

**Structural insight of r(CGG) motif and small
molecule based therapeutics development for
the expanded CGG repeats RNA associated
neurological disorders**

A THESIS

*Submitted in partial fulfillment of the
requirement for the award of the degree*

of

DOCTOR OF PHILOSOPHY

By

Arun Kumar Verma



**DISCIPLINE OF BIOSCIENCES & BIOMEDICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE**

October, 2019



INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **Structural insight of r(CG_nG) motif and small molecule based therapeutics development for the expanded CGG repeats RNA associated neurological disorders** in the partial fulfillment of the requirement for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DISCIPLINE OF BIOSCIENCES AND BIOMEDICAL ENGINEERING, Indian Institute of Technology, Indore**, is an authentic record of my own work carried out during the time period of July 2015 to September 2019 under the supervision of Dr. Amit Kumar, Associate Professor, Indian Institute of Technology, Indore.

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

**Signature of Student with date
(Arun Kumar Verma)**

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

**Signature of Thesis Supervisor with date
(DR. AMIT KUMAR)**

Arun Kumar Verma has successfully given his Ph.D. oral examination held on _____

Signature of chairperson
(OEB)

Date:

Signature of External
Examiner

Date:

Signature of Thesis
Supervisor (s)

Date:

Signature of PSPC
Member # 1

Date:

Signature of PSPC
Member # 2

Date:

Signature of Convener, DPGC

Date:

Signature of Head of Discipline
Date:

ACKNOWLEDGMENTS

Completion of this doctoral dissertation was possible with the support of several people. I would like to express my sincere gratitude to all of them. First of all, I am extremely thankful to my supervisor, Dr. Amit Kumar for his valuable guidance, scholarly inputs and consistent encouragement, I received throughout the research work. Without his invariable leadership and motivation this research work could not have been completed. The combination of his expertise, enthusiasm, and encouragement has been a major factor in my development as a researcher. On the academic level, Dr. Amit Kumar encouraged me to learn the fundamentals of conducting scientific research. Under his supervision, I got the opportunity to learn how to define a research problem, organize the literature search and data, planning of the experiments and time management to find a solution, and finally, translating the bench work into the publication. On a personal level, Dr. Amit Kumar motivated me by his hardworking and passionate attitude and also sharing his own experiences.

I also place on record, my sense of gratitude to the members of the PSPC committee, Dr. Prashant Kodgire and Dr. Sanjay Kumar Singh for their insightful comments and for sharing their tremendous experience with me. I am thankful to Dr. Hem C. Jha (Head, Discipline of Biosciences & Biomedical Engineering) for his continuous support in various aspects. I wish to express my sincere thanks to all faculty members of BSBE discipline for all their suggestions and support. I am also thankful to all non-teaching staff of BSBE family, especially, Mr. Amit Mishra for his technical assistance.

I am very grateful to Prof. Pradeep Mathur, Director, Indian Institute of Technology, Indore for his continuous encouragement, support, and for providing me the research facilities within amenable working environment. I also wish to express a very warm thanks to Prof. Neelesh Kumar Jain, Dean of Academic Affairs, Indian Institute of Technology, Indore for his excellent support, motivation and encouragement throughout my Ph.D. tenure.

I have especially benefited from the genuinely professional support of the technical staffs of Sophisticated Instrument Centre (SIC): Mr. Kinny Pandey, Mr. Ghanashyam A. Bhavsar, Mr. Manish Kushwaha, and Dr. Ravinder Kumar. My sincere thanks to the staff member of the academic office: Mr. Suresh Chandra Thakur, Mr. Tapesih Parihar, Mr. Kailash Jamra, Mr. Nitin Parashar, Mr. Anup Sharma and Ms. Rinki Kukreja for their continuous help and support.

I thank my fellow labmates Dr. Arpita Tawani, Dr. Subodh Kumar Mishra, Dr. Sonali Bhagwat, Dr. Naresh Kumar, Dr. Eshan Khan, Ms. Richa Agrawal, Ms. Prativa Majee, Ms. Nirali Pandya, Ms. Neha Jain, and Mr. Uma Shankar. It was a great delight to have Dr. Subodh Kumar Mishra and Dr. Eshan Khan as a senior cum friend as I could enjoy the sleepless nights we were working together. It was an excellent time spent with all my lab members and will surely miss the cries and laughs shared with you all during the evening snacks. My immense love and care to all my junior labmates for always being there and bearing with me in the good and bad during my wonderful days of Ph.D. I would also like to thank all my seniors and juniors from Discipline of Biosciences and Biomedical Engineering, IIT Indore, for all the love that they have showered upon me. I could not have asked for more than what I got from them.

Lastly, I would like to thank my family for all their care, love, affection and encouragement. I sincerely express my love and gratitude to my father Mr. Dwarka Prasad Verma, the three of my angels, my mother Mrs. Savitri Dwarka Prasad Verma, my wife Mrs. Jayshree Arun Verma and my daughter, Mishka (Bella) for their unconditional love, trust, timely encouragement, and endless patience.

I would like to express my sincere thanks to IIT Indore for infrastructure and Department of Biotechnology (DBT), India, for my fellowship.

Arun Kumar Verma

***DEDICATED
TO
MY FAMILY***

SYNOPSIS

Ribonucleic acid (RNA) is a unique biological polymer for cellular regulation and maintenance. A central principle about RNA is that it transfers the biological information from DNA *via* RNA to protein. Since many years, before the discovery of its catalytic activity, RNA is only known for its three primary functions such as synthesis xerox of the DNA (mRNA), coupler between the amino acid (protein building block) & genetic codes and ribosomal RNA. Such a new discovery of RNA, its catalytic activity has opened the new universe for the scientific community to study the shaded cellular mechanisms. For the effective function of the cell, rigorous and controlled expression of protein-coding (mRNA) and noncoding RNAs is imperative. In addition, different types of RNAs have been discovered which have a crucial role in cell regulatory processes. RNA has been categorized based on their structural as well as functional roles in cellular events. For example, transfer RNA (t-RNAs), ribosomal RNA (r-RNAs), 7SL/SRP RNA and microRNAs (miRNAs) involved in translation processes. Similarly, other RNAs like small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) involved being participating in RNA processing. Some different sets of RNA molecules involved in other cellular processes such as telomere-associated RNAs, Ribonuclease P RNase, and mitochondrial RNA processing (MRP) RNase. Apart from all these, few RNAs are very less elaborated with unknown functions such as vault RNAs, Y RNAs, Piwi-interacting RNAs (piRNAs).

Typically, RNAs possess different topological structure to deliver specific functions. Topological structure with precise folding of RNAs could be a deciding factor for the phenotypic outcomes. Riboswitches, microRNAs and siRNAs are the epitome for the importance of RNA secondary structure in which canonical and non-canonical base pairing places a central role in small RNA processing that leads to the recognition and successive regulation of selected messenger RNA. Thus, the formation of secondary structures of RNA is crucial to regulate cellular

processes. However, changes in the base-pairing are simple but significant for not only for the regulatory RNA (microRNAs, siRNAs) but also for mRNA and long noncoding RNAs. Unlikely, unusual secondary hairpin structures formation by mRNAs is associated with various human diseases. These uncommon hairpin structures are formed due to the unstable and dynamic expansion of repetitive sequences in the human genome. These repetitive sequences are simple sequence repeats (SSR), or tandem repeats also known as microsatellites, and almost 30 percent of the human genome constituted these microsatellites. Alternation in the sequence and the length of repeatome gives rise to diversity in the species during an evolutionary period. Although, a variation of these simple repetitive DNA sequences beyond threshold limit results human disorders. These repetitive sequences are present in the coding (exons) and noncoding (5'UTR, introns, and 3'UTR) region of the gene and could be tri, tetra, penta, hexa and deca repeats in nature. Trinucleotide repeat (TNR) disorders occur when the number of triplet repeats is higher in mutated genes compared to a number of triplets found in the normal gene.

Additionally, TNRs are genetically inherited in nature and become severe with the successive generation. Increase in number of TNRs beyond the specific limit completely changes the expression and functional profile of the gene. The length of the triplet repeats shows a very striking genotypic-phenotypic correlation with disease pathology. Longer the repeat length leads to worsening the disease condition and early onset of symptoms. Depending on the type of trinucleotide expansion and the location (coding and noncoding) on the gene caused different neurological disorders. CGG nucleotide expansion at 5'UTR of *FMR1* gene causes Fragile X syndrome (FXS), Fragile X-associated tremor/ataxia syndrome (FXTAS) and Fragile X-associated primary ovarian insufficiency (FXPOI). Expansion of GCC repeats at 5'UTR of *FMR2* gene cause Fragile XE mental retardation syndrome (FRAXE).

Similarly, expansion of CAG repeats at coding, and noncoding region of the different genes causes different neurodegenerative diseases including, Spinocerebellar ataxia type (1, 2, 3, 6, 7, 12 and 17), Huntington disease, Huntington disease-like 2, Spinobulbar muscular atrophy and Dentatorubral-pallidoluysian atrophy. Further, CTG and CCTG repeat expansion causes Myotonic dystrophy type 1 (DM1), Spinocerebellar ataxia 8 and Myotonic dystrophy type 2 (DM2) respectively. Likewise, GAA repeat causes Friedreich's ataxia. All neurological disorders show a different pattern of TNRs expansion and the proposed models for the repeat expansion suggested that formation of intermediate looped structure incorporate in the DNA. However, why some repeats expansion occurs more in the genome than other repeats are still unanswered. The previous report demonstrated that DNA repair and replication mediates the repeat expansion through polymerase slippage mechanism.

All these neurological disorders possess common pathogenic mechanisms

- 1) RNA gain of function in which mRNA transcript form stable secondary structure that sequester several RNA binding protein for e.g. DiGeorge syndrome critical region 8 (DGCR8), DROSHA, Src associated in mitosis of 68 kDa (Sam68), Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNPA2/B1), CUGBP1 and Pur α . Thus, sequestration of these protein results in the formation of nuclear RNA foci and also influence splicing regulation. Another pathogenic mechanism includes protein gain of function in which repeat expansion promotes the formation of homopolymeric protein aggregates e.g. polyglutamine (polyQ) and polyglycine (polyG). Additionally, protein loss of function mechanism occurs due to over-expansion of TNRs on the corresponding gene induce epigenetic silencings such as Frataxin and FMRP.

FXTAS is caused by the CGG trinucleotide expansion on the 5'UTR of fragile X mental retardation 1 (*FMR1*) gene on the long arm of the X chromosome at Xq27.3, code for fragile X mental retardation protein (FMRP). It is a late-onset monogenic neurological disorder which affects adult over the age of 50. The

numbers of CGG repeats vary 5-54 times at 5'UTR of *FMR1* gene in fragile X carrier. However, due to instability of CGG repeats across generations lead to enhancement in the repeat expansion to 55-200, called fragile X premutation (PM) carriers. The prevalence rate of FXTAS is approximately 1 in 400-800 males and 1 in 260 females. The probability of developing FXTAS in older age (50) is 40% of males and 8-16% of females. Male are more prone for FXTAS than female as extra X chromosome compensate the disease condition. Initially, premutation carriers were thought to be unaffected. However, it is now known to be associated with a huge spectrum of neurodegenerative and clinical symptoms. The major clinical symptoms include with premutation carrier is intention tremor and gait ataxia. Minor clinical symptoms include memory loss, dementia, cognitive decline, Parkinsonism, and psychiatric problems. Neurohistopathological hallmark of FXTAS includes the formation of ubiquitin-positive inclusion bodies inside the brain cells (neurons and astrocytes) of animal and post-mortem brain tissues. *Immunohistochemistry* and *mass spectrometric* analysis of these inclusion bodies disclosed the presence of more than 30 inclusions associated protein for example- DGCR8, DROSHA, Sam68, hnRNPA2/B1, CUGBP1, Pur α , neurofilament, and lamina A/C. Similar intracellular inclusion bodies are also found in the non-CNS organs of the patient, including, thyroid, kidney, heart, pituitary and adrenal gland. Further, degeneration was also reported in the cerebellum, including Purkinje fibers, nerve cells, glial cells, and spongiosis of cerebellar white matter. Mitochondrial dysfunction, disruption of nuclear lamina structure, loss of dendrite and dendritic spine morphology are other pathological features, examined in FXTAS.

The molecular mechanisms underlying the FXTAS diseased condition were explored using cellular, animal models (mouse and drosophila) and patients. Different molecular mechanisms have been reported. 1) RNA “toxic gain-of-function” mechanism was proposed in which the contribution of mutant *FMR1* transcript in diseased condition was examined. Here, an elevated level of CGG repeat-containing *FMR1* mRNA found to be toxic for the cell. The cytotoxicity

appears due to CGG repeat expansion in *FMRI* transcripts not because of an elevated level of *FMRI* transcripts. This supports the fact that alone CGG expansion in the range of premutation results in the formation of intranuclear inclusion bodies (RNA foci) and reduces cell viability. In this proposed model, sequestration of essential RNA binding proteins (splicing regulator) *via* interaction with expanded CGG motif prevents the proteins to perform their normal functions. Several RNA binding proteins associated with intranuclear inclusions have been identified such as DGCR8 & DROSHA, Sam68, Puro α , hnRNPA2/B1 and CUGBP1, etc.

In another mechanism, 2) Sequestration of RNA regulatory proteins to CGG expansion leads to inefficient regulation of micro-RNA processing, abnormal RNA metabolism, apoptosis, and signal transduction. According to this proposed model, the CGG repeats RNA forms the hairpin structure that mimics the pri-miRNAs structure that leads to the sequestration of DGCR8. Binding of DGCR8 and its partner DROSHA to the expanded CGG repeats form a platform which further promotes to binding of other cellular proteins, reducing the formation of mature miRNAs subsequently causing neuron dysfunction. The other mechanism. 3) Repeat associated non-AUG (RAN) translation is the “toxic protein gain-of-function” model in which toxic protein contributes to the cellular cytotoxicity. In this proposed model, toxic CGG repeats transcript influence the normal initiation of protein translation, while it provokes to initiate translation without utilizing canonical start codon (ATG). In FXTAS, RAN translation start at the 5'UTR of *FMRI* mRNA that leads to the formation of cytotoxic homopolymeric protein, FMRpolyGlycine (FMRpolyG) and FMRpolyAlanine (FMRpolyA).

In most of the trinucleotide disorders, including FXTAS complete treatment are still not available. Currently, available therapeutics are based on specific symptoms of particular diseases. Thus, there is an urgent need to develop a therapeutic intervention which targets the pathogenic mechanisms, rather than

symptomatic treatment. The actual pathogenic triggers for all neurological disorders are triplet repeats expansion in DNA and RNA. Different therapeutic approaches have been explored previously based on different binding agents such as antisense oligonucleotide (ASOs), RNA interference (RNAi) and small molecules. ASOs and RNAi mediated correction of diseased pathology were reported in different microsatellite disorders. The ASOs and RNAi bind with the complementary sequence of repeating RNAs and make them prone to endonuclease digestion. In the case of FXTAS, these antisense sequences shield the CGG RNA structure and hinder the interaction of RNA binding proteins found within nuclear inclusion and also prevent the RANT event. However, poor penetrance of blood barrier, immunogenic response, non-specificity makes them less favorable for therapeutics. Also, specifically developed cellular and animal models should be used for testing of these ASOs and RNAi to come out with final conclusion of their efficacy and selectivity.

Small molecule-based therapeutics overcome such limitations of ASOs and RNAi. The repeats containing RNAs typically form A-type conformations with different mismatches which are targeted by either naturally available or chemically synthesized small molecules that selectively bind TNR RNAs structure. The unbiased screening of small molecule and structure-guided chemical synthesis has reported several potent small molecules against distinct TNR RNAs such as CUG, CAG, CGG, and CAG. These small molecules stabilize RNA hairpin structures and disrupt protein binding that results to reverse molecular disease hallmarks such as splicing defects, nuclear RNA foci, and homopolymeric protein aggregates. Here, we demonstrated structural insight of CGG repeat RNA and determined the binding potential of naturally available and synthetic small molecules which selectively interact with GG mismatch internal loops. Such selective targeting the pathogenic trigger (CGG repeat RNA) of FXTAS *via* small molecule would be a promising therapeutic intervention.

Chapter 1 includes a detailed review of the literature of trinucleotide repeat disorders, pathogenic disease mechanisms, and small molecules based therapeutic approaches.

Chapter 2 explains the material methods and techniques used in the present studies. Here, we have discussed the fundamental principle of techniques used to study the screening of small molecules, structural insight of RNA and small molecules interaction using Nuclear Magnetic Resonance (NMR) spectroscopy, Molecular Dynamic (MD) simulation, Umbrella sampling MD simulation and Molecular Docking, Fluorescence-based binding assay, Isothermal Calorimetry (ITC) titration, Circular Dichroism (CD) spectroscopy, CD and/or Ultraviolet & Visible (UV) spectroscopy thermal denaturation study. Some gel-based techniques were also used, such as DNA & RNA electrophoresis and Polymerase Chain Reaction (PCR). Addition to this, some cellular assays were performed to assess the efficacy of small molecules such as western blot assay, semi real-time PCR, protein aggregation reduction assays, cell viability enhancement assay, confocal imaging microscopy, and small molecule uptake assay.

Chapter 3 describes the understanding of molecular dynamic (MD) governed by GG 1x1 non-canonical internal loops in CGG repeats RNA. We evidenced the importance of structural information of the expanded repeats system to develop a potential therapeutic lead molecule for the treatment of these expanded repeats associated neurological disorders. In this study, NMR spectroscopy, conventional MD simulation, and Umbrella sampling were performed to simulate the pathways of dynamic motions involved in non-canonical GG pairs. Apart from the NMR structure sequence [(5'-CCGCGGCGG-3')₂], two other different sequences were also simulated to shed lights on the dynamic behavior of the CGG motif. A (3xCGG) repeat sequence [(5'-GGGCCGGCGGCGGGUCC-3')₂] crystallized by Kumar et al. 2012 and modeled sequence [(5'-GCGCGGCGC-3')₂] was used to run a dynamic study in explicit

solvent using AMBER force field. Simulation of all these sequences was performed up to 1 μ s and elaborated the fluctuation of the GG mismatch base pairs, nature, and the root mean time evolution (RMSD) of the base flipping, ions distribution, and backbone torsional angle. Apart from the 1 μ s long molecular dynamics, a 2D free energy landscape was constructed with the help of umbrella sampling molecular dynamics along with Weighted Histogram Analysis Method (WHAM) to figure out the minimum structures and possible transition pathways between them. NMR spectroscopy data showed the broadening and shifting of G6NH and G5NH imino proton resonance peaks with increasing temperature from 293K to 318K. G5H1' resonance peak was appeared at a higher temperature, while the same peak was not present at a lower temperature. Further, 2D NOESY 1 H NMR spectra of similar sequences were depicted the dynamics near internal GG pair.

Additionally, MD simulation showed a significant change in RMSD value of (CGG)₁ and (CGG)_{1m}, with an initial *anti-anti* conformation. In contrast, the modeled CGG where *anti-syn* was the starting conformation showed a very stable RMSD. In continuation, MD simulation of (CGG)₃ repeats RNA, where initial GG mismatch was in *syn-anti* conformation was remaining in initial position during the simulation. Na⁺ binding pocket was only found in the *syn-anti* conformation of (CGG)_{1m} and (CGG)₃. The *anti-anti* and *syn-anti* conformers possessed one and two hydrogen bonds on an average respectively that makes *anti-syn* structures less flexible. Distribution of the glycosidic angle (χ) suggested the initial state was conserved in all the cases, i.e., *anti-anti* remains in *anti-anti* or same for *syn-anti*, which is a common feature of all the distributions. Guanidine of GG mismatched kept its initial state of conformation and fluctuated only in that state of conformation throughout the simulation.

Further, the direction of base flipping revealed the motion of GG mismatched towards minor groove and major groove (opposite) in the case of (CGG)₁ and (CGG)_{1m} (*anti-anti*). In contrast, (CGG)_{1m} and (CGG)_{1m} (*syn-anti*),

very less flexibility during the simulation, instead flipping was mostly directed around the minor groove. Free energy minima generated by principal component analysis (PCA) showed two slightly separable equiprobable minima structures of the NMR sequence and the model *anti-anti* sequence. On the contrary, the *syn-anti* conformation of (CGG)_{1m} and (CGG)_{1m} was showed only a single minimum. Additionally, umbrella sampling MD simulation was elucidated four distinct minimum regions, P1 & P4 (*anti*), P3 & P4 (*syn*), separated by the high energy barrier. The possible pathway of transition only in between P1 to P4 and P2 to P3 because of fewer energy barriers, apart from this there is no physically possible pathway for the transition from *anti-anti* to *anti-syn* at room temperature. In conclusion, this study would help in understanding the RNA structural dynamics-function relationship. It will also allow designing novel therapeutics compounds, those recognize and bind with the mismatch RNA repeats which eventually leads to the inhibition of the interaction between RNA-protein complexes by blocking the protein binding site on RNA and improved the defects caused by repeats RNA.

Chapter 4 describes the therapeutic potential of a naturally available small molecule against trinucleotide disorders. CGG trinucleotide expansion (55-200) in 5'UTR of fragile X mental retardation 1 (*FMRI*) gene located on the X chromosome causes FXS, FXTAS, and FXPOI. Targeting this toxic RNA *via* small molecules could be a promising therapeutic strategy for these devastating disorders. We explore the therapeutic efficacy of Piperine, an active alkaloid isolated from black pepper (*Piper nigrum*). *Piperine* possesses numerous pharmacological properties like anticarcinogenic, hepatoprotection, antioxidant, *nephroprotective*, anti-inflammatory, antiapoptotic, antidepressant, and neuroprotective. Knowing the significant therapeutic properties of Piperine on human health, we sought to explore its therapeutic efficacy for CGG associated neurological disorders (FXTAS, FXPOI, and FXS) by determine its ability to regulate the CGG RNAs hairpin structure and functions. To achieve this, first binding affinity of Piperine were determined with different combination of single

trinucleotide repeat RNAs using fluorescence titration assay. Conclusively, Piperine found as a strong binder with CGG RNA over other mismatched RNAs and AU duplex RNA. Furthermore, higher repeat duplex RNA & other mismatched RNAs, yeast t-RNA, G-quadruplex forming DNAs and duplex DNA were served as control. Further, the selectivity of Piperine was checked with higher CGG repeat RNAs. Moreover, the binding affinity of Piperine molecule with CGG repeats RNAs was confirmed by Isothermal calorimetry titration (ITC) assay. Additionally, interaction of Piperine with CGG repeat RNAs was examined by other biophysical techniques such as Circular Dichroism (CD) spectroscopy, CD melting assay, Proton Nuclear Magnetic Resonance (NMR) spectroscopy, gel retardation assay, and taq-polymerase inhibition assay.

CD spectroscopy detected the significant spectral changes in the negative peak of CGG repeats RNAs during titrated with Piperine. Interestingly, the thermal profile portrays the stabilizing effect of Piperine with CGG repeats RNA, that suggested the formation of stable complex. NMR spectroscopy and molecular docking of Piperine molecule with CGG repeat RNA demonstrated the atomistic detail of interactions. Gel mobility shift assays also found the considerable retardation in the mobility of CGG RNA with Piperine over duplex RNA control. PCR inhibition assay revealed that Piperine arrested the movement of DNA polymerase movement during amplification. Further selectivity of Piperine for CGG RNA was compared with previously reported nucleic acid binder TMPyP4 and TMPyP2 using similar biophysical assay. Finally, the therapeutic potency of Piperine was extensively characterized by analyzing cellular assays such as improvement of pre-mRNA alternative splicing defect and reducing FMRPolyG protein aggregates formation in CGG expressing FXTAS cell model system. Piperine was also found to enhance the cell viability of CGGx99 plasmid transfecting cells. Cell cytotoxicity of Piperine was also determined using established cell line such as HEK293, normal fibroblast and patient-derived fibroblast. All the findings suggested that Piperine could serve as a potential therapeutic molecule for the ameliorating pathogenic condition of FXTAS.

Chapter 5 describe the similar rational approach as used in chapter 4 to identify the therapeutic potential of an already known natural compound, Curcumin. As previously reported that trinucleotide repeat disorders (TNR) are caused by the repetitive expansion of triplet repeats above the threshold either coding or non-coding region on the gene. Generally, all the TNR disorder shows common pathogenic mechanisms despite different nucleotide repeat expansions. In this study, we have examined a naturally abundant biologically potent small molecule, Curcumin, that selectively target r(CGG)^{exp} repeats RNA. To accomplish that we screened, Curcumin, a polyphenol used as a traditional herbal medicine against different 5'CNG/3'GNCx1 single repeat-containing RNAs, using fluorescence binding assay, where **N** stands for either A, G, U or C bases. Curcumin found as a potent binder with 1x1 GG motifs. Further, binding affinity and selectivity was confirmed with long CGG repeat containing, r(CGGx20), r(CGGx40) and r(CGGx60) RNA. Moreover, different mismatch RNA (AA, CC, and UU), duplex RNA (AUx1 & AUx6), yeast t-RNA, G-quadruplexes (c-kit, cmyc, tel22 and bcl2) and duplex calf thymus (CT) DNA controls were used as a control to check the selectivity of Curcumin for CGG motif. Binding studies revealed Curcumin is a better binder with CGG RNA, supporting a high selectivity of Curcumin for CGG motif RNA in the nanomolar range. ITC, another screening technique was used to confirm the selectivity and affinity of this interaction. The high association constant (K_a) values of Curcumin with different CGG repeats RNAs compared others mismatched and duplex RNA (AUx1/6) evident the selective interaction of Curcumin for CGG repeats RNA motifs.

Moreover, we have further validated the binding potential and selectivity of Curcumin with multiple r(CGG)^{exp} RNAs using various biophysical methods such a gel retardation assay, Circular Dichroism spectroscopy, CD thermal denaturation, and DNA polymerase inhibition assay. Significant retardation in the mobility of CGG RNA in the presence of Curcumin suggested strong complex

formation due to the high interaction. PCR assay also concluded inhibition of polymerase activity during extension in the presence of Curcumin. CD spectroscopic analysis found the significant hypochromic and redshift in the negative peak of different CGG repeat-containing RNA over control RNA. In continuation, the thermal denaturation profile of CGG RNA also suggested that Curcumin thermodynamically stabilize 5'CGG/3'GGC repeats RNA hairpin structure. Further, we assessed the cellular potency of Curcumin for their capability to correct both toxic phenotypes in FXTAS established cellular models. Curcumin significantly improved pre-mRNA alternative splicing defect and reduced the protein aggregates, FMRpolyG, formation in FXTAS developed cell model. Curcumin also enhanced the cell viability of r(CG) repeat-containing cell models. Our studies conclude that Curcumin potently improves FXTAS associated defects via interacting with CGG motif and could be used for the therapeutic development of FXTAS.

Chapter 6 describes the screening of synthetic compound for the therapeutic development of CGG associated trinucleotide repeats disorders. Here, we utilized the shape similarity approach, a key concept in drug discovery. Bisantrene hydrochloride, an anthracene derivative, is a small molecule reported for the effective treatment of metastatic cancer. Previous studies in our lab identified bisantrene as a better binder for CGG repeat RNA. It is considered that structurally similar molecule generally have similar properties. Synthetic drug with very less cytotoxicity and high binding affinity could be used for the therapeutic of FXTAS and FXPOI. Thus, Bisantrene has taken as a query molecule for the shape-based similarity search in national cancer institute (NCI) small molecules library. The molecules which found the highest totipotency score were taken for further studies. Bisantrene similar molecules were screened against 5'CGG/3'GGCx1 and 5'CAG/3'GUCx1 RNA. Five bisantrene similar lead molecules were found to have better binding compared to AU duplex control. Further, the binding affinity of five lead molecules was scrutinized 5'CGG/3'GGCx6 and 5'CAG/3'GUCx6 RNA. Out of five three lead molecules

B1, B4 and B11 showed effective and consistent binding with CGG RNA over AU duplex RNA. In addition, the selectivity of these three lead molecules was determined with higher CGG repeat RNA. Yeast t-RNA, different G-quadruplex DNA and CT duplex DNA were served as control.

To reaffirm the binding affinity of lead molecules with CGG RNA isothermal calorimetry titration assay was performed. Lead molecules showed approximately a hundred folds better binding with CGG RNA compared to AU duplex RNA. Structural variations in CGG RNA hairpin structure in the presence of lead molecules were analyzed using CD spectroscopy. CD spectrum depicts the hypochromic and redshift in negative peak of different CGG repeats RNA with the gradual addition of lead molecules whereas no significant changes were detected in AU duplex RNA. UV-thermal denaturation study of CGG RNA in presence of lead molecules also showed incrementing melting temperature (T_m) profile of CGG RNA over duplex RNA suggested lead compounds was stabilizing the CGG RNA hairpin motif. Interaction of CGG RNA with lead molecules was further evaluated by utilizing gel-based biophysical techniques such as gel mobility shift assay and Taq-polymerase stop assay. Data of both the assay confirmed the interaction between CGG RNA and lead molecules is significant compared to AU duplex RNA.

In-cellulo potency of lead molecules for inhibition of FMRpolyG protein aggregates and improvement of pre-mRNA alternative mis-splicing defect in FXTAS cultured cell. Treatment of lead molecules correct the pre-mRNA splicing defects near to wild type higher concentration without affecting the global splicing efficiency of different genes. Similar to splicing defect correction, treatment of lead molecules reduced the number and size of polyG protein aggregate formation without affecting the expression of control plasmid, lacking CGG repeat. Importantly, the cell viability of toxic plasmid transfected cell was enhanced with the addition of lead molecules. Furthermore, these lead compounds exert no cytotoxicity at the concentration used in biological assays. Cellular data

implies that binding of lead molecules interrupt the sequestrations of splicing regulator and also impair translational machinery to start translation from non-canonical mechanism. In conclusion, our data suggest that Bisantrene similar lead molecules could be used as a potential candidate for the therapeutics of FXTAS.

Chapter 7 describes the future scope and conclusion of the present work. As trinucleotide repeat disorders follow the common pathogenic mechanism despite different triplet repeats expansion. Since past years, numerous underlying pathogenic mechanisms of trinucleotide disorders explored vigorously to target TNR pathogenesis. There is still a need to identify the unanswered mechanistic and chronological best targets for promising therapeutics. Current, treatment scenario of TNR disorders is symptomatic. Several case studies show improvement of particular symptom in diseased individuals. It would be beneficial if we could target the actual trigger of these diseases. Repeat containing gene and its transcript is the straight forward trigger for TNR disorders. The most encouraging therapeutic approaches are reducing the expression of mutant RNA and mutant protein by aiming for DNA or RNA transcripts. TNR repeats length and its probable secondary structure in DNA and/or RNA is concern between normal and toxic transcripts for the selective and specific development of therapeutic agents. Different types of agents were explored to target the nucleotide repeat-containing nucleic acid, overcome the disease pathology.

The small molecule-based therapeutic strategy was used to selectively target the different mismatched hairpin RNAs motifs. Many of such small molecules are in clinical trial phases. Now, current therapeutic intervention strategies of many diseases like cancer, infectious disease, and immune disorders, etc., utilize a combination of drugs for different selective target instead of targeting single component in regulatory circuits. Thus, combination therapies for TNR associated neurological disorders would be one of an advantageous approach. Other targeted strategies have been using antisense reagents such as

RNA interference (RNAi) and antisense oligonucleotides (ASOs) also called cutter, bind to complementary sequences and prone them for endonucleases digestion. However, lack of optimum cellular penetration, intracellular trafficking, and immunogenic response makes them less efficient agents. Additionally, the CRISPR/Cas9 gene-editing system was used to remove TNR expansions from mutant allele with the aim of reducing the toxic condition. However, this a novel approach and still needed development.

The general aim of this thesis is to investigate the structural characteristic of GG mismatched RNA and develops a small molecule-based therapeutic intervention for FXTAS/FXPOI. In this study, several structural specific studies like NMR spectroscopy and molecular dynamics simulation were successfully used to resolve the high-resolution structure and gaining dynamic involve during non-canonical interaction such as the interaction between mismatched base pairs and between RNA & ligand. These small structures may simulate the understanding of various biological complex interactions in the future. Further improvement in this direction would help for the better understanding of RNA-trigger pathophysiology in TREDs and also advance the rational design of repeat targeted small molecules. Till date, there is no targeted treatment for CGG associated disorders. We have also explored already known neuroprotective naturally available small molecules which have the potential to ameliorate FXTAS disease pathology. Further efforts are going on to enhance association rate and sequence & structure-specific selectivity of small molecules for therapeutic intervention of FXTAS. Apart from this, targeting multiple pathogenic pathways simultaneously in the regulatory circuit via combination drug therapy would produce stronger treatment for the individual.

LIST OF PUBLICATIONS

(A) Publication from the Thesis

(a) Published

1. Arun Kumar Verma; Eshan Khan; Subodh Kumar Mishra; Neha Jain; and Amit Kumar*. Piperine modulates protein mediated toxicity in Fragile X-associated tremor/ataxia syndrome through interacting with expanded r(CGG)^{exp} RNA, *ACS Chemical Neuroscience* DOI: 10.1021/acschemneuro.9b002822.

2. **Arun Kumar Verma**; Eshan Khan; Soniali Bhagwat; and Amit Kumar*. Exploring the potential of small molecules based therapeutic approaches for targeting trinucleotide repeats Disorders, *Molecular Neurobiology*, DOI: 10.1007/s12035-019-01724-4.

(b) Under peer review

1. **Arun Kumar Verma**; Subodh Mishra; Rajarshi Roy; Eshan Khan; Arpita Tawani; Biplob Ghosh; Parimal Kar; and Amit Kumar*. Structural insights reveal the dynamics of the repeating r(CGG)^{exp} transcript found in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome (FXTAS), *Journal of Chemical Information and Modeling*, (Manuscript ID: ci-2019-00583s).

2. **Arun Kumar Verma**; Eshan Khan; Subodh Kumar Mishra; Amit Kumar* Curcumin selectively target toxic r(CGG)^{exp} RNA and ameliorates Fragile X-associated tremor ataxia pathogenesis, *ACS Chemical Neuroscience* (Manuscript id: cn-2019-002102).

(B) Publication apart from the Thesis

(a) Published

Eshan Khan; Arpita Tawani; Subodh Kumar Mishra; **Arun Kumar Verma**; Arun Upadhyay; Mohit Kumar; Rajat Sandhir; Amit Mishra; and Amit Kumar* (2018). Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease

(HD) and Spino Cerebellar Ataxia (SCAs), *ACS Chemical Biology*, 13(1), 180-188; doi: 10.1021/acscchembio.7b00699.

TABLE OF CONTENT

1. LIST OF FIGURES	xxxvii-liv
2. LIST OF TABLES	iv-lviii
3. ACRONYMS	lix-lxi

Chapter 1: Introduction.....	1-71
1.1 Microsatellites and repeat instability	1
1.2 Microsatellite repeats expansion diseases.....	2
1.3 Fragile X Syndromes (FXS)	11
1.3.1 Clinical features and symptoms	14
1.3.2 Molecular mechanism/pathophysiology of FXS	16
1.3.3 Therapeutic approaches	19
1.3.3.1 Targeting Glutamatergic system	21
1.3.3.2 Targeting GABAergic system	24
1.3.3.3 Targeting Epigenetic modulator	27
1.4 Fragile X-associated tremor/ataxia syndrome (FXTAS)	35
1.4.1 Clinical features of FXTAS/Premutation diseased condition	38
1.4.2 Molecular mechanism and diseased pathology of FXTAS.....	39
1.4.2.1 RNA toxicity via sequestration of RNA binding proteins	39
1.4.2.2 Repeat associated non-AUG translation (RANT)	43
1.4.3 Therapeutic development of FXTAS	46

1.5 Aim of the thesis	50
1.6 Reference	54

Chapter 2: Material, Methods and Instrumentation.....73-112

2.1 Material	73
2.2 Sample preparation and methodology	77
2.2.1 Nuclear Magnetic Resonance (NMR) spectroscopy study	77
2.2.2 Preparation of model systems for Molecular Dynamics simulations	78
2.2.3 Umbrella Sampling Molecular Dynamics simulations	79
2.2.4 RNA preparation using in-vitro transcription	80
2.2.5 Fluorescence titration assay	80
2.2.6 Isothermal Calorimetric titration assay (ITC).....	81
2.2.7 Circular Dichroism spectroscopy assay	81
2.2.8 Thermal denaturation study by UV spectroscopy.....	82
2.2.9 Gel retardation assay	82
2.2.10 PCR stop assay.....	83
2.2.11 Real time PCR.....	83
2.2.12 Western blot assay and protein aggregate quantification	84
2.3 Principle behind the assay and the related Instrumentation.....	85
2.3.1 In-vitro transcription for RNA synthesis	85
2.3.2 Fluorescence binding assay.....	88
2.3.3 Isothermal titration calorimetric assay (ITC).....	90
2.3.4 Circular Dichroism spectroscopy.....	92

2.3.5 Nuclear Magnetic Resonance (NMR) spectroscopy	94
2.3.6 Gel retardation assay	96
2.3.7 Polymerase Chain Reaction (PCR) stop assay.....	98
2.3.8 Molecular Docking	100
2.3.9 Cell viability assay (MTT assay)	100
2.3.10 Western blot assay	103
2.4 Reference	106

Chapter 3: Structural insights reveal the dynamics of the repeating r(CGG)^{exp} transcript found in Fragile X syndrome and Fragile-X associated tremor/ataxia syndrome (FXTAS).....113-158

3.1 Introduction.....	113
3.2 Results & Discussion	117
3.2.1 One dimensional proton (1H) NMR spectra of [(5'-CCGCGGCGG-3') ₂] at variable temperatures revealed the dynamics at noncanonical GG hair-pin loop	117
3.2.2 Dynamics of (CGG) ₁ , (CGG) _{1m} and (CGG) ₃ structure.....	121
3.2.3 Principle component analysis (PCA) revealed the transitional dynamics	140
3.2.4 Umbrella sampling revealed the probable pathways for Anti-Anti ⇌ Anti-Anti and Anti-Syn ⇌ Anti-Syn GG transformation in (CGG) ₁	143
3.2.5 Target CGG repeat motif RNA with small molecule	145
3.2.6 Insights into CGG motif binding proteins	145

3.3 Material and Methods	146
3.3.1 RNA sample preparation.....	146
3.3.2 Nuclear Magnetic Resonance (NMR).....	146
3.3.3 Preparation of Model systems for Molecular Dynamics simulations ..	147
3.3.4 Umbrella Sampling Molecular Dynamics Simulations	148
3.3.5 Analysis.....	149
3.4 Conclusion	150
3.5 References.....	152

Chapter 4: Piperine modulates protein mediated toxicity in Fragile X-associated tremor/ataxia syndrome through interacting expanded (r(CGG)^{exp})RNA.....159-221

4.1 Introduction.....	159
4.2 Results & Discussion	161
4.2.1 Probing binding of Piperine to 1x1 RNA motifs and r(CGG)exp RNAs	161
4.2.2 Isothermal titration calorimetry (ITC) studies for the interaction of Piperine with r(CGG)exp RNAs.....	175
4.2.3 Elucidation of conformational changes of r(CGG)exp RNA with Piperine using Circular Dichroism spectroscopy.....	184
4.2.4 Thermal denaturation analysis of r(CGG)exp RNAs with Piperine	187
4.2.5 NMR broadening experiments to analyze the interactions of Piperine with r(CGG)exp RNAs	191

4.2.6 Gel mobility shift assay and PCR stop assay of r(CGG) ^{exp} with Piperine	193
4.2.7 Improvement of alternative pre-mRNA splicing defects in FXTAS cellular models	198
4.2.8 Studying the cellular potency of Piperine: Inhibition of RAN translation	202
4.3 Material and Methods	206
4.3.1 Reagents	206
4.3.2 RNA Preparation and Purification	206
4.3.3 Fluorescence titration assay	207
4.3.4 Isothermal titration calorimetry experiment	207
4.3.5 Circular Dichroism spectroscopy assay	208
4.3.6 RNA thermal denaturation study	208
4.3.7 Nuclear magnetic resonance (NMR) spectroscopy	209
4.3.8 Molecular Docking	209
4.3.9 Gel retardation assay	210
4.3.10 PCR stop assay	210
4.3.11 Improvement of splicing defect in FXTAS cellular model	211
4.3.12 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of Piperine by using MTT assay	212
4.3.13 Quantification of canonical GFP and non-canonical FMRpolyG-GFP protein Inclusions	213
4.4 Conclusion	214
4.5 Reference	215

Chapter 5: Curcumin regulates the r(CGG)^{exp} RNA hairpin structure and ameliorate defects in Fragile X-associated tremor ataxia syndrome.....223-268

5.1 Introduction..... 223

5.2 Results and Discussion 225

 5.2.1 Binding affinity and selectivity of Curcumin with different trinucleotide RNA repeats..... 225

 5.2.2 Isothermal calorimetry titration of Curcumin with CGG repeats RNA233

 5.2.3 Gel mobility shift assay, PCR stop assay, and Molecular docking of r(CGG)^{exp} with Curcumin 236

 5.2.4 Curcumin induced conformational changes in CGG repeats RNA 239

 5.2.5 CD thermal denaturation study of CGG repeats RNA with Curcumin 241

 5.2.6 Curcumin improves alternative pre-mRNA splicing defects in FXTAS cellular models 243

 5.2.7 Curcumin reduces RAN translation of the CGG repeats 247

5.3 Material and Methods 253

 5.3.1 Reagents..... 253

 5.3.2 RNA Preparation and Purification 253

 5.3.3 Fluorescence titration assay 254

 5.3.4 Circular Dichroism spectroscopy assay 254

 5.3.5 RNA thermal denaturation study 255

 5.3.6 Isothermal titration calorimetry experiment 255

 5.3.7 Molecular Docking 256

 5.3.8 Gel retardation assay 256

5.3.9 PCR stop assay.....	257
5.3.10 Improvement of splicing defect by Curcumin in FXTAS cellular model	257
5.3.11 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of Curcumin by using MTT assay.....	260
5.3.12 Quantification of RAN translation by western blot.....	260
5.3.13 Quantification of canonical GFP and non canonical FMRpolyG-GFP protein Inclusions.....	261
5.4 Conclusion	261
5.5 Reference	263

**Chapter 6: Chemical screening identifies compounds that mitigate
r(CGG)exp RNA mediated toxicity in Fragile X-associated tremor
ataxia models.....269-334**

6.1 Introduction.....	269
6.2 Results and Discussion	271
6.2.1 Virtual screening for Bisantrene similar molecules.....	271
6.2.2 In-vitro selectivity of Bisantrene similar compound against different trinucleotide repeats RNAs.....	272
6.2.3 Isothermal calorimetry analysis of B1, B4 and, B11 with CGG repeats RNAs.....	289
6.2.4 Conformational analysis of CGG RNA in the presence of the lead molecules	295
6.2.5 Thermal denaturation profile of CGG repeats RNAs with B1, B4 and B11 molecules.....	298

6.2.6 Electrophoretic mobility shift and polymerase chain reaction (PCR) stop assay of r(CGG) ^{exp} RNAs with lead compounds	301
6.2.7 Structural insight of CGG repeats RNA and B1, B4, and B11 molecule complex using NMR spectroscopy and molecular docking	304
6.2.8 In-cellulo potency of lead molecules inhibition of RAN translation in FXTAS developed cell model.....	307
6.2.9 Improvement of a splicing defect in FXTAS cellular model.....	312
6.3 Material and Methods	317
6.3.1 Reagents	317
6.3.2 Invitro transcription and RNA Purification	317
6.3.3 Fluorescence binding assay.....	318
6.3.4 Isothermal calorimetry titration (ITC) assay.....	318
6.3.5 Circular Dichroism spectroscopy experiment.....	319
6.3.6 RNA thermal denaturation experiment.....	319
6.3.7 Nuclear magnetic resonance (NMR) spectroscopy titration assay	320
6.3.8 Molecular Docking of lead molecule with CGG RNA.....	320
6.3.9 Electrophoretic mobility shift assay.....	321
6.3.10 Polymerase chain reaction inhibition assay	321
6.3.11 Quantification of RANT product by western blot	322
6.3.12 Quantifying of non canonical FMRpolyG-GFP protein aggregates ..	322
6.3.13 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of lead compound by using MTT assay	323
6.3.14 Assess splicing defect correction by lead compounds in cell cellular model.....	324
6.4 Conclusion	327

6. 5 Reference 328

Chapter 7: Conclusion and Future Perspective335-336

7.1 Conclusion of the thesis 335

7.2 Future scope 336

LIST OF FIGURES

Chapter 1: Introduction

Figure 1.1 Diagram representing position of repetitive sequence found on the gene that contributes to repeat-associated neurological disorders.	10
Figure 1.2 Overview of FMR1 gene comprises 17 exons that can undergo pre-mRNA alternative splicing in the 3' region of the gene and produce FMR1 transcript and FMRP protein.....	13
Figure 1.3 Diagram showing clinical features of fragile X syndrome patient at the different periods.	15
Figure 1.4 Schematic representation of protein loss of function mechanism.	18
Figure 1.5 Effect of neuronal receptor signalling pathway on FMRP mediated regulation at synapsis.	20
Figure 1.6 Small molecules that target the glutamatergic signalling pathway.	23
Figure 1.7 Small molecules that target the GABAergic signaling pathway.	26
Figure 1.8 Small molecules that target the Epigenetic modulator.	29
Figure 1.9 Diagram represent FXS therapy development. a) Shows epigenetic modification by compound reactivate FMR1 gene expression. b) Depicting compound restoring deregulated pathway due to lack of FMRP.	30
Figure 1.10 Diagram represent CGG repeat associated neurological diseased condition.....	37
Figure 1.11 Diagrammatic representation of RNA gain of function mechanism.	42
Figure 1.12 Schematic representation of protein gain of function mechanism. ..	45
Figure 1.13 Diagram display therapeutic interventions for FXTAS treatment. ...	48

Figure 1.14 Small molecules that target the mGluR and GABA signaling pathway in FXTAS.	49
---	----

Chapter 2: Material, Methods and Instrumentation

Figure 2.1 (a) Consensus sequence of (class III and class II) T7 RNA polymerase (b) Representation of invitro transcription.	87
Figure 2.2 The Jablonski diagram depicting the fluorescence phenomenon.	89
Figure 2.3 Schematic illustration of Isothermal calorimetry instrumentation (left). Showing the example of the binding isotherm (Right).	91
Figure 2.4 Schematic representation of the Circular Dichroism spectroscopy instrumentation.	93
Figure 2.5 Schematic representation of Nuclear Magnetic Resonance spectroscopy.	95
Figure 2.6 Precessional motion of a spin in absence and presence of external magnetic field (M).	95
Figure 2.7 Diagrammatic illustration of gel retardation working mechanism.	97
Figure 2.8 Diagram showing principle of PCR stop assay.	99
Figure 2.9 Reduction of MTT into insoluble purple color crystals by mitochondrial reductase.	102
Figure 2.10 Schematic illustration of western blot assay and sample preparation.	105

Chapter 3: Structural insights reveal the dynamics of the repeating r(CGG)^{exp} transcript found in Fragile X syndrome and Fragile-X associated tremor/ataxia syndrome (FXTAS)

Figure 3.1 A) Showing sequence of RNA molecules used in molecular dynamics and umbrella sampling simulations. B) Temperature-dependent 1H NMR spectra for 5' r(CCGCGGCGG)₂ showing H' region (5.0-6.3 ppm) and imino region (10.5-13.5 ppm). 119

Figure 3.2 The portion of 300ms NOESY spectrum showing base to 10 regions for the sequence for 5'r(CCGCGGCGG)₂ at different temperatures A. 298 K and B. 318 K. 120

Figure 3.3 RMSD and Na⁺ distribution around mismatch base pairs of (CGG)₁ and (CGG)_{1m}. A1 and A2 are RMSD of GG mismatch pair and distance of Na⁺ ions from the GG mismatch pair of (CGG)₁. Similarly, B1, B2 and C1, C2 are RMSD and Na⁺ distribution of (CGG)_{1m}. Initial GG pairing in B1 and B2 is anti-anti and whereas syn-anti in case of C1 and C2. For each Na⁺ ion we use different colors for better visualization. 3D conformation of all three systems is extracted from the different region as shown in RMSD plot. Red, black and blue color in 3D conformation depicts three different colored regions in the plot. A3, B3 and C3 represents (CGG)₁, (CGG)_{1m} (anti-anti) and (CGG)_{1m} (syn-anti). 122

Figure 3.4 Na⁺ bound state in the different orientation of RNA structure from the initial anti-syn GG base pair as found in (CGG)_{1m}. 123

Figure 3.5 RMSD and Na⁺ distribution analysis of (CGG)₃. Root mean square deviation is shown in A backbone atom B1 G6-G29 base-pair C G7C8-G27C28 D1 G9-G26 base-pair E G10C11-G24C25 F1 G12-G23 base pair. B2, D2, F2 are the distance between Na⁺ ions and G6-G29, G9-G26, G12-G23 respectively. The residue numbering scheme is defined in Figure 1. 3D conformation corresponds to the G6-G29

(B3), and G9-G26 (D3) base pairs are also shown in the image. Different states are shown in the same color scheme used in the RMSD plot. 130

Figure 3.6 Na⁺ bound state in anti-syn GG base pairs as found in (CGG)₃. 131

Figure 3.7 Distribution of χ angle in four cases. In left hand side image, A) (CGG)₁ (anti-anti), B) (CGG)_{1m} (anti-anti) and C) (CGG)_{1m} (anti-syn) where brown corresponds to G5 and purple corresponds to G14. In Right hand side image D) G6 (brown)-G29 (purple) E) G9 (brown)-G26 (purple) F) G12 (brown)-G23(purple) of (CGG)₃ sequence. 134

Figure 3.8 Time evolution of centre of mass pseudo dihedral (CMPD) angle as a function of simulation time. Left-hand side: A) (CGG)₁ (anti-anti), B) (CGG)_{1m} (anti-anti) and C) (CGG)_{1m} (syn-anti) where purple is corresponds to G5 and brown corresponds to G14; Right-hand side: A) G6-G29 B) G9-G26 C) G12-G23 where purple corresponds to G6, G9 and G12 and brown corresponds to G23, G26 and G29..... 135

Figure 3.9 Backbone sugar torsion δ and the glycosyl torsion χ angle distribution of (CGG)₁, (CGG)_{1m} (anti-anti) and (CGG)_{1m} (anti-syn) in each row respectively where left hand side is corresponds to G5 and right hand side is G14. 136

Figure 3.10 Backbone sugar torsion δ and the glycosyl torsion χ angle distribution of (CGG)₃. A) G6, B) G29, C) G9, D) G26, E) G12 and F) G23... 136

Figure 3.11 Dynamical cross-correlated map of RNA; where (CGG)₁ and (CGG)₃ are the NMR and crystal structure respectively having single and tri CGG repeat. (CGG)_m and (CGG)_{m1} are the modelled structure with anti-anti and anti-syn initial conformation at the mismatch G base pair. X and y-axis of this plot colour indicates the atom indices. Red square shows the regions corresponds to the atoms of mismatch base pairs. 139

Figure 3.12 Free energy map corresponds to the principal components PC1 and PC2 for four different RNA systems. A) (CGG)₁ (anti-anti), B) (CGG)_{1m} (anti-anti), C) (CGG)_{1m} (anti-syn) and D) (CGG)₃. Two closely separated minima region observed in case of (CGG)₁ and (CGG)_{1m} (anti-anti) which also found from other analysis. 141

Figure 3.13 Principal component analysis showing the predominant motions of the regions: A) (CGG)₁(anti-anti), B) (CGG)_{1m} (anti-anti), C) (CGG)_{1m} (anti-syn), and D) (CGG)₃. 142

Figure 3.14 Two-dimensional potential of mean force (PMF) surface for (CGG)₁ system (PDB ID 2NCR). Here, x-axis represents torsional angle (χ) between base and sugar residues and y axis represents pseudodihedral angle (θ) for stacking. Four different minima structures are obtained through our umbrella sampling simulation denoted as P1, P2, P3, and P4. 144

Chapter 4: Piperine modulates protein mediated toxicity in Fragile X-associated tremor/ataxia syndrome through interacting expanded (r(CGG)^{exp})RNA

Figure 4.1 Secondary structure of different 1x1 nucleotide internal motif RNAs sequence. 164

Figure 4.2 Evaluation of binding affinity of Piperine with nucleic acids. a) Bar graph denotes K_d values of r(CNG)_{x1} RNA motifs with Piperine. b) Systemic representation of bar graph for K_d values analysis of different CGG repeat RNAs with Piperine. c) Bar graph represents K_d value analysis of different RNA and DNA controls with Piperine. d) The plot represents fluorescence titration assays two-mode curve fitting of r(CNG)_{x1} RNA motifs with Piperine. e) Plot depict fluorescence titration assay two-mode curve fitting of r(CGG)^{exp} RNAs and different RNA & DNA controls with Piperine. 165

Figure 4.3 Diagram shows the fluorescence titration assay plots of different RNAs motif with Piperine. Solid line curve represents the two-mode fitting. The K_{d1} and K_{d2} are also mentioned below. 168

Figure 4.4 Diagram shows the fluorescence titration assay plots of $r(\text{CGG})^{\text{exp}}$, $(\text{CNG})^{\text{exp}}$, AU paired RNA and DNA with Piperine. Solid line represents the two-mode fitting. The K_{d1} and K_{d2} are also mentioned below. 169

Figure 4.5 Fluorescence titration assay of TMPyP4 with expanded CGG RNA. a) Bar graph represents the dissociation constant values of TMPyP4 with $r(\text{CGG})^{\text{exp}}$ RNA. b) Binding plot of CGG RNA with TMPyP4. 170

Figure 4.6 Plot showing the binding curve of TMPyP4 with expanded CGG RNA. Dissociation constant values were determined using two mode fitting. 171

Figure 4.7 Fluorescence titration assay of TMPyP2 with expanded CGG RNA. a) Bar graph represents the dissociation constant of TMPyP2 with $r(\text{CGG})^{\text{exp}}$ RNA. b) Binding plot of CGG RNA with TMPyP2. ... 172

Figure 4.8 Plot showing the binding curve of TMPyP2 with expanded CGG RNA. Dissociation constant values were determined using two mode fitting. 173

Figure 4.9 Isothermal calorimetry titrations of Piperine with $r(\text{CNG})^{\text{exp}}$ and $r(\text{AU})^{\text{exp}}$ duplex RNA. (a) $r(\text{CGG})_6$, (b) $r(\text{CGG})_{20}$, (c) $r(\text{CGG})_{40}$, (d) $r(\text{CGG})_{60}$, (e) $r(\text{AU})_6$, (f) $r(\text{CAG})_6$, (g) $r(\text{CCG})_6$ and (h) $r(\text{CUG})_6$ RNA represents the titrated thermogram with Piperine. Solid line curve represents the two-mode binding best fit. ITC data showed significant interaction of Piperine with expanded CGG RNA over other RNA motifs..... 177

Figure 4.10 Isothermal calorimetry titration plot of TMPyP4 and TMPyP2 with expanded CGG repeat RNAs. 181

Figure 4.11 Circular Dichroism spectroscopy titration of r(CNG)^{exp} and r(AU)^{exp} duplex RNAs in the presence of Piperine. (a) r(CG)₁, (b) r(CG)₆, (c) r(CG)₂₀, (d) r(CG)₄₀, (e) r(CG)₆₀, (f) r(AU)₁, (g) r(AU)₆, (h) r(CAG)₆, (i) r(CCG)₆, (j) r(CUG)₆ RNA. In the presence of Piperine significant spectroscopic changes were detected with CGG repeat RNAs compare to other mismatched RNA motifs and AU duplex RNA. D/N denotes the drug by nucleotide ratio. 185

Figure 4.12 Circular dichroism spectroscopic titration of TMPyP4 with expanded CGG repeat RNA. On gradual addition of TMPyP4 caused hypochromic and red shift in the negative peak in CGG repeat RNA. D/N denotes for drug by nucleotide ratio. 186

Figure 4.13 Circular dichroism spectroscopic titration of TMPyP2 with expanded CGG repeat RNA. TMPyP2 caused no significant change in the CGG repeat RNA. D/N denotes for drug by nucleotide ratio. 186

Figure 4.14 A systemic representation of circular dichroism thermal profile of r(CNG)^{exp} and r(AU)^{exp} duplex RNAs with Piperine (a) r(CG)₁, (b) r(CG)₆, (c) r(CG)₂₀, (d) r(CG)₄₀, (e) r(CG)₆₀, (f) r(AU)₆, (g) r(CCG)₆, (h) r(CUG)₆, (i) r(CAG)₆ RNA. Piperine enhanced the melting temperature of expanded CGG RNAs whereas no significant change in T_m was observed with other mismatch RNA motifs and AU duplex RNA. D/N denotes the drug by nucleotide ratio. 188

Figure 4.15 Thermal denaturation curve of expanded CGG repeat RNA with TMPyP4. TMPyP4 decreased the melting temperature (T_m) of CGG repeat RNA that suggested it destabilize the CGG hairpin structure. 189

Figure 4.16 Thermal denaturation curve of expanded CGG repeat RNA with TMPyP2. TMPyP2 showed no significant T_m change as compared to Piperine. 190

Figure 4.17 One dimensional proton NMR spectroscopy of Piperine with r(CGG)_n RNAs and molecular docking study. (a) & (b) represents the NMR titration of 400µM of Piperine with (a) r(CGG)₁ and (b) r(CGG)₆ RNAs. (c) & (d) docking study of Piperine with CGG RNA (c) GxG-5 CGG RNA (d) GxG-2 CGG RNA. 192

Figure 4.18 Gel retardation assay and PCR stop assay. (a) Gel retardation plot shows an increasing concentration of Piperine significantly retards the mobility of CGG repeat RNAs over AU duplex RNA. (b) Gel images show the decreased intensity of PCR product with increasing concentration of Piperine than AU paired template. 194

Figure 4.19 Plot representing the percentage reduction in mobility of r(CGG)^{exp} RNA and PCR inhibition in presence of Piperine. Plot (a) r(CGG)₂₀, (b) r(CGG)₄₀, (c) r(CGG)₆₀ and (d) r(AU)₆ shows the retardation in mobility of expanded CGG RNA over AU duplex RNA. (e) (CGG)₁, (f) (CGG)₆, (g) (AU)₁ and (h) (AU)₆ shows the inhibition of PCR product of CGG template in presence of Piperine over AU duplex template. 195

Figure 4.20 Gel retardation and PCR stop assay of other mismatch RNA repeats with Piperine. (a) Piperine showed no significant retardation compared to CGG repeat RNA. (b) Similarly, no appreciable PCR inhibition detected with other mismatched RNAs. 196

Figure 4.21 Gel retardation and PCR stop assay of TMPyP4 and TMPyP2 with expanded CGG RNAs. (a-b) TMPyP4 showed significant retardation in bands while no retardation observed with TMPyP2. (c) TMPyP4 reduced the band intensity similar to Piperine whereas no inhibition detected with TMPyP2. 197

Figure 4.22 In-vitro efficacy of Piperine in FXTAS cellular models. Piperine improves FXTAS associated alternative pre-mRNA splicing defect of SMN2 and Bcl-x minigenes. (a) Improvement of SMN2 splicing

defects was found as a function of Piperine concentration (top, left). (b) Similarly, Bcl-x alternative splicing were improved with Piperine treatment in FXTAS cell model (top, right). (c) Interestingly, Piperine does not affect the GFP mRNA expression level. Error bar indicates standard deviation. One way ANOVA was utilized for statistical significance analysis where * $p < 0.05$; ** $p < 0.01$ 200

Figure 4.23 Quantification of alternative splicing defects of cTNT and IR minigenes in presence of Piperine (top gel image). CGGx99 and cTNT/IR co-transfected in HEK293 cell. Piperine treatment not shows any significant change in cTNT and IR splicing. 201

Figure 4.24 Effects of Piperine on protein aggregation & accumulation and its cytotoxicity caused by CGG repeats RNA (a-b) Micrographs of EGFP-CGGx99, transiently transfected HEK293 cells (a) 20x image and (b) 60x image, treated with control (DMSO) & Piperine for 24 hours. The arrowheads represent protein aggregates. (c) Few same sets of cells expressing expanded CGG repeats were used for quantitative analysis of the aggregates. (d) Cell viability assay for CGGx20-pcDNA3.1, CGGx99-GFP-pcDNA3.1 and GFP-pcDNA3.1 plasmid transfected cell (HEK293) in the presence of Piperine (e-f) Images represented the inhibition of FMRpolyG protein aggregates for 24 hrs in the presence of Piperine and vehicle (DMSO) while it does not significantly affect canonical translation of EGFP at same concentration. 204

Figure 4.25 Plot represents percentage cell viability of normal established cell line (HEK293), fibroblast cells (GM00357) and patient-derived cells (GM04026). 205

Chapter 5: Curcumin regulates the r(CGG)^{exp} RNA hairpin structure and ameliorate defects in Fragile X-associated tremor ataxia syndrome

Figure 5.1 Secondary structure of different 1x1 internal nucleotide motif (5'CNG/3'GNC) RNAs sequence..... 227

Figure 5.2 Elucidation of binding affinity (K_d) of Curcumin with nucleic acids using fluorescence based assay. a) Bar graph representing the binding constant (K_d) values of Curcumin with different 1x1 nucleotide internal motif RNAs. b) Bar graph showing the K_d values of Curcumin with different CGG repeat containing RNAs. c) Bar graph showing K_d values of Curcumin with different RNA and DNA controls. d) Plot illustrates fitting for fluorescence binding assay of Curcumin with different 1x1 nucleotide internal loop RNAs. e) Plot illustrate fitting for fluorescence binding assay of Curcumin with different CGG repeat containing RNAs along with different RNA and DNA controls. 228

Figure 5.3 Plots show the fluorescence titration assay of different 1 x 1 5'(CNG/3'GNC) RNAs motif with Curcumin. The dissociation constant (K_d) values are also mentioned below..... 229

Figure 5.4 Diagram shows the fluorescence titration assay plots of Curcumin with different CGG repeat RNAs and RNA ((AU)_{6x}, r(CAG)_{6x}, r(CCG)_{6x} & r(CUG)_{6x}) & DNA controls (c-kit DNA, cmyc DNA, bcl2 DNA, tel22 DNA & ct DNA). The dissociation constant (K_d) values are also mentioned below..... 230

Figure 5.5 Isothermal titration calorimetry of Curcumin with r(CGG)^{exp}, r(CNG)^{exp} and r(AU) duplex RNA. a) r(CGG)_{6x}, b) r(CGG)_{20x}, c) r(CGG)_{40x}, d) r(CGG)_{60x} and e) r(AU)_{6x}, f) r(CAG)_{6x} g) r(CCG)_{6x}, h) r(CUG)_{60x} represents the titrated thermogram of respective RNA with Curcumin. The exothermic peaks represent the

favorable contribution of energy during interactions. The following plots were fitted using two-mode binding fitting model. 234

Figure 5.6 Gel mobility shift assay, and PCR inhibition assay. Gel mobility shift assays plot shows significant retardation in mobility of CGG RNA with the increasing concentration of Curcumin; a) r(CG_G)_{20x} b) r(CG_G)_{40x} and c) r(CG_G)_{60x} RNA. In contrast no significant retardation detected in d) r(AU)_{6x} duplex RNA. PCR inhibition assay of CGG DNA template were performed with Curcumin as a function of concentration; decrease in the PCR band intensity suggest the binding of Curcumin with CGG Repeat (e) (CG_G)_{1x} and (f) (CG_G)_{6x}. In contrast, there is no such inhibition observed in PCR band intensity of AU paired duplex template g) (AU)_{1x} and (h) (AU)_{6x}..... 237

Figure 5.7 Curcumin docking study with CGG_{x3} RNA (PDB ID: 3JS2). a) Represent the dock image of G_xG-2 position. b) Represent the dock image G_xG-5 position. The best binding energy of G_xG-2 & G_xG-5 is -7.35 and -6.71 kcal/mol respectively..... 238

Figure 5.8 Circular Dichroism spectroscopy study of different r(CG_G) repeats containing RNAs and r(AU) duplex RNA as a function of Curcumin concentration. Plot a) r(CG_G)_{1x}, b) r(CG_G)_{6x}, c) r(CG_G)_{20x}, d) r(CG_G)_{40x} and e) r(CG_G)_{60x} shows spectral variation with the increasing concentration of Curcumin. Arrow heads denotes the shifting of negative and positive peaks upon drug titration. D/N denotes drug by nucleotide ratio. Each spectrum was recorded three times and average changes were plotted. 240

Figure 5.9 Diagram represents CD thermal denaturation profile of r(CG_G)_n repeats containing RNAs and r(AU) duplex RNA as a function of Curcumin concentration. Graph shows titration plot of a) r(CG_G)_{1x}, b) r(CG_G)_{6x}, c) r(CG_G)_{20x}, d) r(CG_G)_{40x}, e) r(CG_G)_{60x} and f) r(AU)_{6x} duplex RNA with increasing concentration of Curcumin.

Bar graph represents the melting temperature (T_m) values of respective RNA at different drug by nucleotide ratio (D/N). All the experiments were performed three times and the average values are presented..... 242

Figure 5.10 In-vitro efficacy of Curcumin in FXTAS cellular models. Curcumin improves FXTAS associated alternative pre-mRNA splicing defect of SMN2 and Bcl-x mini gene. Briefly, HEK cell were co-transfected SMN2 & Bcl-x mini gene with plasmid containing CGG repeat (CGGx99). SMN2 splicing regulation is controlled by DGCR8 and Sam68. Upon sequestration and inactivation of these proteins the splicing of SMN2 is dysregulated (a. top, left). When FXTAS cell model system treated with Curcumin, improvement of SMN2 splicing defects were restored to near to wild type at 100 μM as determined by semi-quantitative PCR. Similarly, Bcl-x gene alternative splicing is regulated by DGCR8 and Sam68 and mis-splicing of Bcl-x gene caused due to presenece of r(CGG)exp. were improved with Curcumin treatment in FXTAS cell models (b. top, right). Interestingly, Curcumin does not affect the GFP mRNA expression level (c. bottom). 245

Figure 5.11 Gel image (top) and quantification of alternative pre-mRNA splicing defect of cTNT mini-gene in CGGx99 transfected and healthy cells as a function of Curcumin concentration. Briefly, cTNT and CGGx99 plasmid (toxic) co-transfected in HEK293 cells. curcumin does not affect alternaive splicing defect of cTNT mini-gene..... 246

Figure 5.12 Diagram shows in-vitro efficacy of Curcumin in FXTAS cell models. In FXTAS, CGG expansion triggers the non-canonical protein translation leads to protein aggregate formation inside the cell. (a-b) Micrograph showing the effect of Curcumin treatment in transiently transfected ((CGGx99)-EGFP) HEK cell. (a) 20x image and (b) 60x image of DMSO (control) and Curcumin for 24hrs. (c) Blot showing

the reduction of polyG aggregates formation as a function of Curcumin concentration with control GAPDH and EGFP plasmid (lack CGG repeat). (d) Bargraph showing the percentage reduction of polyG aggregate in presence of Curcumin. (e) Plot denotes Curcumin treatment reduces number of protein aggregates in the cell. (f) Bar graph showing improvement of cell viability of CGGx20-pcDNA3.1, CGGx99-GFP-pcDNA3.1 and GFP-pcDNA3.1 plasmid transfected cell (HEK293) in presence of Curcumin. 249

Figure 5.13 (a) Diagram showing the EGFP expression of pcDNA3.1 plasmid (lack r(CGG)^{exp}) of HEK293 transfected cell in absence and presence of Curcumin. (b) Bar graph represents the normalized EGFP intensity of Curcumin treated and untreated cells. 250

Figure 5.14 Systemic representation of cell viability plot of Curcumin on eshtablished cell line for 24 hrs. 251

Figure 5.15 Cellular intake of Curcumin in patient derived cells (GM04026). a) 20x magnification b) 40x magnification. 252

Chapter 6: Chemical screening identifies compounds that mitigate r(CGG)exp RNA mediated toxicity in Fragile X-associated tremor ataxia models

Figure 6.1 Secondary strucutre of different 1x1 internal nucleotide motif (5'CNG/3'GNC) RNAs sequence. 277

Figure 6.2 Fluorescence-based screening of small molecules against single GG mismatched RNA and different RNA and DNA controls. a) Bargraph depicts dissociation constant (Kd) value analysis of Bisantrene with different mismatch single repeat RNA internal loop. b) Bargraph represents Kd values comparison of Bisantrene similar molecules with different mismatch single repeat RNA motif. c) Dissociated

constant value plot of five lead molecules with 5'CGG/3'GGCx6. d) Bargraph representing K_d values of final lead molecules with 5'CGG/3'GGCx20/40/60. e) Representation of K_d values plot of final lead compounds with different RNA and DNA controls. 278

Figure 6.3 Plots show the fluorescence titration assay of different 1 x 1 5'(CNG/3'GNC) RNAs motif with Bisantrene. The dissociation constant (K_d) values are also mentioned below..... 279

Figure 6.4 Diagram shows the fluorescence based titration assay plots of Bisantrene similar molecules with single repeat CGG RNA and AU duplex RNA. The dissociation constant (K_d) values are also mentioned below. 280

Figure 6.4 Conti.. Diagram shows the fluorescence based titration assay plots of Bisantrene similar molecules with single repeat CGG RNA and AU duplex RNA. The dissociation constant (K_d) values are also mentioned below. 281

Figure 6.5 Diagram shows the fluorescence based titration assay plots of lead molecules with CGGx6 repeats and AUx6 repeats duplex RNAs. The dissociation constant (K_d) values are also mentioned below..... 284

Figure 6.6 Diagram shows the fluorescence based titration assay plots of lead molecules with higher repeats (CGGx20, CGGx40, and CGGx60) RNAs. The dissociation constant (K_d) values are also mentioned below. 285

Figure 6.7 Plots represents the fluorescence based titration assay plots of lead compounds with different RNA (yeast t-RNA) and DNA controls (C-myc DNA, C-kit DNA, Bcl2 DNA, & CT DNA). The dissociation constant (K_d) values are also mentioned below..... 286

Figure 6.8 Plot represents the isothermal calorimetry titration normalized thermogram of B1, B4, and B11 with multiple CGG repeat RNA. The top panel of each thermogram indicates the raw data for automated

sequential injection of small molecules solution in the RNA solution. The bottom panel of each thermogram showed the normalized data point of each injection. a-d), e-f), and i-k) represent the ITC plots of B1, B4, and B11 with different CGG repeat RNAs and AUx6 RNA respectively. The exothermic peaks of all the plots represent the favorable contribution of energy during an interaction between the small molecule and CGG RNAs. The following plots were fitted using two-mode of binding model and solid curve showing is the calculated best fit data. 291

Figure 6.9 Plots represent the Circular Dichroism spectroscopy titrations of B1, B4, and B11 lead molecules with r(CGGe)exp RNA. Plots a) r(CGGe)20, b) r(CGGe)40, c) r(CGGe)60, and d) r(AU)6 RNAs represent titration spectrums with B1 molecule. Plots e) r(CGGe)20, f) r(CGGe)40, g) r(CGGe)60, and h) r(AU)6 RNAs showed titrated spectrums with B4 molecule. Plots i) r(CGGe)20, j) r(CGGe)40, k) r(CGGe)60, and l) r(AU)6 RNAs represent titration spectrums with B11 molecule. A significant conformation variation observed in CGG repeats RNAs hairpin structure in the presence of B1, B4, and B11. D / N denotes the drug by nucleic acid ratio. 296

Figure 6.10 Representative CD spectroscopy titration plots of CGG RNA with lead molecules 297

Figure 6.11 A systemic representation of thermal denaturation profile of r(CGGe)exp RNAs with B1, B4, and B11 molecules. Graphs a) r(CGGe)20, b) r(CGGe)40, c) r(CGGe)60, and d) r(AU)6 RNAs represent titrated melting curve with B1 molecule. Graphs e) r(CGGe)20, f) r(CGGe)40, g) r(CGGe)60, and h) r(AU)6 RNAs showed thermal denaturation profile in presence of B4 molecule. Graphs i) r(CGGe)20, j) r(CGGe)40, k) r(CGGe)60, and l) r(AU)6 represent the melting profile with B11 molecule. m) Bargraph showed the ΔT_m values of expanded CGG repeats RNAs with B1, B4, and B11

molecules. The enhancement in the T_m of CGG repeats RNAs was observed with the addition of B1, B4, and B11 molecules whereas no significant change detected in fully paired AUx6 RNA. D / N denote the drug by nucleic acid ratio. 299

Figure 6.12 Electrophoretic mobility shift assay of expanded CGG RNAs with lead compounds. a) B1, b) B4, and c) B11 images showed the increasing concentration of molecules retard in the mobility of r(CG_Gx20/40/60) repeats RNAs and AUx6 duplex RNA 302

Figure 6.13 a), b), & c) Representing gel image of PCR inhibition assay using CGGx6 and AUx6 as a template with increasing concentration of lead compounds..... 303

Figure 6.14 One dimension proton NMR spectroscopy titration of B1, B4, and B11 with CGGx6 RNA. a) B1, b) B4, and c) B11 has shown NMR titration with CGGx6 RNA. d) B1, e) B4, and f) B11 images represent the docking study with CGGx3 RNA (PDB ID: 3JS2). .. 306

Figure 6.15 In cellular efficacy of lead compounds in FXTAS developed cell models. In FXTAS, CGG triplet repeats expansion mediates the formation of homopolymeric protein aggregates, FMR1polyG inside the cells. a) B1, B4, and B11 treatment reduced the numbers and aggregation of FMRpolyG-GFP inclusions. Left, images represent the FMRpolyG-GFP inclusions in r(CG_G)x99-EGFP expressing COS7 cells with the vehicle (DMSO) and compounds B1, B4, and B11 at the mentioned concentration. Right, bar graph represents percentage reduction of inclusion numbers 24 and 48hrs post-treatment of a vehicle and lead compounds. b) FMRpolyG-GFP protein aggregates reduction was further confirmed using western blots. Blots represent the inhibition of protein inclusions with the treatment of B1, B4, & B11 and vehicle (DMSO) for 24hrs while it not considerably inhibits the canonical translation of EGFP. Normalized bar graphs have shown the quantification of respective blots. 309

Figure 6.16 Cell viability enhancement of toxic plasmid transfected cells (COS7) in presence of lead compounds.	310
Figure 6.17 Cell cytotoxicity assessment of lead molecules using patient derived (GM04026) cell line.	310
Figure 6.18 Cellular intake of lead compounds in patient derived cells (GM04026). a) 10x magnification b) 20x magnification.	311
Figure 6.19 In-vitro potency of lead compounds in developed FXTAS cellular models. Bisantrene similar molecules improve pre-mRNA splicing defect of SMN2 and Bcl-x minigenes associated with sequestration of Sam68 and DGCR8 by r(CGG)exp in a concentration-dependent manner. Briefly, COS7 cells were co-transfected with SMN2 & Bcl-x minigenes with CGGx99 repeat-containing plasmid. Splicing event of both SMN2 and Bcl-x genes regulated by DGCR8 and Sam68. a) & b) B1 improves both SMN2 and Bcl-x splicing defect as a function of concentration. Statistically, significant improvement was observed of B1 at a higher concentration. B4 and B11 also improve the SMN2 and Bcl-x splicing defect towards wild-type phenotype in a similar manner as B1. c) Represent the normalized CGGx99-GFP mRNA expression in the presence of lead compounds.	314
Figure 6.20 Gel image (top) and quantification of pre-mRNA splicing defect of SMN2 and Bcl-x mini-genes with plasmid lacking CGG repeats in transfected and untransfected cells as a function of compound concentration. Briefly, SMN2 and Bcl-x and EGFP plasmid (lacking CGG repeat) co-transfected in COS7 cells. lead compounds does not significantly affect wildtype expression of SMN2 and Bcl-x mini-genes.	315
Figure 6.21 Gel image (top) and quantification of alternative pre-mRNA splicing defect of cTNT mini-gene in CGGx99 transfected and healthy cells as a function of lead compounds concentration. Briefly, cTNT and	

CGGx99 plasmid (toxic) co-transfected in COS7 cells. lead
compounds does not affect alternaive splicing defect of cTNT mini-
gene. 316

LIST OF TABLES

Chapter 1: Introduction

Table 1.1 Neurological disorders caused by unstable nucleotide expansions.	4
Table 1.2 Trinucleotide repeat expansion disorders sorted according to repeats found in coding and non-coding region of the gene.....	8
Table 1.3 Small molecules reported for the treatment of FXS and FXTAS.....	31

Chapter 2: Material, Methods and Instrumentation

Table 2.1 DNA sequence template used for in-vitro transcription.	75
---	----

Chapter 3: Structural insights reveal the dynamics of the repeating r(CGG)^{exp} transcript found in Fragile X syndrome and Fragile-X associated tremor/ataxia syndrome (FXTAS)

Table 3.1 Sugar and backbone torsional angles (°) calculated for (CGG) ₁	124
Table 3.2 Distances (Å) and angles (°) of atoms for different base pairs of (CGG) ₁	125
Table 3.3 Global helical parameters calculated for the base pairs of (CGG) ₁ ...	126
Table 3.4 Helical parameters for different base pairs and steps of (CGG) ₁	127
Table 3.5 Minor and major groove widths according to refined P-P distances for the direction of sugar-phosphate backbone in (CGG) ₁	127
Table 3.6 Helical parameters for different base pairs and steps of (CGG) ₁	128
Table 3.7 H-bond percentage occupation different CGG mismatch sequences .	138

Chapter 4: Piperine modulates protein mediated toxicity in Fragile X-associated tremor/ataxia syndrome through interacting expanded (r(CGG)^{exp})RNA

Table 4.1	The dissociation constant (Kd) values obtained from flouresecne titration assay of Pipeirne with different (5'CNG/3'CNG)x1 RNA motifs with two mode curve fitting.	166
Table 4.2	The dissociation constant (Kd) values obtained from flouresecne titration assay of Piperine with r(CGG)exp, r(CNG)6, and AU paired RNAs and DNA controls with two mode curve fitting.	167
Table 4.3	The dissociation constant (Kd) values obtained from flouresecne titration assay of TMPyP4 with r(CGG)exp RNAs. The curves were fitted using two mode binding.	174
Table 4.4	The dissociation constant (Kd, μ M) values obtained after titrating r(CGG)exp RNAs and TMPyP2 using flourosence based binding assay. Two mode binding method were used for curve fitting.	174
Table 4.5	Thermodynamic parameter values of r(CGG)exp, r(CNG)exp and r(AU)6 duplex RNAs with Piperine obtained from ITC using two mode curve fitting.	178
Table 4.6	Sequential mode of fitted association constant values of expanded CGG repeats RNA with Piperine obtained from ITC.	180
Table 4.7	Thermodynamic parameter values of r(CGG) ^{exp} RNAs with TMPyP4.	182
Table 4.8	Thermodynamic parameter values of r(CGG)exp RNAs with TMPyP2.	183

Chapter 5: Curcumin regulates the r(CGG)^{exp} RNA hairpin structure and ameliorate defects in Fragile X-associated tremor ataxia syndrome

Table 5.1 The dissociation constant (K _d) values was determined from flouresecne titration assay of Curcumin with different (5'CNG/3'CNG) _{x1} RNAs motif.	231
Table 5.2 The dissociation constant (K _d) values was obtained from flouresecne titration assay of Curcumin with different r(CGG) ^{exp} RNA, r(CNG) ^{exp} RNA, AU paired RNA and DNA controls.	232
Table 5.3 Thermodynamic parameter values of different r(CGG) _{exp} RNAs, r(CNG) _{exp} RNAs & r(AU) ₆ duplex RNA (control) and with Curcumin.	235
Table 5.4 PCR primer sequences for splicing defect.	259

Chapter 6: Chemical screening identifies compounds that mitigate r(CGG)_{exp} RNA mediated toxicity in Fragile X-associated tremor ataxia models

Table 6.1 Bisantrene similar molecules color and shape tanimoto score values and structure of best similar molecules.	274
Table 6.2 The dissociation constant (K _d) values was determined from flouresecne based titration assay of Bisantrene with different (5'CNG/3'CNG) _{x1} RNAs motif.	282
Table 6.3 The dissociation constant (K _d) values of Bisantrene similar molecules with (5'CGG/3'GGC) _{x1} and (5'CAG/3'GUC) _{x1} RNAs was obtained from flouresecne based titration experiment.	283
Table 6.4 The dissociation constant (K _d) values of lead molecules with higher repeat target and control RNA, (5'CGG/3'GGC) _{x6} and	

(5'CAG/3'GUC)x6 respectively. The values was obtained from fluorescence based titration experiment.	287
Table 6.5 The dissociation constant (Kd) values of lead molecules with higher repeat (5'CGG/3'GGC)x20, (5'CGG/3'GGC)x40, (5'CGG/3'GGC)x60 and with RNA and DNA controls. The values was obtained from fluorescence based titration experiment.	288
Table 6.6 a Thermodynamic parameter values of different r(CGG)exp RNAs & r(AU)6 duplex RNA (control) with B1 compound.	292
Table 6.6 b Thermodynamic parameter values of different r(CGG)exp RNAs & r(AU)6 duplex RNA (control) with B4 compound.	293
Table 6.6 c Thermodynamic parameter values of different r(CGG)exp RNAs & r(AU)6 duplex RNA (control) with B11 compound.	294
Table 6.7 Thermal melting denaturation assay; T _m values of CGG repeat RNAs in presence and absence of lead compound.....	300
Table 6.8 PCR primer sequences for splicing defect.....	326

ACRONYMS

AR	Androgen Receptor
ASOs	Antisense Oligonucleotides
AONs	Anti-sense oligonucleotides
ATXN3	Ataxin 3
BSA	Bovine serum albumin
CJD	Creutzfeldt-Jakob disease
CT-DNA	Calf thymus Deoxyribonucleic acid
CD	Circular Dichroism
cDNA	Complementary DNA
D/N	Drug/nucleotide or drug/DNA
DMEM	Dulbecco's Modified Eagle's medium
DRPLA	Dentatorubral–pallidoluysian atrophy
DM	Myotonic Dystrophy
DMSO	Dimethyl sulphoxide
EGCG	Epigallocatechin gallate
EMSA	Electrophoretic mobility shift assay
EDTA	Ethylenediaminetetraacetic acid
FRDA	Friedreich ataxia
FXTAS	Fragile X associated tremor ataxia syndrome
FXPOI	Fragile X-associate primary ovarian insufficiency

FMRP	Fragile X mental retardation protein
FXS	Fragile X syndrome
GFP	Green fluorescent protein
HSPs	Heat shock proteins
HTT	Huntingtin protein
HDAC	Histone deacetylase
HD	Huntington Disease
HDL-2	Huntington's disease-like 2
IC ₅₀	Ligand concentration that inhibits fifty percent of the total
ITC	Isothermal titration calorimetry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
mRNA	messenger ribonucleic acid
MJD	Machado-Joseph Disease
MBNL	Muscleblind
NMR	Nuclear magnetic resonance
OPMD	Oculopharyngeal Muscular Dystrophy (poly(A)-binding protein-2)
PABPN1	Polyadenylate-binding protein nuclear1 gene
PVDF	Polyvinylidene fluoride
PCR	Polymerase chain reaction
PBS	Phosphate buffer saline

RNAi	RNA interference
RT-PCR	Reverse Transcriptase PCR
RISC	RNA-induced silencing complex
RAN	Repeat-associated non-ATG
SAR	Structure-activity relationships
SPR	Surface Plasmon Resonance
SBMA	Spinal and bulbar muscular atrophy
SCA	Spinocerebellar ataxia
shRNA	Short hairpin ribonucleic acid
siRNA	Small interfering ribonucleic acid
SDS-PAGE	SDS polyacrylamide gel electrophoresis
T _m	Melting Temperature
TNRs	Trinucleotide Repeats Disorders
TREDS	Trinucleotide repeat expansion disorders
TBST	Tris-buffered saline tween
UV	Ultraviolet-visible
UTRs	Untranslated region

Chapter 1

Introduction

1.1 Microsatellites and repeat instability

Microsatellite sequences contain almost 3 percent of the genome, which is approximately double the protein-coding sequences [1,2]. Microsatellites are one to six nucleotides simple sequence motifs that are repeated often. Microsatellites are also known as simple or short tandem repeat sequences [3,4]. These microsatellites could be present anywhere in the genome but predominantly found in noncoding regions of the genome like 5', 3' untranslated regions (UTRs), promoter region, and intergenic regions [5,6]. The reason for their origin and evolution is still under investigation and not yet fully explored, however, they might have originated from simple and repetitive sequences of mobile genetic elements like non-LTR (long terminal repeats) retrotransposons such as Alu and L1 [7-9]. Small sequence duplication of these microsatellites could further produce simple sequences with multiple repeats like microsatellites expanded in Friedreich ataxia (FRDA), and Myotonic dystrophy type 2 (DM2) contains traces of Alu element origin [10,11]. However, additional processes such as random mutation, duplication of some sequences, and replication slippage could be alternative causes for microsatellite expansion [4,12].

