases beyond 37 repeats and its severity is directly correlated with the length of repeats. Protein product from HTT gene acts as a scaffold protein in autophagy, however in HD, an expanded polyglutamine tract is found near amino

terminus of the protein which forms aggregates and lead to cellular toxicity.

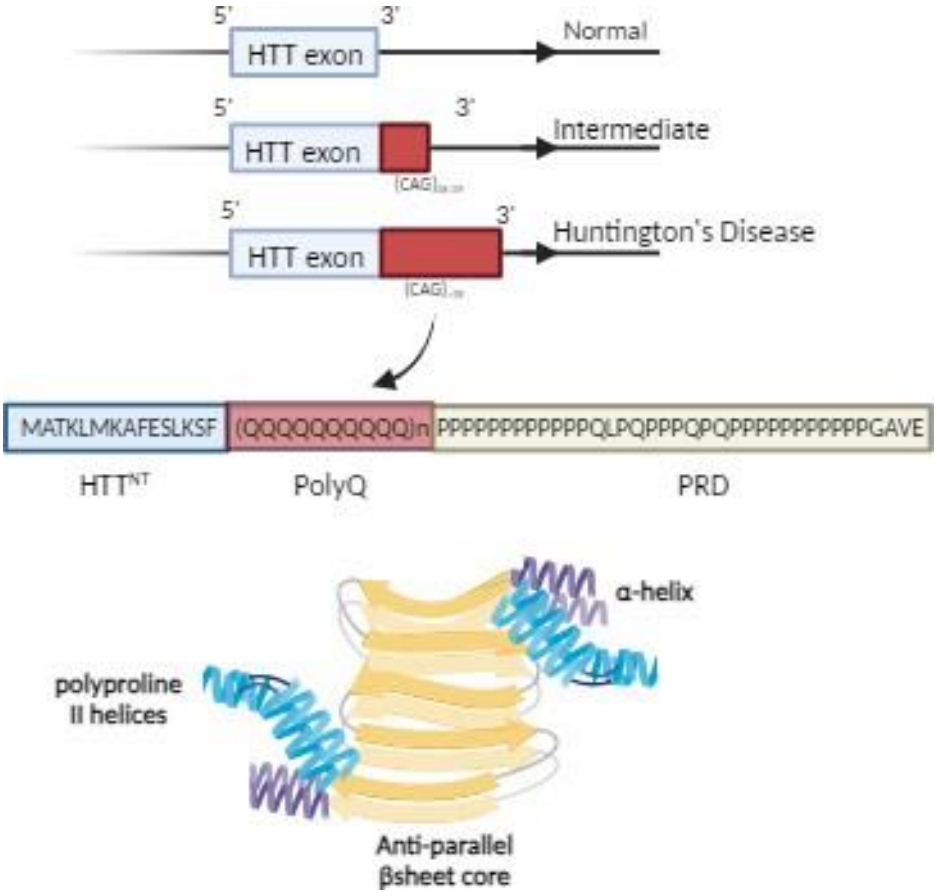


Figure 1.3. Represents PolyQ formation in Huntington's disease

RAN translation is also observed in HD (Newell et al. 1999). The degeneration of the striatum, particularly the caudate nucleus and putamen, is the most obvious neuropathological aspect of HD. The GABAergic medium spiny neurons (MSNs), which serve as the primary projection neurons in the striatum, are the most significantly affected. In more severe cases of illness, this causes the caudate and putamen to atrophize and shrink (Vonsattel et al. 1985). In addition to striatum, there is considerable cortical atrophy. In the later stages of the illness, the cortex thins more noticeably. Neuronal degeneration and gliosis occur in cortical layers III, V, and VI (Rüb et al. 2016). GABA and substance P

levels are decreased because of striatal degeneration since these neurotransmitters are mostly located in MSNs. Additionally, there is a considerable reduction in dopamine receptor density, notably D2 receptors (Albin et al. 1989). HD is associated with the formation of intranuclear inclusions (DiFiglia et al. 1997). White matter is also affected in HD as the volume decreases and microstructural changes are observed (Magnotta et al. 2009). Symptoms for Huntington's disease range from involuntary movements to motor abnormalities and hyperkinetic movements (Bates et al. 2015; Andrich et al. 2007).

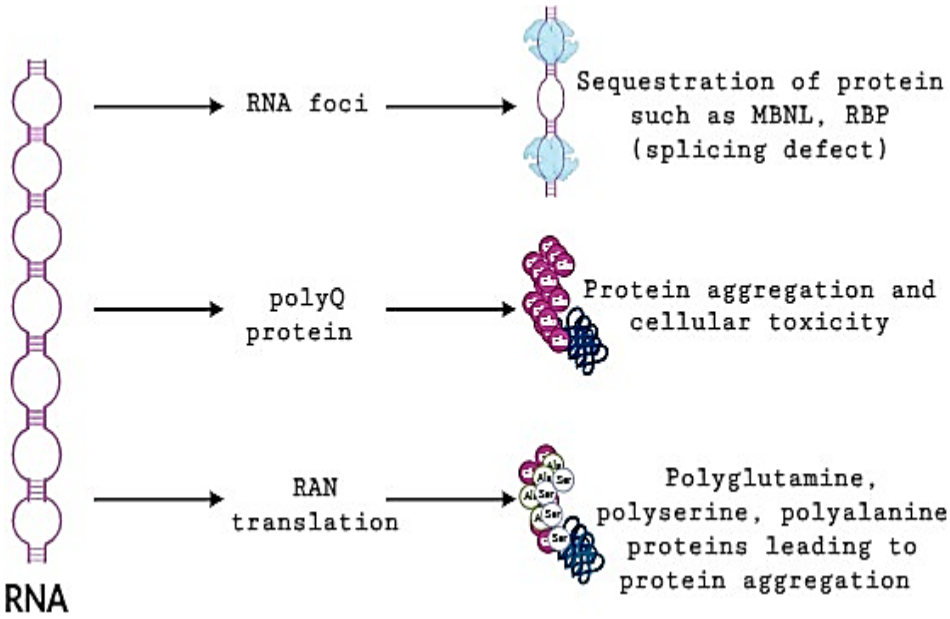


Figure 1.4. Modes of toxicity associated with r(CAG) expansion

Spinal and Bulbar Muscular Atrophy is a repeat expansion disorder in which expansion of CAG repeats in the coding region of the Androgen Receptor of the X chromosome. Normally, 5-34 repeats are present, but expansion of these repeats to 38-70 causes SMBA. DRLPA is also a CAG repeat expansion disorder, which is caused by an expansion in the atrophin-1 Gene. There is no known cure for DRLPA. Other major CAG repeat expansion disorders are Spinocerebellar ataxia 1, Spinocerebellar ataxia 2, SCA3, SCA6, SCA7, and SCA17. They are caused due to expansion in the coding region of ATXN1, ATXN2, MJD1, VGCC, SCA7 gene, TATA-box binding protein (TBP) respectively.

1.4 Therapeutic development against repeat expansion disorders:

Repeat expansion disorder involving trinucleotide mini satellites such as CAG, CGG, CCG are a being toll on human life, thus their therapeutic development is necessary. Earlier no definitive treatment of these expansion disorders was present aside from drugs which alleviated the symptoms caused by the pathophysiology of these repeats such as Tetrabenzaine, Allopreganolone. However, development of therapeutics which directly affect the root cause of these expansion disorders are not present in the market. Directly targeting these expansions requires molecules which can interact with nucleic acids. A major hurdle in designing therapeutics for expansion disorder is selectivity for the gene harboring expansion. Thus, three main strategies for targeting repeat expansion disorders are: a) use of antisense oligonucleotides, b) use of RNAi technology to degrade mRNA transcript containing expanded repeats forming several secondary structures, c) use of small molecules that are specific to the trinucleotide repeats.

RNAi and antisense oligonucleotide can be used to reduce the pathogenic effect of expansion disorder either by inhibiting the transcript or by transcriptional gene silencing. However, there are several shortcomings as the targeted delivery to multiple organ systems and elimination of dominant disease-inducing RNA through these therapeutics is difficult. siRNA has also been used to inhibit mutant protein synthesis. This approach has been utilized in the case of HTT transcripts and ATXN3 transcripts. Several approaches such as inhibiting mutant protein aggregation, inducing autophagy or by intracellular bodies. Protein aggregation can also be reduced by using ubiquitin- tagged proteasome-mediated clearance.

Small molecules are a promising therapeutic against expansion disorders, as they can impede the interaction of protein with RNA. Small molecules are generally chemically synthesized drug molecules with molecular weight less than 500 Da (Li and Kang 2020). Small molecules can form complexes with several proteins and can affect their functioning. Small molecules have also been identified to bind RNA structures which can lead to cleavage or even recruitment of several nucleases. Several small molecules have been identified which can modulate variegated biological processes such as inhibiting mRNA translation, and inhibiting biogenesis of microRNA (Childs- Disney et al. 2022). RNA binding small molecules are evolved from a class of aminoglycosides. CrisprCas9 technology can also be engineered to precisely cleave the specific repeat expansions, facilitating gene correction via NHEJ, causing suppression of transcription of repeat containing RNA.

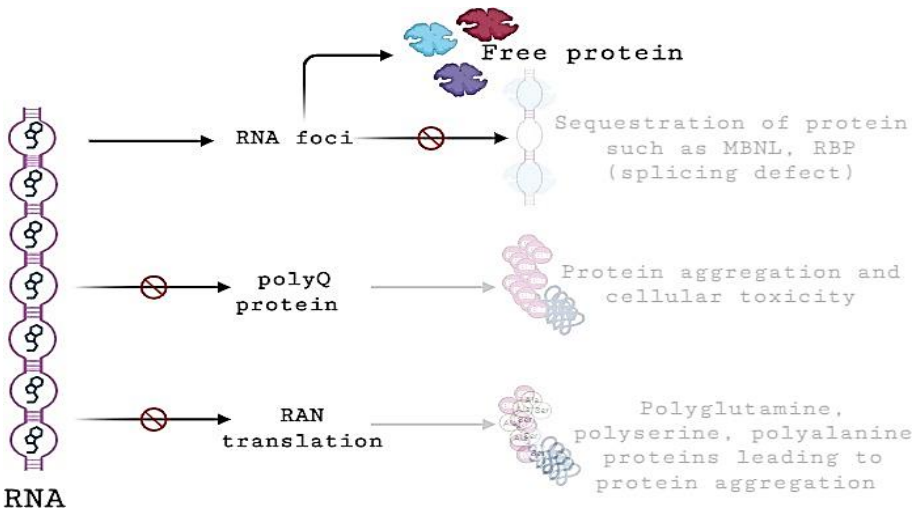


Figure 1.5. Therapeutic development using small molecules

1.5 *Drosophila* as a model for neurological disorders

Drosophila melanogaster, commonly known as fruit fly, is a holometabolous insect that has emerged as an especially effective tool to study the molecular details of repeat expansion disorders. Being an extensively studied model system for genetics and developmental biology for decades, the *Drosophila* genome sequence has shown that 77% of human illness genes are preserved in the fly

Drosophila melanogaster has been a staple in genetic research for over a century. Surprisingly, despite its tiny size and seemingly distant relation to humans, Drosophila shares many genetic similarities with us, making it an excellent model organism for studying various genetic phenomena, including repeat expansion disorders. Here's how Drosophila can be used to study repeat expansion disorders:

1. Conserved Genetic Pathways: Many of the genetic pathways involved in fundamental biological processes are conserved between Drosophila and humans. This includes pathways related to DNA replication, repair, and gene expression regulation, which are crucial in understanding repeat expansion disorders.

2. Genetic Manipulation: Drosophila offers powerful genetic manipulation techniques. Researchers can easily create transgenic flies carrying expanded repeat sequences associated with specific disorders. By introducing these repeats into Drosophila, scientists can study the effects of repeat expansion on gene expression, protein aggregation, and cellular toxicity.

3. Behavioral Studies: Drosophila exhibits complex behaviors that can be easily observed and quantified in the laboratory. Researchers can assess how repeat expansion affects various behaviors in fruit flies, such as locomotion, learning, and memory. These studies provide insights into the neurological aspects of repeat expansion disorders.

4. High Throughput Screening: Drosophila allows for high throughput genetic and pharmacological screens. Researchers can quickly screen large numbers of flies for modifiers of repeat expansion phenotypes. This approach can identify potential therapeutic targets or drugs that alleviate the symptoms associated with repeat expansion disorders.

5. Neurodegeneration Models: Several repeat expansion disorders, such as Huntington's disease and certain types of spinocerebellar ataxias, are characterized by progressive neurodegeneration. Drosophila models of these disorders recapitulate key features of neurodegeneration, including neuronal loss, protein aggregation, and motor dysfunction. Studying these models in fruit flies can uncover molecular mechanisms underlying neurodegeneration and identify potential interventions.

Overall, Drosophila provides a tractable and cost-effective system for studying repeat expansion disorders, offering valuable insights into disease mechanisms and potential therapeutic strategies. Numerous studies using Drosophila have examined basic biological processes such cell death as well as cell proliferation, development, and migration.

Fly complex behaviors also include circadian rhythms, memory and learning, sleep, and aggression. Studies have shown that some of these once believed to be exclusive in humans also have genetic roots in other species (Guo, 2012).

Drosophila melanogaster mimics the cellular environment associated with genetic versions of the diseases utilizing transgenic techniques, while offering the convenience of

a) small size
b) low chromosome number
c) small genome size
d) giant salivary gland chromosomes
e) quick generation time,
f) and the capacity to produce number of flies at a relatively low cost

Drosophila has long been regarded as a key organism for investigating both transmission genetics and animal development, owing to its array of experimental benefits. Through genetic analysis, scientists have uncovered that many developmental genes and processes observed in flies are conserved across various species. Notably, a significant proportion of genes influencing human development have been identified as counterparts to those initially identified in Drosophila.

1.6 The Drosophila genome

The haploid genome of Drosophila comprises approximately 170,000 kilobases of DNA, housing around 13,600 genes, making it roughly 5% of the size of the human genome. Its genetic material is organized into four chromosomes: chromosome 1 represents the X chromosome, while chromosomes 2 through 4 are autosomes. In Drosophila, sex determination follows the XY system, where females possess two X chromosomes (XX) and males have one X and one Y chromosome (XY). The ratio of X chromosomes to autosomes (the X:A ratio) governs the determination of sex.

Classification of *Drosophila melanogaster*

Drosophila melanogaster also known as 'Fruit fly' has the following classification.

Table 1: Classification of Drosophila melanogaster

Kingdom	Animalia
Phylum	Arthropoda
Class	Insecta
Order	Diptera
Family	Drosophilidae
Genes	*Drosophila*
Species	*melanogaster*

1.7 Life Cycle of Drosophila

Flies breeding on a suitable medium allows for a 10-day life cycle, from fertilization to adult fly emergence. Adult Drosophila has a very high capacity for reproduction. A female can have 3000 offspring in her lifetime. Normally, fertilization takes place as the egg is being laid. Each egg is a millimeter long and consists of plenty of yolk that supports embryonic development.

The Drosophila embryo develops within 24 hours and then hatches into a first instar larva which is worm-like and specific for feeding. 24-48 hours after hatching, the larva molts to produce second and third instar larvae. Third instar larvae complete their growth three days after the second moult and then move out of the food. These third instar larvae furtherundergo pupariation as it pupates.

Within the pupal case, the larvae metamorphosis takes about 4 days. This includes disintegration of most larval tissues and the differentiation of cells into adults.

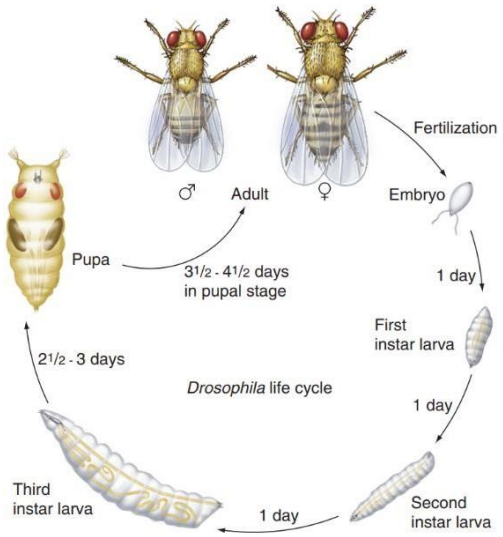


Figure 1.6. Life cycle of Drosophila melanogaster (Figure adapted from T.H. Morgan 1910)

Different Stages in the life cycle of *Drosophila melanogaster* –

Egg: Life cycle begins when the adult flies lay eggs after mating. In general, the egg of *Drosophila melanogaster* is about 0.5 mm long. The ovum in the egg is adjoined by an inner, very thin vitelline envelope and an outer, tough extracellular coat called a chorion, which is opaque and shows a hexagonal marking pattern. An egg consists of two respiratory spiracles and a micropyle. Micropyle is the small opening through which penetration of spermatozoa occurs into the egg. Following the penetration, meiotic (reduction) divisions are completed and female pronucleus is formed. Sperm and egg come into a side-by-side position to produce zygote nucleus, which divides to form the first two cleavage nuclei, the initial stage of development of embryo. Embryo takes 24 hours to mature completely

Figure1.7. Drosophila egg (Adapted from T.H. Morgan 1910)

Larval Stages: Following embryogenesis, a wormlike larva hatches that is particular for feeding and grows dramatically. They voraciously feed and grow. This is commonly known as the mobile phase. The larvae show three growth stages: 1st instar, 2nd instar and 3rd instar. 24 and 48 hours post hatching, the larva moults produce 2nd and 3rd instar larvae. The 3rd instar larvae complete its growth after 3 days of 2nd molt. Pulses of ecdysone which is a steroid molting hormone controls the timing of each larval molt and pupation.

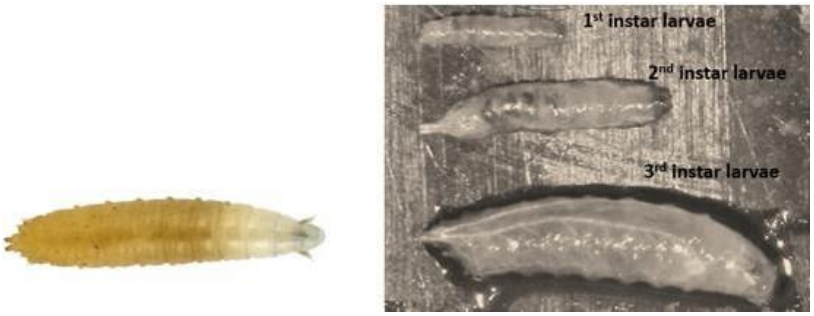


Figure 1.8. Drosophila larvae (Adapted from T.H. Morgan 1910)

Pupa: Larval stages are followed by pupa. This is also known as the static phase. Pupa consists of three stages – pre-pupal stage, mid-pupal stage, and late pupal stage.Once under the pupal case, the larvae undergo metamorphosis. On getting pupate, they creep from the culture medium and adhere to some dry place, such as at the sideof the bottle.

Figure 1.9. Drosophila pupa (Adapted from T.H. Morgan 1910)

Adult: At the end of the metamorphosis, the adult emerges from the pupal case through eclosion, its wings expand, and the entire exoskeleton hardens and becomes pigmented. Then adult fly forms with three predominant body parts- Head, thorax, and abdomen.

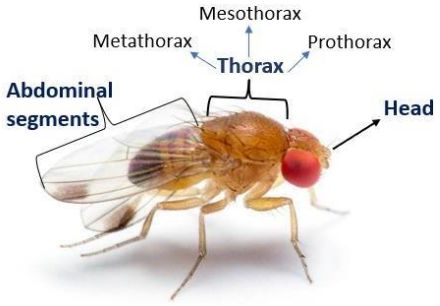


Figure 1.10. Adult Drosophila fly (Adapted from T.H. Morgan 1910)

Identification of male and female *Drosophila melanogaster*

Female flies are larger than the male.

Tip of the abdomen is pointed in the female, while males have a round abdomen tip. The color of the last two segments of the abdomen is darker in males as compared to the females. Females have 7 abdominal segments, whilethe males have 6 abdominal segments.

A group of bristles called sex combs are present on the 3rd tarsal segments ofthe first pair of legs in male *drosophila*.

The ovipositor of the female is pointed, while the clasper of the male is darkblack and circular.