The primary cause for microsatellite expansion in dividing cells has been considered as DNA replication, but additional factors like mitotic recombination could play a crucial role in these events [13-16]. During replication, repetitive elements affect several factors at the replication fork, subsequently resulting in the reversal of replication fork or template switching, leading to the insertion of extra repeats [14,17,18]. Additional events include polymerase slippage, imperfect Okazaki ligation, and addition of bases at the lagging strands [13,18,19]. After the

addition of extra repeats to one or both the strands, re-annealing at daughter strands occurs, and the misalignment at this step could form typical non-canonical DNA, structures such as quadruplex and/or hairpins due to the bulge created by extra repeats. Consequently, these structures may escape various repair pathways and are transferred to next-generation, leading to permanent expansions [20-24].

Different DNA damage and repair pathways are involved in repeat expansion events. DNA damage pathways such as base excision repair (BER) and nucleotide excision repair (NER), replacement of DNA bases are the primary source of repeats expansion. However, mismatch repair pathways have still been considered as the prime cause of repeats expansion[22,23,25-28]. The MMR enzymes MutS β (MSH2-MSH3 complex), recognise and process the unusual DNA structures, with the help of some other proteins like MutL γ (MLH1-MLH3 complex) and MutL α (MLH1-PMS2 complex) endonucleases that exclude the DNA lesions [22,23,29]. Subsequently, polymerases such as Pol β insert extra repeats due to the imprecise priming or templating [22,27,30]. Many intriguing questions like how expanded repeats expand beyond the threshold limit in hundreds or thousands are still unanswerable till date. Although mechanisms like template switching replication models and other appealing mechanisms like break induced replication (BIR), a homologous pathway that helps to prevent collapsed or broken replication fork [31], have answered these questions to some extent further investigations are still needed to explore these repeats instability[32,33].

1.2 Microsatellite repeats expansion diseases

Microsatellite expansion has been the cause for several neurological and neuromuscular diseases since long and the number is increasing with recent reports of autism spectrum disorders that are caused due to the expansion of CGG repeats in fragile 7A (FRA7) and fragile2A (FRA2)[34,35] (Table 1.1) . These neurological disorders are caused due to the expansion of different repeats such as CAG, CGG, CUG and share diverse similarities as well as dissimilarities. Almost half of these microsatellite expansion diseases are caused due to CAG repeats

expansion. All these microsatellites repeat expansion diseases contain GC rich sequences with exceptions being FRDA, spinocerebellar ataxia 10 (SCA10) and 31 (SCA31) that are caused due to trinucleotide GAA, tetranucleotide repeats like ATTCT & TGGAA respectively [36]. Similarly hexa-nucleotide expansion caused Amyotrophic lateral sclerosis and frontotemporal dementia (GGGGCC) [37].

Table 1.1 Neurological disorders caused by unstable nucleotide expansions.

Disorder	Gene locus	Repeats type	Repeat size	
			Normal	Disease
FRAXA: fragile X syndrome (fragile site mental retardation 1 gene)	Xq27.3	(CGG) ^{exp}	6-52	230-2000
FXTAS: fragile X-associated tremor/ataxia syndrome (fragile site mental retardation 1 gene)	Xq27.3	(CGG) ^{exp}	7-54	55-200
FXPOI: fragile X-associated primary ovarian insufficiency	Xq27.3	(CGG) ^{exp}	7-54	55-200
FRAXE: fragile X syndrome (fragile site mental retardation 2 gene & fragile site mental retardation 3 gene)	Xq28	(CGG) ^{exp}	4-39	200-900
FRAXF: No confirmed disease association (family with sequence similarity 11, member A)	Xq28	(CGG) ^{exp}	7-40	306-1008
FRA10A: No confirmed disease association (FRA10A candidate gene 1)	10q23	(CCG) ^{exp}	8-14	> 200
FRA11B: Jacobsen syndrome (CAS-BR-M murine ectopic retroviral transforming sequence homology)	11q23	(CCG) ^{exp}	11	100-1000
MAB21L1: no confirmed disease association (MAB21, C. elegans, homologue-like 1)	13q13	(CTG) ^{exp}	6-31	> 50
SCA8: Spinocerebellar ataxia 8 (SCA8 gene)	13q21	(CTG) ^{exp}	2 >130	> 110
DM1: myotonic dystrophy type 1 (dystrophia myotonia protein kinase)	19q13.2-q13.3	(CTG) ^{exp}	5-37	> 90
CTG18.1: no confirmed disease association (transcription factor)	18q21	(CTG) ^{exp}	10-37	800-2100

4)				
FRDA: Friedreich ataxia (frataxin)	Pq13	(GAA)^{exp}	6-32	> 200
SCA12: Spinocerebellar ataxia 12 (protein phosphatase 2, regulatory subunit B)	5q31-q33	(CAG)^{exp}	7-45	55-78
HDL2: Huntington's disease-like 2 (junctophilin-3)	16q24.3	(CAG)^{exp}	6-27	51-57
Spinal and bulbar muscular atrophy (SBMA)	Xq11-q12	(CAG)^{exp}	9-36	40-55
Dentatorubral-pallidoluysian atrophy (DRPLA)	12p13.31	(CAG)^{exp}	7-25	49-88
Huntington Disease (HD)	4p16.3	(CAG)^{exp}	10-34	>35
Spinocerebellar ataxia 1 (SCA1)	6p23	(CAG)^{exp}	6-39	39-81
Spinocerebellar ataxia 2 (SCA2)	12p24	(CAG)^{exp}	13-33	>34
Spinocerebellar ataxia 3(SCA3)	14p24.3-q31	(CAG)^{exp}	13-44	>55
Spinocerebellar ataxia 6 (SCA6)	19p13	(CAG)^{exp}	4-18	20-29
Spinocerebellar ataxia 7 (SCA7)	3p21.1-p12	(CAG)^{exp}	4-35	37-306
Spinocerebellar ataxia 17 (SCA17)	6q27	(CAG)^{exp}	25-42	47-63
potassium calcium-activated channel subfamily N member 3 (KCNN3) No confirmed disease	1q21.3	(CAG)^{exp}	7-28	ND
AIB-I: increased prostate cancer risk Nuclear receptor coactivator 3	20q13	(CAG)^{exp}	29/29 or 28/29	>29
ZIC2: holoprosencephaly (zinc finger protein of cerebellum 2)	13q32	(GCG)^{exp}	15	25
HOXA13: Hand-Foot-Genital Syndrome (homeobox A13)	7p15-p14.2	(GCG)^{exp}	18	24-26
FOXL2: Blepharophimosis/Ptosis/Epicant hus inversus syndrome type II (forkhead transcription factor FOXL2)	3q23	(GCG)^{exp}	14	22-24

ARX: infantile spasm syndrome (aristaless-related homeobox, X linked)	Xp22.13	(GCG)^{exp}	10-16	17-23
HOXD13: Synpolydactyly (Homeobox D13)	2q31-q32	(GCG)^{exp}	15	22-29
OPMD: Oculopharyngeal Muscular Dystrophy (poly(A)-binding protein-2)	14q11.2-q13	(GCG)^{exp}	10	12-17
CBFA1:cleidocranial dysplasia (runt-related transcription factor 2)	6p21	(GCG)^{exp}	17	27
CJD:Creutzfeldt-Jakob disease (Prion)	20pter-p12,6p21.3	24-bp repeat	4	2,5-13
COMP: Multiple Skeletal dysplasias (cartilage oligomeric matrix protein)	19p13.1	(GAC)^{exp}	5	4,6,7

Furthermore, among different microsatellite expansion, trinucleotide repeats (TNR) expansions are the most common and caused more than 20 neurological disorders (Table 1.2). These neurological disorders have been classified into three different classes based on the repeats location and pattern of diseases. 1) TNR found in protein-coding sequences and causing dominantly inherited diseases. 2) Dominantly inherited diseases in which TNR repeats are present outside of protein-coding sequence. 3) The third type is found in recessive or X-linked genes. Generally, the TNR repeats expansion causes disease pathogenesis via protein-mediated toxicity such as Huntington's Diseases (HD), however, recently, RNA mediated toxicity has also been reported with these trinucleotide repeats expansion disorders such as Myotonic dystrophy type 1 (DM1). Additionally, TNR repeats expansion also exerts loss of function mutation within the repeats harbouring genes and Friedreich's ataxia (FRDA), autosomal recessive ataxia is a classic example for this event [38]. TNRs can form stable RNA structures in certain diseases. Disease pathology mediated by gain of function via binding of these microsatellites expanded RNA with proteins has been extensively researched (REF??). The newly recognised and very unconventional mode of translation utilising non-canonical start codon known as repeat associated non-AUG (RAN) translation has recently been discovered [39]. This RAN translation has also been identified to be one of the disease mechanisms and different *cis* and *trans* factors involved in the disease pathology have been recognized, yet the exact mechanism is still unexplored.

The genetic testing for trinucleotide repeats expansion diseases is simple compared to other diseases due to the fact these diseases are caused by specific repeats like CGG, CAG, CUG, etc. The directed PCR-based screening of specific repeats in most of the microsatellite expansion diseases is easy, simple, straightforward, accurate and cheap. However, in certain diseases that are caused due to massive repeats expansion further advancement in testing methods like use of Southern blot hybridisation are warranted [38].

Table 1.2 Trinucleotide repeat expansion disorders sorted according to repeats found in coding and non-coding region of the gene.

Disease	Gene/Locus	Codons	Functional disease group	Repeats length	
				Normal	Disease
Diseased caused by repetitive sequence present in Non coding region of gene					
Fragile X syndrome	<i>FMRI</i> (<i>FRAXA</i>) Xq27.3	(CGG) ^{exp}	Loss of function or gene silencing	6-53	>230
Fragile X-associated tremor/ataxia syndrome (FXTAS)	<i>FMRI</i> (<i>FRAXA</i>) Xq27.3	(CGG) ^{exp}	Toxic RNA and Protein	6-53	55-200
Fragile X-associate primary ovarian insufficiency (FXPOI)	<i>FMRI</i> (<i>FRAXA</i>) Xq27.3	(CGG) ^{exp}	Toxic RNA and Protein	6-53	55-200
Spinocerebellar ataxia type 8	<i>SCA8</i> 13q21	(CTG) ^{exp}	Toxic RNA	16-37	110 to < 250
Myotonic Dystrophy 1	<i>DMPK</i> 19q13	(CTG) ^{exp}	Toxic RNA	5-37	> 50
Spinocerebellar ataxia type 12	<i>SCA12</i> 5q31-33	(CAG) ^{exp}	Toxic RNA	7-28	66-78
Friedreich ataxia	<i>X25</i> 9q13-21.1	(GAA) ^{exp}	Loss of function	7-34	>100
Fragile XE syndrome	<i>FMR2</i> (<i>FRAXE</i>) Xq28	(CCG) ^{exp}	Loss of function or gene silencing	6-35	>200
Diseased caused by repetitive sequence present in coding region of gene					

Spinocerebellar ataxia type 1	<i>SCA1</i> 6p23	(CAG)^{exp}	Toxic protein	6-44	39-82
Spinocerebellar ataxia type 2	<i>SCA2</i> 12q24.1	(CAG)^{exp}	Toxic protein	15-31	36-63
Spinocerebellar ataxia type 3 (Machado–Joseph disease)	<i>SCA3</i> (<i>MJD1</i>) 14q32.1	(CAG)^{exp}	Toxic protein	12-40	55-84
Spinocerebellar ataxia type 6	<i>SCA6</i> 19p13	(CAG)^{exp}	Toxic protein	4-18	21-33
Spinocerebellar ataxia type 7	<i>SCA7</i> 13p12-13	(CAG)^{exp}	Toxic protein	4-35	37-306
Spinobulbar muscular atrophy (Kennedy disease)	<i>AR</i> Xq13-21	(CAG)^{exp}	Toxic protein	9-36	38-62
Huntington’s disease	<i>HD</i> 4p16.3	(CAG)^{exp}	Toxic protein	6-35	36-121
Dentatorubral-pallidoluysian atrophy. (Haw–River syndrome)	<i>DRPLA</i> 12p13.31	(CAG)^{exp}	Toxic protein	6-35	49-88

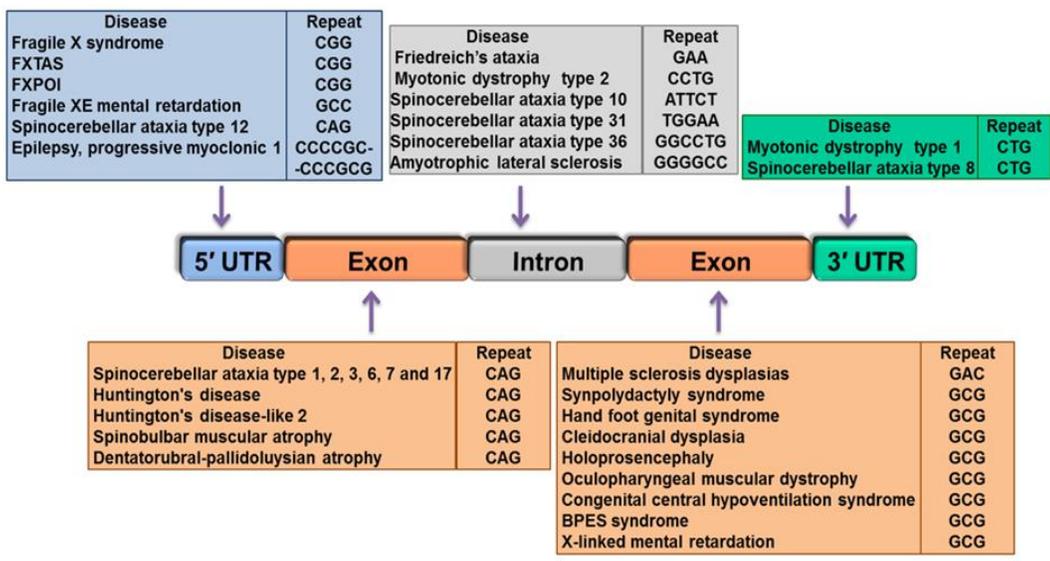


Figure 1.1 Diagram representing position of repetitive sequence found on the gene that contributes to repeat-associated neurological disorders.

Diseases associated with CGG nucleotide repeats expansion

CGG triplet nucleotide repeat expansion is alone responsible for different neurological disorders namely, Fragile X syndrome, Fragile X-associated tremor/ataxia syndrome, and Fragile X-associated primary ovarian insufficiency. All these neuropsychiatric diseases are caused by the repetition of CGG nucleotide repeats in the 5'UTR position of *FMR1* gene located on fragile site (A-site) of the long arm of the X chromosome (X.27).

1.3 Fragile X Syndromes (FXS)

Fragile X syndrome (FXS) also termed Martin-Bell syndrome, is an intellectual disability (ID) disorder, first recognized in 1943 [40]. Lubs and his colleagues first described different fragile sites present on the X chromosome [41]. Further, fragile site, Xq27.3 mapping was confirmed in 1991. [42] FXS is identified as the second most common inherited monogenic form of mild and severe mental retardation diseased condition after Down syndrome.

FXS is caused by the unstable and dynamic CGG nucleotide expansion in the 5' UTR and promoter region of the *FMR1* gene, coding for fragile X mental retardation protein (FMRP). *FMR1* is a highly conserved gene, 38 kilo base (kb) in length and contains 17 exons [43,44]. In a healthy individual the CGG expansion is extremely polymorphic in nature often interrupted by AGG nucleotide found every 10 to 20 repeat position and stably inherited to the next generation [45]. AGG interruption has been believed to play a vital role in the maintenance of repeat stability. The CGG expansion is also found at the 3' UTR of gene. Depending on the CGG triplet length, *FMR1* allele is categorized into four types with distinct clinical manifestation. 1) In normal healthy population the CGG repeat ranges from 5-44 repeats, with most frequently 30 repeat found on the allele. 2) CGG repeat range from 45-54 CGG is termed as the "gray zone", also called intermediate zone. The repeats at this zone are vulnerable for premutation condition, potentially due to parental meiotic instability across

generation [46]. 3) CGG repeat range 55-200 CGG repeat is the premutation condition (PM) leading to FXTAS and/or FXPOI, a pathophysiological condition different from FXS. The chances for a premutation to convert to full mutation exponentially increases when the numbers of CGG triplet are 65 to 100 in the coming generation. 4) In FXS, the number of CGG repeats increases to more than 200 in the affected individual and the condition referred as full mutation (FM) [47]. Of note, full mutation condition occurs when the premutation allele is transmitted from mother, not from father [48]. Additionally, in some fragile X patients, the FM carrier is unmethylated, apparently producing FMRP. Also some FXS patients harbour evidence regarding the methylation status of the full mutation allele, unmethylated or partial methylated. Further studies reported that CGG expanded *FMR1* transcript is not translated. These examinations suggested that CGG nucleotide expansion and methylation are two independent processes. Although, their interconnection is still not clear. Thus, CGG expansion length and its methylation condition determine the production of FMRP which is crucial during cognition development [49].

FXS has been extensively studied mostly in western countries. The estimated incidence of FXS is 1 in 5000-7000 males and 1 in 4000-6000 female [50]. Although the prevalence rate of FXS can vary according to racial, ethnic, and founder effect [51,52]. The FXS phenotypes are more severe in male as compared to female, because of extra X chromosome compensating for the expression of affected allele. Fragile X syndrome is caused due to CGG repeat trigger epigenetic silencing of the *FMR1* gene, which is characterised by the massive expansion and hypermethylation of the CGG repeat of the promoter region. *FMR1* gene is the sole gene firmly responsible for FXS pathogenesis. Thus, suppression of FMRP expression is the typical cause of FXS. FMRP is RNA binding protein that regulates the translational process of several messenger RNAs of the postsynaptic site of neuron [53,54]. The size of FMRP is 70-80 kDa size and has 12 different isoforms generated through alternative splicing of 4.4 kb *FMR1* transcript.

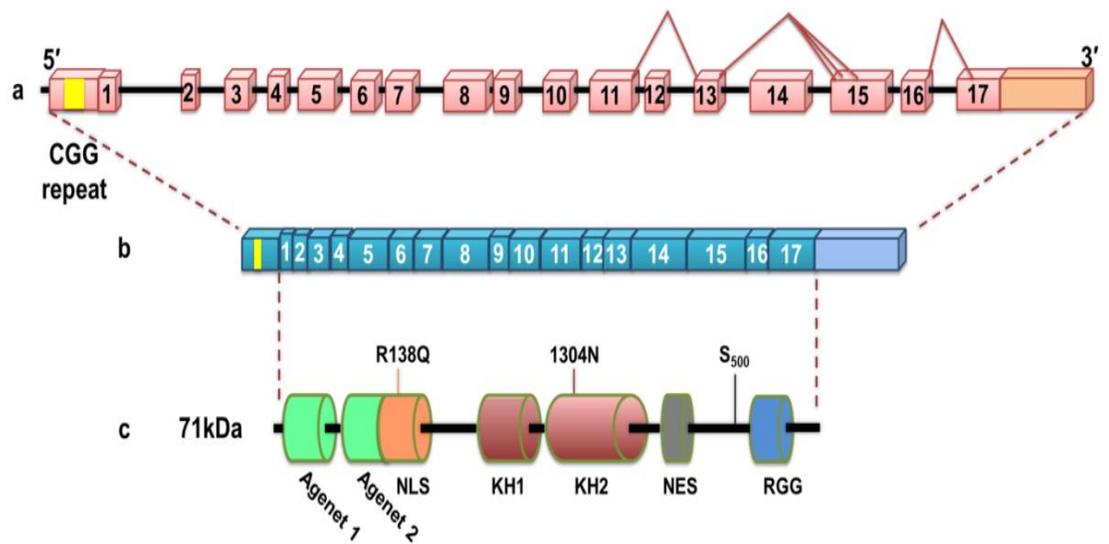


Figure 1.2 Overview of FMR1 gene comprises 17 exons that can undergo pre-mRNA alternative splicing in the 3' region of the gene and produce FMR1 transcript and FMRP protein.

1.3.1 Clinical features and symptoms

The typical clinical symptoms of FXS are loss of memory, cognitive decline, attention deficit, anxiety, hyperactivity, autistic behaviour, speech problems and macro-orchidism [55]. The severity of symptoms depends on the FMRP concentration that depends on the percentage of methylated CGG repeat [56]. The symptoms are less severe with moderate cognitive disability and normal intelligence quotient when the FMRP concentration is not very low. When the level of FMRP protein completely decreases and severe intellectual disability and cognitive impairment result. The specific phenotypic features also depend on the amount of FMRP produced. 80% of the affected individuals have a typical clinical features such as long face, large and protruded ears, cleft palate, mandibular prognathism, cardiac abnormalities and macroorchidism [57]. Female with the full mutation, FXS condition shows broad range of phenotypic features than men, strongly dependent on the activation percentage of the affected X chromosome. In FXS women, there are 30% chances of having normal IQ and 25% chances of having intelligence disability with IQ less than 70, however learning deficit (60%) and emotional difficulties (70%) still persist [58,59].

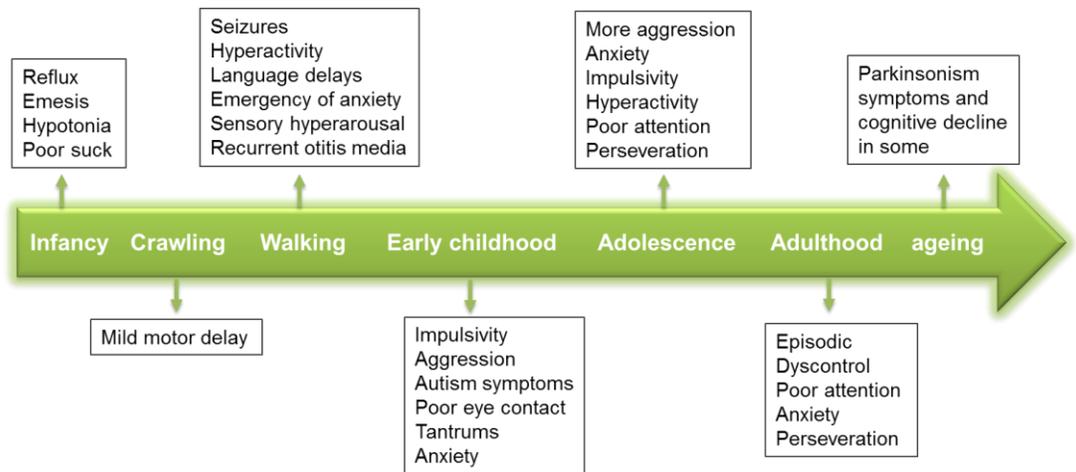


Figure 1.3 Diagram showing clinical features of fragile X syndrome patient at the different periods.

1.3.2 Molecular mechanism/pathophysiology of FXS

FMR1 gene silencing occurs in full mutation (CGG repeat expansion above 200) condition. It is a complex phenomenon requiring several levels of epigenetic modification such as chromatin remodeling, DNA methylation, histone modification and RNA interference. The FXS silencing allele shows heterochromatic configuration, not transcriptionally permissive compared to the euchromatic configuration of the transcriptionally active normal allele. Usually, such dynamic transformation from transcriptionally active gene to silenced state is directly related to the size of CGG nucleotide expansion, leading to epigenetic modification. One exception to this rule is the unmethylated full mutation (UFM) where allele is not silenced even when the CGG repeats are expanded above 200.

In FXS, methylated cytosines are found approximately 1kb upstream from *FMR1* promoter including CGG repeat. Methylation usually occurs in the promoter region of the *FMR1* allele serving as a boundary to prevent further spreading. In contrast, in full mutation and premutation this boundary is lost and CGG nucleotide repeat becomes methylated at approximately 11 weeks of gestation [60]. In full mutation carrier different signature patterns of methylation, demethylation and deacetylation of histones have been documented for chromatin modification. Histones H3 and H4 are deacetylated. Methylation and demethylation are found on lysine 9 and lysine 4 of H3 (H3K9 & H3K4) respectively. Furthermore, trimethylation of 27th lysine is present on histone 3 (H3K27me3) and also increased methylation of 20th lysine on histone 4 (H4K20me3) near the CGG repeat. All these epigenetic modifications are typical marks of heterochromatin and facultative configuration, transcriptionally not active. A similar type of chromatin modification has also been reported in other neurological disorders including Myotonic dystrophy type 1 [61], Friedreich ataxia [62], and Amyotrophic lateral sclerosis [63].

FMR1 gene silencing can be partially reactivated and treated through DNA methylation inhibitor in the patient-derived cells which are not influenced by

demethylation (H3K9me2 or H3K9me3) [64]. This observation indicated that DNA methylation is an independent event occurring downstream of the chromatin modification process. Similar epigenetic modification patterns were also observed with the unmethylated full mutation individuals where gene silencing does not occur due to an unmethylated CpG region [65]. Furthermore, the FMR1 silencing can also be partially reactivated by using inhibitor of the enzyme Sirtuin 1 (SIRT1) that modulates the deacetylation of lysine 9 on histone 3 and lysine 16 on histone 4. It has been reported that DNA methylation followed by the acetylation of H4K16, and unacetylated H3K9 was observed [66]. Thus, it would be speculated that the deacetylation H3K9 followed by DNA methylation is the downstream process after methylation compared to deacetylation of H4K9. In comparison to the deacetylation of the full mutation carrier, premutation carrier have 1.5-2.0 times more acetylated histone than the wild type. This epigenetic modification might be responsible for the open chromatin leading to overproduction of FMR1 transcript.

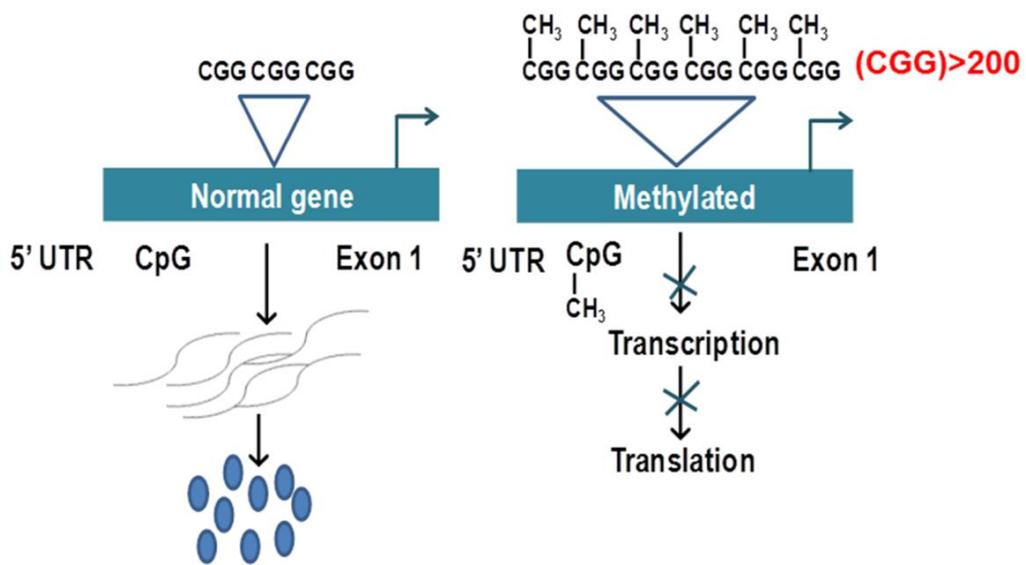


Figure 1.4 Schematic representation of protein loss of function mechanism.

1.3.3 Therapeutic approaches

There are two main mechanisms involved in the silencing of the FMR1 gene. First is the expansion of CGG repeat (above 200) and second is the epigenetic modification (DNA methylation) that leads to the absence of FMRP protein, ultimately leading to FXS phenotypes. Thus, two different possible approaches were considered for the effective treatment of FXS. **a)** To compensate the abnormal function due to the loss of FMRP. **b)** Reactivation of the affected FMR1 gene which is transcriptionally silenced due to the chromatin modification. Both approaches have been tested using cellular models (patient-derived cell lines and brain slice from FMR1 knock-out mouse), animal models (mouse and drosophila) and clinical trials. But all the reported drugs were found to be not effective enough in clinical trial as compared to their respective results from animal models. Thus, it would be essential to develop and test the new therapeutic drugs on human cellular models. Induced pluripotent stem cells from Fragile X syndrome (FXS-iPS) patients and neuronal cells derived from FXS-iPS would provide advanced model systems to test the drugs.

The first therapeutic strategy would be to normalise the functional deficit condition caused by the loss of FMRP. Several signalling pathway systems were found to be affected due to the abolishing of FMRP such as mGluRs, AMPA, GABA, NMDA, and TrkB and contribute to FXS phenotypes. Different individualised therapeutic small molecules have been identified to target specific signalling pathways that can potently address the physiological and psychological aspect of FXS.

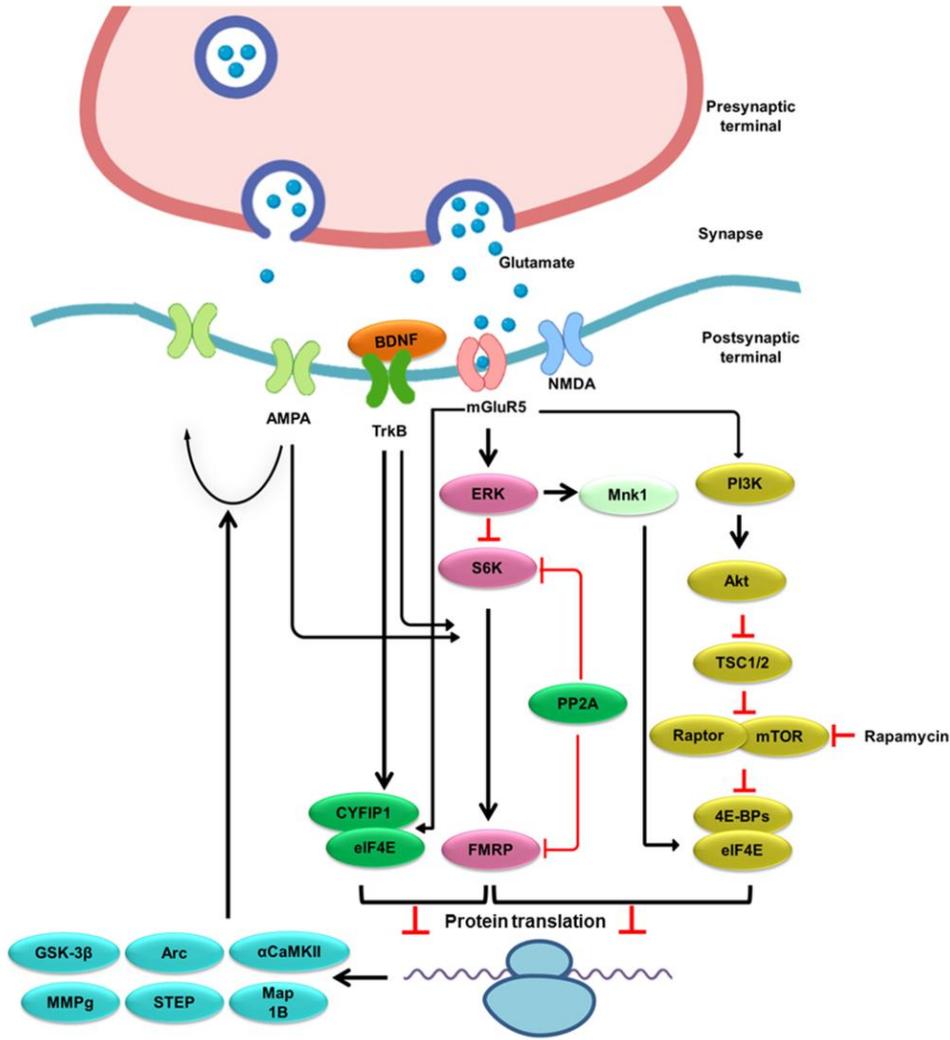


Figure 1.5 Effect of neuronal receptor signalling pathway on FMRP mediated regulation at synapsis.

1.3.3.1 Targeting Glutamatergic system

mGluRs is the protein of G protein couple receptor class. Lack of FMRP increases the expression of mGluRs having roles in learning, speech and language. The mGluR hypothesis suggested that antagonising the signalling pathway may results to reduction of the FXS phenotype. The first antagonist, fenobam, was used to selectively target mGluR5. Fenobam has been previously used in anxiety treatment, and recently open-label trial was done in 12 adult FXS patients in which single dose was given orally. Fenobam significantly improved hyperexcitability, anxiety, prepulse inhibition, behaviour disorders and better precision in task performance was also reported with no clinical adverse side effects. Longer treatment would be expected to improve cognitive dysfunction [67]. Another mGluR5 antagonist MPEP, tested on FMR1KO mice resulted in decreased audiogenic seizures and rescued spine dysmorphogenesis. Further MPEP was found to improve cognitive and behavioural deficits in drosophila, zebrafish and mouse model [68,69]. Similarly, chronic administration of pharmacological inhibitor CTEC in adult animal model (FMR1KO mice) alleviated the FXS clinical outcomes including memory and learning deficits, uncontrolled protein synthesis, hippocampus LTD, audiogenic seizures, partially ameliorated macroorchidism, improved dendritic spine abnormalities in the cortex, enhanced locomotor activity and ameliorated the hyperactive ERK and mTOR signalling [70].

Another published data from double-blind controlled placebo clinical trial focused on assessing mGluR5 antagonist, mavoglurant (AFQ056) from Novartis, a selective inhibitor tested in adult FXS patient age from 18-35 years. Mavoglurant significantly improved the stereotypic behaviour, inappropriate speech and hyperactivity [71]. The efficacy of AFQ056 was also assessed in subgroups comprising full methylation at the FMR1 promoter region and non-responder partially methylated subgroup. Recently two different large organisations namely, Novartis (AFQ056) and Hoffmann-La Roche (RO4917523) conducted controlled clinical trial on both adolescent and adult up to two years, and results of these

studies are still pending [72,73]. Mice studies (FMR1 KO) demonstrated that AFQ056 ameliorated pathological startle response after prepulse inhibition [74]. AFQ056 showed better pharmacokinetic profile and stress-induced hyperthermia test in mice over MPEP [75]. Furthermore, AFQ056 treated hippocampal neuron cultured cells from FMR1 KO mice showed shorted dendritic spines while untreated resulted in elongated spines [75].

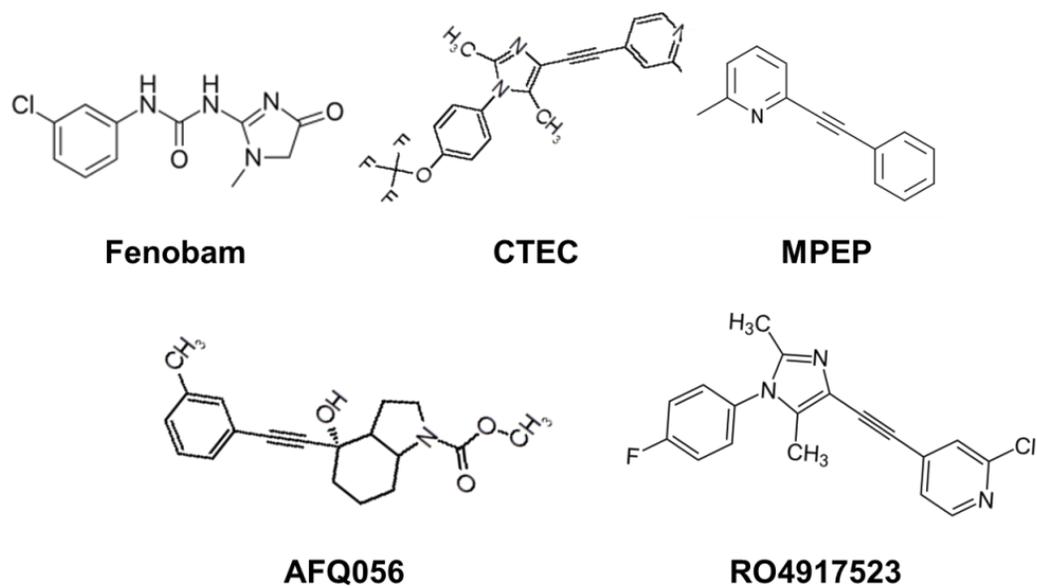


Figure 1.6 Small molecules that target the glutamatergic signalling pathway.

1.3.3.2 Targeting GABAergic system

GABAergic (Gamma-Aminobutyric Acid) system was believed as the down regulate in the FMR1 knockout mice and drosophila due to underexpression mRNA code for GABA subunit receptor proteins compared to wildtype [76] [77] [78]. It is a major inhibitory receptor which controls several central nervous process like insomnia, depression, learning, memory, anxiety, and epilepsy. Thus, impairment of GABA neurotransmission system in different location of brain such as cerebral cortex, striatum and amygdala results to contribution of FXS abnormalities and phenotypes [78]. Several pharmacological interventions targeting gamma-aminobutyric system have been explored to rescue amygdala associated symptoms [79]. Acamprosate, previously used for abstinence from alcohol, act as an antagonist of NMDA receptor and agonism for GABA receptor. Recently, acamprosate was tested in three young adults with of FXS and autism phenotype, significant clinical benefits (evaluated by clinical global impression improvement) and marked in linguistic communication [80]. Baclofen, agonist of GABA receptor inhibit the susceptibility audiogenic seizures incidence in the double knockout mice. Additionally, a large clinical trial with arbaclofen, an R-isomer of baclofen in 63 FXS patient of age 6-40 years demonstrated improvement in sever social impairment [81,82].

Furthermore, a double-blind, randomized, placebo-controlled trial on 59 children and adolescents with FXS symptoms, age 6-17 years old revealed that Ganaxolone (analogue of neuroactive allopregnanolone), a positive allosteric regulator of GABA receptor improved overall primary outcomes measures (rated by CGI-I) and secondary outcomes (anxiety) without having any serious adverse effects [83]. Preclinical studies of ganaxolone in FXS KO mice demonstrated the improvement in audiogenic seizures, repetitive and stereotypic behaviour [84,85]. Similarly longer treatment of gaboxadol, a superagonist of GABA receptor target specifically δ -subunit restore neuronal excitability deficit in the FXS mice and also restore behaviours abnormalities and significantly decrease hyperexcitability [79,86]. Taken together, all these drugs are target GABA receptor system,

henceforth; any of these drugs could show high potency in FXS then ultimately became clinically available.

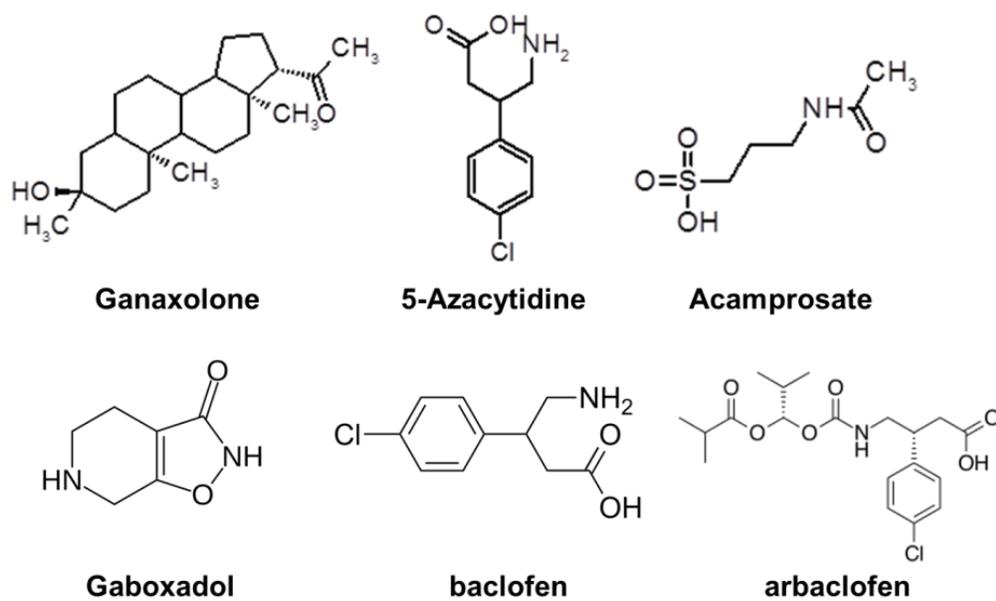


Figure 1.7 Small molecules that target the GABAergic signaling pathway.

1.3.3.3 Targeting Epigenetic modulator

The second strategy is to restore the epigenetic marks that maintain the mutated inactive gene. It is hypothesized that inhibition of the epigenetic modulator (primarily DNA methyltransferase) via small molecules could allow the FMR1 transcription. This idea is further supported by unmethylated full mutation condition where CGG triplet expansion is unmethylated due to unknown reason and allows for normal expression of FMR1 gene. In this regard, the first tested compound was 5-azacytidine (5-azaC), a methyltransferase inhibitor on FXS patient-derived cells (lymphoblastoid). 5-azaC prevented the DNA methylation and restored FMR1 transcription and translation.

Further, more efficient methyltransferase inhibitor 5-azadC, an analogue of deoxycytidine incorporated during the replication and irreversibly inhibit the DNA methylation was also tested. Several studies have been reported to assess the effect of DNA demethylation on the mutated FMR1 gene using FXS patient-derived lymphoblastoid cell line. The reactivation was accompanied with slight DNA demethylation and partially recovered the FMRP production level. This indicates that DNA hypermethylation is principle phenomena for FMRP and not the expansion of CGG repeats. This reactivation of FMR1 gene changes its allele methylation status similar to unmethylated full mutation where no methylations occur due to unidentified reasons and lead to normal production of FMRP. Administration of 5-azadC transiently and partially restored the allele until the 3-4 week after treatment withdrawal. Acetylation and methylation status of the different parts of the gene, promoter, exon 1, and exon 16 were evaluated using bisulphite sequencing methods before and after treatment of 5-azadC for the 7 consecutive days. Partial acetylation, as well as DNA methylation of H3K4, was achieved along with slightly reduced H3K9 methylation. It was found that the effect of 5-azadC is not random on genomic DNA. Specificity for the FMR1 promoter region could not be proven as the administration did not affect the methylated sequence of the other part of gene. However this cannot exclude the effect of methylation of status of other gene promoter region on 5-azadC.

Unfortunately, 5-azadC is not used in vivo and in clinical trials due to the adverse effects of cytotoxicity and apoptosis induction. However, 5-azadC has been widely used for the treatment of haematological malignancies; also it is expected to be potent only on dividing cells. Thus, it would be good to discover whether the compound could be used reactivation with minimal/no cytotoxicity. In this case butyrate and phenylbutyrate, deacetylase inhibitors were also checked in peripheral lymphocytes of FXS patient but were found to be less promising. Combined treatment of 5-azadC, butyrate and phenylbutyrate improve the percentage of FMRP production compared to 5-azadC alone, indicating a synergetic effect. L-acetylcarnitine is a natural available small molecule that improves cell metabolism and inhibits the cytogenetic expression of FXS cultured lymphocytes. L-acetylcarnitine administration reduced attention deficit and hyperactive behaviour of FXS treated boys compared with placebo group. Another tested molecule, valproic acid (VPA) is the histone deacetylases inhibitor and modest reactivator of silenced FMR1 gene. For the safe open-label trial, 10 FXS boys were treated for 6 months and exhibited improvement in attention deficit and hyperactivity phenotypes. Recently, researchers have been able to selectively reactivate the mutated FMR1 gene using the gene-editing technique CRISPR/Cas9. Reactivation of FMR1 expression was achieved alongside promoter and repeats methylation in FXS embryonic stem cells (FXS-ESC) and neuronal cells obtained from induced pluripotent stem cells (iPS) derived. Table 1.3 showing the small molecules have been discovered till date for FXS and FXTAS treatment.

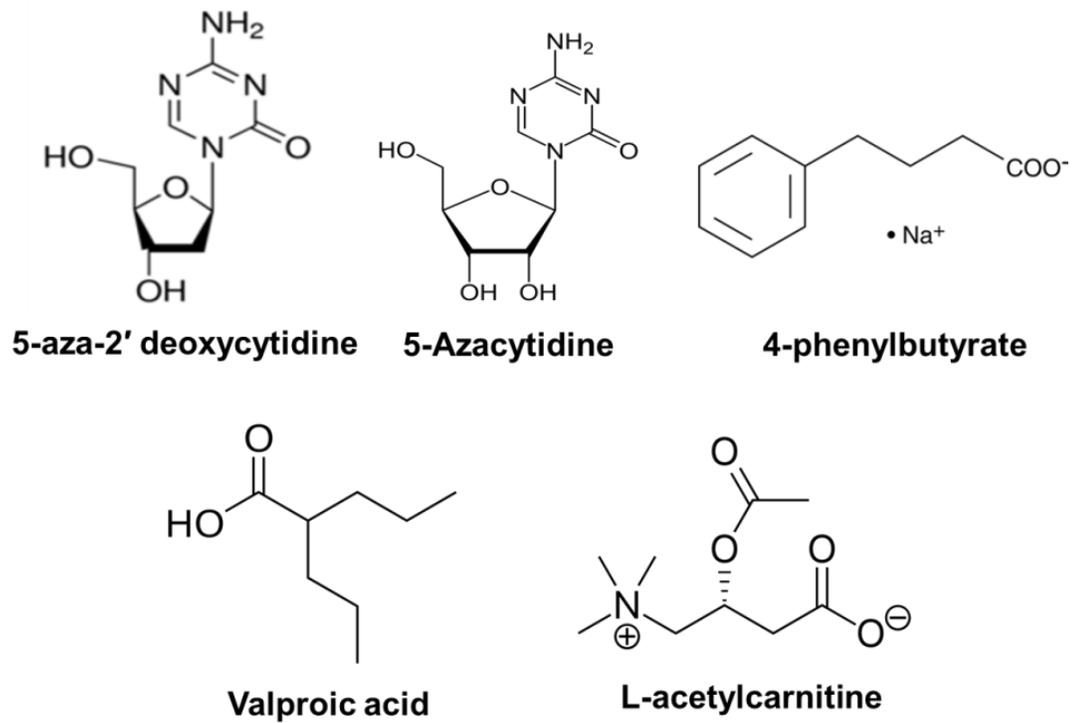


Figure 1.8 Small molecules that target the Epigenetic modulator.

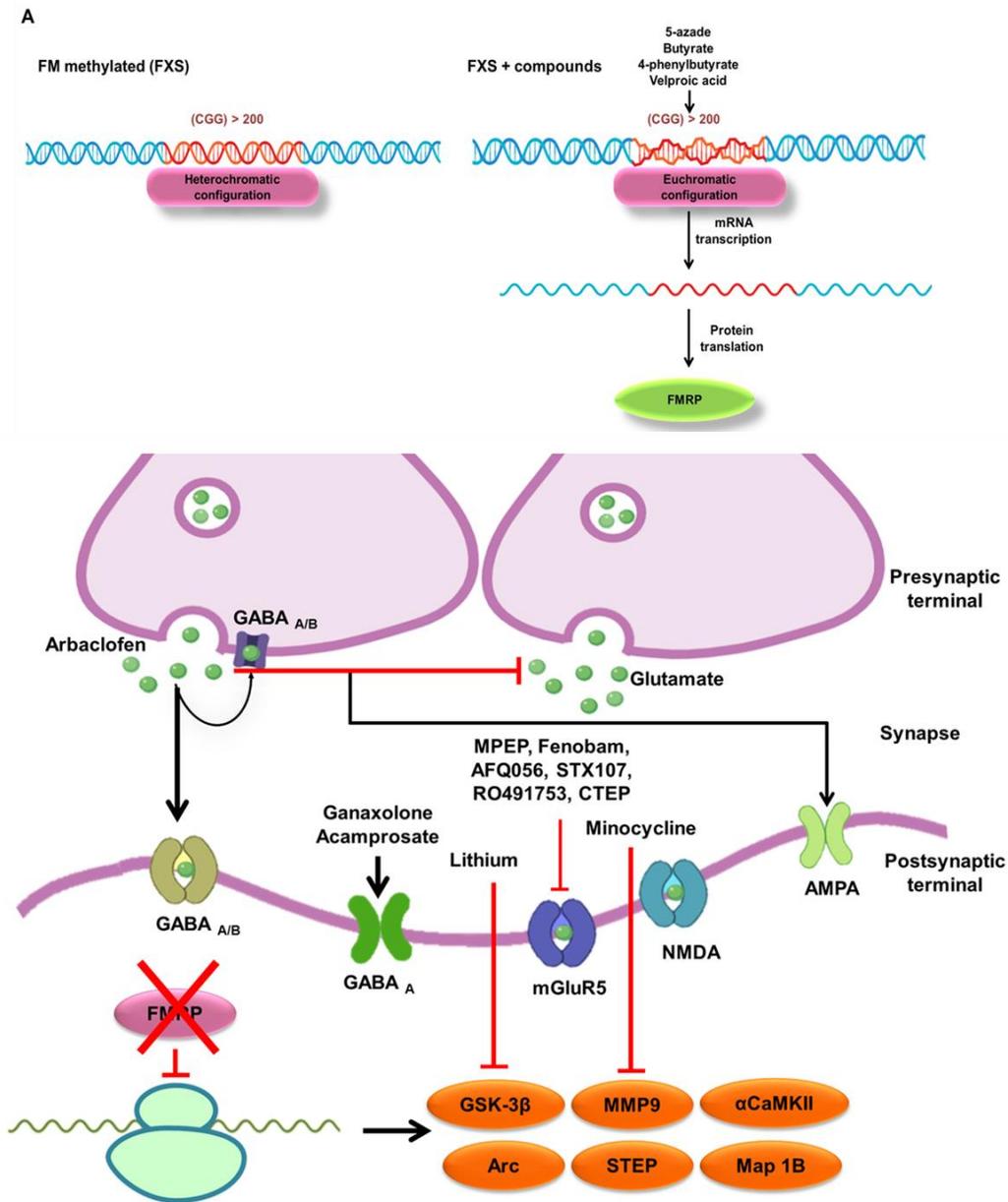


Figure 1.9 Diagram represent FXS therapy development. a) Shows epigenetic modification by compound reactivate FMRI gene expression. b) Depicting compound restoring deregulated pathway due to lack of FMRP.

Table 1.3 Small molecules reported for the treatment of FXS and FXTAS.

Clinical trail				Targeted pathway
Therapeutically tested small molecules for Fragile X syndrome				
Preclinical	Phase I or open label	Phase II	Phase III	
Baclofen*	Ganaxolone	Arbaclofen	Arbaclofen	Modulate signalling via GABA and/or other non-mGluR receptor to restore or balance abnormal signalling due to absence FMRP
Arbaclofen	Gaboxadol	Ganaxolone		
Acamprosate*	Donepezil*	Acamprosate*		
Ganaxolone	Acamprosate*	Metadoxine		
Metadoxine		Donepezil*		
Minocycline*	Minocycline*	Minocycline*	None	Inhibit excess activity of produced protein in bulk due to lack of FMRP
STEP inhibitor	Rolipram			
Rolipram				
PDE inhibitor				
CX516	None	CX516	None	Enhance activity of deficient AMPA receptor

Lithium*	Lithium*	Metadoxine	None	Inhibiting excess mGluRI signalling by blocking pathway directing from receptor activation to protein production
Lovastatin*	Lovastatin*	Lovastatin*		
Metadoxine		NNZ-2256		
PIKE inhibitor				
PAK inhibitor				
NNZ-2256				
GSK3β inhibitor				
PI3K inhibitor				Block excess mGluRI signalling pathways directly by blocking receptor
AFQ056	AFQ056	AFQ056	None	
Fenobam	Fenobam	RO4917526		
CTEP	STX107			
MPEP				
STX107				Regulate aberrant insulin signalling pathway in the absence of FMRP
Metformin*	Metformin*	None	None	
Cannabidiol	None	None	None	Control abnormal endocannabinoid signalling pathway
Endocannabinoid blockers	None			
miR-125a	None	None	None	Inhibit excess production of specific protein

				with miRNAs
5-aza-deoxytidine				Inhibiting epigenetic modulators (DNA methyltransferase) to reactivate FMR1 gene
TSA				
Butyrate				
4-phenylbutyrate				
Valproic acid				
Acetyl-L-carnitine				
Therapeutically tested small molecules for Fragile X-associated tremor/ataxia syndrome				
MPEP				Inhibiting mGluR signalling and activate GABA signalling pathway
allopregnone	allopregnone*	allopregnone*		
Citocoline	Citocoline	Citocoline		Ameliorate neurodegeneration
Fluocinolone acetonide				Suppress locomotion deficits condition by inhibiting Phospholipase A2
Quinacrine				
Arachidonyl trifluoromethyl ketone				

rapamycin				Blocking mTOR signalling pathway leads to attenuate neurodegeneration
1a				Targeting CGG hairpin structure and improve FXTAS associated defect
2H-5				
2H-5-CA-Biotin				
2HE-5NMe				
Piperine				

1.4 Fragile X-associated tremor/ataxia syndrome (FXTAS)

Fragile X-associated tremor/ataxia syndrome is the late onset of progressive neurological and neuromuscular disorder caused by the expansion of CGG trinucleotide repeat in the 5'UTR of *FMR1* gene positioned on the long arm of X-chromosome (Xq27.3) [87-89]. In normal individuals, the number of CGG repeat varies in fragile X gene from 5-54. Due to the genomic instability of repeat across the generation, a large number of the population carries the CGG expansion from 55-200. A fragile X permutation carrier occurs with a prevalence of 1 of 400-800 in males and 1 of 300 in females respectively which is more frequent than full mutation [90,91]. Female are less susceptible to developing FXTAS as random inactivation of X chromosome in female lead to few female patients. Approximately, 45-75% of males and 16-20% of female's permutation carrier have chance to developing FXTAS, which is dramatically increased over the age of 50 [50]. Currently ongoing large population neonatal screening studies demonstrated that the chance prevalence may increase as diagnostic techniques have improved. Once FXTAS was characterized, neurologist tested genomic DNA and scrutinized features sequentially to recognize premutation carrier in the population [92]. Several studies focused on phenotypic expressions of patients that were originally reported for FXTAS, like, tremor, ataxia, Parkinsonism. There are some common clinical characteristics in most of the neurological disorders and also associated with FXTAS. Around 29 studies have reported the observance of movement and locomotor disability results with CGG expansion in *FMR1*. These meta-analysis screening studies include patients of several other neurodegenerative diseases like Friedreich's ataxia, spinocerebellar ataxia, autonomic dysfunction, Huntington disease-like phenotypes, multiple system atrophy, ataxia, essential tremor [93-95].

As a matter of fact, FXTAS is considered as the most common cause of tremor and gait ataxia in aged individuals. Apart from the CGG repeat length and age, the other risk factors for the development of FXTAS disorder in premutation carriers is still unknown. However, likely candidates include epigenetic modification and

environmental factors (addictive substances, environmental toxins, chemotherapy and other illness) [96-98]. Identification of these risk factors could further provide the idea about the penetrance of diseased allele. The length of CGG repeat and late-onset of age positively correlates with the phenotypic expression of diseased individuals. As the number of individuals reaching to 65 years of age increases it is presumed that the number of FXTAS patient will increase accordingly and help in further identifications of the underlying pathogenic molecular mechanism of FXTAS [99].

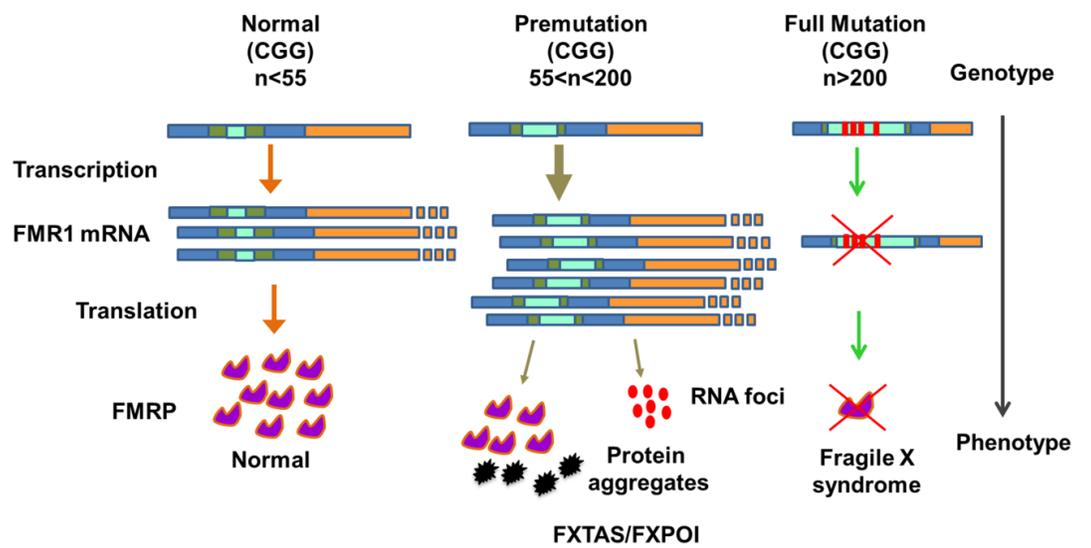


Figure 1.10 Diagram represent CGG repeat associated neurological diseased condition.

1.4.1 Clinical features of FXTAS/Premutation diseased condition

Initially, when the *FMRI* gene was discovered in 1991, premutation carrier was considered to be clinically asymptomatic [100]. However, it is now considered to present a vast spectrum of neurological symptoms such as memory problems, depression, anxiety, difficulties with numerical and magnitude estimation and executive function deficits. They are also very susceptible to develop late-onset progressive neurological and neuromuscular disorder Fragile X associated tremor/ataxia syndrome. The major clinical features of FXTAS include prominent intention tremor and cerebellar gait ataxia (balance problem), difficulties in movement and difficulty to perform specific task (holding something). Intention tremor is very commonly found in FXTAS patients with variable severity [101]. In premutation male carriers, 50 percent showed mild tremor and 17 percent showed moderate tremor with similar appearance. The minor clinical features involve progressive cognitive defects, Parkinsonism, dementia, multiple systemic atrophy, white matter abnormalities, mood lability, social phobias, neuropsychological problems and, white matter abnormalities in some patients [102-104].

Furthermore, along with principle features of FXTAS (tremor & ataxia) some undefined medical comorbidity have also been reported such as peripheral neuropathy (60%), bowel and bladder dysfunction (30–55%), impotence (80%), thyroid diseases[105], autoimmune dysfunction[105], migraine[106], seizure [107], erectile dysfunction, limb muscle weakness, cardiac arrhythmias [105], hearing loss, fibromyalgia [105] and dysphagia [101,108,109]. FXTAS clinical symptoms start with the slow progression of tremor and/or ataxia to motor disabilities and cognitive impairment that remarkably affect the individual ability to independently face society. Neuroimaging and neuropathological profile of FXTAS affected person shows discrete brain imaging features including cerebrum atrophy and ventricular enlargement. Several other studies reported that approximately 60 percent of FXTAS affected male have white matter lesion in the center of cerebellar peduncles [109-111].

1.4.2 Molecular mechanism and diseased pathology of FXTAS

The two broadly reported molecular mechanisms for the pathogenesis of FXTAS are RNA-gain-of-function and protein-gain-of-function where cellular toxicity is caused due to RNA and protein aggregates, respectively [112-115]. Studying animal (mouse and drosophila) models is undoubtedly very useful for exploring the underlying pathogenic mechanism of the FXTAS [115,116].

1.4.2.1 RNA toxicity via sequestration of RNA binding proteins

Several animal and cellular models have been supported by the evidence of RNA toxicity [117,118]. In adult older males with FXS, full mutation, silenced FMR1 gene, does not develop FXTAS. Secondly, 2-8 fold increased FMR1 mRNA level bearing CGG repeat with no or slightly reduced level of FMRP protein is cytotoxic [118,119]. The contribution of RNA to cellular toxicity is primarily due to the CGG repeats expansion and not because of the elevated level of FMR1 transcript [117]. This signifies that the ectopic expression of expanded CGG repeat is sufficient for causing cellular toxicity such as intracellular aggregate formation, decreased cell viability and promote neuronal cell death whereas overexpression of FMR1 transcript without CGG repeat expansion does not cause such defects [120-122]. A similar idea of FXTAS pathogenesis came from the underlying pathogenic mechanisms of several neurological diseases. A myotonic disorder where RNA toxic gain of function is mediated by expanded CUG triplet repeat at the 3' UTR of DMPK gene [123]. In this model CUG transcript sequester important RNA binding proteins and interrupt the protein function. A similar model for RNA pathogenesis states that CGG repeat-containing RNA sequesters important RNA binding protein and other proteins accumulate in the intranuclear RNA foci. Different animal and human studies demonstrated that a number of RNA binding protein have been discovered such as DiGeorge syndrome critical region 8 (DGCR8) and DROSHA [124], pura [117,125], Src-associated substrate during mitosis of 68-kDa (Sam68) [126], heterogeneous

nuclear ribonucleoprotein A2/B1 (hnRNP A2/B1) [127], and CUG-binding protein 1 (CUGBP1)[127].

While there is no strong proof for CGG repeat mediated sequestration of cellular protein within RNA foci, the consequence of such sequestration on the cell remain unexplored. Additionally, along with FMR1 mRNA, RNA binding proteins are found in the form of ubiquitin inclusion body inside the brain of FXTAS patient. Mass spectrometry and immunohistochemical analysis revealed more than 20 cellular proteins have been detected in the cortex region of FXTAS patient. RNA binding proteins include RNAs hnRNP A2/B1, transactive response binding protein 43kDa (TDP-43), muscleblind-like 1 protein (MBNL), Pur α , Sam68, neurofilament protein like lamin (A/C) & α -internexin, Rm62, CUGBP1, and DGCR8, typically involved in neurological disorders. In the *Drosophila* model of FXTAS overexpression of Pur- α and hnRNP A2/B1 ameliorated the CGG associated neurological phenotype. Studies reported that Pur- α and hnRNP A2/B1 proteins showed direct interaction with CGG repeats RNA. Sequestration of proteins Rm62, Sam68, CUGBP1 and DGCR8 resulted in abnormal regulation of splicing micro RNAs. All these studies indicate that toxic RNA gain of function is mediated by expanded CGG repeat in the FMR1 transcripts.

Heterogeneous nuclear ribonucleoprotein (hnRNP A2/B1), is the RNA biogenic protein that modulates the constitutive and alternative splicing of mRNA. hnRNP A2/B1 has a crucial role in maturing mRNA and is directly associated with spliceosome [128,129]. In the case of FXTAS, hnRNP A2/B1 is sequestered directly on the CGG repeat RNA structure, and this sequestration impedes the normal carrying function of hnRNP A2/B1, leading to abnormal neuronal mRNA transport. In a recent study, FXTAS transgenic *drosophila* model demonstrated that overproduction of hnRNP A2/B1 protein mitigates expanded CGG RNA-mediated eye phenotypes [127]. HnRNP also mediates the interaction of other proteins with CGG RNA sequence such as CUGBP1 and Pur α . CUGBP1 is the CUG triplet repeat RNA binding protein which has key role in DM1 disease. HnRNP directly interacts with premutation CGG RNA and then CUGBP1 interact

with CGG RNA through hnRNP [127,130-132]. The authors demonstrated that increased expression of CUGBP1 protein ameliorates toxic eye phenotypes caused due to CGG repeat-containing fly model. Furthermore, hnRNP A2/B1 also regulates the expression of micro-RNA277, a miRNA whose increased expression could modulate FXTAS associated physiological defects [133].

Pur α is the single standard specific DNA and RNA binding protein, playing an important role in replication, cell proliferation, transcription regulation, oncogenic transformation, and neuronal mRNA transport [134]. Pur α deleted mice developed severe neurological symptoms such as action tremor and spontaneous seizures, megaloccephaly, lower down neuron count in cerebellum and hippocampus region after two weeks of birth, and altered expression and arrangement of dendritic and axonal proteins [134,135]. Pur α is also found in intranuclear RNA aggregates of FXTAS patient and animal models. Several findings suggested that Pur α with hnRNP A2/B1 interact with CGG repeat RNA in a sequence-specific fashion, leading to neuronal cytotoxicity and cell death. Overexpression of Pur α in the FMR1 KO drosophila model resulted in suppression of neurodegeneration in a dose-dependent manner [125,127].

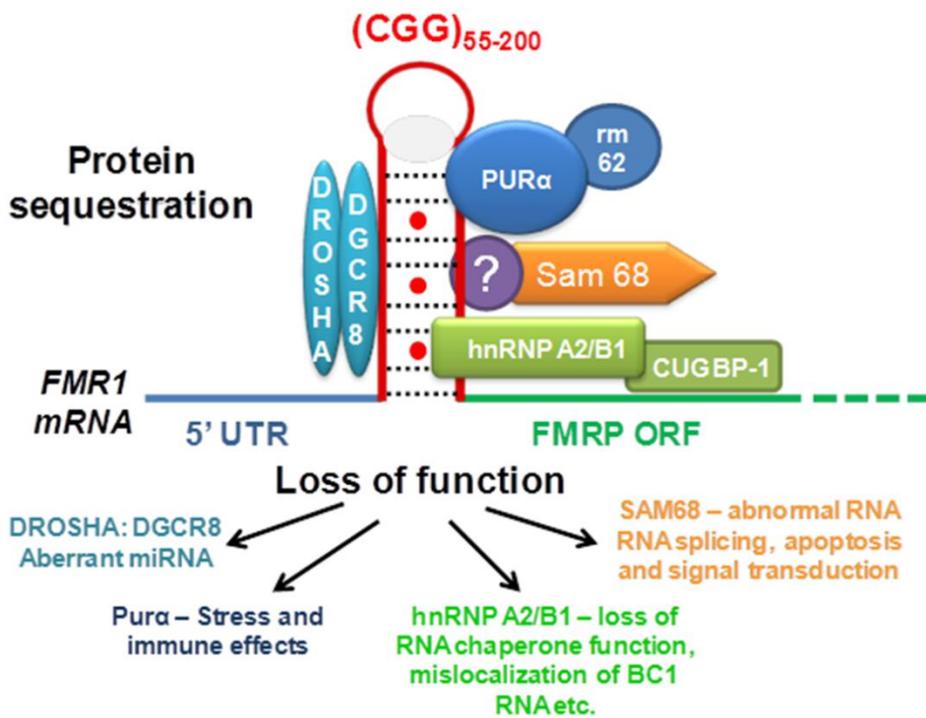


Figure 1.11 Diagrammatic representation of RNA gain of function mechanism.

1.4.2.2 Repeat associated non-AUG translation (RANT)

Another additional mechanism which involves in FXTAS pathogenesis is shown in figurexxx. In this proposed model of a cell, cytotoxicity is provoked by trinucleotide repeat-associated (RAN) translation of protein [136,137]. It is different from RNA intranuclear inclusion which includes non RNA binding protein supported by the cellular and animal models and human brain studies [114,138]. RAN translation first observed in CAG repeat-associated disorders and also found in different nucleotide expansion disorder includes Amyotrophic lateral sclerosis / frontal dementia, Huntington disease, spinocerebellar ataxia type 8 (SCA-8), and Myotonic dystrophy type 1 [136,139-141]. It is believed that CGG triplet repeats translated into proteins even if start codon (AUG) is not present in open reading frame (ORF) of gene. This translation occurs in all three possible combinations of ORFs and generated potentially toxic homopolymeric proteins from single repeats. In the proposed model of FXTAS, CGG repeat triggered RANT initiated in the 5'UTR of FMR1 transcripts and produce cytotoxic FMRpolyGlycine (FMRpolyG) and FMRpolyalanine (FMRpolyA) containing protein aggregates [115]. Several studies reported the FMRpolyG inclusions, specifically in postmortem brain tissue and confirmed by western blot of brain lysate of FXTAS patients. Staining of brain tissue slice particularly for FMRpolyG aggregates was only found in FXTAS patients compared to control patients of Alzheimer's and SCA-3[142]. FMRpolyG positive aggregates also found in other tissues of FXTAS patients such as kidney, heart, adrenal gland, pituitary, and thyroid gland [142,143]. FMRpolyG were observed in two different inducible mouse model (CGG_{dup} KI and CGG_{nih} KI) after 8 weeks of dox induction [144]. Similarly, primary hippocampal neurons derived from FXTAS mouse model observed FMRpolyG inclusion after two weeks of dox induction. The actual initiative mechanism of RAN translation is still unknown. However, Todd group provide evidenced that initiation of RAN translation requires m⁷G-cap, eIF4E (translation initiation factor), EIF4A helicase, and 40S ribosome scanning to initiate translation near cognate region or within CGG repeat of

FMR1 transcripts and is strongly influence by the CGG repeat length and sequential context. Furthermore, CGG repeats associated non-canonical translation (RANT) is only 30-40% efficient as compared to canonical translation [145].

FMRpolyG inclusion produces proteotoxicity due to impairment of the ubiquitin-proteasome system (UPS) [146]. In FXTAS drosophila model study evidenced that UPS impairment led to increased neurodegeneration while over-expression of HSP70 chaperon protein reduced this toxicity. Similar outcomes observed in transfected mammalian cells, FMRpolyG triggers the accumulation of UPS in a length-dependent manner whereas prevention of FMRpolyG production diminishes UPS impairment [146]. In conclusion these findings conclude that FMRpolyG aggregates induce FXTAS pathogenesis *via* bypassing UPS protein quality control pathway [146]. Furthermore, apart from sense *FMRI* mRNA, antisense RNA (*ASFMRI*) also mediate repeat-associated non-aug translation and produce polyproline (ASFMRpolyP), polyalanine (ASFMRpolyA), and polyarginine proteins aggregates. Interestingly, these novel proteins accumulate and colocalize in positive ubiquitin neuronal inclusions [147,148].

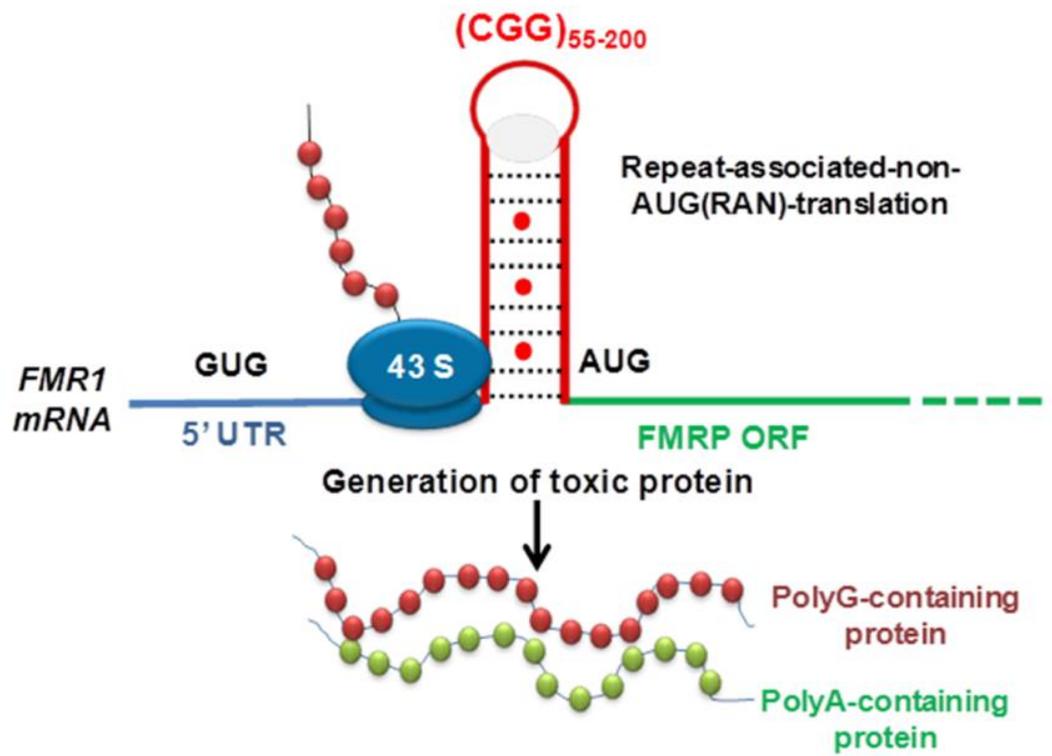


Figure 1.12 Schematic representation of protein gain of function mechanism.

1.4.3 Therapeutic development of FXTAS

Currently, there is no targeted effective therapeutic interventions has been developed that can restore and arrest the FXTAS phenotype. However, numbers of therapeutics strategies are limited potentially befitted specific symptoms of FXTAS patients. All these drugs are neuroprotective drugs which may slow down the progression of FXTAS. We want to put emphasis that till date no randomized, placebo-controlled clinical trial has not been carried out on the patient specifically affected by FXTAS for any drugs agent or procedure.

For the treatment of essential tremor and psychosis, β beta-blockers, primidone and propranolol agents are very frequently used and most probable candidate to start treatment of FXTAS. Propranolol is the most effective agent for the treatment of essential tremor and it may also improve physiologic tremors [149,150]. However, these drugs showed side effects such as impotence, fatigability, sedation, lightheadedness, and depression. A placebo-controlled trial demonstrated effectiveness primidone in essential tremor [151]. It also showed side effect (vomiting, nausea, poor motor coordination, sedation, and confusion) that why it is started with low dose, 25 mg/day or less. Furthermore, topiramate, an antiepileptic with tremor is the second-line treatment, efficacy reported in the placebo-controlled trial. Other β -blockers are sotalol and atenolol are also effective and may have fewer side effects. Alprazolam, a benzodiazepine also studies in some FXTAS patients can be included for therapeutic [152]. Usually in most of the neurological disorder stress condition (anxiety) exacerbated the tremor, benzodiazepines, helping to reduced anxiety symptoms and tremor secondly. Some other medications include nadolol, clozapine, botulinum toxin, clonazepam levetiracetam and nimodipine effectively ameliorate the tremor studied in open-label trial [153-155].

Pre-mutation hippocampal neuronal culture from mice display reduced expression of glutamate transporters and glutamate uptake and increased spontaneous asynchronous Ca^{2+} oscillations [156]. MPEP, an antagonist of mGluR5 receptor

decreased the intracellular Ca^{2+} intake induced by the Glu transporter [156,157]. Moreover, MPEP is showed neuronal toxicity in the human cannot be used for the therapeutic purpose. Thus, this finding marked that inhibition of glutamate transporter of Ca^{2+} signalling could be used as a potential target. Additionally, allopregnanolone, a natural neurosteroid and GABA agonist mitigate the clustered burst firing neuronal cells of FXTAS mice model [156]. Allopregnanolone safely rescues memory and learning deficit phenotype in Alzheimer's and it also stimulates the neurogenesis and survival of neuron cells [158-160]. A clinical trial of allopregnanolone has undergone for the therapeutic of traumatic brain injury and Alzheimer. Thus, there is high hope that allopregnanolone shows high efficacy in clinical trial of FXTAS.

Interestingly, the above-discussed therapeutics approaches target neuronal signalling pathway to treat specific symptoms of FXTAS. Henceforth, it would be better if target the actual trigger of the FXTAS pathogenesis. Few small molecules are developed with high affinity to target CGG hairpin structure of FMR1 transcripts and prevent the sequestration of RNA binding protein and formation of intranuclear RNA aggregates that result to improve FXTAS associated defects. Disney and his group reported small molecule (9-hydroxy-5,11-dimethyl-2-(2-(piperidin-1-yl)ethyl)-6H-pyrido[4, 3-b]carbazol-2-ium) which improved pre-mRNA alternative splicing defect and decreased the number and size of FMRpolyG aggregates [161]. Recently, modularly assembled small molecules (1a and 2H-5-CA-Biotin) reported that reduced the FMRpolgG-GFP protein aggregates and RNA inclusion inside the FXTAS cellular model. It also improved the splicing defect of SMN2 minigenes [162-164]. Apart from synthetic compound natural molecule, Piperine also binds with high affinity and rescue FXTAS associated defect in cell model [165]. Additionally, Qurashi and groups reported a chemical screening approach that revealed small molecules which suppress locomotion deficit condition and improve neurodegeneration in drosophila model [166]. Similarly, rapamycin, a mTOR inhibitor mitigate neurotoxicity via activation of autophagy [167].

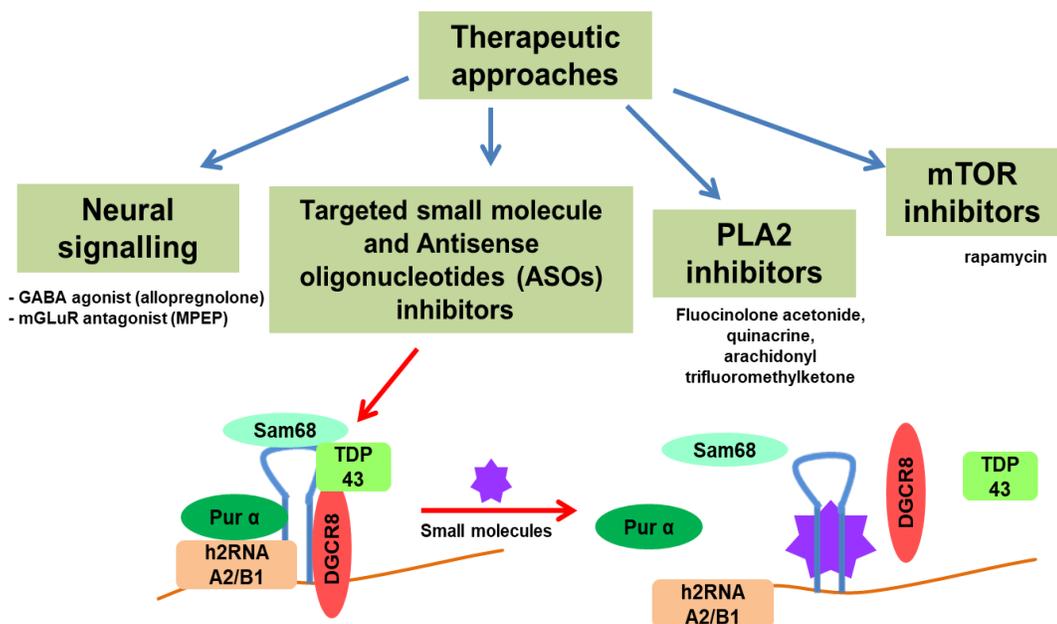


Figure 1.13 Diagram display therapeutic interventions for FXTAS treatment.

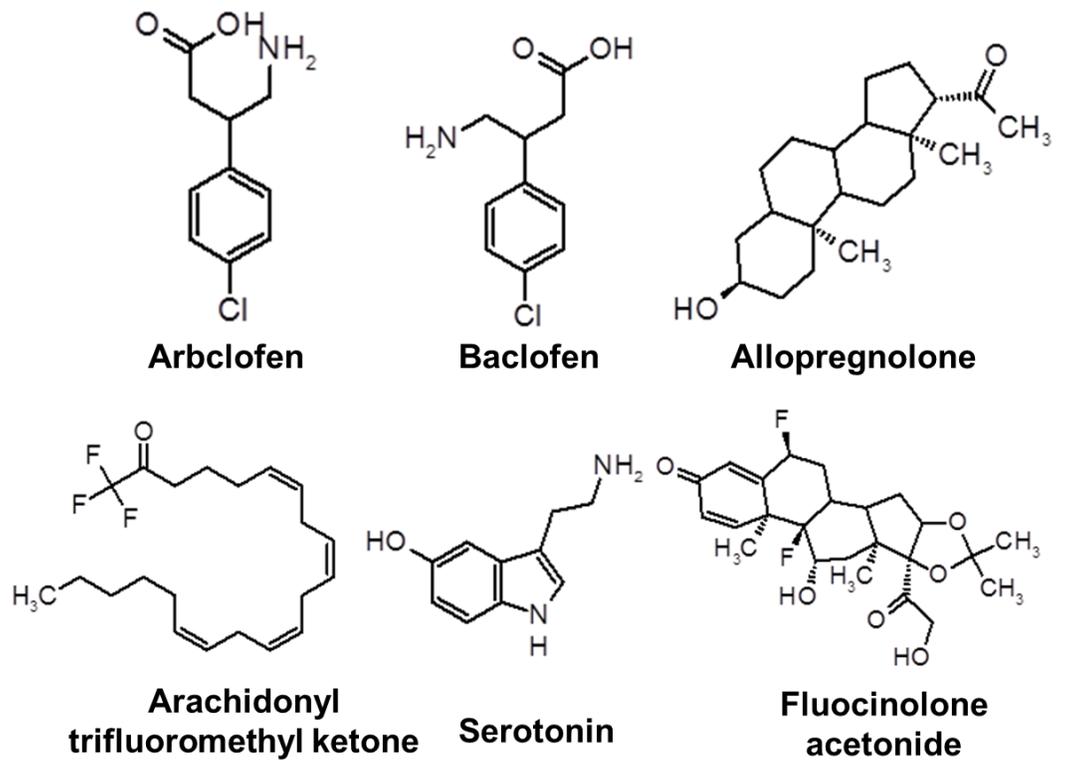


Figure 1.14 Small molecules that target the mGluR and GABA signaling pathway in FXTAS.

1.5 Aim of the thesis

The presence of a dynamic mutation in simple repetitive sequences was reported two decades back when the CAG•CTG repeat expansion was first reported to be associated with spino-bulbar muscular atrophy (SBMA)[168]. After that, more than 40 neurological and neuromuscular diseases have been reported to be associated with these expanded repeats till date[169]. Although most of the disease is caused due to trinucleotide repeats like CAG, CGG, CTG, etc., however, other repeats like tetra, penta and hexa repeats also contribute to some other neurological disorders. These expanded repeats disorders include very common diseases like myotonic dystrophy type 1 (DM1) or fragile X syndrome (FRAXA) to very rare diseases such as spinocerebellar ataxias (SCAs) [170-173]. Some of the trinucleotide repeats such as CGG•CCG and CAG•CTG which contains CG rich regions are very common unstable repeats. These repeats are very short in the normal population, but in the diseased condition, these increased to beyond the threshold limit. Moreover, there is a group of people that carries permutation allele in a manner that the repeats present are longer than the normal person but not enough to show the disease pathology. Although these permutation alleles are non-pathogenic, the propensity to become the pathologic repeats is very high in the future generations [169].

These expanded repeats are present everywhere in the genome like 5' or 3' untranslated region (UTRs), exonic sequences as well as introns. The disease severity is directly correlated to the repeat length, which indicates the early onset and severity of diseases increases with an increase in repeat length. Although in significant cases, the repeats expansion is a prime cause of pathology, some of the rare cases are also evidenced by disease pathology caused due to the repeats contractions[171,172,174,175]. Many model systems including in vitro cellular model system to in vivo model system, have shown that several *cis*-elements and *trans* factors are involved in these repeats instability. Although several advance research in the past two decades has changed our perception to the view of expansion or contraction theory, all these studies have proved the initial

hypothesis, which states the formation of non-B DNA conformation like hairpin or quadruplexes subsequently leading to diseases pathogenesis[19,169,176].

Expanded CGG repeats results in some common neurological disorders like Fragile X syndrome (FRAXA), Fragile X-associated tremor/ataxia syndrome (FXTAS), and Fragile X-associate primary ovarian insufficiency (FXPOI). In these diseases, the expanded CGG repeats are found in the fragile X mental retardation 1 (FMR1) gene in the 5' untranslated (UTRs) region[177]. The range of CGG repeats in normal, permutation, and full mutation lies in up to 55 repeats, 55-200 and more than 200, respectively. FXS has been extensively studied worldwide, mostly in western countries. The estimated prevalence of FXS is 1 in 5000-7000 males and 1 in 4000-6000 female [50]. Although, the prevalence rate of FXS can vary according to racial, ethnic, and founder effect [51,52].

Although the prevalence of these diseases is worldwide, there is no effective treatment for these diseases to date. Only symptomatic treatments targeting the specific symptoms are available. Therefore, targeting the actual cause and rather than symptoms is a prerequisite for development the therapeutics for these neurological disorders. Various therapeutic approaches that involve the antisense oligonucleotides (ASOs), RNAi, and peptides have been utilized recently; however, some limitations such as blood-brain barrier, poor absorption, and allergic reactions prevented these to be effective treatment options[178]. On the other hand, small molecules based therapeutics overcome these limitations; therefore, exploring the small molecules based therapeutics could provide new hope for therapeutics development for these diseases.

In the present thesis work, I have first characterized insight of CGG RNA via NMR spectroscopy, umbrella sampling, and molecular dynamics simulations to understand the nature of mismatch GG base pair with expanded CGG repeats RNA. Furthermore, I have utilized the same information to screen natural as well as synthetic small molecules that target expanded CGG repeats that cause neurological disorders like FXTAS, FXS, and FXPOI. Specifically, molecular

dynamics studies of CGG repeats RNA showed the dynamic nature of internal GG pair, which along with other electronic distribution in atoms, reflects the α -helical form of nucleic acids which in turn provide the binding site for crucial proteins and small molecules. Umbrella sampling showed the stability of GG mismatch in its *syn-anti* conformation as compared to *syn-syn* or *anti-anti* conformation. Additionally, the presence of Na^+ binding pocket near the GG internal loop also stabilizes the *syn-anti* conformation of $(\text{CGG})_{1m}$ and $(\text{CGG})_3$. Therefore, the small molecules which contain a positive charge group (NH_2 and NH) could utilize the binding pocket of the GG internal loop to bind with higher affinity. Conclusively, our, NMR spectroscopy results, MD simulation, and Umbrella sampling reveal that the active compounds those recognize and bind with the repeats RNA that eventually leads to the inhibition of RNA-protein complexes interactions by blocking the protein binding site on RNA and improved the defects caused due to expanded repeats RNA (Chapter-3).

Next, we sought to discover the role of some natural small molecules that target the expanded CGG repeats RNA causing FXS, FXTAS or FXPOI. The binding affinity and selectivity of naturally available compounds such as piperine (Chapter-4) and curcumin (chapter-5) were first assessed via utilizing several biophysical methods like Isothermal calorimetry (ITC), circular Dichroism (CD), Thermal denaturation assay, and NMR spectroscopy. Moreover, the efficacy and potency of these compounds were assessed in FXTAS cellular models, as well as in patient-derived cells. In conclusion, the present work emphasizes the role of naturally available small molecules like piperine and curcumin in ameliorating the disease pathology caused due to the expanded CGG repeats RNA.

Additionally, the chemical similarity search approach was utilized to screen a diverse set of compounds library, National Cancer Institute (NCI), to obtain potential lead compounds that target expanded CGG repeats RNA. The potential lead compounds showed high affinity and selectivity for expanded CGG repeats RNA in several biophysical assays. Subsequently, lead molecules also improved the pre-mRNA alternative splicing defects with no perturbation in global splicing

efficacy and also reduced the FMRpolyG protein aggregates without affecting the downstream expression of the gene. Taken together, this study reveals that the chemical similarity search approach could identify potential small molecules for therapeutics development.

Conclusively, the present thesis work emphasizes the structural insight of CGG RNA via NMR spectroscopic studies, MD simulations and umbrella sampling to develop small molecules based therapeutics. Additionally, the present thesis has also explored the natural, as well as synthetic small molecules that target expanded CGG, repeats RNA causing neurological disorders. The current thesis provides the information of small molecules as well as a scaffold to further design and screen other potent small molecules targeting these expanded repeats RNA.

1.6 Reference

1. Abecasis, G. R., Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth and G. A. McVean. (2012), An integrated map of genetic variation from 1,092 human genomes, *Nature*, 491, 56-65 (10.1038/nature11632)
2. Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh *et al.* (2001), Initial sequencing and analysis of the human genome, *Nature*, 409, 860-921 (10.1038/35057062)
3. Brinkmann, B., M. Klintschar, F. Neuhuber, J. Huhne and B. Rolf. (1998), Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat, *Am J Hum Genet*, 62, 1408-1415 (10.1086/301869)
4. Ellegren, H. (2004), Microsatellites: simple sequences with complex evolution, *Nat Rev Genet*, 5, 435-445 (10.1038/nrg1348)
5. Payseur, B. A., P. Jing and R. J. Haasl. (2011), A genomic portrait of human microsatellite variation, *Mol Biol Evol*, 28, 303-312 (10.1093/molbev/msq198)
6. Sawaya, S., A. Bagshaw, E. Buschiazzo, P. Kumar, S. Chowdhury, M. A. Black and N. Gemmell. (2013), Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements, *PloS one*, 8, e54710-e54710 (10.1371/journal.pone.0054710)
7. Arcot, S. S., Z. Wang, J. L. Weber, P. L. Deininger and M. A. Batzer. (1995), Alu repeats: a source for the genesis of primate microsatellites, *Genomics*, 29, 136-144 (10.1006/geno.1995.1224)
8. Crouau-Roy, B. and I. Clisson. (2000), Evolution of an Alu DNA element of type Sx in the lineage of primates and the origin of an associated tetranucleotide microsatellite, *Genome*, 43, 642-648 (10.1139/g00-033)
9. Grandi, F. C. and W. An. (2013), Non-LTR retrotransposons and microsatellites: Partners in genomic variation, *Mob Genet Elements*, 3, e25674 (10.4161/mge.25674)
10. Clark, R. M., G. L. Dalglish, D. Endres, M. Gomez, J. Taylor and S. I. Bidichandani. (2004), Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu, *Genomics*, 83, 373-383 (10.1016/j.ygeno.2003.09.001)
11. Kurosaki, T., S. Ueda, T. Ishida, K. Abe, K. Ohno and T. Matsuura. (2012), The unstable CCTG repeat responsible for myotonic dystrophy type 2 originates from an AluSx element insertion into an early primate genome, *PloS one*, 7, e38379-e38379 (10.1371/journal.pone.0038379)

12. Buschiazzo, E. and N. J. Gemmell. (2006), The rise, fall and renaissance of microsatellites in eukaryotic genomes, *Bioessays*, 28, 1040-1050 (10.1002/bies.20470)
13. Gadgil, R., J. Barthelemy, T. Lewis and M. Leffak. (2017), Replication stalling and DNA microsatellite instability, *Biophys Chem*, 225, 38-48 (10.1016/j.bpc.2016.11.007)
14. Kim, J. C. and S. M. Mirkin. (2013), The balancing act of DNA repeat expansions, *Current opinion in genetics & development*, 23, 280-288 (10.1016/j.gde.2013.04.009)
15. Mirkin, S. M. (2007), Expandable DNA repeats and human disease, *Nature*, 447, 932-940 (10.1038/nature05977)
16. Polleys, E. J., N. C. M. House and C. H. Freudenreich. (2017), Role of recombination and replication fork restart in repeat instability, *DNA Repair (Amst)*, 56, 156-165 (10.1016/j.dnarep.2017.06.018)
17. Follonier, C., J. Oehler, R. Herrador and M. Lopes. (2013), Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions, *Nat Struct Mol Biol*, 20, 486-494 (10.1038/nsmb.2520)
18. Kerrest, A., R. P. Anand, R. Sundararajan, R. Bermejo, G. Liberi, B. Dujon, C. H. Freudenreich and G.-F. Richard. (2009), SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination, *Nature structural & molecular biology*, 16, 159-167 (10.1038/nsmb.1544)
19. Kang, S., A. Jaworski, K. Ohshima and R. D. Wells. (1995), Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*, *Nat Genet*, 10, 213-218 (10.1038/ng0695-213)
20. Pearson, C. E. and R. R. Sinden. (1996), Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci, *Biochemistry*, 35, 5041-5053 (10.1021/bi9601013)
21. Zhao, J., A. Bacolla, G. Wang and K. M. Vasquez. (2010), Non-B DNA structure-induced genetic instability and evolution, *Cell Mol Life Sci*, 67, 43-62 (10.1007/s00018-009-0131-2)
22. Iyer, R. R., A. Pluciennik, M. Napierala and R. D. Wells. (2015), DNA triplet repeat expansion and mismatch repair, *Annu Rev Biochem*, 84, 199-226 (10.1146/annurev-biochem-060614-034010)
23. Schmidt, M. H. M. and C. E. Pearson. (2016), Disease-associated repeat instability and mismatch repair, *DNA Repair (Amst)*, 38, 117-126 (10.1016/j.dnarep.2015.11.008)

24. Usdin, K., N. C. House and C. H. Freudenreich. (2015), Repeat instability during DNA repair: Insights from model systems, *Crit Rev Biochem Mol Biol*, 50, 142-167 (10.3109/10409238.2014.999192)
25. McMurray, C. T. (2010), Mechanisms of trinucleotide repeat instability during human development, *Nat Rev Genet*, 11, 786-799 (10.1038/nrg2828)
26. Foiry, L., L. Dong, C. Savouret, L. Hubert, H. te Riele, C. Junien and G. Gourdon. (2006), Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice, *Hum Genet*, 119, 520-526 (10.1007/s00439-006-0164-7)
27. Guo, J., L. Gu, M. Leffak and G. M. Li. (2016), MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)_n or (CTG)_n hairpins for error-prone DNA synthesis, *Cell Res*, 26, 775-786 (10.1038/cr.2016.66)
28. Slean, M. M., G. B. Panigrahi, A. L. Castel, A. B. Pearson, A. E. Tomkinson and C. E. Pearson. (2016), Absence of MutSbeta leads to the formation of slipped-DNA for CTG/CAG contractions at primate replication forks, *DNA Repair (Amst)*, 42, 107-118 (10.1016/j.dnarep.2016.04.002)
29. Zhao, X. N., R. Lokanga, K. Allette, I. Gazy, D. Wu and K. Usdin. (2016), A MutSbeta-Dependent Contribution of MutSalpha to Repeat Expansions in Fragile X Premutation Mice?, *PLoS Genet*, 12, e1006190 (10.1371/journal.pgen.1006190)
30. Pluciennik, A., V. Burdett, C. Baitinger, R. R. Iyer, K. Shi and P. Modrich. (2013), Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLalpha endonuclease activation, *Proc Natl Acad Sci U S A*, 110, 12277-12282 (10.1073/pnas.1311325110)
31. Malkova, A. and J. E. Haber. (2012), Mutations arising during repair of chromosome breaks, *Annu Rev Genet*, 46, 455-473 (10.1146/annurev-genet-110711-155547)
32. Kim, J. C., S. T. Harris, T. Dinter, K. A. Shah and S. M. Mirkin. (2017), The role of break-induced replication in large-scale expansions of (CAG)_n/(CTG)_n repeats, *Nat Struct Mol Biol*, 24, 55-60 (10.1038/nsmb.3334)
33. Leffak, M. (2017), Break-induced replication links microsatellite expansion to complex genome rearrangements, *BioEssays*, 39, 1700025 (10.1002/bies.201700025)
34. Metsu, S., J. K. Rainger, K. Debacker, B. Bernhard, L. Rooms, D. Grafodatskaya, R. Weksberg, E. Fombonne, M. S. Taylor, S. W. Scherer

- et al.* (2014), A CGG-repeat expansion mutation in ZNF713 causes FRA7A: association with autistic spectrum disorder in two families, *Hum Mutat*, 35, 1295-1300 (10.1002/humu.22683)
35. Metsu, S., L. Rooms, J. Rainger, M. S. Taylor, H. Bengani, D. I. Wilson, C. S. Chilamakuri, H. Morrison, G. Vandeweyer, E. Reyniers *et al.* (2014), FRA2A is a CGG repeat expansion associated with silencing of AFF3, *PLoS Genet*, 10, e1004242 (10.1371/journal.pgen.1004242)
 36. Rohilla, K. J. and K. T. Gagnon. (2017), RNA biology of disease-associated microsatellite repeat expansions, *Acta Neuropathologica Communications*, 5, 63 (10.1186/s40478-017-0468-y)
 37. Murmann, A. E., J. Yu, P. Opal and M. E. Peter. (2018), Trinucleotide Repeat Expansion Diseases, RNAi, and Cancer, *Trends Cancer*, 4, 684-700 (10.1016/j.trecan.2018.08.004)
 38. Paulson, H. (2018) In Geschwind, D. H., Paulson, H. L. and Klein, C. (eds.), *Handbook of Clinical Neurology*. Elsevier, Vol. 147, pp. 105-123.
 39. Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. C. Ingram *et al.* (2011), Non-ATG-initiated translation directed by microsatellite expansions, *Proceedings of the National Academy of Sciences*, 108, 260-265 (10.1073/pnas.1013343108)
 40. Martin, J. P. and J. Bell. (1943), A PEDIGREE OF MENTAL DEFECT SHOWING SEX-LINKAGE, *Journal of neurology and psychiatry*, 6, 154-157 (10.1136/jnnp.6.3-4.154)
 41. Lubs, H. A. (1969), A marker X chromosome, *American journal of human genetics*, 21, 231-244
 42. Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, Y. H. Fu, D. P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria, F. P. Zhang *et al.* (1991), Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, *Cell*, 65, 905-914 (10.1016/0092-8674(91)90397-h)
 43. Ashley, C. T., J. S. Sutcliffe, C. B. Kunst, H. A. Leiner, E. E. Eichler, D. L. Nelson and S. T. Warren. (1993), Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat, *Nat Genet*, 4, 244-251 (10.1038/ng0793-244)
 44. Eichler, E. E., S. Richards, R. A. Gibbs and D. L. Nelson. (1993), Fine structure of the human FMR1 gene, *Hum Mol Genet*, 2, 1147-1153 (10.1093/hmg/2.8.1147)
 45. Kunst, C. B. and S. T. Warren. (1994), Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles, *Cell*, 77, 853-861 (10.1016/0092-8674(94)90134-1)

46. Hall, D. A. (2014), In the Gray Zone in the Fragile X Gene: What are the Key Unanswered Clinical and Biological Questions?, *Tremor and other hyperkinetic movements (New York, N.Y.)*, 4, 208-208 (10.7916/D8NG4NP3)
47. Lozano, R., C. A. Rosero and R. J. Hagerman. (2014), Fragile X spectrum disorders, *Intractable & rare diseases research*, 3, 134-146 (10.5582/irdr.2014.01022)
48. McLennan, Y., J. Polussa, F. Tassone and R. Hagerman. (2011), Fragile x syndrome, *Curr Genomics*, 12, 216-224 (10.2174/138920211795677886)
49. Feng, Y., F. Zhang, L. K. Lokey, J. L. Chastain, L. Lakkis, D. Eberhart and S. T. Warren. (1995), Translational suppression by trinucleotide repeat expansion at FMR1, *Science*, 268, 731-734 (10.1126/science.7732383)
50. Hunter, J., O. Rivero-Arias, A. Angelov, E. Kim, I. Fotheringham and J. Leal. (2014), Epidemiology of fragile X syndrome: a systematic review and meta-analysis, *Am J Med Genet A*, 164a, 1648-1658 (10.1002/ajmg.a.36511)
51. Coffee, B., K. Keith, I. Albizua, T. Malone, J. Mowrey, S. L. Sherman and S. T. Warren. (2009), Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA, *Am J Hum Genet*, 85, 503-514 (10.1016/j.ajhg.2009.09.007)
52. Levesque, S., C. Dombrowski, M. L. Morel, R. Rehel, J. S. Cote, J. Bussieres, K. Morgan and F. Rousseau. (2009), Screening and instability of FMR1 alleles in a prospective sample of 24,449 mother-newborn pairs from the general population, *Clin Genet*, 76, 511-523 (10.1111/j.1399-0004.2009.01237.x)
53. Sethna, F., C. Moon and H. Wang. (2014), From FMRP function to potential therapies for fragile X syndrome, *Neurochemical research*, 39, 1016-1031 (10.1007/s11064-013-1229-3)
54. Sidorov, M. S., B. D. Auerbach and M. F. Bear. (2013), Fragile X mental retardation protein and synaptic plasticity, *Mol Brain*, 6, 15 (10.1186/1756-6606-6-15)
55. Garber, K. B., J. Visootsak and S. T. Warren. (2008), Fragile X syndrome, *Eur J Hum Genet*, 16, 666-672 (10.1038/ejhg.2008.61)
56. Saldarriaga, W., F. Tassone, L. Y. González-Teshima, J. V. Forero-Forero, S. Ayala-Zapata and R. Hagerman. (2014), Fragile X syndrome, *Colombia medica (Cali, Colombia)*, 45, 190-198
57. Ciaccio, C., L. Fontana, D. Milani, S. Tabano, M. Miozzo and S. Esposito. (2017), Fragile X syndrome: a review of clinical and molecular diagnoses, *Italian journal of pediatrics*, 43, 39-39 (10.1186/s13052-017-0355-y)

58. Hagerman, R. J., E. Berry-Kravis, W. E. Kaufmann, M. Y. Ono, N. Tartaglia, A. Lachiewicz, R. Kronk, C. Delahunty, D. Hessel, J. Visootsak *et al.* (2009), Advances in the treatment of fragile X syndrome, *Pediatrics*, 123, 378-390 (10.1542/peds.2008-0317)
59. Tolmie, J. (2002), Fragile X Syndrome - Diagnosis, Treatment and Research, *Journal of Medical Genetics*, 39, 783-783 (10.1136/jmg.39.10.783)
60. Naumann, A., C. Kraus, A. Hoogeveen, C. M. Ramirez and W. Doerfler. (2014), Stable DNA methylation boundaries and expanded trinucleotide repeats: role of DNA insertions, *J Mol Biol*, 426, 2554-2566 (10.1016/j.jmb.2014.04.025)
61. Cho, D. H., C. P. Thienes, S. E. Mahoney, E. Analau, G. N. Filippova and S. J. Tapscott. (2005), Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF, *Mol Cell*, 20, 483-489 (10.1016/j.molcel.2005.09.002)
62. Herman, D., K. Jenssen, R. Burnett, E. Soragni, S. L. Perlman and J. M. Gottesfeld. (2006), Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia, *Nat Chem Biol*, 2, 551-558 (10.1038/nchembio815)
63. Belzil, V. V., P. O. Bauer, M. Prudencio, T. F. Gendron, C. T. Stetler, I. K. Yan, L. Pregent, L. Daugherty, M. C. Baker, R. Rademakers *et al.* (2013), Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood, *Acta Neuropathol*, 126, 895-905 (10.1007/s00401-013-1199-1)
64. Kumari, D. and K. Usdin. (2014), Polycomb group complexes are recruited to reactivated FMR1 alleles in Fragile X syndrome in response to FMR1 transcription, *Hum Mol Genet*, 23, 6575-6583 (10.1093/hmg/ddu378)
65. Tabolacci, E., U. Moscato, F. Zalfa, C. Bagni, P. Chiurazzi and G. Neri. (2008), Epigenetic analysis reveals a euchromatic configuration in the FMR1 unmethylated full mutations, *Eur J Hum Genet*, 16, 1487-1498 (10.1038/ejhg.2008.130)
66. Biacsi, R., D. Kumari and K. Usdin. (2008), SIRT1 inhibition alleviates gene silencing in Fragile X mental retardation syndrome, *PLoS Genet*, 4, e1000017 (10.1371/journal.pgen.1000017)
67. Berry-Kravis, E., D. Hessel, S. Coffey, C. Hervey, A. Schneider, J. Yuhas, J. Hutchison, M. Snape, M. Tranfaglia, D. V. Nguyen *et al.* (2009), A pilot open label, single dose trial of fenobam in adults with fragile X syndrome, *Journal of medical genetics*, 46, 266-271 (10.1136/jmg.2008.063701)
68. Yan, Q. J., M. Rammal, M. Tranfaglia and R. P. Bauchwitz. (2005), Suppression of two major Fragile X Syndrome mouse model phenotypes

- by the mGluR5 antagonist MPEP, *Neuropharmacology*, 49, 1053-1066 (10.1016/j.neuropharm.2005.06.004)
69. Su, T., H. X. Fan, T. Jiang, W. W. Sun, W. Y. Den, M. M. Gao, S. Q. Chen, Q. H. Zhao and Y. H. Yi. (2011), Early continuous inhibition of group 1 mGlu signaling partially rescues dendritic spine abnormalities in the Fmr1 knockout mouse model for fragile X syndrome, *Psychopharmacology (Berl)*, 215, 291-300 (10.1007/s00213-010-2130-2)
70. Michalon, A., M. Sidorov, T. M. Ballard, L. Ozmen, W. Spooren, J. G. Wettstein, G. Jaeschke, M. F. Bear and L. Lindemann. (2012), Chronic pharmacological mGlu5 inhibition corrects fragile X in adult mice, *Neuron*, 74, 49-56 (10.1016/j.neuron.2012.03.009)
71. Jacquemont, S., A. Curie, V. des Portes, M. G. Torrioli, E. Berry-Kravis, R. J. Hagerman, F. J. Ramos, K. Cornish, Y. He, C. Paulding *et al.* (2011), Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056, *Sci Transl Med*, 3, 64ra61 (10.1126/scitranslmed.3001708)
72. Berg, D., J. Godau, C. Trenkwalder, K. Eggert, I. Csoti, A. Storch, H. Huber, M. Morelli-Canelo, M. Stamelou, V. Ries *et al.* (2011), AFQ056 treatment of levodopa-induced dyskinesias: results of 2 randomized controlled trials, *Mov Disord*, 26, 1243-1250 (10.1002/mds.23616)
73. Bagni, C., F. Tassone, G. Neri and R. Hagerman. (2012), Fragile X syndrome: causes, diagnosis, mechanisms, and therapeutics, *The Journal of clinical investigation*, 122, 4314-4322 (10.1172/JCI63141)
74. Vranesic, I., S. Ofner, P. J. Flor, G. Bilbe, R. Bouhelal, A. Enz, S. Desrayaud, K. McAllister, R. Kuhn and F. Gasparini. (2014), AFQ056/mavoglurant, a novel clinically effective mGluR5 antagonist: identification, SAR and pharmacological characterization, *Bioorg Med Chem*, 22, 5790-5803 (10.1016/j.bmc.2014.09.033)
75. Levenson, J., S. Hayashi, F. M. de Vrij, S. K. Koekkoek, H. C. van der Linde, I. Nieuwenhuizen, C. Song, R. A. Buijsen, A. S. Pop, B. Gomez-mancilla *et al.* (2011), AFQ056, a new mGluR5 antagonist for treatment of fragile X syndrome, *Neurobiol Dis*, 42, 311-317 (10.1016/j.nbd.2011.01.022)
76. Centonze, D., S. Rossi, V. Mercaldo, I. Napoli, M. T. Ciotti, V. De Chiara, A. Musella, C. Prosperetti, P. Calabresi, G. Bernardi *et al.* (2008), Abnormal striatal GABA transmission in the mouse model for the fragile X syndrome, *Biol Psychiatry*, 63, 963-973 (10.1016/j.biopsych.2007.09.008)

77. Curia, G., T. Papouin, P. Seguela and M. Avoli. (2009), Downregulation of tonic GABAergic inhibition in a mouse model of fragile X syndrome, *Cereb Cortex*, 19, 1515-1520 (10.1093/cercor/bhn159)
78. D'Hulst, C., N. De Geest, S. P. Reeve, D. Van Dam, P. P. De Deyn, B. A. Hassan and R. F. Kooy. (2006), Decreased expression of the GABAA receptor in fragile X syndrome, *Brain Res*, 1121, 238-245 (10.1016/j.brainres.2006.08.115)
79. Olmos-Serrano, J. L., S. M. Paluszkiwicz, B. S. Martin, W. E. Kaufmann, J. G. Corbin and M. M. Huntsman. (2010), Defective GABAergic neurotransmission and pharmacological rescue of neuronal hyperexcitability in the amygdala in a mouse model of fragile X syndrome, *J Neurosci*, 30, 9929-9938 (10.1523/jneurosci.1714-10.2010)
80. Erickson, C. A., J. E. Mullett and C. J. McDougle. (2010), Brief report: acamprosate in fragile X syndrome, *J Autism Dev Disord*, 40, 1412-1416 (10.1007/s10803-010-0988-9)
81. Pacey, L. K., S. P. Heximer and D. R. Hampson. (2009), Increased GABA(B) receptor-mediated signaling reduces the susceptibility of fragile X knockout mice to audiogenic seizures, *Mol Pharmacol*, 76, 18-24 (10.1124/mol.109.056127)
82. Berry-Kravis, E. M., D. Hessler, B. Rathmell, P. Zarevics, M. Cherubini, K. Walton-Bowen, Y. Mu, D. V. Nguyen, J. Gonzalez-Heydrich, P. P. Wang *et al.* (2012), Effects of STX209 (arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: a randomized, controlled, phase 2 trial, *Sci Transl Med*, 4, 152ra127 (10.1126/scitranslmed.3004214)
83. Ligsay, A., A. Van Dijck, D. V. Nguyen, R. Lozano, Y. Chen, E. S. Bickel, D. Hessler, A. Schneider, K. Angkustsiri, F. Tassone *et al.* (2017), A randomized double-blind, placebo-controlled trial of ganaxolone in children and adolescents with fragile X syndrome, *Journal of neurodevelopmental disorders*, 9, 26-26 (10.1186/s11689-017-9207-8)
84. Braat, S., C. D'Hulst, I. Heulens, S. De Rubeis, E. Mientjes, D. L. Nelson, R. Willemsen, C. Bagni, D. Van Dam, P. P. De Deyn *et al.* (2015), The GABAA receptor is an FMRP target with therapeutic potential in fragile X syndrome, *Cell Cycle*, 14, 2985-2995 (10.4161/15384101.2014.989114)
85. Heulens, I., C. D'Hulst, D. Van Dam, P. P. De Deyn and R. F. Kooy. (2012), Pharmacological treatment of fragile X syndrome with GABAergic drugs in a knockout mouse model, *Behav Brain Res*, 229, 244-249 (10.1016/j.bbr.2012.01.031)
86. Olmos-Serrano, J. L., J. G. Corbin and M. P. Burns. (2011), The GABA(A) receptor agonist THIP ameliorates specific behavioral deficits

- in the mouse model of fragile X syndrome, *Dev Neurosci*, 33, 395-403 (10.1159/000332884)
87. Renoux, A. J. and P. K. Todd. (2012), Neurodegeneration the RNA way, *Progress in neurobiology*, 97, 173-189 (10.1016/j.pneurobio.2011.10.006)
 88. Hagerman, R. J. and P. Hagerman. (2016), Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management, *Nat Rev Neurol*, 12, 403-412 (10.1038/nrneurol.2016.82)
 89. Willemsen, R., J. Levenega and B. A. Oostra. (2011), CGG repeat in the FMR1 gene: size matters, *Clin Genet*, 80, 214-225 (10.1111/j.1399-0004.2011.01723.x)
 90. Jacquemont, S., R. J. Hagerman, M. A. Leehey, D. A. Hall, R. A. Levine, J. A. Brunberg, L. Zhang, T. Jardini, L. W. Gane, S. W. Harris *et al.* (2004), Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population, *Jama*, 291, 460-469 (10.1001/jama.291.4.460)
 91. Rodriguez-Revenga, L., I. Madrigal, J. Pagonabarraga, M. Xunclà, C. Badenas, J. Kulisevsky, B. Gomez and M. Milà. (2009), Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families, *European journal of human genetics : EJHG*, 17, 1359-1362 (10.1038/ejhg.2009.51)
 92. Hall, D. A. and J. A. O'Keefe. (2012), Fragile x-associated tremor ataxia syndrome: the expanding clinical picture, pathophysiology, epidemiology, and update on treatment, *Tremor and other hyperkinetic movements (New York, N.Y.)*, 2, tre-02-56-352-351 (10.7916/D8HD7TDS)
 93. Park, J.-H., W. Jang, J. Youn, C.-S. Ki, B. J. Kim, H.-T. Kim, E. D. Louis and J. W. Cho. (2019), Prevalence of fragile X-associated tremor/ataxia syndrome: A survey of essential tremor patients with cerebellar signs or extrapyramidal signs, *Brain and behavior*, 9, e01337-e01337 (10.1002/brb3.1337)
 94. Hagerman, P. J. and R. J. Hagerman. (2007), Fragile X-associated tremor/ataxia syndrome—an older face of the fragile X gene, *Nature Clinical Practice Neurology*, 3, 107-112 (10.1038/ncpneuro0373)
 95. Jacquemont, S., M. A. Leehey, R. J. Hagerman, L. A. Beckett and P. J. Hagerman. (2006), Size bias of fragile X premutation alleles in late-onset movement disorders, *Journal of medical genetics*, 43, 804-809 (10.1136/jmg.2006.042374)
 96. Muzar, Z., R. Lozano, A. Schneider, P. E. Adams, S. M. H. Faradz, F. Tassone and R. J. Hagerman. (2015), Methadone use in a male with the FMRI premutation and FXTAS, *American journal of medical genetics. Part A*, 167, 1354-1359 (10.1002/ajmg.a.37030)

97. O'Dwyer, J. P., C. Clabby, J. Crown, D. E. Barton and M. Hutchinson. (2005), Fragile X-associated tremor/ataxia syndrome presenting in a woman after chemotherapy, *Neurology*, 65, 331-332 (10.1212/01.wnl.0000168865.36352.53)
98. Chen, Y., F. Tassone, R. F. Berman, P. J. Hagerman, R. J. Hagerman, R. Willemsen and I. N. Pessah. (2010), Murine hippocampal neurons expressing Fmr1 gene premutations show early developmental deficits and late degeneration, *Hum Mol Genet*, 19, 196-208 (10.1093/hmg/ddp479)
99. Jacquemont, S., F. Farzin, D. Hall, M. Leehey, F. Tassone, L. Gane, L. Zhang, J. Grigsby, T. Jardini, F. Lewin *et al.* (2004), Aging in individuals with the FMR1 mutation, *American journal of mental retardation : AJMR*, 109, 154-164 (10.1352/0895-8017(2004)109<154:AIWTF>2.0.CO;2)
100. Hagerman, P. (2013), Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms, *Acta neuropathologica*, 126, 1-19 (10.1007/s00401-013-1138-1)
101. Leehey, M. A., E. Berry-Kravis, S. J. Min, D. A. Hall, C. D. Rice, L. Zhang, J. Grigsby, C. M. Greco, A. Reynolds, R. Lara *et al.* (2007), Progression of tremor and ataxia in male carriers of the FMR1 premutation, *Mov Disord*, 22, 203-206 (10.1002/mds.21252)
102. Leehey, M. A. (2009), Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment, *Journal of investigative medicine : the official publication of the American Federation for Clinical Research*, 57, 830-836 (10.2310/JIM.0b013e3181af59c4)
103. Bourgeois, J. A., A. L. Seritan, E. M. Casillas, D. Hessel, A. Schneider, Y. Yang, I. Kaur, J. B. Cogswell, D. V. Nguyen and R. J. Hagerman. (2011), Lifetime prevalence of mood and anxiety disorders in fragile X premutation carriers, *The Journal of clinical psychiatry*, 72, 175-182 (10.4088/JCP.09m05407blu)
104. Cornish, K. M., C. S. Kogan, L. Li, J. Turk, S. Jacquemont and R. J. Hagerman. (2009), Lifespan changes in working memory in fragile X premutation males, *Brain and cognition*, 69, 551-558 (10.1016/j.bandc.2008.11.006)
105. Coffey, S. M., K. Cook, N. Tartaglia, F. Tassone, D. V. Nguyen, R. Pan, H. E. Bronsky, J. Yuhas, M. Borodyanskaya, J. Grigsby *et al.* (2008), Expanded Clinical Phenotype of Women With the FMR1 Premutation, *American journal of medical genetics. Part A*, 146A, 1009-1016 (10.1002/ajmg.a.32060)
106. Au, J., R. S. Akins, L. Berkowitz-Sutherland, H. T. Tang, Y. Chen, A. Boyd, F. Tassone, D. V. Nguyen and R. Hagerman. (2013), Prevalence

- and risk of migraine headaches in adult fragile X premutation carriers, *Clin Genet*, 84, 546-551 (10.1111/cge.12109)
107. Bailey, D. B., Jr., M. Raspa, M. Olmsted and D. B. Holiday. (2008), Co-occurring conditions associated with FMR1 gene variations: findings from a national parent survey, *Am J Med Genet A*, 146a, 2060-2069 (10.1002/ajmg.a.32439)
 108. Hagerman, R. J., M. Leehey, W. Heinrichs, F. Tassone, R. Wilson, J. Hills, J. Grigsby, B. Gage and P. J. Hagerman. (2001), Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X, *Neurology*, 57, 127-130 (10.1212/wnl.57.1.127)
 109. Jacquemont, S., R. J. Hagerman, M. Leehey, J. Grigsby, L. Zhang, J. A. Brunberg, C. Greco, V. Des Portes, T. Jardini, R. Levine *et al.* (2003), Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates, *Am J Hum Genet*, 72, 869-878 (10.1086/374321)
 110. Loesch, D. Z., K. Kotschet, N. Trost, C. M. Greco, G. Kinsella, H. R. Slater, A. Venn and M. Horne. (2011), White matter changes in basis pontis in small expansion FMR1 allele carriers with parkinsonism, *Am J Med Genet B Neuropsychiatr Genet*, 156b, 502-506 (10.1002/ajmg.b.31189)
 111. Brunberg, J. A., S. Jacquemont, R. J. Hagerman, E. M. Berry-Kravis, J. Grigsby, M. A. Leehey, F. Tassone, W. T. Brown, C. M. Greco and P. J. Hagerman. (2002), Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction, *AJNR Am J Neuroradiol*, 23, 1757-1766
 112. Tassone, F., R. J. Hagerman, A. K. Taylor, L. W. Gane, T. E. Godfrey and P. J. Hagerman. (2000), Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome, *Am J Hum Genet*, 66, 6-15 (10.1086/302720)
 113. Pretto, D. I., G. Mendoza-Morales, J. Lo, R. Cao, A. Hadd, G. J. Latham, B. Durbin-Johnson, R. Hagerman and F. Tassone. (2014), CGG allele size somatic mosaicism and methylation in FMR1 premutation alleles, *J Med Genet*, 51, 309-318 (10.1136/jmedgenet-2013-102021)
 114. Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent *et al.* (2017), Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome, *Neuron*, 93, 331-347 (10.1016/j.neuron.2016.12.016)
 115. Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur *et al.* (2013), CGG repeat-associated translation mediates neurodegeneration in fragile X

- tremor ataxia syndrome, *Neuron*, 78, 440-455 (10.1016/j.neuron.2013.03.026)
116. Berman, R. F., R. A. Buijsen, K. Usdin, E. Pintado, F. Kooy, D. Pretto, I. N. Pessah, D. L. Nelson, Z. Zalewski, N. Charlet-Bergeurand *et al.* (2014), Mouse models of the fragile X premutation and fragile X-associated tremor/ataxia syndrome, *Journal of neurodevelopmental disorders*, 6, 25-25 (10.1186/1866-1955-6-25)
 117. Jin, P., D. C. Zarnescu, F. Zhang, C. E. Pearson, J. C. Lucchesi, K. Moses and S. T. Warren. (2003), RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*, *Neuron*, 39, 739-747 (10.1016/s0896-6273(03)00533-6)
 118. Fernandez, J. J., R. Martinez, E. Andujar, M. Perez-Alegre, A. Costa, V. Bonilla-Henao, F. Sobrino, C. O. Pintado and E. Pintado. (2012), Gene expression profiles in the cerebellum of transgenic mice over expressing the human FMR1 gene with CGG repeats in the normal range, *Genet Mol Res*, 11, 467-483 (10.4238/2012.March.1.4)
 119. Handa, V., D. Goldwater, D. Stiles, M. Cam, G. Poy, D. Kumari and K. Usdin. (2005), Long CGG-repeat tracts are toxic to human cells: Implications for carriers of Fragile X premutation alleles, *FEBS Letters*, 579, 2702-2708 (10.1016/j.febslet.2005.04.004)
 120. Hoem, G., C. R. Raske, D. Garcia-Arocena, F. Tassone, E. Sanchez, A. L. Ludwig, C. K. Iwahashi, M. Kumar, J. E. Yang and P. J. Hagerman. (2011), CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS, *Hum Mol Genet*, 20, 2161-2170 (10.1093/hmg/ddr101)
 121. Hashem, V., J. N. Galloway, M. Mori, R. Willemsen, B. A. Oostra, R. Paylor and D. L. Nelson. (2009), Ectopic expression of CGG containing mRNA is neurotoxic in mammals, *Hum Mol Genet*, 18, 2443-2451 (10.1093/hmg/ddp182)
 122. Arocena, D. G., C. K. Iwahashi, N. Won, A. Beilina, A. L. Ludwig, F. Tassone, P. H. Schwartz and P. J. Hagerman. (2005), Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells, *Hum Mol Genet*, 14, 3661-3671 (10.1093/hmg/ddi394)
 123. Meola, G. and R. Cardani. (2015), Myotonic dystrophies: An update on clinical aspects, genetic, pathology, and molecular pathomechanisms, *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1852, 594-606 (<https://doi.org/10.1016/j.bbadis.2014.05.019>)
 124. Sellier, C., F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page *et al.* (2013), Sequestration of

- DROSHA and DGCR8 by Expanded CGG RNA Repeats Alters MicroRNA Processing in Fragile X-Associated Tremor/Ataxia Syndrome, *Cell reports*, 3, 869-880 (10.1016/j.celrep.2013.02.004)
125. Jin, P., R. Duan, A. Qurashi, Y. Qin, D. Tian, T. C. Rosser, H. Liu, Y. Feng and S. T. Warren. (2007), Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome, *Neuron*, 55, 556-564 (10.1016/j.neuron.2007.07.020)
 126. Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott *et al.* (2010), Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *Embo j*, 29, 1248-1261 (10.1038/emboj.2010.21)
 127. Sofola, O. A., P. Jin, Y. Qin, R. Duan, H. Liu, M. de Haro, D. L. Nelson and J. Botas. (2007), RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS, *Neuron*, 55, 565-571 (10.1016/j.neuron.2007.07.021)
 128. Geuens, T., D. Bouhy and V. Timmerman. (2016), The hnRNP family: insights into their role in health and disease, *Human genetics*, 135, 851-867 (10.1007/s00439-016-1683-5)
 129. Chaudhury, A., P. Chander and P. H. Howe. (2010), Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles, *RNA (New York, N.Y.)*, 16, 1449-1462 (10.1261/rna.2254110)
 130. Timchenko, L. T., J. W. Miller, N. A. Timchenko, D. R. DeVore, K. V. Datar, L. Lin, R. Roberts, C. T. Caskey and M. S. Swanson. (1996), Identification of a (CUG)_n triplet repeat RNA-binding protein and its expression in myotonic dystrophy, *Nucleic Acids Res*, 24, 4407-4414 (10.1093/nar/24.22.4407)
 131. Timchenko, N. A., Z. J. Cai, A. L. Welm, S. Reddy, T. Ashizawa and L. T. Timchenko. (2001), RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1, *J Biol Chem*, 276, 7820-7826 (10.1074/jbc.M005960200)
 132. Timchenko, N. A., R. Patel, P. Iakova, Z. J. Cai, L. Quan and L. T. Timchenko. (2004), Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis, *J Biol Chem*, 279, 13129-13139 (10.1074/jbc.M312923200)

133. Tan, H., M. Poidevin, H. Li, D. Chen and P. Jin. (2012), MicroRNA-277 modulates the neurodegeneration caused by Fragile X premutation rCGG repeats, *PLoS Genet*, 8, e1002681 (10.1371/journal.pgen.1002681)
134. Khalili, K., L. Del Valle, V. Muralidharan, W. J. Gault, N. Darbinian, J. Otte, E. Meier, E. M. Johnson, D. C. Daniel, Y. Kinoshita *et al.* (2003), Puralpha is essential for postnatal brain development and developmentally coupled cellular proliferation as revealed by genetic inactivation in the mouse, *Mol Cell Biol*, 23, 6857-6875 (10.1128/mcb.23.19.6857-6875.2003)
135. Hokkanen, S., H. M. Feldmann, H. Ding, C. K. Jung, L. Bojarski, I. Renner-Muller, U. Schuller, H. Kretzschmar, E. Wolf and J. Herms. (2012), Lack of Pur-alpha alters postnatal brain development and causes megalencephaly, *Hum Mol Genet*, 21, 473-484 (10.1093/hmg/ddr476)
136. Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. Ingram *et al.* (2011), Non-ATG-initiated translation directed by microsatellite expansions, *Proc Natl Acad Sci U S A*, 108, 260-265 (10.1073/pnas.1013343108)
137. Cleary, J. D., A. Pattamatta and L. P. W. Ranum. (2018), Repeat-associated non-ATG (RAN) translation, *J Biol Chem*, 293, 16127-16141 (10.1074/jbc.R118.003237)
138. Pearson, C. E. (2011), Repeat associated non-ATG translation initiation: one DNA, two transcripts, seven reading frames, potentially nine toxic entities!, *PLoS Genet*, 7, e1002018 (10.1371/journal.pgen.1002018)
139. Zu, T., Y. Liu, M. Banez-Coronel, T. Reid, O. Pletnikova, J. Lewis, T. M. Miller, M. B. Harms, A. E. Falchook, S. H. Subramony *et al.* (2013), RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia, *Proc Natl Acad Sci U S A*, 110, E4968-4977 (10.1073/pnas.1315438110)
140. Ash, P. E., K. F. Bieniek, T. F. Gendron, T. Caulfield, W. L. Lin, M. Dejesus-Hernandez, M. M. van Blitterswijk, K. Jansen-West, J. W. Paul, 3rd, R. Rademakers *et al.* (2013), Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS, *Neuron*, 77, 639-646 (10.1016/j.neuron.2013.02.004)
141. Mori, K., S. M. Weng, T. Arzberger, S. May, K. Rentzsch, E. Kremmer, B. Schmid, H. A. Kretzschmar, M. Cruts, C. Van Broeckhoven *et al.* (2013), The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS, *Science*, 339, 1335-1338 (10.1126/science.1232927)

142. Buijsen, R. A. M., C. Sellier, L.-A. W. F. M. Severijnen, M. Oulad-Abdelghani, R. F. M. Verhagen, R. F. Berman, N. Charlet-Berguerand, R. Willemsen and R. K. Hukema. (2014), FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome, *Acta Neuropathologica Communications*, 2, 162 (10.1186/s40478-014-0162-2)
143. Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K.-c. Chen, K. M. Scaglione, V. Basrur *et al.* (2013), CGG Repeat Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome, *Neuron*, 78, 10.1016/j.neuron.2013.1003.1026 (10.1016/j.neuron.2013.03.026)
144. Hukema, R. K., R. A. M. Buijsen, M. Schonewille, C. Raske, L.-A. W. F. M. Severijnen, I. Nieuwenhuizen-Bakker, R. F. M. Verhagen, L. van Dessel, A. Maas, N. Charlet-Berguerand *et al.* (2015), Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS, *Human molecular genetics*, 24, 4948-4957 (10.1093/hmg/ddv216)
145. Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm and P. K. Todd. (2016), CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins, *Mol Cell*, 62, 314-322 (10.1016/j.molcel.2016.02.034)
146. Oh, S. Y., F. He, A. Krans, M. Frazer, J. P. Taylor, H. L. Paulson and P. K. Todd. (2015), RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome, *Human molecular genetics*, 24, 4317-4326 (10.1093/hmg/ddv165)
147. Krans, A., M. G. Kearse and P. K. Todd. (2016), Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome, *Annals of neurology*, 80, 871-881 (10.1002/ana.24800)
148. Glineburg, M. R., P. K. Todd, N. Charlet-Berguerand and C. Sellier. (2018), Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome, *Brain Research*, 1693, 43-54 (<https://doi.org/10.1016/j.brainres.2018.02.006>)
149. Calzetti, S., E. Sasso, M. Baratti and R. Fava. (1990), Clinical and computer-based assessment of long-term therapeutic efficacy of propranolol in essential tremor, *Acta Neurol Scand*, 81, 392-396 (10.1111/j.1600-0404.1990.tb00982.x)
150. Caccia, M. R., M. Osio, V. Galimberti, G. Cataldi and A. Mangoni. (1989), Propranolol, clonidine, urapidil and trazodone infusion in essential

- tremor: a double-blind crossover trial, *Acta Neurol Scand*, 79, 379-383 (10.1111/j.1600-0404.1989.tb03804.x)
151. Koller, W. C. and V. L. Royse. (1986), Efficacy of primidone in essential tremor, *Neurology*, 36, 121-124 (10.1212/wnl.36.1.121)
 152. Gunal, D. I., N. Afsar, N. Bekiroglu and S. Aktan. (2000), New alternative agents in essential tremor therapy: double-blind placebo-controlled study of alprazolam and acetazolamide, *Neurol Sci*, 21, 315-317
 153. Zesiewicz, T. A., R. Elble, E. D. Louis, R. A. Hauser, K. L. Sullivan, R. B. Dewey, Jr., W. G. Ondo, G. S. Gronseth and W. J. Weiner. (2005), Practice parameter: therapies for essential tremor: report of the Quality Standards Subcommittee of the American Academy of Neurology, *Neurology*, 64, 2008-2020 (10.1212/01.wnl.0000163769.28552.cd)
 154. Striano, P., A. Coppola, G. Vacca, F. Zara, V. Brescia Morra, G. Orefice and S. Striano. (2006), Levetiracetam for cerebellar tremor in multiple sclerosis: an open-label pilot tolerability and efficacy study, *J Neurol*, 253, 762-766 (10.1007/s00415-006-0112-4)
 155. Bushara, K. O., T. Malik and R. E. Exconde. (2005), The effect of levetiracetam on essential tremor, *Neurology*, 64, 1078-1080 (10.1212/01.wnl.0000154596.21335.2e)
 156. Cao, Z., S. Hulsizer, F. Tassone, H. T. Tang, R. J. Hagerman, M. A. Rogawski, P. J. Hagerman and I. N. Pessah. (2012), Clustered burst firing in FMR1 premutation hippocampal neurons: amelioration with allopregnanolone, *Hum Mol Genet*, 21, 2923-2935 (10.1093/hmg/dds118)
 157. Cao, Z., S. Hulsizer, Y. Cui, D. L. Pretto, K. H. Kim, P. J. Hagerman, F. Tassone and I. N. Pessah. (2013), Enhanced asynchronous Ca(2+) oscillations associated with impaired glutamate transport in cortical astrocytes expressing Fmr1 gene premutation expansion, *J Biol Chem*, 288, 13831-13841 (10.1074/jbc.M112.441055)
 158. Wang, J. M., C. Singh, L. Liu, R. W. Irwin, S. Chen, E. J. Chung, R. F. Thompson and R. D. Brinton. (2010), Allopregnanolone reverses neurogenic and cognitive deficits in mouse model of Alzheimer's disease, *Proc Natl Acad Sci U S A*, 107, 6498-6503 (10.1073/pnas.1001422107)
 159. Chen, S., J. M. Wang, R. W. Irwin, J. Yao, L. Liu and R. D. Brinton. (2011), Allopregnanolone promotes regeneration and reduces beta-amyloid burden in a preclinical model of Alzheimer's disease, *PLoS One*, 6, e24293 (10.1371/journal.pone.0024293)
 160. Singh, C., L. Liu, J. M. Wang, R. W. Irwin, J. Yao, S. Chen, S. Henry, R. F. Thompson and R. D. Brinton. (2012), Allopregnanolone restores hippocampal-dependent learning and memory and neural progenitor

- survival in aging 3xTgAD and nonTg mice, *Neurobiol Aging*, 33, 1493-1506 (10.1016/j.neurobiolaging.2011.06.008)
161. Disney, M. D., B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney. (2012), A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome, *ACS Chem Biol*, 7, 1711-1718 (10.1021/cb300135h)
 162. Yang, W. Y., F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney. (2016), Small Molecule Recognition and Tools to Study Modulation of r(CGG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chem Biol*, 11, 2456-2465 (10.1021/acscchembio.6b00147)
 163. Yang, W. Y., H. D. Wilson, S. P. Velagapudi and M. D. Disney. (2015), Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA, *J Am Chem Soc*, 137, 5336-5345 (10.1021/ja507448y)
 164. Tran, T., J. L. Childs-Disney, B. Liu, L. Guan, S. Rzuczek and M. D. Disney. (2014), Targeting the r(CGG) Repeats That Cause FXTAS with Modularly Assembled Small Molecules and Oligonucleotides, *ACS Chemical Biology*, 9, 904-912 (10.1021/cb400875u)
 165. Verma, A. K., E. Khan, S. K. Mishra, N. Jain and A. Kumar. (2019), Piperine Modulates Protein Mediated Toxicity in Fragile X-Associated Tremor/Ataxia Syndrome through Interacting Expanded CGG Repeat (r(CGG)(exp)) RNA, (10.1021/acscemneuro.9b00282)
 166. Qurashi, A., H. Liu, L. Ray, D. L. Nelson, R. Duan and P. Jin. (2012), Chemical screen reveals small molecules suppressing fragile X premutation rCGG repeat-mediated neurodegeneration in *Drosophila*, *Hum Mol Genet*, 21, 2068-2075 (10.1093/hmg/dds024)
 167. Lin, Y., C. Tang, H. He and R. Duan. (2013), Activation of mTOR ameliorates fragile X premutation rCGG repeat-mediated neurodegeneration, *PLoS One*, 8, e62572 (10.1371/journal.pone.0062572)
 168. La Spada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding and K. H. Fischbeck. (1991), Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature*, 352, 77-79 (10.1038/352077a0)
 169. Lopez Castel, A., J. D. Cleary and C. E. Pearson. (2010), Repeat instability as the basis for human diseases and as a potential target for therapy, *Nat Rev Mol Cell Biol*, 11, 165-170 (10.1038/nrm2854)
 170. Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas *et al.* (2011), A hexanucleotide repeat expansion in

- C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD, *Neuron*, 72, 257-268 (10.1016/j.neuron.2011.09.010)
171. Liquori, C. L., K. Ricker, M. L. Moseley, J. F. Jacobsen, W. Kress, S. L. Naylor, J. W. Day and L. P. Ranum. (2001), Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9, *Science*, 293, 864-867 (10.1126/science.1062125)
 172. Matsuura, T., T. Yamagata, D. L. Burgess, A. Rasmussen, R. P. Grewal, K. Watase, M. Khajavi, A. E. McCall, C. F. Davis, L. Zu *et al.* (2000), Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10, *Nat Genet*, 26, 191-194 (10.1038/79911)
 173. Kobayashi, H., K. Abe, T. Matsuura, Y. Ikeda, T. Hitomi, Y. Akechi, T. Habu, W. Liu, H. Okuda and A. Koizumi. (2011), Expansion of intronic GGCCTG hexanucleotide repeat in NOP56 causes SCA36, a type of spinocerebellar ataxia accompanied by motor neuron involvement, *Am J Hum Genet*, 89, 121-130 (10.1016/j.ajhg.2011.05.015)
 174. Delot, E., L. M. King, M. D. Briggs, W. R. Wilcox and D. H. Cohn. (1999), Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene, *Hum Mol Genet*, 8, 123-128 (10.1093/hmg/8.1.123)
 175. Wells, R. D. (2009), Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders, *J Biol Chem*, 284, 8997-9009 (10.1074/jbc.X800010200)
 176. Pearson, C. E., K. Nichol Edamura and J. D. Cleary. (2005), Repeat instability: mechanisms of dynamic mutations, *Nat Rev Genet*, 6, 729-742 (10.1038/nrg1689)
 177. Groh, M., M. M. Lufino, R. Wade-Martins and N. Gromak. (2014), R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome, *PLoS Genet*, 10, e1004318 (10.1371/journal.pgen.1004318)
 178. Marelli, C. and F. Maschat. (2016), The P42 peptide and Peptide-based therapies for Huntington's disease, *Orphanet J Rare Dis*, 11, 24 (10.1186/s13023-016-0405-3)

Chapter 2

Material, Methods and Instrumentation

2.1 Material

All the DNA oligonucleotide used (Table 2.1) in all of our studies and used for RNA preparation were procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA). The Calf thymus (CT) DNA taken as control was also purchased from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA). The RNA samples used in our study was procured from Integrated DNA Technologies Private Limited.

The chemical reagents used in making buffers like KCl, NaCl, K_2HPO_4 , KH_2PO_4 , NaH_2PO_4 , Na_2HPO_4 , NaOH, $MgCl_2$, Ethylenediaminetetraacetic acid (EDTA), Tris, Boric acid, Dimethyl sulfoxide (DMSO), glycine, methanol, Sodium Dodecyl Sulfate (SDS) were of HPLC grade and procured from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA). For the PCR reagents and in-vitro transcription, dNTPs and rNTPs were procured from Sigma-Aldrich Chemicals Ltd. and Taq polymerase enzyme was obtained from the HiMedia Laboratories Pvt. Ltd. (India) while T7 RNA polymerase was purified in-house. The components required for gel preparation including Acrylamide, bis-acrylamide, Ammonium per sulfate (APS), Tetramethylethylenediamine (TEMED), Agarose, Ethidium bromide, Urea were bought from the HiMedia Laboratories Pvt. Ltd. (India). All the purification kits including plasmid, DNA, RNA and gel elution kits were purchased from Thermo Fisher Scientific, USA. The submerged horizontal electrophoresis unit and the Mini-PROTEAN Tetra cell unit used for gel electrophoresis and western blot were from Bio-Rad Laboratories, Inc. All the small molecules or ligands used in the studies including Piperine, Curcumin, TMPyP4 and Bisantrone etc. were purchased from Sigma-Aldrich Chemicals Ltd. (St. Louis, MO, USA) while the

NSC compounds were received from National Cancer Institute (NCI), USA on request. MilliQ water was obtained from Sartorius water purification system, Sartorius Corporate, Germany.

The growth media used for cell culture which includes Dulbecco's Modified Eagle's Medium (DMEM), Ham's F12 media, RPMI, etc as well as the Fetal Bovine serum (FBS), phosphate buffer saline (PBS), and antibiotic solution were purchased from GE Healthcare and Gibco Life Technologies, USA. The other chemical reagents used in the cell based assays were of molecular grade quality and were procured accordingly - 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) and Dimethyl sulfoxide (DMSO) from HiMedia Laboratories Pvt. Ltd. (India), Lipofectamine 3000 reagent from Thermo Fisher Scientific, . The RNA isolation kit and cDNA synthesis kit were purchased from Invitrogen and Bio-Rad respectively. The Human embryonic kidney cell lines (HEK293) and the Monkey kidney fibroblasts cell lines (COS-7) were obtained from the National Center for Cell Sciences (NCCS), Pune, India.

Ethics statement

The following cell lines GM00537 and GM04026 were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research. The Coriell Institute maintains the written consent forms and privacy of the donors of the samples, and the authors had no contact or interaction with the donors.

Table 2.1 DNA sequence template used for in-vitro transcription.

S. no.	Oligo Name	Oligo sequence 5' to 3'
1.	(5' <u>C</u> <u>A</u> G/3' <u>G</u> <u>G</u> C)x1	GGGAGAGGGTTTAAT <u>CAG</u> TACGAAA GTAC <u>GG</u> ATTGGATCCGCAAGG
2.	(5' <u>C</u> <u>C</u> G/3' <u>G</u> <u>A</u> C)x1	GGGAGAGGGTTTAAT <u>CCG</u> TACGAAA GTAC <u>CAG</u> ATTGGATCCGCAAGG
3.	(5' <u>C</u> <u>G</u> G/3' <u>G</u> <u>A</u> C)x1	GGGAGAGGGTTTAAT <u>CGG</u> TACGAAA GTAC <u>CAG</u> ATTGGATCCGCAAGG
4.	(5' <u>C</u> <u>C</u> G/3' <u>G</u> <u>U</u> C)x1	GGGAGAGGGTTTAAT <u>CCG</u> TACGAAA GTAC <u>CUG</u> ATTGGATCCGCAAGG
5.	(5' <u>C</u> <u>A</u> G/3' <u>G</u> <u>C</u> C)x1	GGGAGAGGGTTTAAT <u>CAG</u> TACGAAA GTAC <u>CG</u> ATTGGATCCGCAAGG
6.	(5' <u>C</u> <u>U</u> G/3' <u>G</u> <u>C</u> C)x1	GGGAGAGGGTTTAAT <u>CUG</u> TACGAAA GTAC <u>CG</u> ATTGGATCCGCAAGG
7.	(5' <u>C</u> <u>G</u> G/3' <u>G</u> <u>G</u> C)x1	GGGAGAGGGTTTAAT <u>CGG</u> TACGAAA GTAC <u>GG</u> ATTGGATCCGCAAGG
8.	(5' <u>C</u> <u>A</u> G/3' <u>G</u> <u>A</u> C)x1	GGGAGAGGGTTTAAT <u>CAG</u> TACGAAA GTAC <u>CAG</u> ATTGGATCCGCAAGG
9.	(5' <u>C</u> <u>C</u> G/3' <u>G</u> <u>C</u> C)x1	GGGAGAGGGTTTAAT <u>CCG</u> TACGAAA GTAC <u>CG</u> ATTGGATCCGCAAGG
10.	(5' <u>C</u> <u>U</u> G/3' <u>G</u> <u>U</u> C)x1	GGGAGAGGGTTTAAT <u>CUG</u> TACGAAA GTAC <u>CUG</u> ATTGGATCCGCAAGG

11.	(5' <u>CAG</u> /3' <u>GUC</u>)x1	GGGAGAGGGTTTAAT <u>CAG</u> TACGAAA GTAC <u>UGA</u> TTGGATCCGCAAGG
12.	(5' <u>CGG</u> /3' <u>GGC</u>)x2	GGGAGAGGGTTTAAT(<u>CGG</u>)x2TACGA AAGTA(<u>CGG</u>)x2ATTGGATCCGCAAGG
13.	(5' <u>CGG</u> /3' <u>GGC</u>)x3	GGGAGAGGGTTTAAT(<u>CGG</u>)x3TACGA AAGTA(<u>CGG</u>)x3ATTGGATCCGCAAGG
14.	(5' <u>CGG</u> /3' <u>GGC</u>)x4	GGGAGAGGGTTTAAT(<u>CGG</u>)x4TACGA AAGTA(<u>CGG</u>)x4ATTGGATCCGCAAGG
15.	(5' <u>CGG</u> /3' <u>GGC</u>)x6	GGGAGAGGGTTTAAT(<u>CGG</u>)x6TACGA AAGTA(<u>CGG</u>)x6ATTGGATCCGCAAGG
16.	(5' <u>CGG</u> /3' <u>GGC</u>)x2 0	GGGAGAGGGTTTAAT(<u>CGG</u>)x20TACG AAAGTA(<u>CGG</u>)x20ATTGGATCCGCAA GG
17.	(5' <u>CGG</u> /3' <u>GGC</u>)x4 0	GGGAGAGGGTTTAAT(<u>CGG</u>)x40TACG AAAGTA(<u>CGG</u>)x40ATTGGATCCGCAA GG
18.	(5' <u>CGG</u> /3' <u>GGC</u>)x4 0	GGGAGAGGGTTTAAT(<u>CGG</u>)x60TACG AAAGTA(<u>CGG</u>)x60ATTGGATCCGCAA GG
19.	(5' <u>CAG</u> /3' <u>GUC</u>)x6	GGGAGAGGGTTTAAT(<u>CAG</u>)x6TACGA AAGTA(<u>CUG</u>)x6ATTGGATCCGCAAGG
20.	NMR RNA	CCG <u>CGG</u> CGG

2.2 Sample preparation and methodology

2.2.1 Nuclear Magnetic Resonance (NMR) spectroscopy study

The desalted RNA sample containing 5'CGG/3'CGG RNA repeats motif was procured (sequence 5'r(CCGCGGCGG)₂ from Integrated DNA Technologies, Inc. The final concentration of the RNA duplex was calculated using its absorbance at 260 nm. Hyther Server was used to determine the molar extinction coefficients of RNA sample that is based on nearest neighbour thermodynamics [1,2]. Further RNA was lyophilised in the Christ Alpha 1-2 LD Freeze Dryer and dissolved in 1X phosphate buffer. All NMR experiments were performed on Bruker NMR spectrometer with the field strength of 400 MHz. Double-stranded RNA sequence containing a single CGG motif, 5'r(CCGCG- GCGG)₂ was used for NMR spectroscopy. Lyophilised RNA sample was prepared by resuspending lyophilised RNA in 10 mM phosphate buffer, pH 7.2, 0.1 M NaCl, and 50 mM EDTA in 10% D₂O. Variable temperature proton spectra were acquired from 25 °C to 65 °C. Spectral assignments were performed using NOESY (Nuclear Overhauser Effect Spectroscopy). All the spectra were processed by the program TopSpin3.4.

Two-dimensional (2D) NOESY was performed acquiring 4096 data points with 64 K transients for each of the 296 FID signals. The spectra were recorded at 288, 298 and 308 K with mixing times of 300, 200 and 100 ms and 128 number of scans. A NOESY spectrum was recorded using an excitation exculpating pulse sequence to suppress residual water signals. ¹H-¹H NOE distances were calculated using the SPARKY software.

For the NMR titration experiment of small molecules with different CGG repeats RNAs. The RNA samples were heated at 95°C for 10min followed by slow cooling at room temperature for proper annealing of the nucleic acid sequence. The RNA samples were prepared in 1x potassium phosphate buffer (10 mM phosphate buffer, pH 7.2, 0.1 M KCl, and 50 mM EDTA). The ligand solution

was also made in the same buffer and the RNA was successively added to the ligand during the NMR titration experiments. The NMR solution contained H₂O + D₂O in a 9:1 ratio while DSS (4, 4-dimethyl-4-silapentane-1-sulfonic acid) was used as a reference for the NMR samples. The obtained NMR spectra were consequently processed for the baseline and phase correction. The Topspin 3.5 was used to analyze and process the data obtained.