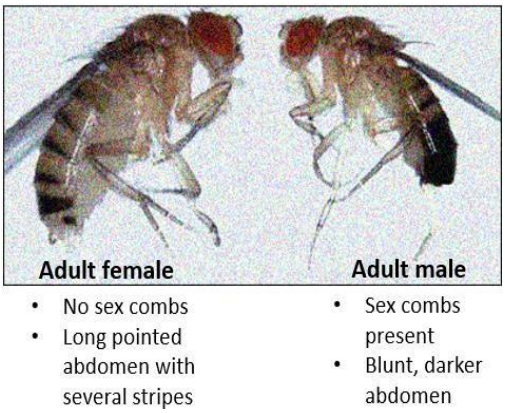


Figure 1.11. Male and female adult flies (Adapted from T.H. Morgan 1910)

1.8 Organization of the thesis

Chapter 1 offers a succinct introduction to the thesis, providing a comprehensive overview of pathogenic repeat expansions and their underlying molecular mechanisms. The chapter delves into CAG repeat expansion disorders, detailing the resultant formation of PolyQ proteins. Additionally, it addresses the development of therapeutic approaches targeting these repeat expansion disorders. The utility of *Drosophila melanogaster* as a model organism for studying CAG repeat expansion disorders, including Huntington's disease, is examined. Furthermore, an overview of the life cycle of *Drosophila melanogaster* is included to contextualize its relevance in research.

Chapter 2 discusses literature review about development of therapeutics against PolyQ expressing repeat expansion disorders and formulates the research gap giving ample confidence for the research work outlined in the thesis.

Chapter 3 gives a concise overview of materials and methodologies followed in performing the present thesis work. It also provides a basic summary of the instrumentation used in the thesis work and its principles.

Chapter 4 discusses the results of the experiments performed. Elaborative biophysical experiments were performed to characterize the interactions between the lead small molecule and CAG repeat containing RNA. Thereafter, docking analysis and molecular dynamics simulations are discussed. Later cell-based studies are discussed which describe the effects of small molecules on the aggregate formation.

Chapter 5 encompasses the concluding remarks and the future perspective of the thesis.

Chapter 2
Literature Review and Problem Formulation

Nucleotide expansion disorders encompass a wide range of conditions, with over 50 recognized diseases, and more than a dozen are attributed to expansions of the CAG trinucleotide sequence. Huntington's disease (HD) stands out as a prototypical example, resulting from CAG repeat expansion in the HTT gene coding region. The severity and pathogenesis of HD correlate with the length of these CAG expansions; longer expansions produce mutant transcripts that are more prone to aggregation, disrupting cellular homeostasis by forming protein aggregates and sequestering vital proteins, including those involved in RNA splicing. The HD-related expansion of CAG repeats in RNA (r (CAG)exp) is central to HD pathogenesis, not only translating into the harmful polyQ HTT protein but also disrupting cellular processes through other routes. For instance, extended r (CAG)exp can hijack vital splicing proteins, resulting in splicing errors. It can also lead to the production of a truncated HTT mRNA with disease implications due to the misappropriation of splicing factors. The discovery that Huntington's disease (HD) stemmed from a specific chromosomal mutation was a pioneering advancement. However, even after two decades of identifying the mutation, a cure remains elusive (Bates et al. 2015; DiFiglia et al. 1997; Rüb et al. 2016; Vonsattel et al. 1985; MacDonald 1993).

The prevalence of HD has notably increased in the Western hemisphere, rising from 6 to approximately 12.3 individuals per 100,000 (Evans et al. 2013). It shows a marked prevalence in Caucasian populations, whereas Asian populations exhibit significantly fewer cases. Traditional therapeutic approaches have focused on

symptom alleviation, employing drugs like reserpine, butyrophenones, and tetrabenazine, which primarily modulate dopaminergic pathways in the brain. Newer agents, such as the dopamine D2 receptor antagonist pridopidine, have entered clinical trials, reflecting ongoing efforts to improve treatment outcomes.

Beyond symptomatic relief, strategies targeting post-transcriptional modifications of mRNA, such as ribozymes, antisense oligonucleotides, and siRNA, aim to mitigate the mutant protein's aggregation. Chromatin-remodeling drugs, specifically anthracyclines, have shown promise in addressing behavioral and neuropathological symptoms (Stack et al. 2007). However, challenges

persist with these drugs due to inconsistent results, issues with blood-brain barrier penetration, poor absorption, and potential for allergic reactions. Some success has been reported with peptides that inhibit caspase-6, which reduces mutant protein levels, and with antisense oligonucleotides that modulate mRNA processing.

Given these challenges, small molecules offer several advantages, including oral bioavailability and an enhanced ability to cross the blood-brain barrier. The pharmacokinetic profiles and potency of small molecules can be more readily optimized. Their interaction with RNA benefits from typically having a lower polar surface area and fewer hydrogen bond donors and acceptors, which contributes to higher RNA binding affinity. Despite the previous development of protein-targeting small molecules like phenothiazine, tetrabenazine, butyrophenones, EGCG, and coenzyme Q10 for trinucleotide disorders, their long-term use raises concerns due to potential adverse effects, which are critical considering the need for prolonged treatment in neurological conditions (de Yebenes et al. 2011; Dickey and La Spada 2018).

These compounds reflect the diverse pharmacological strategies attempted to address the complex pathology of HD (Dickey and La Spada 2018).

a) Tetrabenazine (TBZ): FDA-approved for suppressing involuntary movements in HD, acts as a VMAT2 inhibitor depleting synaptic monoamines.

b) Deutetrabenazine (SD-809): Like TBZ, it has a longer half-life, FDA-approved for chorea in HD.

c) Olanzapine: Atypical neuroleptic, used to manage chorea in HD.

d) Amantadine: Glutamate antagonist has shown some efficacy in reducing chorea.

e) Riluzole: Initially showed promise but later studies revealed it to be ineffective for HD.

f) Coenzyme Q10: Initially promising in mouse models, but human trials have not demonstrated effectiveness.

g) Creatine: Mixed results from trials, with some showing treatment-dependent slowing of brain atrophy but no significant clinical benefit.

h) Ethyl-EPA, Acetyl-L-carnitine, OPC-14117: Trials did not demonstrate effectiveness in HD treatment.

i) Cannabinoids: Some evidence of efficacy in small trials, but overall, not significant.

j) Donepezil: No significant improvement noted in trials.

Given the root of the disease lies at the RNA level, small molecules that target RNA are increasingly favored over protein-targeting counterparts. Extensive research is needed to identify and develop small molecules that specifically target CAG repeat-containing RNA. Such targeted therapeutic agents could revolutionize the treatment of HD and similar trinucleotide repeat disorders by addressing the underlying genetic cause rather than just the downstream Proteinopathy. This research must also consider the drug candidates' therapeutic windows, minimizing off-target effects while maximizing their ability to modulate disease-related pathways. **However, up until very few RNA targeting small molecules are described, such as (HTT-C1, HTT-D1, branapam, and HTT-D3, Myricetin, CP13, however, none of them include FDA approved drug** (Ahamad and Bhat 2022)**. Thus, using an FDA-approved drug against CAG RNA can lead to the development of better therapies to tackle repeat expansion disorder.**

In summary, as the quest for effective treatments for HD and other CAG repeat-associated disorders including SBMA, SCA1, 2, 3, 6, 7, 17 continues, the focus has shifted toward small molecules that can engage directly with the RNA, potentially offering a more direct means of altering the disease course. By specifically binding to RNA sequences, small molecules offer a multi-faceted therapeutic approach. Such compounds have been pinpointed using computational methods and shown to release sequestered splicing proteins in HD patient cells, correcting splicing irregularities. "**This thesis will delve into the identification and characterization of FDA- approved small molecules with the potential to bind and modulate CAG repeat-containing RNA, representing a pivotal step in the journey towards novel and more effective therapies for these challenging genetic disorders"**

Chapter 3
Materials, Methods, and Instrumentation

3.1 Materials

The DNA oligonucleotides utilized in biophysical and molecular studies were acquired from Integrated DNA Technologies, located in Iowa, USA. **Chemical substances** such as NaCl, KCl, $MgCl_2$, K_2HPO_4, KH_2PO_4, NaH_2PO_4, Na_2HPO_4, NaOH, HCl, Tris-base, EDTA, DMSO, APS, TEMED, Acrylamide, Bis-Acrylamide, ethanol, isopropanol, 2-butanol, methanol, urea, Triton-X 100, and others, were sourced from Sigma-Aldrich Chemicals Ltd., based in St. Louis, Missouri, USA and Sisco Research Laboratories Pvt. Ltd. in Mumbai, India. These were predominantly of HPLC purified or molecular grade quality. The dNTPs and rNTPs used for PCR and in-vitro transcription reactions were also obtained from Sigma-Aldrich Chemicals Ltd.

Additional materials such as Agarose, Luria broth, Luria agar, GC agar, and antibiotics like Ampicillin, Kanamycin, Chloramphenicol, Taq polymerase, and others were procured from Himedia Laboratories in India. Standard plastic ware was supplied by Tarsons Products Pvt. Ltd. from Kolkata, India, while glassware was sourced from Borosil. The water purification system from Sartorius Corporate, Germany, was employed to produce molecular biology-grade water essential for preparing buffers and media in various experiments.

3.2 Sample preparation

3.2.1 Oligonucleotide preparation

The oligonucleotides, once desalted and purified, were dissolved in Milli-Q water following the guidelines provided by the manufacturer. To determine their final concentration, their absorbance at 260 nm was measured using Lambert's law, conducted in a denaturing buffer environment. For storage, the oligonucleotides were kept as stock solutions at 4°C for short-term needs and at -20 °C for long-term

preservation, until they were needed for subsequent applications.

3.2.2. Sample Preparation for Fluorescence Intercalator Displacement Assay

For Fluorescence measurements, the working solution of each RNA (500 nM) was prepared in 1X Phosphate buffer containing 50 mM KCl and 1µM Thiazole orange. A final volume of 50 µl was prepared by adding MQ water.

3.2.3. Sample Preparation for Fluorescence Binding Assay

For fluorescence binding measurements, RNA samples were prepared by dissolving RNA oligonucleotides in 1x phosphate buffer containing 50 mM KCl. The solution was heated at 95 º C for 10 minutes and allowed to slowly cool. Binding reactions were done in the final volume using 50 µM of RNA oligonucleotides and varying concentrations of small molecules.

3.2.4. Sample Preparation for Electrophoretic Mobility Shift Assay

The RNA specimens, prepared in a 1x phosphate buffer with 50 mM KCl, underwent annealing through a process of heating them to 92 °C for 10 minutes, followed by a gradual cooling period of 30 minutes. For interaction experiments, these RNA samples were mixed in a solution containing 8 µM RNA oligonucleotides and varying levels of lead small molecules, in a specified final volume. This mixture was then incubated for 30 minutes at a temperature of 25 °C. A 3% agarose gel, made in 1X TBE (Tris-Borate-EDTA) buffer without any added cation, was prepared. After it solidified, the gel was placed in a horizontal electrophoresis tank filled with 1X TBE buffer. Prior to their application to the gel, each sample was combined with 6X orange dye, and then loaded into the designated wells.

3.2.5. Sample Preparation for Circular Dichroism

In the experiment monitored via Circular Dichroism (CD) spectroscopy,

varying amounts of the lead small molecule were incrementally introduced to a fixed concentration of RNA samples, adhering to the designated Drug: RNA (D/N) ratio. Following this, the CD spectra were recorded. The RNA oligonucleotides were initially dissolved in a 1x KPO_4 buffer containing 50 mM KCl. Prior to conducting the CD spectral analysis, these samples underwent a heating process at 92 ° C for 10 minutes and were then allowed to cool slowly to room temperature.

3.2.6. Sample Preparation for NMR studies

RNA samples were prepared in 10mM phosphate buffer, pH 7.2, 0.1 M NaCl, and 50 mM EDTA in 10% D2O. H2O + D2O solvent at a 9:1 ratio was used for all titration studies and 64k data points were recorded for 1D proton NMR spectra. EDTA was added in buffer to avoid the effect of paramagnetic impurities that cause the line broadening during NMR measurement. Sodium trimethylsilylpropanesulfonate (DSS) is the organosilicon compound with the formula $(CH_3)_3SiCH_2CH_2CH_2SO_3{-}Na+$ was taken as internal reference for each set of experiments.

3.2.7. Sample Preparation for Isothermal Titration Calorimetry

RNA dissolved in 10 mM phosphate buffer and 3 M KCl and annealed by heating at 90 ºC for 10 minutes followed by slow cooling at room temperature. The working solution of both RNA and ligand were degassed just before each binding experiments.

3.2.8. Sample Preparation for Thermal studies

RNA samples were prepared in 10 mM phosphate buffer containing 3 M KCl. The RNA samples were heated at 90 ºC for 10 minutes and re-annealed by slow cooling at room temperature. RNA was heated from 25 ºC to 95 ºC and the absorbance at 260 nm was monitored at a rate of 1 ºC/min, in absence and presence of compound of interest at varying D/N ratios.

3.3 RNA oligo synthesis using T7-RNA polymerase-mediated *In-vitro* transcription reactions

3.3.1. Expression and purification of T7-polymerase

A single colony of transformed cells was cultured in LB broth with 50 μg/mL ampicillin and 34 μg/mL chloramphenicol, incubated overnight at 37 °C. For the secondary culture, fresh LB broth was used, and the cells were incubated with shaking at 37 °C until they reached an OD600 of 0.5. This culture was then stimulated for 4 hours by adding 0.5 mM IPTG at 37 °C. After induction, the cells were gathered by centrifugation at 9000 rpm, followed by a wash in lysis buffer. Cell lysis was achieved through sonication with particulars as amplitude 55%, cycle -on for 1 second and off for 2 seconds and total duration of 1 minute in the same buffer, and the lysate was then centrifuged at 9500 rpm for 1 hour 30 minutes at 4 °C. The resulting supernatant was combined with pre-cleaned, stripped, and recharged Ni-NTA beads and gently agitated for 3-4 hours at 4 °C. This mixture was then applied to an affinity chromatography column and sequentially washed with different buffers: a lysis buffer (4-5 column volumes), a low salt buffer (3-4 column volumes), and a high salt buffer (1-2 column volumes). The target protein was finally eluted using lysis buffer with 300 mM imidazole and preserved in storage buffer (100 mM Nacl 50 mM, Tris-HCl, pH 8.0 10 mM DTT, and 50% glycerol) at -20°C for later use.

3.3.2. Polymerase chain Reaction amplification

The template DNAs, as specified in the appendix were subjected to amplification using gene- specific forward and reverse primers, utilizing Taq DNA Polymerase. The master mix, composed of these primers, 4.25 mM $MgCl_2$, 0.33 mM dNTPs, template DNA, and Taq DNA Polymerase, was distributed into individual PCR tubes. The thermocycler was programmed for a PCR process that began with an initial denaturation at 95 °C, followed by 30 cycles each consisting of a denaturation phase at 92 °C 30 seconds, an annealing phase at a temperature suited to the primers for 30 seconds, and an extension phase at 72 °C for a duration determined by the length of the target gene. The final extension phase was set for 10 minutes at 72 °C. After completion of the PCR process, the resulting products were analyzed using agarose gel electrophoresis and subsequently utilized for in-vitro transcription reactions.

3.3.3. *In-vitro* Transcription Reaction

For all templates, 25 µL transcription reaction was prepared at room temperature sequentially adding: (i) 2.5 µL of 10× transcription buffer, (ii) 2 µL of 10× rTNP mix, (iii) 12.75 µL of template DNA, and (iv) 7.75 µL of T7 RNA Polymerase. This mixture was then incubated for 3-4 hours at 37°C. The transcription reaction was analyzed using a 12% denaturing urea-PAGE gel. Post- electrophoresis, RNA bands were identified by UV shadowing against a TLC plate. The bands corresponding to full-length transcripts were excised using a scalpel. These gel fragments were then placed into tubes with 300 mM NaCl and agitated overnight at 4 °C. Following this, the tubes were centrifuged, and the liquid phase was transferred to a new tube and subjected to two washes with equal volume of 1-butanol. The RNA was then precipitated by adding absolute chilled ethanol.

RNA was allowed to incubate at 80 °C for 2-3 hours and centrifuged at

15600 rpm for 30 minutes at 4 °C. After the removal of excess alcohol through rota-evaporation the RNA precipitate was re- dissolved in MQ water. This was followed by the desalting of RNA using PD10 size exclusion column. The elutions collected were lyophilized to acquire high-purity RNA.

3.4 Fluorescence Intercalator displacement Assay

In the realm of nucleic acid binding ligand identification, the Fluorescence Indicator Displacement (FID) assay emerges as a proficient methodology. This assay employs Thiazole Orange (TO) as a fluorescent indicator. The assay was systematically executed with an RNA sequence of interest screened against a set of small molecules, all of which have received FDA approval, obtained after in-silico studies using Schrodinger. The binding of TO to RNA induces a fluorescence signal, quantifiable via a microplate reader. The excitation and emission wavelengths for the assay were meticulously set at 480nm and 530nm, respectively.

Control measurements were integral to the assay's robustness. These included the fluorescence intensities of the buffer alone, the buffer in conjunction with TO, and a combination of TO and RNA. To enhance the accuracy of the results, all measurements were conducted in duplicate. After the initial fluorescence measurements, various small molecules were introduced into the wells. This addition was followed by a ten-minute incubation period, post which, fluorescence readings were again recorded to evaluate any changes.

For data representation, the fluorescence intensity of the RNA-TO control was designated as the 100% reference point. Comparative fluorescence intensities of other samples were then calculated as percentages relative to this control. Molecules that displaced TO and bound to the RNA, indicated by a decrease in fluorescence, were identified as potential lead compounds **(Figure 3.1).** The 5'CAG/3'CUG RNA served as a specific control; any molecules causing a decrease in

fluorescence with this RNA were excluded from subsequent binding studies. This meticulous approach ensures a targeted selection of small molecules for further exploration in binding studies.

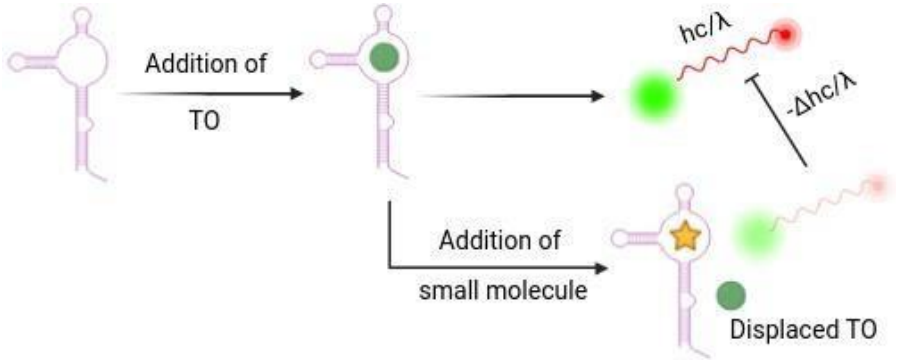


Figure 3.1. Fluorescent indicator displacement assay to detect RNA: small molecule interactions

3.5 Fluorescence Binding Assay

In this study, we examined the fluorescence intensity alterations in FDA-approved small molecules following their interaction with RNA. Fluorescence, a phenomenon wherein an electron moves from a higher to a lower energy state after being excited by high-energy electromagnetic radiation, results in visible light emission. The energy emitted during this process is contingent on the electron's chemical surroundings, and the duration for the electron's excitation and subsequent return to a lower energy state is approximately 10^{-8} seconds.

Due to the presence of transient states, an unbound fluorophore gives very low fluorescence, but a bound fluorophore with a macromolecule, adopting a conformational change generates very high fluorescence or adopts a photo quenched state. The fluorescence binding assays were conducted at room temperature using Corning half-area black 96-well plates.

The excitation and emission wavelengths for these molecules were predetermined. To ensure consistency, each well in the assay contained an equal concentration of the small molecule. The experiment was replicated for accuracy. Initially, the ligand was added to all wells, followed by a serial dilution of the RNA sample up to the 11th well. The 12th well, devoid of RNA, served as a control. Readings were taken using a Synergy H1 multi-mode microplate reader post-assay. These assays involved lead molecules identified through the FID. Data analysis was conducted using Sigma Plot 12.0 software, employing a binding equation for data fitting: both one mode and two mode data fits were used.

$$df = \frac{Bmax * abs\,(x)}{kd1 * abs(x)} + \frac{Bmax2 * abs\,(x)}{kd2 * abs(x)} \qquad \textbf{(1)}$$

Here, B_{max} represents the maximum number of binding sites, and KD denotes the equilibrium dissociation constant.