2.2.2 Preparation of model systems for Molecular Dynamics simulations

We have considered three sequences in the current study as shown in Figure 1. The structure of the RNA duplex, (CGG)₁ [(5'-CCGCGGCGG-3')₂] having a single mismatch was solved by NMR (PDB code: 2NCR) and employed for the MD simulation. The model system, (CGG)_{1m} [(5'-GCGCGGCGC-3')₂] having a single mismatch was prepared from (CGG)₁ using the software, chimera [3]. It should be noted that the mismatch base pair of above-mentioned sequence was in *anti-anti* state for both cases.

Further, (CGG)_{1m} was also modified by changing the orientation of one of the mismatch base, i.e. G5 from *anti* to *syn* conformation creating a *syn-anti* conformation of the same systems. Thus (CGG)_{1m} sequence represent two different systems, (CGG)_{1m} (*anti-anti*) and (CGG)_{1m} (*syn-anti*). Finally, atomic coordinates of (CGG)₃ or triple repeat of CGG [(5'-GGGCCGGCGGCGGGUCC-3')₂] were taken from the previously published X-ray crystallographic structure (PDB ID: 3SJ2) [4] where all the mismatch base pairs were in *anti-syn* conformations. All the systems were solvated in TIP3P [5] water molecules in a truncated octahedral box having 3624, 3853 and 8210 water molecules in (CGG)₁, (CGG)_{1m} and (CGG)₃, respectively. A suitable number of Na⁺ ions were added to neutralise the solvated systems. Molecular dynamics simulations for all the solvated systems were carried out using the AMBER16 program [6]. The AMBER force field [7] with revised χ [8] and α/γ [9] torsional parameters was used to simulate RNAs. The same force field was also used in previous studies [10]. The long-range electrostatic interactions were treated using the particle-

mesh Ewald (PME) scheme [11]. The non-bonded cut-off was set to 8 Å, and the non-bonded pair list was updated every 25 fs. All the bonds involving hydrogen atoms were constrained using the SHAKE algorithm [12]. The minimisation was done in two steps: firstly with RNA held fixed with a restraint force of 500 kcal/mol Å² and then without restraint. The systems were minimized by 1000 steps of steepest descent followed by another 1500 steps of conjugate gradient algorithm. After the minimisation, the temperature was raised from 0 to 300 K by using a Langevin thermostat where the RNA molecule was held fixed with a constraint force of 10 kcal/mol Å². Constant pressure dynamics with isotropic position scaling was turned on in the next step where reference pressure was 1 atm with a pressure relaxation time of 2 ps and RNA was held fixed with 2 kcal/mol Å². After that 100 ps equilibration was done without restraint with all the above conditions. A total of 1 μs MD run was performed at 300 K with 1 fs time-step for all cases under NPT ensembles. Atomic coordinates were saved every 5 ps resulting 200000 configurations for each simulation.

2.2.3 Umbrella Sampling Molecular Dynamics simulations

To determine the pathways for the *anti-syn* transformation, a two-dimensional potential of mean force (PMF) surface was constructed for the (CGG)₁ system from umbrella sampling molecular dynamics simulations. Since the *syn* and *anti*-base orientations were determined by the rotation of the glycosidic bond concerning sugar, the torsional angle χ was chosen as one of the reaction coordinates for the umbrella sampling simulation. This dihedral angle consists of O4, C1 atoms of sugar and N9, C8 atoms of base. The centre of mass pseudo dihedral angle (θ) was chosen as another reaction coordinate. Pseudo torsions are used to determine the base unstacking. Here we used four groups of atoms to define it. As for example, in our (CGG)₁ or (CGG)_{1m} cases four points are as follow; centre of mass of sugar atoms of 15, 4, 6 bases and the centre of mass of mismatch base atoms, i.e. G5 base. These reaction coordinates were used previously in different base flipping studies [10,13-15]. Both the coordinates were rotated with an increment of 10 degrees (°) creating a total of 1296 combinations.

All the combinations were minimised and equilibrated in two steps using a protocol previously used by Yildirim et al [10]. Throughout the simulation, χ and θ were held fixed in a particular combination using a square bottom well with a parabolic side. During the minimization and equilibration stages, a constraint force of 50 000 kcal/mol.rad² and 500 kcal/mol.rad² were employed, respectively. For all the simulations, the long-range cut off was always kept at 8 Å, and SHAKE [16] was used in all bonds involving hydrogen atoms. A 20 ps run was used to raise the temperature from 0 to 300 K using the Langevin dynamics with a collision frequency of 1 ps⁻¹. Constant pressure dynamics with isotropic positional scaling was turned on in the second stage of equilibration having a reference pressure of 1 atm with a pressure relaxation of 2 ps. All the production runs were conducted for 2 ns using a time-step of 2 fs. A potential with square bottom well and parabolic sides was used to restraint the reaction coordinates. For restraining χ and θ , force constants of 200 kcal/mol.rad² and 250 kcal/mol.rad² were used, respectively

2.2.4 RNA preparation using in-vitro transcription

The synthetic DNA templates for in-vitro transcription were either amplified by polymerase chain reaction (PCR) or cloned in a plasmid. Then they were used to prepare the RNA sequences by run-off transcription method using T7 RNA polymerase enzyme. The transcribed products were then purified by running in a 15% denaturing PAGE. The RNA band was visualized by UV shadowing and the RNA band was excised out and extracted with the help of 0.3M NaCl by tumbling down for overnight at 4 °C. The RNA was then ethanol precipitated for further purification.

2.2.5 Fluorescence titration assay

The excitation and emission wavelengths of each ligand were calculated using Synergy™ H1 multi-mode microplate reader, which were further used to perform the Fluorescence titration assays. The target RNA sequences along with the

control sequences which included RNA 5'CAG/3'GUC duplex pair, yeast t-RNA and different DNAs like CT-DNA, c-Myc DNA, Bcl2 DNA were heated in the 10mM phosphate buffer at 95°C for 10mins and then slowly cooled down to room temperature before each set of experiment. A constant concentration of ligand in 10mM phosphate buffer was added to each of the twelve wells of the black coming 96 well plate and the target RNA/DNA solution was titrated keeping the last well blank. The fluorescence change on addition of the RNA/DNA was measured at the particular excitation and emission of the particular ligand with the help of the microplate reader at 25 °C. The obtained data were then analyzed using Sigma Plot 12.0 software (Systat Software, Chicago, USA).

2.2.6 Isothermal Calorimetric titration assay (ITC)

The working solution of the RNA were prepared by dissolving it in the 10mM potassium phosphate buffer and then properly annealed by heating it at 95°C for 10mins followed by slow cooling at room temperature. The RNA sample was loaded into the cell while the working stock of the ligand was prepared in the same buffer and loaded to the syringe of the ITC instrument. Both the RNA as well as the ligand solutions was degassed before loading into the instrument to avoid any bubble. The ITC experiments were executed using the MicroCal™ isothermal titration calorimeter iTC200 instrument (GE healthcare) at a constant temperature of 298K where the ligand solution was injected to the RNA solution containing cell for a particular number of injections. A constant stirring speed of 750 rpm was maintained throughout the experiment. The binding data obtained due to ligand-RNA interaction was plotted using the MicroCal Origin software. The binding thermogram provided the dissociation constant (Kd) and other thermodynamic parameters obtained due to the ligand-RNA interaction.

2.2.7 Circular Dichroism spectroscopy assay

All the circular dichroism (CD) experiments were performed with the help of J-815 Spectropolarimeter (JASCO) instrument. The CD spectra experiments were

carried out at the constant temperature of 25°C which was maintained using the peltier junction temperature controller. A constant supply of nitrogen was flushed into the cuvette containing chamber to avoid the condensation of water outside the cuvette. The spectrum for each experiment was measured using the quartz cuvette of 0.2cm path length and a scanning speed of 20nm/min between the wavelength of 200nm to 320nm. The RNA samples were heated at 95°C for 10mins and then slowly cooled down to the room temperature before each set of experiment in 10mM phosphate buffer (50mM KCl, pH 7.0). In the CD melting studies, the data was collected along the temperature range of 25°C to 95°C at a particular wavelength. Buffer subtraction was done in each set of experiment to avoid signal contribution from the buffer. The data were plotted using the SigmaPlot 12.0 software both for the CD spectra and CD melting experiment.

2.2.8 Thermal denaturation study by UV spectroscopy

The samples were annealed in 10mM phosphate buffer before each set of experiment. The change in absorbance at 260 nm was monitored using the UV-Vis spectrophotometer along the temperature range of 25°C to 95°C both in presence and absence of ligand till drug/nucleic acid ratio of 2. The temperature increased at the rate 1°C/min and maintained with the help of the attached Peltier system. The normalized absorbance changes against temperature were plotted using the SigmaPlot 12.0 software.

2.2.9 Gel retardation assay

For each set of experiment, the 20 µM of each RNA samples were heated at 95°C in 10mM potassium phosphate buffer and cooled slowly to room temperature. Then the ligand was added at the optimum concentration and then serially diluted to rest of the tubes keeping the last well as blank. The ligand was incubated for 15 mins and then the 6X orange dye was mixed to the samples for loading into the gel. All samples were analyzed on a 3% agarose gel and run in 1X Tris-Borate-

EDTA (TBE) buffer. The gels were pre-stained with ethidium bromide and visualized using analyzed on ImageQuant LAS 4000 (GE Healthcare).

2.2.10 PCR stop assay

The template and the complementary sequence was added to PCR mastermix containing 1X PCR buffer, 4.25mM MgCl₂, 0.33 mM dNTPs and 2.5 units Taq DNA polymerase. The 25µl of mastermix was distributed in each of the six PCR tubes for each set of experiment and the ligand was added in the optimum concentration in the first tube and serially diluted in the other tubes. The last well was kept blank serving as control as no ligand was added to it. The reaction mixtures were then incubated in thermo cycler (Eppendorf) with following conditions: initial denaturation of 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, extension at 72 °C for 1 minute and finally held at 4 °C following completion. The PCR products were then mixed with 6X orange loading dye and then resolved on 3% agarose gel pre-stained with ethidium bromide. The gels were at last visualized using analyzed on ImageQuant LAS 4000 (GE Healthcare).

2.2.11 Real time PCR

In order to determine whether Piperine improve pre-mRNA alternative splicing defect *in-vitro*, a FXTAS cellular model were used. Briefly, HEK293 and/or COS7 cells were grown in 24 well plates as monolayer in the growth medium contain 1X DMEM, 10% fetal bovine serum, 1% antibiotic and antimycotic at 37 °C with 5% CO₂. After cell reached 80-90% confluency, HEK and/or COS7 cell were transfected with the equal amount of plasmid contain CGGx99 repeat and targeted mini-genes (*SMN2* and *Bcl-x*) using Lipofectamine 3000 reagent (Thermo Fisher Scientific), according to standard manufacturers protocol. Transfection cocktail was removed after 4-5 hours and drug containing growth medium were added. After 24 hours, cells were lysed in the plate and total RNA

were recovered using RNA isolation kit (Invitrogen), according to standard manufacturers protocol.

All different conditioned RNA sample were subjected to reverse transcribe (RT-PCR) using cDNA synthesis kit from Bio-Rad, followed standard protocol. Out of 500 ng reverse transcribed mRNA 100 ng were subjected for semi-quantitative PCR. The PCR protocol used: denaturation for 95 °C for 1 min, annealing 55 °C for 1 min, extension at 72 °C for 2 min and final extension at 72 for 10 min. The PCR protocol was run for 25-30 cycles. The PCR product were analysed on agarose gel electrophoresis, stained with ethidium bromide and image were recorded using ImageQuant LAS 4000 (GE Healthcare). The splicing isoform intensity was quantified using imageJ software. The primer sequence mentioned in the table S4 for each construct[17].

Two control condition were used to study pre-mRNA splicing defect. 1) Targeted mini-genes (*SMN2* and *Bcl-x*) were co-transfected with plasmid lacking CGG repeat at 5'UTR as describe above. 2) Co-transfection of CGG repeat plasmid with mini-genes whose pre-mRNA splicing not govern by sam68 (cTNT and IR) [18].

2.2.12 Western blot assay and protein aggregate quantification

HEK/COS7 cells were seeded in 6 well plate in a monolayer form and transfected with 2.5µg of plasmid coding CGGx99-GFP using lipofectamine 3000 as per manufacturer protocol from Thermo Fisher Scientific. The transfection cocktail was removed after 5 hrs incubation and added fresh compound containing media and the cell were incubated for 24hrs at 37 °C. After incubation, cells were lysed using RIPA buffer 200 µL/well containing 1 µL of Halt Protease Inhibitor Cocktail from Sigma. Cellular proteins concentration were calculated using Bradford assay. Same amount of protein were loaded in SDS-PAGE for separation and transferred to a PVDF membrane. Blot were incubated with antiFMR1-polyG antibody from Merck Millipore as a primary antibodies. After primary incubation

anti-IgG-horseradish peroxidase conjugate used as the secondary antibody. A chemiluminescent signal was detected using Luminata Crescendo Western HRP substrate from Merck Millipore in ImageQuant LAS 4000 (GE Healthcare).

In order to determine the effect of compounds on polyG-GFP protein aggregates two different constructs were used that contain r(CGG)_{X99}-GFP and pcDNA-GFP. The above mentioned construct was designed and validated by the Nicolas group earlier [17]. Briefly, HEK/COS7 cells were seeded in six well plates in 10% DMEM media. After cells were reached 80-90 percent confluency transfection was done with above mentioned plasmid using lipofectamine 3000 reagent from Thermo Fisher Scientific as per manufacturer standard protocol. The transfection cocktail was removed after 4hrs incubation and fresh media were applied which contain respective compound concentration and the cells were incubated for 18-24 hrs at 37°C. Cells after washing with PBS, fixed with 4% paraformaldehyde for 15 minutes. Images were captured and protein aggregates visualised at higher magnification using fluorescence confocal microscopy and processed to remove background signals. 100 positive transfected cells from each well were selected and manually graded as “with protein aggregates” and “no aggregates”. The number of GFP protein inclusions were manually counted from three independent experiments and standard deviation was calculated.

2.3 Principle behind the assay and the related Instrumentation

2.3.1 In-vitro transcription for RNA synthesis

RNA sequences can be synthesized from its corresponding DNA template using the In-vitro transcription method (Figure 2.1). The template-directed synthesis of the RNA can be conveniently achieved by using the RNA polymerase enzyme like T3, T7, and SP6 RNA polymerase enzymes in presence of rNTPs. Here, T7 RNA polymerase enzyme has been used for RNA synthesis. This 100 kDa enzyme is highly efficient in producing transcripts of length ranging from smaller than 30nt to more than 10⁴nt and to the amount stretching from µg to mg [19]. Both the plasmid DNA bearing the T7 promoter upstream of the desired sequence

or the PCR product having the T7 promoter sequence at the 5' end of the oligonucleotide can serve as the DNA template for in-vitro transcription process. The promoter of T7 polymerase is of 23bp length and covers the position -17 to +6 of the sequence while +1 position being the first transcribed nucleotide followed by the sequence of interest. The T7 polymerase enzyme supports the run-off transcription thus no additional feature is required for the stopping of the transcription process. One can purify the transcribed RNA from its DNA template by simply treating it with DNaseI enzyme [20].

2.3.2 Fluorescence binding assay

The fluorescence binding assay is based on the fluorescence intensity change on interaction of the ligand with the DNA/RNA molecules (Figure 2.2). Fluorescence as such is defined as the emission of light by a substance as an electron jumps from a higher energy state to lower energy state that has been previously excited to higher energy state due to absorbance of the high energy electromagnetic radiation.[21] Ideally it takes about 10^{-8} seconds for an electron to get excited and again revert back to its ground energy state and this phenomenon is also governed by its chemical environment. Moreover the fluorescence intensity is directly proportional to the number of fluorophores present in the excited state.[21] When a fluorophore binds to biological macromolecule, it leads to conformational changes which may either result in high fluorescence intensity or complete quenching of the fluorescence [22-25]. This principle has been exploited to detect change in the fluorescence intensities of the small molecules used in the study upon interaction with the nucleic acid sequences. The DNA or the RNA is serially diluted along the eleven wells of the 96 well black plate while the last well is kept blank where no DNA/RNA was added. On interaction with nucleic acid molecule, the fluorescence intensity either gradually increases or decreases due to quenching, thus giving a pattern for the interaction of the nucleic acid sequence and ligand. The data obtained from the Synergy H1 multi-mode microplate reader was further analyzed using SigmaPlot 12.0 software which fitted the data into the following binding equation.

$$df = \frac{B_{max1} * abs(x)}{Kd1 * abs(x)} + \frac{B_{max1} * abs(x)}{kd2 * abs(x)} \quad (1)$$

Where B_{max} = maximum number of binding sites.

K_d = equilibrium dissociation constant

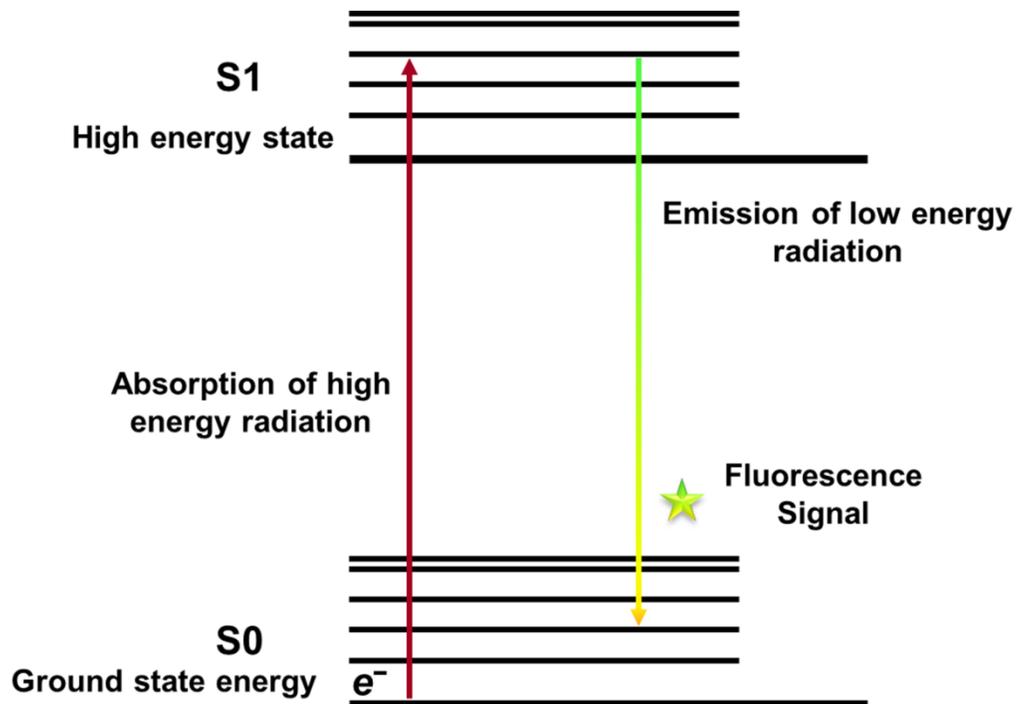


Figure 2.2 The Jablonski diagram depicting the fluorescence phenomenon.

2.3.3 Isothermal titration calorimetric assay (ITC)

Isothermal titration calorimetric (ITC) is one of the essential tools in ligand interaction studies especially in the field of drug discovery (Figure 2.3). This instrument is based on the measurement of the heat released or absorbed during the interaction between the ligand and the biomolecule.[26] It provides the advantage of direct and label free measurement of the thermodynamic profile of binding interactions between two molecules[25-27]. The ITC plot gives a precise value for the binding constant between the drugs or ligands and their target molecule may DNA, RNA or the protein along with the other thermodynamic parameters including association constant (K_a), reaction stoichiometry (n), enthalpy (ΔH), and entropy (ΔS).[28-31] In principle, this assay follows the below equation to calculate the thermodynamic parameters depending on the heat absorbed or released during the interaction event:

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S \quad (2)$$

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

Depending on the data obtained, one can decide how strongly a ligand binds to the macromolecule and thereby choose the lead molecule.

The ITC instrument consists of two cells and a stirring syringe when one of the cell acts as reference cell containing water while the other cell contains the sample, in our the DNA or the RNA sample in the appropriate buffer. The ligand in the same buffer is loaded into the stirring syringe. The syringe is inserted into the cell where the ligand is gradually injected into the sample. Each injection results in release of the ligand into the DNA/RNA solution resulting in the heat change which is detected by the instruments. The heat change measured is integrated with respect to time and a titration curve of kcal/mol vs molar ratio is generated using the MicroCal Origin software. This provides the user with an isotherm which can be fitted into different binding models which gives the affinity value (KD), stoichiometry value (n) and value for enthalpy change (ΔH).

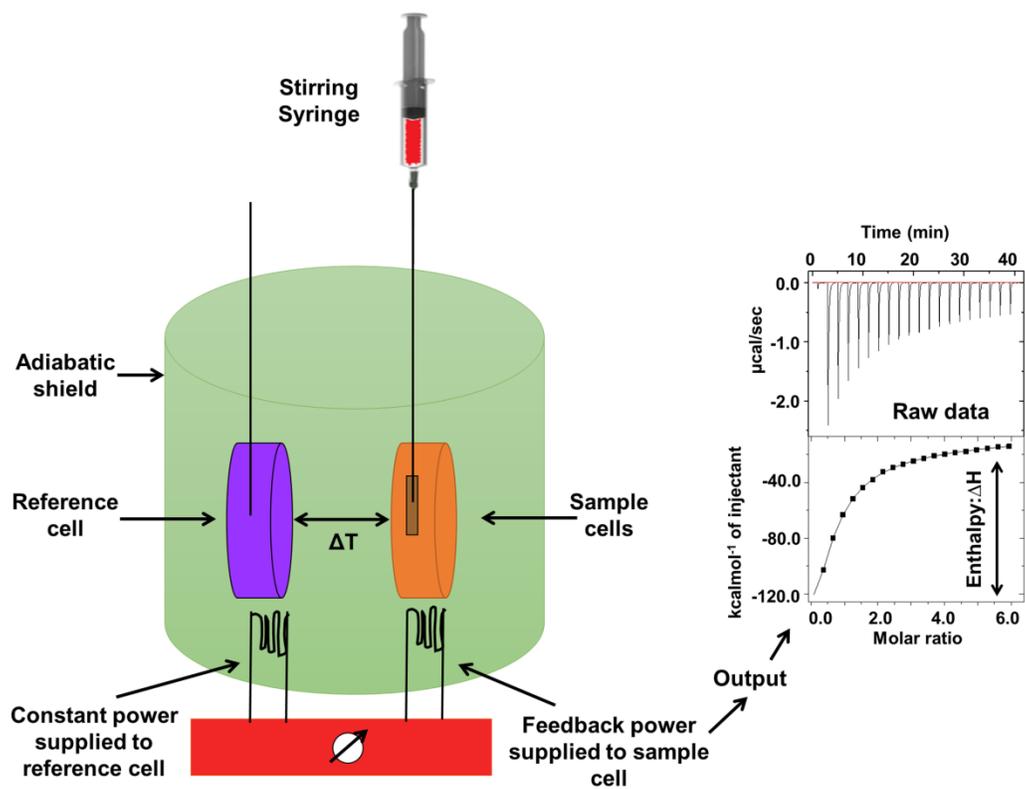


Figure 2.3 Schematic illustration of Isothermal calorimetry instrumentation (left). Showing the example of the binding isotherm (Right).

2.3.4 Circular Dichroism spectroscopy

Circular Dichroism spectroscopy plays a significant role in the structural studies as gives a preliminary idea regarding the conformational changes occurring during the binding event between the macromolecules and the ligand (Figure 2.4). Although it is unable to provide detailed structural like Nuclear Magnetic Resonance (NMR) and X-ray crystallography, but it provides the advantage of measuring the structural changes in physiological conditions in very less concentration. The optically active samples are capable to divert the circularly polarized light either to the right or to left. CD instrument measures the difference in absorption between the left and right circularly polarized light.[32-35] The conformational attributes provided by the sugar moiety and the nitrogenous bases in the nucleic acid sequences ensue the difference in the absorption of circularly polarized light that can denoted as following: [34]

$$CD = A^l - A^r \quad (3)$$

Where A^l and A^r stand for the absorptions of left and right circularly polarized light respectively by the sample.

Every biological macromolecule gives a characteristic CD pattern depending on its conformation.[36-39]. Typical single standard CGG RNA and duplex RNA CD signature spectrum are shown in the chapters (4, 5 and 6). The fundamental instrumentation of the CD techniques are illustrated in the figure XXX..

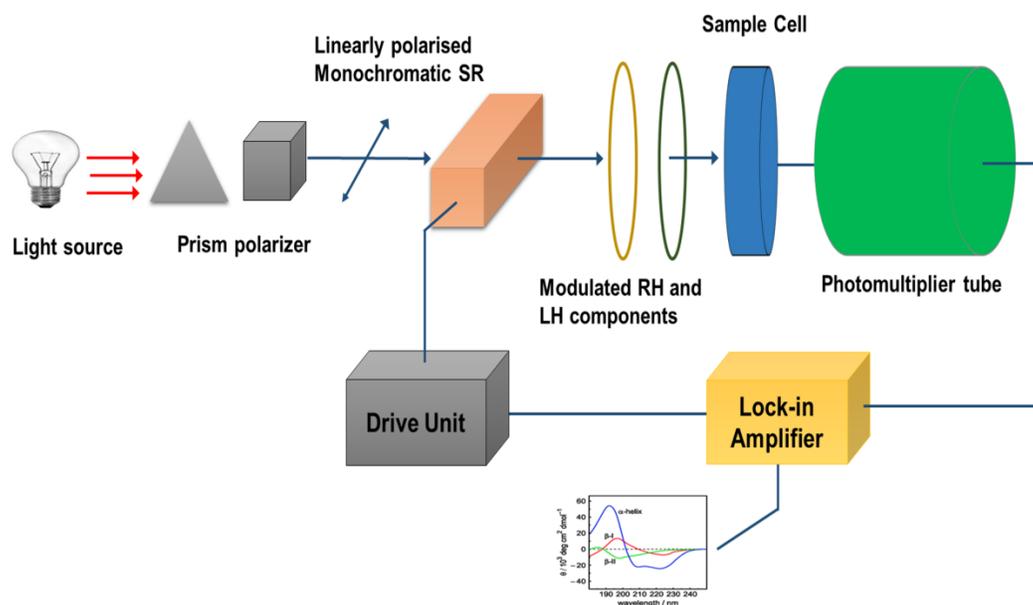


Figure 2.4 Schematic representation of the Circular Dichroism spectroscopy instrumentation.

2.3.5 Nuclear Magnetic Resonance (NMR) spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a very robust and strong technique which helps to detect the structure of a molecule based on the resonance of the sub-atomic particles of its atomic nucleus (Figure 2.5) [40-42]. Two molecules differ in their structure though their composition may be same. So when a magnetic field is applied, each atom behaves differently depending on the structure of the molecule having it i.e. because of the magnetic shielding effect of the neighboring sub-atomic particle including the nuclei and the protons. NMR exploits this property of the atoms to determine the molecular structure of a molecule. The nucleus of an atom like nucleus of an atom like hydrogen (^1H), carbon (^{13}C) possesses an intrinsic magnetic moment (μ) which when placed in a magnetic field results in the spinning of the atom.[43] When an external magnetic field is applied, the spinning nucleus starts precession about the axis of the external magnetic field. Therefore the positively charged nucleus generates an intrinsic spin angular momentum denoted by I (Figure 2.6).

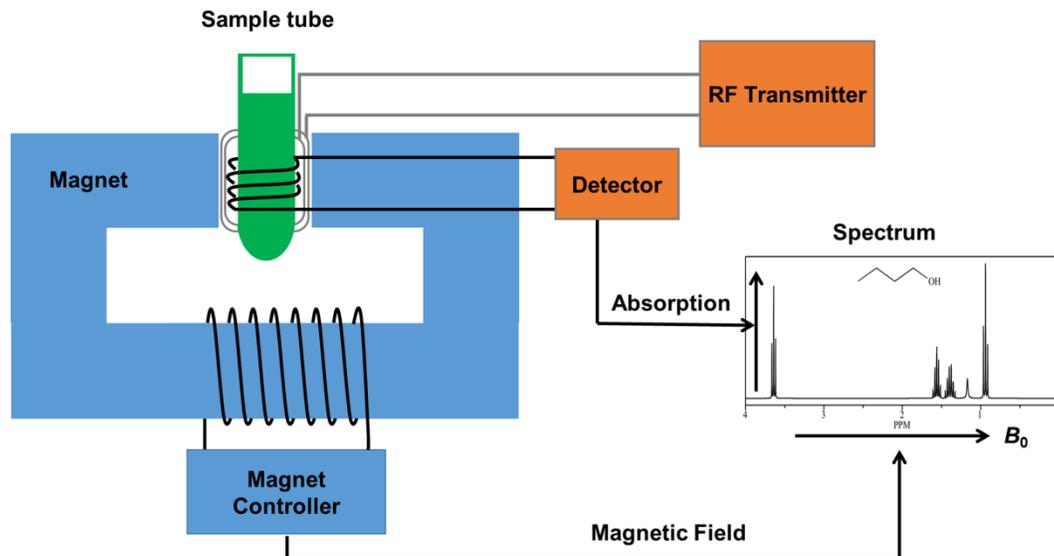


Figure 2.5 Schematic representation of Nuclear Magnetic Resonance spectroscopy.

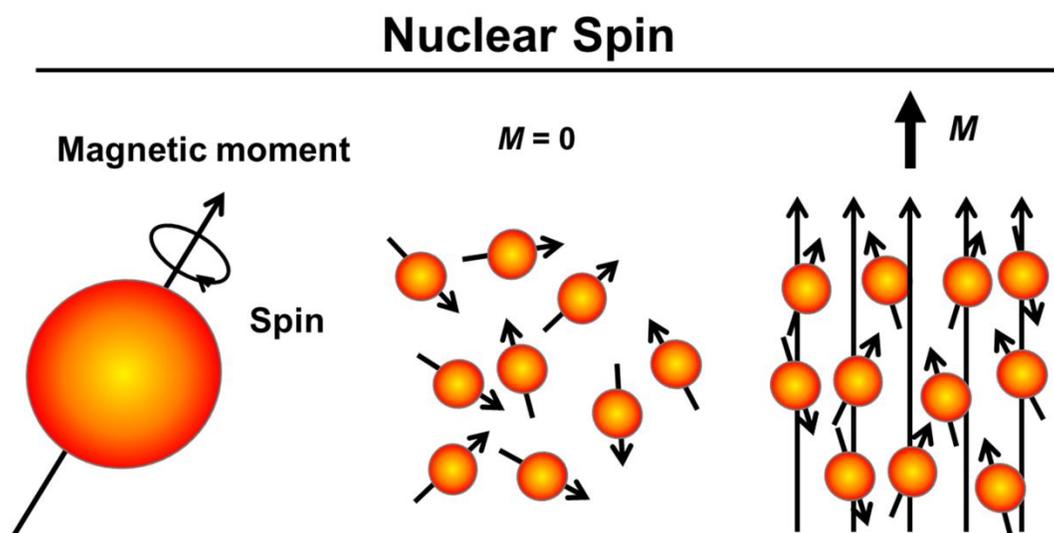


Figure 2.6 Precessional motion of a spin in absence and presence of external magnetic field (M).

2.3.6 Gel retardation assay

The gel retardation assay is based on the shift in movement of the bands due to interaction of the DNA/RNA sequences with the ligand molecule (Figure 2.7). The ligand molecule intercalates or stacks within the DNA/RNA grooves and thereby increases the molecular weight of the sequences. As the DNA/RNA molecules are resolved through the gel electrophoresis, the bands migrate in the gel according to their molecular weight. The bands with higher molecular weight move slower as compared to the ones having comparatively low molecular weight. Thus, when the ligand interacts with the nucleic acid molecules, they form a strong association thereby increasing the molecular weight of the ligand-nucleic acid complex. This leads to the retarded movement of the band interacting with ligand as compared to that which does not have any ligand interaction.

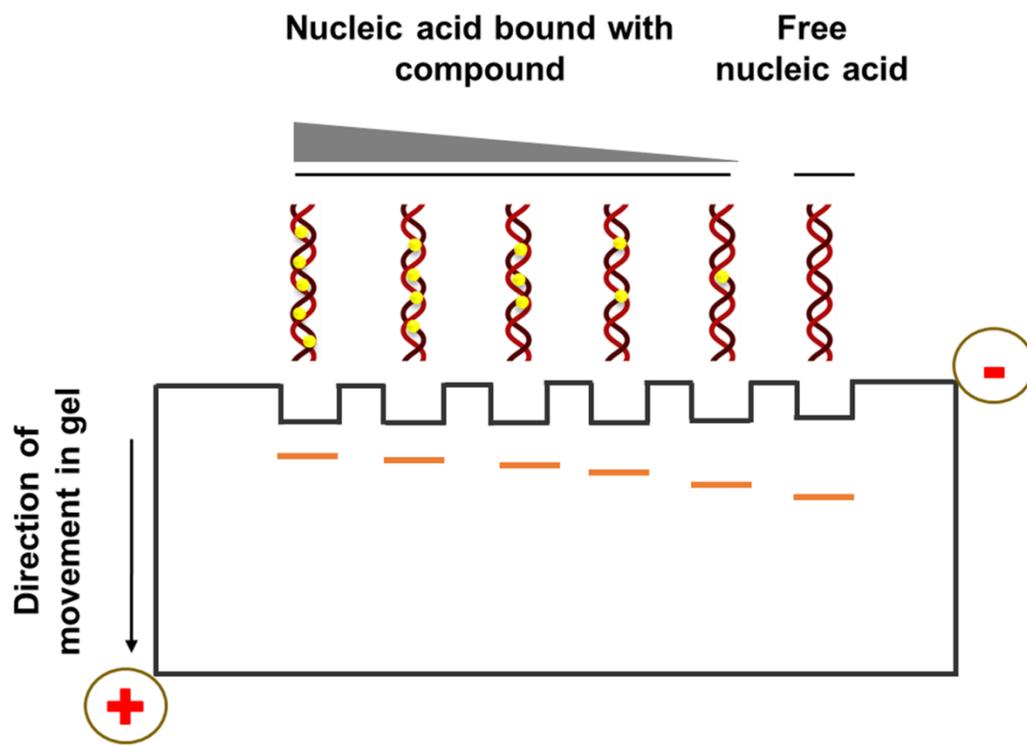


Figure 2.7 Diagrammatic illustration of gel retardation working mechanism.

2.3.7 Polymerase Chain Reaction (PCR) stop assay

Polymerase Chain Reaction (PCR) is a standard molecular method through which a DNA template sequence is amplified to greater number of copies using the Taq polymerase enzyme (Figure 2.8) [44]. While the PCR stop assay is based on the situational stalling of the replication machinery which inhibits the further amplification of the DNA molecule [45]. The DNA replication process occurs in a semi-conservative method, the two strands open up to form a replication fork while the DNA polymerase enzyme goes on adding the dNTPs complementary to the template strand leading to the formation of a new strand in the 5' to 3' direction [46]. But formation of any stable secondary structures like hair-pin loops, G-quadruplex structure, etc. in the template strand stalls the movement of the replication machinery eventually leading to the inhibition of the replication process. On the search of finding novel small molecules modulating the stability of the secondary structure, this assay pose to be a strong evidence whether this structures effect the replication process or not. Thus, the small molecules involved in stabilization of these DNA secondary structures will inhibit the DNA polymerase movement thus terminating the replication process, which results in less number of amplified products during the PCR reaction.

We see the same phenomena with our ligand of interest, as we find a decrease in the intensity of the DNA band with the increase in concentration of the ligand due to stabilization of the CGG repeats forming secondary structure.

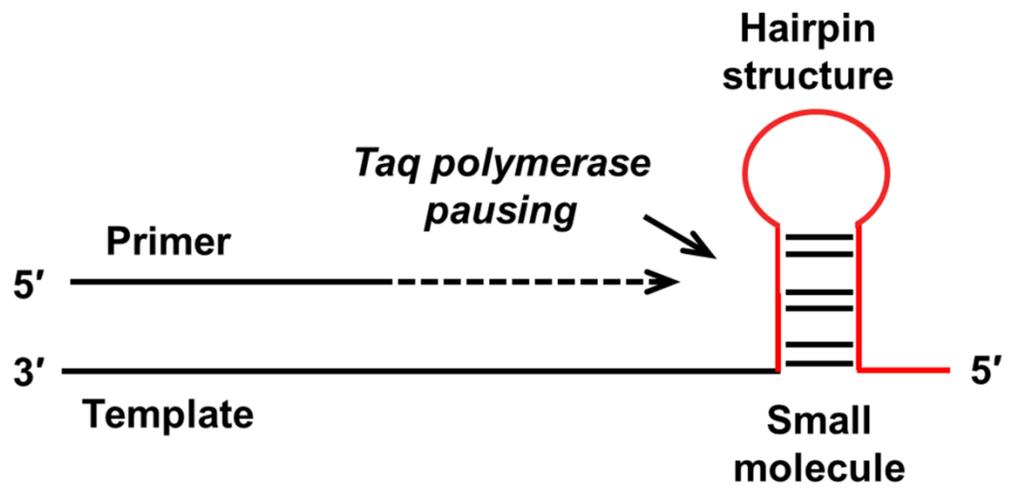


Figure 2.8 Diagram showing principle of PCR stop assay.

2.3.8 Molecular Docking

Molecular docking is a valuable method which involves the prediction of ligand orientation and conformation within a target site. Molecular docking is a multi-step complex process with two important aims: 1) Correct prediction of activity and 2) Accurate structural modelling [47]. Computational tools such as Autodock, Discovery studio and Schrodinger, etc. have emerged as the worthwhile tools, for drug discovery purpose [48]. Out of these tools, Autodock is a non-commercial docking tool, which uses stochastic Lamarckian genetic algorithm to deduce the ligand conformations and different scoring functions. These two values help to elucidate the thermodynamic stability of ligand-macromolecule complex [49,50]. Rizvi et al., have discussed the detailed description of protocol to perform the docking of ligand and macromolecules [51].

2.3.9 Cell viability assay (MTT assay)

Cell viability assay is widely used to check the toxicity of compounds on different cell lines as this helps us to decide whether a molecule is safe to be used or not. They use the metabolic or enzymatic activities as a marker of the viable cells and come with different variations like dye reduction by cellular enzyme, Protease Viability Marker Assay, Luminescence ATP assay, etc. [52,53] The most commonly used is the MTT assay. This assay quantifies the amount of the pale yellow colored substrate, MTT (3-(4, 5-Dimethylthiazol-2-yl)-2, 5-Diphenyltetrazolium Bromide) is being converted to its purple colored product, formazan by the mitochondrial succinate dehydrogenase enzyme of the viable cells. The formazan is an insoluble compound which is solubilized in DMSO and then the fluorescence is observed using a plate reader. [54] This assay is based on the principle that the amount of signal generated by the formation of the product is proportional to the number of live cells present. When the cells die, the metabolic activity of the cells reduces and therefore the rate of conversion of the reporter substrate to its colored product decreases (Figure 2.9).

All the small molecules used in our study have been initially screened through the MTT assay to check their effect on the cell viability for various cell lines including normal and patient derived fibroblast cell lines. The cells were treated for 24 hours with the small molecule and then the MTT solution was added to each well of the 96-wll plate. After four hours of incubation with the MTT, DMSO was added to each well for solubilization of the formazon formed and then the absorbance was measured at 540nm using the Synergy™ H1 multi-mode microplate reader.

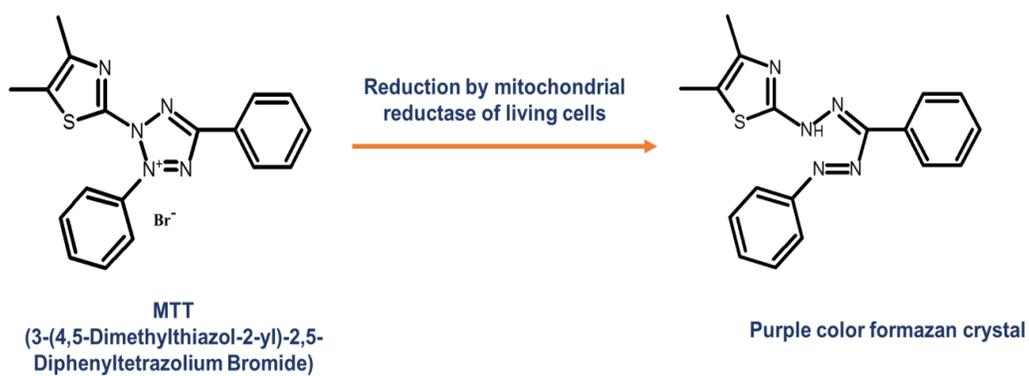


Figure 2.9 Reduction of MTT into insoluble purple color crystals by mitochondrial reductase.

2.3.10 Western blot assay

Western Blotting is an essential technique to identify specific protein from pool of all cellular proteins. This technique is achieved through three steps, first being the separation of all the proteins by SDS-PAGE, next is the transfer of the proteins to the solid support like PVDF membrane or nitrocellulose membrane and the last step includes the labeling of the target protein using the particular primary and subsequent secondary antibody which helps in the visualization (Figure 2.10). The specific antibodies bind to the protein of interest and thereby showing a specific band. The thickness of the bands is directly proportional to the amount of protein present in the sample.[55]

The cells were plated on 6-well tissue culture plate and once they achieve more than 70% confluency, they were transiently transfected with the plasmids of interest. The transfected cells were then treated with desired concentration of ligand for 24hours. After treatment, the cells were lysed using RIPA lysis buffer (Thermo scientific-89900) and cell debris was pelleted down at 4 °C at 14000 g for 15 min. The concentrations of the protein samples were determined using Bradford method (Bio-Rad)[56] where BSA was used as a standard to plot the standard curve. The proteins were then resolved in 10% SDS-PAGE according to standard protocol.[57] The samples were heated at 95°C for 5mins after mixing with 6X loading dye (Laemmli buffer). Pre-stained protein marker (Himedia-MBT092) was used as a reference. The running buffer (0.3% Tris base, 1.44% glycine and 0.1% SDS) was filled into the tank of the Mini PROTEAN III Cell electrophoresis unit (Bio-Rad) and run at 70 V until the bromophenol blue dye front reaches the bottom of the gel. The proteins were then transferred to PVDF membrane (Millipore) using the Mini Trans-Blot®Electrophoretic Transfer Cell unit. The transfer buffer (25 mM Tris; 193 mM glycine; 20 % methanol) was used for the process and it was performed at 250 V for 70-80 min at 4°C. To check the efficiency of transfer of the proteins, the blotting membrane was stained with Ponceau S (0.1% (w/v) in 5% acetic acid). After destaining the membrane by washing with H₂O, the blotting membrane was blocked in 3 % blocking buffer

(3% bovine serum albumin (BSA) in TBST) at room temperature for 1 hour on a shaking platform followed by incubation with primary antibody (primary antibody is diluted in 5% milk PBS-T before use per manufacturer's instructions) for overnight at 4°C. The membrane was washed with TBS-T 3-5 times at RT and then incubated with the appropriate HRP-conjugated secondary antibody (Diluted as per manufacturer's instructions) for 2 hours at RT. The specific protein bands were then detected using chemiluminescence reagents (Serva pvt. Ltd. -catalog no. 42582.01)[58] and the images were acquired and analyzed using ImageQuant LAS 4000 (GE Healthcare, Biosciences Ltd., Sweden).

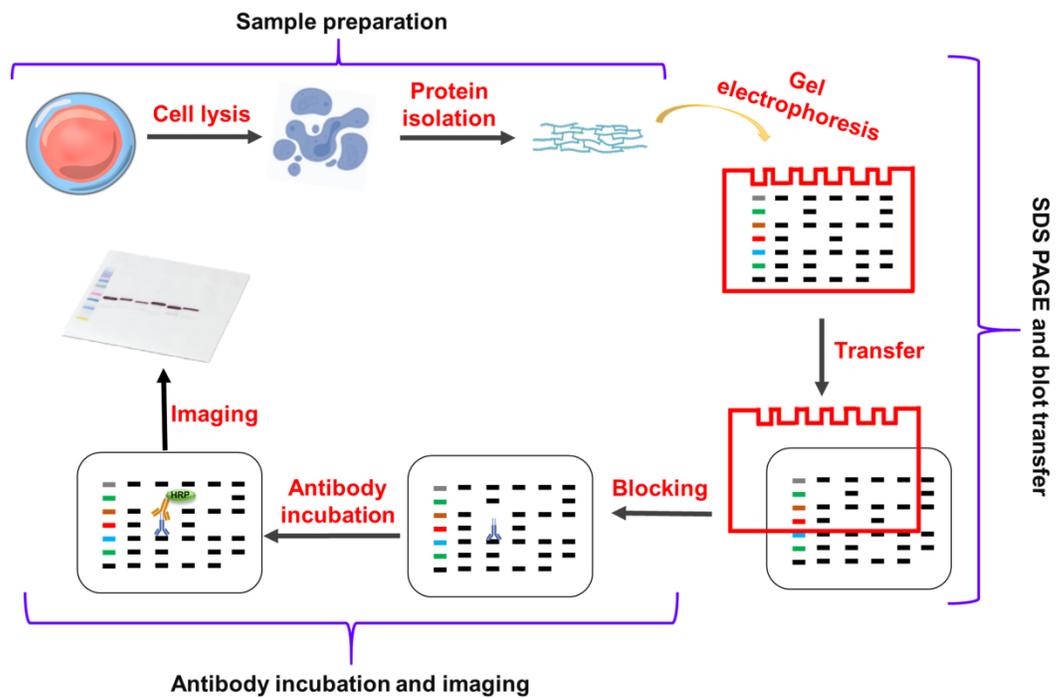


Figure 2.10 Schematic illustration of western blot assay and sample preparation.

2.4 Reference

1. Peyret, N., P. A. Seneviratne, H. T. Allawi and J. SantaLucia. (1999), Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches, *Biochemistry*, 38, 3468-3477
2. SantaLucia, J. (1998), A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proceedings of the National Academy of Sciences*, 95, 1460-1465
3. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin. (2004), UCSF Chimera—a visualization system for exploratory research and analysis, *Journal of computational chemistry*, 25, 1605-1612
4. Kumar, A., P. Fang, H. Park, M. Guo, K. W. Nettles and M. D. Disney. (2011), A crystal structure of a model of the repeating r (CGG) transcript found in fragile X syndrome, *ChemBioChem*, 12, 2140-2142
5. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein. (1983), Comparison of simple potential functions for simulating liquid water, *The Journal of chemical physics*, 79, 926-935
6. Case, D. A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods. (2005), The Amber biomolecular simulation programs, *Journal of computational chemistry*, 26, 1668-1688
7. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman. (1995), A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of the American Chemical Society*, 117, 5179-5197
8. Yildirim, I., H. A. Stern, S. D. Kennedy, J. D. Tubbs and D. H. Turner. (2010), Reparameterization of RNA χ torsion parameters for the AMBER

- force field and comparison to NMR spectra for cytidine and uridine, *Journal of chemical theory and computation*, 6, 1520-1531
9. Pérez, A., I. Marchán, D. Svozil, J. Spöner, T. E. Cheatham III, C. A. Laughton and M. Orozco. (2007), Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers, *Biophysical journal*, 92, 3817-3829
 10. Yildirim, I., H. Park, M. D. Disney and G. C. Schatz. (2013), A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations, *Journal of the American Chemical Society*, 135, 3528-3538
 11. Darden, T., D. York and L. Pedersen. (1993), Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems, *The Journal of chemical physics*, 98, 10089-10092
 12. Ryckaert, J.-P., G. Ciccotti and H. J. Berendsen. (1977), Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics*, 23, 327-341
 13. Banavali, N. K. and A. D. MacKerell Jr. (2002), Free energy and structural pathways of base flipping in a DNA GCGC containing sequence, *Journal of molecular biology*, 319, 141-160
 14. Song, K., A. J. Campbell, C. Bergonzo, C. de los Santos, A. P. Grollman and C. Simmerling. (2009), An improved reaction coordinate for nucleic acid base flipping studies, *Journal of chemical theory computation*, 5, 3105-3113
 15. Priyakumar, U. D. and A. D. MacKerell. (2006), Computational approaches for investigating base flipping in oligonucleotides, *Chemical reviews*, 106, 489-505
 16. Ryckaert, J.-P., G. Ciccotti and H. J. J. J. o. C. P. Berendsen. (1977), Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, 23, 327-341

17. Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott *et al.* (2010), Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *Embo j*, 29, 1248-1261 (10.1038/emboj.2010.21)
18. Disney, M. D., B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney. (2012), A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome, *ACS Chem Biol*, 7, 1711-1718 (10.1021/cb300135h)
19. Beckert, B. and B. Masquida. (2011), Synthesis of RNA by in vitro transcription, *Methods Mol Biol*, 703, 29-41 (10.1007/978-1-59745-248-9_3)
20. Arnaud-Barbe, N., V. Cheynet-Sauvion, G. Oriol, B. Mandrand and F. Mallet. (1998), Transcription of RNA templates by T7 RNA polymerase, *Nucleic Acids Research*, 26, 3550-3554 (10.1093/nar/26.15.3550)
21. Wolf, D. E. (2013), Fundamentals of fluorescence and fluorescence microscopy, *Methods Cell Biol*, 114, 69-97 (10.1016/b978-0-12-407761-4.00004-x)
22. Joo, C., H. Balci, Y. Ishitsuka, C. Buranachai and T. Ha. (2008), Advances in single-molecule fluorescence methods for molecular biology, *Annu Rev Biochem*, 77, 51-76 (10.1146/annurev.biochem.77.070606.101543)
23. Mathies, R. A., K. Peck and L. Stryer. (1990), Optimization of high-sensitivity fluorescence detection, *Analytical Chemistry*, 62, 1786-1791 (10.1021/ac00216a012)
24. Olmsted, J. and D. R. Kearns. (1977), Mechanism of ethidium bromide fluorescence enhancement on binding to nucleic acids, *Biochemistry*, 16, 3647-3654 (10.1021/bi00635a022)
25. Shivalingam, A., M. A. Izquierdo, A. L. Marois, A. Vyšniauskas, K. Suhling, M. K. Kuimova and R. Vilar. (2015), The interactions between a small molecule and G-quadruplexes are visualized by fluorescence

- lifetime imaging microscopy, *Nature Communications*, 6, 8178 (10.1038/ncomms9178 <https://www.nature.com/articles/ncomms9178#supplementary-information>)
26. Duff, M. R., J. Grubbs and E. E. Howell. (2011), Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity, *Journal of Visualized Experiments : JoVE*, 2796 (10.3791/2796)
 27. Pochetti, G. and R. Montanari. (2012), Isothermal titration calorimetry to determine the association constants for a ligand bound simultaneously to two specific protein binding sites with different affinities,
 28. Ward, W. H. and G. A. Holdgate. (2001), Isothermal titration calorimetry in drug discovery, *Prog Med Chem*, 38, 309-376
 29. Velazquez-Campoy, A., S. A. Leavitt and E. Freire. (2015), Characterization of protein-protein interactions by isothermal titration calorimetry, *Methods Mol Biol*, 1278, 183-204 (10.1007/978-1-4939-2425-7_11)
 30. Draczkowski, P., D. Matosiuk and K. Jozwiak. (2014), Isothermal titration calorimetry in membrane protein research, *J Pharm Biomed Anal*, 87, 313-325 (10.1016/j.jpba.2013.09.003)
 31. Freyer, M. W. and E. A. Lewis. (2008), Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions, *Methods Cell Biol*, 84, 79-113 (10.1016/s0091-679x(07)84004-0)
 32. Greenfield, N. J. (2006), Using circular dichroism spectra to estimate protein secondary structure, *Nat Protoc*, 1, 2876-2890 (10.1038/nprot.2006.202)
 33. Kypr, J., I. Kejnovska, D. Renciuik and M. Vorlickova. (2009), Circular dichroism and conformational polymorphism of DNA, *Nucleic Acids Res*, 37, 1713-1725 (10.1093/nar/gkp026)
 34. Vorlickova, M., I. Kejnovska, K. Bednarova, D. Renciuik and J. Kypr. (2012), Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes, *Chirality*, 24, 691-698 (10.1002/chir.22064)

35. Micsonai, A., F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Refregiers and J. Kardos. (2015), Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, *Proc Natl Acad Sci U S A*, 112, E3095-3103 (10.1073/pnas.1500851112)
36. del Villar-Guerra, R., R. D. Gray and J. B. Chaires. (2017), Quadruplex DNA Structure Characterization by Circular Dichroism, *Current protocols in nucleic acid chemistry*, 68, 17.18.11-17.18.16 (10.1002/cpnc.23)
37. Greenfield, N. J. (2006), Using circular dichroism spectra to estimate protein secondary structure, *Nature protocols*, 1, 2876-2890 (10.1038/nprot.2006.202)
38. Wei, Y., A. A. Thyparambil and R. A. Latour. (2014), Protein Helical Structure Determination Using CD Spectroscopy for Solutions with Strong Background Absorbance from 190-230 nm, *Biochimica et biophysica acta*, 1844, 2331-2337 (10.1016/j.bbapap.2014.10.001)
39. Siligardi, G. and R. Hussain. (2015), CD spectroscopy: an essential tool for quality control of protein folding, *Methods Mol Biol*, 1261, 255-276 (10.1007/978-1-4939-2230-7_14)
40. Lee, W., G. Cornilescu, H. Dashti, H. R. Eghbalnia, M. Tonelli, W. M. Westler, S. E. Butcher, K. A. Henzler-Wildman and J. L. Markley. (2016), Integrative NMR for biomolecular research, *J Biomol NMR*, 64, 307-332 (10.1007/s10858-016-0029-x)
41. Marion, D. (2013), An introduction to biological NMR spectroscopy, *Mol Cell Proteomics*, 12, 3006-3025 (10.1074/mcp.O113.030239)
42. Tognarelli, J. M., M. Dawood, M. I. F. Shariff, V. P. B. Grover, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson and M. J. W. McPhail. (2015), Magnetic Resonance Spectroscopy: Principles and Techniques: Lessons for Clinicians, *Journal of Clinical and Experimental Hepatology*, 5, 320-328 (10.1016/j.jceh.2015.10.006)
43. Marion, D. (2013), An Introduction to Biological NMR Spectroscopy, *Molecular & Cellular Proteomics : MCP*, 12, 3006-3025 (10.1074/mcp.O113.030239)

44. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich. (1988), Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, *Science*, 239, 487 (10.1126/science.239.4839.487)
45. Witte, A. K., P. Mester, S. Fister, B. Süß, M. Wagner and P. Rossmannith. (2018), PCR-Stop analysis as a new tool for qPCR assay validation, *Scientific Reports*, 8, 8275 (10.1038/s41598-018-26116-x)
46. Hanawalt, P. C. (2004), Density matters: the semiconservative replication of DNA, *Proc Natl Acad Sci U S A*, 101, 17889-17894 (10.1073/pnas.0407539101)
47. Brooijmans, N. and I. D. Kuntz. (2003), Molecular recognition and docking algorithms, *Annu Rev Biophys Biomol Struct*, 32, 335-373 (10.1146/annurev.biophys.32.110601.142532)
48. Gilbert, D. (2004), Bioinformatics software resources, *Brief Bioinform*, 5, 300-304
49. Forli, S., R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson. (2016), Computational protein-ligand docking and virtual drug screening with the AutoDock suite, *Nature protocols*, 11, 905-919 (10.1038/nprot.2016.051)
50. Morris, G. M., R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson. (2009), AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J Comput Chem*, 30, 2785-2791 (10.1002/jcc.21256)
51. Rizvi, S. M. D., S. Shakil and M. Haneef. (2013), A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians, *EXCLI journal*, 12, 831-857
52. Riss, T. L., R. A. Moravec, A. L. Niles, S. Duellman, H. A. Benink, T. J. Worzella and L. Minor. (2016), *Assay Guidance Manual [Internet]*. Eli Lilly & Company and the National Center for Advancing Translational Sciences.

53. Tominaga, H., M. Ishiyama, F. Ohseto, K. Sasamoto, T. Hamamoto, K. Suzuki and M. Watanabe. (1999), A water-soluble tetrazolium salt useful for colorimetric cell viability assay, *Analytical Communications*, 36, 47-50
54. Sylvester, P. W. (2011), Optimization of the tetrazolium dye (MTT) colorimetric assay for cellular growth and viability, *Methods Mol Biol*, 716, 157-168 (10.1007/978-1-61779-012-6_9)
55. Mahmood, T. and P.-C. Yang. (2012), Western blot: technique, theory, and trouble shooting, *North American journal of medical sciences*, 4, 429-434 (10.4103/1947-2714.100998)
56. Carlsson, N., A. Borde, S. Wolfel, B. Kerman and A. Larsson. (2011), Quantification of protein concentration by the Bradford method in the presence of pharmaceutical polymers, *Anal Biochem*, 411, 116-121 (10.1016/j.ab.2010.12.026)
57. Malke, H. (1990), J. SAMBROCK, E. F. FRITSCH and T. MANIATIS, *Molecular Cloning, A Laboratory Manual (Second Edition)*, Volumes 1, 2 and 3. 1625 S., zahlreiche Abb. und Tab. Cold Spring Harbor 1989. Cold Spring Harbor Laboratory Press. \$ 115.00. ISBN: 0-87969-309-6, *Journal of Basic Microbiology*, 30, 623-623 (10.1002/jobm.3620300824)
58. Thorpe, G. H. and L. J. Kricka. (1986), Enhanced chemiluminescent reactions catalyzed by horseradish peroxidase, *Methods Enzymol*, 133, 331-353

Chapter 3

Structural insights reveal the dynamics of the repeating r(CGG)^{exp} transcript found in Fragile X syndrome and Fragile-X associated tremor/ataxia syndrome (FXTAS)

3.1 Introduction

The expansion of simple sequence repeats (SSRs) is a dynamic genetic mutation that involves the stable expansion of a particular set of nucleotides (1 to 12) in various region of genes and causes more than 40 neurological diseases [1,2]. Among all expanded SSRs, expansion of trinucleotide repeats (TNRs) causes more than 30 neurological diseases [3]. Different TNRs create different neurological disease conditions by their expansion in various gene locations (5'UTR, promoter, exon, intron, and, 3'UTR) through a series of overlapping pathogenic mechanisms [4]. Huntington disease (HD), Dentatorubral-pallidoluysian atrophy (DRPLA), Spinocerebellar ataxia type 1/2/3/7, Spinal and Bulbar Muscular Atrophy (SBMA) are caused due to the expansion of a set of CAG repeats in *Huntingtin (HTT)*, atrophin-1 (*ATN1*), ataxin (*ATXN1/2/3/7*), and androgen receptor (*AR*) genes respectively [5,6]. Another genetic disorders, Myotonic dystrophy type 1 (DM1) and Myotonic dystrophy type 2 (DM2) are caused due to the expansion of CUG repeats at the (3'-UTR) of dystrophia myotonica protein kinase (DMPK) gene and r(CCUG) repeats expansion in the intronic region of zinc-finger 9 (ZNF9) gene, respectively. The expansion of CGG repeats in 5'UTR of fragile X mental retardation (*FMR1*) gene causes Fragile X syndrome (FXS) and Fragile X associated tremor ataxia syndrome (FXTAS) [7]. In a normal individual, the typical range of CGG repeats is 5-54 whereas the CGG expansion in between 55-200 repeats causes Fragile X-associated tremor/ataxia syndrome (FXTAS) and Fragile X-associated primary ovarian insufficiency

syndrome (FXPOI) [7-9]. However, when the numbers of CGG repeats exceed 200 (called full mutation) results in transcriptional silencing of the *FMR1* gene and causes Fragile X syndrome (FXS).

Despite different types of expanded repeats and their gene locations, all these diseases share two common pathogenic mechanisms i) Protein loss of function and ii) Protein/RNA toxic gain of function. The protein loss of function occurs due to inhibited and abortive expression of a specified protein that eventually causes the haploinsufficiency of repeat harboring gene [10]. Whereas the RNA gain of function results in the formation of RNA foci as well as the sequestration of cytoplasmic proteins. Additionally, the protein gain of function results in the repeat-associated non-AUG (RAN) translation of expanded repeats that forms the aggregates of toxic homopolymeric proteins [2,11]. All these pathomechanisms work as a toxic event for neuronal cells that leads to neurodegeneration and subsequently the disease condition.

To develop a potential therapeutics for the treatment of these expanded repeats associated neurological diseases, it is essential to study the structural insight of the transcript of expanded repeats. Previously, a structural insight of expanded CAG repeats RNA structure has been reported [12]. The NMR spectroscopy, X-Ray crystallographic, molecular dynamic simulation and umbrella sampling were employed to indicate the dynamic interaction between non-canonical AA pair that form in the transcript of expanded CAG repeat [13]. These studies revealed the involvement of multiple orientations of adenine nucleotide at AA pair and this information is extensively utilized by the investigators in developing the therapeutics for CAG repeat-associated neurological diseases [14-16]. Similarly, we also previously reported a crystal structure of CGG repeats-containing RNA that revealed the *syn-G/anti-G* pairing between each GG pairs [17]. However, this study did not identify the fluxional behavior of large amplitude motions that might be anticipated for 1×1 GG pair. Information of these dynamic motions could be significant in understanding the conformations of GG nucleotide that

governs the sequestration of cytoplasmic proteins or binding of small molecules. Therefore, the conformational flexibility in non-canonical GG pairs of expanded r(CG_n) repeats-containing RNA is the matter of significant interest. It will open the new dimension for the discovery of small bioactive molecules that can target 5'CGG/3'GGC motifs and could stop both the sequestration of RNA binding proteins to the RNA and non-AUG (RAN) translation that would lead to a normal phenotype.

The NMR technique is being extensively utilized by the investigators to resolve the solution structure of biological macromolecule and to reveal the dynamics involved in internal residues [18,19]. Thus, to understand the dynamic behavior of internal GG pairs, we studied the NMR based solution structure at different temperature and found the equilibration between *syn*-G/*anti*-G and *anti*-G/*anti*-G conformations of internal GG pairs along with closing standard Watson-Crick GC pairing. Along with the NMR, molecular dynamic (MD) simulation and enhanced sampling techniques are also an excellent choice to study the dynamic motion involved in expanded nucleotide harboring RNA/DNA [20,21]. In case of expanded CAG and CUG repeats RNA, the possible routes for the *anti-anti* to *anti-syn* transformation were elucidated using umbrella sampling and molecular dynamic simulation [22,23]. Umbrella sampling is one of the analytical techniques for this kind of system as it can explore the high energy conformational barriers within a short time by an adequately chosen collective variables (CV). This sampling technique works with the addition of bias potential that is a function of CVs of that system and produces the system as a multiple overlapping umbrella of the transition state. In the field of RNA dynamics, Umbrella sampling uses one or two CV [24-28]. Weighted Histogram Analysis Method (WHAM) is a widely used technique which is merged with umbrella sampling technique to compute the potential mean force (PMF) surfaces [29]. In this context, PMF and free energy surfaces are easily interchangeable. In the field of RNA repeat-associated neurological diseases, computational studies on different trinucleotide repeat [22,23,30,31] show the dynamic behavior of those

mismatch pairs. There is also one study depicted upon the binding of the ligands to CAG repeats using the well-tempered metadynamics method [32]. Stability and conformational preference of CGG and CCG mismatch in the DNA and RNA is a function of the nature of sequence environment which was studied by Pan et al. using adaptive biased molecular dynamics method [33]. Along with the trinucleotide repeats one hexanucleotide repeat was simulated by Zhang et al., and the dynamicity was quantified by using principal component analysis, the open-angle of the base flipping and other torsion angles [34].

Therefore, along with NMR, we employed the conventional MD simulation and umbrella sampling to simulate the dynamic motions involved in non-canonical GG pairs. Two other different sequences along with the NMR structure was simulated to shed lights on the dynamical behavior. A (3xCGG) repeats sequence (5'-GGGCCGCGCGGGUCC-3') crystallized by Kumar et al. [17] and another model sequence (5'-GCCGCGCGCG-3') was used to run a dynamic study in explicit solvent using AMBER force field. Dynamic simulations for all these sequences were run up to 1 μ s and characterize the fluctuation of the mismatch base pairs, nature and the time evolution of the base flipping, ion distribution and backbone torsional angle. Apart from the 1 μ s long molecular dynamics, a 2D free energy landscape was constructed with the help of umbrella sampling, molecular dynamics as well as utilizing the WHAM to figure out the minimum structures and possible transition pathways between them.

This study reveals the internal dynamics involved in expanded CGG repeats by employing the NMR, molecular dynamics simulation and umbrella sampling methods. These dynamic motions rule the binding of cytoplasmic RNA interacting proteins and can be utilized in the rational discovery of a small molecule modulator. These small molecule modulators may further be used as potential therapeutic molecules to provide a possible treatment of debilitating diseases like FXTAS and FXS.

3.2 Results & Discussion

3.2.1 One dimensional proton (^1H) NMR spectra of [(5'-CCGCGGCGG-3')₂] at variable temperatures revealed the dynamics at noncanonical GG hair-pin loop

NMR spectroscopy has been recognized as one of the most robust techniques to explore the structural dynamics in aqueous solution near the physiological conditions [18] [12,35] [36]. Therefore, We recorded the one-dimensional ^1H NMR spectra at variable temperatures (from 298K to 318K with a gradient of 5K) to assess the structural dynamics at internal noncanonical GG pairs of 5'r(CCGCGGCGG)₂ RNA (Figure 3.1 A & B). With an increase of temperature, a broadening of G6 and G5 imino proton resonance peak was observed. The G5NH imino proton resonance also showed a clear upfield shift in the imino region of proton NMR spectra. The peaks G5NH and G6NH broadened and eventually disappeared as the temperature was increased from 293 to 318 K (Figure 3.1 B). This broadening and shifting signifies the exchange of G6NH and G5NH protons with solvent due to a possibility of nucleotide base flipping and affirms the dynamics involved in non-canonical GG pair. Additionally, an appearance of G5H1' resonance peak was observed at a higher temperature (318K) however; the same peak was found to be absent at lower temperature (293K) (Figure 3.1 B). Moreover, a slight upfield shift was observed in the resonance peak of G6H1' atom. Both events evident the dynamics near an internal GG loop (Figure 3.1 B).

To reaffirm the dynamic nature of GG pairs, a 2D NOESY ^1H NMR spectra were also recorded at 298 K and 318 K in H₂O:D₂O (9:1) ratio. 2D NOESY ^1H NMR spectra examine the sequential connectivity of sugar and bases at two different temperature, 298K, and 318K respectively (Figure 3.2). The disappearance of the cross peak of G6H1'- C7H6 atoms has been observed at higher temperature (318 K) (marked as an asterisk in the figure) that depicts the

dynamics near internal GG pair. Moreover, change in the intra nucleotide (G5) resonance peak intensity shows the movement of atoms and dynamic behavior at a higher temperature (318 K) (Figure 3.2).

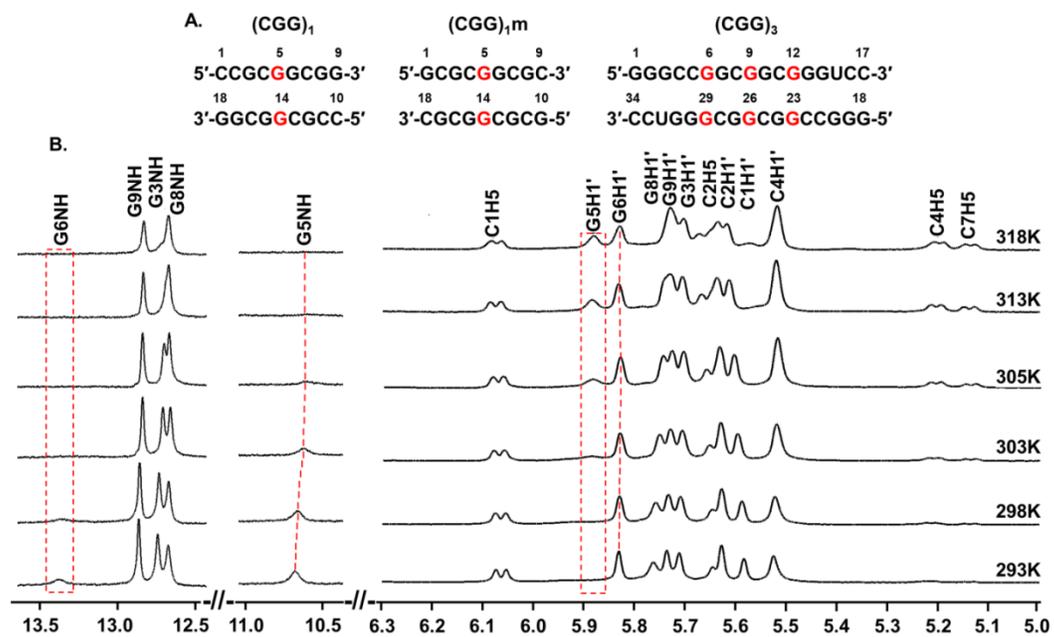


Figure 3.1 A) Showing sequence of RNA molecules used in molecular dynamics and umbrella sampling simulations. B) Temperature-dependent ^1H NMR spectra for 5' r(CCGCGGCGG)₂ showing H' region (5.0-6.3 ppm) and imino region (10.5-13.5 ppm).

3.2.2 Dynamics of (CGG)₁, (CGG)_{1m} and (CGG)₃ structure

The helical architecture of the NMR structure, (CGG)₁ was analyzed using the 3DNA software [37] and all data involving dihedral angles, global helical parameters, widths of the major and minor grooves were reported in the Supporting Information (Table 3.1 - 3.6). Molecular dynamics (MD) simulations of the NMR structure, (CGG)₁ was carried out for 1 μ s to study the dynamicity of the RNA structure where the initial GG mismatch was in *anti-anti* conformation. Further, a model structure (CGG)_{1m} was also built into *anti-anti* and *syn-anti* conformations and also simulated for 1 μ s. The time evolution of the root-mean-square deviation (RMSD) of every heavy atom of the mismatch nucleotides, i.e. G₅-G₁₄ from the respective initial structure was calculated for all three systems ((CGG)₁ *anti-anti*, (CGG)_{1m} *anti-anti* & *syn-anti*) (Figure 3.3 A1 - C1). Furthermore, the distribution of Na⁺ ions around the mismatch region for all cases is shown in the same figure (Figure 3.3 A2 - C2). As is evident from Figure 3.3, the RMSDs of (CGG)₁ and (CGG)_{1m}, with an initial *anti-anti* conformation, show frequent jumps from the initial minimum position just after few ns and fluctuated between these two states throughout the time scales. The modeled CGG where *anti-syn* is the starting conformation shows a very stable RMSD up to 700 ns. After that a sudden jump in the RMSD from the initial structure was observed and remained in that state for nearly 100 ns. It may be due to the formation of two hydrogen bonds in *syn* orientation of base that leads to the more stabilized state. In the case of *anti-anti* conformations, several *anti-anti* conversions similar to our data were also observed in CUG repeat molecular simulation causing Myotonic Dystrophy disease [23]. We extracted the structures for each state and examined the glycosidic angle to determine their respective state. All the structures were shown in Figure 3.3 A3 - C3. In spite of such variations, stacking to the initial conformational stage, i.e. *anti-anti* or *anti-syn* is a general tendency of CGG repeats that also been observed in other study involving these same repeats in different sequences [33].

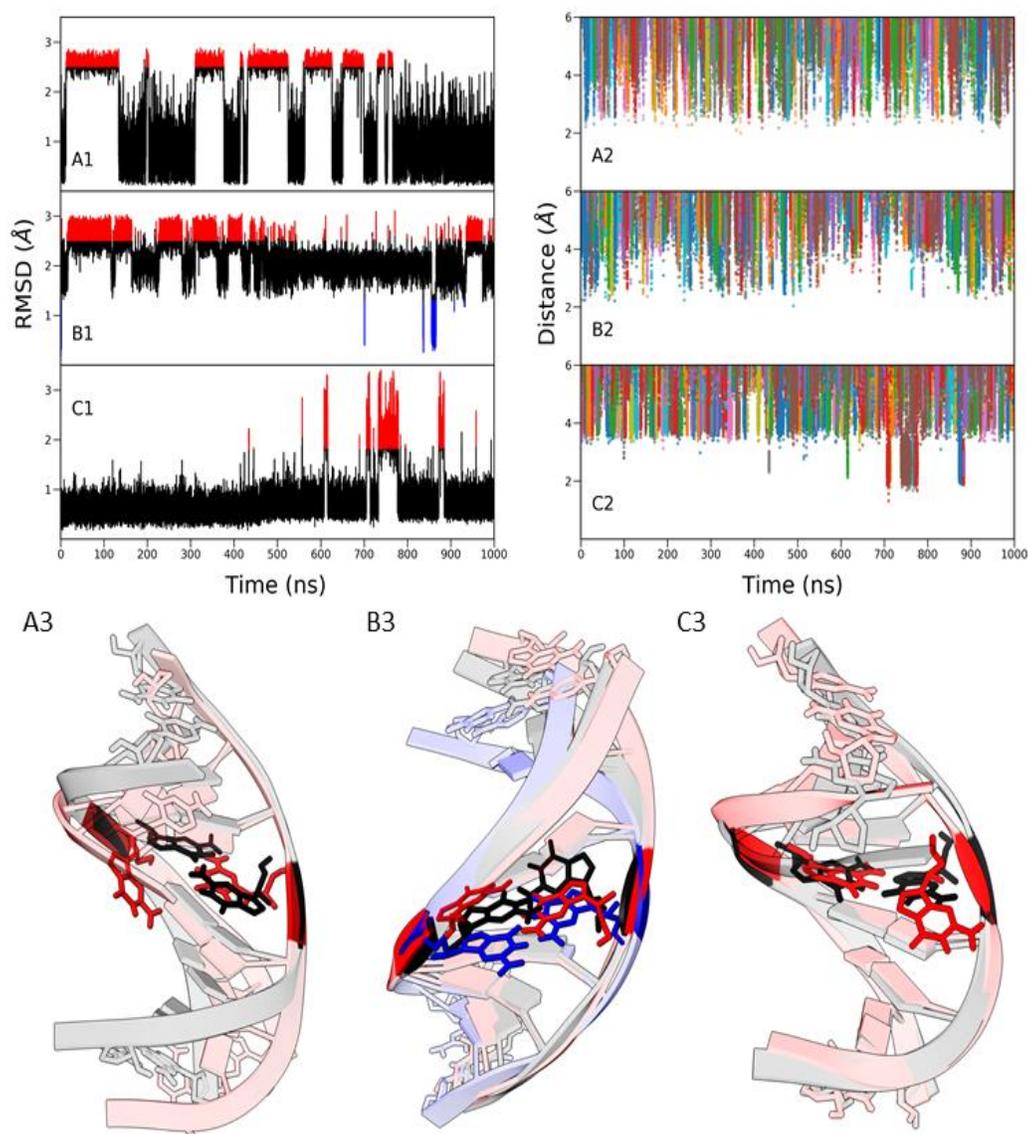


Figure 3.3 RMSD and Na^+ distribution around mismatch base pairs of $(\text{CGG})_1$ and $(\text{CGG})_{1m}$. A1 and A2 are RMSD of GG mismatch pair and distance of Na^+ ions from the GG mismatch pair of $(\text{CGG})_1$. Similarly, B1, B2 and C1, C2 are RMSD and Na^+ distribution of $(\text{CGG})_{1m}$. Initial GG pairing in B1 and B2 is anti-anti and whereas syn-anti in case of C1 and C2. For each Na^+ ion we use different colors for better visualization. 3D conformation of all three systems is extracted from the different region as shown in RMSD plot. Red, black and blue color in 3D conformation depicts three different colored regions in the plot. A3, B3 and C3 represents $(\text{CGG})_1$, $(\text{CGG})_{1m}$ (anti-anti) and $(\text{CGG})_{1m}$ (syn-anti).

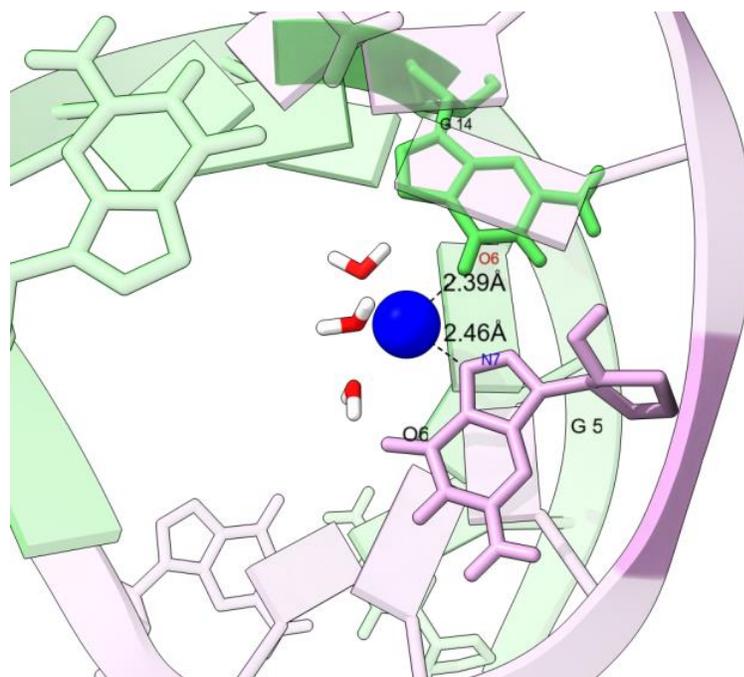


Figure 3.4 Na^+ bound state in the different orientation of RNA structure from the initial anti-syn GG base pair as found in $(\text{CGG})_{1m}$.