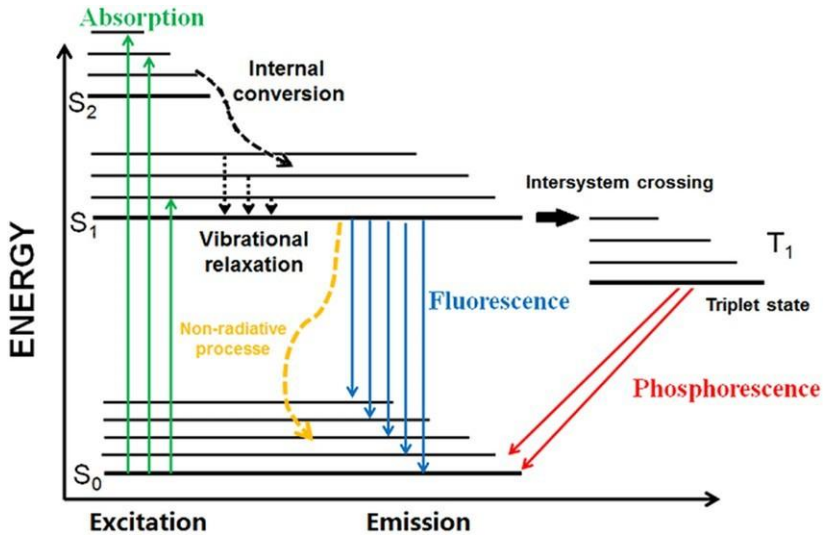


Figure 3.2. Jablonski diagram depicting fluorescence phenomenon where the molecule is excited to a higher energy state followed by the release of energy (Schweizer et al. 2021)

3.6 Circular Dichroism Spectroscopy

Circular Dichroism (CD) spectroscopy is a highly esteemed analytical technique extensively utilized for discerning the differential absorption of left and right circularly polarized light, thereby facilitating the assessment of the global structure and conformational dynamics of biological

macromolecules in varied chemical contexts. This methodology is particularly adept at detecting distinct conformational states of nucleic acids, attributable to the asymmetrical configuration of sugars in their backbone and the helical organization of nitrogenous bases. This differential absorption is quantified as CD,

$$CD = A^{l} - A^{r} \qquad \textbf{(2)}$$

Where A^{r}, A^{l} represents the absorptions of right and left circularly polarized light, respectively. The core apparatus of a CD spectropolarimeter includes components for generating left and right circularly polarized light (LCP and RCP), and a photomultiplier (PMT) for detection. Biological macromolecules exhibit characteristic CD signals, termed signature CD spectra, which are unique to their specific conformational states. This signature is particularly evident in nucleic acid samples, where the CD signal arises from glycosidic linkages between sugar units and nitrogenous bases, leading to chiral perturbations.

Notably, the CD spectroscopy of nucleic acids provides insights into their secondary structure conformations and their interactive patterns with ligands or proteins. The positioning and intensity of the CD spectral peaks vary with the topology and are also influenced by the DNA/RNA concentration in the sample.

In this study, CD titration experiments were conducted on *in-vitro* synthesized RNA samples, incorporating incremental additions of a lead small molecule at a controlled temperature of 25°C and a scanning rate of 20 nm/min, ranging from 190 to 300 nm. These experiments utilized a

Jasco J185 Spectropolarimeter, with spectra recorded in a cuvette of 1 mm path length. RNA samples were maintained at constant concentrations, while small molecule concentrations were varied. To ensure accuracy, a blank spectrum of the buffer (1X KPO_4, KCl, MQ water) was recorded before each measurement and subtracted from the sample's CD spectrum. Data analysis was performed using SigmaPlot 13.0, allowing for a detailed interpretation of ligand interactions and their stabilizing or destabilizing effects on loop structures.

Thermal profile assay was also performed using CD spectroscopy, in which the Thermal denaturation experiments were conducted using a Perkin Elmer Lambda 35 spectrophotometer equipped with a Peltier temperature programmer (PTP 6+ 6) and a water Peltier system PCB-1500. RNA samples were gradually heated from 25°C to 95°C, while monitoring the absorbance at 260 nm at a rate of 1°C per minute. These experiments were performed both in the absence and presence of compounds of interest at a varying D/N ratio. The resulting data on normalized absorbance changes at 260 nm versus temperature were analyzed and plotted using SigmaPlot 12.0 software.

3.7 Electrophoretic Mobility Shift Assay

Electrophoretic Mobility Shift Assay (EMSA) represents a rapid and highly sensitive method employed for discerning the size and molecularity of biological macromolecules. This assay is notably utilized in the examination of interactions between such macromolecules. It operates on the principle that the electrophoretic mobility of a molecule in a native polyacrylamide or agarose gel is influenced by its charge, shape, and size. Consequently, EMSA finds extensive application in the study of ligand-nucleic acid interactions. The interaction between a ligand and nucleic acid often results in an alteration of the nucleic acid's molecular weight, thereby inducing a shift in the mobility of the corresponding band on the gel.

In the current research, EMSA was conducted using a 3% agarose gel in 1X TBE buffer. The RNA samples under study were dissolved in a 1X phosphate buffer containing 50 mM KCl. These samples were then treated with varying concentrations of the lead drug molecule and allowed to equilibrate for 30 minutes at a temperature of 25°C.

The electrophoresis was performed in a Bio-Rad Mini-Sub Cell GT Electrophoresis Cell. Each sample, amounting to 20 µl, was loaded into the cell, and electrophoresis was conducted at room temperature. The voltage was maintained at 72 V until the dye front traversed three-fourths of the gel length. Post-electrophoresis, the gel was extracted from its cassette and stained using Ethidium Bromide solution. The results of the EMSA were documented and analyzed utilizing the ImageQuant LAS4000, a sophisticated imaging system provided by GE Healthcare Biosciences Ltd, Sweden.

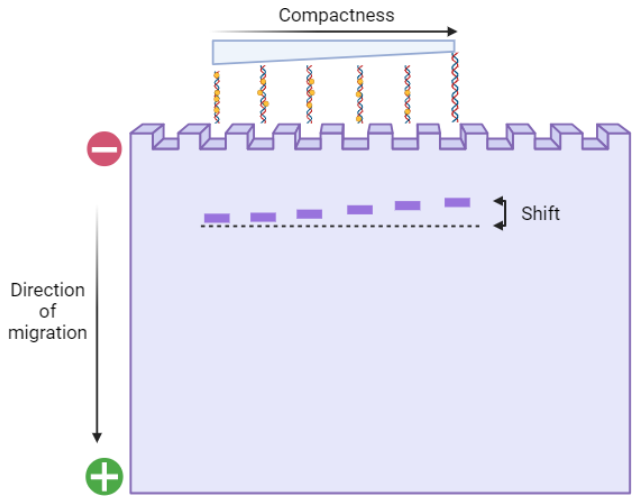


Figure 3.3. Schematic representation of Electrophoretic shift mobility

3.8 PCR Stop Assay

The template and complementary sequence used in this study is mentioned in appendix table S1 and table S2 at number 12 and 1 respectively and were procured from Sigma-Aldrich Chemicals Ltd. Based in USA. Master mix for the PCR stop assay contained 1X PCR reaction buffer, 8 mM $MgCl_2$, 3 µM oligonucleotides, 1 mM dNTPs, and 2.5 units of Taq polymerase and serial dilutions of small molecule. The PCR cycle was as follows: 95 ºC for 3 min, followed by 30 cycles of 95 ºC for 30 sec, 70 ºC for 30 sec and 72 ºC for 1-minute, final extension at 72 ºC for 10 minute and finally held at 4 ºC for infinity. The products were mixed with 6X loading dye and resolved on 2.5 % acrylamide resolving gel without any added cations. The Gel image was captured using Image Quant LAS 4000 and analyzed using ImageJ.

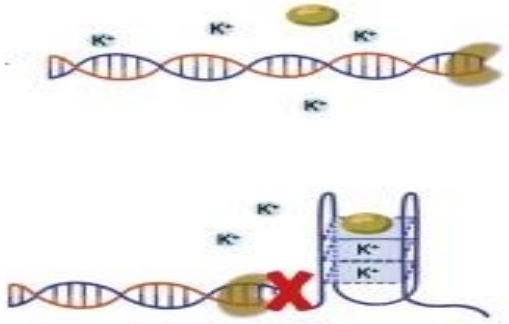


Figure 3.4. Schematic representation of PCR Stop assay (Jamroskovic et al. 2019)

3.9 Isothermal Titration Calorimetry assay

Isothermal titration calorimetry (ITC) is a technique used to quantify the thermodynamics of molecular interactions by measuring the heat changes associated with binding events. It involves titrating one reactant into a solution containing the other reactant while maintaining a constant temperature throughout the experiment. This assay provides valuable insights into various aspects of molecular interactions, including binding affinity, stoichiometry, enthalpy, and entropy changes. By precisely

measuring the heat released or absorbed during binding, ITC enables researchers to determine the strength and specificity of interactions between biomolecules, such as proteins, nucleic acids, and small molecules. The information obtained from ITC experiments is crucial for understanding fundamental biological processes, elucidating the mechanisms of disease, and facilitating drug discovery efforts. For example, ITC can be used to study protein-ligand interactions, receptor-ligand binding, enzyme-substrate interactions, DNA-protein interactions, RNA-small molecule interactions.

One of the key advantages of ITC is its ability to provide quantitative data without requiring any labeling or modification of the molecules under investigation. This allows for the direct measurement of binding parameters under physiologically relevant conditions, making ITC an invaluable tool in both academic research and pharmaceutical development. The measured heat helps in the accurate determination of several thermodynamic parameters such as association constant, reaction stoichiometry, enthalpy, and entropy using the following equation.

$$\Delta G = -RTlnKa = \Delta H - T\Delta S \qquad \textbf{(3)}$$

where R is the gas constant and T is the temperature.

The setup of an isothermal titration calorimeter (ITC) encompasses several critical components within a temperature-regulated environment. At its core are the sample and reference cells, with the former housing one reactant and the latter serving as a baseline. An injection system meticulously dispenses controlled volumes of the second reactant into the sample cell, ensuring precision and accuracy. Stirring mechanisms maintain homogeneity within the solution, vital for consistent results. Temperature control mechanisms sustain a stable environment, preventing fluctuations that could compromise data integrity. An integrated data acquisition system continuously monitors heat changes within the sample cell, capturing them over time. This comprehensive setup enables the precise measurement of heat changes

associated with molecular interactions, facilitating the analysis of binding kinetics, stoichiometry, and thermodynamic parameters. Such insights are invaluable across diverse scientific domains, from elucidating fundamental biological processes to advancing drug discovery endeavors.

In academic research, ITC is used to investigate molecular recognition events involved in biological processes, such as signal transduction, gene regulation, and metabolic pathways. It helps researchers understand the underlying principles governing molecular recognition and provides insights into the structure-function relationships of biomolecules.

In this study, ITC experiments were conducted using a MicroCal iTC200 isothermal titration calorimeter (GE Healthcare) at 25 °C. Data acquisition and analysis were performed using Origin scientific software version 7 (Microcal Software Inc.). RNA samples were dissolved in 10 mM potassium phosphate buffer (pH=7.2) containing 0.1 M NaCl and 50 mM EDTA. The ITC cell was loaded with 7 μM RNA, while the syringe contained 50 μM Myricetin in the same buffer. A total of 32 injections (1.6 μL each) of Myricetin were sequentially added to the ITC cell, with an initial injection volume of 0.4 μL and a 60-second initial equilibrium delay. The reference power was set to 8 μcal/s, and a 120-second pause was maintained between injections. Stirring was maintained at a speed of 750 rpm throughout the experiment. The thermal titration data were fitted to a two-site binding model to determine the dissociation constant (K_d).

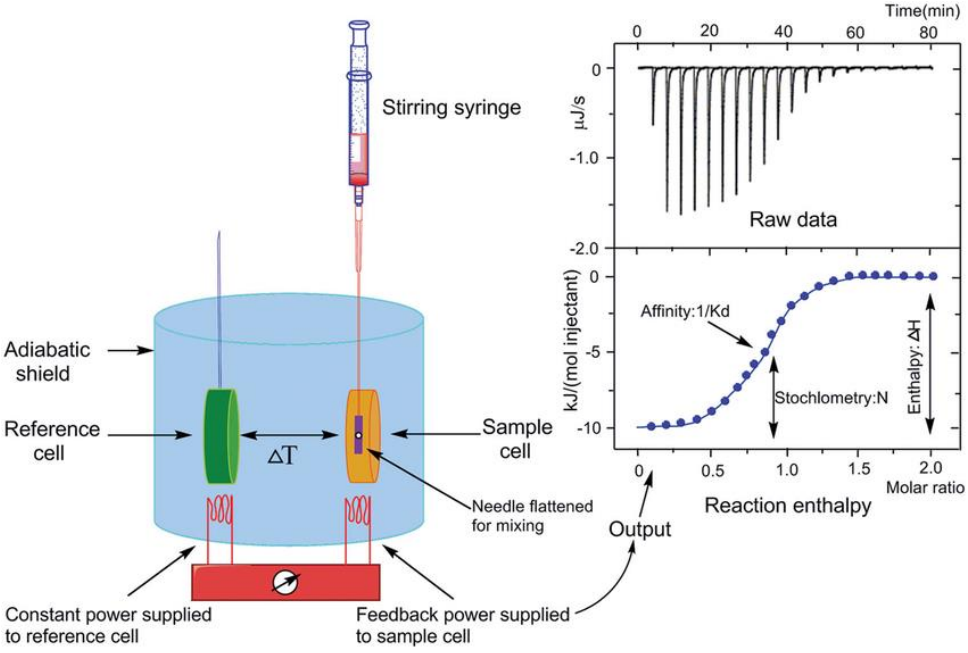


Figure 3.5. Schematic representation of the heat calorimeter in ITC instrument along with example of binding isotherm. (Song et al. 2015)

3.10 Nuclear Magnetic Resonance spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy stands as one of the most indispensable analytical techniques in both chemical and biological research due to its unparalleled ability to unveil molecular structures, elucidate dynamics, and probe molecular interactions at the atomic level. The core principle underlying NMR spectroscopy revolves around the magnetic properties of atomic nuclei. When subjected to an external magnetic field B^o, nuclei possessing an odd number of protons or neutrons exhibit a magnetic moment and align either parallel or anti-parallel to the field.

Upon exposure to radiofrequency (RF) pulses at a frequency matching the energy difference between nuclear spin states, nuclei undergo nuclear magnetic resonance, as depicted by the Larmor equation:

$$\omega = \gamma B \qquad \textbf{(4)}$$

Here, ω represents the angular frequency of the RF pulse, γ signifies the gyromagnetic ratio (a characteristic of the nucleus), and B denotes the strength of the magnetic field. This resonance phenomenon is profoundly influenced by the chemical environment surrounding the nucleus, leading to chemical shift, which in turn furnishes critical information about molecular structure. Moreover, interactions between adjacent nuclei with magnetic moments result in spin-spin coupling, which manifests as signal splitting in the NMR spectrum, providing insights into molecular connectivity.

The instrumental setup of an NMR spectrometer is intricate and encompasses several key components. A high-field magnet, often superconducting, is employed to generate a stable and uniform magnetic field. RF transmitter and receiver coils are utilized to administer RF pulses for excitation and to capture resulting NMR signals from the sample. The sample, typically dissolved in a solvent, is housed within a sample probe inserted into the magnet, where it experiences the homogeneous magnetic field. Sophisticated pulse sequences, orchestrated by a pulse sequence controller, manipulate nuclear spins to extract specific information about the sample. Subsequent data acquisition and processing are facilitated by the spectrometer's computer system, allowing for comprehensive analysis.

The versatility of NMR spectroscopy extends across various scientific domains, underpinning a myriad of applications. In chemical research, NMR is instrumental in elucidating the structures of organic compounds,

characterizing complex mixtures, and studying reaction kinetics. In the realm of biology, it serves as a vital tool for probing protein structures and dynamics, unraveling ligand-receptor interactions, and unraveling metabolic pathways in metabolomics studies. Additionally, in materials science, NMR spectroscopy aids in characterizing polymers, catalysts, nanoparticles, and other materials, offering valuable insights into their composition, structure, and properties. With its non-destructive nature, high resolution, and ability to provide detailed structural information, NMR spectroscopy remains an indispensable cornerstone in scientific research and development across a broad spectrum of disciplines.

3.11 In-silico Molecular docking

In this study, the three-dimensional structure of CAG duplex RNA (Protein Data Bank ID: 4J50) served as the initial model for receptor preparation. Structure preparation for docking analysis was carried out using Auto Dock tools. Following receptor preparation, ligand 3D structure was downloaded from the PubChem database in sdf format (PubChem CID ID: 65028). Further ligand preparation was meticulously executed using Discovery Studio 3.5 (San Diego, Dassault Systèmes, USA), a sophisticated molecular modeling and simulation platform.

After receptor and ligand preparation, docking studies were conducted utilizing Auto Dock Vina. (The Scripps Research Institute, La Jolla, CA, USA). This software is distinguished for its capability to treat RNA as a rigid body in docking simulations. Such an approach is imperative for understanding the interaction dynamics between RNA and ligand molecules. In preparation for docking, the ligand was carefully optimized to ensure its readiness for interaction studies.

A crucial aspect of the docking procedure involved the creation of a grid box encompassing the entire RNA structure.

This strategic placement allowed for the comprehensive exploration of the ligand within the complete conformational space of the RNA. Doing so achieved a more accurate and holistic representation of potential ligand-RNA interactions.

The outcomes of the docking studies were primarily analyzed based on the binding affinities of the ligand to the duplex RNA. These binding affinities indicate the strength and stability of the ligand-RNA interactions, providing valuable insights into the molecular mechanisms at play. The analysis of these interactions is instrumental in elucidating the structural and functional dynamics of the RNA-ligand complexes, offering significant contributions to RNA biology and drug design.

3.12 Molecular dynamic simulation studies

Stability of the generated RNA structure and the docked complexes were further evaluated by an all-atom MD simulation using GROMACS 2023.3. The topology details for RNA were acquired using the CHARMM27 all-atom force field. The complexes were solvated by a TIP3P water model in a cubic box. They were then neutralized by adding a suitable amount of counter ions, with the concentration of KCl set to 0.15 M. To avoid atomic clashes in the system, the system's energy was minimized using the steepest descent method and the maximum forces were converged below 1000 kJ/mol/nm. Electrostatics were treated with Particle Mesh Ewald, and the cutoff for both Coulomb and van der Waals interactions was 1.0 nm. Further NVT, and NPT ensembles run for 1 ns each was carried out at temperature of 298.15 K and pressure of 1 bar respectively. For these cases, the modified Berendsen thermostat was used to regulate temperature and a Parinello-Rahman barostat was used to regulate pressure equilibrations. Finally, 100 ns production run was executed. A timestep of 2 fs was used and the frames were updated every 5000 steps. RMSD of the RNA alone and RNA-ligand complex were calculated after square fitting to the system.

3.13 Binding free energy calculation

The binding free energy calculations in the dynamic state were calculated using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method and the gmx_MMPBSA tool. A total of 9999 frames of the 100-ns simulation were subjected to the calculations. The calculations were carried out at a temperature of 298.15 K and the PB radii was 4. In this method, all individual energy components are calculated, including polar and non-polar solvation energy, the Van der Waals forces, and the electrostatic contacts of the complex as well as the receptor and ligand individually. The final free-binding energy for a complex can then be estimated by the following:

$$\Delta G_{bind} = \langle G_{COM} \rangle - \langle G_{REC} \rangle - \langle G_{LIG} \rangle \qquad \textbf{(5)}$$

$\Delta Gbind$ can also be represented as $\Delta G_{bind} = \Delta H - T\Delta S$. where ΔH corresponds to the enthalpy of binding and $-T\Delta S$ to the conformational entropy after ligand binding. When the entropic term is dismissed, the computed value is the effective free energy, which is usually sufficient for comparing the relative binding free energies of related ligands.

3.14 MTT Assay

To evaluate the toxicity of small compounds on cell lines, a test for cell viability is employed. A cell's metabolic activity begins to diminish several times as it gets closer to death, which implies that the cell's ability for redox reactions decreases significantly at that point. MTT is a water-soluble reporter chemical with a pale-yellow tint that measures a cell's metabolic activity based on its redox potential. MTT is reduced by the mitochondrial reductase in living cells to an insoluble purple formazon moiety. Hence, a purple color signal in the MTT assay indicates active or living cells. We have tested the small molecule toxicity in this study. After being exposed to small molecules for 24

hours in 96-well plates, the cells were treated with MTT for 4 hours, and the absorbance at 540 nm was used to measure the toxicity of the small molecules.

3.15 Cell culture and protein aggregates visualization

COS-7 cell lines were cultured at 37 degrees Celsius in a 5% CO_2 environment using DMEM medium supplemented with 10% heat-inactivated FBS and antibiotics. Cells were cultured on chamber slides as well as 6-well and 96-well tissue culture plates. At a 60% confluence stage, transient transfections were performed using plasmids (EGFP – HDQ74) and (EGFP – HDQ23). Cells were then treated with DMSO as a control and different concentrations of lead small molecule for 24 hours each. After washing cells with PBS, they were fixed with 4% paraformaldehyde for 15 minutes. Fluorescence microscopy was utilized to observe polyQ protein aggregates. Confocal microscopy was utilized to visualize protein aggregation at a higher magnification. 100 positive transfected cell from each well were selected and manually graded as "with protein aggregates" and "no aggregates". The number of GFP protein inclusion was manually counted from three independent experiment and standard deviations were calculated.

3.16 Western blot analysis

COS-7 cells cultured in a 6-well plate were transfected with EGFP-HDQ74 and EGFP-HDQ23 plasmids and exposed to lead compound for 24 hours. Cell lysates were gathered using RIPA lysis buffer and separated using SDS page at 80 V, then transferred onto a nitrocellulose membrane in transfer buffer containing methanol for 80 minutes at 4 °C. Blots were blocked for 1 hour in 5% skimmed milk in TBST, then probed overnight at 4°C with primary antibodies at a 1:1000 dilution in TBST. Horseradish peroxidase-conjugated secondary antibodies were used at a 1:1000 dilution prepared in TBST. The Luminata Crescendo Western HRP substrate was utilized for blot detection. Analysis was

performed using ImageJ.

3.17 Drosophila fly stocks, food, and their maintenance

Flies were reared on standard cornmeal-agar (HiMedia Drosophila Diet) supplemented with antibiotic and propionic acid, with a 12-hour on–off light cycle at 25 °C.

The following fly stocks have been used:

1) UAS-Httex1Q20 (non-pathogenic): Expresses human HTT exon 1 with 20 PolyQ repeats under the control of UAS.

2) UAS-Httex1Q93 (pathogenic): Expresses human HTT exon 1 with 93 PolyQ repeats under the control of UAS.

3) GMR-Gal4: A Gal-4 stock that expresses GMR (Glass Multiple Reporter) in all cell's posterior to the morphogenetic furrow and later it becomes active throughout most of the pupal eye.

4) Elav-Gal4c155 driver line: The "Elav-Gal4" driver line in Drosophila melanogaster drives expression of genes specifically in neurons when combined with Gal4-responsive promoters.

Males of all the transgenic lines used in the study were crossed with virgins of elav-Gal4 and GMR-Gal4 to drive their expression in the fly brain and the eye-imaginal discs, respectively. The flies were maintained on a typical cornmeal-agar diet at a temperature of 25±1°C with consistent light-dark cycles and humidity ranging from 70% to 80%. Drug of choice was added to the fly food at concentrations of 10 μM, 25 μM, and 50 μM to rear experimental larvae or flies. Control groups were raised on regular food without the drug at a temperature of 25°C.

3.18 Rough eye phenotype assessment

The rough eye phenotype of flies expressing Httex1p Q93 under the control of gmr-GAL4 was evaluated. These flies were cultured at 25 °C on standard food and food supplemented with either 5, 10 and 25 μM drug compound. The outer surface of the fly eye was scrutinized using a Nikon SMZ745T microscope on days 1, 7, and 14 after emergence. A total of 30 flies were examined for each condition.

3.19 Negative geotaxis

To evaluate the locomotor capabilities of flies, a negative geotaxis assay was conducted. Each condition was tested with two groups of 10 flies each, across a total of six trials conducted in the designated tube. Flies were introduced into a vertical column measuring 25 cm in length and 1.5 cm in diameter. After gently tapping the column three times, flies that reached the top and those remaining at the bottom were counted separately. The outcome was quantified as the number of flies surpassing a specified distance within a 20-second interval.

3.20 Crawling Assay

To measure the crawling ability, a petri dish measuring 100 mm × 10 mm and made of 3.3% agar served as the apparatus. Within this dish, a track measuring 2 mm wide, 30 mm long, and 5 mm deep was formed. The locomotion of each third-instar larva placed in the track was observed for 20 seconds, with 10 larvae per condition monitored across two trials.

3.21 Scanning Electron Microscopy Images

Critical Point Dried (CPD) method was used to prepare samples for attaining high-magnification analysis through Scanning electron microscopy as per Kimmel at al. (1990). First, whole flies were immersed in a fixative composed of (1% glutaraldehyde, 1% formaldehyde and 1M sodium cacodylate (pH 7.2)) for 2 hours. Further, to submerse the flies in the fixative, single drops of 0.2% Tween 20

(diluted in H_20) was used. This was followed by rinsing in H_20. Then, the flies were dehydrated in an Ethanol series (once in 25%, 50% and 75% ethanol and twice in 100% ethanol), each for 12 hours, at room temperature. The flies were then fixed, and images are taken followed by CPD and sputter-coat.

3.22 Dead cell detection

Acridine Orange (AO), a vital dye is used for the identification of the apoptotic cells. Living cells being selectively permeable, exclude the dye. Dead cells whereas, lose their permeability barrier and the dye molecules easily permeate and intercalate between the base pairs of dsDNA, causing it to fluoresce green.

Here, the Acridine Orange (AO) was used to observe the presence of apoptotic cells in the eye imaginal discs of the third-instar larvae. Larval eye imaginal discs were first dissected in 1X PBS. The tissues were then incubated in 5 μg/ml acridine orange solution (Sigma Cat. No. A6014) solution for 2 minutes, followed by washing in 1X PBS and were then mounted in 1X PBS. The tissues were then immediately photographed with a fluorescent microscope with a minimum of 5 discs captured for each condition.

Chapter 4
Results and discussion

A small molecule modulator that targets RNA has been promoted as a potential therapeutic approach. Thus, we used several biophysical experiments to determine the specificity and selectivity of several FDA-approved small molecules against r (CAG)exp to alleviate the toxic effects of Poly-Q protein.

4.1 Fluorescence Intercalator Displacement assay used as primary screening of FDA molecules

4.1.1. FID – RNA 5'CAG/3'CAG against 33 compounds

Fluorescence Intercalator Displacement assay has been found viable in discerning the interaction of a molecule with nucleic acid by displacing a fluorophore through the change in the fluorescence intensity.

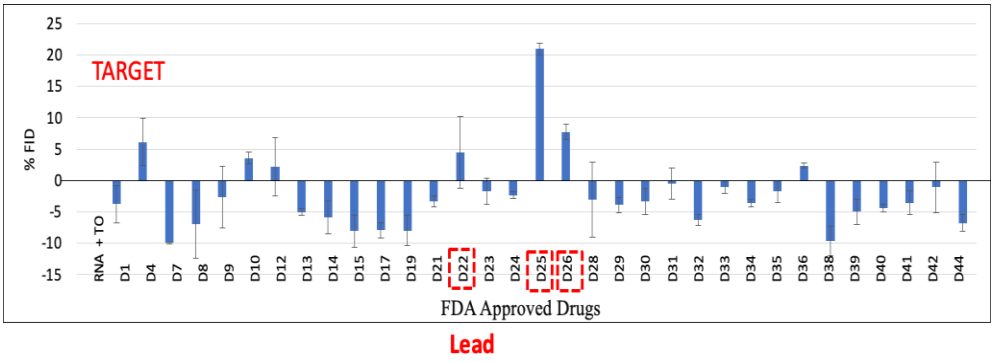


Figure 4.1.1. Represent %change in fluorescence of Thiazole orange for RNA 5'CAG/3'CAG after incubation with FDA-approved small molecules

The data represent a change in the fluorescence intensity for RNA containing A-A mismatch, 7 drugs showcased a decrease in the fluorescence intensity. D22, D25, D26 were selected for further experimentation. D25

showed maximum change in the intensity.

4.1.2. FID – RNA 5’CAG/3’CUG against 33 compounds:

RNA 5’CAG/3’CUG is used as a control RNA because it does not contain any mismatch because of A-U canonical matching, and thus, change in the fluorescence intensity after incubation of FDA-approved small molecules would indicate binding with control and thus would not be considered for further experimentation.

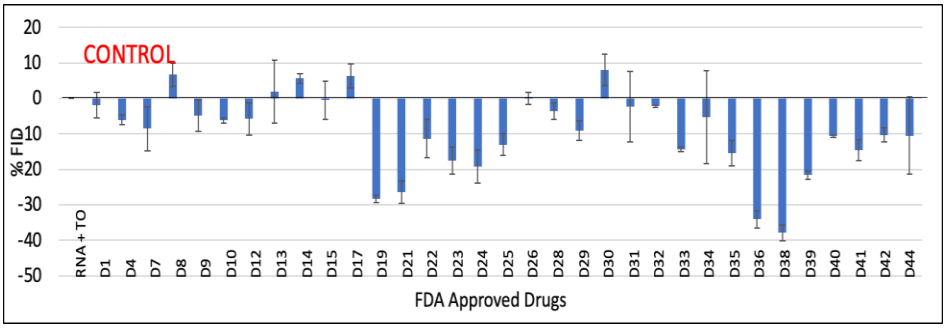


Figure 4.1.2. Represent %change in fluorescence of Thiazole orange for RNA 5’CAG/3’CUG after incubation with FDA-approved small molecules

Molecules which show decrease in fluorescence such as D8, D30 will not be used for further experimentation, however our molecules of interest including D22, D25, D26 does not indicate any decrease in the fluorescence intensity and thus can be used for further experimentation.

4.2 Fluorescence Binding Assay:

Fluorescence Binding Assay is used as primary screening tool along with the FID, which helps to determine the specificity of the molecule for the RNA motif. It has been reported that planar molecules bind to the RNA motifs effectively. In this study, Fluorescence binding was performed with lead molecules from FID namely D22, D25, D26 with RNA library containing 1x1 nucleotide mismatch loop like A-A mismatch loop. Primarily FBA was performed with 5’CAG/3’CAG and 5’CAG/3’CUG

to determine the specificity of the molecule for AA mismatch. Molecules with similar or more affinity for the duplex RNA were not carried forward for further experimentation elucidating the conformational changes of the RNA motif after binding with the molecule. Compounds with high affinity were proceeded for further screening with r $(CAG)_6$ RNA loop.

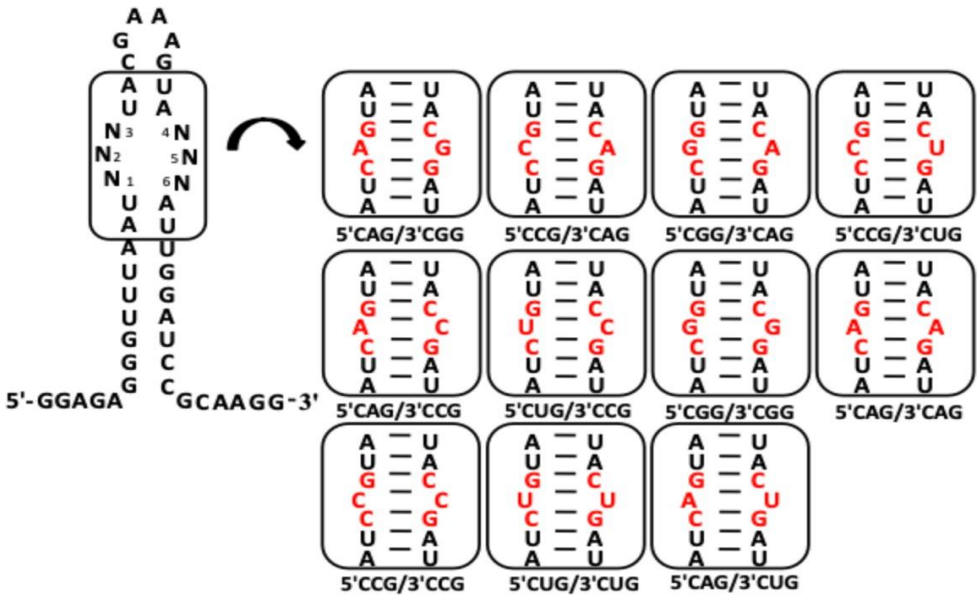
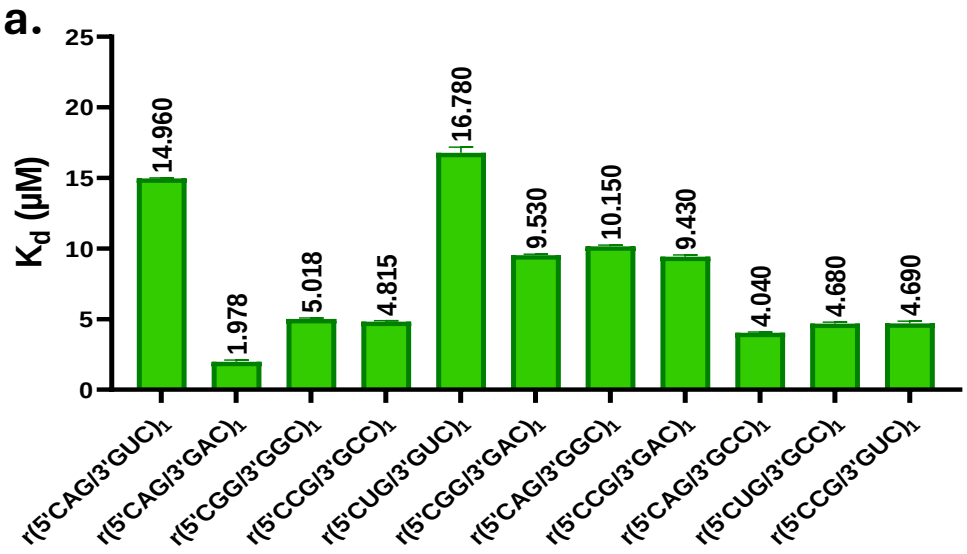
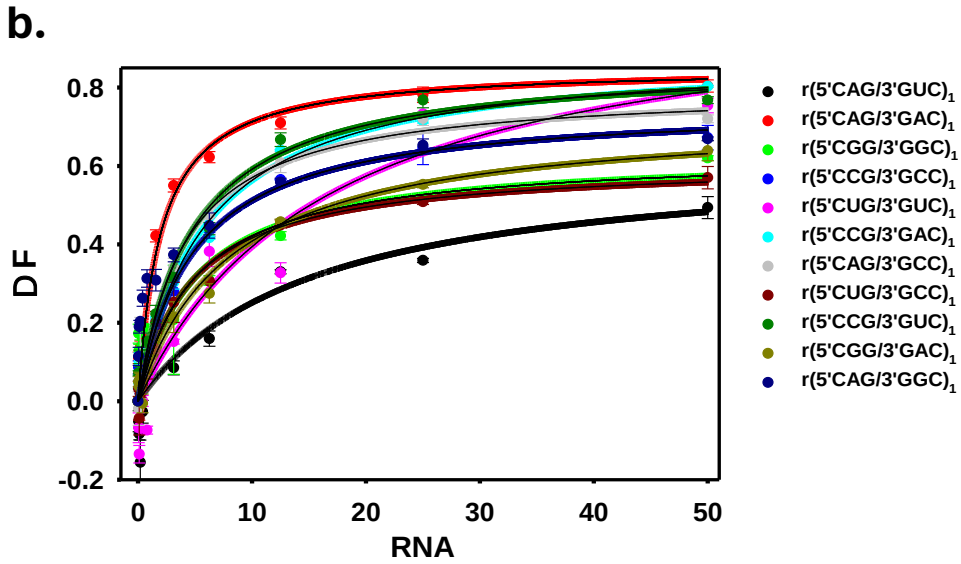


Figure 4.2. Secondary structures of different RNA sequences with internal loops comprise one nucleotide each. DNA molecule binding is studied using the secondary structure of RNA sequences. Below their secondary structure is their nomenclature.
(Khan et al. 2018)

4.2.1. Fluorescence Binding assay of D25 with 11 loop mismatch RNA library, r(CAG)^{exp} RNA and control:

Prior to conducting experiments with D25, the drug's excitation and emission profiles were characterized using UV-visible and fluorescence spectrophotometry. These profiles were essential for determining the optimal wavelengths to excite the drug and to monitor any changes in its fluorescence upon interacting with the RNA motif.





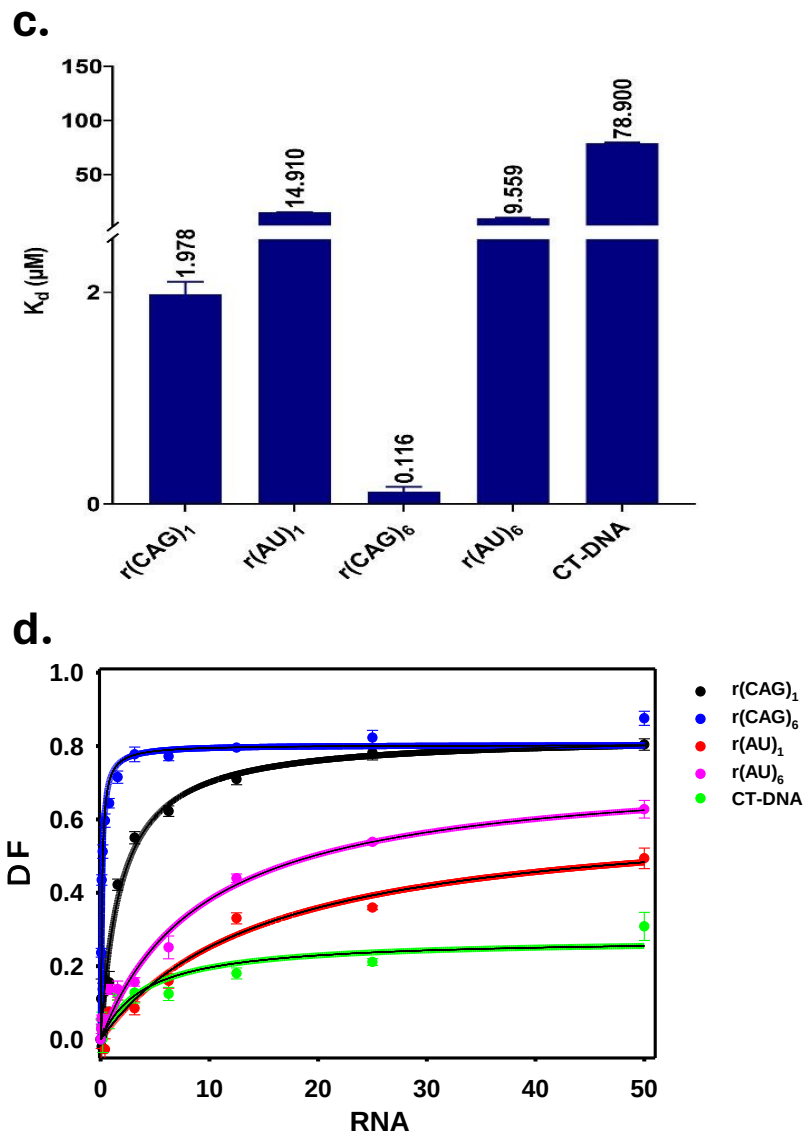


Figure 4.2.1. Evaluation of binding affinity of D25 with nucleic acids. (a) Bar graph denotes K_d values of r(CNG)×1 RNA motifs with D25. (b) The plot represents fluorescence titration assay two-mode curve fitting of r(CNG)×1 RNA motifs with D25. (c) Bar graph represents K_d value analysis of different RNA and DNA controls with D25. (d) Plot depicts fluorescence titration assay two-mode curve fitting of r(CAG)^{exp} RNAs and different RNA and DNA controls with D25.

Thus, it is evident from all titration curves of D25, including AA mismatch, control duplex RNA, internal loop library containing all different combinations of mismatch, Calf thymus DNA, and r $(CAG)_6$ mismatch RNA, that it can be considered as a lead molecule for further experimentation. Additionally, D25 displayed selectivity for CAG RNA over other fully base-paired AU×6 RNA exhibiting approximately 86-fold higher affinity for the r$(CAG)_6$ repeat RNA. The difference between AA mismatch loop K_d and control RNA K_d is significant with p-value of less than 0.01. Moreover, with increased CAG repeats, K_d value further decreases and is significantly higher for control duplex r(AU) and CT-DNA. D25 shows interaction with CT-DNA sequence, however its affinity was approximately 700-fold lower as compared to binding affinity of r(CAGx6) RNA.