Table 3.1 Sugar and backbone torsional angles ($^{\circ}$) calculated for $(CGG)_1$.

Strand I							
Base	α	β	γ	δ	ϵ	ζ	χ
C-1	---	---	50.8	90.4	-155.6	-73.6	-163.4
C-2	-75.9	162.1	64.5	79.1	-155.1	-74.2	-166.7
G-3	-66.8	159.5	59.3	81.0	-155.1	-72.2	165.4
C-4	-71.9	164.9	70.2	76.8	-121.3	-89.9	-167.7
G-5	-60.2	111.1	81.1	79.9	-140.1	-79.0	-177.6
G-6	-78.9	176.9	49.3	85.1	-155.1	-71.1	-162.5
C-7	-72.2	153.0	67.6	79.9	-153.5	-75.2	-162.4
G-8	-69.4	166.3	59.3	83.6	-153.3	-76.5	-167.2
G-9	-74.9	168.1	63.7	85.5	---	---	-162.5
Strand II							
G-1	-55.4	152.6	60.2	80.2	---	---	-161.8
G-2	-70.6	160.0	65.1	77.5	-149.6	-79.3	-163.1
C-3	-68.4	162.1	66.7	75.4	-157.3	-70.5	-158.7
G-4	-70.8	177.3	55.7	81.0	-152.4	-73.3	-166.8
G-5	-69.2	112.1	86.4	66.5	-140.6	-64.9	-178.3
C-6	-69.4	155.1	74.0	75.2	-123.7	-88.9	-166.3
G-7	-70.3	157.4	65.2	74.4	-148.6	-75.9	-163.3
C-8	-67.9	154.2	67.7	75.8	-149.1	-79.9	-160.9
C-9	---	---	57.9	84.8	-155.4	-72.7	-167.9

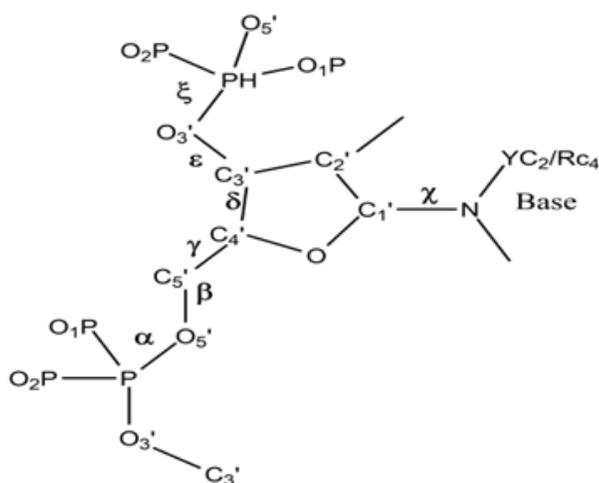


Table 3.2 Distances (Å) and angles (°) of atoms for different base pairs of (CGG)₁.

Local base-pair parameter						
Base pair	Shear (Å)	Stretch(Å)	Stagger(Å)	Buckle (°)	Propeller (°)	Opening (°)
C-G	0.12	-0.08	-0.33	-1.13	-2.41	0.55
C-G	0.09	-0.16	-0.28	0.39	-4.31	-0.25
G-C	-0.06	-0.13	-0.14	-2.69	-0.16	-1.10
C-G	0.25	-0.10	-0.22	1.84	-2.30	-0.80
G-G	3.65	0.87	-0.64	-2.57	-34.05	-34.10
G-C	-0.24	-0.12	-0.15	-2.42	-3.12	-0.59
C-G	0.30	-0.15	-0.18	-0.49	-1.00	1.09
G-C	-0.10	-0.19	-0.25	-2.79	-2.74	-1.07
G-C	-0.06	-0.10	-0.29	3.27	-2.61	-2.09
Average	0.44	-0.02	-0.27	-0.73	-5.86	-4.26
Std. Dev.	1.22	0.33	0.15	2.19	10.64	11.23

Table 3.3 Global helical parameters calculated for the base pairs of (CGG)₁.

Base Pair	λ (I) (°)	λ (II) (°)	C1'-C1' (Å)	RN9-YN1 (Å)	RC8-YC6 (Å)
C1-C18	54.7	55.8	10.8	9.1	9.9
C2-G17	54.2	54.0	10.7	9.0	9.9
G3-C16	54.3	53.5	10.8	9.0	9.9
C4-G15	51.9	56.1	10.8	9.1	10.0
G5-G14	60.4	25.7	13.1	11.0	11.0
G6-C13	55.0	53.0	10.7	9.0	9.9
C7-G12	52.1	51.7	10.8	9.0	9.9
G8-C11	55.4	50.8	10.7	8.9	9.8
G9-C10	56.4	56.1	10.8	9.1	9.9

Lambda is the virtual angle between C1'-YN1 or C1'-RN9 glycosidic bonds and the base-pair C1'-C1' line.

C1'-C1' is the distance between C1' atoms for each base-pair.

RN9-YN1 is the distance between RN9-YN1 atoms for each base-pair.

RC8-YC6 is the distance between RC8-YC6 atoms for each base-pair.

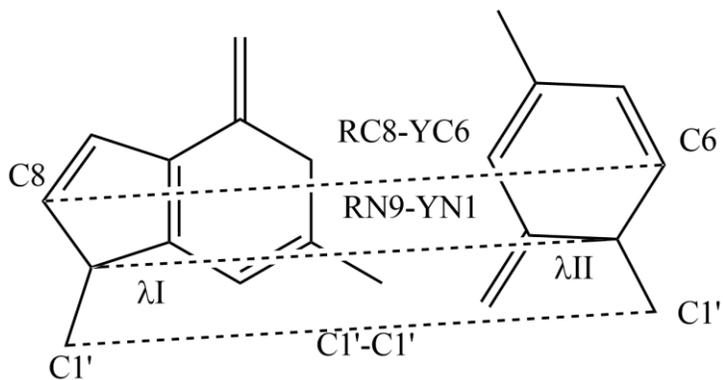


Table 3.4 Helical parameters for different base pairs and steps of (CGG)₁.

Base pair	Displacement (Å)	Angle (°)	Twist (°)	Rise (Å)
C1-G18	6.45	14.91	33.63	2.80
C2-G17	6.38	14.10	32.91	2.77
G3-C16	6.33	14.06	31.74	2.92
C4-G15	6.28	13.05	28.20	2.91
G5-G14	5.63	13.20	38.19	2.85
G6-C13	6.17	13.76	31.63	2.84
C7-G12	6.08	14.30	33.23	2.79
G8-C11	6.12	14.54	33.02	2.78
G9-C10	6.29	14.95	---	---

Table 3.5 Minor and major groove widths according to refined P-P distances for the direction of sugar-phosphate backbone in (CGG)₁.

Step	Refined Minor Groove Widths (Å)	Refined Major Groove Widths (Å)
CC/GG	---	---
CG/CG	---	---
GC/GC	17.9	15.7
CG/GG	17.7	15.6
GG/CG	17.8	15.6
GC/GC	17.9	15.7
CG/CG	---	---
GG/CC	---	---

Table 3.6 Helical parameters for different base pairs and steps of (CGG)₁.

Local base-pair parameters						
Step	Shift (Å)	Slide (Å)	Rise(Å)	Tilt (°)	Roll (°)	Twist (°)
CC/GG	0.03	-1.33	3.33	0.92	9.37	31.66
CG/CG	0.03	-1.36	3.36	-0.55	11.30	31.06
GC/GC	0.06	-1.33	3.22	1.25	6.05	32.67
CG/GG	-0.99	-1.35	3.81	5.44	10.00	41.20
GG/CG	0.80	-1.82	3.09	-0.58	13.41	18.58
GC/GC	0.07	-1.26	3.30	-0.04	7.42	33.94
CG/CG	-0.14	-1.37	3.32	0.78	11.10	29.24
GG/CC	-0.05	-1.40	3.18	-0.59	8.97	32.02
Average	-0.02	-1.40	3.33	0.83	9.70	31.30
Std. Dev.	0.48	0.17	0.21	2.00	2.31	6.25
Local base-pair parameters						
Step	Shift (Å)	Slide (Å)	Rise (Å)	Tilt (°)	Roll (°)	Twist (°)
CC/GG	0.03	-1.33	3.33	0.92	9.37	31.66
CG/CG	0.03	-1.36	3.36	-0.55	11.30	31.06
GC/GC	0.06	-1.33	3.22	1.25	6.05	32.67
CG/GG	-0.99	-1.35	3.81	5.44	10.00	41.20
GG/CG	0.80	-1.82	3.09	-0.58	13.41	18.58
GC/GC	0.07	-1.26	3.30	-0.04	7.42	33.94
CG/CG	-0.14	-1.37	3.32	0.78	11.10	29.24
GG/CC	-0.05	-1.40	3.18	-0.59	8.97	32.02
Average	-0.02	-1.40	3.33	0.83	9.70	31.30
Std. Dev.	0.48	0.17	0.21	2.00	2.31	6.25

A total of 16 Na⁺ ions were used in all the three cases to neutralize the systems. All the ions were displayed using different colours to show (Figure 3.3) their occupancy near to the GG mismatch base pair. In case of (CGG)₁ and (CGG)₁m (*anti-anti*), there were no sodium binding pockets or grooves throughout the simulation period of 1 μs as shown in Figure 3.3 (A2 and B2). A steady sodium binding region was also observed in the case of (CGG)₁m (*syn-anti*) around 700 ns, where the sequence takes different orientation from the initial. (Figure 3.3 (C2)). Similarly, in the case of CAG repeat[22], the sequence involving an initial *anti-anti* conformation also shows no steady sodium binding pocket throughout their simulation. On the other hand, their *syn-anti* structure was found to possess a regular Na⁺ binding pocket throughout the simulation, which is not found in our cases. However, we observed a less probable but stable secondary state with sodium binding groove. A snapshot of the nature of the sodium binding after 700 ns is shown in Figure 3.4. The (CGG)₃ was an r(3xCGG)^{X-RAY} repeat RNA crystal structure (PDB ID 3SJ2) [17] wherein all three GG internal loops were in *syn-anti* conformation, i.e. in GG pair, while one G was *anti*, the other one was in *syn* conformation. There were two hydrogen bonds in each GG pair between Hoogsteen and Watson-Crick faces. Root-mean-squared deviations of backbone atoms of RNA and GG pairs as well as distributions of Na⁺ around mismatch base pairs are shown in Figure 3.5.

The RMSD of backbone atoms compared to the crystal structure was stable throughout the simulation. Overall, in all three GG pairs, initial *syn-anti* conformation was remaining in initial position. In the case of G6-G29 at about 900 ns there was a certain transition in a base pair (Figure 3.5 B1); however, the other two GG mismatch pair G9-G26 & G12-G23, was showing very less transition throughout the 1 μs simulation (Figure 3.5 D1 and F1). Out of total 32 Na⁺ ions in the simulation box shown in different colors, had no stable binding near to the GG base pair. In the case of transition, stable binding of Na⁺ ions indicates the role of ions to stabilize the other orientation from the crystal.

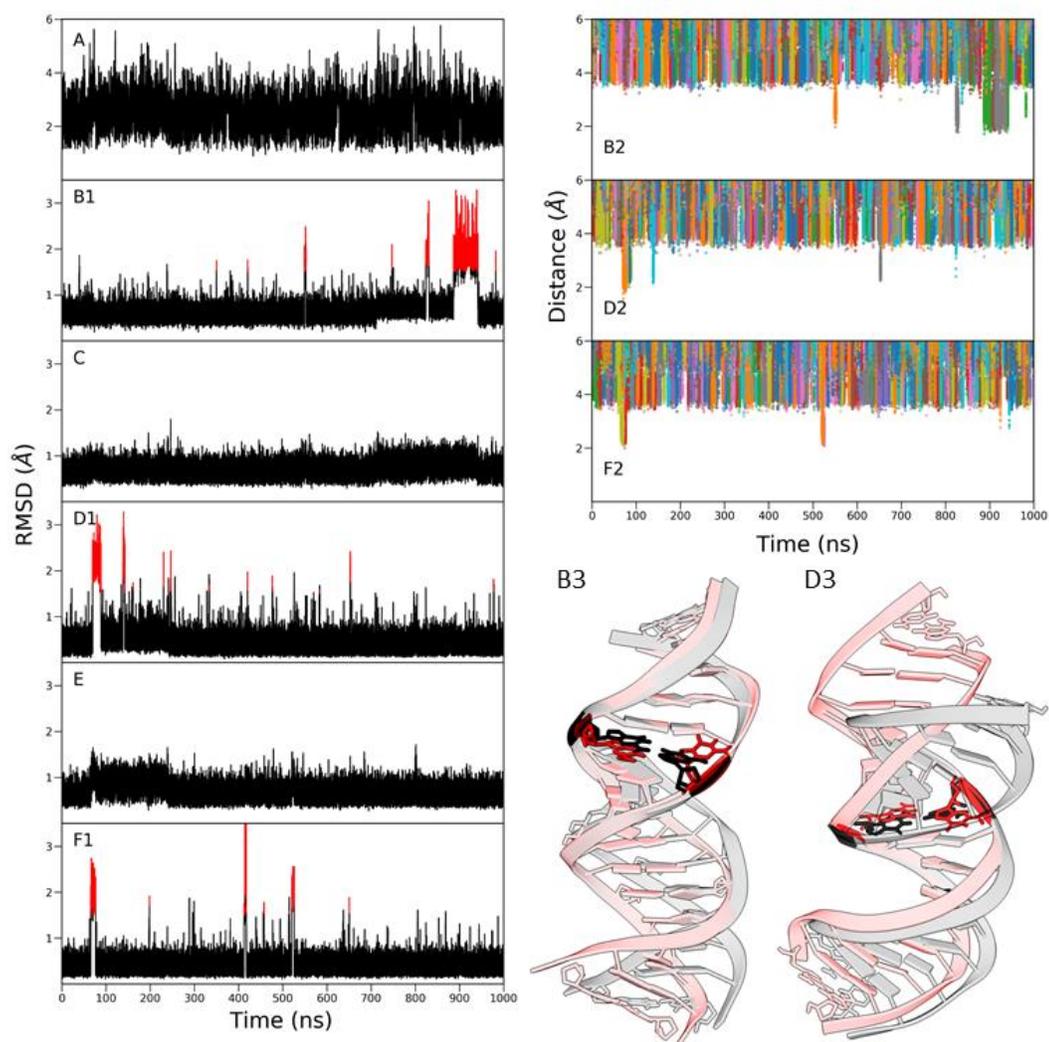


Figure 3.5 RMSD and Na⁺ distribution analysis of (CGG)₃. Root mean square deviation is shown in A backbone atom B1 G6-G29 base-pair C G7C8-G27C28 D1 G9-G26 base-pair E G10C11-G24C25 F1 G12-G23 base pair. B2, D2, F2 are the distance between Na⁺ ions and G6-G29, G9-G26, G12-G23 respectively. The residue numbering scheme is defined in Figure 1. 3D conformation corresponds to the G6-G29 (B3), and G9-G26 (D3) base pairs are also shown in the image. Different states are shown in the same color scheme used in the RMSD plot.

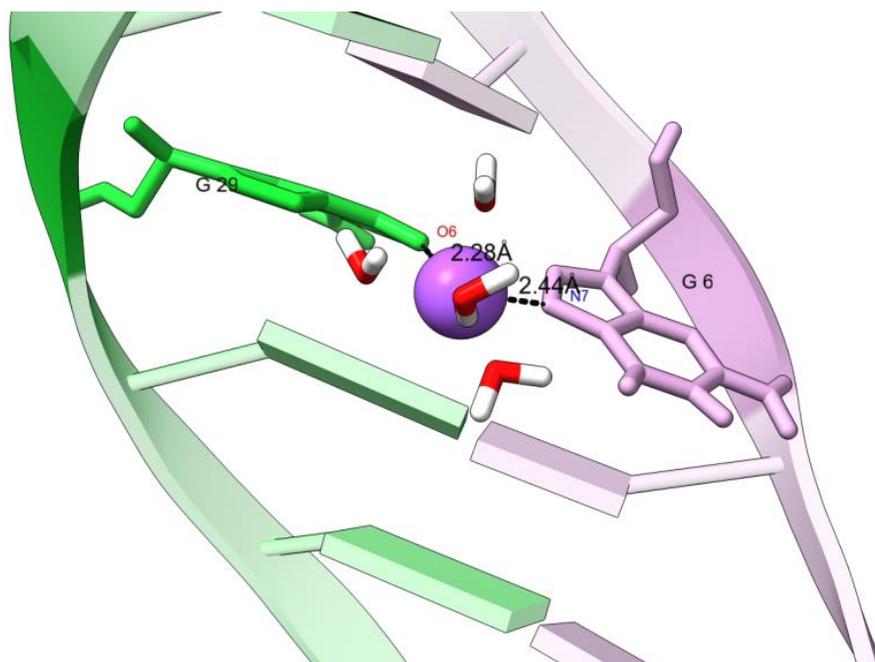


Figure 3.6 Na⁺ bound state in anti-syn GG base pairs as found in (CGG)₃.

Figure 3.6 also shows the interaction between Na^+ and nucleic acid bases. Another study on CGG repeat also shows a stable conformation throughout the 1 μs long simulation [33]. This is in contrast to what has been observed by Yildirim and coworkers when they simulated a $\text{r}(3\text{xCAG})^{\text{X-RAY}}$ motifs for 500 ns [22]. In the case of $\text{r}(\text{CAG})_3$, the conversion between *anti-anti* to *anti-anti* or even *anti-syn* to *anti-anti* was observed. Further, the transition of *anti-syn* was involved with sodium binding states. Therefore, the nature of the sequence, especially the nature of the repeats is highly responsible for this kind of transformation. The mismatch base pairs which includes guanine nucleotide is trying to remain in *anti-syn* states which were also evident from another study [33].

To elucidate the nature of the transition, we plotted the distribution of the glycosidic angle (χ) for all cases. The initial state is conserved in all the cases, i.e., *anti-anti* remains in *anti-anti* or same for *syn-anti*, which is a common feature of all the distributions. The NMR structure showed a very narrow distribution of χ (Figure 3.7 A) for both the GG mismatch bases with peaks around -160° and 170° . In contrast to the NMR structure, two distinct peaks were observed (Figure 3.7 B) around -175° and -80° in the case of $(\text{CGG})_{1\text{m}}$. Therefore, the behavior of the same mismatch sequence can vary depending upon the sequence environment. The distribution of χ angle for the *syn-anti* model ($(\text{CGG})_{1\text{m}}$) structure showed that the *syn* base (G5) remains in *syn* conformation having a bi-modal distribution with peaks around 30° and 70° , respectively (Figure 3.7 C). Overall, a single guanidine in GG mismatch maintains its initial state of conformation and fluctuates in that state of conformation. The same nature of the χ angle also found in the case of $(\text{CGG})_3$ sequence, as is evident from Figure 3.7 also. Here, initially, G6, G9, and G23 were in *syn* states and throughout the simulation distributions of these bases were also in the *syn* region. G6 and G9 fluctuate around 35° with a little hump around 75° , whereas G23 shows perfect Gaussian distribution (Figure 3.7 D, E, and F).

Further, we calculated the centre of mass pseudo dihedral (CPD) angle [38] to estimate the direction of base flipping, i.e. along with the major groove or around the minor groove (Figure 3.8). The positive definition value of CPD angle signifies flipping through the minor groove while negative values indicate flipping *via* the major groove. The time evolution and the occupancy of the CPD angle for all the GG mismatch nucleotides are shown in Figure 3.8. In cases of (CGG)₁ and (CGG)_{1m} (*anti-anti*), one base was flipping around the minor groove whereas the other base was flipping around the major groove. Therefore, it shows an entirely uncorrelated motion. On an average, three distinct states were obtained: around the major and minor grooves while the third state showed the least flipped state of the two bases, which also slightly around minor groove (Figure 3.8 A and B). In contrast, the single repeat *syn-anti* model sequences showed very less flexibility during the simulation (Figure 3.8 C). In the case of (CGG)₃, the base flipping angles for all three GG mismatches indicate that the flipping was mostly directed around the minor groove (Figure 3.8 D, E, and F).

As the magnitude is very less for all the bases, it also indicates a stable mismatch agreeing with our previous calculations. Further, we have investigated the backbone sugar torsional angle (δ) along with the glycosyl torsion (χ). From Figures 3.9 and 3.10, a large dispersion in the torsion angle is observed for some cases. In our 1 μ s long MD simulation, there is only one stable hydrogen bond observed in the case of *anti-anti* GG mismatch, and two hydrogen bonds are formed in the case of *syn-anti* GG mismatch which may result in the higher stability of *syn-anti* GG mismatch base-pair. In *anti-anti* GG mismatch, the hydrogen bond is formed between O6 atom and N2 atom of the G5 and G14 residues which interchange throughout our simulation time scale. In a *syn-anti* GG mismatch, H-bonds are formed between N1 (*anti*-base) with O6 (*syn* base), N2 (*anti*-base) with N7 (*syn* base). All these hydrogen bonds were present with an average occupancy of 50% throughout the simulations, which results in a stable conformation of those sequences. A detailed percentage occupancy is listed in the supporting information (Table 3.7). From Table 3.7, it is evident that the

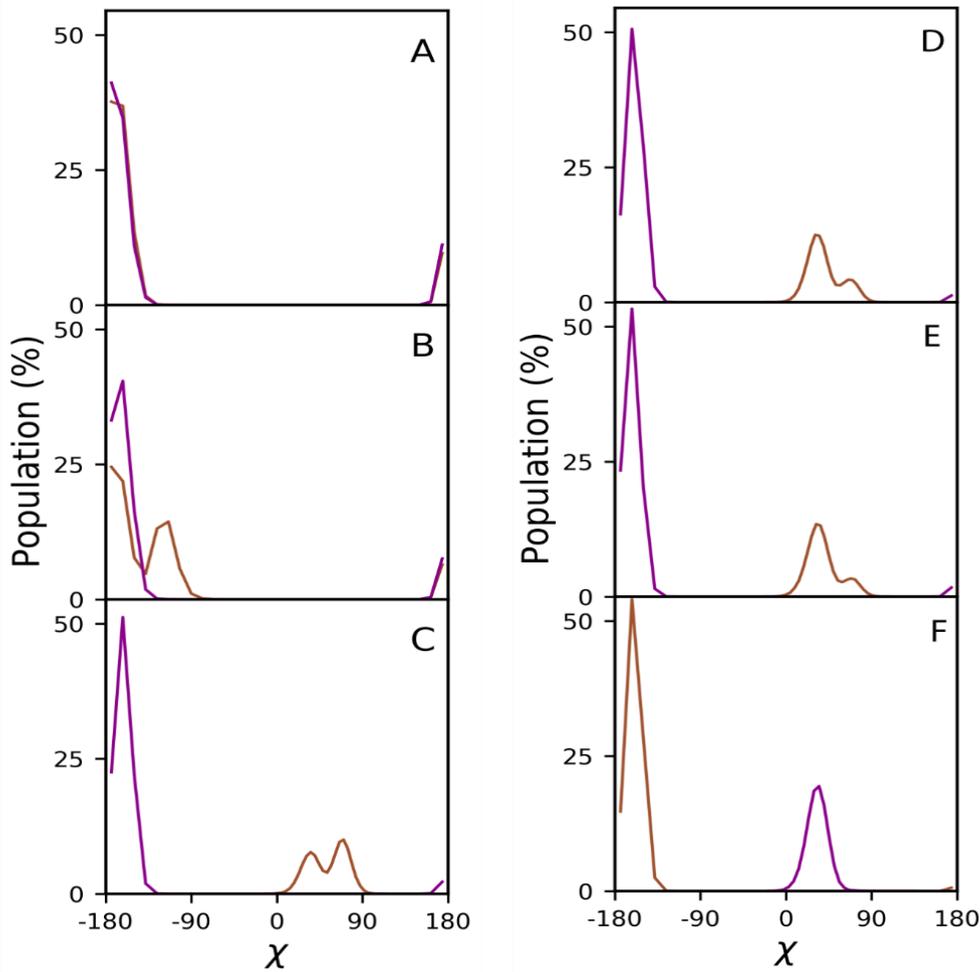


Figure 3.7 Distribution of χ angle in four cases. In left hand side image, A) (CGG)1 (anti-anti), B) (CGG)1m (anti-anti) and C) (CGG)1m (anti-syn) where brown corresponds to G5 and purple corresponds to G14. In Right hand side image D) G6 (brown)-G29 (purple) E) G9 (brown)-G26 (purple) F) G12 (brown)-G23(purple) of (CGG)3 sequence.

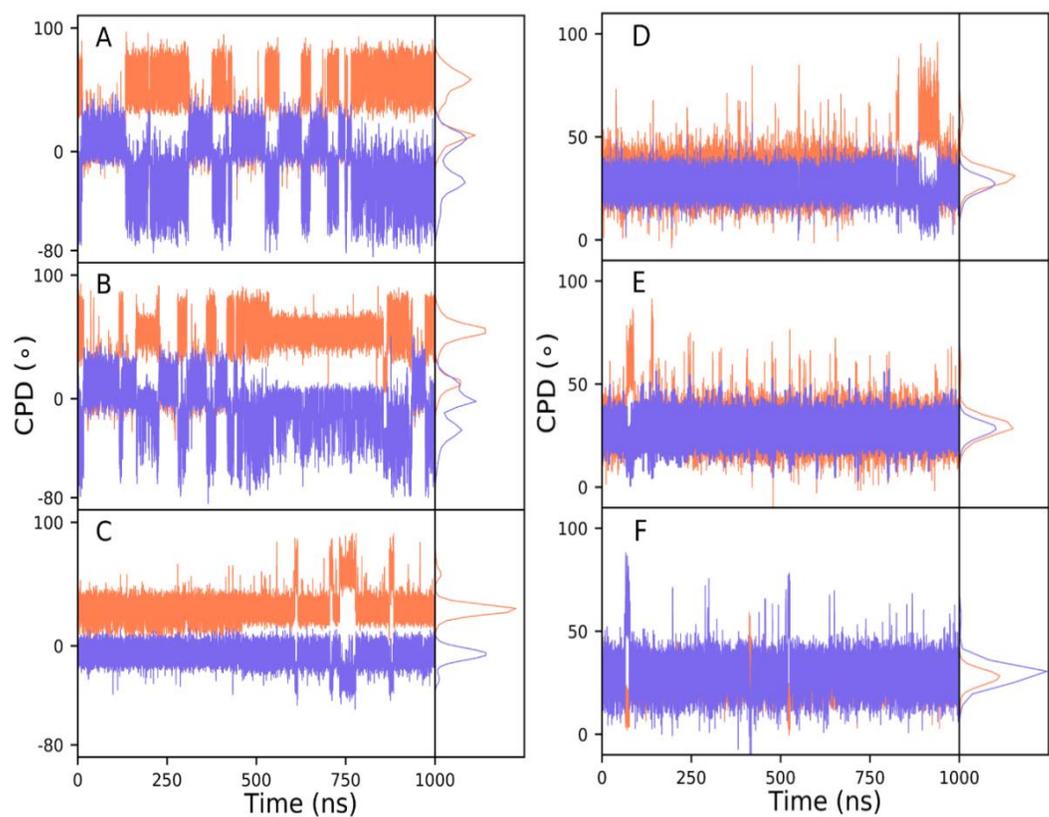


Figure 3.8 Time evolution of centre of mass pseudo dihedral (CMPD) angle as a function of simulation time. Left-hand side: A) (CGG)1 (anti-anti), B) (CGG)1m (anti-anti) and C) (CGG)1m (syn-anti) where purple is corresponds to G5 and brown corresponds to G14; Right-hand side: A) G6-G29 B) G9-G26 C) G12-G23 where purple corresponds to G6, G9 and G12 and brown corresponds to G23, G26 and G29.

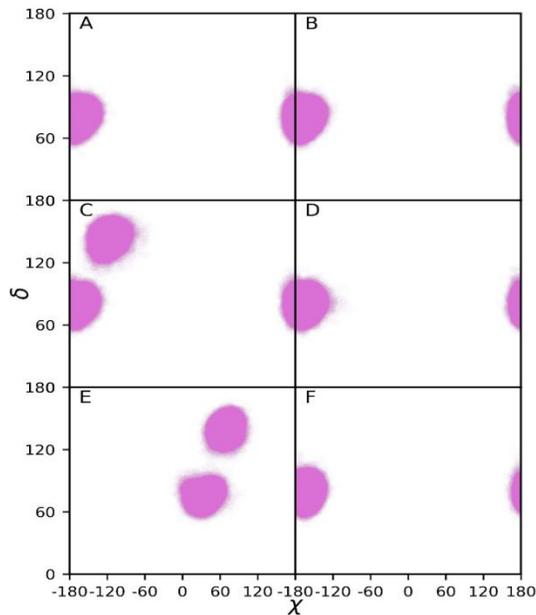


Figure 3.9 Backbone sugar torsion δ and the glycosyl torsion χ angle distribution of (CGG)1, (CGG)1m (anti-anti) and (CGG)1m (anti-syn) in each row respectively where left hand side corresponds to G5 and right hand side is G14.

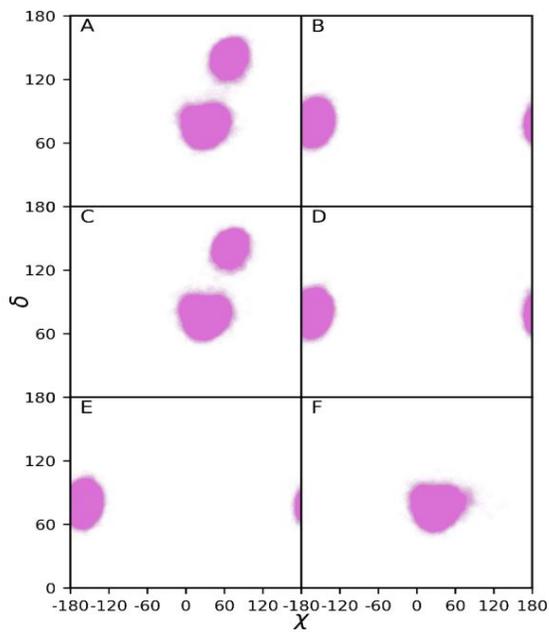


Figure 3.10 Backbone sugar torsion δ and the glycosyl torsion χ angle distribution of (CGG)3. A) G6, B) G29, C) G9, D) G26, E) G12 and F) G23

percentage of H-bond occupancy is much higher in *syn-anti* mismatch compared to the *anti-anti* mismatch. This is one of the main reasons for the higher stability of *syn-anti* mismatch compared to the *anti-anti* mismatch.

The dynamical behavior of the mismatch GG base pair can also be further investigated by constructing the cross-correlation map, as shown in Figure 3.11. Atoms involved in the mismatched pair are indicated in red-colored rectangles. A similar result was obtained for the (CGG)₁ and modeled structures. A slight decrease in correlation was observed in modeled *anti-anti* structures compared to *syn-anti*, which indicates the higher flexibility of those base pair. Another exciting nature of these sequences was found around atom number 300 to 400, which corresponds to 10th to 12th bases which reside in the terminal. These regions show an increased order of anti-correlation with the other end of the RNA duplex. The anti-correlation increases around two-fold in the NMR structure to the model structure, implying a global deformation of conformation. The span of these anti-correlation regions further increases in the model structure, where the starting conformation of the mismatch base pair was *anti-syn*. From that, we can conclude that not only sequence, initial states also keep a trace in global conformation in long time scale simulation. On the other map of the (CGG)₃ shows a very stable nearly four-fold symmetry. Here, the relation between the mismatch base pair showing an anti-correlated motion which was also been found from the different angle calculation. For a long sequence, single mismatch in regular interval may not be able to change the global conformation.

Table 3.7 H-bond percentage occupation different CGG mismatch sequences

Acceptor	Donor	Occupancy (%)
(CGG)₁ (anti-anti)		
G14@O6	G5@N2-H	29.23
G5@O6	G14@N2-H	21.89
(CGG)_{1m} (anti-anti)		
G14@O6	G5@N2-H	19.41
G5@O6	G14@N2-H	15.53
(CGG)_{1m} (anti-syn)		
G5@O6	G14@N1-H	56.26
G5@N7	G14@N2-H	42.14
(CGG)₃		
G6@O6	G29@N1-H	56.84
G6@N7	G29@N2-H	49.97
G9@O6	G26@N1-H	62.11
G6@N7	G26@N2-H	47.89
G23@O6	G12@N1	67.80
G23@N7	G12@N2-H	47.67

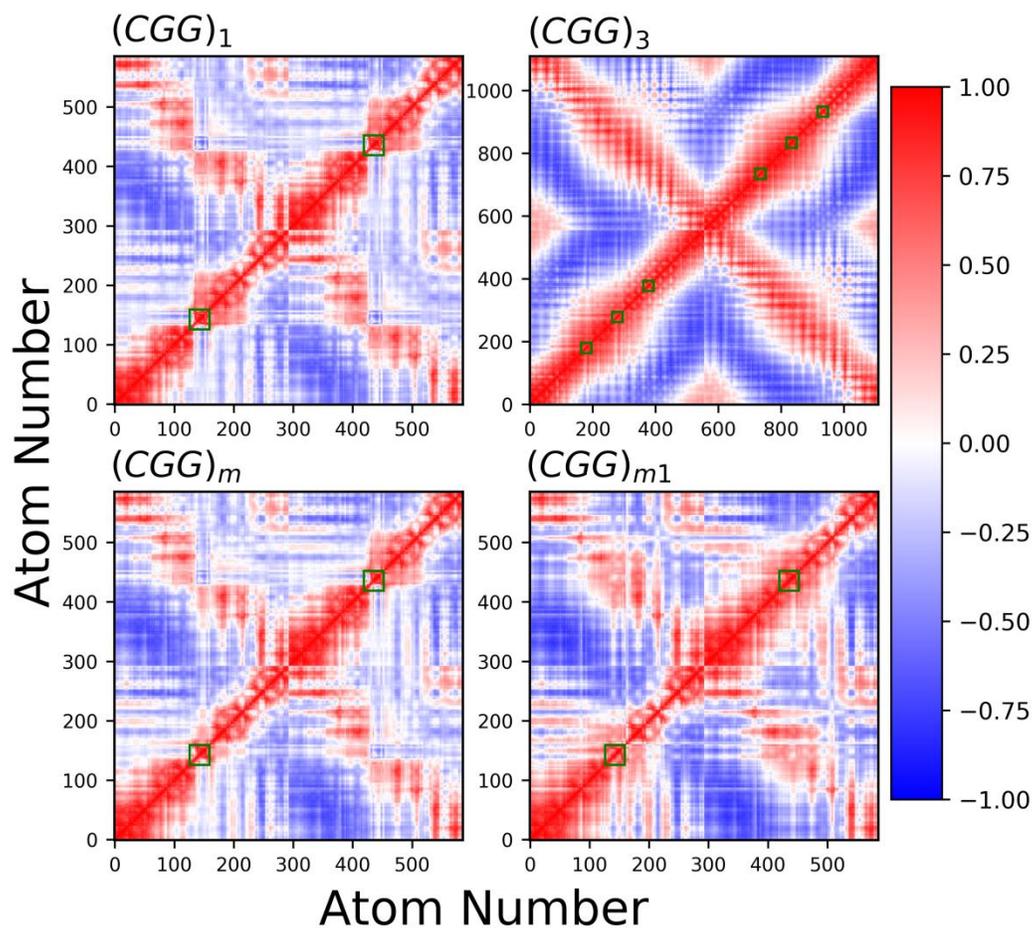


Figure 3.11 Dynamical cross-correlated map of RNA; where $(CGG)_1$ and $(CGG)_3$ are the NMR and crystal structure respectively having single and tri CGG repeat. $(CGG)_m$ and $(CGG)_{m1}$ are the modelled structure with anti-anti and anti-syn initial conformation at the mismatch G base pair. X and y-axis of this plot colour indicates the atom indices. Red square shows the regions corresponds to the atoms of mismatch base pairs.

3.2.3 Principle component analysis (PCA) revealed the transitional dynamics

We have generated a free energy map corresponding to the principal component 1 (PCA1) and 2 (PCA2) and is shown in Figure 3.12. It is evident from Figure 3.12 that there exists two slightly separable equiprobable minima structures of the NMR sequence and the model *anti-anti* sequence (Figure 3.12 A -B). On the contrary, only a single minimum was found for the modeled *syn-anti* conformation (Figure 3.12 C). For the (CGG)₃ repeat, we get a large conformational basin with a single broad free energy minimum (Figure 3.12 D). Further, we have generated the first mode of the motion in a vector representation with a certain cut-off to visualize only the prominent motions (Figure 3.13). In cases of (CGG)₁ and (CGG)₁m (*anti-anti*), along with flexible end residues, we observed an opposite motion of the two mismatch base pairs (Figure 3.13A-B). In the mismatch region, the motion of the two bases was completely anti-correlated. On the other hand, the *syn-anti* sequences, such as (CGG)₁m (*anti-syn*) and (CGG)₃ showed only fluctuations in their terminal base because of the two hydrogen bonds involves with *syn* mismatch base pairs (Figure 3.13C-D).

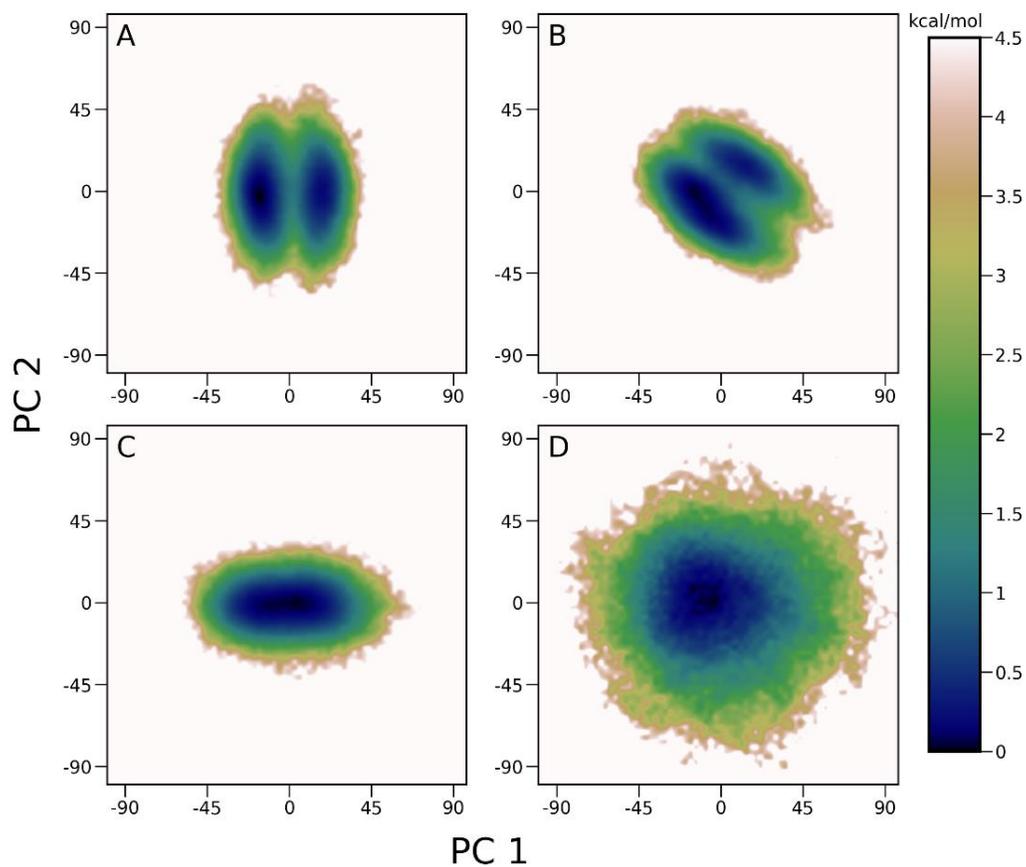


Figure 3.12 Free energy map corresponds to the principal components PC1 and PC2 for four different RNA systems. A) (CGG)1 (anti-anti), B) (CGG)1m (anti-anti), C) (CGG)1m (anti-syn) and D) (CGG)3. Two closely separated minima region observed in case of (CGG)1 and (CGG)1m (anti-anti) which also found from other analysis.

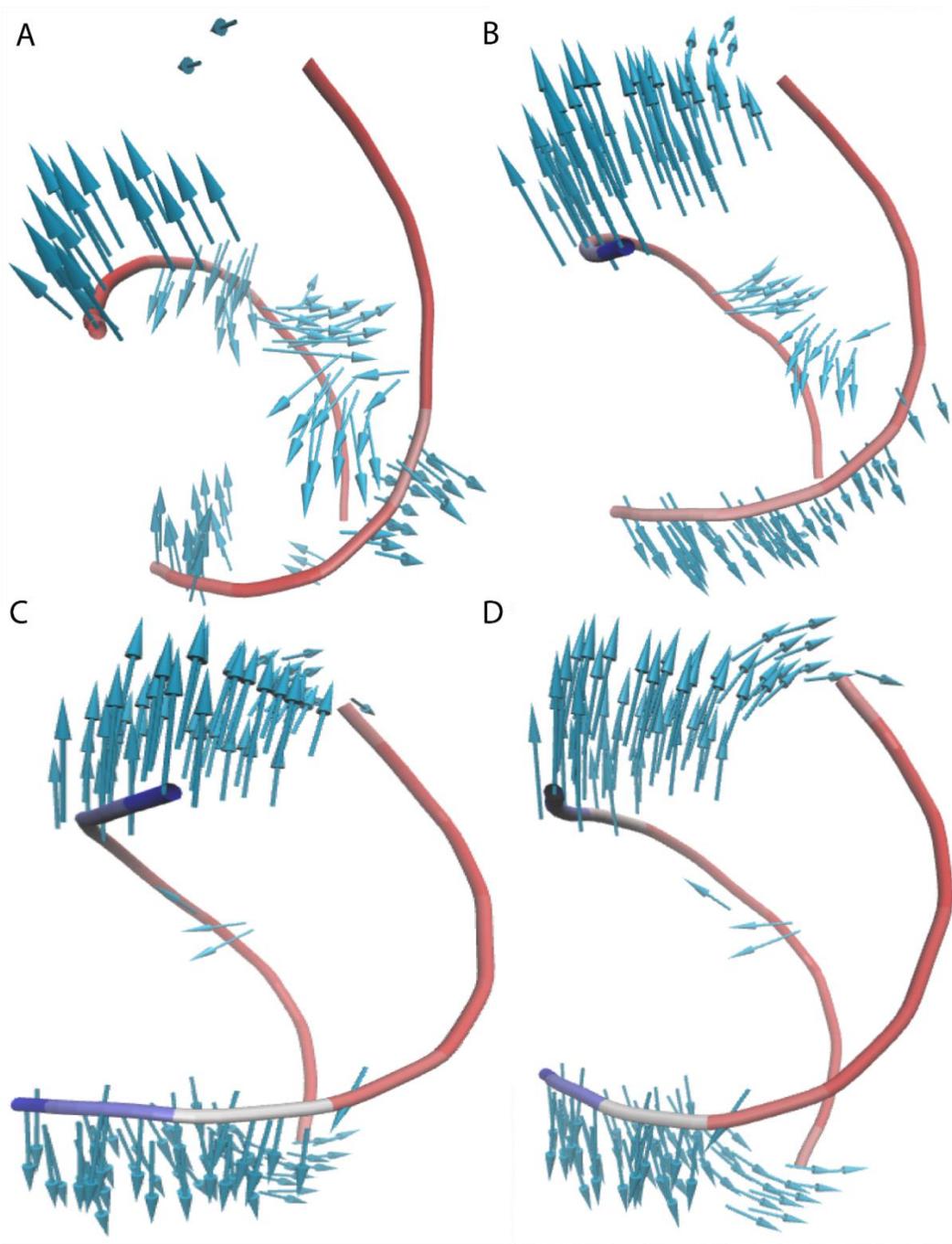


Figure 3.13 *Principal component analysis showing the predominant motions of the regions: A) (CGG)I(anti-anti), B) (CGG)Im (anti-anti), C) (CGG)Im (anti-syn), and D) (CGG)3.*

3.2.4 Umbrella sampling revealed the probable pathways for Anti-Anti \leftrightarrow Anti-Anti and Anti-Syn \leftrightarrow Anti-Syn GG transformation in (CGG)₁

An umbrella sampling molecular dynamics simulation was carried out to build the 2D potential mean force (PMF) surface to describe the possible pathways for *anti-anti* to *syn-anti* transformation. In the following figure (Figure 3.14) a 2D PMF was shown regarding two reactions coordinate base orientation concerning ribose sugar of G5 (χ) and pseudo dihedral angle (θ) in x and y-axis respectively. Here we have observed four distinct minimum regions (P1, P2, P3, and P4). In contrast, P1 (-125° , 40°), P4 (-125° , -140°) and P2 (55° , 40°), P3 (55° , -145°) both were separated by the high energy barrier. Overall, there are two minima corresponding to the *syn* conformation, and the other two minima correspond to the *anti*-structure. As the mismatch bases are entirely equivalent, a mirror symmetry across the diagonal in the free energy map is observed. This is in agreement with previous studies [33]. It can be seen from Figure 10 that both the *anti* and *syn* states are separated by a high energy barrier (~ 15 kcal/mol) which is merely impossible to overcome at room temperature. Therefore P1, P4, and P2, P3 are connected through a fewer energy barriers (2.5 kcal/mol to 5 kcal/mol) which may be a possible pathway for the *anti-anti* to *anti-anti* or *anti-syn* to *anti-syn* pathway. However, there are no physically possible pathways for the transition from *anti-anti* to *anti-syn*. This is because of the presence of two hydrogen bonds in *anti-syn* mismatch throughout our simulation length (Table 3.7).

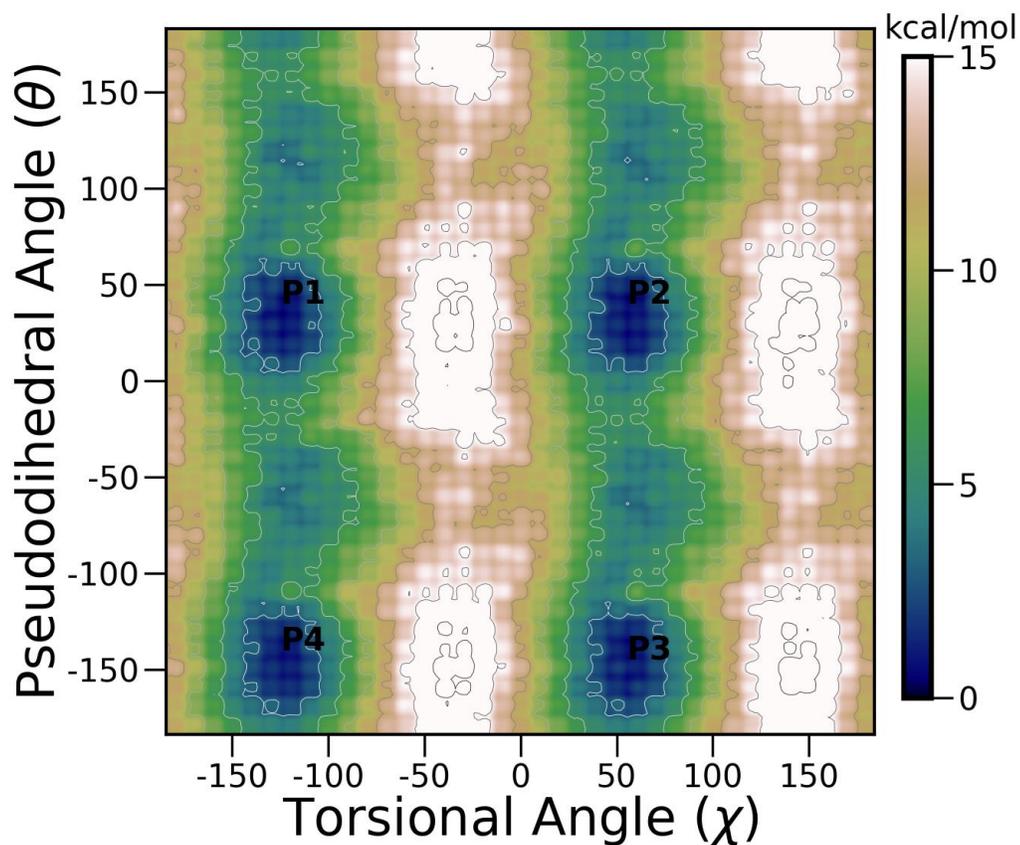


Figure 3.14 Two-dimensional potential of mean force (PMF) surface for (CGG)₁ system (PDB ID 2NCR). Here, x-axis represents torsional angle (χ) between base and sugar residues and y axis represents pseudodihedral angle (θ) for stacking. Four different minima structures are obtained through our umbrella sampling simulation denoted as P1, P2, P3, and P4.

3.2.5 Target CGG repeat motif RNA with small molecule

Targeting CGG repeats DNA/RNA *via* small molecule improved CGG associated defects such as *fmr1* gene silencing delay in FXS and RNA foci formation, alternative splicing defect and RANT in FXTAS [39-41]. Inhibition of DGCR8/DROSHA complex with CGG RNA interaction through small molecule could improve defect in FXTAS [40]. It has been reported that the small molecule preferentially binds to 1x1 GG internal loops. Henceforth, understating the conformational and thermodynamic properties of 1x1 GG mismatched motif could help to elucidate the underlying binding mechanism of a small molecule with G rich RNA internal loop. Further, by understanding this structure dynamics small molecule-based drug could be designed. Previously, naphthyridine dimer a small molecule reported for the specific binding with the GG loop. The results suggested that RNA conformation transformation was found during the naphthyridine binding process. NMR spectroscopy reveals that four naphthyridine molecules were stacked in the helix and form hydrogen bond with guanidine proton (H5' & H5'') [42]. Similarly, molecular modelling simulation study reported the binding of naphthyridine carbamate dimer cysteinylcystein (NCD-CC) with 5'CGG/3'GGC triad and formed intermolecular disulfide bond [43]. These previous results state that the final bound structure of small molecule is thermodynamically favorable and create stable structure because it increases the melting temperature of the GG loop.

3.2.6 Insights into CGG motif binding proteins

It is very well known that DGCR8 binds to CGG repeat RNA. Kwon and co-workers have solved the crystal structure of DGCR8 which contains non-canonical zinc finger motif [44]. This study showed that DGCR8 specifically interact with CGG repeats as compared to other TNRs such as CUG [45]. Studies have suggested that binding of DICER may cleave the CGG hairpin structure. Similar studies have been suggested that interaction of CGG binding protein change the conformational changes that melt the hairpin structure [46]. The presence of secondary structure in the 5'UTR may promote mRNA translation by

strong binding with translation factor or completely inhibit by binding with cellular proteins. According to our data, GG mismatched pair is stable most than other mismatched. Here in this study stable GG mismatched motif control the translation efficiency of amyloid precursor protein [47]. All these findings suggested that small molecules that stabilize the CGG repeat RNA structure could effectively alter the cellular function.

3.3 Material and Methods

3.3.1 RNA sample preparation

The desalted RNA sample containing 5'CGG/3'CGG RNA repeats motif was procured (sequence 5'r(CCGCGGCGG)₂ from Integrated DNA Technologies, Inc. The final concentration of the RNA duplex was calculated using its absorbance at 260 nm. Hyther Server was used to determine the molar extinction coefficients of RNA sample that is based on nearest neighbour thermodynamics [48,49]. Further RNA was lyophilised in the Christ Alpha 1-2 LD Freeze Dryer and dissolved in 1X phosphate buffer.

3.3.2 Nuclear Magnetic Resonance (NMR)

All NMR experiments were performed on Bruker NMR spectrometer with the field strength of 400 MHz. Double-stranded RNA sequence containing a single CGG motif, 5'r(CCGCG- GCGG)₂ was used for NMR spectroscopy. Lyophilised RNA sample was prepared by resuspending lyophilised RNA in 10 mM phosphate buffer, pH 7.2, 0.1 M NaCl, and 50 mM EDTA in 10% D₂O. Variable temperature proton spectra were acquired from 25 °C to 65 °C. Spectral assignments were performed using NOESY (Nuclear Overhauser Effect Spectroscopy). All the spectra were processed by the program TopSpin3.4.

Two-dimensional (2D) NOESY was performed acquiring 4096 data points with 64 K transients for each of the 296 FID signals. The spectra were recorded at 288, 298 and 308 K with mixing times of 300, 200 and 100 ms and 128 number of scans. A NOESY spectrum was recorded using an excitation exculpating pulse sequence to suppress residual water signals. ^1H - ^1H NOE distances were calculated using the SPARKY software.

3.3.3 Preparation of Model systems for Molecular Dynamics simulations

We have considered three sequences in the current study as shown in Figure 1. The structure of the RNA duplex, $(\text{CGG})_1 [(5'-\text{CCGCGGCGG}-3')_2]$ having a single mismatch was solved by NMR (PDB code: 2NCR) and employed for the MD simulation. The model system, $(\text{CGG})_{1\text{m}} [(5'-\text{GCGCGGCGC}-3')_2]$ having a single mismatch was prepared from $(\text{CGG})_1$ using the software, chimera [50]. It should be noted that the mismatch base pair of above-mentioned sequence was in *anti-anti* state for both cases.

Further, $(\text{CGG})_{1\text{m}}$ was also modified by changing the orientation of one of the mismatch base, i.e. G5 from *anti* to *syn* conformation creating a *syn-anti* conformation of the same systems. Thus $(\text{CGG})_{1\text{m}}$ sequence represent two different systems, $(\text{CGG})_{1\text{m}}$ (*anti-anti*) and $(\text{CGG})_{1\text{m}}$ (*syn-anti*). Finally, atomic coordinates of $(\text{CGG})_3$ or triple repeat of CGG [(5'-GGGCCGGCGGCGGGUCC-3')₂] were taken from the previously published X-ray crystallographic structure (PDB ID: 3SJ2) [17] where all the mismatch base pairs were in *anti-syn* conformations. All the systems were solvated in TIP3P [51] water molecules in a truncated octahedral box having 3624, 3853 and 8210 water molecules in $(\text{CGG})_1$, $(\text{CGG})_{1\text{m}}$ and $(\text{CGG})_3$, respectively. A suitable number of Na^+ ions were added to neutralise the solvated systems. Molecular dynamics simulations for all the solvated systems were carried out using the AMBER16 program [52]. The AMBER force field [53] with revised χ [54] and α/γ [55] torsional parameters was used to simulate RNAs. The same force field was also used in previous studies [22]. The long-range electrostatic interactions were treated using the

particle-mesh Ewald (PME) scheme [56]. The non-bonded cut-off was set to 8 Å, and the non-bonded pair list was updated every 25 fs. All the bonds involving hydrogen atoms were constrained using the SHAKE algorithm[57]. The minimisation was done in two steps: firstly with RNA held fixed with a restraint force of 500 kcal/mol Å² and then without restraint. The systems were minimized by 1000 steps of steepest descent followed by another 1500 steps of conjugate gradient algorithm. After the minimisation, the temperature was raised from 0 to 300 K by using a Langevin thermostat where the RNA molecule was held fixed with a constraint force of 10 kcal/mol Å². Constant pressure dynamics with isotropic position scaling was turned on in the next step where reference pressure was 1 atm with a pressure relaxation time of 2 ps and RNA was held fixed with 2 kcal/mol Å². After that 100 ps equilibration was done without restraint with all the above conditions. A total of 1 μs MD run was performed at 300 K with 1 fs time-step for all cases under NPT ensembles. Atomic coordinates were saved every 5 ps resulting 200000 configurations for each simulation.

3.3.4 Umbrella Sampling Molecular Dynamics Simulations

To determine the pathways for the *anti-syn* transformation, a two-dimensional potential of mean force (PMF) surface was constructed for the (CGG)₁ system from umbrella sampling molecular dynamics simulations. Since the *syn* and *anti*-base orientations were determined by the rotation of the glycosidic bond concerning sugar, the torsional angle χ was chosen as one of the reactions coordinates for the umbrella sampling simulation. This dihedral angle consists of O4, C1 atoms of sugar and N9, C8 atoms of base. The centre of mass pseudo dihedral angle (θ) was chosen as another reaction coordinate. Pseudo torsions are used to determine the base unstacking. Here we used four groups of atoms to define it. As for example, in our (CGG)₁ or (CGG)_{1m} cases four points are as follow; centre of mass of sugar atoms of 15, 4, 6 bases and the centre of mass of mismatch base atoms, i.e. G5 base. These reaction coordinates were used previously in different base flipping studies[22,58-60]. Both the coordinates were rotated with an increment of 10 degrees (°) creating a total of 1296 combinations.

All the combinations were minimised and equilibrated in two steps using a protocol previously used by Yildirim et al [22]. Throughout the simulation, χ and θ were held fixed in a particular combination using a square bottom well with a parabolic side. During the minimization and equilibration stages, a constraint force of 50 000 kcal/mol.rad² and 500 kcal/mol.rad² were employed, respectively. For all the simulations, the long-range cut off was always kept at 8 Å, and SHAKE [61] was used in all bonds involving hydrogen atoms. A 20 ps run was used to raise the temperature from 0 to 300 K using the Langevin dynamics with a collision frequency of 1 ps⁻¹. Constant pressure dynamics with isotropic positional scaling was turned on in the second stage of equilibration having a reference pressure of 1 atm with a pressure relaxation of 2 ps. All the production runs were conducted for 2 ns using a time-step of 2 fs. A potential with square bottom well and parabolic sides was used to restraint the reaction coordinates. For restraining χ and θ , force constants of 200 kcal/mol.rad² and 250 kcal/mol.rad² were used, respectively.

3.3.5 Analysis

All the analyses including root mean square deviation (RMSD) of backbone atoms and the mismatch base pairs, distance of ions from the mismatch basepair, distribution of glycosidic angle were performed using the Cpptraj [62] module of AMBER. The different parameters of the RNA structure were analysed with the 3DNA [37] software package (Tables, Supporting Information). The centre of mass pseudodihedral angle, which has been used to calculate the direction of base flipping [38] was calculated over the complete trajectory. Cross-correlation map also plays an essential role in the recognition of the biological macromolecules [63]. The correlated cross values range from positive to negative unity. The positive value represents the correlated motion or motion in the same direction whereas negative implies anticorrelated motion or motion in entirely opposite direction. In this study, covariance matrix for atomic fluctuations of all the RNA systems was constructed using Cpptraj [62] module. To extract the global motion using dimensionality reduction method, we use principal component analysis (PCA) in our RNA systems.

The covariance matrix obtained from the cross-correlation analysis was diagonalised to calculate orthogonal eigenvectors and the corresponding eigenvalues. The obtained eigenvectors are termed as a principal components (PC) and first few of them represents the slow and the global motion of the systems [64,65]. Here, PCA was performed using the same Cpptraj module [62]. For constructing 2D PMF (potential of mean force) surface, the weighted histogram analysis method (WHAM) was used [29,66].

3.4 Conclusion

Targeting the non-canonical structures formed by expanded repeats RNA has been proved as a promising therapeutic approach for the treatment of repeat-associated neurological diseases. Structural insights of these non-canonical structures become essential for designing small molecule modulators that can target these RNA motifs. An atomistic understating of this structure may rule out the new strategy for developing of such small molecules that can bind to target RNA at very low concentration. Therefore, here we have employed NMR, classical molecular dynamic simulation, umbrella sampling, and principal component analysis to reveal the dynamic behavior of internal mismatched GG pair in the RNA that contains a single repeat motif 5'CGG/3'GGC. NMR spectroscopy study showed the dynamic nature of internal GG pair. This type of special geometry and other electronic distribution of atoms in RNA molecule display the properties of the α -helical form of nucleic acids and provide the site of binding of cellular proteins and a small molecule. Molecular dynamics simulations study revealed the two orientation of non-canonical GG pair i.e. *anti-anti* and *syn-anti*. The *anti-anti* and *syn-anti* conformers possess one and two hydrogen bonds on an average respectively, that makes *anti-syn* structures less flexible. The nature of the flexibility of mismatch bases is also limited to their initial conformation range. Analysis of the 2D PMF also concluded that there are multiple conformers of 1x1 GG base pair that are interconnected as previous report suggested for 1x1 AA and UU base pair. However, there are no such pathways in between *anti* and *syn* states observed. The umbrella sampling

analysis revealed that both *syn-syn* and *anti-anti* conformation are nearly 15 kcal/mol more favorable than *syn-anti* and *anti-anti* conformation of CAG and CUG RNA respectively. This study also supported the experimental results of (CGG)₃ where no *syn-anti* GG transformation found. Thus, umbrella sampling MD simulation generated local minima conformation of 1x1 GG base pair is the most suitable target for small molecule binding.

Additionally, our results also suggested the *syn-anti* conformation of (CGG)_{1m} and (CGG)₃ stabilized by the Na⁺ binding pocket present near the GG internal loop similar to one detected in r(GGGGCC) RNA. Thus, designing of small molecule that contains positive charge group (NH₂ and NH) could utilize binding pocket of GG internal loop for higher affinity. The motion of the two-mismatch pair is completely anti-correlated, and the flipping of these bases are completely in the opposite direction. As a whole, the GG mismatch pair is quite stable as compared to CAG or CUG, which have also been found in other computational studies. This study implies the designing of novel therapeutically active compounds that recognise and bind with the repeats RNA that eventually leads to the inhibition of RNA-protein complexes interactions by blocking the protein binding site on RNA and improved the defects caused due to long repeats RNA.

3.5 References

1. Ellegren, H., Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics* **2004**, 5, 435-45.
2. Paulson, H., Repeat expansion diseases. *Handbook of clinical neurology* **2018**, 147, 105-123.
3. Orr, H. T.; Zoghbi, H. Y., Trinucleotide repeat disorders. *Annual review of neuroscience* **2007**, 30, 575-621.
4. Kovtun, I. V.; McMurray, C. T., Features of trinucleotide repeat instability in vivo. *Cell Research* **2008**, 18, 198.
5. Nalavade, R.; Griesche, N.; Ryan, D. P.; Hildebrand, S.; Krauss, S., Mechanisms of RNA-induced toxicity in CAG repeat disorders. *Cell death & disease* **2013**, 4, e752-e752.
6. Li, L. B.; Bonini, N. M., Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends in neurosciences* **2010**, 33, 292-8.
7. Hagerman, R. J.; Berry-Kravis, E.; Hazlett, H. C.; Bailey Jr, D. B.; Moine, H.; Kooy, R. F.; Tassone, F.; Gantois, I.; Sonenberg, N.; Mandel, J. L.; Hagerman, P. J., Fragile X syndrome. *Nature Reviews Disease Primers* **2017**, 3, 17065.
8. Hagerman, R. J.; Hagerman, P., Fragile X-associated tremor/ataxia syndrome — features, mechanisms and management. *Nature Reviews Neurology* **2016**, 12, 403.
9. Fink, D. A.; Nelson, L. M.; Pyeritz, R.; Johnson, J.; Sherman, S. L.; Cohen, Y.; Elizur, S. E., Fragile X Associated Primary Ovarian Insufficiency (FXPOI): Case Report and Literature Review. *Frontiers in genetics* **2018**, 9, 529.
10. Krzyzosiak, W. J.; Sobczak, K.; Wojciechowska, M.; Fiszer, A.; Mykowska, A.; Kozlowski, P., Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic acids research* **2012**, 40, 11-26.
11. Gatchel, J. R.; Zoghbi, H. Y., Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nature Reviews Genetics* **2005**, 6, 743.

12. Tawani, A.; Kumar, A., Structural Insights Reveal the Dynamics of the Repeating r(CAG) Transcript Found in Huntington's Disease (HD) and Spinocerebellar Ataxias (SCAs). *PLoS one* **2015**, 10, e0131788.
13. Chen, J. L.; VanEtten, D. M.; Fountain, M. A.; Yildirim, I.; Disney, M. D., Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1. *Biochemistry* **2017**, 56, 3463-3474.
14. Matthes, F.; Massari, S.; Bochicchio, A.; Schorpp, K.; Schilling, J.; Weber, S.; Offermann, N.; Desantis, J.; Wanker, E.; Carloni, P.; Hadian, K.; Tabarrini, O.; Rossetti, G.; Krauss, S., Reducing Mutant Huntingtin Protein Expression in Living Cells by a Newly Identified RNA CAG Binder. *ACS Chemical Neuroscience* **2018**, 9, 1399-1408.
15. Khan, E.; Tawani, A.; Mishra, S. K.; Verma, A. K.; Upadhyay, A.; Kumar, M.; Sandhir, R.; Mishra, A.; Kumar, A., Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs). **2018**, 13, 180-188.
16. Khan, E.; Biswas, S.; Mishra, S. K.; Mishra, R.; Samanta, S.; Mishra, A.; Tawani, A.; Kumar, A., Rationally designed small molecules targeting toxic CAG repeat RNA that causes Huntington's disease (HD) and spinocerebellar ataxia (SCAs). *Biochimie* **2019**, 163, 21-32.
17. Kumar, A.; Fang, P.; Park, H.; Guo, M.; Nettles, K. W.; Disney, M. D., A crystal structure of a model of the repeating r (CGG) transcript found in fragile X syndrome. *ChemBioChem* **2011**, 12, 2140-2142.
18. Al-Hashimi, H. M., NMR studies of nucleic acid dynamics. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **2013**, 237, 191-204.
19. Kim, H. J.; Howell, S. C.; Van Horn, W. D.; Jeon, Y. H.; Sanders, C. R., Recent Advances in the Application of Solution NMR Spectroscopy to Multi-Span Integral Membrane Proteins. *Progress in nuclear magnetic resonance spectroscopy* **2009**, 55, 335-360.
20. Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M., Exploring the energy landscape of a small RNA hairpin. *Journal of the American Chemical Society* **2006**, 128, 1523-1530.

21. Chen, S.-J., RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **2008**, *37*, 197-214.
22. Yildirim, I.; Park, H.; Disney, M. D.; Schatz, G. C., A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. *Journal of the American Chemical Society* **2013**, *135*, 3528-3538.
23. Yildirim, I.; Chakraborty, D.; Disney, M. D.; Wales, D. J.; Schatz, G. C.; computation, Computational investigation of RNA CUG repeats responsible for myotonic dystrophy 1. *Journal of chemical theory computation* **2015**, *11*, 4943-4958.
24. Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell Jr, A. D., Impact of 2'-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of computational chemistry* **2011**, *32*, 1929-1943.
25. HART, K.; NYSTRÖM, B.; ÖHMAN, M.; NILSSON, L., Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *Rna* **2005**, *11*, 609-618.
26. Curuksu, J.; Sponer, J.; Zacharias, M., Elbow flexibility of the kt38 RNA kink-turn motif investigated by free-energy molecular dynamics simulations. *Biophysical journal* **2009**, *97*, 2004-2013.
27. Allnér, O.; Nilsson, L.; Villa, A., Loop-loop interaction in adenine-sensing riboswitch: A molecular dynamics study. *RNA* **2013**.
28. Sun, Z.; Wang, X.; Zhang, J. Z., Protonation-dependent base flipping in the catalytic triad of a small RNA. *Chemical Physics Letters* **2017**, *684*, 239-244.
29. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* **1992**, *13*, 1011-1021.
30. Pan, F.; Man, V. H.; Roland, C.; Sagui, C., Structure and dynamics of DNA and RNA double helices of CAG and GAC trinucleotide repeats. *Biophysical journal* **2017**, *113*, 19-36.

31. Chen, J. L.; VanEtten, D. M.; Fountain, M. A.; Yildirim, I.; Disney, M. D., Structure and dynamics of RNA repeat expansions that cause Huntington's disease and Myotonic Dystrophy type 1. *Biochemistry* **2017**, 56, 3463-3474.
32. Bochicchio, A.; Rossetti, G.; Tabarrini, O.; Krauß, S.; Carloni, P., Molecular view of ligands specificity for CAG repeats in anti-Huntington therapy. *Journal of chemical theory computation* **2015**, 11, 4911-4922.
33. Pan, F.; Man, V. H.; Roland, C.; Sagui, C., Structure and Dynamics of DNA and RNA Double Helices Obtained from the CCG and GGC Trinucleotide Repeats. *The Journal of Physical Chemistry B* **2018**, 122, 4491-4512.
34. Zhang, Y.; Roland, C.; Sagui, C., Structure and dynamics of DNA and RNA double helices obtained from the GGGGCC and CCCC GG hexanucleotide repeats that are the hallmark of C9FTD/ALS diseases. *ACS chemical neuroscience* **2016**, 8, 578-591.
35. Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J., Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A·A, C·C, G·G, and T·T Mismatches. *Biochemistry* **1999**, 38, 3468-3477.
36. SantaLucia, J., A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences* **1998**, 95, 1460-1465.
37. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* **2004**, 25, 1605-1612.
38. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, 79, 926-935.
39. Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005**, 26, 1668-1688.

40. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, 117, 5179-5197.
41. Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H., Reparameterization of RNA χ torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *Journal of chemical theory and computation* **2010**, 6, 1520-1531.
42. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal* **2007**, 92, 3817-3829.
43. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, 98, 10089-10092.
44. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, 23, 327-341.
45. Banavali, N. K.; MacKerell Jr, A. D., Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *Journal of molecular biology* **2002**, 319, 141-160.
46. Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C., An improved reaction coordinate for nucleic acid base flipping studies. *Journal of chemical theory computation* **2009**, 5, 3105-3113.
47. Priyakumar, U. D.; MacKerell, A. D., Computational approaches for investigating base flipping in oligonucleotides. *Chemical reviews* **2006**, 106, 489-505.
48. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. J. o. C. P., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **1977**, 23, 327-341.

49. Roe, D. R.; Cheatham III, T. E., PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory computation* **2013**, 9, 3084-3095.
50. Lu, X. J.; Olson, W. K., 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic acids research* **2003**, 31, 5108-5121.
51. Huang, N.; MacKerell, A. D., Atomistic view of base flipping in DNA. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical Engineering Sciences* **2004**, 362, 1439-1460.
52. Kormos, B. L.; Baranger, A. M.; Beveridge, D. L., Do collective atomic fluctuations account for cooperative effects? Molecular dynamics studies of the U1A- RNA complex. *Journal of the American Chemical Society* **2006**, 128, 8992-8993.
53. Maisuradze, G. G.; Liwo, A.; Scheraga, H. A., Relation between free energy landscapes of proteins and dynamics. *Journal of chemical theory and computation* **2010**, 6, 583-595.
54. Chang, S.; He, H. Q.; Shen, L.; Wan, H., Understanding peptide competitive inhibition of botulinum neurotoxin a binding to SV2 protein via molecular dynamics simulations. *Biopolymers* **2015**, 103, 597-608.
55. Roux, B., The calculation of the potential of mean force using computer simulations. *Computer physics communications* **1995**, 91, 275-282.
56. Campagne, S.; Gervais, V.; Milon, A., Nuclear magnetic resonance analysis of protein-DNA interactions. *Journal of the Royal Society, Interface* **2011**, 8, 1065-1078.
57. Tawani, A.; Kumar, A., Structural Insight into the interaction of Flavonoids with Human Telomeric Sequence. *Scientific Reports* **2015**, 5, 17574.
58. Kumari, D.; Usdin, K., Sustained expression of FMR1 mRNA from reactivated fragile X syndrome alleles after treatment with small molecules that prevent trimethylation of H3K27. *Human molecular genetics* **2016**, 25, 3689-3698.

59. Yang, W. Y.; He, F.; Strack, R. L.; Oh, S. Y.; Frazer, M.; Jaffrey, S. R.; Todd, P. K.; Disney, M. D., Small Molecule Recognition and Tools to Study Modulation of r(CG_nG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome. *ACS Chem Biol* **2016**, 11, 2456-65.
60. Verma, A. K.; Khan, E.; Mishra, S. K.; Jain, N.; Kumar, A., Piperine Modulates Protein Mediated Toxicity in Fragile X-Associated Tremor/Ataxia Syndrome through Interacting Expanded CGG Repeat (r(CG_nG)exp) RNA. *ACS Chemical Neuroscience* **2019**.
61. Nomura, M.; Hagihara, S.; Goto, Y.; Nakatani, K.; Kojima, C., NMR structural analysis of the G.G mismatch DNA complexed with naphthyridine-dimer. *Nucleic acids symposium series (2004)* **2005**, 213-4.
62. Yamada, T.; Miki, S.; Ni, L.; Nakatani, K., CGG repeat DNA assisted dimerization of CGG/CGG binding molecule through intermolecular disulfide formation. *Chemical communications (Cambridge, England)* **2018**, 54, 13072-13075.
63. Kwon, S. C.; Nguyen, Tuan A.; Choi, Y.-G.; Jo, Myung H.; Hohng, S.; Kim, V. N.; Woo, J.-S., Structure of Human DROSHA. *Cell* **2016**, 164, 81-90.
64. Sellier, C.; Freyermuth, F.; Tabet, R.; Tran, T.; He, F.; Ruffenach, F.; Alunni, V.; Moine, H.; Thibault, C.; Page, A.; Tassone, F.; Willemsen, R.; Disney, M. D.; Hagerman, P. J.; Todd, P. K.; Charlet-Berguerand, N., Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell reports* **2013**, 3, 869-80.
65. Fukuda, H.; Katahira, M.; Tanaka, E.; Enokizono, Y.; Tsuchiya, N.; Higuchi, K.; Nagao, M.; Nakagama, H., Unfolding of higher DNA structures formed by the d(CG_nG) triplet repeat by UP1 protein. *Genes to cells : devoted to molecular & cellular mechanisms* **2005**, 10, 953-62.
66. Mbella, E. G.; Bertrand, S.; Huez, G.; Octave, J. N., A GG nucleotide sequence of the 3' untranslated region of amyloid precursor protein mRNA plays a key role in the regulation of translation and the binding of proteins. *Molecular and cellular biology* **2000**, 20, 4572-4579.

Chapter 4

Piperine modulates protein mediated toxicity in Fragile X-associated tremor/ataxia syndrome through interacting expanded (r(CGG)^{exp})RNA

4.1 Introduction

RNA plays a central role in diverse expression and functions of the biological processes [1,2]. Expansion of certain set of nucleotides in RNA causes formation of tandem repeat sequences that cause several neurological diseases such as Fragile X-associated tremor/ataxia syndrome (FXTAS), Huntington's disease (HD) and Myotonic Dystrophy (MD) etc [3,4]. The expansion of CGG nucleotides at 5' UTR region of Fragile X mental retardation 1 (*FMRI*) gene results in Fragile X-associated tremor/ataxia syndrome (FXTAS), Fragile X syndrome (FXS) and Fragile X-associated primary ovarian insufficiency (FXPOI)[5]. A healthy individual possess 5 to 55 sets of CGG repeats. However, the increased number from 55 repeats causes the premutation condition that results in FXTAS and FXPOI [6,7]. Further expansion of CGG repeats to ≥ 200 is classified as full mutation allele and results in Fragile X syndrome. The clinical feature of FXTAS includes multisystemic atrophy, intention tremor, Parkinsonism, cerebellar & gait ataxia, dysautonomia, autonomic dysfunction and dementia [8] [9].

The expanded CGG repeats RNA causes pathogenicity *via* two different mechanisms. First, the expanded repeats RNA folds in a GG 1x1 array of non-canonical hairpin structure that causes sequestration of cellular RNA binding proteins like DiGeorge syndrome critical region 8 (DGCR8),[10] an Src-associated in mitosis protein (Sam68)[11], heterogeneous nuclear

ribonucleoproteins (hnRNPs) A1/B2 and muscleblind-like 1 (MBNL1) protein etc [11], that lead to the formation of toxic RNA-foci. These proteins are key players in splicing regulation, therefore their sequestration to toxic RNA results in pre-mRNA splicing defects inside the cell [11]. Another mechanism of the disease includes the expanded repeats RNA that undergoes the repeat-associated non-AUG translation (RANT) producing toxic aggregates of homopolymeric proteins such as polyglycine (polyG) and polyalanine (polyAla) [12,13]. Hence, destabilising or stabilising these toxic RNAs could be a promising therapeutic approach for the FXTAS, FXPOI and FXS diseases. Indeed, antisense oligonucleotide-based tools have been designed to target these toxic RNAs, however, the lack of their optimal cellular penetrance, non-specific immunogenic responses, and intracellular trafficking make them a less efficient tool [14-17]. On the other hand, small molecules have been offered as a better tool to target these toxic RNA due to their pharmacological advantages [18-20]. We have recently reported the therapeutic potential of Myricetin and some small synthetic molecules with high binding affinity and selectivity to the r(CAG)^{exp} RNA that causes Huntington disease (HD) and spinocerebellar ataxia (SCA) [20] [21].