4.2.2. Fluorescence Binding assay of D22 and D26 with 11 loop mismatch RNA library

Before experimenting with D22 and D26, the drug's excitation and emission characteristics were analyzed using UV-visible and fluorescence spectrophotometry. The profiles were crucial for identifying the best wavelengths to activate the small molecule and to observe any alterations in its fluorescence while interacting with the RNA motif.

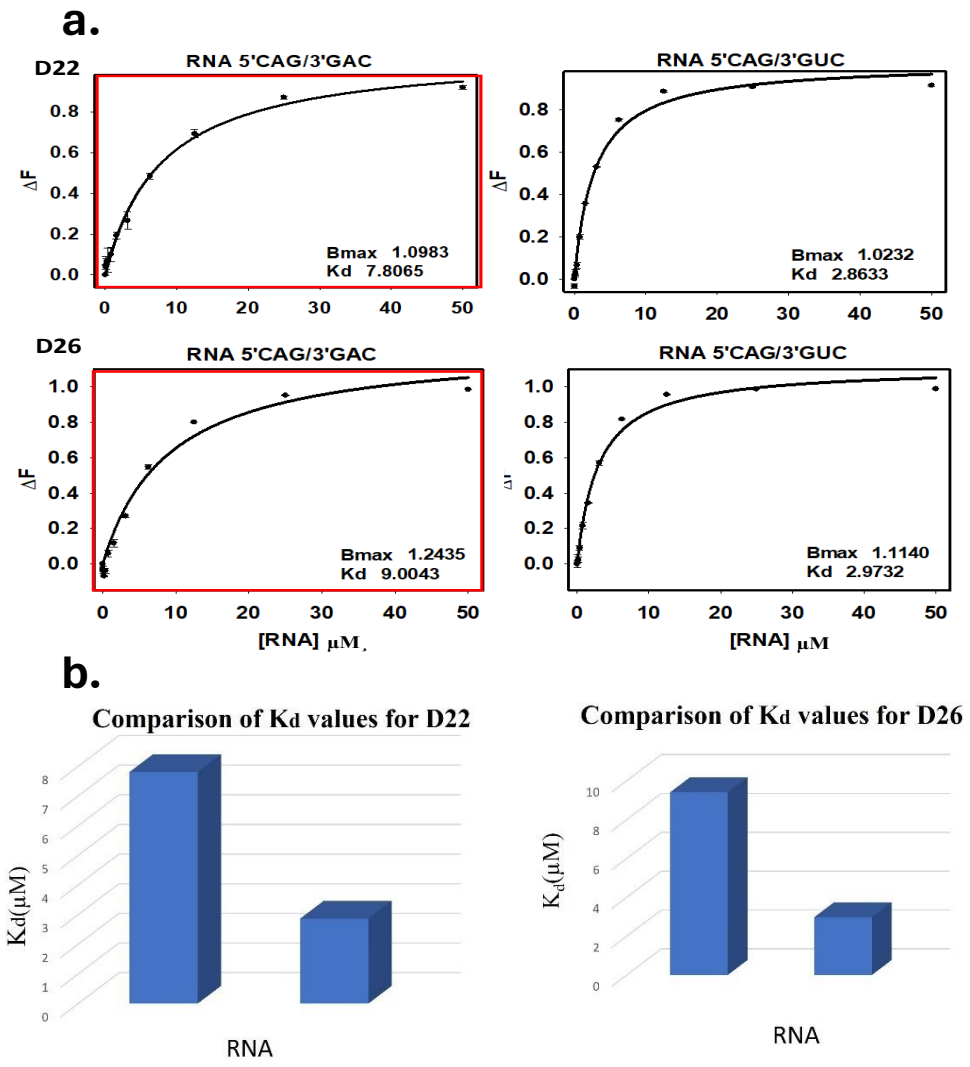


Figure 4.2.2. (a). Fluorescence Binding Assay of r(5'CAG/3'GAC), r(5'CAG/3'GUC) with D22 and D26 (b) Bar graphs represent K_d value analysis for D22 and D26.

After titration curve fitting using one mode, it was evident that D22 and D26 binding affinity with mismatch AA is not lower as compared to canonical AU paired RNA motif. Thus, neither molecule is carried forward.

4.3 Isothermal Titration calorimetry assay

Isothermal Titration Calorimetry (ITC) is a powerful analytical technique used to study the thermodynamics of molecular interactions in solution. By directly measuring the heat absorbed or released during a binding event, ITC provides detailed insights into the binding affinity, stoichiometry, enthalpy (ΔH), and entropy (ΔS) of the interaction between molecules.

The core principle of ITC involves titrating a ligand into a solution containing the target molecule, typically a protein or nucleic acid, within a precisely controlled, isothermal environment. As the ligand binds to the target, heat is either absorbed or released, and this thermal change is detected by the calorimeter. The resulting thermogram consists of a series of peaks corresponding to the heat changes associated with each injection of the ligand.

The Isothermal titration calorimetry (ITC) study is used to corroborate the binding affinity of D25 with r(CAGx6) and r(Aux6) RNA. The data were fitted using a two-site binding model for both RNAs. ITC data showed exothermic peaks due to the increased interaction of D25 with RNA. The dissociation constants obtained for the r(CAGx6) pair and r(Aux6) pair were 707 nM and 0.180 mM, respectively, which suggests a higher affinity and specificity of D25 with the r(CAGx6) motif-RNA over AU paired duplex RNA control. This ensures strong and tight binding as well as confirms the nanomolar range of binding obtained from the fluorescence binding assay with 5′CAG/3′GAC motif-RNA. The top panel shows the power versus time curve while bottom panel shows the thermogram of the integrated peak intensities, plotted against the molar ratio.

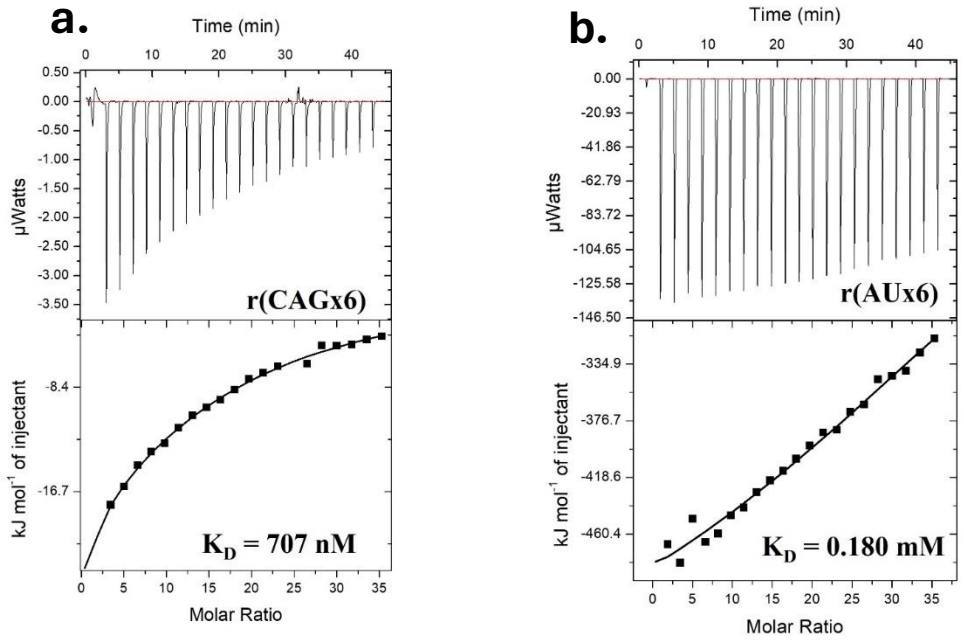


Figure 4.3. Isothermal calorimetry titrations of D25 with a) $r(CAG)_6$ and b) $r(AU)_6$ duplex RNA. It represents the titrated thermogram with D25. Solid line curve represents the two-mode binding best fit.

4.4 Electrophoretic mobility shift assay assesses the interaction of small molecules with RNA

Electrophoretic mobility shift assay or gel retardation assay is an important tool used to confirm the binding of the D25 with the mismatch RNA. This technique provides insights into the specificity and the selectivity of the molecule with the RNA motif.

$r(CAG)_1$ RNA and $r(CAG)_6$ RNA contains 1 mismatch and 6 mismatch loops, respectively. Thus, the binding of D25 must be more prominent in the later RNA motif due to more binding sites for interaction. Thus, a prominent shift must be observed for the RNA motif. $(AU)_1$ RNA and $(AU)_6$ RNA contains no mismatch loops, respectively. Thus, the binding of D25 must not produce any significant shift in the migration of the RNA.

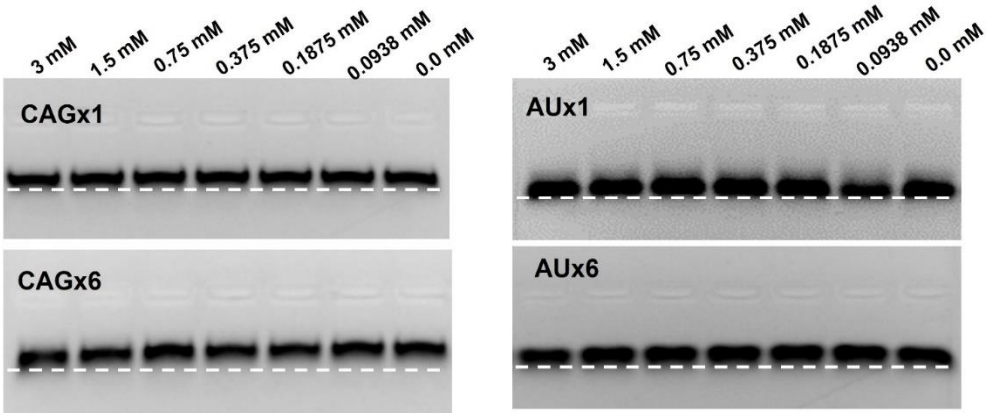


Figure 4.4.1. Represent Gel retardation assay increasing concentration of D25 significantly increases the mobility of CAG repeat RNAs over AU duplex RNA

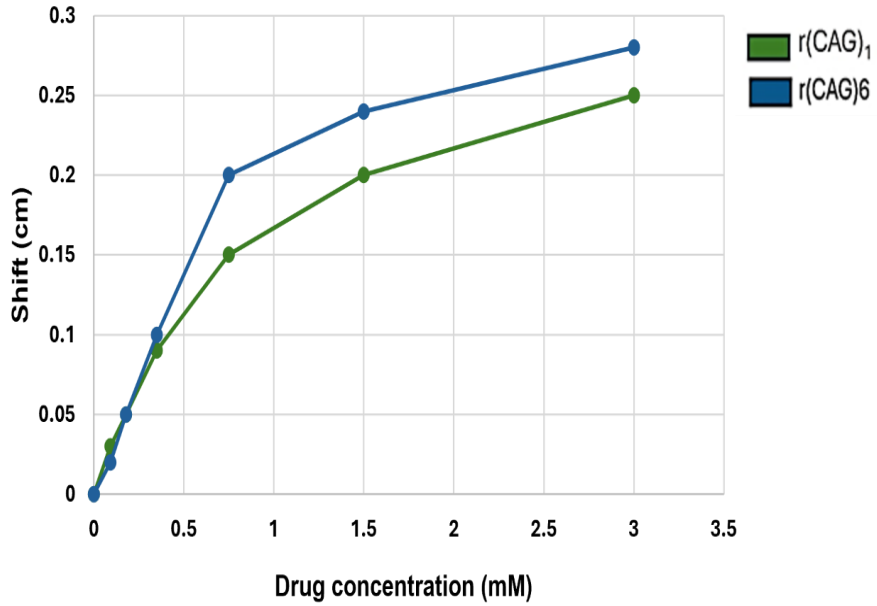


Figure 4.4.2. Graph showing increasing shift with increasing concentration of D25 for $r(CAG)_1$ and $r(CAG)_6$

The above graph shows a strong correlation between increasing concentration of D25 and increasing shift with R^2 of 0.858 fitted using linear equation, thus exhibiting strong interaction between drug and the AA mismatch containing RNA.

Agarose gel images clearly show that the migration speed of the RNA containing a six-mismatch loop increases with higher drug concentrations. This acceleration in migration suggests that the drug enhances the RNA motif's stability. This observation aligns with the findings from CD spectroscopy, which indicates that the D25 molecule stabilizes the loop's double-stranded structure. The increased stability results in a more compact macromolecule that migrates through the gel more rapidly than the unbound RNA.

Agarose gel images clearly show that the migration speed of the RNA containing a six-mismatch loop is not affected by higher drug concentrations. This suggests that the drug does not enhance the RNA motif's stability, as observed with r $(CAG)_6$. This observation aligns with the findings from CD spectroscopy, which indicates that the D25 molecule does not interact with the control duplex RNA. This signifies the specificity and selectivity along with the strong interaction of D25 with AA mismatch RNA concerning control duplex RNA.

Moreover, the shift observed with respect to canonical paired RNA is also significant with p value of $0.001614 < 0.01$, thus giving confidence of 99% that D25 interacts with AA mismatch RNA but not with canonical paired RNA.

4.5 Polymerase Chain Reaction stop assay

Polymerase chain reaction stop assay is an important tool used to confirm the binding of the D25 with the mismatch RNA. This technique provides insights into the specificity and the selectivity of the molecule with the RNA motif. Binding of small molecules with the template containing loops lead to hinderance in the activity of the *Taq* polymerase which leads to reduction in the amplification of the template.

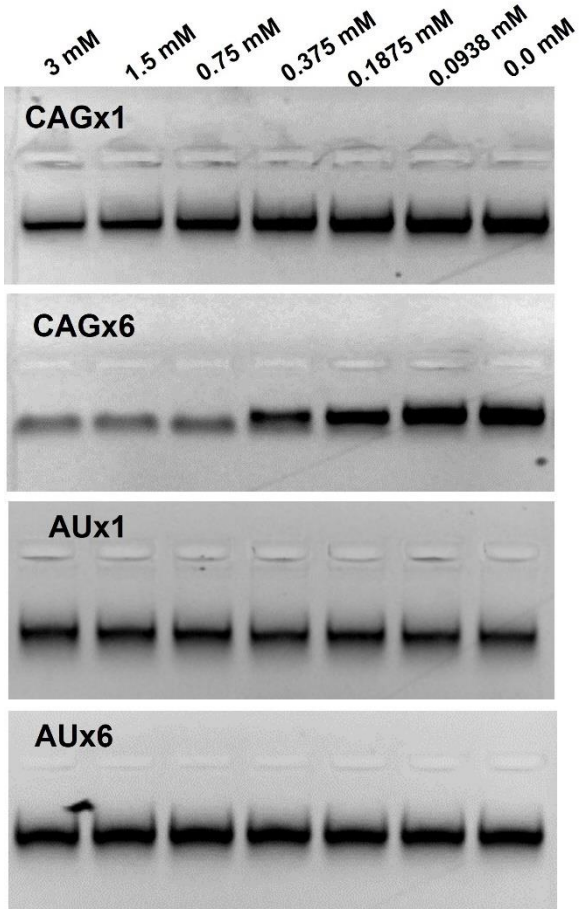


Figure 4.5.1. Represent PCR stop assay for AA mismatch containing template and duplex AU in presence of increasing concentration of D25

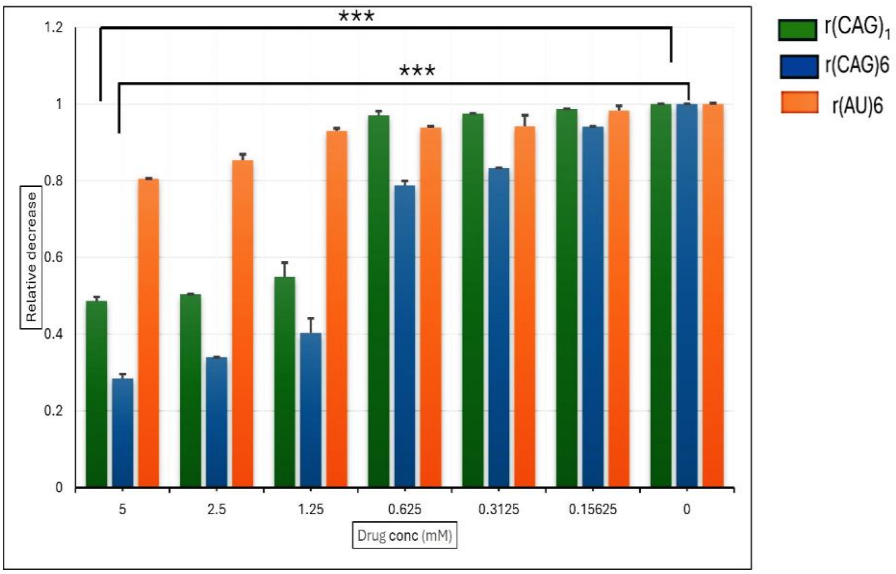


Figure 4.5.2. Represent Bar graph showing relative decrease of PCR stop assay for AA mismatch containing template and AU duplex in presence of increasing concentration of D25

As visible from the gel images for PCR stop assay performed for the template (5'CAG/3'GAC), $(5'CAG/3'GAC)_6$, $(5'CAG/3'GUC)_6$, it is clearly evident that with increasing concentration of small molecule the amplified PCR products as identified by the intensity of bands has decreased significantly, whereas no significant decrease is observed for the canonically paired duplex AU DNA PCR amplification, specifying the selectivity of the D25 molecule to bind with AA mismatch containing nucleic acids. Further analysis was conducted using ImageJ software to ascertain the differences between each set and find the differences as shown in the bar graph. After analysis, a significant decrease in the intensity of bands for AA mismatch DNA is observed with p value of $0.002879 < 0.1$.

4.6 Circular Dichroism spectroscopy evaluates the topological effects of binding of small molecules to RNA

Circular Dichroism (CD) spectroscopy is utilized to investigate how ligand binding influences the loop structures in RNA that arise from mismatched bases. This study examined RNA sequences with A- A mismatches and those with six similar mismatches. RNA with standard A-U base pairing was also compared to the mismatched pairs to evaluate the small molecule's selectivity and specificity.

CD spectra of RNA containing 5’CAG/3’CAG and r$(CAG)_6$ showed a large positive peak at around 265 nm and a large negative peak at around 220 nm, and a smaller peak at around 240 nm, which was consistent with double-stranded A-form structures. With the addition of D25 up to a D/N ratio of 7.0, the intensity of the positive peak does not change, however, there is an observed red shift in the negative peak also the intensity of the negative peak increases known as hyperchromatic shift, defining the stabilizing effect of the lead molecule by binding with the AA mismatch, thus defining change in the stacking interaction of the motif RNA. The change was more profound in the RNA motif containing 6 (AA) mismatches, showing stronger molecule interaction with the RNA as the number of repeats increased.

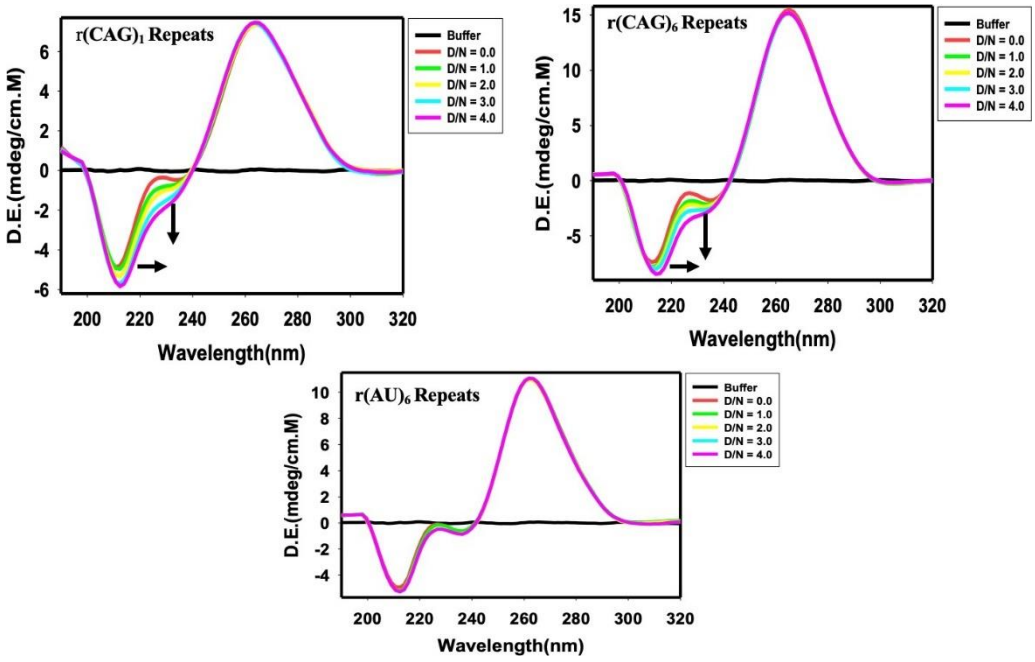


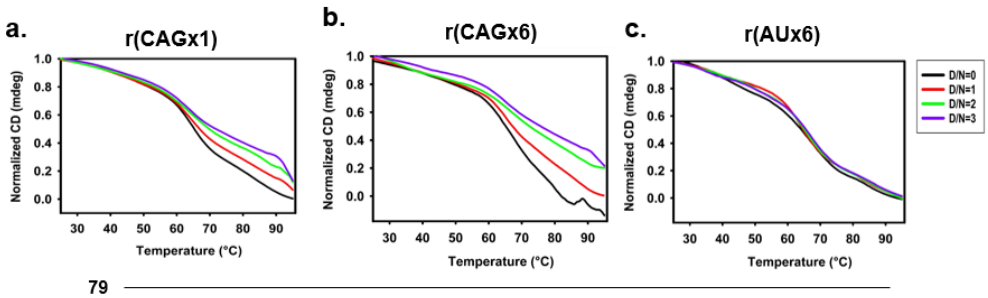
Figure 4.6. Circular dichroism spectrum of free RNA and in the presence of increasing concentration of D25. Vertical arrow represents hyperchromicity and horizontal arrow represents red shift in CD spectra after titration using D25.

There was no discernible change in the positive or negative peak of the control RNA after titrating with the D25 molecule. This implies that D25 has little to no affinity for canonical paired nucleic acid. Moreover, with the addition of the drug, there is no discernable change in the overall structure of the RNA motifs.

4.7 Thermal Denaturation Analysis of r(CAG)^{exp} RNAs with D25

CD thermal denaturation studies of RNA with ligand have been used to determine the folding pattern of RNA as a function of temperature. The CD melting curves were recorded at 260 nm wavelength for all CAG repeat RNAs up to ratio D/N = 3.0. On addition of D25 at 3.0 molar ratio the change in melting temperature (ΔT_m) was observed as 6.28, 10.51 °C for r(CAG)$_1$, r(CAG)$_6$ respectively. In comparison to CAG RNA, no change in T_m was observed with AU paired RNA. Thermal profile data implies that D25 strongly and selectively interacts with

CAG repeats RNA and forms a stable complex, over AUx6 duplex control RNA



	CAGx1	CAGx6	Aux6
T_m °C	64.9	64.89	64.20
	66.99	66.87	64.38
	69.8	71.9	64.92
	71.18	75.4	64.93

Figure 4.7. Systematic representation of circular dichroism thermal profile of r(CAG)exp and r(AU)exp duplex RNAs with D25: (a) $r(CAG)_1$, (b) $r(CAG)_6$, (c) $r(AU)_6$ RNA. D25 enhanced the melting temperature of expanded CAG RNAs, whereas no significant change in T_m was observed with other mismatch RNA motifs and AU duplex RNA. D/N denotes the drug to nucleotide ratio.

4.8 Molecular study to gather insights about the binding of CAG RNA with small molecules

To append our findings from the biophysical experiments, docking analyses were performed to generate models of drug-RNA complex Insilco. Our mismatch AA RNA motif was found to interact with the drug molecules. After docking the small molecule with the duplex RNA complex, the top seven affinity scores range from -6.1 to -5.6 Kcal/mol. The presence of 3 (CAG) repeats causes the strong interaction between the molecule and Duplex, mainly the A8 and A11 in the RNA duplex complex, to form conventional hydrogen bonds with small molecules, justifying the interaction of molecule at the AA mismatch loop and preliminary in-vitro studies.

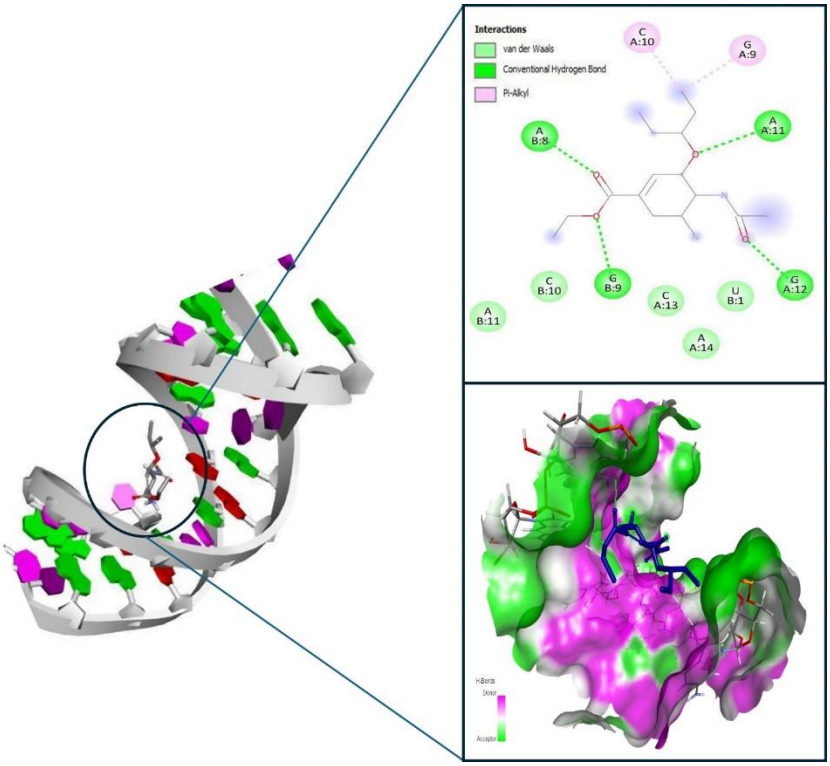


Figure 4.8. Represents 3-D and 2-D interaction of small molecule D25 with duplex RNA

4.9 Interaction studies using molecular dynamic simulations

Following a 100 ns long molecular dynamics simulation, the trajectory was examined to evaluate the stability of the RNA structure as well as RNA-ligand complex. The root mean square deviation (RMSD) shows the structural variations in the dynamic trajectory. The consistent RMSD observed in molecular dynamics (MD) simulation confirmed the stable nature of the RNA structure (Figure 4.7.1.) and the RNA-Ligand complexes. Further comparison between the RMSD of RNA and complex shows the complex to be more stable than only RNA. This further indicates the stable nature of the interactions between the RNA and the small molecule.

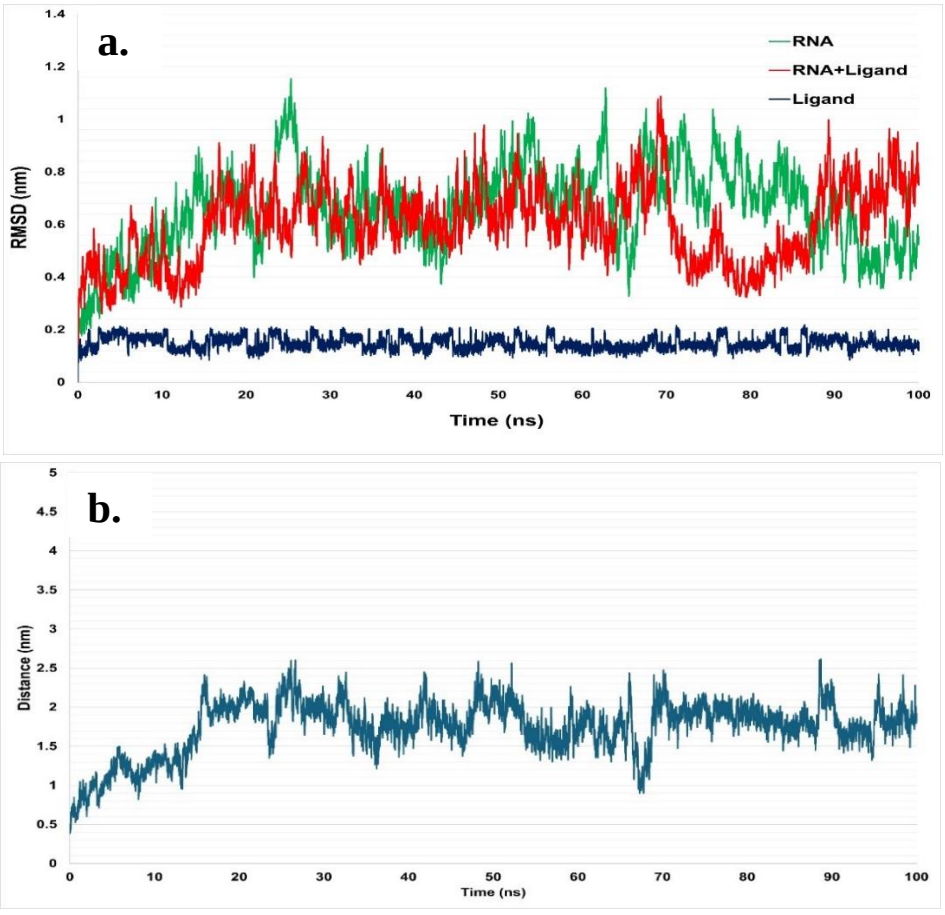


Figure 4.9. a) Plot representing RMSD of CAG repeat RNA alone, RNA and ligand in RNA-D25 complex. b) Plot representing distance between ligand and RNA during simulation

From the above graphs, it is evident that the fluctuations observed during MD run for the RNA ligand complex is due to structural instability of the RNA itself as observed during RNA without docked ligand MD simulation run, moreover the overall stability of RNA is also increased as compared to alone RNA as observed through decrease in periodic fluctuations as well as the intensity of these fluctuations. The stable RMSD of the ligand signifies the stable interaction of the ligand with the RNA forming a stable RNA ligand complex, which is in conjunction with the overall results. This result is further solidified using the distance plot showing the overall difference between the RNA motif and the ligand of interest throughout the simulation run of 100 ns. The distance is comparably constant throughout the run between 1.5 to 2.5 nm showing a stable interaction between ligand and the RNA motif.

4.10 Binding energy calculation

The Relative Binding free energies calculated from the MD trajectory give further insights into the nature of the interactions between RNA and ligand. From the MMPBSA calculation, it is evident that the binding affinity of the D25 with AA mismatch containing RNA (PDB id: 4J50) is -12.95 kJ/mol, which signifies strong interactions between the RNA and the ligand as observed through the docking studies where D25 binds to Adenine base involved in the AA loop formation. The interaction is mainly dependent on electrostatic forces as observed through MMPBSA calculations.

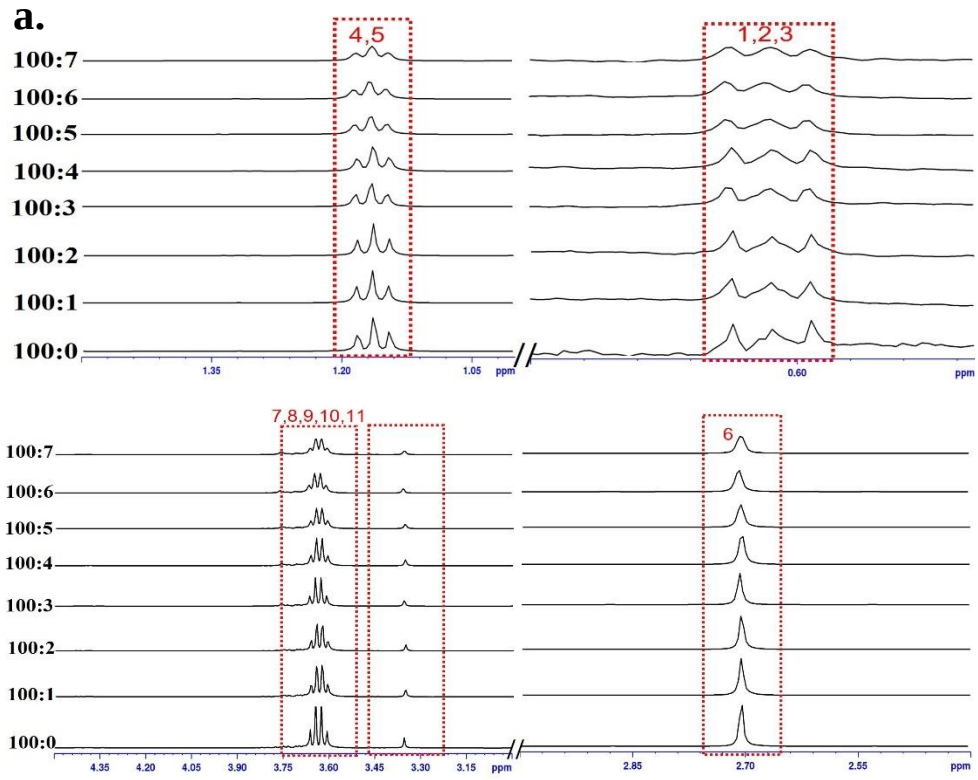
Energy Component (Kcal/mol)	Average	SD	SEM
ΔVDWAALS	-7.40	5.89	0.06
ΔE_{EL}	-71.02	19.07	0.19
ΔE_{GB}	67.19	17.70	0.18
ΔE_{SURF}	-1.72	0.75	0.01
ΔG_{GAS}	-78.42	22.58	0.23
ΔG_{SOLV}	65.47	17.04	0.17
ΔTOTAL	-12.95	6.79	0.07

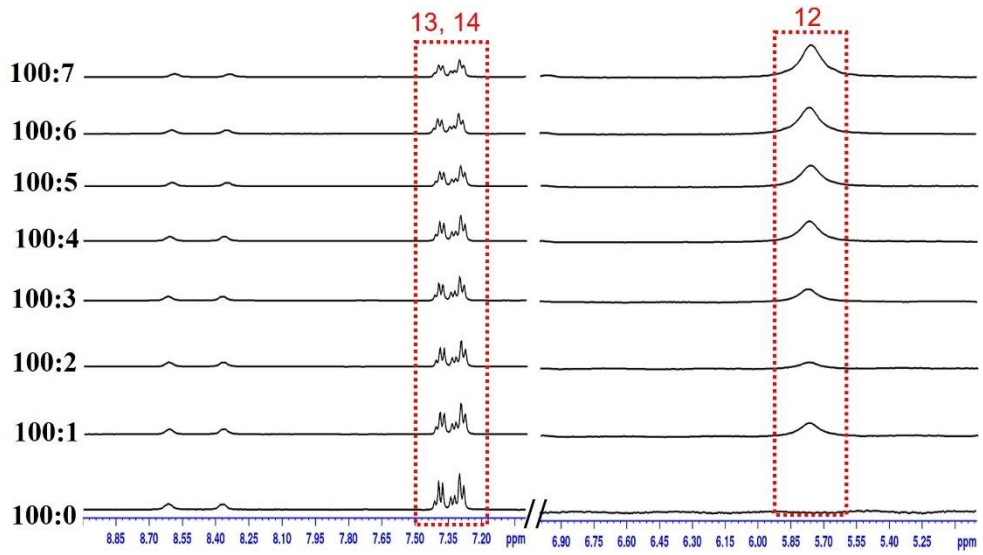
Figure 4.10. Table showing MMPBSA calculation for RNA docked with D25

4.11 Understanding the Binding of D25 to CAG×6 Motif containing RNA by NMR Spectroscopy

The peaks at approximately 1.20 ppm and 0.60 ppm decrease in intensity with increasing titration ratios. These peaks correspond to aliphatic protons, likely methyl (–CH3) and methylene (–CH2–) groups. Protons in the 1.20 ppm region are in slightly electron-rich environments, such as alkyl chains. The decrease in intensity suggests that the drug molecule interacts with these aliphatic protons, causing deshielding. Protons showing 0.60 ppm shift are in highly shielded environments, such as terminal methyl groups. The intensity decrease indicates similar deshielding effects due to drug binding. The peaks at approximately 3.60 ppm and 2.70 ppm also decrease in intensity as the titration ratio increases. These peaks correspond to aliphatic protons, likely involved in methoxy groups ($–OCH_3$) and methylene groups (–CH2–) near electronegative atoms. Protons in the 3.60 ppm region are typically found in environments such as $–OCH_3$ or protons attached to oxygen or nitrogen atoms such as NH_2. The decrease in intensity suggests that the

drug molecule interacts with these aliphatic protons, causing deshielding. Protons at 2.70 ppm are in environments adjacent to electronegative atoms, such as carbonyls (C=O. The intensity decrease indicates similar deshielding effects due to drug binding. The drug molecule likely interacts with hydrophobic regions of the target molecule, driven by van der Waals forces and hydrophobic interactions. The binding causes deshielding of aliphatic protons, reflected in reduced peak intensity. The peak at approximately 5.75 ppm increases in intensity with increasing titration ratios. This peak corresponds to protons in environments such as hydroxyl (–OH) or amide (–NH) groups. The peaks at approximately 7.35 ppm intensity with increasing titration ratios. The decreased intensity indicates that the drug binds near the aromatic rings, causing deshielding. The drug molecule likely forms π-π stacking interactions with aromatic rings stabilizing the complex and causing deshielding. These interactions result in decreased peak intensity due to changes in the electronic environment.





b.

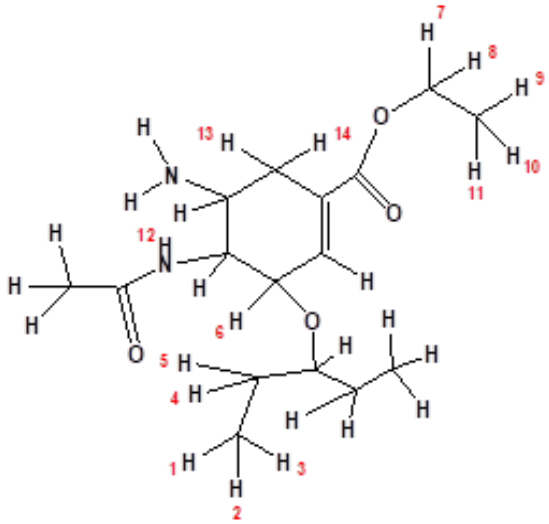
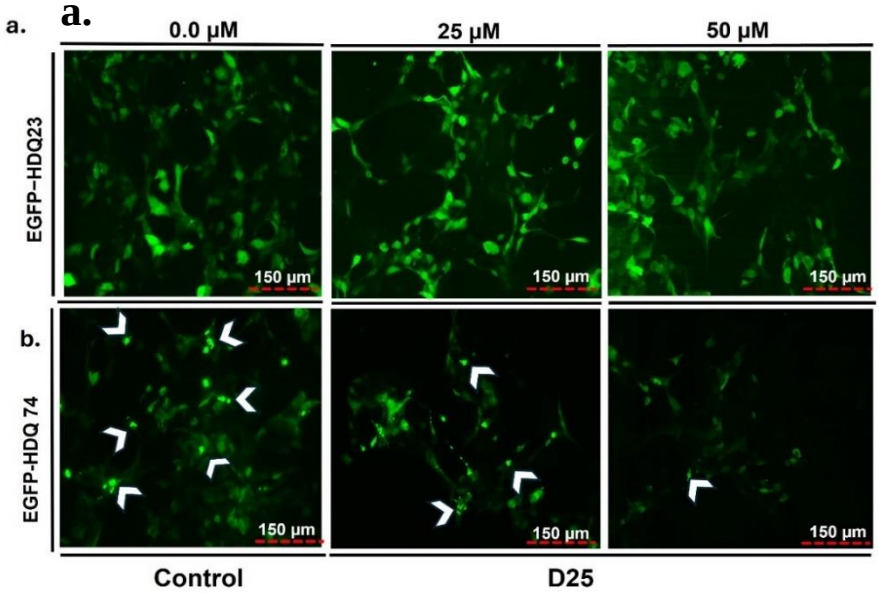


Figure 4.11. Structural insight into the interaction of 5′r(CAGx6)3′ with D25 using NMR spectroscopy. One dimensional proton spectra of RNA 5′r(CAGx6)3′ as a function of increasing concentration of D25. (b) Represents marked Proton of D25 taking part in interaction with CAG repeat RNA motif.