Piperine is a bioactive alkaloid obtained from black pepper (*Piper nigrum*) [22] that possess various pharmacological properties such as anticarcinogenic activity [23] hepatoprotection [24], antioxidant [24], nephroprotective [25], antiinflammatory [26], antiapoptotic [27] antidepressant [28] and neuroprotective [27,29,30]. Given the significant medicinal properties of Piperine on human health, we sought to evaluate the therapeutic potential of Piperine for FXTAS, FXPOI and FXS by assessing its ability to modulate the r(CGCG)^{exp} RNAs structure and function. To this, the affinity of the Piperine for r(CGCG)^{exp} RNA motifs was first determined using fluorescence binding and isothermal titration calorimetry (ITC) assay. Also, the interaction of Piperine with r(CGCG)^{exp} RNAs was evaluated using various biophysical methods such as Circular Dichroism (CD) spectroscopy, CD thermal denaturation assay, ¹H NMR spectroscopy, gel mobility shift assay and polymerase chain reaction (PCR) stop assay. Finally, the

therapeutic potential was extensively characterised by analysing reduction in protein aggregates formation in r(CGG)^{exp} expressing cellular models and FXTAS patient-derived cell lines. Piperine was also found to improve the cell viability of FXTAS cellular models. These findings suggest Piperine could be a potential molecule for the FXTAS.

4.2 Results & Discussion

4.2.1 Probing binding of Piperine to 1x1 RNA motifs and r(CGG)^{exp} RNAs

The discovery and designing of RNA based biomedicines have changed the facets of drug development since the last few decades. Expanded repeats RNA have emerged as suitable targets for therapeutic development in various neurological and neuromuscular diseases like FXTAS, FXPOI, HD and several SCAs. Herein, we have explored, Piperine, a naturally available small molecule as a potent compound targeting selectively the r(CGG)^{exp} RNA.

Given the suitability of r(CGG)^{exp} RNA as a promising drug target, first, we have evaluated the binding ability of Piperine to different r(CNG)^{exp} RNA motif. We have performed the fluorescence titration assay for Piperine against all possible combinations of 5'CNG/3'GNC RNA motifs where N denotes A, U, C and G bases (Figure 4.1). These trinucleotide repeats RNA motifs are enclosed by GC bases which allow the proper folding of RNA hairpin casket [31]. The gradual addition of RNA in Piperine solution depicts a gradual increase in the fluorescence intensity at emission maxima (496 nm) of Piperine that illustrates the formation of RNA-Piperine complex. The change in fluorescence intensity was plotted against the concentration of RNA and which gives the binding constant value for each RNA that elucidates the binding affinity of Piperine with each RNA.

Interestingly, Piperine showed higher binding affinity and selectivity to G-rich (5'CGG/3'GGC) RNA motifs compared to other mismatch RNA motifs and 5'CAG/3'GUC (AU paired) RNA that forms a normal duplex structure and served

as control (Figure 4.2 a & d, Figure 4.3 and Table 4.1). The variable binding behavior of Piperine with different RNA motifs could be speculated to arise from the distinct binding affinity of the different repeat sequences as well as due to the topological differences between these RNAs [32]. However, the high affinity and selectivity of Piperine with (5'CGG/3'GGC) RNA motifs over other RNA compelled to further exploration of binding behavior with (5'CGG/3'GGC) RNA with large number of CGG repeats RNA including CGG_{x6}, CGG_{x20}, CGG_{x40} and CGG_{x60} repeats RNA. Fluorescence titration assay two-mode curve fitting of r(CG_nG)^{exp} RNAs and different RNA & DNA controls with Piperine.

As anticipated, the high affinity with these large repeats G-rich motif RNA was observed. Moreover, Piperine also showed the selectivity for CGG RNA over other fully base-paired AU_{x6} RNA and yeast t-RNA as Piperine showed ~315 and ~334 fold higher affinity to the r(CG_nG)₆₀ repeat RNA compared to AU_{x6} paired RNA and yeast t-RNA respectively (Figure 4.2 c & e Figure 4.4 and Table 4.2). Previous studies reported that r(CAG)^{exp}, r(CCG)^{exp} and r(CUG)^{exp} produces toxic homopolymeric protein aggregates via RAN translation. Similar to r(CG_nG)^{exp}, r(CAG)^{exp}, r(CCG)^{exp} and r(CUG)^{exp} also form AA, CC and UU 1x1 internal hairpin structure respectively [33]. Binding studies reveal that Piperine binds more tightly with CGG repeats RNA than other mismatch RNA motifs (Figure 4.2 c & e Figure 4.4 and Table 4.2). Furthermore, to assess the applicability of Piperine based therapeutics the binding affinity of Piperine was also evaluated with different other G-quadruplex forming DNA, c-Myc & Bcl-2 and duplex calf thymus (CT) DNA present in the cells. Although Piperine also binds with these sequences, the affinity of Piperine was ~286, ~815, ~2504-fold lesser with c-Myc, Bcl-2 and duplex CT DNA respectively (Figure 4.2 c & e Figure 4.4 and Table 4.2). TMPyP4 & TMPyP2 are nucleic acids binder [34,35] and therefore we assessed the binding affinity of TMPyP4 and TMPyP2 with CGG repeats RNA to further affirming the selectivity of Piperine with CGG repeats RNA. Interestingly, Piperine showed ~10-fold excess binding affinity over TMPyP4 however, no significant binding was observed with TMPyP2

(Figures 4.5 - 4.8 and Table 4.3 & 4.4). Conclusively, Piperine showed the high affinity (nanomolar) and selectivity with CGG repeats RNA as compared with other mismatch RNA and G-quadruplex DNA.

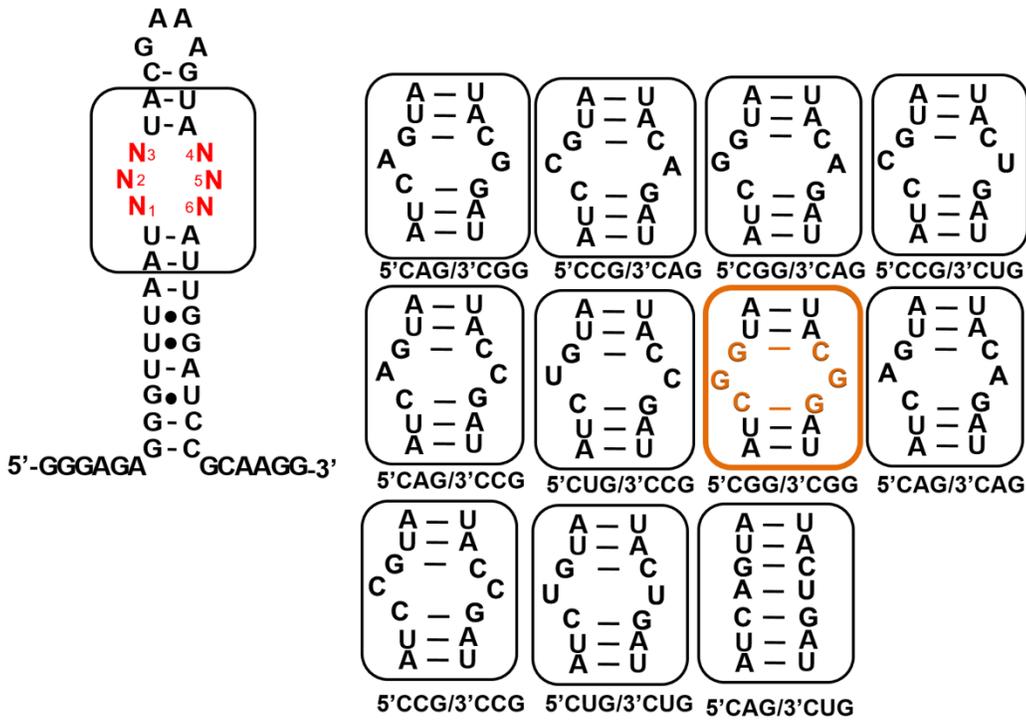


Figure 4.1 Secondary structure of different 1x1 nucleotide internal motif RNAs sequence.

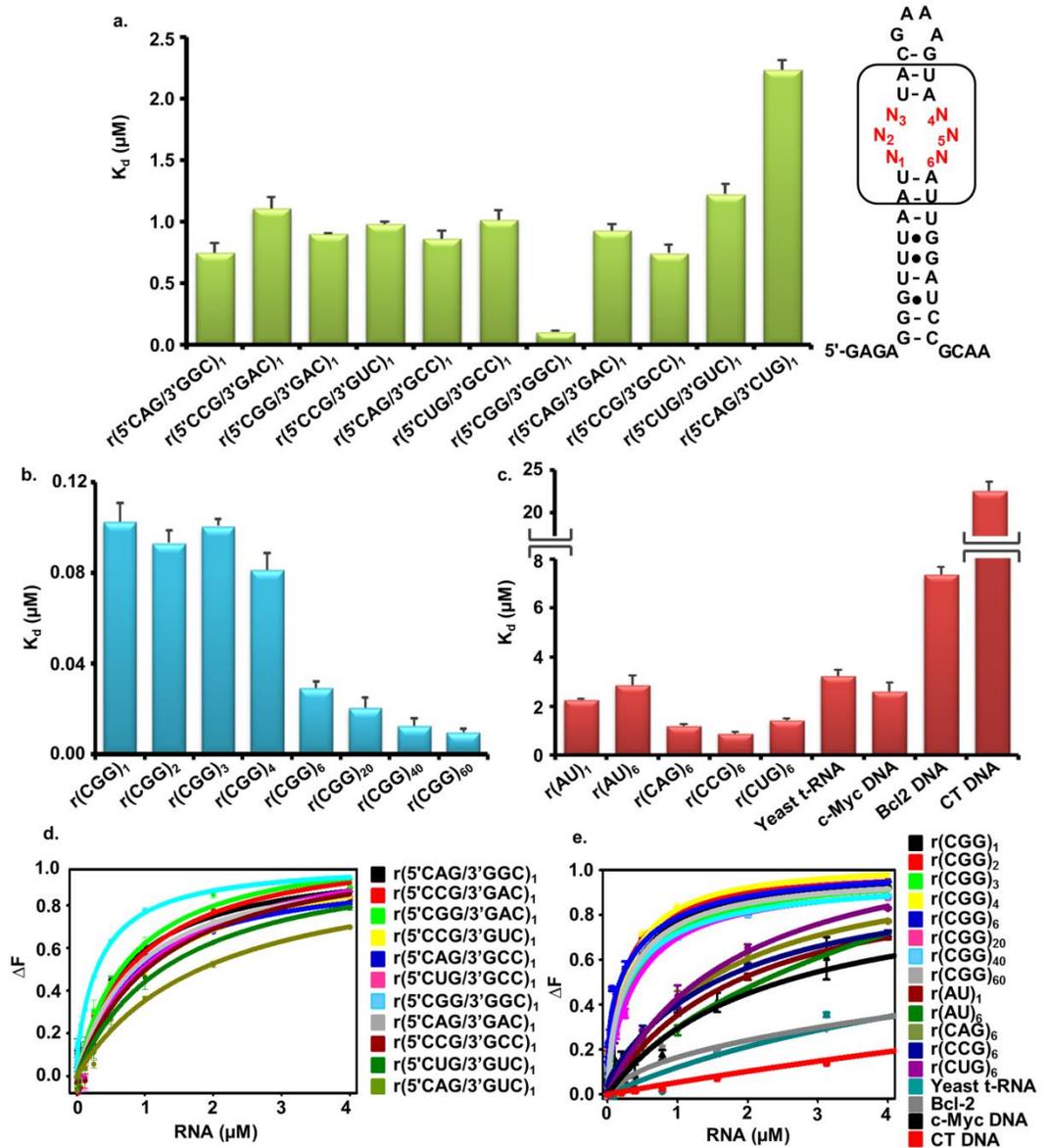


Figure 4.2 Evaluation of binding affinity of Piperine with nucleic acids. *a*) Bar graph denotes K_d values of r(CNG)x1 RNA motifs with Piperine. *b*) Systemic representation of bar graph for K_d values analysis of different CGG repeat RNAs with Piperine. *c*) Bar graph represents K_d value analysis of different RNA and DNA controls with Piperine. *d*) The plot represents fluorescence titration assays two-mode curve fitting of r(CNG)x1 RNA motifs with Piperine. *e*) Plot depict fluorescence titration assay two-mode curve fitting of r(CG)x1 RNAs and different RNA & DNA controls with Piperine.

Table 4.1 The dissociation constant (K_d) values obtained from fluorescence titration assay of Pipeirne with different (5'CNG/3'CNG) $_x$ 1 RNA motifs with two mode curve fitting.

S. No.	RNA	K_d^1 (μ M)	K_d^2 (μ M)
1.	(5'CAG/3'GGC) $_x$ 1	0.66 \pm 0.078	8.5\pm 1.4
2.	(5'CCG/3'GAC) $_x$ 1	0.50 \pm 0.089	2.0\pm 0.12
3.	(5'CGG/3'GAC) $_x$ 1	0.40 \pm 0.007	13.3\pm 1.9
4.	(5'CCG/3'GUC) $_x$ 1	0.81 \pm 0.017	2.5\pm 0.54
5.	(5'CAG/3'GCC) $_x$ 1	0.45 \pm 0.060	30.7\pm 3.65
6.	(5'CUG/3'GCC) $_x$ 1	0.69 \pm 0.079	1.5\pm 0.69
7.	(5'CGG/3'GGC) $_x$ 1	0.102 \pm 0.008	0.40\pm 0.14
8.	(5'CAG/3'GAC) $_x$ 1	0.47 \pm 0.052	1.6\pm 0.76
9.	(5'CCG/3'GCC) $_x$ 1	0.63 \pm 0.070	34.2\pm2.3
10.	(5'CUG/3'GUC) $_x$ 1	0.58 \pm 0.082	1.9\pm 0.37
11.	(5'CAG/3'CUG) $_x$ 1	2.23 \pm 0.077	1.94\pm 0.98

Table 4.2 The dissociation constant (K_d) values obtained from fluorescence titration assay of Piperine with r(CGG)_{exp}, r(CNG)₆, and AU paired RNAs and DNA controls with two mode curve fitting.

S No.	DNA/RNA	K_d^1 (μ M)	K_d^2 (μ M)
1.	r(CGG) ₁	0.102 ± 0.0082	0.41 ± 0.10
2.	r(CGG) ₂	0.093 ± 0.0056	0.45 ± 0.094
3.	r(CGG) ₃	0.10 ± 0.0031	0.47 ± 0.14
4.	r(CGG) ₄	0.081 ± 0.0071	2.68 ± 0.096
5.	r(CGG) ₆	0.029 ± 0.0027	0.82 ± 0.028
6.	r(CGG) ₂₀	0.020 ± 0.0045	0.51 ± 0.028
7.	r(CGG) ₄₀	0.012 ± 0.0031	0.74 ± 0.038
8.	r(CGG) ₆₀	0.009 ± 0.0012	0.33 ± 0.045
9.	r(AU) ₁	2.23 ± 0.077	2.23 ± 0.16
10.	r(AU) ₆	2.84 ± 0.39	2.84 ± 0.17
11.	r(CAG) ₆	1.16 ± 0.095	8.74 ± 0.73
12.	r(CCG) ₆	0.86 ± 0.070	1.86 ± 0.75
13.	r(CUG) ₆	1.41 ± 0.075	6.78 ± 0.55
14.	Yeast t-RNA	3.21 ± 0.27	9.88 ± 0.81
15.	Cymc DNA	2.58 ± 0.39	158.98 ± 6.0
16.	Bcl-2 DNA	7.34 ± 0.87	6.67 ± 0.11
17.	CT DNA	22.54 ± 1.06	23.04 ± 1.5

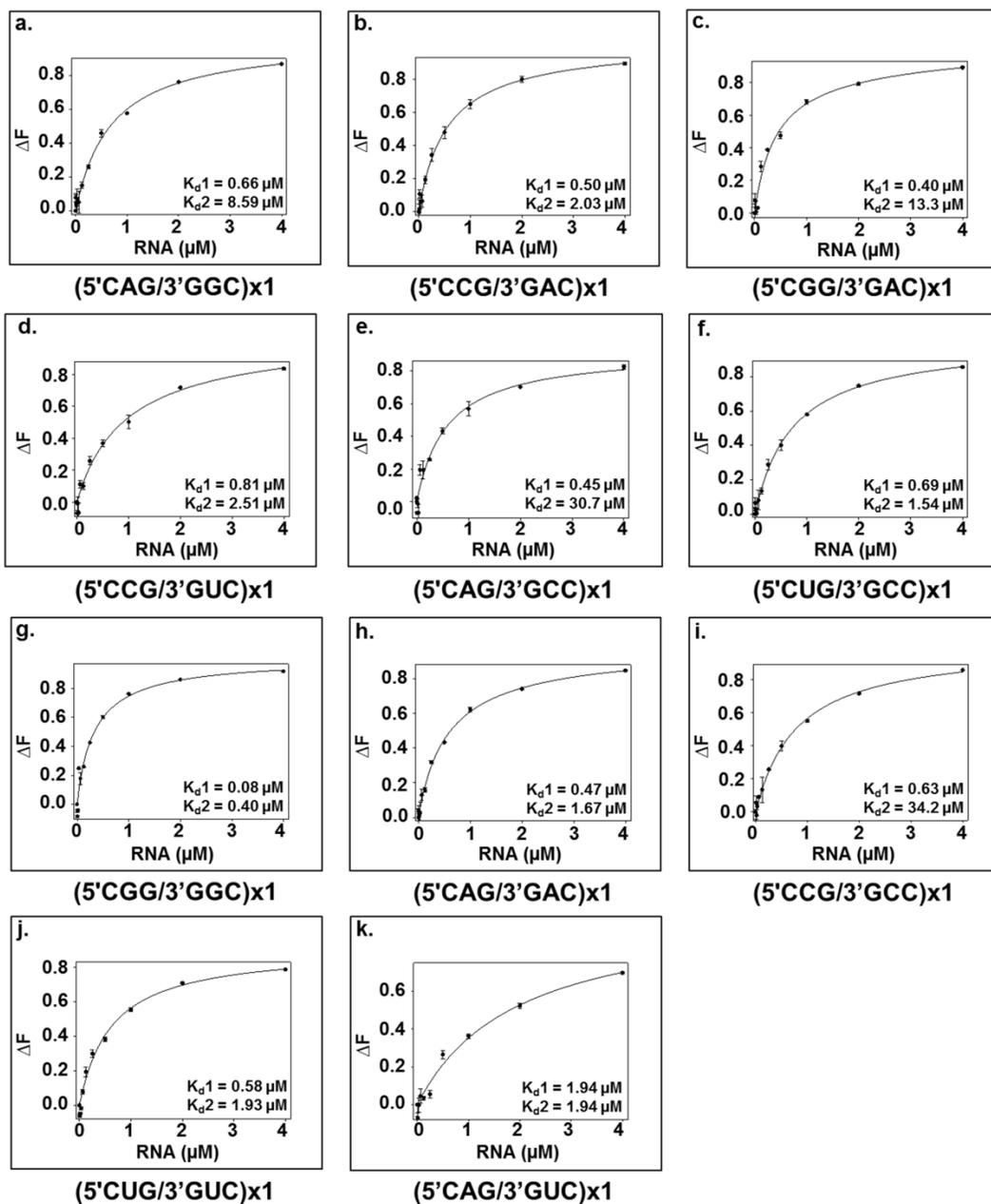


Figure 4.3 Diagram shows the fluorescence titration assay plots of different RNAs motif with Piperine. Solid line curve represents the two-mode fitting. The K_{d1} and K_{d2} are also mentioned below.

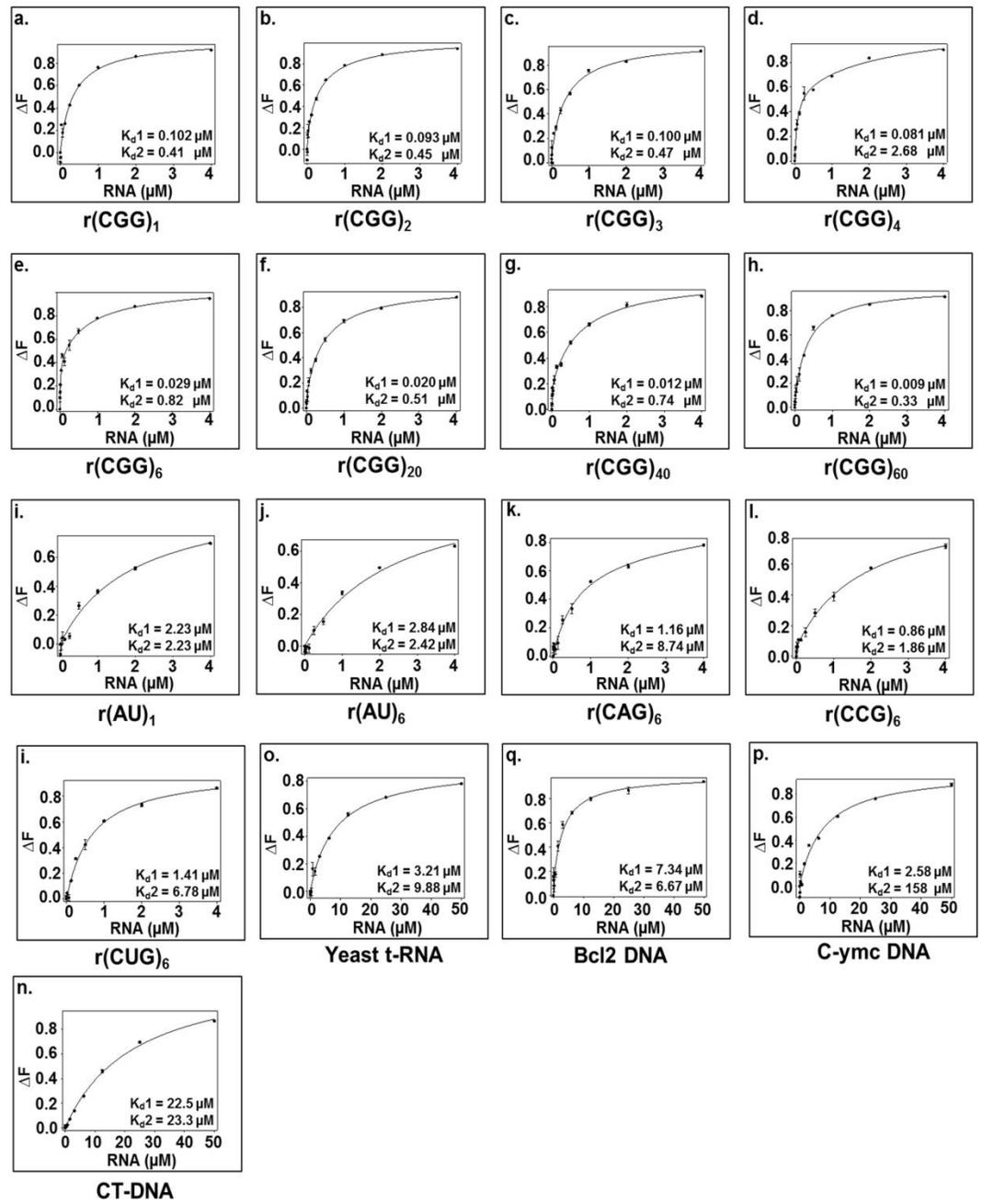


Figure 4.4 Diagram shows the fluorescence titration assay plots of r(CGG)^{exp}, (CNG)^{exp}, AU paired RNA and DNA with Piperine. Solid line represents the two-mode fitting. The K_{d1} and K_{d2} are also mentioned below.

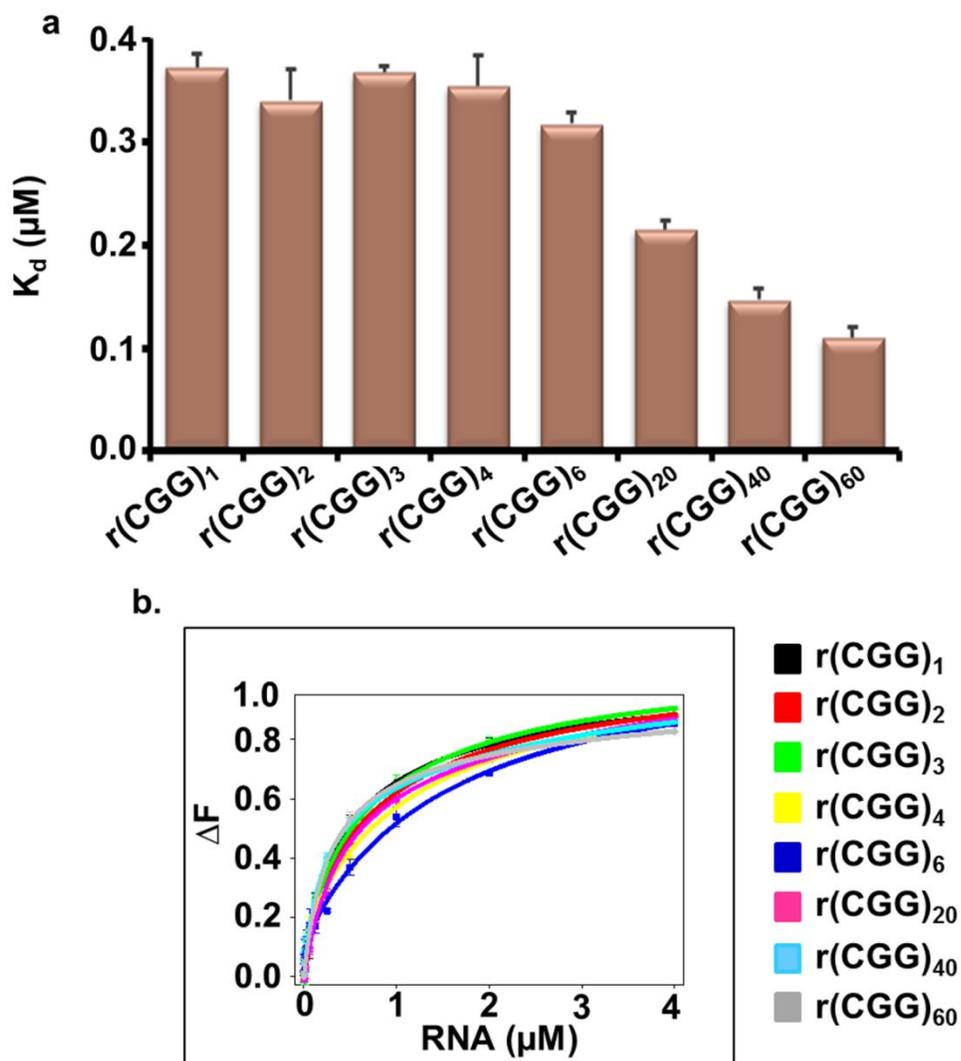


Figure 4.5 Fluorescence titration assay of TMPyP4 with expanded CGG RNA. a) Bar graph represents the dissociation constant values of TMPyP4 with r(CGGe)_{exp} RNA. b) Binding plot of CGG RNA with TMPyP4.

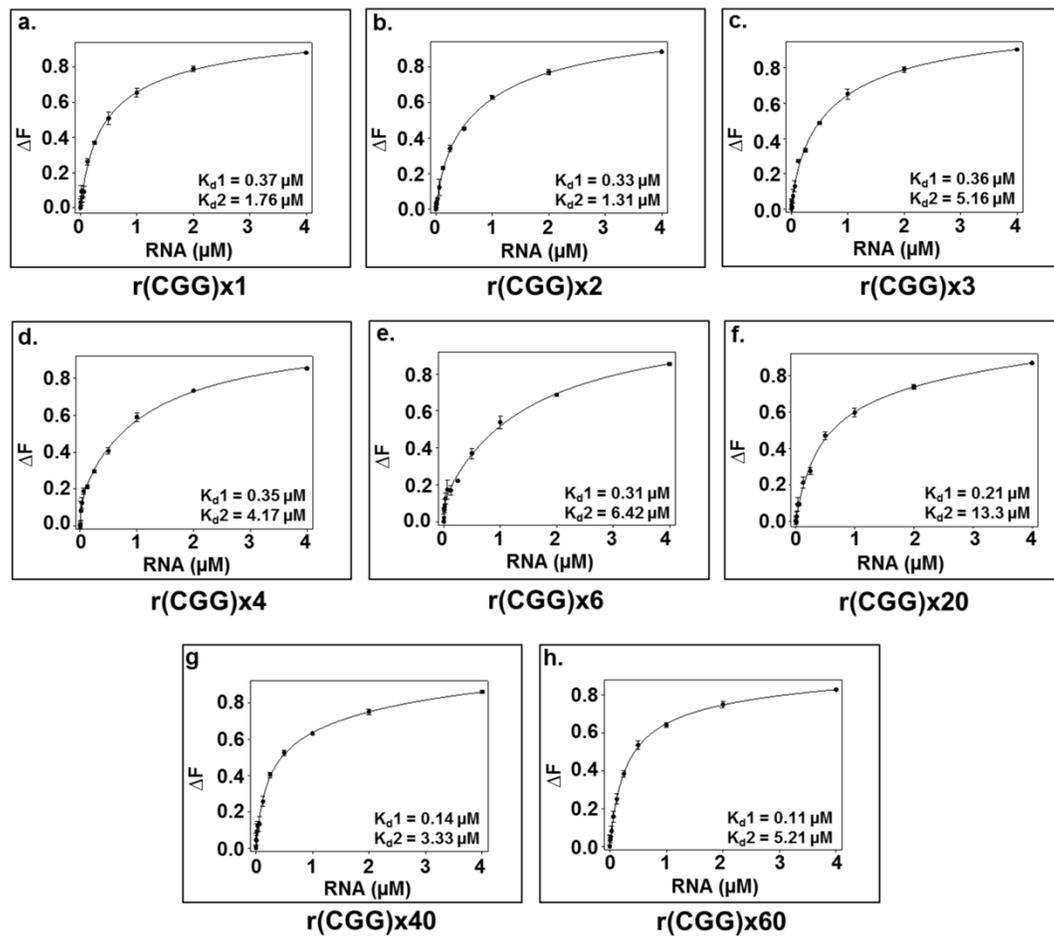


Figure 4.6 Plot showing the binding curve of TMPyP4 with expanded CGG RNA. Dissociation constant values were determined using two mode fitting.

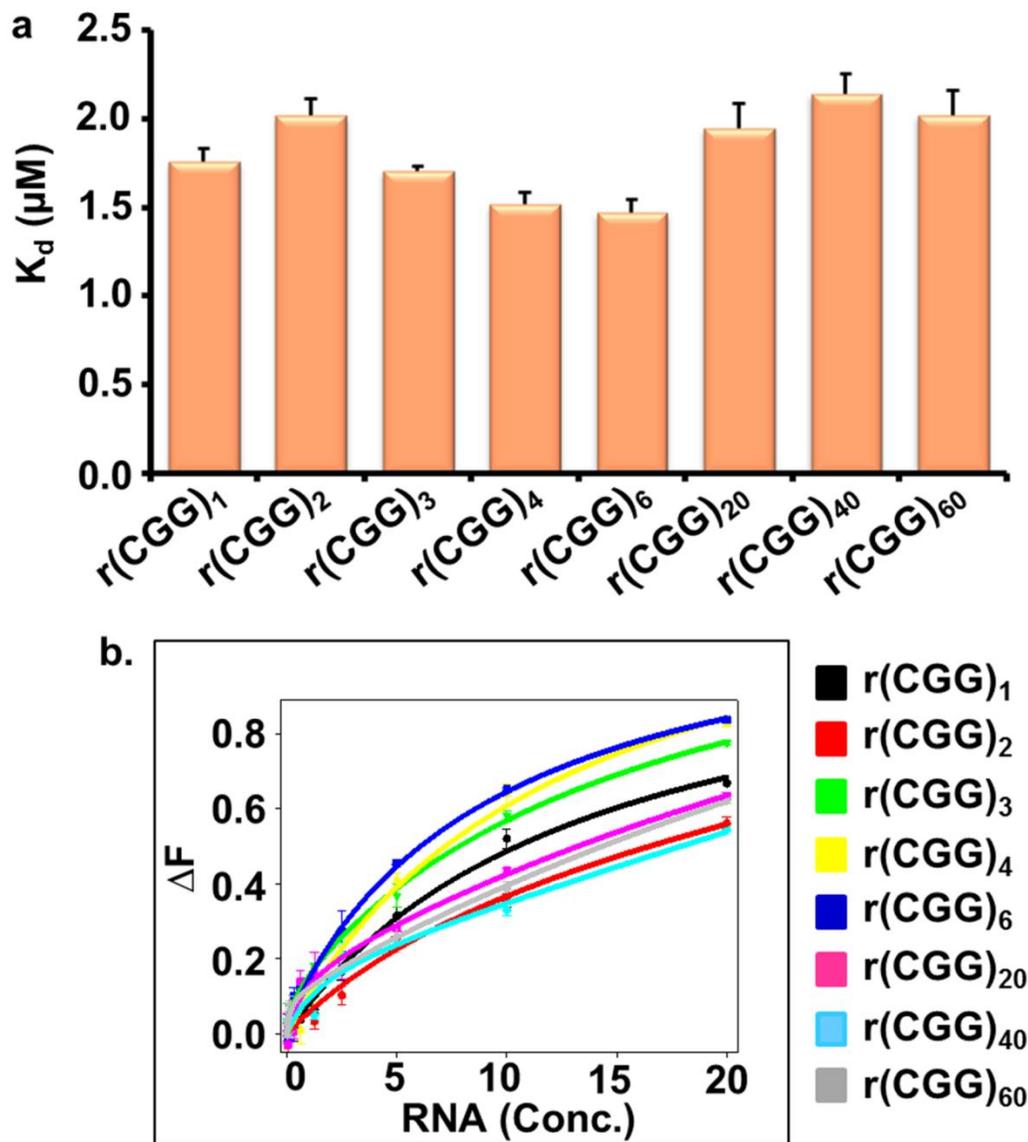


Figure 4.7 Fluorescence titration assay of TMPyP2 with expanded CGG RNA. **a)** Bar graph represents the dissociation constant of TMPyP2 with r(CGGe)_{exp} RNA. **b)** Binding plot of CGG RNA with TMPyP2.

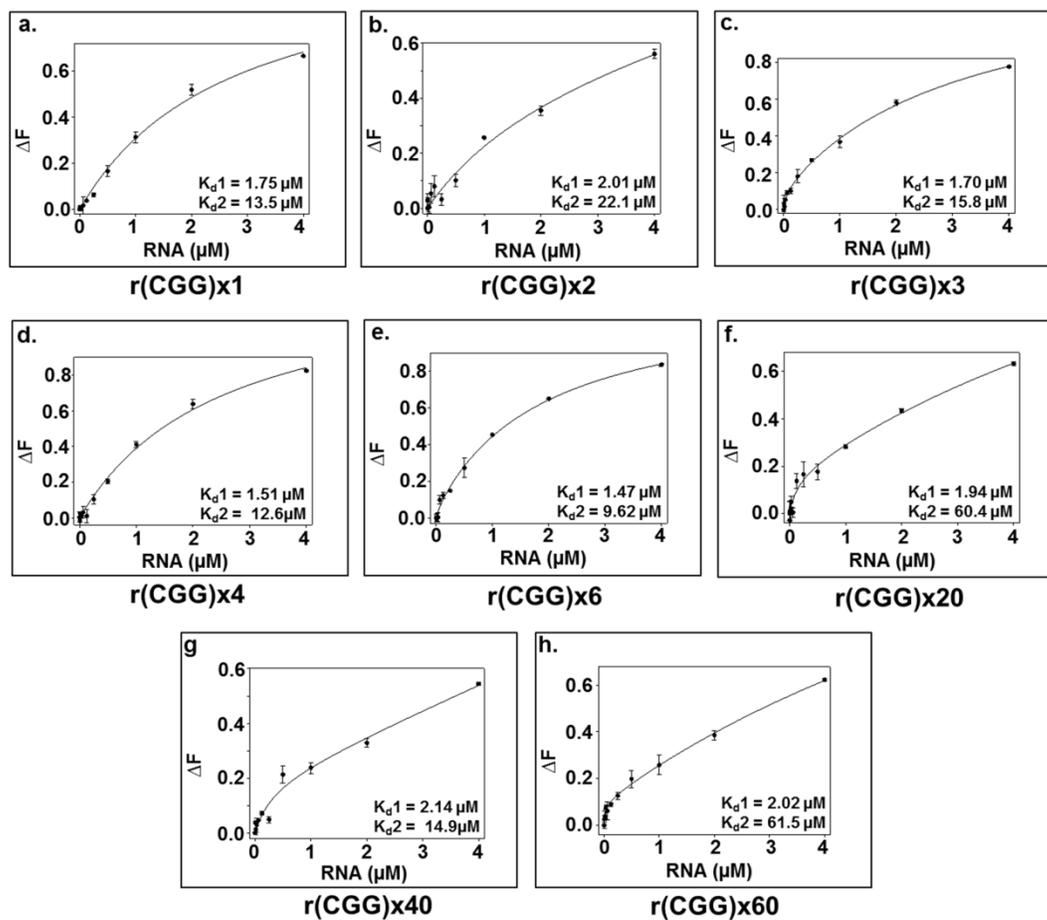


Figure 4.8 Plot showing the binding curve of TMPyP2 with expanded CGG RNA. Dissociation constant values were determined using two mode fitting.

Table 4.3 The dissociation constant (K_d) values obtained from fluorescence titration assay of TMPyP4 with r(CGG)_{exp} RNAs. The curves were fitted using two mode binding.

S No.	RNA	K_{d1} (μ M)	K_{d2} (μ M)
1.	r(CGG) ₁	0.37 ± 0.012	1.76 ± 0.13
2.	r(CGG) ₂	0.33 ± 0.031	1.31 ± 0.19
3.	r(CGG) ₃	0.36 ± 0.006	5.16 ± 0.09
4.	r(CGG) ₄	0.35 ± 0.029	4.17 ± 0.19
5.	r(CGG) ₆	0.31 ± 0.011	6.42 ± 0.11
6.	r(CGG) ₂₀	0.21 ± 0.007	13.3 ± 0.65
7.	r(CGG) ₄₀	0.14 ± 0.011	3.33 ± 0.42
8.	r(CGG) ₆₀	0.11 ± 0.01	5.2 ± 0.45

Table 4.4 The dissociation constant (K_d , μ M) values obtained after titrating r(CGG)_{exp} RNAs and TMPyP2 using fluorescence based binding assay. Two mode binding method were used for curve fitting.

S No.	RNA	K_{d1} (μ M)	K_{d2} (μ M)
1.	r(CGG) ₁	1.75 ± 0.075	13.5 ± 0.72
2.	r(CGG) ₂	2.01 ± 0.090	18.2 ± 0.66
3.	r(CGG) ₃	1.70 ± 0.023	13.7 ± 1.3
4.	r(CGG) ₄	1.51 ± 0.065	12.6 ± 0.25
5.	r(CGG) ₆	1.47 ± 0.072	9.6 ± 0.75
6.	r(CGG) ₂₀	1.94 ± 0.136	60.5 ± 3.8
7.	r(CGG) ₄₀	2.14 ± 0.111	24.2 ± 2.6
8.	r(CGG) ₆₀	2.02 ± 0.137	61.5 ± 4.3

4.2.2 Isothermal titration calorimetry (ITC) studies for the interaction of Piperine with r(CGG)_n RNAs

ITC assays help in understanding the thermodynamics of interactions between two molecules such as protein-ligand, DNA-protein interactions, ligand-nucleic acid interactions etc [36,37]. For instance, the negative change in the enthalpy (ΔH) indicating energetically favourable interaction between two molecules [38]. Herein, automated addition of Piperine to RNA solution showed the continuous release of heat that reached saturation at optimum concentration of Piperine. Interestingly, the enthalpy changes (ΔH_1) for CGG₆, CGG₂₀, CGG₄₀ and CGG₆₀ repeat RNAs were observed as $-2.37E4$ cal/mol, $-3.08E5$ cal/mol, $-2.02E6$ cal/mol and $-4.47E6$ cal/mol respectively which affirms the favourable contribution of energy during the interactions (Figure 4.9 a-d and Table 4.5). Similarly, the high association constant (K_a) values $8.41 \times 10^6 M^{-1}$, $3.21 \times 10^7 M^{-1}$, $9.48 \times 10^7 M^{-1}$ and $1.92 \times 10^8 M^{-1}$ with r(CGG)₆, r(CGG)₂₀, r(CGG)₄₀ and r(CGG)₆₀ RNAs respectively also corroborates the high affinity of Piperine with CGG repeats RNA (Figure 4.9 a-d and Table 4.5). The higher CGG repeats RNA consist of more binding sites which provides a high saturation point at the same concentration of ligand. Interestingly, Piperine showed ~ 20 , ~ 72 , ~ 213 and ~ 432 -fold higher binding with r(CGG)₆, r(CGG)₂₀, r(CGG)₄₀, and r(CGG)₆₀ repeat RNAs than AU paired RNA respectively (Figure 4.9 a-e, Table 4.5). Additionally, Piperine also exhibited ~ 100 -fold higher affinity with CGG repeat RNA as compared with other mismatch RNAs that also corroborates the selectivity of Piperine for CGG RNA (Figure 4.9 f-h, Table 4.5). In addition to that, the ITC of CGG repeat RNAs with TMPyP4 & TMPyP2 also demonstrated that Piperine has high affinity with CGG repeat RNAs as compared to TMPyP4, TMPyP2 (Figure 4.10 and Table 4.7 & 4.8). In conclusion, our ITC results were in-line with the fluorescence titration assay which showed that Piperine has a very high affinity (nanomolar) and selectivity with CGG repeat RNAs as compared with other mismatch RNA. Moreover, the exothermic peaks inferences the pattern of

interaction of Piperine with CGG repeat as it could be due to the π - π stacking of the aromatic ring structure of Piperine with r(CGG)^{exp} RNA base [39].

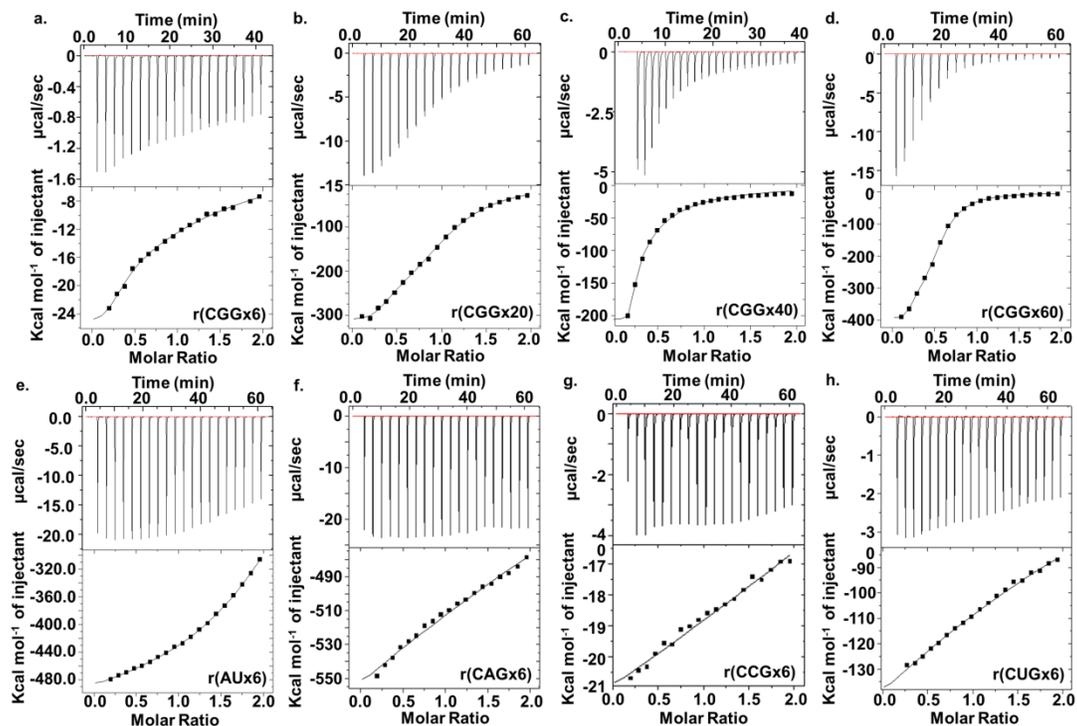


Figure 4.9 Isothermal calorimetry titrations of Piperine with $r(\underline{CNG})^{exp}$ and $r(AU)^{exp}$ duplex RNA. (a) $r(CGG)_6$, (b) $r(CGG)_{20}$, (c) $r(CGG)_{40}$, (d) $r(CGG)_{60}$, (e) $r(AU)_6$, (f) $r(CAG)_6$, (g) $r(CCG)_6$ and (h) $r(CUG)_6$ RNA represents the titrated thermogram with Piperine. Solid line curve represents the two-mode binding best fit. ITC data showed significant interaction of Piperine with expanded CGG RNA over other RNA motifs.

Table 4.5 Thermodynamic parameter values of r(CGG)exp, r(CNG)exp and r(AU)6 duplex RNAs with Piperine obtained from ITC using two mode curve fitting.

S. No.	Parameters	RNAs			
		r(CGG) ₆	r(CGG) ₂₀	r(CGG) ₄₀	r(CGG) ₆₀
1.	N1 (sites)	0.233 ± 0.048	0.256 ± 0.054	0.191 ± 0.064	0.472 ± 0.064
2.	K1 (M ⁻¹)	8.41E6 ± 1.8	3.21E7 ± 1.1	8.654E7 ± 1.02	1.92E8 ± 0.56
3.	ΔH1 cal/mol	-2.370E4 ± -0.14	-3.080E5 ± 1.3	-2.02E6 ± 1.7E6	-4.47E6 ± 1.48
4.	ΔS1 cal/mol/deg	-47.7 ± 5.7	-998 ± 471	-3.5E3 ± 3.5	-1.46E3 ± 1.8
5.	N2 (sites)	1.10 ± 0.35	0.738 ± 0.14	1.73E-4 ± 0.219	51.1 ± 0.209
6.	K2 (M ⁻¹)	1.91E4 ± 0.79	2.87E5 ± 1.4	8.95E4 ± 4.4	2.16E5 ± 1.3
7.	ΔH2 (cal/mol)	-4.02E4 ± 1.7	-3.35E5 ± 2.85E5	-5.078E8 ± 5.8E7	-4.0E5 ± 2.3
8.	ΔS2 (cal/mol/deg)	-307 ± 206	-1.10E3 ± 535	-1.70E6 ± 1.0E5	-1.32E3 ± 1.4
		r(AU) ₆	r(CAG) ₆	r(CCG) ₆	r(CUG) ₆
1.	N1 (sites)	0.759 ± 0.047	0.853 ± 0.06	2.64 ± 0.91	0.49 ± 0.37
2.	K1 (M ⁻¹)	4.44E5 ± 2.1	3.39E4 ± 2.5	2.75E4 ± 1.5	5.17E3 ± 2.2E3
3.	ΔH1 cal/mol	-1.04E7 ±	-2.58E6 ± 1.1	3.16E6 ±	-2.01E7

		1.97E6		4.3E5	±1.9E7
4.	ΔS1 cal/mol/deg	-3.52E4 ± 1.3	-8.65E3 ± 4.0E3	1.06E4 ± 1.1E4	-6.74E4 ± 6.5E4
5.	N2 (sites)	1.73 ± 0.15	3.49 ± 2.3	8.5 ± 3.3	0.7135 ± 0.392
6.	K2 (M⁻¹)	4.40E5 ± 1.4	0.47 ± 0.32	2.07E5 ± 2.6E5	9.3E5 ± 2.1E6
7.	ΔH2 (cal/mol)	3.920E6 ± 2.2	-4.79E5 ± 1.6	-1.59E6 ± 2.0	-5.49E4 ± 9.9E4
8.	ΔS2 (cal/mol/deg)	1.3E4 ± 0.5	-1.61E3 ± 58.5	-772.5 ± 335	-155 ± 320

Table 4.6 Sequential mode of fitted association constant values of expanded CGG repeats RNA with Piperine obtained from ITC.

S. No.	Parameters	RNAs			
		r(CGG) ₆	r(CGG) ₂₀	r(CGG) ₄₀	r(CGG) ₆₀
1.	K ₁	1.67E6	1.44E7	1.24E7	8.03E7
2.	K ₂	9.29E5	2.98E6	7.79E5	1.25E7
3.	K ₃	6.92E5	6.33E5	2.94E8	5.87E6
4.	K ₄	1.97E5	7.8E5	6.80E5	2.40E7
5.	K ₅	4.82E4	1.75E5	6.80E5	3.06E8
6.	K ₆	7.36E4	7.43E4	6.8E8	1.17E8
7.	K ₇	1.27E5	1.31E4	5.96E5	1.78E5
8.	K ₈	3.55E5	961	4.13E4	2.77E4
9.	K ₉	3.71E5	9.09E4	1.89E7	1.32E7
10.	K ₁₀	9.94E4	9.15E5	2.85E9	1.56E9

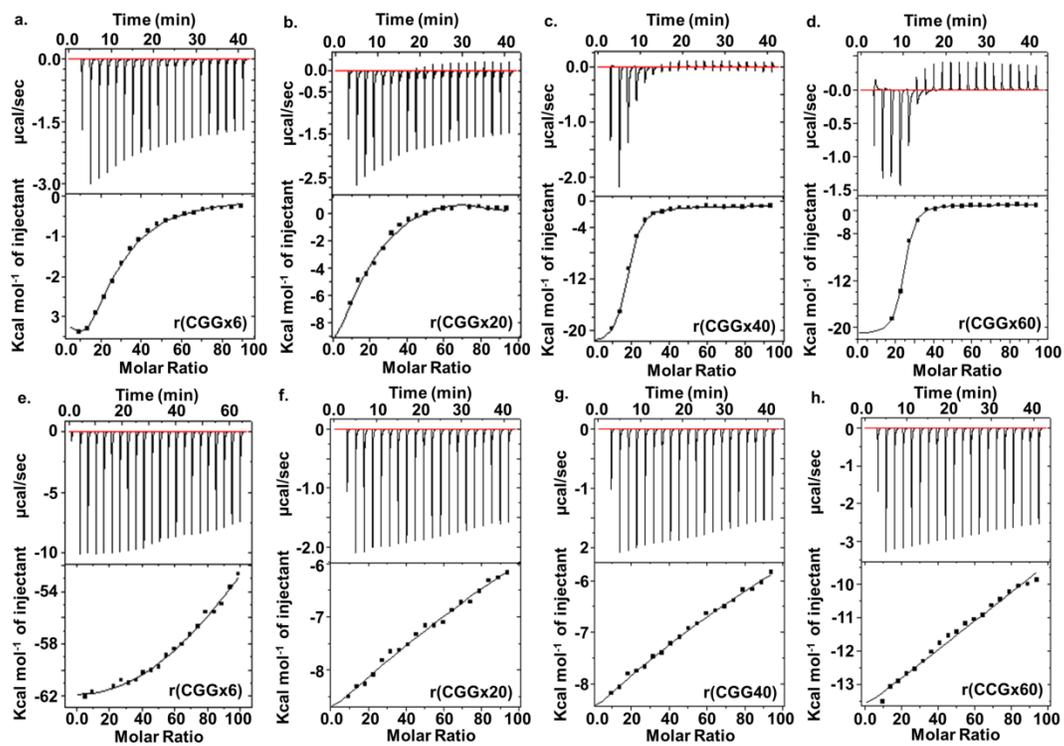


Figure 4.10 Isothermal calorimetry titration plot of TMPyP4 and TMPyP2 with expanded CGG repeat RNAs.

Table 4.7 Thermodynamic parameter values of $r(\text{CGG})^{\text{exp}}$ RNAs with TMPyP4.

S. No.	Parameters	RNAs			
		$r(\text{CGG})_6$	$r(\text{CGG})_{20}$	$r(\text{CGG})_{40}$	$r(\text{CGG})_{60}$
1.	N1 (sites)	6.55 ± 1.5	10.66 ± 3.0	16.0 ± 4.0	78.7 ± 18.8
2.	K1 (M^{-1})	$4.5\text{E}6 \pm 4.1$	$3.17\text{E}6 \pm 2.5$	$4.78\text{E}7 \pm 3.0$	$5.96\text{E}7 \pm 3.3$
3.	$\Delta\text{H}1$ cal/mol	$-2.87\text{E}3 \pm 0.74$	$-1.2\text{E}5 \pm 1.6\text{E}5$	-5690 ± 4.1	-713.5 ± 248
4.	$\Delta\text{S}1$ cal/mol/deg	21.6 ± 3.5	-59.83 ± 22.5	-61.16 ± 23.2	37.7 ± 1.8
5.	N2 (sites)	19.9 ± 1.6	66.33 ± 3.35	87.03 ± 15.40	30.33 ± 13.2
6.	K2 (M^{-1})	$3.17\text{E}4 \pm 1.15$	$7.97\text{E}5 \pm 7.0\text{E}5$	$7.93\text{E}5 \pm 3.2\text{E}5$	$2.54\text{E}10 \pm 0.99$
7.	$\Delta\text{H}2$ (cal/mol)	$-5.23\text{E}3 \pm 1.1$	2207 ± 150.6	-601.23 ± 247	$-1.77\text{E}4 \pm 0.61$
8.	$\Delta\text{S}2$ (cal/mol/deg)	1.28 ± 0.63	132.76 ± 162.7	30.03 ± 8.9	20.26 ± 6.56

Table 4.8 Thermodynamic parameter values of *r(CGG)*_{exp} RNAs with TMPyP2.

S. No.	Parameters	RNAs			
		<i>r(CGG)</i> ₆	<i>r(CGG)</i> ₂₀	<i>r(CGG)</i> ₄₀	<i>r(CGG)</i> ₆₀
1.	N1 (sites)	8.5 ± 2.1	18 ± 2.8	27.5 ± 3.5	39 ± 1.4
2.	K1 (M ⁻¹)	3.75E4 ± 1.4	3.58E5 ± 1.4	1.82E4 ± 0.41	5.3E4 ± 1.6
3.	ΔH1 cal/mol	2.15E4 ± 1.5	-9.01E3 ± 0.65	-3.14E4 ± 2.7	-4.83E6 ± 6.3
4.	ΔS1 cal/mol/deg	-16.35 ± 2.4	-12.75 ± 1.2	-18.8 ± 3.2	5.06E4 ± 2.7
5.	N2 (sites)	48.45 ± 10.2	34.5 ± 4.3	385 ± 22.62	473 ± 32.52
6.	K2 (M ⁻¹)	5.66E4 ± 1.8	2.74 ± 2.07	1.62E3 ± 0.9	2.01 ± 0.22
7.	ΔH2 (cal/mol)	-8.70E3 ± 4.6E3	-1.03E4 ± 0.69	-1.29E4 ± 2.1	-5.89E5 ± 3.4
8.	ΔS2 (cal/mol/deg)	-3.10E ± 1.5E4	-3.27E4 ± 0.33	-28.5 ± 1.2	-4.51E3 ± 2.4

4.2.3 Elucidation of conformational changes of r(CG₂)^{exp} RNA with Piperine using Circular Dichroism spectroscopy

Optics-based qualitative analysis is a crucial investigation to monitor the conformational properties of nucleic acids under different chemical environmental conditions [40]. In the current study, we performed CD spectroscopic titration studies for investigating the structural changes of the target RNA due to the presence of Piperine (Figure 4.11). A positive band around 265-270 nm and a negative band around 215-220 nm typically characterises the CD spectrum of CGG repeats RNA, which is similar to A type RNA confirmation [41,42]. The gradual addition of Piperine to CGG repeats RNA caused a major hypochromic shift and red shift in the negative peak whereas the overall RNA structure was found to remain the same (Figure 4.11 a-e). The CD titration of Piperine with other r(CNG)^{exp}, and r(AU) duplex RNAs also affirms the selectivity of Piperine with CGG repeats RNA. As the other r(CNG)^{exp} and r(AU) duplex RNA also interacted with Piperine nonetheless the changes were not very much significant (Figure 4.11 f-j). Similarly, structural changes were also observed with TMPyP4, but the differences were even lesser (Figure 4.12) and no significant interaction was found with TMPyP2 (Figure 4.13) that further affirms the selectivity of Piperine for r(CG₂)^{exp} RNAs. In general, the presence of major red shift and hypochromicity in negative peak signify $\pi - \pi$ stacking interactions between ligand and target RNA [23]. These constant changes in the CD spectrum of RNA have corroborated the intercalation mode of binding to form a strong Piperine and CGG RNA complex [23,43,44].

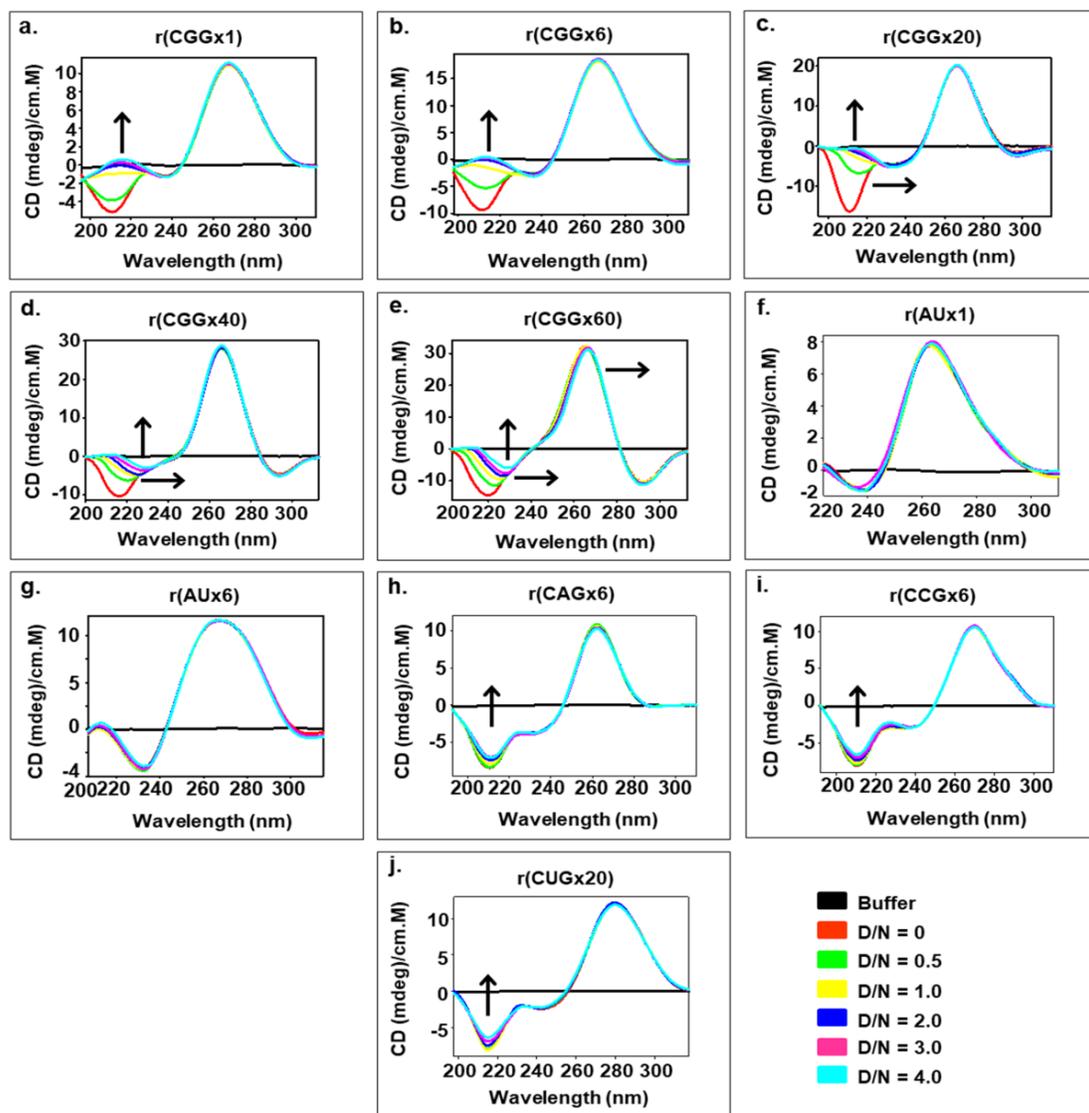


Figure 4.11 Circular Dichroism spectroscopy titration of $r(\text{CNG})^{\text{exp}}$ and $r(\text{AU})^{\text{exp}}$ duplex RNAs in the presence of Piperine. (a) $r(\text{CGG})_1$, (b) $r(\text{CGG})_6$, (c) $r(\text{CGG})_{20}$, (d) $r(\text{CGG})_{40}$, (e) $r(\text{CGG})_{60}$, (f) $r(\text{AU})_1$, (g) $r(\text{AU})_6$, (h) $r(\text{CAG})_6$, (i) $r(\text{CCG})_6$, (j) $r(\text{CUG})_6$ RNA. In the presence of Piperine significant spectroscopic changes were detected with CGG repeat RNAs compare to other mismatched RNA motifs and AU duplex RNA. D/N denotes the drug by nucleotide ratio.

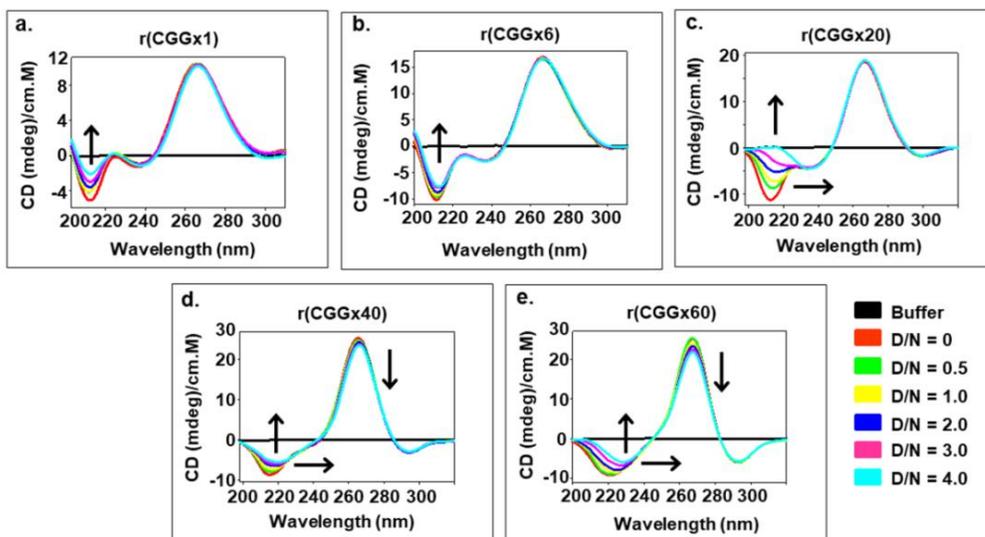


Figure 4.12 Circular dichroism spectroscopic titration of TMPyP4 with expanded CGG repeat RNA. On gradual addition of TMPyP4 caused hypochromic and red shift in the negative peak in CGG repeat RNA. D/N denotes for drug by nucleotide ratio.

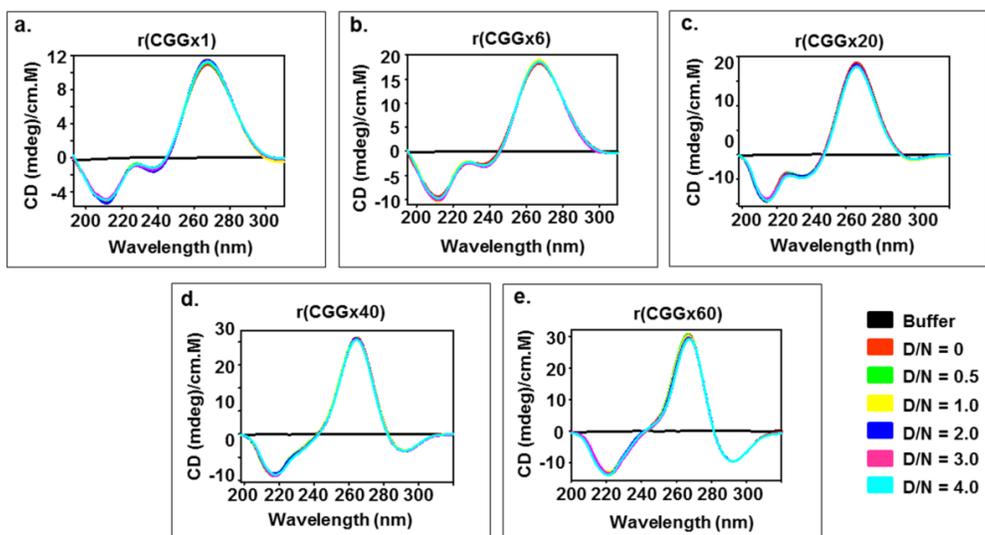


Figure 4.13 Circular dichroism spectroscopic titration of TMPyP2 with expanded CGG repeat RNA. TMPyP2 caused no significant change in the CGG repeat RNA. D/N denotes for drug by nucleotide ratio.

4.2.4 Thermal denaturation analysis of r(CGG)^{exp} RNAs with Piperine

CD thermal denaturation studies of RNA with ligand have been used to determine the folding pattern of RNA as a function of temperature [45]. The CD melting curves were recorded at 267 nm wavelength for all CGG repeat RNAs up to ratio D/N = 2.0. On addition of Piperine at 2.0 molar ratio the change in melting temperature (ΔT_m) was observed as 1.71 °C, 2.8 °C, 3.68 °C, 4.64 °C and 5.36 °C for r(CGG)₁, r(CGG)₆, r(CGG)₂₀, r(CGG)₄₀ and r(CGG)₆₀ respectively (Figure 4.14 a-e & 4.14 j). In comparison to CGG RNA, a minimal rise in T_m was found with other mismatch RNA motifs and no change in T_m was observed with AU paired RNA (Figure 4.14 f-i & 4.14 j). Similarly, TMPyP4 decreased the T_m of CGG repeat RNAs at the same molar ratio whereas no significant changes were found with TMPyP2 (Figure 4.15 and 4.16). Previous studies have demonstrated that an increase in T_m upon addition of ligand is related to target RNA stabilization [19,20]. Thus, our thermal profile data implies that Piperine strongly and selectively interacts with CGG repeats RNA and forms a stable complex. Recently, our group reported some small molecules that stabilize 5'CAG/3'GAC repeats containing RNA hairpin structure which rescue the pathogenic features of HD [20,21]. Similarly, Piperine could also exhibit a neuroprotective effect by stabilizing the hairpin structure of r(CGG)^{exp} RNAs and could potentially aid in the therapeutics development for FXTAS.

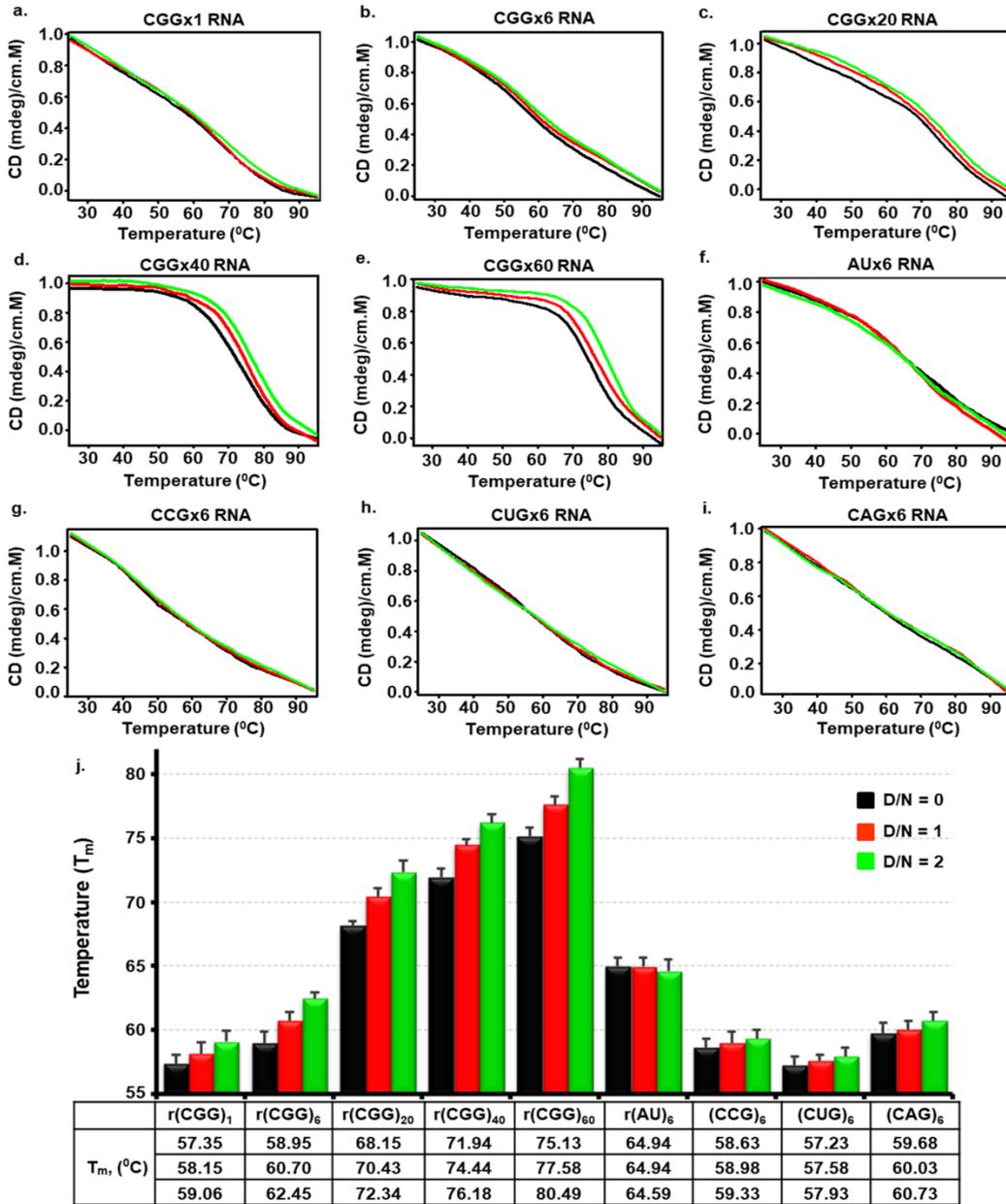


Figure 4.14 A systemic representation of circular dichroism thermal profile of *r(CNG)exp* and *r(AU)exp* duplex RNAs with Piperine (a) *r(CGG)1*, (b) *r(CGG)6*, (c) *r(CGG)20*, (d) *r(CGG)40*, (e) *r(CGG)60*, (f) *r(AU)6*, (g) *r(CCG)6*, (h) *r(CUG)6*, (i) *r(CAG)6* RNA. Piperine enhanced the melting temperature of expanded CGG RNAs whereas no significant change in T_m was observed with other mismatch RNA motifs and AU duplex RNA. D/N denotes the drug by nucleotide ratio.

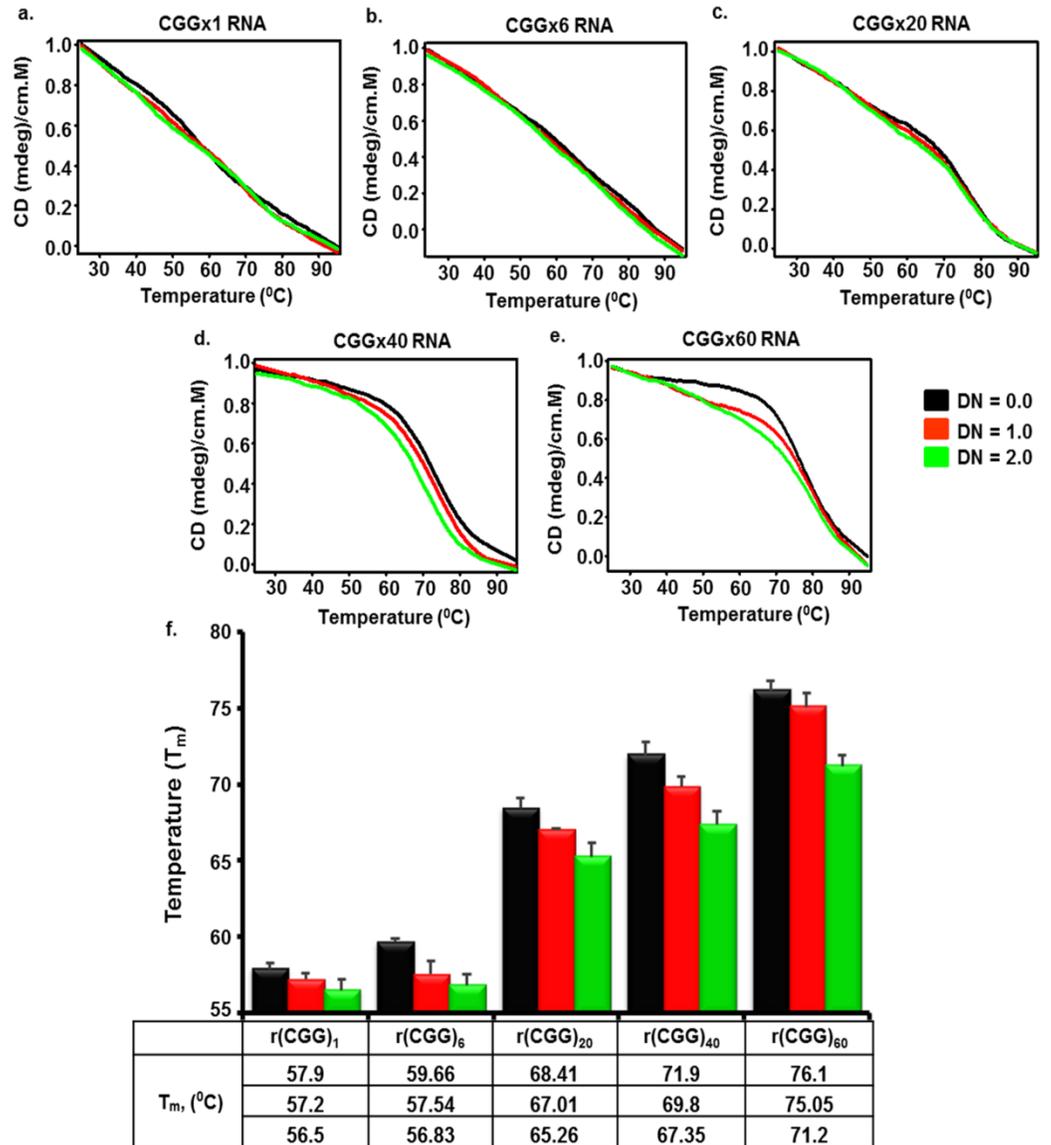


Figure 4.15 Thermal denaturation curve of expanded CGG repeat RNA with TMPyP4. TMPyP4 decreased the melting temperature (T_m) of CGG repeat RNA that suggested it destabilize the CGG hairpin structure.