The proton NMR spectra analysis reveals that the drug molecule interacts with the target molecule at multiple binding sites, including aliphatic (3.60 ppm, 2.70 ppm, 1.20 ppm, 0.60 ppm) and aromatic/vinyl (7.35 ppm, 5.75 ppm) regions. The nature of these interactions involves hydrophobic forces, π-π stacking, hydrogen bonding, and electrostatic interactions, leading to deshielding effects and reduced peak intensities. The increasing peak at 5.75 ppm suggests the involvement of hydrogen bonding interactions facilitated by the drug molecule. Aromatic protons are affected by π-π stacking interactions with the drug molecule. These findings provide valuable insights into the binding mechanism, which can inform the design of targeted therapeutic interventions**.**

4.12 Aggregation counting in cellular models

COS-7 cells were transfected with a plasmid encoding a pathogenic repeat expansion of CAG repeats labeled with GFP, resulting in green fluorescence under 395 nm illumination. Transfected cells were exposed to D25 for 12 hours. Fluorescence micrographs were captured at 10X and 60X magnification to study the impact of D25 on the development of polyQ aggregates within the cell, which are the primary factor contributing to cell damage. Arrowheads indicate the creation of aggregates within the cells.



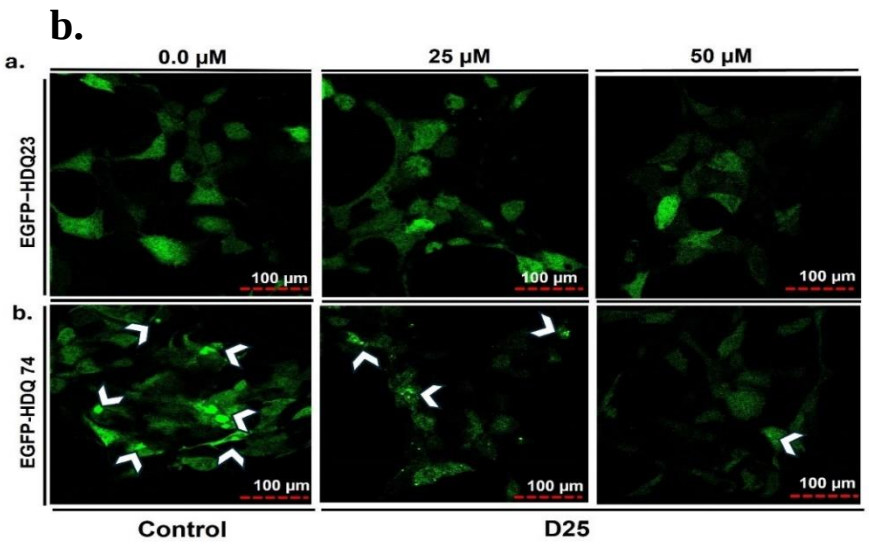


Figure 4.12.1. Represent micrographs of EGFP-HDQ23 and EGFP-HDQ74 transfected cells treated with increasing concentrations of the D25 at different magnifications ((a) 20X and (b) 40X). Arrowhead represents the formation of protein aggregates

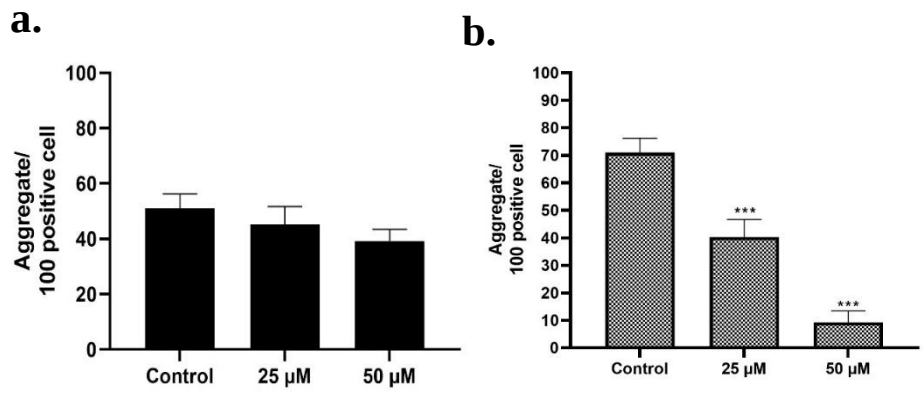


Figure 4.12.2. (a) Plot representing aggregate formation in EGFP-HDQ23 transfected Cos-7 cell line. (b) Plot representing aggregate formation in EGFP-HDQ74 transfected Cos-7 cell line.

From aggregate count plot, it is evident, treatment with D25 significantly reduces the aggregate formation inside the cells transfected with CAGx74 repeats. The decrease in aggregate count per 100 positive cells was significant with increase in the concentration of the drug. From

the micrographs, as the concentration of D25 increases, the number of positive cells displaying aggregation formation drops markedly. Additionally, a reduction in the production of aggregates within a single cell is also noted, hence reducing cellular toxicity. Quantitative examination of the tests shows a decrease in polyQ production within the cells. D25 can reduce both the size and quantity of these aggregates within the cells.

4.13 Western blot analysis

Western blotting is used to measure protein levels within cells. Similarly, this method was employed to confirm the expression of plasmids bearing pathogenic CAG repeat expansions as EGFP-HDQ74 and non-pathogenic CAG repeat expansions as EGFP-HDQ23. Cells were transfected temporarily using these plasmids, and complete cell lysates were used as samples. The validity of the expression was confirmed using anti-GFP and anti-β actin antibodies. Cells that had foreign genetic material introduced into them were exposed to D25 and the contents of the cells were used for an immunoblot experiment. Encouragingly, treatment with D25 demonstrated a dose-dependent reduction in protein aggregates with a 40% reduction observed at 25.0 μM and a further reduction to 75% at 50.0 μM. Notably, no significant inhibition of Beta-actin and control containing r(CAGx23) repeats was observed at the same concentrations.

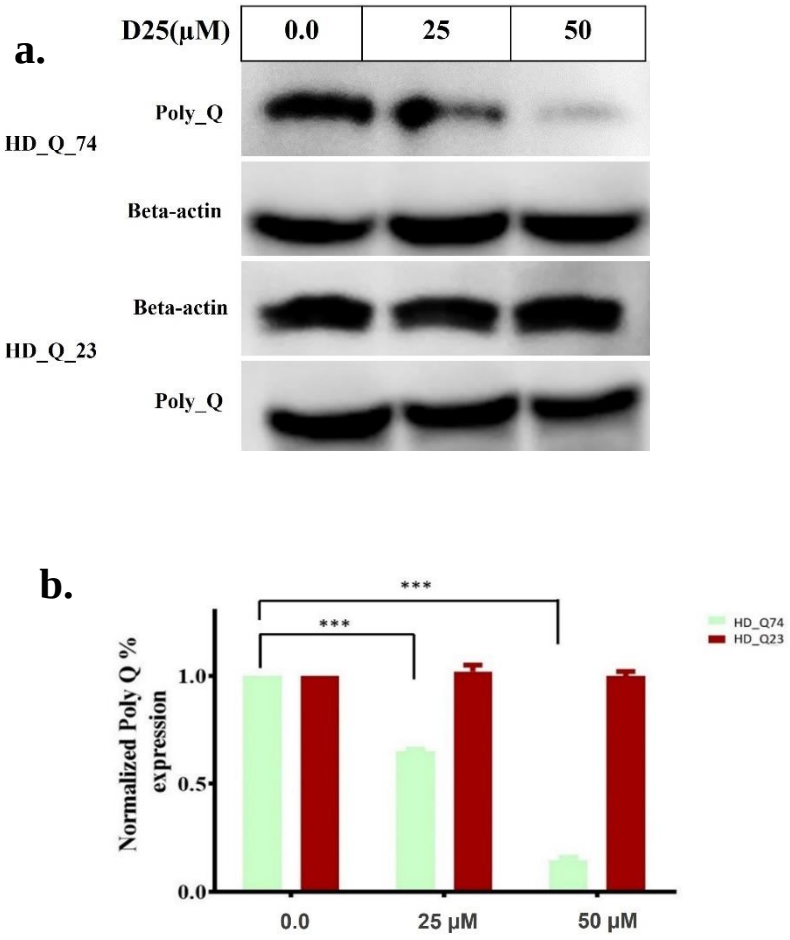


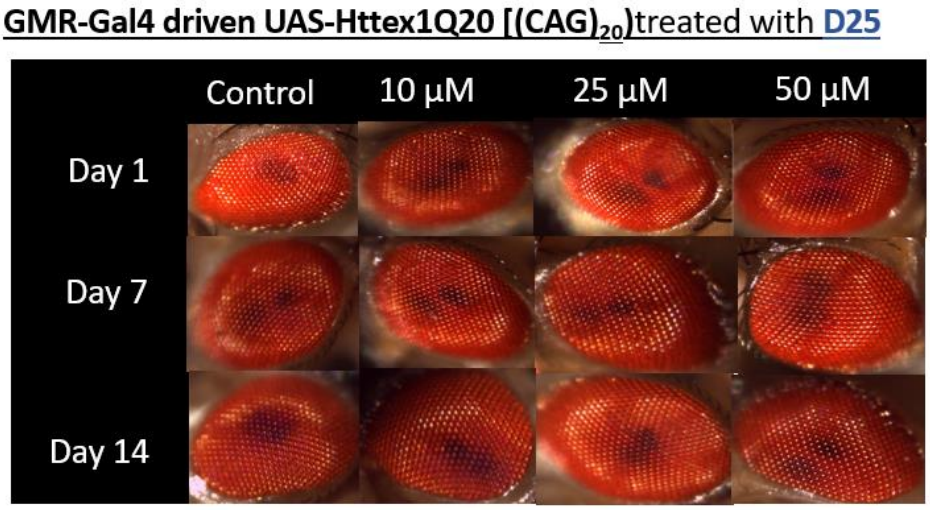
Figure 4.13.1. a) Represents blot image for cells transfected with plasmid containing 74 and 23 repeats and treated with increasing concentration of D25, later subjected to immunoblotting with anti-GFP and anti β- actin antibodies b) Represents bar graph showing significance of the decrease in the polyQ expression inside the transfected cells.

The analysis of blot images was done using ImageJ software after which a bar graph was plotted and significance of the decrease in polyQ expression was plotted. From the blot images it is evident that as the concentration of D25 molecule is increased, there is significant decrease in the polyQ expression with p-value of 0.000089 for 100 µM and p-value of 0.00010 for 200 µM.

4.14 D25 helps alleviate the rough eye phenotype and pigment loss associated with polyQ-induced cytotoxicity.

We employed the gmr-GAL4 driver to activate the expression of both the Httex1pQ20 and Httex1p Q93 transgene within the compound eye of the fly. The expression of mutant proteins through GMR leads to significant eye degeneration and noticeable external abnormalities. The findings strongly suggests that the external morphological abnormalities including a rough eye phenotype, necrotic lesions and pigment loss showed an observable reduction with the increasing concentration of the D25 drug for the Httex1pQ93 flies. However, no significant change was observed for the control type (Httex1pQ20). The drug was mixed in the food according to different concentrations. Roughness and pigmentation loss are indicative of gradual deterioration of the pigmented cells. Incorporation of the specific drug in the food however, clearly alleviates the impaired eye phenotype at days 7 and 14 post eclosion for the diseased phenotype (Httex1pQ93).

a.



b.

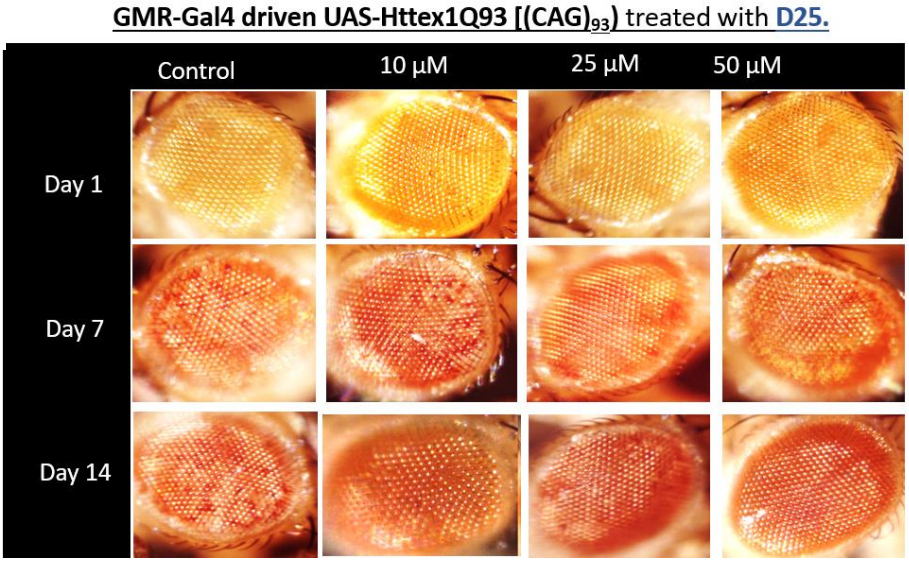


Figure 4.14.1. The figure shows eye images of adult flies a) UAS-Httex1pQ20 and b) UAS-Httex1Q93, both driven by the eye-specific driver line GMR-Gal4. The flies were fed with different concentrations (10 µM, 25 µM and 50 µM) of D25 drug. a) The control type did not show any significant change. The images were taken at 8X magnification.

The Scanning Electron Microscopy (SEM) images of the control and drug treated flies showed a considerable improvement in the eye morphology, that too in a dosage dependent manner. The compound eye of Drosophila comprises approximately 800 ommatidia, which are organized in a stereotypical pattern. These hexagonal units exhibit precise arrangement with consistent spacing between interommatidial bristles. The presence of disease-associated proteins or peptides in the eye results in the manifestation of a characteristic phenotype known as the rough eye phenotype. The treatment with D25 showed suppression of the rough eye phenotype, as visible through the SEM images. The overall eye appearance could be seen to have improved upon dosing with higher concentrations of the drug molecules, compared to the untreated ones. The structural abnormalities seen in the form of rough eye for the untreated one can be seen to

significantly reduce in the drug treated flies, with the most substantial improvement observed in the flies fed with the highest dose (50 µM) of the drug.

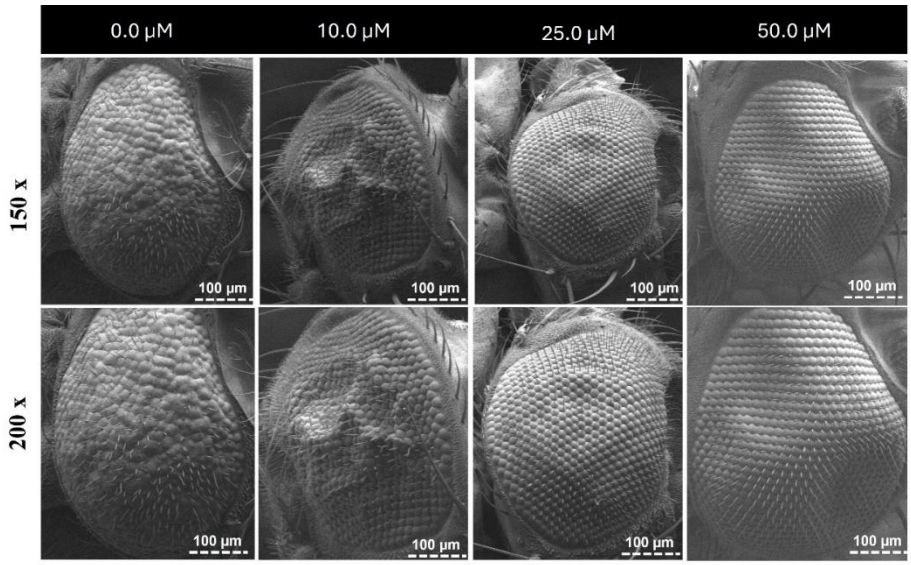


Figure 4.14.2. SEM images of GMR-Gal4 driven UAS-Httex1Q93 [$(CAG)_{93}$] eye structure treated with D25.

4.15 D25 diminishes the cytotoxicity and apoptotic effects inflicted by Poly-Q

Acridine Orange (AO) staining was performed to compare the intensity of cell death in both control and HD model flies. AO staining basically provides a means to visualise the intensity of apoptosis via visualising the acidic vesicular organelles in the tissues of transgenic flies. The third-instar larval eye discs expressing UAS-Httex1pQ93 under the influence of neuronal specific Elav-GAL4 was treated as control against the Elav-GAL4> UAS-Httex1pQ93 fed with different concentrations of D25. Eye discs of control type clearly showed a greater number of apoptotic cells as compared to the drug treated larval eye discs. Infact, a significant reduction has been clearly observed in the number of apoptotic cells on increasing the concentration of the drug. Analyzing the fluorescence intensity of eye discs stained with AO further validated

that providing D25-supplemented food significantly reduced the presence of AO-positive cells. Thus, it can be decisively concluded that a potent D25 concentration significantly aleviates the cytotoxic and apoptotic effects triggered by PolyQ.

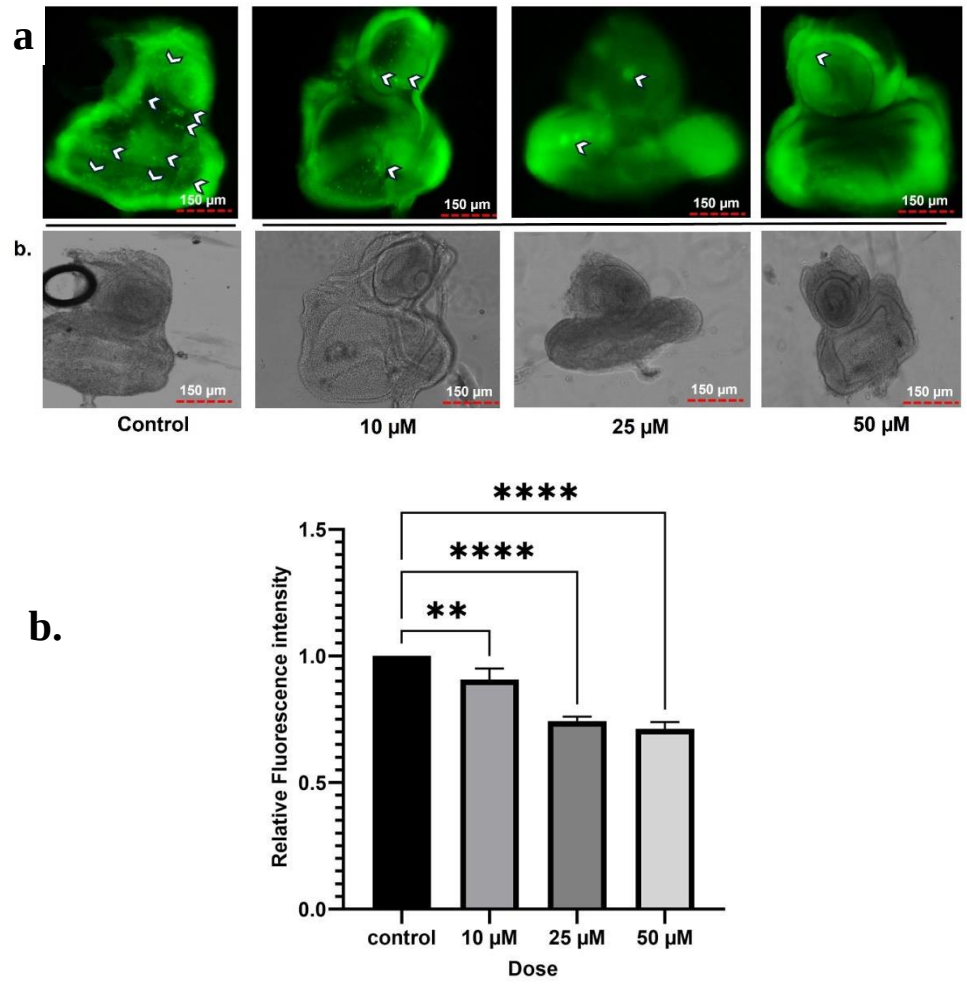


Figure 4.15. a) Acridine orange (AO) staining of third-instar eye discs from elav-GAL4> UAS-Httex1p Q93 larvae (pathogenic) displayed varying levels of cell death. Larvae reared on 10 and 25 µM D25 exhibited a partial reduction in cell death, whereas those exposed to 50µM D25 showed further alleviation. Arrows are indicative of cell death. b) Relative fluorescence intensity also showed a significant decrease with the increasing concentration of D25. Data represents mean ± SEM of values obtained in 5 larvae per condition (n = 5). Statistical significance: ****p < 0.001; **p < 0.001, compared with control. Analysis of data was performed using an analysis of variance (ANOVA) followed by Tukey's post hoc test.