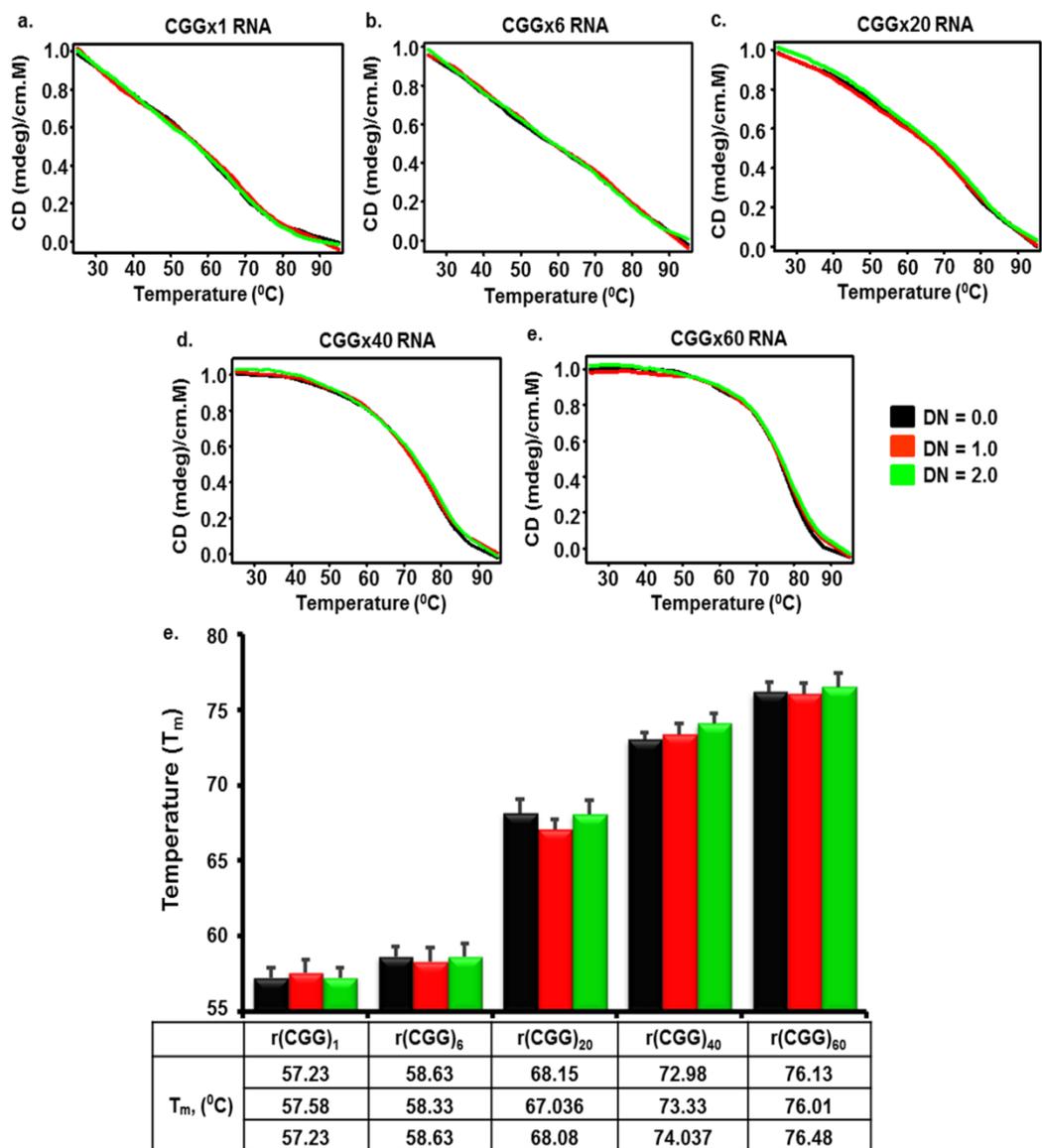


Figure 4.16 Thermal denaturation curve of expanded CGG repeat RNA with TMPyP2. TMPyP2 showed no significant T_m change as compared to Piperine.

4.2.5 NMR broadening experiments to analyze the interactions of Piperine with r(CG_G)_n RNAs

NMR spectroscopy is an attractive tool to visualize RNA ligand interaction at atomic level. NMR chemical shift perturbation and line broadening of ligand resonances upon addition of RNA could provide structural information about the involvement of ligand protons in un-complexed and complexed condition [46]. Considering all this, we performed NMR titration assays of Piperine with r(CG_G)_{1/6} repeat RNAs (Figure 4.17 a - b). The change in the resonances of Piperine protons was measured at each step with the addition of CGG RNA to Piperine. The resonances of H2, H3 and H5 proton of Piperine were found to be completely broadened and H1 & H6 proton were suppressed under the increasing concentration of r(CG_G)_{1/6} RNAs (Figure 4.17 a - b). Protons of the aromatic and aliphatic conjugated system of Piperine were involved in the interaction with CGG RNA.

Further, most of the exchangeable protons of Piperine, involved in binding lie in the region of 6.4 to 7.2 ppm which corroborates our earlier statement of $\pi - \pi$ interaction of Piperine and CGG RNA. Previously, it has been reported that resonance of the conjugated system usually involves $\pi - \pi$ interaction [47]. Conclusively, we demonstrated here that Piperine binds with CGG RNA with high affinity and the intercalation binding mode could be one of the modes of binding. After deducing the binding affinity and selectivity of Piperine with CGG RNA. We performed the molecular docking studies to get a better insight of Piperine binding with CGG RNA (Figure 4.17 c - d). Molecular docking of Piperine with CGG RNA was done using the full grid to provide flexibility at its most preferable binding site. Interestingly, Piperine binds with GxG-5 and GxG-2 (Figure 4.17 c - d) with the best binding energy of -7.88 and -7.98 kcal/mol respectively.

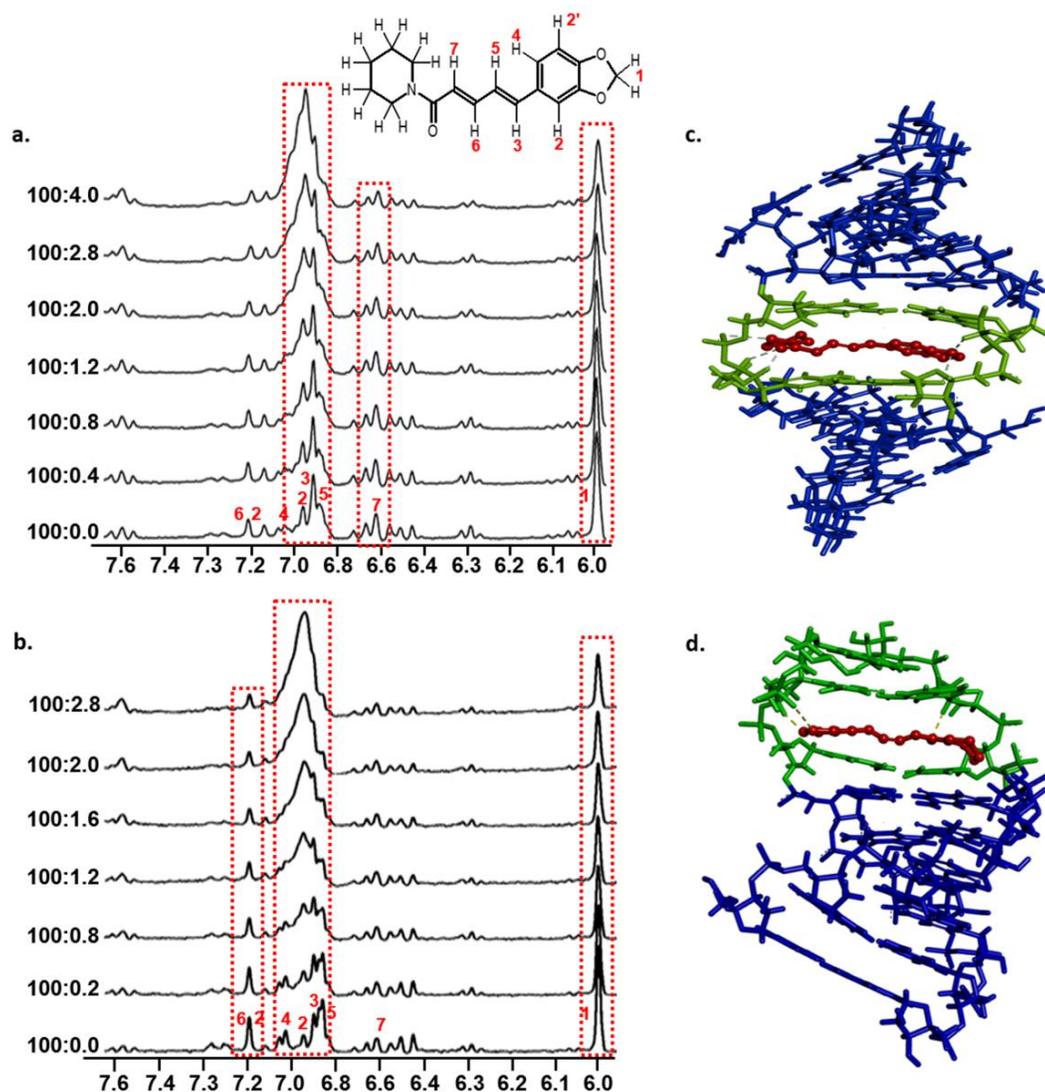


Figure 4.17 One dimensional proton NMR spectroscopy of Piperine with $r(CG)_n$ RNAs and molecular docking study. (a) & (b) represents the NMR titration of 400 μ M of Piperine with (a) $r(CG)_1$ and (b) $r(CG)_6$ RNAs. (c) & (d) docking study of Piperine with CGG RNA (c) GxG-5 CGG RNA (d) GxG-2 CGG RNA.

4.2.6 Gel mobility shift assay and PCR stop assay of r(CGG)^{exp} with Piperine

The above mentioned results provide an idea about the Piperine binding properties with expanded r(CGG)^{exp} repeat RNAs. Furthermore, to affirm these results we performed gel retardation and PCR stop assay. On increasing the concentration of Piperine, the retardation in the migration of band was observed in r(CGG)^{exp} RNAs (Figure 4.18 a & Figure 4.19 a-c). The shift in the mobility was due to the sequence and motif specific interactions that allow the formation of a stable complex between Piperine and r(CGG)^{exp} RNAs [23]. Furthermore, we also found that higher repeats CGG RNA, r(CGG)₄₀ and r(CGG)₆₀ showed prominent retardation in the migration. In contrast to CGG RNA, other mismatched RNAs motifs and AU paired duplex RNA showed no such significant retardation (Figure 4.18 a, Figure 4.19 d, & Figure 4.20 a). TMPyP4 showed less retardation with CGG repeats RNA whereas no considerable changes were observed with TMPyP2 (Figure 4.21 a - b).

Furthermore, we also performed the polymerase chain reaction (PCR) stop assay to further elucidate the binding affinity of Piperine with CGG repeats sequence using DNA templates. Piperine bind to the CGG repeat motif and impair the polymerase activity during extension [23]. Similar to gel retardation assay, with the gradual addition of Piperine, the intensity of PCR bands decreased suggesting the arrest of polymerase enzyme movement during extension (Figure 4.18 b & Figure 4.19 e - h). On the other hand, no significant changes were observed in the intensity of the PCR product of other mismatch & AU pair DNA (Figure 4.20 b). These observations effectively characterize Piperine as a CGG-interactive compound that stabilizes the CGG repeat RNA hairpin structure. Furthermore, TMPyP4 also showed the inhibition of CGG PCR amplification appreciable while no significant inhibition was detected with TMPyP2 (Figure 4.21 c).

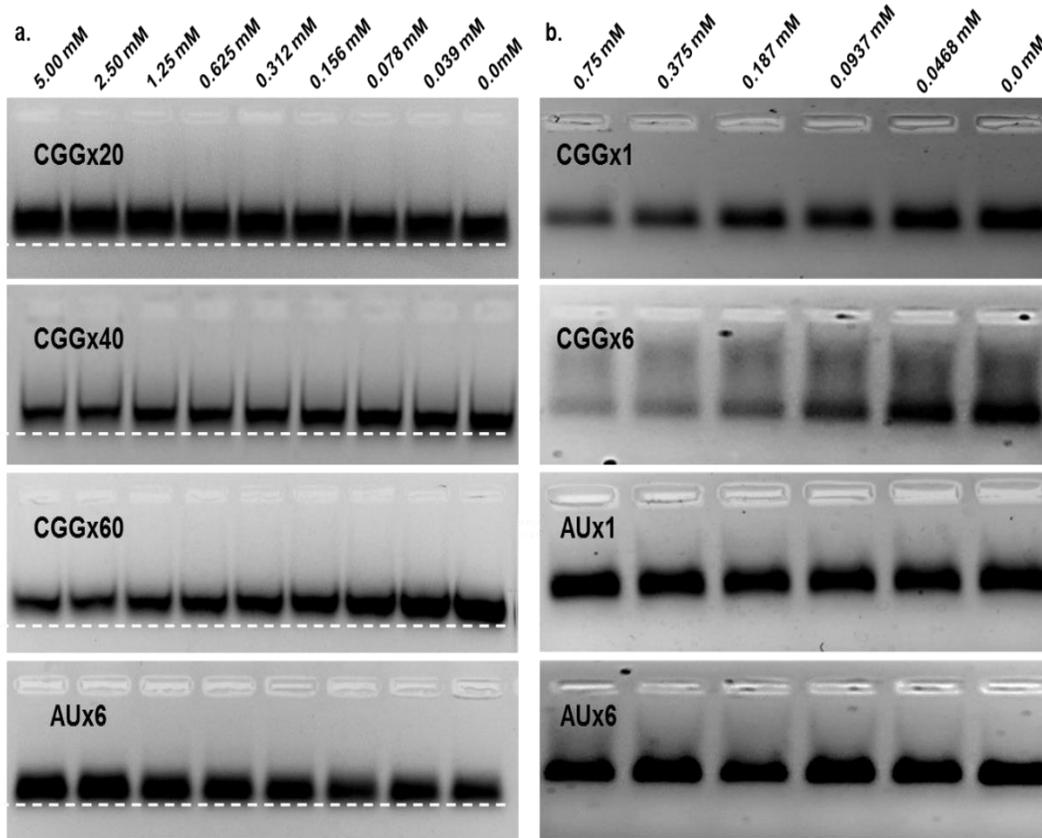


Figure 4.18 Gel retardation assay and PCR stop assay. (a) Gel retardation plot shows an increasing concentration of Piperine significantly retards the mobility of CGG repeat RNAs over AU duplex RNA. (b) Gel images show the decreased intensity of PCR product with increasing concentration of Piperine than AU paired template.

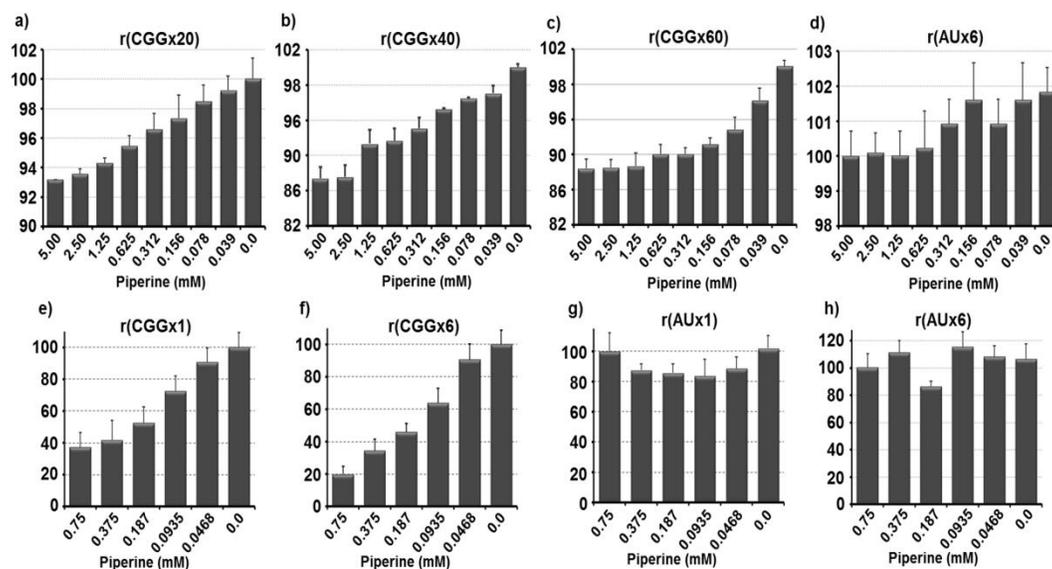


Figure 4.19 Plot representing the percentage reduction in mobility of $r(CG)_x^{exp}$ RNA and PCR inhibition in presence of Piperine. Plot (a) $r(CG)_20$, (b) $r(CG)_40$, (c) $r(CG)_60$ and (d) $r(AU)_6$ shows the retardation in mobility of expanded CGG RNA over AU duplex RNA. (e) $(CG)_1$, (f) $(CG)_6$, (g) $(AU)_1$ and (h) $(AU)_6$ shows the inhibition of PCR product of CGG template in presence of Piperine over AU duplex template.

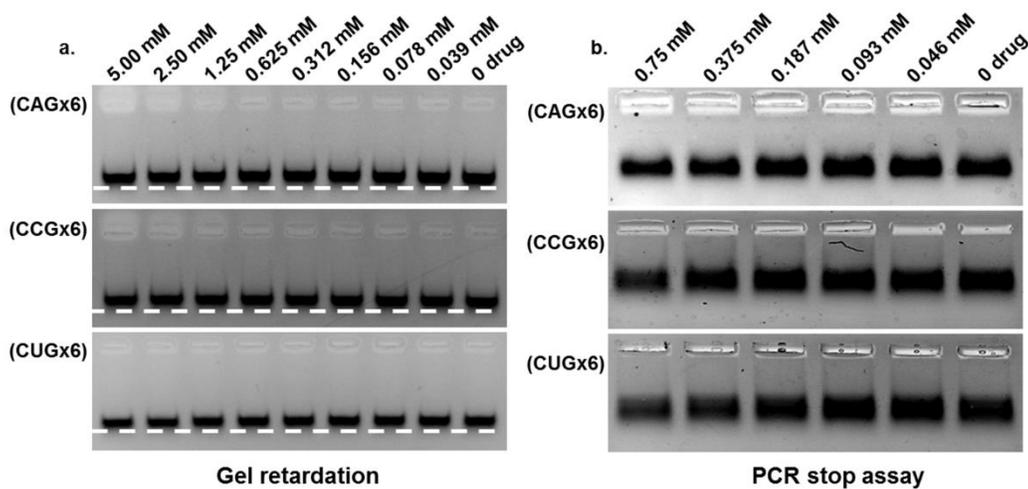


Figure 4.20 Gel retardation and PCR stop assay of other mismatch RNA repeats with Piperine. (a) Piperine showed no significant retardation compared to CGG repeat RNA. (b) Similarly, no appreciable PCR inhibition detected with other mismatched RNAs.

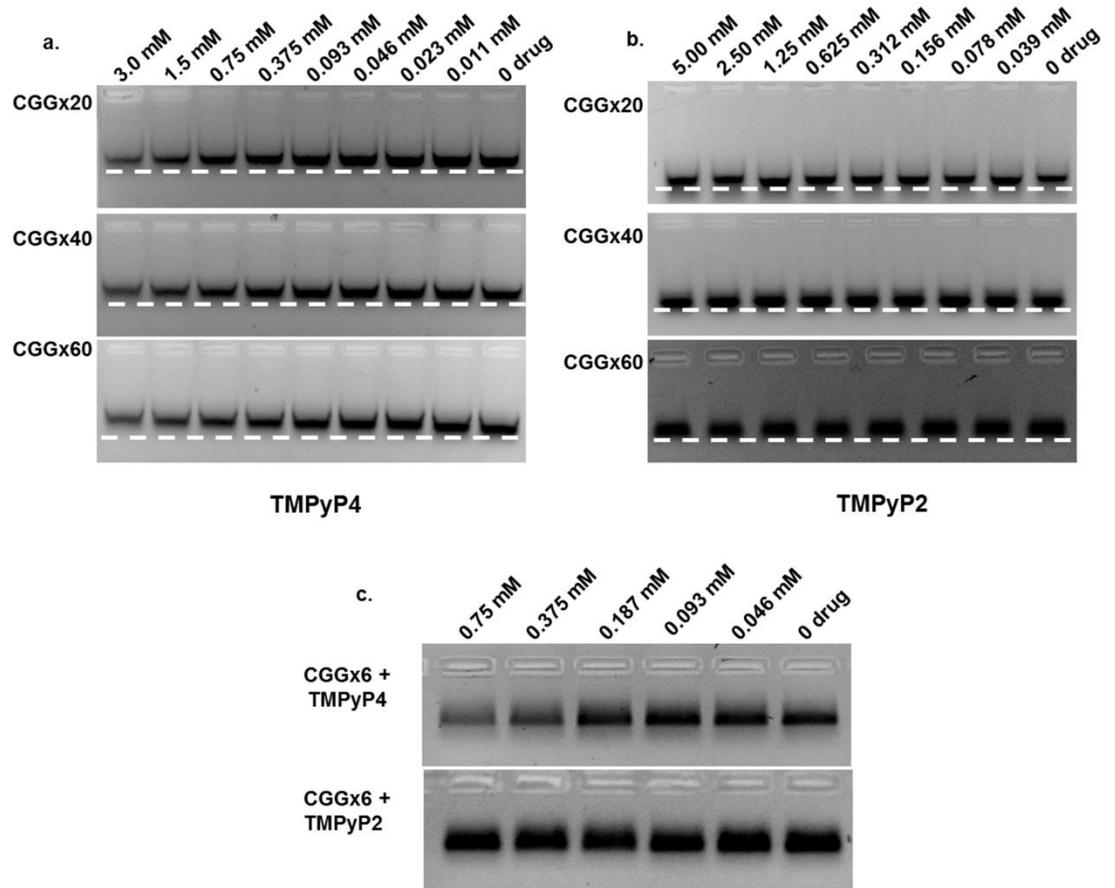


Figure 4.21 Gel retardation and PCR stop assay of TMPyP4 and TMPyP2 with expanded CGG RNAs. (a-b) TMPyP4 showed significant retardation in bands while no retardation observed with TMPyP2. (c) TMPyP4 reduced the band intensity similar to Piperine whereas no inhibition detected with TMPyP2.

4.2.7 Improvement of alternative pre-mRNA splicing defects in FXTAS cellular models

The expression of expanded CGG repeats RNA leads to sequestration of specific RNA binding proteins, ultimately resulting in pre-mRNA splicing defects notably in survival motor neuron 2 (*SMN2*) and B-cell lymphoma x (*Bcl-x*) pre-mRNAs [11]. So, we tested the potency of Piperine to correct alternative pre-mRNA splicing defects in FXTAS cell model system using *SMN2* and *Bcl-x* mini-genes. Herein, HEK293 cells were co-transfected with a plasmid containing 99 CGG repeats and *SMN2* exon 7 minigene. Then, increasing doses of Piperine were added to the cell growth medium for 24 hours. In the absence of r(CGG)^{exp}, *SMN2* exon 7 is included in 35% of mRNAs, while in the presence of r(CGG)^{exp}, the inclusion of *SMN2* exon 7 increases to 80% (Figure 4.22 a). Importantly, Piperine at a concentration of 20 μ M improves splicing defects of *SMN2* mRNA, and 60 μ M of Piperine completely restores alternative pre-mRNA splicing defects. As a control, Piperine does not show any significant effect on alternative splicing of *SMN2* in the absence of r(CGG)^{exp} (Figure 4.22 a). Similarly, *Bcl-x* mRNA has two isoforms *Bcl-xL* and *Bcl-xS*. In normal healthy cell only ~40% of *Bcl-xL* isoform expressed while in r(CGG)^{exp} containing cell model ~60% of the *Bcl-x* mRNA contain *Bcl-xL* isoform [48]. Similarly, Piperine also improves pre-mRNA splicing defect of *Bcl-x* in a dose-dependent manner (Figure 4.22 b). Significant improvement in *Bcl-x* mis-splicing defects was also observed when cell treated with 50 μ M Piperine. Further significant improvement of *Bcl-x* pre-mRNA splicing defect was detected at 100 μ M Piperine. Moreover, Piperine does not affect the pre-mRNA alternative splicing defect of *Bcl-x* in the absence of CGG repeat (Figure 4.22 b).

In addition, we tested the effect of Piperine on EGFP mRNA expression and found no significant effect (Figure 4.22 c). As a further control, we also investigated the alternative splicing of a cardiac troponin T (*cTNT*) exon 5 and insulin receptor (*IR*) exon 11 minigenes, which are not affected by expanded CGG repeats. Further, Piperine treatment does not affect splicing of *cTNT* and *IR*

minigenes (Figure 4.23). These data suggest that Piperine does not globally affect pre-mRNA splicing efficiency and Piperine specifically corrects alternative splicing events mis-regulated by the pathogenic expression of expanded CGG repeats.

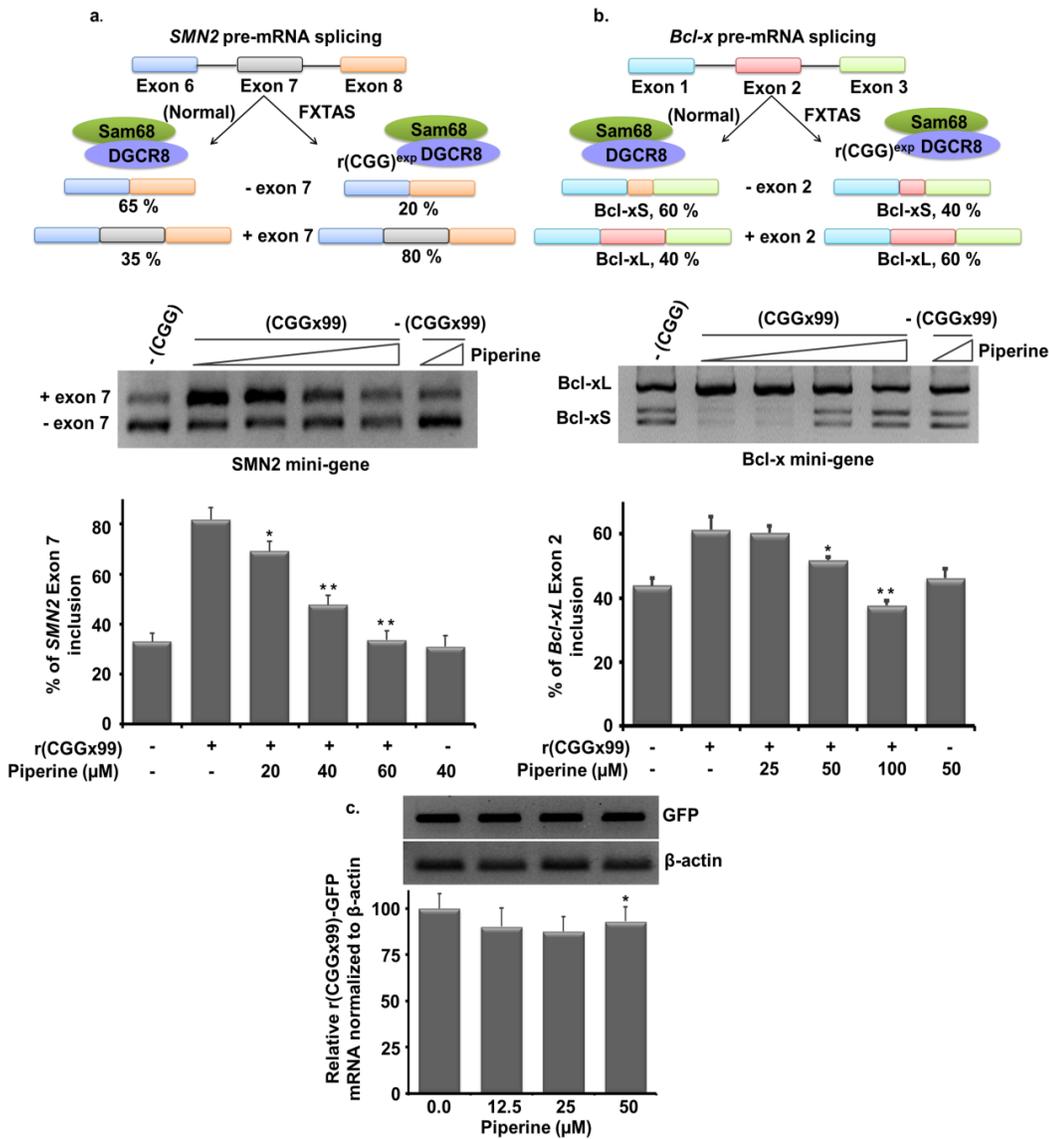


Figure 4.22 In-vitro efficacy of Piperine in FXTAS cellular models. Piperine improves FXTAS associated alternative pre-mRNA splicing defect of SMN2 and Bcl-x minigenes. (a) Improvement of SMN2 splicing defects was found as a function of Piperine concentration (top, left). (b) Similarly, Bcl-x alternative splicing were improved with Piperine treatment in FXTAS cell model (top, right). (c) Interestingly, Piperine does not affect the GFP mRNA expression level. Error bar indicates standard deviation. One way ANOVA was utilized for statistical significance analysis where * $p < 0.05$; ** $p < 0.01$.

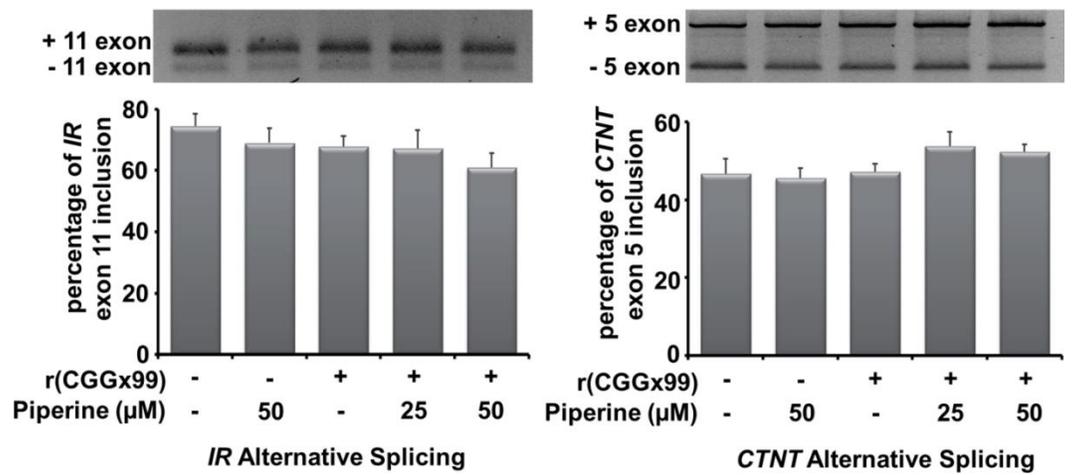


Figure 4.23 Quantification of alternative splicing defects of cTNT and IR minigenes in presence of Piperine (top gel image). CGGx99 and cTNT/IR co-transfected in HEK293 cell. Piperine treatment not shows any significant change in cTNT and IR splicing.

4.2.8 Studying the cellular potency of Piperine: Inhibition of RAN translation

After assessing the affinity and selectivity of Piperine to r(CGG)^{exp} motif containing target RNA, we further sought to determine the ability of Piperine to reduce cellular toxicity caused by r(CGG)^{exp} RNA in the FXTAS cellular models. To develop a FXTAS cellular model, we utilised r(CGG)₉₉ repeats embedded in 5' untranslated region (5' UTR) of the enhanced green fluorescence protein containing plasmid (CGG_{x99}-EGFP-pcDNA3.1) that mimic the natural orientation of r(CGG)^{exp} in *FMRI* gene [11]. A small molecule that could potentially modulate the r(CGG)^{exp} RNA motif will bind to the 5' UTR of the EGFP gene and reduce the expanded repeat-mediated toxicity inside the cell. A plasmid expressing only EGFP (EGFP-pcDNA3.1) was taken as a control for this study. Similar rational approaches have been utilized for designing selective compounds that specifically target different repeat containing mRNAs like r(CAG)^{exp} and r(CUG)^{exp} causes HD, SCA10 and DM1 disease respectively [20,21,49]. Encouragingly, Piperine treatment was found to reduce the protein aggregates in a dose-dependent manner (Figure 4.24 a - b). Arrowheads in Figure 4.24 a - b represent protein inclusion inside the cell. Piperine reduced 27% FMRpolyG aggregates at 50 μM and further reduced to 60% at 100 μM. Interestingly, no significant inhibition of EGFP protein translation was found at the same concentration (Figure 4.24 e - f). Moreover, we manually performed protein aggregates quantification in treated cells. Statistically significant reduction of protein aggregates was found, maximal 67% reduction at 100 μM (Figure 4.24 c). This signifies that Piperine inhibited the translational machinery to initiate the non-canonical translational. Moreover, the increase in the thermal stability of r(CGG)^{exp} hairpin by small molecule corresponds to the inhibitory effect RAN translational. However, many other factors are involved in RAN translation and here in this case increase in the thermal stability of CGG repeat RNA through Piperine binding provide the required inhibitory effect of RAN translation [50].

Our biophysical studies showed the binding affinity in nanomolar concentration however the effects in cellular systems were observed in micromolar range as the

effects in cellular systems is affected due to various factors. It has been previously reported that Piperine has a protective effect against cellular toxicity by reducing the inflammation, apoptosis and oxidative stress load in neurological disorder [30]. Therefore, we checked the effects of Piperine on the cellular cytotoxicity produced by expanded CGG repeats RNA using the MTT assay on cells expressing CGG_{x20}-pcDNA3.1, CGG_{x99}-GFP-pcDNA3.1 and GFP-pcDNA3.1 in the presence of Piperine. Piperine was found to enhance the cell viability of CGG_{x20} and CGG_{x99}-GFP expressing cells (Figure 4.24 d). We also assessed the cytotoxic effect of Piperine in established cell lines (HEK293 & normal fibroblast (GM00357) and in FXTAS patient-derived fibroblast (GM04026). Piperine exerted no significant cytotoxic effect in the established cell lines (GM00357 & HEK293 cells). However, it has shown less cytotoxicity in FXTAS patient-derived cell line (Figure 4.25). Conclusively, our results suggested that Piperine ameliorate the proteotoxicity caused by r(CGG)^{exp} RNA and could be used as a small molecule for therapeutics of FXTAS. Error bar indicates standard deviation. The data was analyzed using one way ANOVA except for cell viability enhancement assay (d) where student t-test was performed. Asterisk denotes the statistical significant difference as compared to control where *p < 0.05; **p < 0.01; and ***p < 0.001

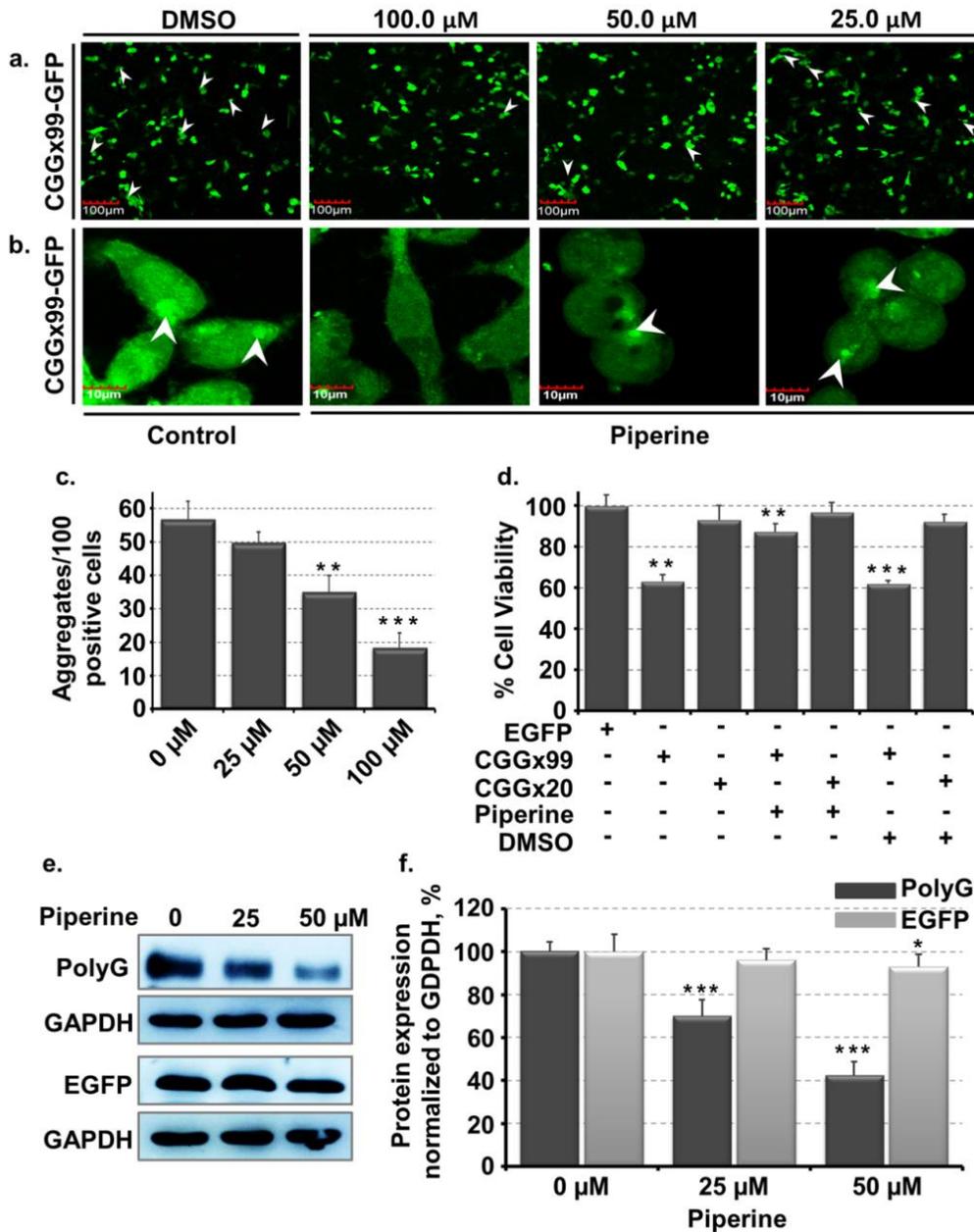


Figure 4.24 Effects of Piperine on protein aggregation & accumulation and its cytotoxicity caused by CGG repeats RNA (a-b) Micrographs of EGFP-CGGx99, transiently transfected HEK293 cells (a) 20x image and (b) 60x image, treated with control (DMSO) & Piperine for 24 hours. The arrowheads represent protein aggregates. (c) Few same sets of cells expressing expanded CGG repeats were used for quantitative analysis of the aggregates. (d) Cell viability assay for CGGx20-pcDNA3.1, CGGx99-GFP-pcDNA3.1 and GFP-pcDNA3.1 plasmid transfected cell (HEK293) in the presence of Piperine (e-f) Images represented the inhibition of FMRpolyG protein aggregates for 24 hrs in the presence of Piperine and vehicle (DMSO) while it does not significantly affect canonical translation of EGFP at same concentration.

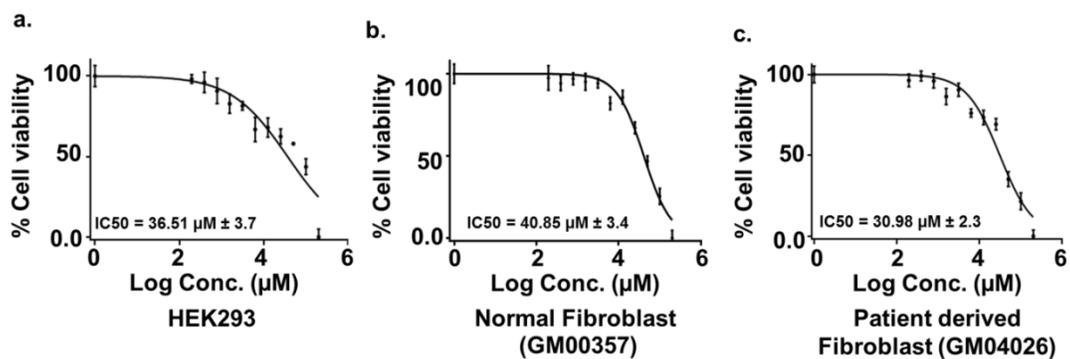


Figure 4.25 Plot represents percentage cell viability of normal established cell line (HEK293), fibroblast cells (GM00357) and patient-derived cells (GM04026).

4.3 Material and Methods

4.3.1 Reagents

Piperine, TMPyP4, and TMPyP2 molecules and other reagents used for the buffer preparation such as NaCl, KCl, NaH₂PO₄, Na₂HPO₄, KH₂PO₄ and K₂HPO₄ were purchased from Sigma Aldrich chemical Pvt. Ltd. The Piperine, TMPyP4, and TMPyP2 molecules were used without further purification. The solvent reagent HPLC grade such as dimethyl sulfoxide (DMSO), 2-butanol, ethanol, deuterium oxide, deuterated DMSO were procured from sigma Aldrich. Midiprep plasmid isolation kit purchased from Invitrogen. All the PCR reagent including Taq polymerase, dNTP mix, primers, DNA template (d(CGG)₁₋₆ repeats) were also obtained from sigma Aldrich. Invitro transcription kit (MEGAscript® T7 Kit) were purchased from Thermo Fisher Scientific pvt. Ltd. For CGG hairpin motif formation, CGG repeat RNAs were dissolved in phosphate buffer (10 mM (K⁺), pH 7.0) with 50 mM KCl. The CGG RNAs were annealed by heating at 90 °C for 5 - 10 mins, and then allowed to cool down at the room temperature for 1-2 hrs. All the biophysical experiment was carried out using above mentioned buffer composition otherwise mentioned separately. For the cell culture experiment, growth media (Dulbecco's Modified Eagle Medium (DMEM), Fetal bovine serum (FBS), Phosphate buffer saline (PBS), Antibiotic solution (ABAM) were purchased from Gibco. The Human Kidney cell line (HEK293) was purchased from National Centre for Cell Science (NCCS), Pune, India.

4.3.2 RNA Preparation and Purification

RNA sequences used for biophysical studies and gel retardation assay were prepared by run off transcription method using synthetic DNA template as previously reported.[51] Briefly, synthetic DNA templates either amplified by PCR or cloned in a plasmid, were transcribed using T7 RNA polymerase, and transcribed products were purified by denaturing 15% PAGE. After UV

shadowing, RNAs were extracted using 0.3 M NaCl by tumbling down for overnight at 4 °C.

4.3.3 Fluorescence titration assay

The Fluorescence titration assays were performed on Synergy™ H1 multi-mode microplate reader using corning 96 well microplate at 25 °C. The excitation and emission wavelength of Piperine, TMPyP4, and TMPyP2 were calculated in 1x potassium phosphate buffer by performing its absorption and fluorescence scan. Each sample was measured in duplicates in 75 µl. The various RNA motifs (5'CNG/3'GNC) and different repeat r(CGG)^{exp} RNAs and (5'CAG/3'GUC), control at the final concentration of 4 µM were serially diluted and last well taken as blank (No RNA). Similarly, RNAs 5'CAG/3'GUC duplex pair (4 µM), yeast t-RNA (50 µM) and different DNAs like CT-DNA, cymc DNA, bcl2 DNA (50 µM) was served as control. For the detection of fluorescence change of Piperine with RNA titration, Piperine excited at the emission wavelength of 341 nm and reading were taken at the emission wavelength of 496 nm. The obtained data were analyzed using Sigma Plot 12.0 software (Systat Software, Chicago, USA) according to the following equation.

$$f = \frac{B_{max}^1 \times \text{abs}(x)}{k_d1 \times \text{abs}(x)} + \frac{B_{max}^2 \times \text{abs}(x)}{k_d2 \times \text{abs}(x)}$$

B_{max} = maximum number of binding sites. K_d = equilibrium binding constant

4.3.4 Isothermal titration calorimetry experiment

The Isothermal titration calorimetry (ITC) measurements were carried out using a MicroCal™ isothermal titration calorimeter iTC200 (GE healthcare) at a constant temperature of 298K. RNA sample were prepared in potassium buffer, 50 mM KCl, 10 mM potassium phosphate, 50 mM EDTA (K +) at pH 7.0. The ITC cell was filled with 20 µM RNA whereas 200 µM Piperine sample was filled in syringe in the same buffer. 1.86 µL of the Piperine added at the each step to the

sample with initial 0.4 μL injection having 60 sec initial equilibrium delays. Total 21 injections were used with 120 sec pause to homogenous the heat between injections and reference power kept it 6 $\mu\text{cal/s}$. The samples stirred speed during the entire titration experiment was kept 750 rpm. The heats of dilution were also calculated by titrating same concentration of Piperine sample with the same buffer only. The binding thermogram of Piperine RNA complex was plotted by two site binding model to determine the dissociation constant (K_d) after the subtracting with the heat of dilution isotherm. Other thermodynamic parameters were also calculated using MicroCal Origin software [52].

4.3.5 Circular Dichroism spectroscopy assay

Circular dichroism experiments were performed on J-815 Spectropolarimeter (JASCO). To maintain the constant temperature 298K, during entire experiment, peltier junction temperature controller was equipped. A constant stream of dry nitrogen with the rate of 5L/min, flushed into the cuvette-holding chamber during whole experiment to avoid the condensation of water droplet outside the cuvette. Spectrums were recorded using the quartz cuvette with 0.2 cm path length containing 10.0 μM CGG, CNG and AU RNA samples and titration were done with increasing concentrations of Piperine, TMPyP4, and TMPyP2 in 50.0 mM KCl, 10.0 mM phosphate buffer (K^+) at pH 7.0. Scan rate 20 nm/min, 0.1 nm intervals from 200 nm to 350 nm with a 1 nm-slit width and averaged over three scans. Lastly, CD spectrum of RNA and RNA-Piperine complex plotted after subtraction with the same buffer.

4.3.6 RNA thermal denaturation study

RNA thermal denaturation experiments were carried out using J-815 Spectropolarimeter (JASCO) equipped with Peltier temperature programmer and water Peltier system PCB-1500. CGG and AU paired RNA (10.0 μM) were diluted in 1x potassium phosphate buffer (50.0mM KCl, 10.0mM phosphate buffer (K^+)) and heated at 90-92°C for 5-10 min then allow it to cool down at

room temperature for next 30 min before start of the experiment. The Melting curve of RNAs were monitored from 25 °C to 95 °C at the rate of 1 °C/min with the titration of Piperine, TMPyP4, and TMPyP2 till drug/nucleic acid ratio 2.0. The normalized absorbance changes at 260 nm against temperature were plotted using the SigmaPlot 12.0 software.

4.3.7 Nuclear magnetic resonance (NMR) spectroscopy

NMR experiment for r(CG_G)₁ and r(CG_G)₆ repeat RNA were performed on the high resolution advance III 400 and 500 MHz BioSpin International AG, Switzerland equipped with a 5/8/10 mm broad band inverse probe able to deliver z-field gradients. Drug-RNA sample were prepared in 1x potassium phosphate buffer (10 mM phosphate buffer, pH 7.2, 0.1 M KCl, and 50 mM EDTA) and 10% D₂O. NMR titration experiments of Piperine were performed with the successive addition of RNA. H₂O + D₂O solvent at a 9:1 ratio were used to lock of the radio frequency during entire titration studies. The 64K data points were recorded for 1D proton NMR spectra with relaxation delay 8 sec, numbers of scan 64 - 128 at 298K and a digital resolution of 0.15 - 0.3 Hz/point was applied. The obtained NMR spectra were consequently processed for the baseline and phase correction. To avoid the sample handling variations for each sample during successive titrations tuning and matching and shimming were performed. Topspin, versions 3.5 was used to process, integrate and analyzed the data. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used to referenced NMR samples.

4.3.8 Molecular Docking

The crystal structure of duplex CGG motif (PDB ID: 3JS2) were used for the docking.[53] The addition, replacements of residues and optimization of duplex RNA structure with Piperine were built in Discovery studio 3.5 (Accelrys Inc.,USA) using CHARMM force field. Piperine docking studies were performed on Autodock 4.0 where duplex RNA treated as solid rigid structure, RNA and

Piperine file prepared in PDBQT format. The other parameters kept as a default values during docking. CGG duplex RNA and Piperine were converted to AD4 format files and Gasteiger charges were assigned to the atoms. The grid box was established in such way which covers the complete RNA structure so that Piperine can explore the whole conformational space. The grid centre was placed by centering the grid box so that both minor groove or intercalation site utilise equally. Lamarckian genetic algorithm [54] was run to carry out energy evaluation of the location of the ligand according to target energy grid and the results were analyzed based on binding energy. The best dock result was further processed to prepare image in Discovery studio 3.5 [23].

4.3.9 Gel retardation assay

Gel mobility shift assay of Piperine were performed with variable length r(CGG)^{exp} RNA. 10 μ M of (CGG)_{20/40/60} RNAs were prepared in potassium phosphate (50mM KCL) buffer prior to heating at 95 °C then allow to cool down at room temperature. The reaction mixture incubates with the increasing concentration of Piperine, Piperine, TMPyP4, and TMPyP2 (0.0 to 5mM) for 30 min at room temperature. The products were resolved on 3% agarose gel prepared in 1X TBE buffer and stained by ethidium bromide. Gel image was visualised and analysed using ImageQuant LAS 4000 (GE Healthcare).

4.3.10 PCR stop assay

The PCR stop assay were performed according to previously describe protocol.[23] Template for GGx1 (5'- GGA GAG GGU UUA AUC GGU ACG AAA GUA CGG AUU GGA UCC GCA AGG - 3'), GGx6 (5'- GGA GAG GGU UUA AUC GGC GGC GGC GGC GGC GGU ACG AAA GUA CGG CGG CGG CGG CGG AUU GGA UCC GCA AGG - 3') and complementary sequence (5'- GGC CGG ATC CTA S5 ATA CGA CTC ACT ATA GGG AGA GGG TTT AAT - 3') were procured from Integrated DNA Technologies pvt ltd. For each concentration of drug final reaction volume of assay is 25 μ l containing

reaction cocktail includes 1X PCR buffer, 4.25 mM MgCl₂, 10 pmol oligonucleotide, 0.33 mM dNTPs, 2.5 units Taq DNA polymerase (sigma aldrich) and volume make up by milli Q water. Piperine was serially added from 0.0 to 700 μM then reaction mixtures were incubated in thermo cycler (Eppendorf) with following conditions: 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 minute and finally held at 4 °C following completion. The obtained PCR product was mixed with 6X DNA loading dye and resolved on 3 % agarose gel stained with ethidium bromide. Gel image was analyzed on ImageQuant LAS 4000 (GE Healthcare).

4.3.11 Improvement of splicing defect in FXTAS cellular model

In order to determine whether Piperine improve pre-mRNA alternative splicing defect *in-vitro*, a FXTAS cellular model were used. Briefly, HEK293 cells were grown in 24 well plates as monolayer in the growth medium contain 1X DMEM, 10% fetal bovine serum, 1% antibiotic and antimycotic at 37 °C with 5% CO₂. After cell reached 80-90% confluency, HEK cell were transfected with the equal amount of plasmid contain CGGx99 repeat and targeted mini-genes (*SMN2* and *Bcl-x*) using Lipofectamine 3000 reagent (Thermo Fisher Scientific), according to standard manufacturers protocol. Transfection cocktail was removed after 4-5 hours and Piperine containing growth medium were added. After 24 hours, cells were lysed in the plate and total RNA were recovered using RNA isolation kit (Invitrogen), according to standard manufacturers protocol.

All different conditioned RNA sample were subjected to reverse transcribe (RT-PCR) using cDNA synthesis kit from Bio-Rad, followed standard protocol. Out of 500 ng reverse transcribed mRNA 100 ng were subjected for semi-quantitative PCR. The PCR protocol used: denaturation for 95 °C for 1 min, annealing 55 °C for 1 min, extension at 72 °C for 2 min and final extension at 72 for 10 min. The PCR protocol was run for 25-30 cycles. The PCR product were analysed on agarose gel electrophoresis, stained with ethidium bromide and image were recorded using ImageQuant LAS 4000 (GE Healthcare). The splicing isoform

intensity was quantified using imageJ software. The primer sequence mentioned in the table S4 for each construct[11].

Two control condition were used to study pre-mRNA splicing defect. 1) Targeted mini-genes (*SMN2* and *Bcl-x*) were co-transfected with plasmid that lack CGG repeat at 5'UTR as describe above. 2) Co-transfection of CGG repeat plasmid with mini-genes whose pre-mRNA splicing not govern by sam68 (cTNT and IR)[48].

4.3.12 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of Piperine by using MTT assay

In order to determine the cytotoxic effect of Piperine in vivo, MTT (3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyltetrazolium bromide dye) assay were performed on normal cell line (HEK293 and GM00357) as well as patient derived cell line (GM04026). Both normal and patient derived cell line ($5 - 10 \times 10^3$ cells/well) were plated in the 96 well culture plates as mono layer in triplicate and allowed to grow in the complete media (1X DMEM, 10% FBS), at 37 °C with 5% CO₂. After cell were attached properly and reached to optimum confluence, cell were treated with Piperine (200 μM to 0.2 μM) with DMSO taken as a vehicle for 24 hrs. After 24 hrs incubation, 10 μL of 5 mg/mL MTT in PBS was added in each well and further incubated for next four hours at 37 °C. The insoluble yellow coloured MTT into dark coloured formazon crystals by intracellular reduction. 100μL DMSO were added to dissolve the formazon crystal structure and absorbance was taken at 570 nm using a microplate reader (Synergy™ H1 multi-mode microplate reader). Half-maximal inhibitory concentration (IC₅₀) of the Piperine was determined by the using formula.[23] In addition, HEK cell were plated to check the cell viability in CGG_{x99} plasmid transfected cell. Cell were transfected with pcDNA-EGFP plasmid (control), CGG_{x20} normal repeat plasmid and CGG_{x99} diseased plasmid using lipofectamine 3000 reagent from Thermo Fisher Scientific as per manufacturer protocol. Piperine treatment was given after

4hrs incubation of transfection for 24 hrs and similar MTT protocol were followed to calculate cell viability.

$$\% \textit{inhibition} = \frac{\textit{Control absorbance} - \textit{sample absorbance}}{\textit{Control absorbance}} \times 100$$

4.3.13 Quantification of canonical GFP and non-canonical FMRpolyG-GFP protein Inclusions

In order to determine the effect of Piperine on polyG-GFP inclusions two different construct were used that contain r(CGG)_{x99}-GFP and pcDNA-GFP. The above mentioned construct was designed and validated by Nicolas group earlier[11]. Briefly, HEK cell were seeded in six well plate in 10% DMEM media. After cell were reached 80-90 percent confluency Transfection were done with above mentioned plasmid using lipofectamine 3000 reagent from Thermo Fisher Scientific as per manufacturer standerd protocol. The transfection cocktail were removed after 4hrs incubation and fresh media were applied which contain respective Piperine concentration and the cells were incbate for 18-24 hrs at 37°C. Cell after washing with PBS, fixed with 4% paraformaldehyde for 15 minutes. Image were captured and protein aggregates visualised at hihger magnification using flurosence confocal microscopy and processed to remove background signals. 100 positve transfected cell from each well were selected and manually graded as “with protein aggregates” and “no aggregates”. The number of GFP protein inclusion were manually counted from three independent experiment and standerd deviation were calculated.

4.4 Conclusion

The development of efficient and versatile therapeutics for CGG repeats is very crucial for the treatment of diseases like FXTAS and FXPOI. The targeting of expanded CGG repeats RNA with naturally available small molecules could provide potential therapeutics lead molecules. A lower dose for long term is always a requisite for the treatment of neurological diseases. Henceforth, naturally available small molecules are always beneficial for the treatment of these diseases. Herein, we demonstrated the selective binding of Piperine with CGG repeats RNA. The selectivity of Piperine was assessed using several biophysical methods like ITC, CD, NMR spectroscopy and the mode of binding was characterized through molecular docking. Moreover, the efficacy of Piperine in ameliorating the disease pathology in cellular models provides potential promise for its therapeutic efficiency hopes. However additional studies involving animal models are anticipated to explore the mechanism of action of Piperine in these cells. Conclusively, this study provides a hope that Piperine could be used as a potential candidate to ameliorate the toxic effect of pathogenic CGG repeats RNA.

4.5 Reference

1. Sharp, P. A. (2009), The Centrality of RNA, *Cell*, 136, 577-580 (10.1016/j.cell.2009.02.007)
2. Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss. (2008), RNA-binding proteins and post-transcriptional gene regulation, *FEBS letters*, 582, 1977-1986 (10.1016/j.febslet.2008.03.004)
3. McMurray, C. T. (2010), Mechanisms of trinucleotide repeat instability during human development, *Nature reviews. Genetics*, 11, 786-799 (10.1038/nrg2828)
4. Orr, H. T. and H. Y. Zoghbi. (2007), Trinucleotide repeat disorders, *Annu Rev Neurosci*, 30, 575-621 (10.1146/annurev.neuro.29.051605.113042)
5. Hagerman, R. J. and P. Hagerman. (2016), Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management, *Nat Rev Neurol*, 12, 403-412 (10.1038/nrneurol.2016.82)
6. Rousseau, F., P. Rouillard, M. L. Morel, E. W. Khandjian and K. Morgan. (1995), Prevalence of carriers of premutation-size alleles of the FMR1 gene--and implications for the population genetics of the fragile X syndrome, *American Journal of Human Genetics*, 57, 1006-1018
7. Coffey, S. M., K. Cook, N. Tartaglia, F. Tassone, D. V. Nguyen, R. Pan, H. E. Bronsky, J. Yuhas, M. Borodyanskaya, J. Grigsby *et al.* (2008), Expanded Clinical Phenotype of Women With the FMR1 Premutation, *American journal of medical genetics. Part A*, 146A, 1009-1016 (10.1002/ajmg.a.32060)
8. Hagerman, P. J. and R. J. Hagerman. (2015), Fragile X-associated tremor/ataxia syndrome, *Ann N Y Acad Sci*, 1338, 58-70 (10.1111/nyas.12693)
9. Hagerman, R. J., M. Leehey, W. Heinrichs, F. Tassone, R. Wilson, J. Hills, J. Grigsby, B. Gage and P. J. Hagerman. (2001), Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X, *Neurology*, 57, 127-130

10. Sellier, C., F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page *et al.* (2013), Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome, *Cell reports*, 3, 869-880 (10.1016/j.celrep.2013.02.004)
11. Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott *et al.* (2010), Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *Embo j*, 29, 1248-1261 (10.1038/emboj.2010.21)
12. Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K.-c. Chen, K. M. Scaglione, V. Basrur *et al.* (2013), CGG Repeat Associated Translation Mediates Neurodegeneration in Fragile X Tremor Ataxia Syndrome, *Neuron*, 78, 10.1016/j.neuron.2013.1003.1026 (10.1016/j.neuron.2013.03.026)
13. Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent *et al.* (2017), Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome, *Neuron*, 93, 331-347 (<https://doi.org/10.1016/j.neuron.2016.12.016>)
14. Crooke, S. T. (2004), Progress in antisense technology, *Annu Rev Med*, 55, 61-95 (10.1146/annurev.med.55.091902.104408)
15. Boggs, R. T., K. McGraw, T. Condon, S. Flournoy, P. Villiet, C. F. Bennett and B. P. Monia. (1997), Characterization and modulation of immune stimulation by modified oligonucleotides, *Antisense Nucleic Acid Drug Dev*, 7, 461-471 (10.1089/oli.1.1997.7.461)
16. Juliano, R. L. (2016), The delivery of therapeutic oligonucleotides, *Nucleic Acids Research*, 44, 6518-6548 (10.1093/nar/gkw236)
17. Juliano, R. L., X. Ming and O. Nakagawa. (2012), Cellular uptake and intracellular trafficking of antisense and siRNA oligonucleotides, *Bioconjug Chem*, 23, 147-157 (10.1021/bc200377d)

18. Kumar, A., R. Parkesh, L. J. Sznajder, J. Childs-Disney, K. Sobczak and M. D. Disney. (2012), Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG) transcripts, *ACS Chemical Biology*, 7, 496-505 (10.1021/cb200413a)
19. Yang, W.-Y., F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney. (2016), Small Molecule Recognition and Tools to Study Modulation of r(CGG)_{exp} in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chemical Biology*, 11, 2456-2465 (10.1021/acschembio.6b00147)
20. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), 13, 180-188 (10.1021/acschembio.7b00699)
21. Khan, E., S. Biswas, S. K. Mishra, R. Mishra, S. Samanta, A. Mishra, A. Tawani and A. Kumar. (2019), Rationally designed small molecules targeting toxic CAG repeat RNA that causes Huntington's disease (HD) and spinocerebellar ataxia (SCAs), *Biochimie* (<https://doi.org/10.1016/j.biochi.2019.05.001>)
22. Leila, G., M. Maedeh, N. G. D. and N. Maryam. (2017), Piperine—The Bioactive Compound of Black Pepper: From Isolation to Medicinal Formulations, *Comprehensive Reviews in Food Science and Food Safety*, 16, 124-140 (doi:10.1111/1541-4337.12246)
23. Tawani, A., A. Amanullah, A. Mishra and A. Kumar. (2016), Evidences for Piperine inhibiting cancer by targeting human G-quadruplex DNA sequences, *Scientific Reports*, 6, 39239 (10.1038/srep39239
<https://www.nature.com/articles/srep39239#supplementary-information>)
24. Sethiya, N. K., P. Shah, A. Rajpara, P. A. Nagar and S. H. Mishra. (2015), Antioxidant and hepatoprotective effects of mixed micellar lipid

- formulation of phyllanthin and piperine in carbon tetrachloride-induced liver injury in rodents, *Food Funct*, 6, 3593-3603 (10.1039/c5fo00947b)
25. Sudjarwo, S. A., K. Eraiko, G. W. Sudjarwo and Koerniasari. (2017), Protective effects of piperine on lead acetate induced-nephrotoxicity in rats, *Iran J Basic Med Sci*, 20, 1227-1231 (10.22038/ijbms.2017.9487)
 26. Bang, J. S., D. H. Oh, H. M. Choi, B.-J. Sur, S.-J. Lim, J. Y. Kim, H.-I. Yang, M. C. Yoo, D.-H. Hahm and K. S. Kim. (2009), Anti-inflammatory and antiarthritic effects of piperine in human interleukin 1 β -stimulated fibroblast-like synoviocytes and in rat arthritis models, *Arthritis Research & Therapy*, 11, R49-R49 (10.1186/ar2662)
 27. Shrivastava, P., K. Vaibhav, R. Tabassum, A. Khan, T. Ishrat, M. M. Khan, A. Ahmad, F. Islam, M. M. Safhi and F. Islam. (2013), Anti-apoptotic and anti-inflammatory effect of Piperine on 6-OHDA induced Parkinson's rat model, *J Nutr Biochem*, 24, 680-687 (10.1016/j.jnutbio.2012.03.018)
 28. Li, S., C. Wang, W. Li, K. Koike, T. Nikaido and M. W. Wang. (2007), Antidepressant-like effects of piperine and its derivative, antiepilepsirine, *J Asian Nat Prod Res*, 9, 421-430 (10.1080/10286020500384302)
 29. Yang, W., Y. H. Chen, H. Liu and H. D. Qu. (2015), Neuroprotective effects of piperine on the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-induced Parkinson's disease mouse model, *Int J Mol Med*, 36, 1369-1376 (10.3892/ijmm.2015.2356)
 30. Bi, Y., P. C. Qu, Q. S. Wang, L. Zheng, H. L. Liu, R. Luo, X. Q. Chen, Y. Y. Ba, X. Wu and H. Yang. (2015), Neuroprotective effects of alkaloids from *Piper longum* in a MPTP-induced mouse model of Parkinson's disease, *Pharm Biol*, 53, 1516-1524 (10.3109/13880209.2014.991835)
 31. Sobczak, K., G. Michlewski, M. de Mezer, E. Kierzek, J. Krol, M. Olejniczak, R. Kierzek and W. J. Krzyzosiak. (2010), Structural diversity of triplet repeat RNAs, *J Biol Chem*, 285, 12755-12764 (10.1074/jbc.M109.078790)

32. Connelly, Colleen M., Michelle H. Moon and John S. Schneekloth, Jr. (2016), The Emerging Role of RNA as a Therapeutic Target for Small Molecules, *Cell Chemical Biology*, 23, 1077-1090 (10.1016/j.chembiol.2016.05.021)
33. Sobczak, K., M. de Mezer, G. Michlewski, J. Krol and W. J. Krzyzosiak. (2003), RNA structure of trinucleotide repeats associated with human neurological diseases, *Nucleic acids research*, 31, 5469-5482 (10.1093/nar/gkg766)
34. Weisman-Shomer, P., E. Cohen, I. Hershco, S. Khateb, O. Wolfovitz-Barchad, L. H. Hurley and M. Fry. (2003), The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)_n, *Nucleic Acids Res*, 31, 3963-3970 (10.1093/nar/gkg453)
35. Han, H., D. R. Langley, A. Rangan and L. H. Hurley. (2001), Selective interactions of cationic porphyrins with G-quadruplex structures, *J Am Chem Soc*, 123, 8902-8913
36. Falconer, R. J. and B. M. Collins. (2011), Survey of the year 2009: applications of isothermal titration calorimetry, *J Mol Recognit*, 24, 1-16 (10.1002/jmr.1073)
37. Islam, M. M., R. Sinha and G. S. Kumar. (2007), RNA binding small molecules: Studies on t-RNA binding by cytotoxic plant alkaloids berberine, palmatine and the comparison to ethidium, *Biophysical Chemistry*, 125, 508-520 (<https://doi.org/10.1016/j.bpc.2006.11.001>)
38. Du, X., Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji and S. Q. Liu. (2016), Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods, *Int J Mol Sci*, 17 (10.3390/ijms17020144)
39. Tse, W. C. and D. L. Boger. (2004), Sequence-selective DNA recognition: natural products and nature's lessons, *Chem Biol*, 11, 1607-1617 (10.1016/j.chembiol.2003.08.012)

40. Sosnick, T. R. (2001), Characterization of tertiary folding of RNA by circular dichroism and urea, *Curr Protoc Nucleic Acid Chem*, Chapter 11, Unit 11.15 (10.1002/0471142700.nc1105s04)
41. Kypr, J., I. Kejnovská, D. Renciuik and M. Vorlícková. (2009), Circular dichroism and conformational polymorphism of DNA, *Nucleic acids research*, 37, 1713-1725 (10.1093/nar/gkp026)
42. Stefl, R., L. Trantirek, M. Vorlickova, J. Koca, V. Sklenar and J. Kypr. (2001), A-like guanine-guanine stacking in the aqueous DNA duplex of d(GGGGCCCC), *J Mol Biol*, 307, 513-524 (10.1006/jmbi.2001.4484)
43. Das, A., K. Bhadra and G. Suresh Kumar. (2011), Targeting RNA by small molecules: comparative structural and thermodynamic aspects of aristololactam-beta-D-glucoside and daunomycin binding to tRNA(phe), *PLoS One*, 6, e23186 (10.1371/journal.pone.0023186)
44. Giri, P. and G. S. Kumar. (2008), Self-structure induction in single stranded poly(A) by small molecules: Studies on DNA intercalators, partial intercalators and groove binding molecules, *Arch Biochem Biophys*, 474, 183-192 (10.1016/j.abb.2008.03.013)
45. Shaw, N. N., H. Xi and D. P. Arya. (2008), Molecular recognition of a DNA:RNA hybrid: Sub-nanomolar binding by a neomycin-methidium conjugate, *Bioorganic & Medicinal Chemistry Letters*, 18, 4142-4145 (<https://doi.org/10.1016/j.bmcl.2008.05.090>)
46. Chen, B., X. Zuo, Y. X. Wang and T. K. Dayie. (2012), Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy, *Nucleic Acids Res*, 40, 3117-3130 (10.1093/nar/gkr1154)
47. Boehr, D. D., A. R. Farley, G. D. Wright and J. R. Cox. (2002), Analysis of the pi-pi stacking interactions between the aminoglycoside antibiotic kinase APH(3')-IIIa and its nucleotide ligands, *Chem Biol*, 9, 1209-1217
48. Disney, M. D., B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney. (2012), A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome, *ACS Chem Biol*, 7, 1711-1718 (10.1021/cb300135h)

49. Bernat, V. and Matthew D. Disney. (2015), RNA Structures as Mediators of Neurological Diseases and as Drug Targets, *Neuron*, 87, 28-46 (<https://doi.org/10.1016/j.neuron.2015.06.012>)
50. Yang, W. Y., F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney. (2016), Small Molecule Recognition and Tools to Study Modulation of r(CGG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chem Biol*, 11, 2456-2465 (10.1021/acscchembio.6b00147)
51. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), *ACS chemical biology*, 13, 180-188 (10.1021/acscchembio.7b00699)
52. Mishra, S. K., N. Jain, U. Shankar, A. Tawani, T. K. Sharma and A. Kumar. (2019), Characterization of highly conserved G-quadruplex motifs as potential drug targets in *Streptococcus pneumoniae*, *Scientific Reports*, 9, 1791 (10.1038/s41598-018-38400-x)
53. Kumar, A., P. Fang, H. Park, M. Guo, K. W. Nettles and M. D. Disney. (2011), A Crystal Structure of a Model of the Repeating r(CGG) Transcript Found in Fragile × Syndrome, *Chembiochem*, 12, 2140-2142 (10.1002/cbic.201100337)
54. M., M. G., G. D. S., H. R. S., H. Ruth, H. W. E., B. R. K. and O. A. J. (1998), Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*, 19, 1639-1662 (doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B)

Chapter 5

Curcumin regulates the r(CGG)^{exp} RNA hairpin structure and ameliorate defects in Fragile X-associated tremor ataxia syndrome

5.1 Introduction

Many, if not all, cellular functions are controlled by changes in mRNA or non-coding RNA expression and regulations [1]. Thus, alterations in RNA metabolism can lead to many pathologies, including neuro-muscular disorders and cancer [2,3]. In that aspect, expression of RNA containing an expansion of trinucleotide repeats (TNR) at different gene locations such as 5' untranslated, exons, introns, and 3' untranslated regions causes various disorders, like Fragile X-associated tremor ataxia syndrome (FXTAS) [4], Huntington's disease (HD) [5], Amyotrophic Lateral Sclerosis (ALS), Spinocerebellar ataxia type 10 (SCA10) [6], myotonic dystrophy type 2 (DM2) and Myotonic dystrophy type 1 (DM1), respectively [7,8]. These repeat expansions (CGG, CAG, GGGGCC, AUUCU, CCUG, CUG, etc.) elicit toxicity *via* overlapping pathogenic mechanisms such as RNA gain of function, protein loss of function and protein gain of function [9]. Specifically, FXTAS is a late onset neurodegenerative disorder caused by an expansion of 55 to 200 CGG trinucleotide repeats in the 5' untranslated region (5'UTR) of the Fragile X-mental retardation 1 (*FMR1*) gene that encodes the Fragile X-mental retardation protein (FMRP) [10,11]. Clinical symptoms of FXTAS are intention tremor, parkinsonism, dysautonomia and cerebellar and gait ataxia [12,13]. Of interest, expansions longer than 200 CGG repeats trigger CpG methylation, resulting in transcription silencing of the *FMR1* gene and subsequently loss of expression of the fragile X-protein (FMRP), lead to a

different pathology, the neurodevelopmental Fragile X syndrome characterized by autism and intellectual disability.

CGG repeats in FXTAS are pathogenic through two main mechanisms, RNA gain of function and protein gain of function. First, transcribed CGG repeats form RNA foci that sequester specific RNA binding proteins such as Src-associated in mitosis of 68 kDa (Sam68) [14], TAR DNA-binding protein (TDP-43) [15] and heterogeneous nuclear ribonucleoprotein (hnRNP A2/B1) [16], which are involved in alternative pre-mRNA splicing. Sequestration of these proteins by the expanded CGG RNA repeats results in specific splicing defects [14,17] [18]. A second pathogenic mechanism in FXTAS is the repeat-associated non-AUG translation (RANT) of the expanded CGG repeats into toxic proteins [19-21]. CGG repeats are notably translated through initiation at non canonical ACG and GUG start codons into a polyglycine containing protein, FMRpolyG, which is found in ubiquitin positive intranuclear inclusions in FXTAS cell and mouse models as well as in FXTAS patient tissues [19,20,22].

Given their roles in multiple disorders, repeat containing RNAs are an attractive target for chemical probe and drug development [23]. However, there are very few drugs designed against RNA yet. In this study, we focused on the identification of small molecules targeting expanded CGG RNA repeats (r(CGG)^{exp}) causing FXTAS. Ideally, a positive compound would prevent non-AUG translation of the CGG repeats into the toxic FMRpolyG protein, as well as correct the alternative splicing defects caused by titration of RNA binding proteins by expanded CGG repeats. Here, we found that Curcumin, a polyphenol used as a traditional herbal medicine, binds to CGG RNA repeats. We validated the binding affinity and selectivity of Curcumin to multiple r(CGG)^{exp} RNA using various biophysical methods such as isothermal titration calorimetry (ITC), Circular Dichroism (CD) spectroscopy, CD thermal denaturation, gel shift and PCR stop assays. Importantly, Curcumin corrects pre-mRNA alternative splicing defects caused by expanded CGG repeats expression. Furthermore, Curcumin also reduces formation of FMRpolyG protein aggregate and improves cell viability in

FXTAS cell models. In conclusion, our study suggests that Curcumin could be considered as a potential candidate for therapeutic approach in the neurodegenerative Fragile X-associated tremor ataxia syndrome.

5.2 Results and Discussion

5.2.1 Binding affinity and selectivity of Curcumin with different trinucleotide RNA repeats.

There is an urgent need to develop therapeutic approaches targeting the molecular mechanisms causing FXTAS pathogenesis. As chemical probes binding to r(CGG)^{exp} RNA are able to reverse the toxic consequences of CGG expansions expression. We searched for naturally available and biologically potent small molecules that would bind to CGG RNA repeat [24].

As a first step, we determined the binding selectivity and affinity of Curcumin for different trinucleotide RNA repeated 5'CNG/3'GNC sequence where N stands for either A, G, U or C bases (Figure 5.1). We took advantage of the natural fluorescence of Curcumin to develop a fluorescence based binding assay recorded at the emission maxima 496 nm in presence or absence of different repeat containing RNAs. With gradual addition of RNA to Curcumin, an enhancement in fluorescence intensity was detected, revealing an interaction of Curcumin with RNA. The change in spectral response (ΔF) was plotted against the concentration of RNA and binding constants were determined using a ligand binding two-site saturation model. Binding constant (K_d) values indicate a preferential binding of Curcumin with 5'CGG/3'GGC RNA motif compared to other RNA motifs and AU duplex RNA (Figures 5.2 a – d, Figure 5.3 and Table 5.1). To confirm binding affinity of Curcumin to CGG expanded RNA repeats, we performed fluorescence titration assay with r(CGG)20x, r(CGG)40x and r(CGG)60x RNA. Interestingly, Curcumin showed ~62 and 161 folds better binding for r(CGG)40x and r(CGG)60x RNA over r(AU)1x duplex pair RNA respectively (Figure 5.2 b & e, Figure 5.4 and Table 5.2). As further controls,

binding assay were also performed with r(AU)_{6x} duplex RNA and yeast t-RNA. Curcumin showed ~178 and 273 folds higher affinities with r(CG₆₀)_{60x} RNA over r(AU)_{6x} and yeast t-RNA, respectively (Figure 5.2 c & e, Figure 5.4 and Table 5.2). It has been reported that r(CAG)^{exp}, r(CCG)^{exp} and r(CUG)^{exp} in the transcripts leads to the alternative splicing defects and RAN translation in neuronal cells [25]. Similar to 1x1 GG motif in r(CG₆₀)^{exp}, AA, CC and UU hairpin structure formed in r(CAG)^{exp}, r(CCG)^{exp}, r(CUG)^{exp}. Binding studies demonstrated that Curcumin showed high selectivity with CGG RNA over other mismatched RNA. Finally, different G-quadruplexes (c-kit, cmyc, tel22 and bcl2) and duplex calf thymus (CT) DNA controls were also tested (Figure 5.2 c & e, Figure 5.4 and Table 5.2). Binding studies revealed that Curcumin binds hundred folds better CGG RNA than duplex DNA, supporting a specificity of Curcumin for CGG motif RNA in the nanomolar range.

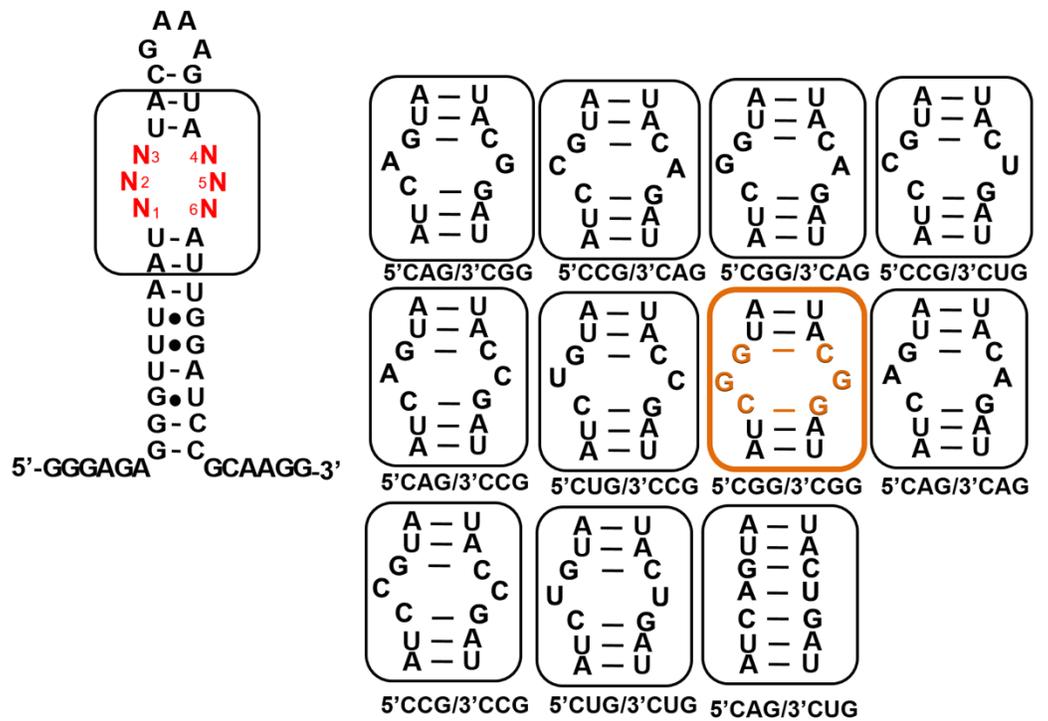


Figure 5.1 Secondary structure of different 1x1 internal nucleotide motif (5'CNG/3'GNC) RNAs sequence.