4.16 D25 improves the locomotor dysfunctions caused by PolyQ

Impaired neuronal functioning is also associated with deficiencies in mobility and climbing proficiency. In order to examine this, we assessed the impact of D25 on the motor abilities of flies expressing Httex1Q20 and Httex1Q93 driven by the elav-GAL4 (neuron specific) driver line. Larvae from both elav-gal4>Httex1pQ93 and elav-gal4>Httex1pQ93 were supplemented with varying concentrations of D25. These larvae then underwent crawling assay, to assess their locomotor functioning. Our results suggest that feeding D25 enhanced the crawling ability of both Httex1pQ93 and Httex1pQ20 larvae, with the degree of improvement depending on the concentration.

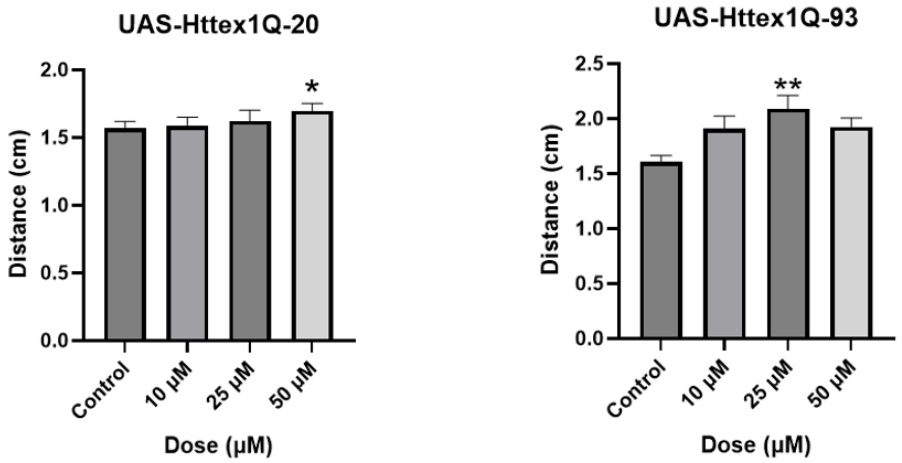
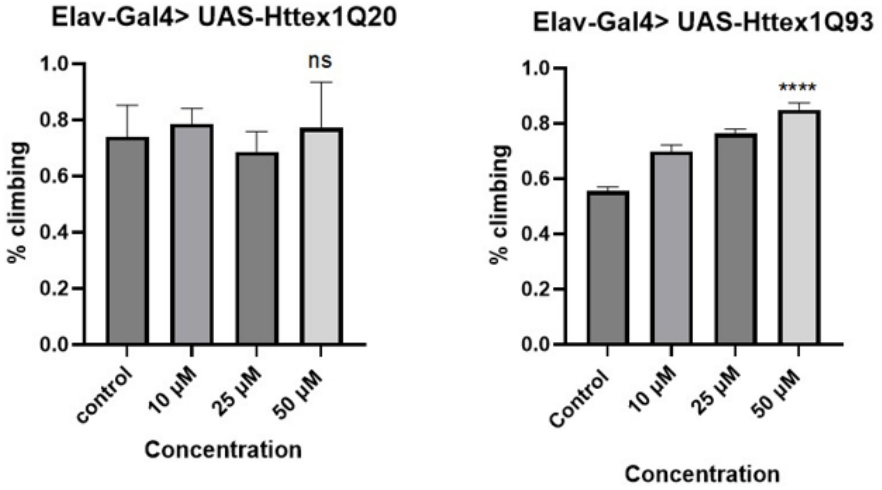


Figure 4.16.1. Data represents the crawling ability of flies at varying concentration of the drug. For each condition, the average of 10 larvae was calculated for two trials (n= 10).

Similarly, to assess the motor functioning of the adult flies, we examined the climbing ability of same age flies (7 days old) on being supplemented with different doses of D25.

The drug particularly did not exhibit any significant impact on the locomotor functioning of control flies (Httex1Q20) that were fed with drug containing food since the larval stage. In contrast, a significant improvement in the motor functioning was observed for the Httex1pQ93 flies, with increasing concentration of the drug. Thus, administration of



different concentrations of D25 fairly improved the locomotor ability at both the larval as well as adult stage of flies.

Figure 4.16.2 Comparison was made between control and drug treated flies of same age. For each condition, the climbing ability of two groups of 10 flies were monitored for a total of 4 trials.

Chapter 5

Conclusion and future perspectives:

Repeat expansion disorders, characterized by various pathological mechanisms such as protein gain or loss of function and RAN translation, have been a significant focus in therapeutic research. Traditionally, the main target for treatment development was pathogenic proteins. However, recent advancements have shifted the focus towards RNA as a key target in mitigating the pathogenesis of these disorders. This shift is primarily due to methods like antisense technology, but the use of small molecules is particularly promising due to their ability to cross the blood-brain barrier effectively.

Disorders caused by CAG repeats, which are prevalent in conditions like Huntington's disease and Spinocerebellar Ataxia (SCA), are typically associated with the production of poly-glutamine (PolyQ) proteins. The binding of small molecules to RNA has emerged as a therapeutic breakthrough in combating these expansion disorders. Studies have increasingly shown that small molecules can stabilize CAG repeats, thereby alleviating their pathogenic effects.

In this study, we focused on screening FDA-approved small molecules for their potential to bind with CAG repeat sequences. Our comprehensive biophysical analyses revealed that molecule D25 exhibited both specificity and selectivity towards RNA motifs containing CAG repeats. This finding suggests its potential as a therapeutic agent against such disorders.

Furthermore, the efficacy of D25 to mitigate cellular toxicity due to formation of Poly-Q protein was exemplified using cell- based studies involving aggregation counting, immunoblotting. Later, Drosophila HD model also showed promising results as dose dependent neurobehavioral improvement was observed along with reduction in cell death. The current work lays the groundwork for screening FDA-approved drugs

for their effectiveness against repeat expansion disorders. It opens new avenues for exploring small molecule therapeutics in the context of these complex genetic conditions. This approach not only enhances our understanding of the molecular mechanisms underlying these disorders but also paves the way for developing more targeted and effective treatment strategies in the future.

The promising results from this study pave the way for several future research directions and potential advancements in the field:

Combination Therapies: Exploring combination therapies involving RNA-targeting small molecules and other therapeutic agents (e.g., antisense oligonucleotides, RNAi technologies, and protein-targeting drugs) could provide synergistic effects and enhance therapeutic outcomes.

Clinical Trials: Based on promising preclinical results, initiating clinical trials to assess the safety and efficacy of RNA-targeting small molecules in human patients is a critical step. Early-phase trials can provide valuable data on dosing, tolerability, and therapeutic potential.

Advanced Delivery Systems: Developing advanced delivery systems, such as nanoparticles, liposomes, and targeted delivery vehicles, can enhance the bioavailability and tissue-specific targeting of RNA-binding small molecules, improving their therapeutic index.

Biomarker Development: Identifying and validating biomarkers for disease progression and therapeutic response can facilitate the clinical development of RNA-targeting therapies. Biomarkers can help monitor treatment efficacy and tailor personalized therapeutic regimens.

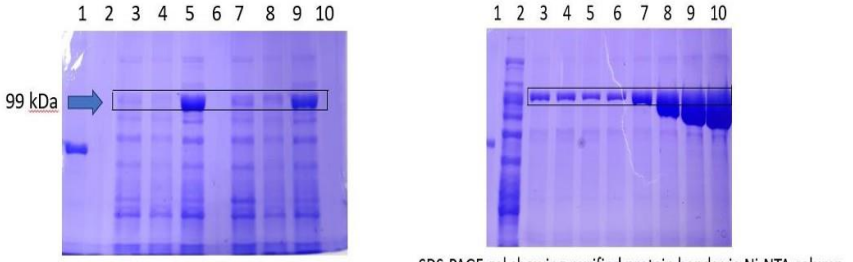
Appendix

Table S1: List of oligonucleotides used in the study

Serial no.	Oligo Name	Sequence
1	5'CAG/3'GGC	GGGAGAGGGTTTAAT**CAG**TACGAAAGTA**CGG**ATTGG ATCCGCAAGG
2	5'CCG/3'GAC	GGGAGAGGGTTTAAT**CCG**TACGAAAGTA**CAG**ATTGG ATCCGCAAGG
3	5'CGG/3'GAC	GGGAGAGGGTTTAAT**CGG**TACGAAAGTA**CAG**ATTGG ATCCGCAAGG
4	5'CCG/3'GUC	GGGAGAGGGTTTAAT**CCG**TACGAAAGTA**CUG**ATTGG ATCCGCAAGG
5	5'CAG/3'GCC	GGGAGAGGGTTTAAT**CAG**TACGAAAGTA**CCG**ATTGG ATCCGCAAGG
6	5'CUG/3'GCC	GGGAGAGGGTTTAAT**CUG**TACGAAAGTA**CCG**ATTGG ATCCGCAAGG
7	5'CGG/3'GGC	GGGAGAGGGTTTAAT**CGG**TACGAAAGTA**CGG**ATTGG ATCCGCAAGG
8	5'CAG/3'GAC	GGGAGAGGGTTTAAT**CAG**TACGAAAGTA**CAG**ATTGG ATCCGCAAGG
9	5'CCG/3'GCC	GGGAGAGGGTTTAAT**CCG**TACGAAAGTA**CCG**ATTGG ATCCGCAAGG
10	5'CUG/3'GUC	GGGAGAGGGTTTAAT**CUG**TACGAAAGTA**CUG**ATTGG ATCCGCAAGG
11	5'CAG/3'GUC	GGGAGAGGGTTTAAT**CAG**TACGAAAGTA**CUG**ATTGG ATCCGCAAGG
12	(5'CAG/3'GAC)6	GGGAGAGGGTTTAAT**CAGCAGCAGCAGCAGCAG**TA CGAAAGTA**CAGCAGCAGCAGCAGCAG**ATTGGATCC GCAAGG
13	(5'CAG/3'GUC)6	GGGAGAGGGTTTAAT**CAGCAGCAGCAGCAGCAG**TA CGAAAGTA**CUGCUGCUGCUGCUG**ATTGGATCC GCAAGG

Table S2: List of primers used in the study

Serial no.	Oligo name	Sequence
1	Forward primer	GGCCGGATCCTAATACGACTCACTATAGGGAGAGGGT TTAAT
2	Reverse primer	CCTTGCGGATCCAA



SDS-PAGE gel showing induction

Well no.	Sample
1	BSA
3	Primary
4	Before IPTG
5	After IPTG(0.5mM)
7	primary
8	Before IPTG
9	After IPTG(0.5mM)

SDS-PAGE gel showing purified protein bands via Ni-NTA column.

Well no.	Elution(mg/ml)
3 (50mM Imidazole)	0.082
4 (3000mM Imidazole)	0.085
5 (3000mM Imidazole)	0.090
6 (3000mM Imidazole)	0.141
7 (3000mM Imidazole)	0.362
8 (3000mM Imidazole)	1.106
9 (3000mM Imidazole)	2.506
10 (3000mM Imidazole)	3.591

Figure S1: SDS-PAGE gel showing induction of T7-RNA Polymerase (left). Alongside is the gel image of the protein after purification via the Ni-NTA column.

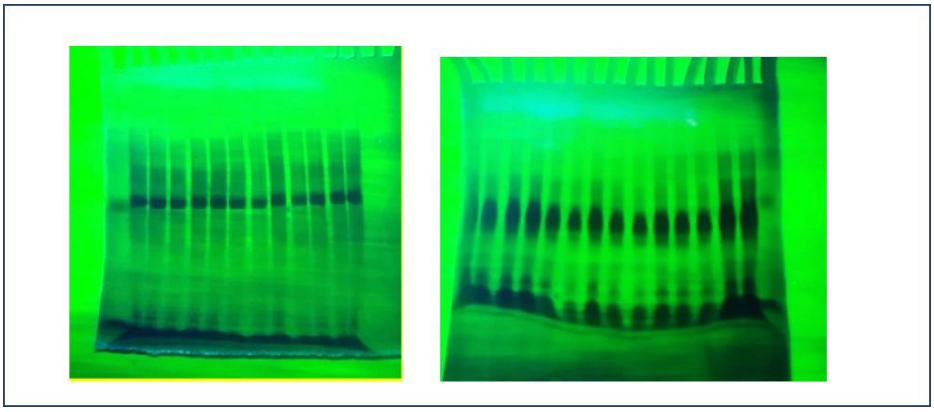


Figure S2. Representative denaturing gel images showing in-vitro transcription of 2 RNA motifs.

References:

Ahamad, Shakir, and Shahnawaz A. Bhat. 2022. 'The Emerging Landscape of Small-Molecule Therapeutics for the Treatment of Huntington's Disease'. *Journal of Medicinal Chemistry* 65 (24): 15993–32. https://doi.org/10.1021/acs.jmedchem.2c00799.

Albin, Roger L., Anne B. Young, and John B. Penney. 1989. 'The Functional Anatomy of Basal Ganglia Disorders'. *Trends in Neurosciences* 12 (10): 366–75. https://doi.org/10.1016/0166-2236 (89)90074- X.

Andrich, Jürgen, Carsten Saft, Natalie Ostholt, and Thomas Müller. 2007. 'Complex Movement Behaviour and Progression of Huntington's Disease'. *Neuroscience Letters* 416 (3): 272–74. https://doi.org/10.1016/j.neulet.2007.02.027.

Armstrong, Richard. 2020. 'What Causes Neurodegenerative Disease?' *Folia Neuropathologica* 58 (2): 93–112. https://doi.org/10.5114/fn.2020.96707.

Bates, Gillian P., Ray Dorsey, James F. Gusella, Michael R. Hayden, Chris Kay, Blair R. Leavitt, Martha Nance, et al. 2015. 'Huntington Disease'. *Nature Reviews Disease Primers* 1 (1): 15005. https://doi.org/10.1038/nrdp.2015.5.

Budworth, Helen, and Cynthia T. McMurray. 2013. 'A Brief History of Triplet Repeat Diseases'. In , 3–17. https://doi.org/10.1007/978-1-62703-411-1_1.

Childs-Disney, Jessica L., Xueyi Yang, Quentin M. R. Gibaut, Yuquan Tong, Robert T. Batey, and Matthew D. Disney. 2022. 'Targeting RNA Structures with Small Molecules'. *Nature Reviews Drug Discovery* 21 (10): 736–62. https://doi.org/10.1038/s41573-022-00521-4.

Dickey, Audrey S., and Albert R. La Spada. 2018. 'Therapy Development in Huntington Disease: From Current Strategies to

Emerging Opportunities'. *American Journal of Medical Genetics Part A* 176 (4): 842–61. https://doi.org/10.1002/ajmg.a.38494.

DiFiglia, Marian, Ellen Sapp, Kathryn O. Chase, Stephen W. Davies, Gillian P. Bates, J. P. Vonsattel, and Neil Aronin. 1997. 'Aggregation of Huntingtin in Neuronal Intranuclear Inclusions and Dystrophic Neurites in Brain'. *Science* 277 (5334): 1990–93. https://doi.org/10.1126/science.277.5334.1990.

Evans, Stephen JW, Ian Douglas, Michael D Rawlins, Nancy S Wexler, Sarah J Tabrizi, and Liam Smeeth. 2013. 'Prevalence of Adult Huntington's Disease in the UK Based on Diagnoses Recorded in General Practice Records'. *Journal of Neurology, Neurosurgery & Psychiatry* 84 (10): 1156–60. https://doi.org/10.1136/jnnp-2012-304636.

Günther Deuschl, Priya Parmar, Michael Brainin, and Christopher Murray. 2020. 'The Global Burden of Neurological Disorders: Translating Evidence into Policy'. *The Lancet Neurology* 19 (3): 255–65. https://doi.org/10.1016/S1474-4422 (19)30411-9.

Fiszer, Agnieszka, and Wlodzimierz J. Krzyzosiak. 2013. 'RNA Toxicity in Polyglutamine Disorders: Concepts, Models, and Progress of Research'. *Journal of Molecular Medicine* 91 (6): 683–91. https://doi.org/10.1007/s00109-013-1016-2.

'Focus on Neurodegenerative Disease'. 2018. *Nature Neuroscience* 21 (10): 1293–1293. https://doi.org/10.1038/s41593-018-0250-x.

Fu, Ying-Hui, Derek P.A. Kuhl, Antonio Pizzuti, Maura Pieretti, James S. Sutcliffe, Stephen Richards, Annemieke J.M.H. Verkert, et al. 1991. 'Variation of the CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox'. *Cell* 67 (6): 1047–58. https://doi.org/10.1016/0092-8674 (91)90283-5.

Li, Qingxin, and CongBao Kang. 2020. 'Mechanisms of Action for Small Molecules Revealed by Structural Biology in Drug Discovery'. *International Journal of Molecular Sciences* 21 (15): 5262.

https://doi.org/10.3390/ijms21155262.

MACDONALD, M. 1993. ‘A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington’s Disease Chromosomes’. *Cell* 72 (6): 971–83. https://doi.org/10.1016/0092-8674 (93)90585-E.

Magnotta, Vincent A., Jinsuh Kim, Tim Koscik, Leigh J. Beglinger, Daisy Espinso, Doug Langbehn, Peg Nopoulos, and Jane S. Paulsen. 2009. ‘Diffusion Tensor Imaging in Preclinical Huntington’s Disease’. *Brain Imaging and Behavior* 3 (1): 77–84. https://doi.org/10.1007/s11682-008-9051-2.

Mahajan, Hitendra S., and Payal H. Patil. 2021. ‘Gel-Based Delivery of Neurotherapeutics via Naso- Brain Pathways’. In *Direct Nose-to-Brain Drug Delivery*, 225–45. Elsevier. https://doi.org/10.1016/B978-0-12-822522-6.00026-6.

Malik, Indranil, Chase P. Kelley, Eric T. Wang, and Peter K. Todd. 2021. ‘Molecular Mechanisms Underlying Nucleotide Repeat Expansion Disorders’. *Nature Reviews Molecular Cell Biology* 22(9): 589–607. https://doi.org/10.1038/s41580-021-00382-6.

Newell, Kathy L., Philip Boyer, Estrella Gomez-Tortosa, Wendy Hobbs, E. Tessa Hedley-Whyte, Jean Paul Vonsattel, and Bradley T. Hyman. 1999. ‘α-Synuclein Immunoreactivity Is Present in Axonal Swellings in Neuroaxonal Dystrophy and Acute Traumatic Brain Injury’. *Journal of Neuropathology & Experimental Neurology* 58 (12): 1263–68. https://doi.org/10.1097/00005072-
199912000-00007.

Paulson, Henry. 2018. ‘Repeat Expansion Diseases’. In, 105–23. https://doi.org/10.1016/B978-0-444- 63233-3.00009-9.

Pearson, Christopher E., Kerrie Nichol Edamura, and John D. Cleary. 2005. 'Repeat Instability: Mechanisms of Dynamic Mutations'. *Nature Reviews Genetics* 6 (10): 729–42. https://doi.org/10.1038/nrg1689.

Ross, Christopher A, and Michelle A Poirier. 2004. 'Protein Aggregation and Neurodegenerative Disease'. *Nature Medicine* 10 (S7): S10–17. https://doi.org/10.1038/nm1066.

Rüb, U., K. Seidel, H. Heinsen, J.P. Vonsattel, W.F. den Dunnen, and H.W. Korf. 2016. 'Huntington's Disease (HD): The Neuropathology of a Multisystem Neurodegenerative Disorder of the Human Brain'. *Brain Pathology* 26 (6): 726–40. https://doi.org/10.1111/bpa.12426.

Stack, E. C., S. J. Del Signore, R. Luthi-Carter, B. Y. Soh, D. R. Goldstein, S. Matson, S. Goodrich, et al. 2007. 'Modulation of Nucleosome Dynamics in Huntington's Disease'. *Human Molecular Genetics* 16 (10): 1164–75. https://doi.org/10.1093/hmg/ddm064.

Vonsattel, Jean-Paul, Richard H. Myers, Thomas J. Stevens, Robert J. Ferrante, Edward D. Bird, And Edward P. Richardson. 1985. 'Neuropathological Classification of Huntington's Disease'. *Journal of Neuropathology and Experimental Neurology* 44(6): 559–77. https://doi.org/10.1097/00005072-198511000-00003.

Yebenes, Justo Garcia de, Bernhard Landwehrmeyer, Ferdinando Squitieri, Ralf Reilmann, Anne Rosser, Roger A Barker, Carsten Saft, et al. 2011. 'Pridopidine for the Treatment of Motor Function in Patients with Huntington's Disease (MermaiHD): A Phase 3, Randomised, Double-Blind, Placebo- Controlled Trial'. *The Lancet Neurology* 10 (12): 1049–57. https://doi.org/10.1016/S1474-4422 (11)70233-2.