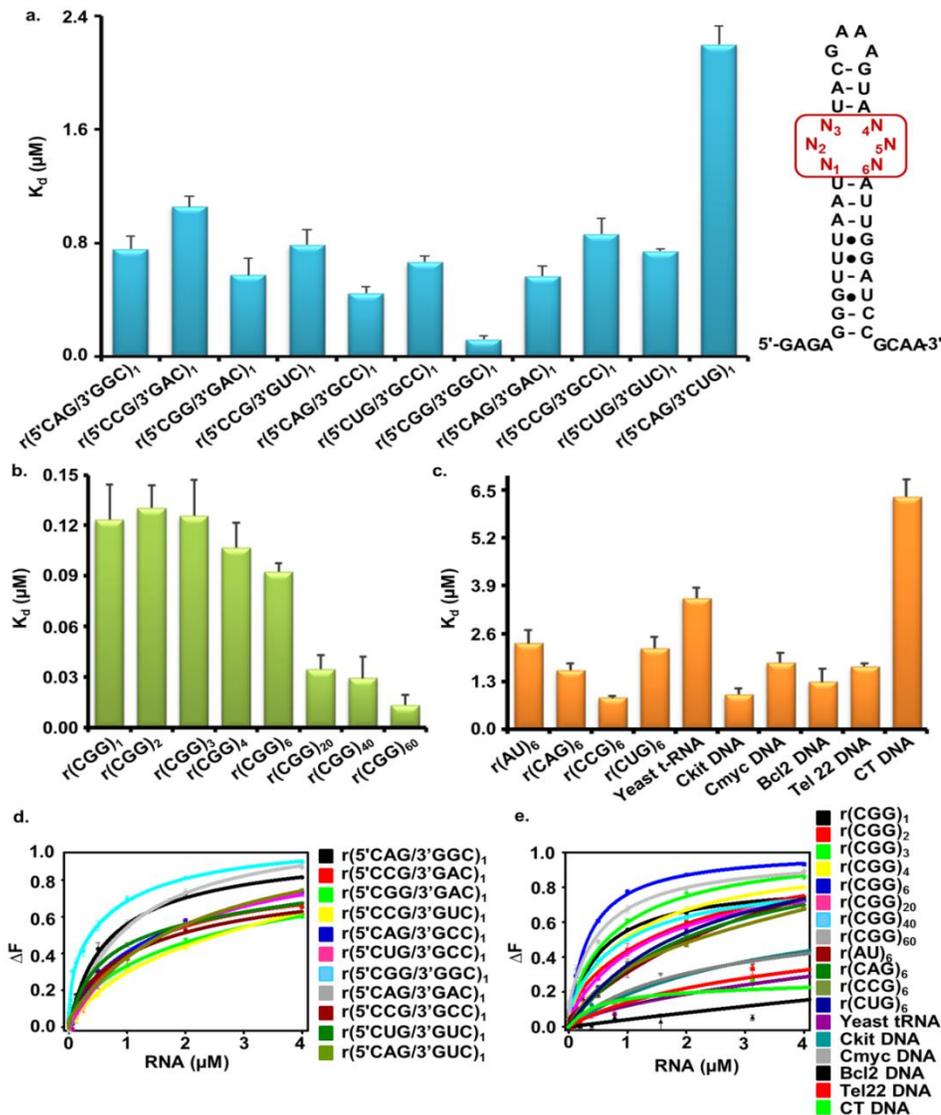


Figure 5.2 Elucidation of binding affinity (K_d) of Curcumin with nucleic acids using fluorescence based assay. *a)* Bar graph representing the binding constant (K_d) values of Curcumin with different 1x1 nucleotide internal motif RNAs. *b)* Bar graph showing the K_d values of Curcumin with different CGG repeat containing RNAs. *c)* Bar graph showing K_d values of Curcumin with different RNA and DNA controls. *d)* Plot illustrates fitting for fluorescence binding assay of Curcumin with different 1x1 nucleotide internal loop RNAs. *e)* Plot illustrate fitting for fluorescence binding assay of Curcumin with different CGG repeat containing RNAs along with different RNA and DNA controls.

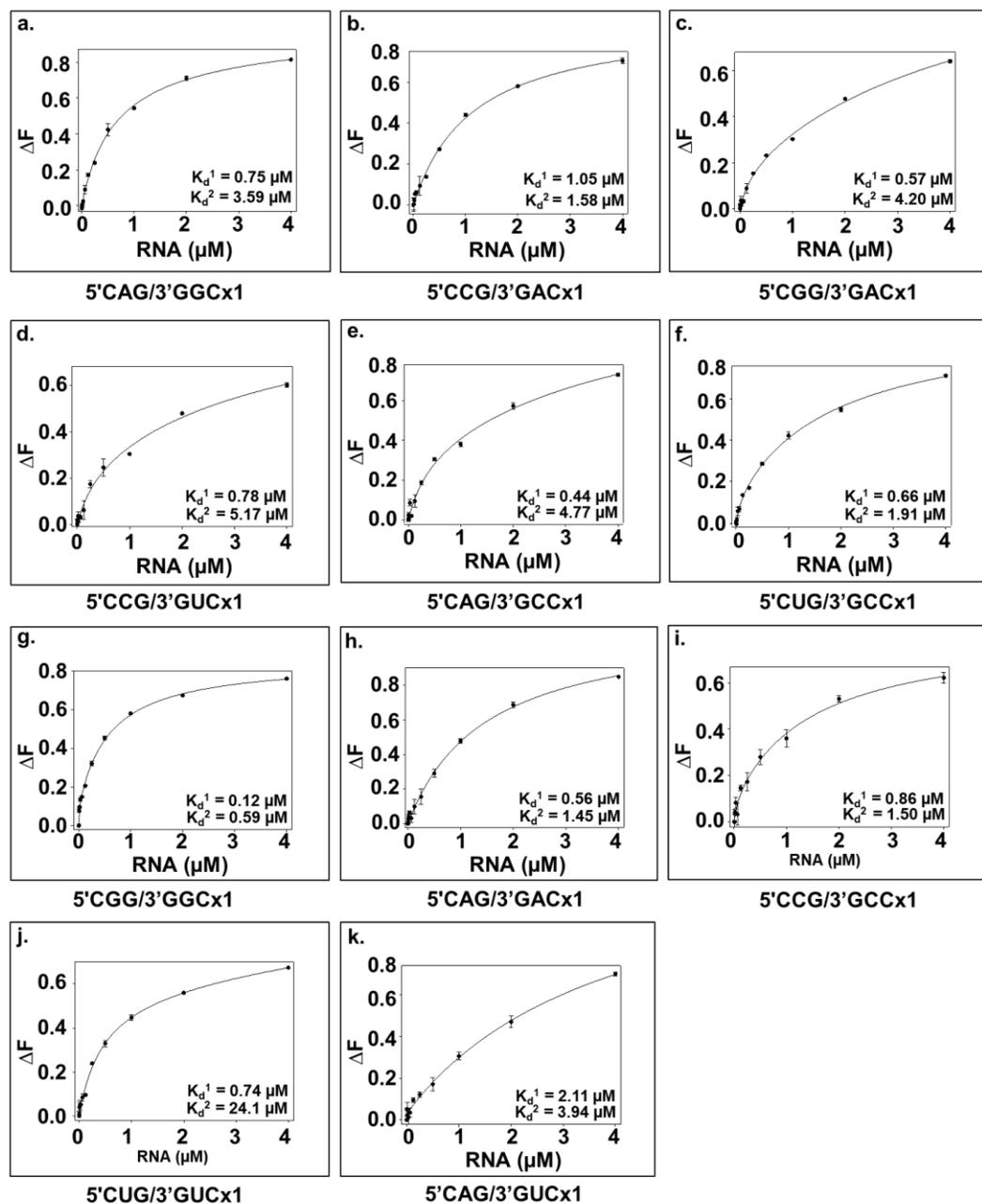


Figure 5.3 Plots show the fluorescence titration assay of different 1 x 1 5'(CNG/3'GNC) RNAs motif with Curcumin. The dissociation constant (K_d) values are also mentioned below.

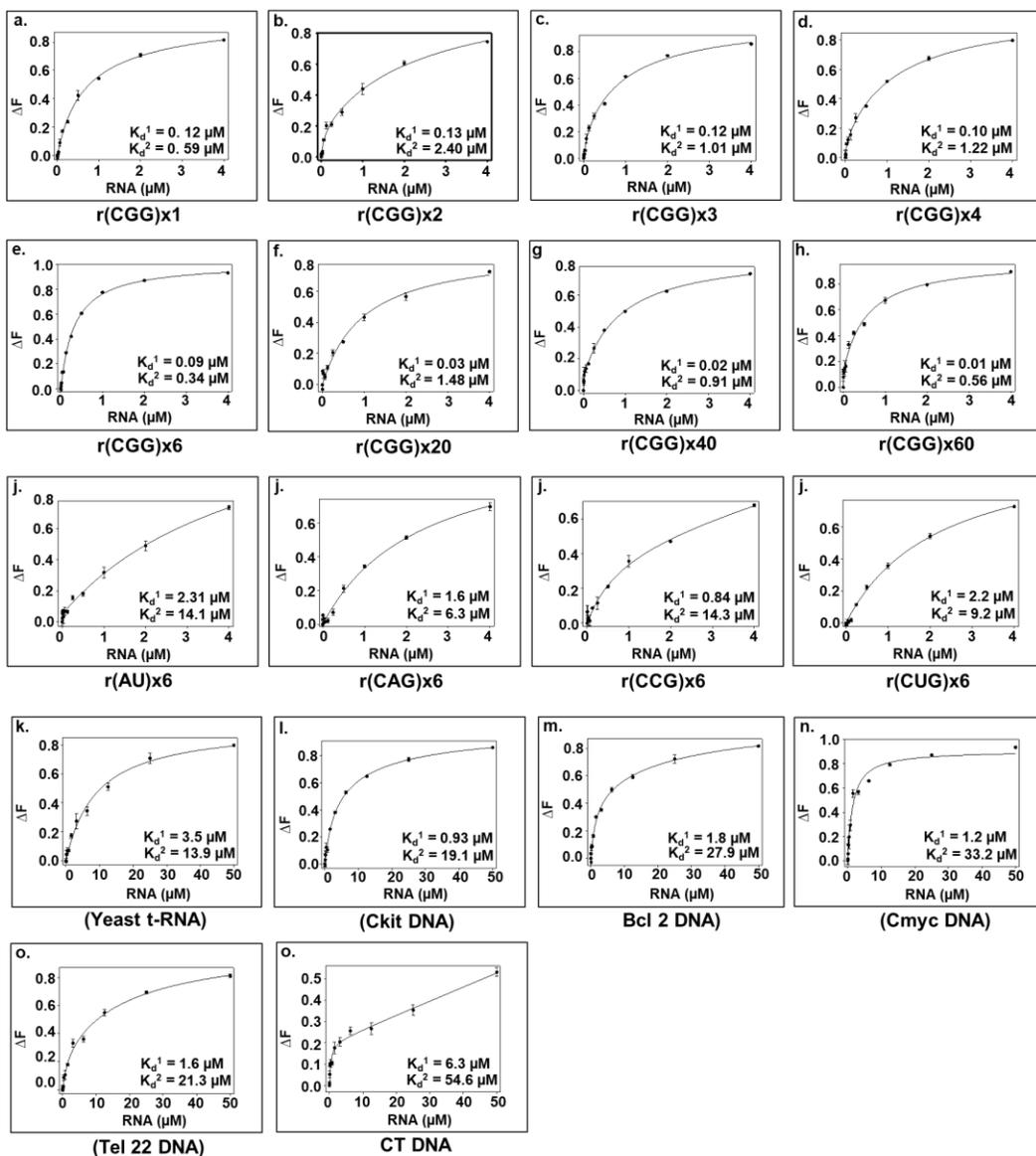


Figure 5.4 Diagram shows the fluorescence titration assay plots of Curcumin with different CGG repeat RNAs and RNA ((AU)6x, r(CAG)6x, r(CCG)6x & r(CUG)6x) & DNA controls (c-kit DNA, cmyc DNA, bcl2 DNA, tel22 DNA & ct DNA). The dissociation constant (K_d) values are also mentioned below.

Table 5.1 The dissociation constant (K_d) values was determined from flourescene titration assay of Curcumin with different (5'CNG/3'CNG)x1 RNAs motif.

S. No.	RNA (5'C <u>N</u> G/3'G <u>N</u> C) x 1	K_d^1 (μ M)	K_d^2 (μ M)
1.	(5'CAG/3'GGC) x 1	0.75 \pm 0.08	3.59 \pm 1.1
2.	(5'CCG/3'GAC) x 1	1.05 \pm 0.07	1.58 \pm 0.26
3.	(5'CGG/3'GAC) x 1	0.57 \pm 0.11	4.2 \pm 0.23
4.	(5'CCG/3'GUC) x 1	0.78 \pm 0.10	5.17 \pm 0.42
5.	(5'CAG/3'GCC) x 1	0.44 \pm 0.04	4.77 \pm 0.34
6.	(5'CUG/3'GCC) x 1	0.66 \pm 0.04	1.91 \pm 0.70
7.	(5'CGG/3'GCC) x 1	0.12 \pm 0.02	0.59 \pm 0.10
8.	(5'CAG/3'GAC) x 1	0.56 \pm 0.07	1.45 \pm 0.52
9.	(5'CCG/3'GGC) x 1	0.86 \pm 0.10	1.5 \pm 0.92
10.	(5'CUG/3'GUC) x 1	0.74 \pm 0.01	24.1 \pm 1.7
11.	(5'CAG/3'CUG) x 1	2.1 \pm 0.12	3.94 \pm 0.84

Table 5.2 The dissociation constant (K_d) values was obtained from flouresecne titration assay of Curcumin with different r(CGG)^{exp} RNA, r(CNG)^{exp} RNA, AU paired RNA and DNA controls.

S. No.	DNA/RNA	K_d^1 (μ M)	K_d^2 (μ M)
1.	r(CGG)1x	0.12 \pm 0.02	0.59 \pm 0.23
2.	r(CGG)2x	0.13 \pm 0.01	2.4 \pm 0.98
3.	r(CGG)3x	0.12 \pm 0.02	1.01 \pm 0.40
4.	r(CGG)4x	0.106 \pm 0.01	1.22 \pm 0.29
5.	r(CGG)6x	0.092 \pm 0.004	0.34 \pm 0.07
6.	r(CGG)20x	0.034 \pm 0.008	1.48 \pm 0.45
7.	r(CGG)40x	0.029 \pm 0.01	0.91 \pm 0.08
8.	r(CGG)60x	0.013 \pm 0.005	0.56 \pm 0.11
9.	r(AU)6x	2.31 \pm 0.36	14.1 \pm 1.6
10.	r(CAG)6x	1.6 \pm 0.180	6.3 \pm 2.9
11.	r(CCG)6x	0.84 \pm 0.05	14.3 \pm 2.5
12.	r(CUG)6x	2.2 \pm 0.3	9.2 \pm 1.7
13.	Yeast t-RNA	3.5 \pm 0.28	13.9 \pm 4.4
14.	Ckit DNA	0.93 \pm 0.17	19.1 \pm 3.1
15.	Cmyc DNA	1.8 \pm 0.36	27.9 \pm 1.7
16.	Bcl2 DNA	1.2 \pm 0.36	33.2 \pm 2.4
17.	Tel22	1.6 \pm 0.08	21.3 \pm 4.3
18.	CT DNA	6.3 \pm 0.48	54.6 \pm 7.2

5.2.2 Isothermal calorimetry titration of Curcumin with CGG repeats RNA

Binding affinity of Curcumin to total yeast t-RNA is in the micromolar range,[24] while we observed a much higher affinity toward CGG repeats RNA (Figure 5.1). To confirm the selectivity and affinity of this interaction, we performed Isothermal calorimetry (ITC) experiments that calculates thermodynamics of non-covalent interaction between two molecules such as nucleic acid-ligand [26]. The thermodynamic parameters such as change in enthalpy (ΔH), change in entropy (ΔS), association (K_a), dissociation constant (K_d) and stoichiometry (N) were derived using Origin 7.0 software (Table 5.3). Our ITC results (Figure 5.5) show negative enthalpy for Curcumin-CGG RNA interaction, which indicate favourable interactions [27]. Furthermore, exothermic peaks throughout titration implies an intercalation mode of binding of Curcumin due to π - π stacking interaction with RNA bases (Figures 5.5 a – d) [28] [29]. The association constant (K_a) values of the highest affinity binding sites for r(CG₆), r(CG₂₀), r(CG₄₀) and r(CG₆₀) RNAs are $1.7 \times 10^6 \text{ M}^{-1}$, $3.8 \times 10^7 \text{ M}^{-1}$, $8.5 \times 10^7 \text{ M}^{-1}$ and $8.6 \times 10^7 \text{ M}^{-1}$, respectively (Table 5.3). These data confirm that increase in repeat length of CGG RNA enhances the binding affinity of Curcumin. As a control, Curcumin showed lower association constant with other mismatched RNAs (r(CAG)₆, r(CCG)₆, r(CUG)₆) and AU paired duplex RNA (Figure 5.5 e – h). Thus, Curcumin binds 24, 535, 1197 and 1211 fold more tightly to r(CG₆), r(CG₂₀), r(CG₄₀) and r(CG₆₀) RNAs, respectively over AU duplex RNA. Overall, ITC experiments complement fluorescence titration binding assays and fully confirmed an interaction of Curcumin with CGG repeats RNAs.

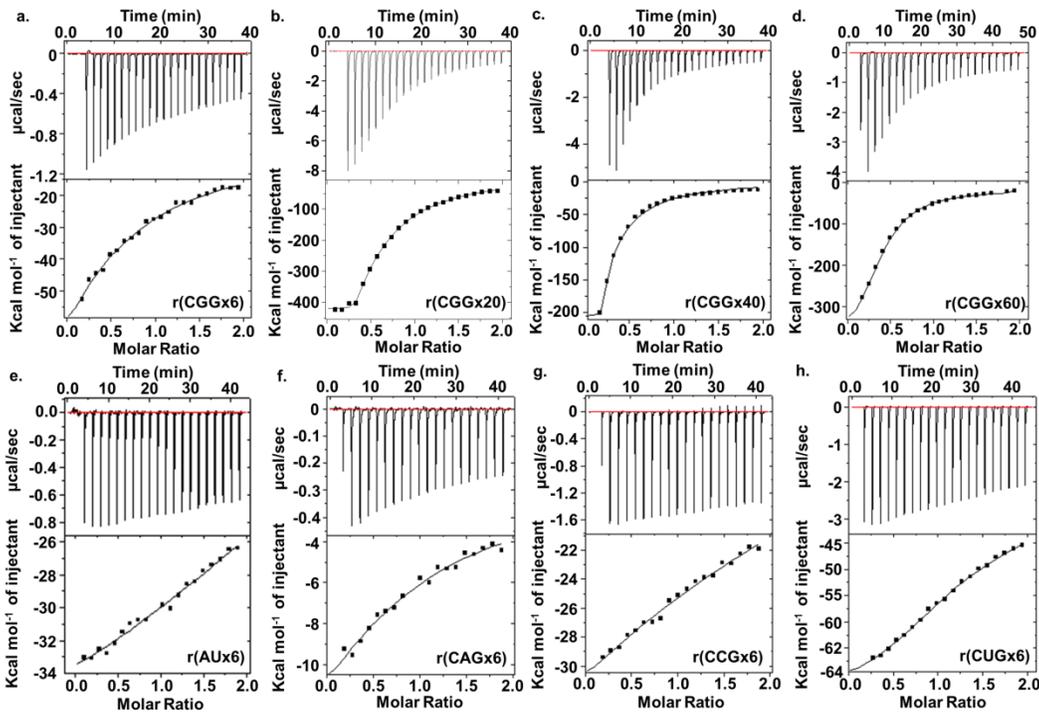


Figure 5.5 Isothermal titration calorimetry of Curcumin with $r(CGg)^{exp}$, $r(CNG)^{exp}$ and $r(AU)$ duplex RNA. a) $r(CGg)6x$, b) $r(CGg)20x$, c) $r(CGg)40x$, d) $r(CGg)60x$ and e) $r(AU)6x$, f) $r(CAG)6x$ g) $r(CCG)6x$, h) $r(CUG)60x$ represents the titrated thermogram of respective RNA with Curcumin. The exothermic peaks represent the favorable contribution of energy during interactions. The following plots were fitted using two-mode binding fitting model.

Table 5.3 Thermodynamic parameter values of different r(CGG)exp RNAs, r(CNG)exp RNAs & r(AU)6 duplex RNA (control) and with Curcumin.

S. No.	Parameters	RNAs			
		r(CGG)6x	r(CGG)20x	r(CGG)40x	r(CGG)60x
1.	N1 (sites)	0.36 ± 0.06	0.34 ± 0.03	0.309 ± 0.08	0.265 ± 0.07
2.	K1 (M ⁻¹)	1.7E6 ± 1.2	3.8E7 ± 1.2	8.5E7 ± 1.4	8.6E7 ± 0.97
3.	ΔH1 cal/mol	-2.1E5 ± 0.48	-4.2E5 ± 0.01	-4.1E5 ± 3.2	-2.7E5 ± 0.73
4.	ΔS1 cal/mol/deg	-693.6 ± 160	-1.3E3 ± 0.04	-393.2 ± 216.5	-906 ± 245.2
5.	N2 (sites)	178.9 ± 160	0.28 ± 0.11	11.1 ± 6.0	6.8 ± 2.5
6.	K2 (M ⁻¹)	1.3E4 ± 7807	1.9E5 ± 0.31	3.2E5 ± 1.5	3.5E5 ± 3.8
7.	ΔH2 (cal/mol)	-7852E7 ± 575	-3.0E6 ± 2.8	-2.4E4 ± 0.53	-3.9E6 ± 4.1
8.	ΔS2 (cal/mol/deg)	-7.6E4 ± 2.2	-3.8E3 ± 1.8	-45.2 ± 35.1	-15.5E5 ± 21
S. No.	Parameters	r(AU)6x	r(CAG)6x	r(CCG)6x	r(CUG)6x
1.	N1 (sites)	0.369 ± 0.11	0.582 ± 0.14	0.75 ± 0.72	0.77 ± 0.20
2.	K1 (M ⁻¹)	7.1E4 ± 0.81	6.8E4 ± 1.8	4.6E4 ± 0.88	3.2 E4 ± 0.46
3.	ΔH1 cal/mol	-4.7E6 ± 4.53E5	-1.7E4 ± 0.37	-3.4E4 ± 1.9	4.6E5 ± 3.8
4.	ΔS1 cal/mol/deg	-2.6E4 ± 1.3E3	-52.1 ± 16.7	-394.5 ± 359.9	5.76E3 ± 0.77
5.	N2 (sites)	0.44 ± 0.05	1.03 ± 0.96	171.5 ± 30.4	0.53 ± .16
6.	K2 (M ⁻¹)	2.5E4 ± 1.0	2.5 ± 0.89	5.2E3 ± 2.2	8.1E4 ± 2.6
7.	ΔH2 (cal/mol)	-1.4E5 ± 0.37	-0.52E5 ± 1.5	-3.3E4 ± 1.7	-83.7 ± 39
8.	ΔS2 (cal/mol/deg)	319 ± 251	1.77E5 ± 0.53	-105.85 ± 75	-673.5 ± 27.5

5.2.3 Gel mobility shift assay, PCR stop assay, and Molecular docking of r(CGG)^{exp} with Curcumin

As further validations of our ITC and fluorescence titration binding assays, we performed electrophoretic gel shift mobility and PCR stop assays. Gel shift performed with 20 μ M r(CGG)^{exp} RNAs incubated at RT for 30 to 60 min with a serial dilution of Curcumin and loaded on agarose gels indicate a migration retardation for all size of r(CGG)^{exp} RNAs (Figures 5.6 a - c). Of interest, higher repeat CGG RNAs such as r(CGG)40x and r(CGG)60x showed maximum migration retardation. In contrast, AU paired duplex RNA showed no gel migration retardation, confirming specific binding of Curcumin with r(CGG)^{exp} RNAs (Figure 5.6 d). In parallel, we also tested Curcumin binding to CGG repeats by polymerase chain reaction (PCR) stop assays. This experiment is based on the hypothesis that if Curcumin binds to CGG repeats, it may impair the polymerase activity during extension. Consistent with this model, a gradual addition of Curcumin decreases intensity of PCR bands (Figure 5.6 e - f). As control, no changes were observed in the intensity of PCR product of AU paired DNA (Figure 5.6 g - h). In addition, molecular docking experiments indicate that Curcumin binding to CGG repeats RNA involves both GxG loop (GxG-2 and GxG-5) with a binding energy of -7.35 and -6.71 kcal/mol, respectively (Figure 5.7). These data confirm an interaction of Curcumin with RNAs containing expansion of CGG repeats.

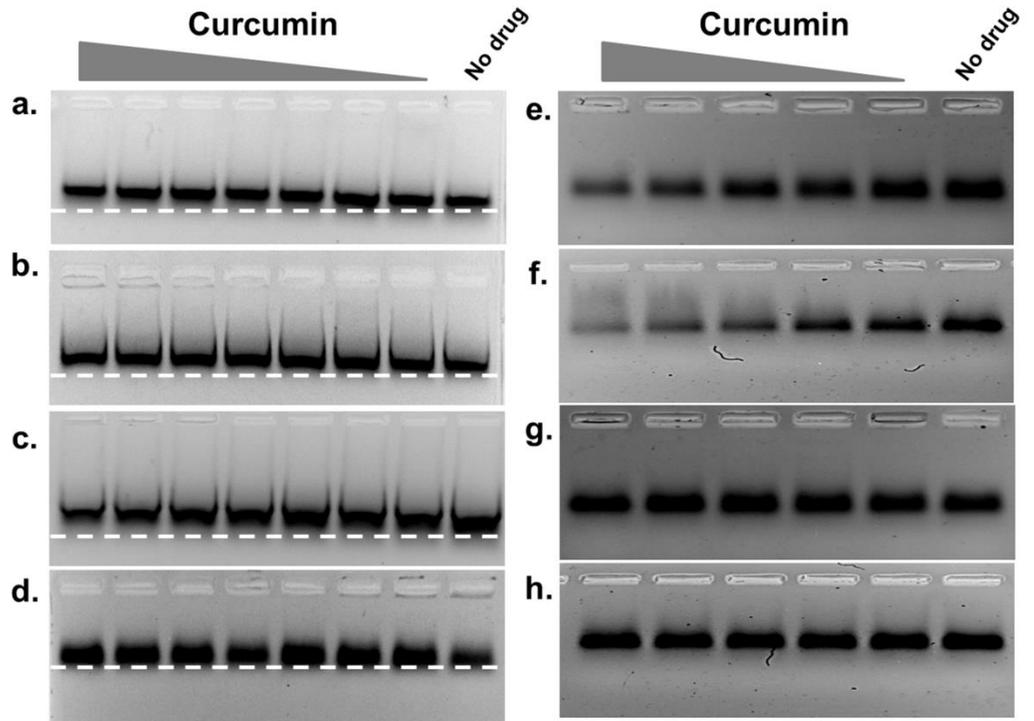


Figure 5.6 Gel mobility shift assay, and PCR inhibition assay. Gel mobility shift assays plot shows significant retardation in mobility of CGG RNA with the increasing concentration of Curcumin; a) r(CGG)20x b) r(CGG)40x and c) r(CGG)60x RNA. In contrast no significant retardation detected in d) r(AU)6x duplex RNA. PCR inhibition assay of CGG DNA template were performed with Curcumin as a function of concentration; decrease in the PCR band intensity suggest the binding of Curcumin with CGG Repeat (e) (CGG)1x and (f) (CGG)6x. In contrast, there is no such inhibition observed in PCR band intensity of AU paired duplex template g) (AU)1x and (h) (AU)6x.

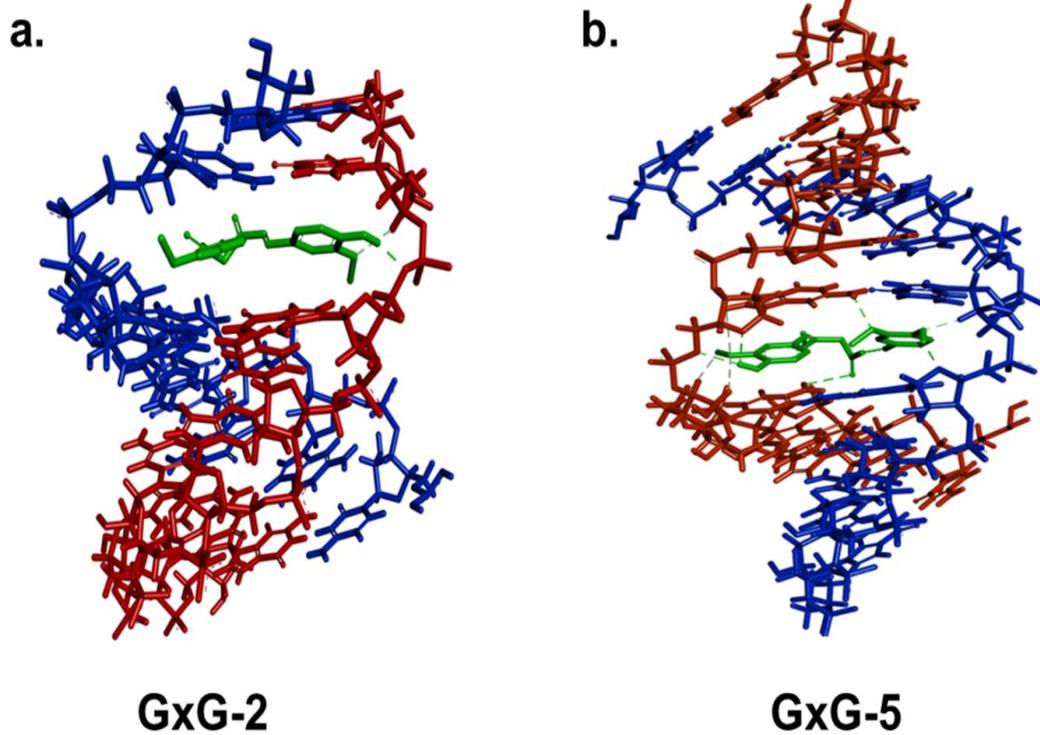


Figure 5.7 Curcumin docking study with CGGx3 RNA (PDB ID: 3JS2). a) Represent the dock image of GxG-2 position. b) Represent the dock image GxG-5 position. The best binding energy of GxG-2 & GxG-5 is -7.35 and -6.71 kcal/mol respectively.

5.2.4 Curcumin induced conformational changes in CGG repeats RNA

Expanded CGG RNA repeats fold into RNA hairpins, which structure can be modified by small molecule ligands [30-32]. Thus, we investigated structural changes in CGG RNA in absence and presence of Curcumin by Circular Dichroism (CD) spectroscopy. Typical CD spectra of r(CG)^{exp} RNAs display positive absorption peak at 265-270 nm and negative absorption peaks at 215-220 nm, which is characteristic of an A type RNA confirmation as previously reported [33]. Gradual addition of Curcumin to CGG RNAs solution causes spectral variations such as hypochromic shift (ellipticity decrease) and red shift in negative peaks in a concentration dependent manner owing to the formation of Curcumin-RNA complex (Figures 5.8 a - e). Of interest, spectral changes were nearly similar with all length of tested r(CG) repeat RNAs, which highlights the similar binding geometry of Curcumin with all r(CG) repeat RNAs (Figures 5.8 a - e). Furthermore, hypochromic shift and red shift are indicative of a $\pi - \pi$ stacking interaction between Curcumin and CGG repeats RNA [29,34]. As a control, no spectral changes were observed in the peaks of r(AU) duplex paired RNAs (Figures 5.8 f - g). Thus, CD analysis confirmed that Curcumin forms a complex with CGG repeat RNAs.

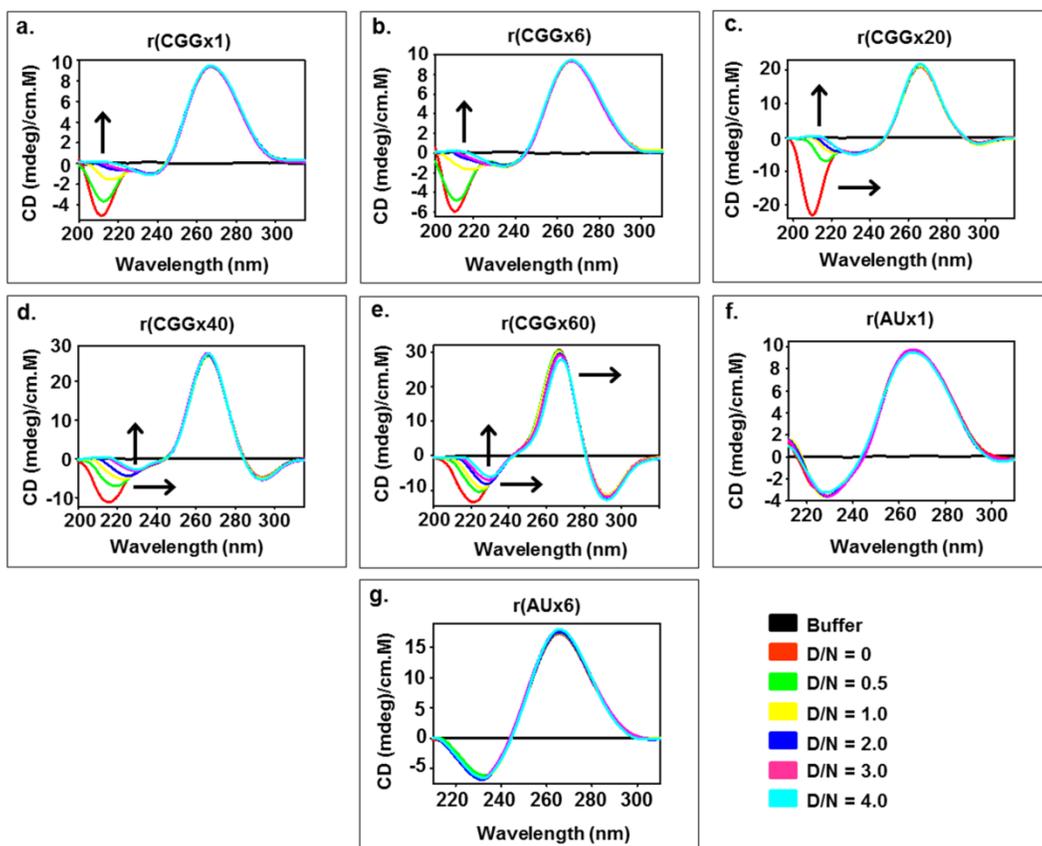


Figure 5.8 Circular Dichroism spectroscopy study of different r(CG) repeats containing RNAs and r(AU) duplex RNA as a function of Curcumin concentration. Plot a) r(CG)1x, b) r(CG)6x, c) r(CG)20x, d) r(CG)40x and e) r(CG)60x shows spectral variation with the increasing concentration of Curcumin. Arrow heads denotes the shifting of negative and positive peaks upon drug titration. D/N denotes drug by nucleotide ratio. Each spectrum was recorded three times and average changes were plotted.

5.2.5 CD thermal denaturation study of CGG repeats RNA with Curcumin

To assess the effect of Curcumin on the thermal stability of CGG repeat RNAs, which could affect CGG associated splicing defects and RAN translation [35], we used CD thermal denaturation assays. CD melting curves were monitored at 267 nm wavelength and the changes in melting temperature (ΔT_m) were of 1.49 °C, 3.48 °C, 4.19 °C, 5.25 °C and 6.28 °C for r(CGG)x1, r(CGG)6x, r(CGG)20x, r(CGG)40x and r(CGG)60x, respectively (Figures 5.9 a – e & g). In contrast, there is no change in melting temperature found as function of Curcumin concentration with AU pair duplex RNA (Figure 5.9 f & g). Interestingly, changes in melting temperature were found larger in RNAs with higher repeat numbers compared to lower repeat RNAs (Figure 5.9 g). In conclusion, Curcumin enhances the thermal stability of all tested CGG targeted RNAs, indicating that Curcumin stabilises the CGG RNA structure, which may prevent ribosomal assembly to initiate RAN translation and/or interaction of the RNA binding proteins that mediate FXTAS-splicing defects.

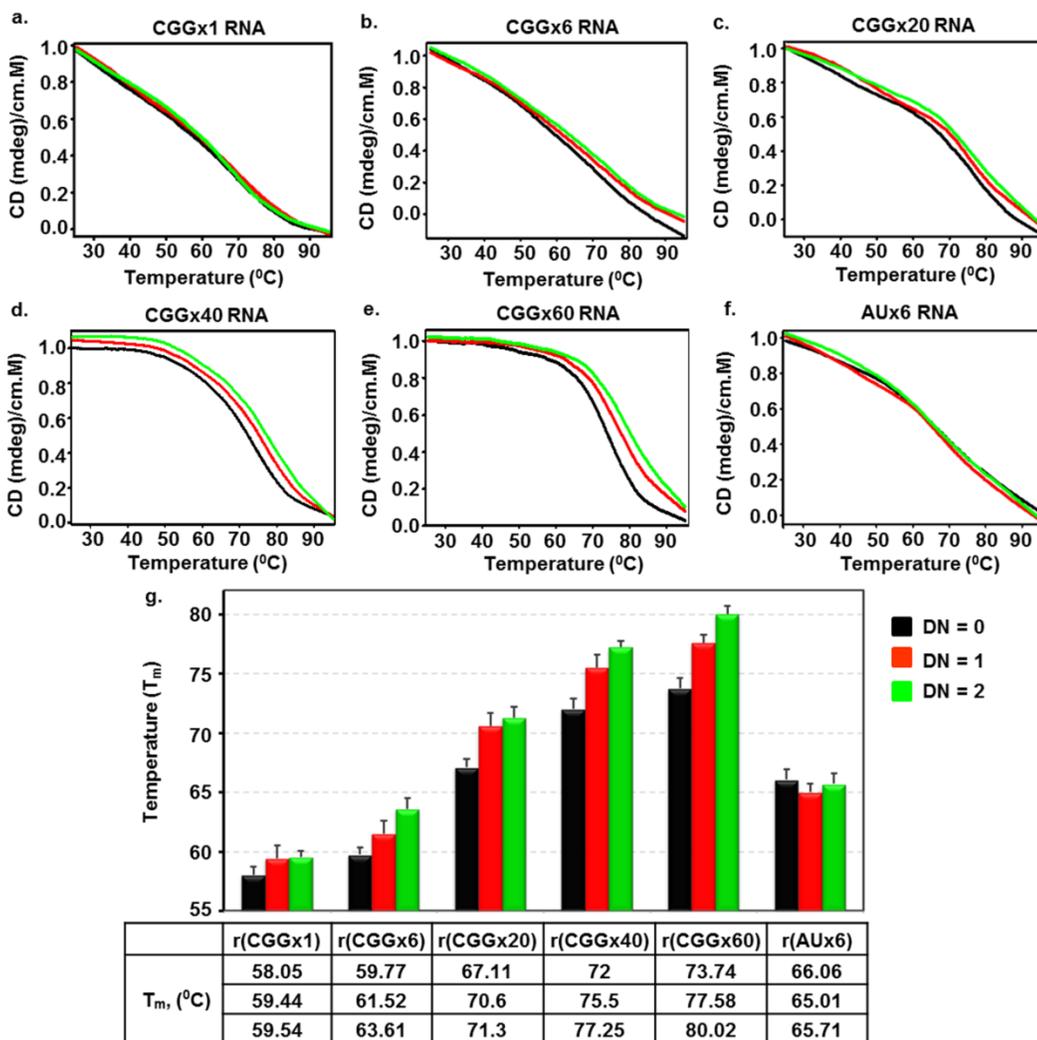


Figure 5.9 Diagram represents CD thermal denaturation profile of r(CG)*n* repeats containing RNAs and r(AU) duplex RNA as a function of Curcumin concentration. Graph shows titration plot of a) r(CG)*1x*, b) r(CG)*6x*, c) r(CG)*20x*, d) r(CG)*40x*, e) r(CG)*60x* and f) r(AU)*6x* duplex RNA with increasing concentration of Curcumin. Bar graph represents the melting temperature (*T_m*) values of respective RNA at different drug by nucleotide ratio (D/N). All the experiments were performed three times and the average values are presented.

5.2.6 Curcumin improves alternative pre-mRNA splicing defects in FXTAS cellular models

Next, we determined cellular potency of Curcumin to improve FXTAS-associated pre-mRNA alternative splicing defects in cultured cell model using survival motor neuron 2 (*SMN2*) and B-cell lymphoma x (*Bcl-x*) minigenes. Previous studies reported that Sam68, a splicing regulator, sequestered by expanded CGG repeat transcripts causes dysregulation of alternative pre-mRNA splicing defect in *SMN2* and *Bcl-x* mRNA [14,30]. In this study, HEK293 cells were co-transfected with plasmid encoding r(CGG)x99 repeat and *SMN2* minigene then increasing concentration of Curcumin were added in the growth media for 24 hours. In the absence of r(CGG)x99, approximately 35% *SMN2* mature mRNA contain exon 7, while in presence of r(CGG)x99 approximately 65% of exon 7 included in *SMN2* mRNA (Figure 5.10 a) [30]. Notably, improvement of *SMN2* splicing defect was observed at a concentration of 25 μ M Curcumin. More significant improvement of splicing defects was observed at higher concentration. Curcumin significantly restored alternative pre-mRNA splicing defects near to wild type (in the absence of r(CGG)^{exp}) pattern of *SMN2* at 100 μ M. In contrast, Curcumin does not affect the alternative splicing defect of *SMN2* minigenes in the absence of CGGx99 expansion (Figure 5.10 a).

Similarly, *Bcl-x* minigene showed two different splicing products *Bcl-xL* and *Bcl-xS*. In FXTAS developed cell model, 68% of *Bcl-x* mRNA showed *Bcl-xL* isoform where as in normal healthy cells, 50% of *Bcl-x* mRNA showed *Bcl-xL* isoform [30]. Curcumin improved missplicing defect of *Bcl-xL* in the concentration dependent fashion. Small improvements of missplicing defect of *Bcl-xL* minigene were observed at 25 μ M (Figure 5.10 b). Further, significant improvement of missplicing was detected at higher concentration (100 μ M). However, it does not significantly alter the pre-mRNA splicing defect of *Bcl-x* minigene with the plasmid lack CGG expansion (Figure 5.10 b).

In addition, control experiments were assessed to check the GFP mRNA expression in presence of Curcumin, till 50 μ M concentration no significant effect were detected (Figure 5.10 c). As further control, cardiac troponin T (*cTNT*) exon 5 minigene was used to investigate the alternative mis-splicing defect, *cTNT* splicing no influenced by the CGG expansion [30]. Curcumin treatment not affects the mis-splicing defect of *cTNT* minigene (Figure 5.11). Thus, our data suggested that Curcumin does not affect global splicing efficiency of genes which is not influenced by CGG expansion. Error bar indicates standard deviation. Statistical significance were calculated by One way ANOVA where * $p < 0.05$; ** $p < 0.01$. Asterisk indicates the statistical difference between control and treated.

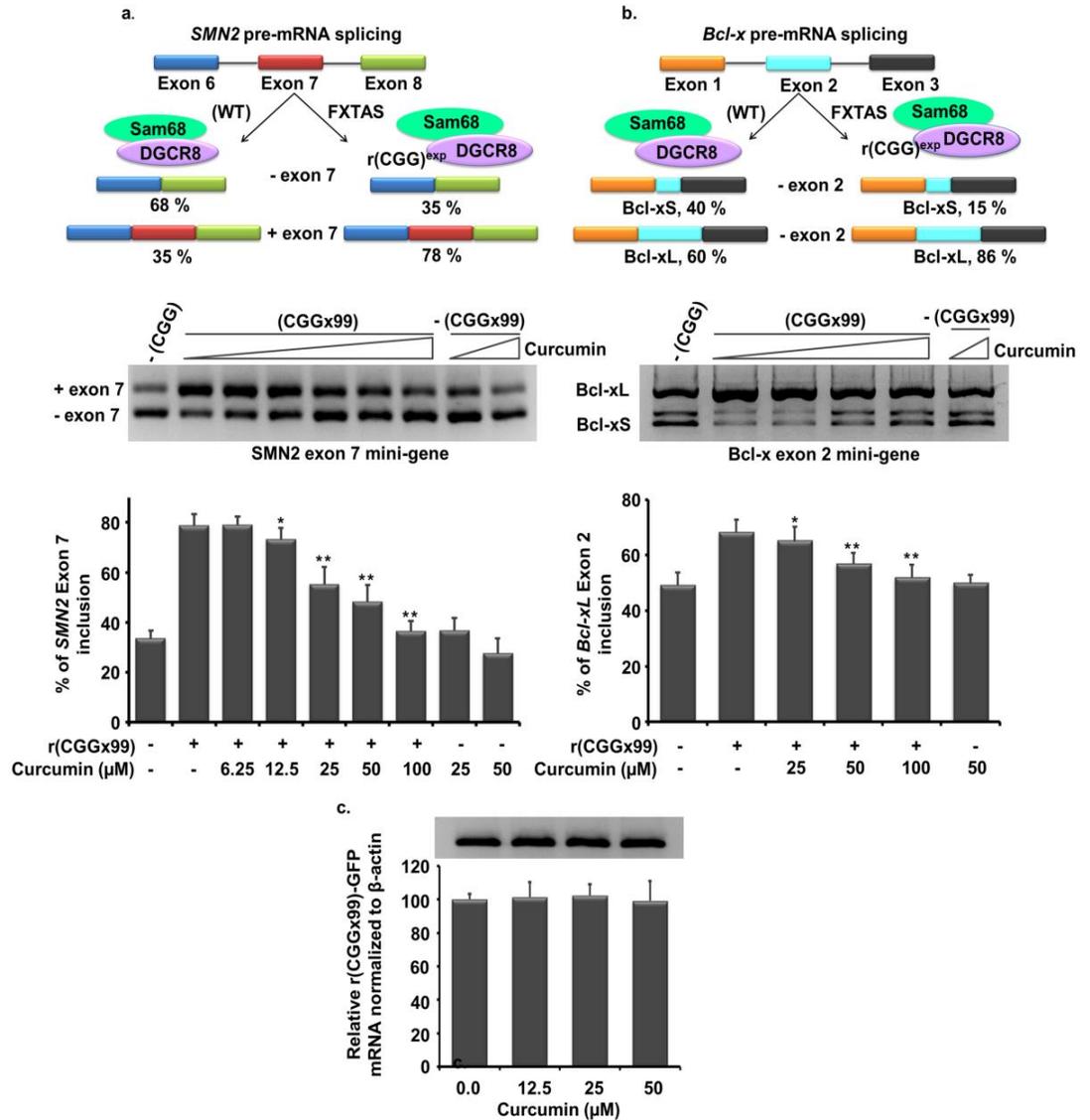


Figure 5.10 *In-vitro* efficacy of Curcumin in FXTAS cellular models. Curcumin improves FXTAS associated alternative pre-mRNA splicing defect of SMN2 and Bcl-x mini gene. Briefly, HEK cell were co-transfected SMN2 & Bcl-x mini gene with plasmid containing CGG repeat (CGGx99). SMN2 splicing regulation is controlled by DGCR8 and Sam68. Upon sequestration and inactivation of these proteins the splicing of SMN2 is dysregulated (a. top, left). When FXTAS cell model system treated with Curcumin, improvement of SMN2 splicing defects were restored to near to wild type at 100 μ M as determined by semi-quantitative PCR. Similarly, Bcl-x gene alternative splicing is regulated by DGCR8 and Sam68 and mis-splicing of Bcl-x gene caused due to presene of r(CGG)^{exp}. were improved with Curcumin treatment in FXTAS cell models (b. top, right). Interestingly, Curcumin does not affect the GFP mRNA expression level (c. bottom).

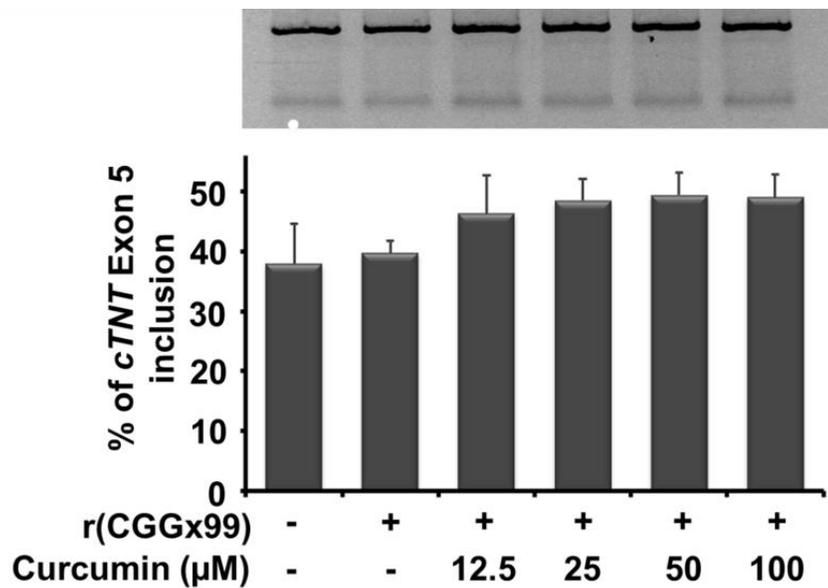


Figure 5.11 Gel image (top) and quantification of alternative pre-mRNA splicing defect of cTNT mini-gene in CGGx99 transfected and healthy cells as a function of Curcumin concentration. Briefly, cTNT and CGGx99 plasmid (toxic) co-transfected in HEK293 cells. curcumin does not affect alternative splicing defect of cTNT mini-gene.

5.2.7 Curcumin reduces RAN translation of the CGG repeats

In addition to correction of splicing defects, we also determined the potency of Curcumin to inhibit the pathogenic non-canonical translational of expanded CGG repeats. Expansions of CGG repeats in *FMRI* are translated into FMRpolyG, a small glycine-rich protein that form intranuclear inclusions, which is an hallmark of FXTAS. HEK293 cells were transfected with a plasmid containing 99 CGG repeats embedded in the human 5'UTR *FMRI* sequence and cloned in the glycine frame with the GFP protein. This construct expresses the FMRpolyG protein from its natural ACG near-cognate start codon¹⁷. Importantly, Curcumin showed a potent reduction in expression of the FMRpolyG proteins with ~60% and ~90% of RAN translation inhibited at 12.5 μ M and 25 μ M of Curcumin, respectively (Figure 5.12 a – e). As a control, Curcumin does not modify translation of the GFP protein from a control plasmid lacking CGG repeats (Figure 5.12 c and Figure 5.13). In addition, protein aggregate quantifications were performed manually in Curcumin treated condition; approx. 40% and 50% of RAN translation were inhibited at 50 μ M and 100 μ M respectively (Figure 5.12 e). FXTAS is a neurodegenerative disease characterized by neuronal cell loss. Thus, we tested whether Curcumin can prevent cell toxicity induced by expanded CGG repeats expression.

Interestingly, Curcumin restores normal cell viability of HEK293 cells expressing 99 CGG repeats embedded within the 5'UTR of *FMRI* (Figure 5.12 f). As controls, we assessed the cytotoxicity of Curcumin in HEK293 cells and normal GM00357 fibroblasts as well as in FXTAS patient derived fibroblasts (GM04026), but found no overt cytotoxicity of Curcumin (Figure 5.14). Furthermore, cellular intake of Curcumin was also assessed after 24hrs in patient derived cells. Confocal microscopy image suggest the Curcumin going inside the cells (Figure 5.15). Taken together these results suggest that Curcumin binds specifically to CGG repeats and ameliorate cell viability in FXTAS cell models without presenting overt toxicity. Error bar denotes standard deviation. One way ANOVA was utilized for determination of statistical significance except cell

viability enhancement (f) assay where student t test were performed. *p< 0.05;
p < 0.01; *p < 0.001.

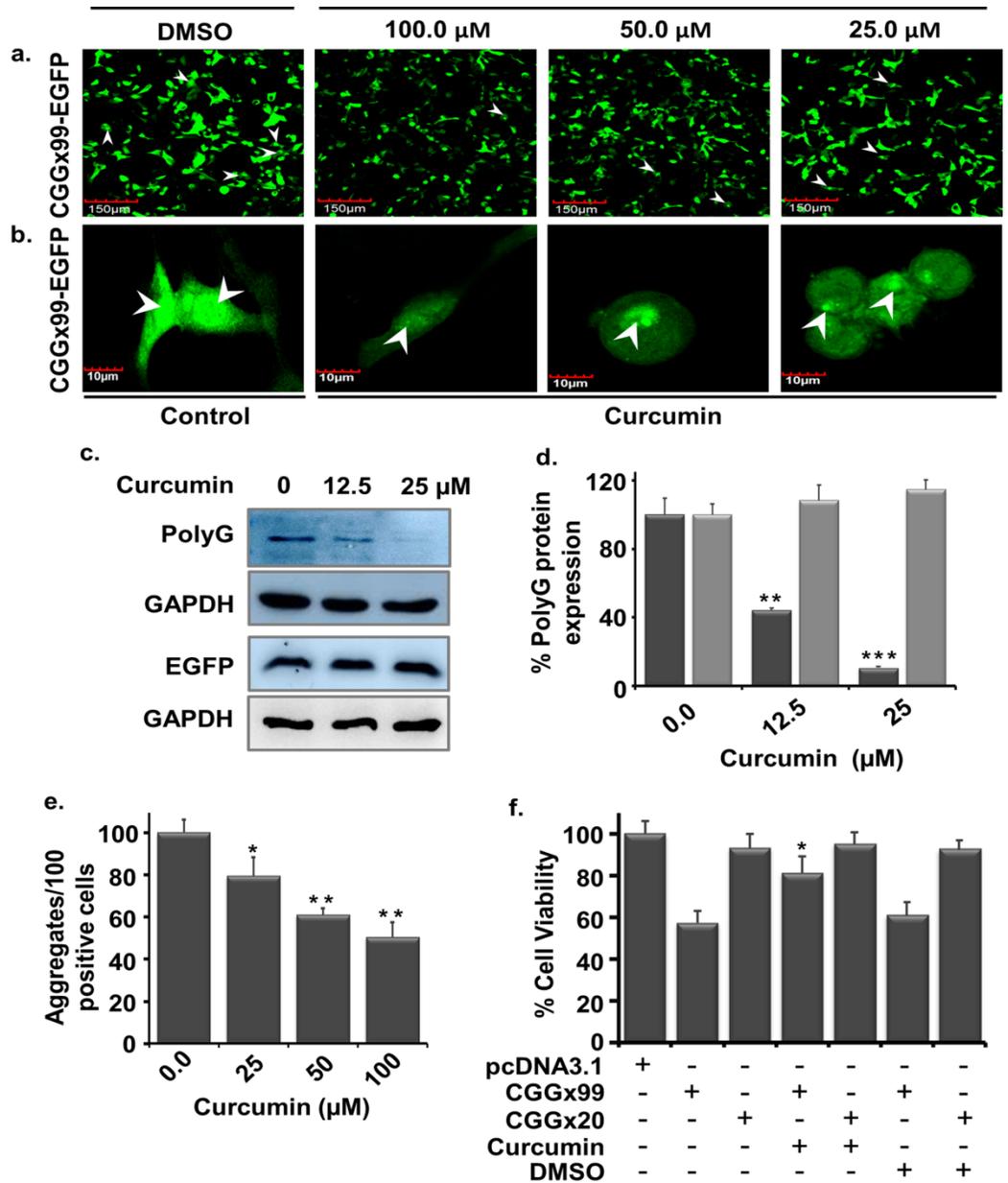


Figure 5.12 Diagram shows *in-vitro* efficacy of Curcumin in FXTAS cell models. In FXTAS, CGG expansion triggers the non-canonical protein translation leads to protein aggregate formation inside the cell. (a-b) Micrograph showing the effect of Curcumin treatment in transiently transfected ((CGGx99)-EGFP) HEK cell. (a) 20x image and (b) 60x image of DMSO (control) and Curcumin for 24hrs. (c) Blot showing the reduction of polyG aggregates formation as a function of Curcumin concentration with control GAPDH and EGFP plasmid (lack CGG repeat). (d) Bargraph showing the percentage reduction of polyG aggregate in presence of Curcumin. (e) Plot denotes Curcumin treatment reduces number of protein aggregates in the cell. (f) Bar graph showing improvement of cell viability of CGGx20-pcDNA3.1, CGGx99-GFP-pcDNA3.1 and GFP-pcDNA3.1 plasmid transfected cell (HEK293) in presence of Curcumin.

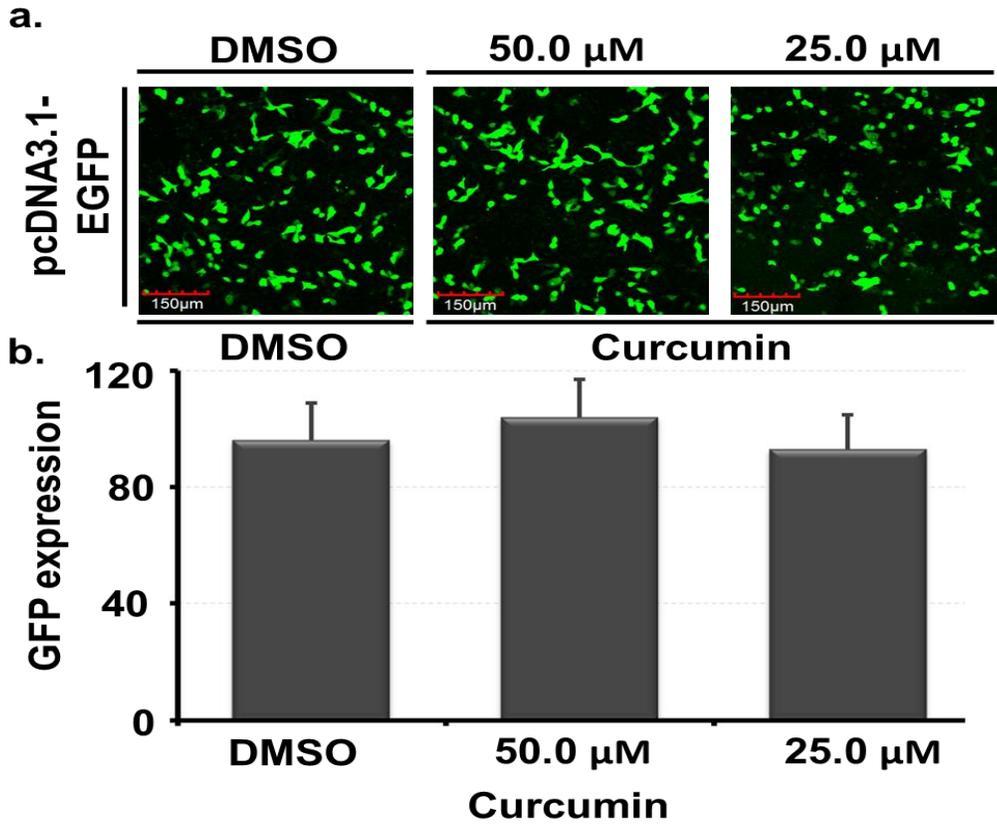


Figure 5.13 (a) Diagram showing the EGFP expression of pcDNA3.1 plasmid (lack $r(\text{CGG})^{\text{exp}}$) of HEK293 transfected cell in absence and presence of Curcumin. (b) Bar graph represents the normalized EGFP intensity of Curcumin treated and untreated cells.

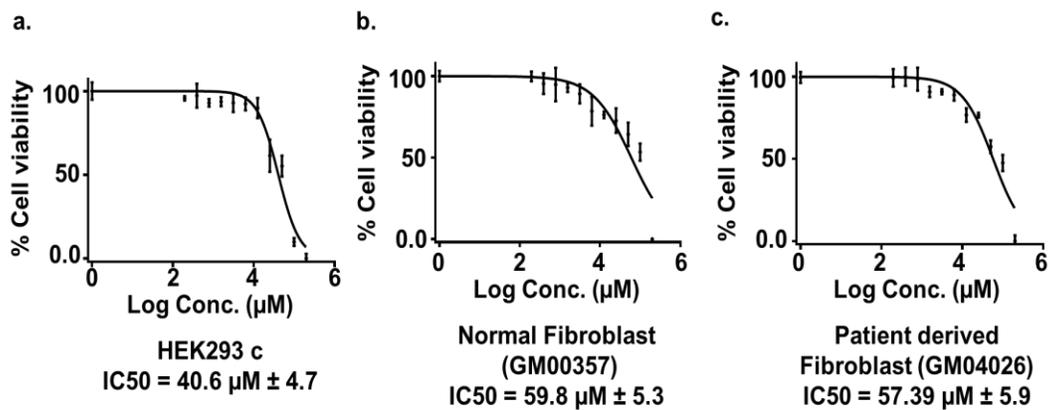


Figure 5.14 Systemic representation of cell viability plot of Curcumin on established cell line for 24 hrs.

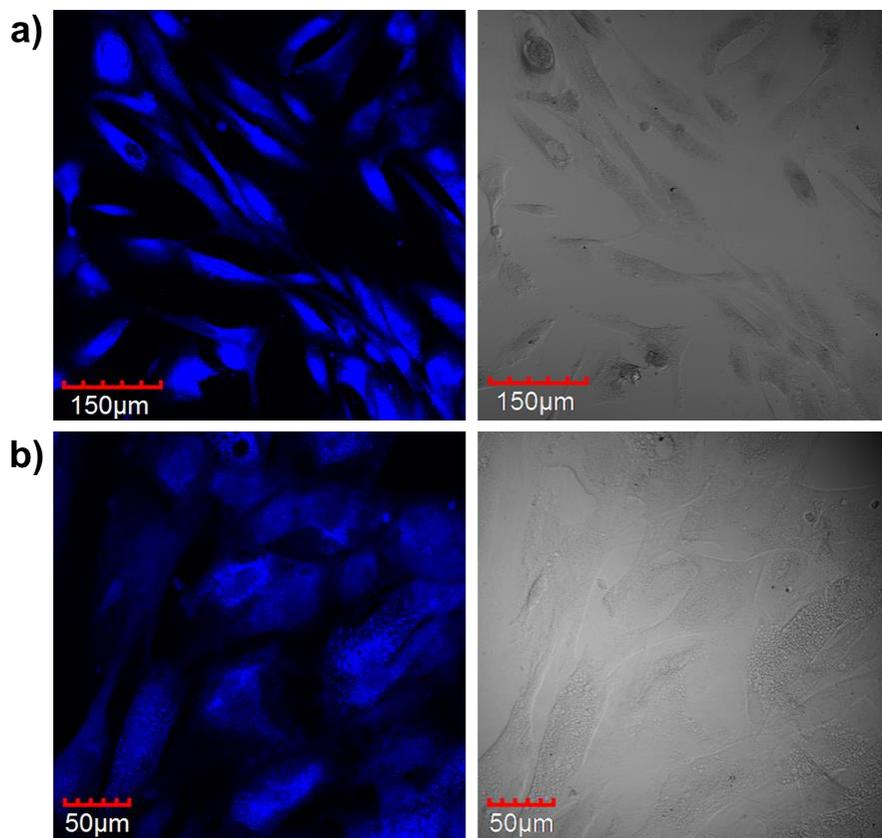


Figure 5.15 Cellular intake of Curcumin in patient derived cells (GM04026). a) 20x magnification b) 40x magnification.

5. 3 Material and Methods

5.3.1 Reagents

Curcumin molecule and other reagent used for the buffer preparation such as NaCl, KCl, NaH₂PO₄, Na₂HPO₄, KH₂PO₄ and K₂HPO₄ were purchased from Sigma Aldrich chemical Pvt. Ltd. The Curcumin molecule was used without further purification. The solvent reagent HPLC grade such as dimethyl sulfoxide (DMSO), 2-butanol, ethanol, deuterium oxide, deuterated DMSO were procured from sigma Aldrich. Midiprep plasmid isolation kit purchased from Invitrogen. All the PCR reagent including Taq polymerase, dNTP mix, primers, DNA template (d(CGG)₁₋₆ repeats) were also obtained from sigma Aldrich. Invitro transcription kit (MEGAscript® T7 Kit) were purchased from Thermo Fisher Scientific pvt. Ltd. For CGG hairpin motif formation, CGG repeat RNAs were dissolved in phosphate buffer (10 mM (K⁺), pH 7.0) with 50 mM KCl. The CGG RNAs were annealed by heating at 90 °C for 5 - 10 mins, and then allowed to cool down at the room temperature for 1-2 hrs. All the biophysical experiment was carried out using above mentioned buffer composition otherwise mentioned separately. For the cell culture experiment, growth media (Dulbecco's Modified Eagle Medium (DMEM), Fetal bovine serum (FBS), Phosphate buffer saline (PBS), Antibiotic solution (ABAM) were purchased from Gibco. The Human Kidney cell line (HEK293) was purchased from National Centre for Cell Science (NCCS), Pune, India.

5.3.2 RNA Preparation and Purification

RNA sequences used for biophysical studies and gel retardation assay were prepared by run off transcription method using synthetic DNA template as previously reported.[36] Briefly, synthetic DNA templates either amplified by PCR or cloned in a plasmid, were transcribed using T7 RNA polymerase, and transcribed products were purified by denaturing 15% PAGE. After UV

shadowing, RNAs were extracted using 0.3 M NaCl by tumbling down for overnight at 4 °C.

5.3.3 Fluorescence titration assay

The Fluorescence titration assays were performed on Synergy™ H1 multi-mode microplate reader using corning 96 well microplate at 25 °C. The excitation and emission wavelength of Curcumin were calculated in 1x potassium phosphate buffer by performing its absorption and fluorescence scan. Each sample was measured in duplicates in 75 µl. The various RNA motifs (5'CNG/3'GNC) and different repeat r(CGG)^{exp} RNAs and (5'CAG/3'GUC), control at the final concentration of 4 µM were serially diluted and last well taken as blank (No RNA). Similarly, RNAs 5'CAG/3'GUC duplex pair (4 µM), yeast t-RNA (50 µM) and different DNAs like CT-DNA, cymc DNA, bcl2 DNA (50 µM) was served as control. For the detection of fluorescence change of Curcumin with RNA titration, Curcumin excited at the emission wavelength of 341 nm and reading were taken at the emission wavelength of 496 nm. The obtained data were analyzed using Sigma Plot 12.0 software (Systat Software, Chicago, USA) according to the following equation.

$$f = \frac{B_{max}^1 \times \text{abs}(x)}{k_d1 \times \text{abs}(x)} + \frac{B_{max}^2 \times \text{abs}(x)}{k_d2 \times \text{abs}(x)}$$

B_{max} = maximum number of binding sites. K_d = equilibrium binding constant

5.3.4 Circular Dichroism spectroscopy assay

Circular dichroism experiments were performed on J-815 Spectropolarimeter (JASCO). To maintain the constant temperature 298K, during entire experiment, peltier junction temperature controller was equipped. A constant stream of dry nitrogen with the rate of 5L/min, flushed into the cuvette-holding chamber during whole experiment to avoid the condensation of water droplet outside the cuvette. Spectrums were recorded using the quartz cuvette with 0.2 cm path length containing 10.0 µM CGG and AU RNA samples and titration were done with

increasing concentrations of Curcumin in 50.0 mM KCl, 10.0 mM phosphate buffer (K⁺) at pH 7.0. Scan rate 20 nm/min, 0.1 nm intervals from 200 nm to 350 nm with a 1 nm-slit width and averaged over three scans. Lastly, CD spectrum of RNA and RNA-Curcumin complex plotted after subtraction with the same buffer.

5.3.5 RNA thermal denaturation study

RNA thermal denaturation experiments were carried out using J-815 Spectropolarimeter (JASCO) equipped with Peltier temperature programmer and water Peltier system PCB-1500. CGG and AU paired RNA (10.0 μ M) were diluted in 1x potassium phosphate buffer (50.0mM KCl, 10.0mM phosphate buffer (K⁺) and heated at 90-92°C for 5-10 min then allow it to cool down at room temperature for next 30 min before start of the experiment. The Melting curve of RNAs were monitored from 25 °C to 95 °C at the rate of 1 °C/min with the titration of Curcumin till drug/nucleic acid ratio 2.0. The normalized absorbance changes at 260 nm against temperature were plotted using the SigmaPlot 12.0 software.

5.3.6 Isothermal titration calorimetry experiment

The Isothermal titration calorimetry (ITC) measurements were carried out using a MicroCalTM isothermal titration calorimeter iTC200 (GE healthcare) at a constant temperature of 298K. RNA sample were prepared in potassium buffer, 50 mM KCl, 10 mM potassium phosphate, 50 mM EDTA (K⁺) at pH 7.0. The ITC cell was filled with 20 μ M RNA whereas 200 μ M Curcumin sample was filled in syringe in the same buffer. 1.86 μ L of the Curcumin added at the each step to the sample with initial 0.4 μ L injection having 60 sec initial equilibrium delays. Total 21 injections were used with 120 sec pause to homogenous the heat between injections and reference power kept it 6 μ cal/s. The samples stirred speed during the entire titration experiment was kept 750 rpm. The heats of dilution were also

calculated by titrating same concentration of Curcumin sample with the same buffer only. The binding thermogram of Curcumin RNA complex was plotted by two site binding model to determine the dissociation constant (K_d) after the subtracting with the heat of dilution isotherm. Other thermodynamic parameters were also calculated using MicroCal Origin software[37].

5.3.7 Molecular Docking

The crystal structure of duplex CGG motif (PDB ID: 3JS2) were used for the docking.[38] The addition, replacements of residues and optimization of duplex RNA structure with Curcumin were built in Discovery studio 3.5 (Accelrys Inc.,USA) using CHARMM force field. Curcumin docking studies were performed on Autodock 4.0 where duplex RNA treated as solid rigid structure, RNA and Curcumin file prepared in PDBQT format. The other parameters kept as a default values during docking. CGG duplex RNA and Curcumin were converted to AD4 format files and Gasteiger charges were assigned to the atoms. The grid box was established in such way which covers the complete RNA structure so that Curcumin can explore the whole conformational space. The grid centre was placed by centering the grid box so that both minor groove or intercalation site utilise equally. Lamarckian genetic algorithm [39]was run to carry out energy evaluation of the location of the ligand according to target energy grid and the results were analysed based on binding energy. The best dock result was further processed to prepare image in Discovery studio 3.5[40].

5.3.8 Gel retardation assay

Gel mobility shift assay of Curcumin were performed with variable length r(CGG)^{exp} RNA. 10 μ M of (CGGx20/40/60) RNAs were prepared in potassium phosphate (50mM KCL) buffer prior to heating at 95 °C then allow to cooling down at room temperature. The reaction mixture incubates with the increasing concentration of Curcumin (0.0 to 5mM) for 30 min at room temperature. The

products were resolved on 3% agarose gel prepared in 1X TBE buffer and stained by ethidium bromide. Gel image was visualised and analysed using ImageQuant LAS 4000 (GE Healthcare)[41].

5.3.9 PCR stop assay

The PCR stop assay were performed according to previously describe protocol.[40] Template for GGx1 (5'- GGAGAGGGUUUAAUCGGUACGAAA GUACGGAUUGGAUCCGCAAGG-3'), GGx6 (5'- GGAGAGGGUUUAAUCG GCGGCGGCGGCGGCGGUACGAAAGUACGGCGGCGGCGGCGGCGGAU UGGAUCCGCAAGG-3') and complementary sequence (5'- GGCCGGATC CTATAATACGACTCACTATAGGGAGAGGGTTTAAT-3') were procured from Integrated DNA Technologies pvt ltd. For each concentration of drug final reaction volume of assay is 25 µl containing reaction cocktail includes 1X PCR buffer, 4.25 mM MgCl₂, 10 pmol oligonucleotide, 0.33 mM dNTPs, 2.5 units Taq DNA polymerase (sigma aldrich) and volume make up by milli Q water. Curcumin was serially added from 0.0 to 700 µM then reaction mixtures were incubated in thermo cycler (Eppendorf) with following conditions: 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 minute and finally held at 4 °C following completion. The obtained PCR product was mixed with 6X DNA loading dye and resolved on 3 % agarose gel stained with ethidium bromide. Gel image was analyzed on ImageQuant LAS 4000 (GE Healthcare)[41].

5.3.10 Improvement of splicing defect by Curcumin in FXTAS cellular model

In order to determine whether Curcumin improve pre-mRNA alternative splicing defect *in-vitro*, a FXTAS cellular model were used. Briefly, HEK293 cells were grown in 24 well plates as monolayer in the growth medium contain 1X DMEM, 10% fetal bovine serum, 1% antibiotic and antimycotic at 37 °C with 5% CO₂. After cell reached 80-90% confluency, HEK cell were transfected with the equal

amount of plasmid contain CGGx99 repeat and targeted mini-genes (*SMN2* and *Bcl-x*) using Lipofectamine 3000 reagent (Thermo Fisher Scientific), according to standard manufacturers protocol. Transfection cocktail was removed after 4-5 hours and Curcumin containing growth medium were added. After 24 hours, cells were lysed in the plate and total RNA were recovered using RNA isolation kit (Invitrogen), according to standard manufacturers protocol.

All different conditioned RNA sample were subjected to reverse transcribe (RT-PCR) using cDNA synthesis kit from Bio-Rad, followed standard protocol. Out of 500 ng reverse transcribed mRNA 100 ng were subjected for semi-quantitative PCR. The PCR protocol used: denaturation for 95 °C for 1 min, annealing 55 °C for 1 min, extension at 72 °C for 2 min and final extension at 72 for 10 min. The PCR protocol was run for 25-30 cycles. The PCR product were analysed on agarose gel electrophoresis, stained with ethidium bromide and image were recorded using ImageQuant LAS 4000 (GE Healthcare). The splicing isoform intensity was quantified using imageJ software. The primer sequence mentioned in the table S4 for each construct[14]. Two control condition were used to study pre-mRNA splicing defect. 1) Targeted mini-genes (*SMN2* and *Bcl-x*) were co-transfected with plasmid that lack CGG repeat at 5'UTR as describe above. 2) Co-transfection of CGG repeat plasmid with mini-genes whose pre-mRNA splicing not govern by sam68 (cTNT)[14,30].

Table 5.4 PCR primer sequences for splicing defect.

S no.	Gene	Forwards primer	Reverse primer
1.	SMN2 mini-gene	5'GGTGTCCACTCCC AGTTCAA	5'GCCTCACCACCGTGCTG G
2.	Bc l-x mini-gene	5'GGAGCTGGTGGTT GACTTTCT	5'TAGAAGGCACAGTCGA GG
3.	cTNT mini-gene	5'GTTCACAACCATC TAAAGCAAGATG	5'GTTGCATGGCTGGTGCA GG
4.	β-Actin	5'CCTGGCACCCAGC ACAAT	5'GGGCCGGACTCGTCATA C
5.	(CGG)₉-GFP	5'GCACGACTTCTTC AAGTCCGCCATGCC	5'GCGGATCTTGAAGTTCA CCTTGATGCC

5.3.11 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of Curcumin by using MTT assay

In order to determine the cytotoxic effect of Curcumin in vivo, MTT (3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyltetrazolium bromide dye) assay were performed on normal cell line (HEK293 and GM00357) as well as patient derived cell line (GM04026). Both normal and patient derived cell line ($5 - 10 \times 10^3$ cells/well) were plated in the 96 well culture plates as mono layer in triplicate and allowed to grow in the complete media (1X DMEM, 10% FBS), at 37 °C with 5% CO₂. After cell were attached properly and reached to optimum confluence, cell were treated with Curcumin (200 μM to 0.2 μM) with DMSO taken as a vehicle for 24 hrs. After 24 hrs incubation, 10 μL of 5 mg/mL MTT in PBS was added in each well and further incubated for next four hours at 37 °C. The insoluble yellow coloured MTT into dark coloured formazon crystals by intracellular reduction. 100μL DMSO were added to dissolve the formazon crystal structure and absorbance was taken at 570 nm using a microplate reader (Synergy™ H1 multi-mode microplate reader). Half-maximal inhibitory concentration (IC₅₀) of the Curcumin was determined by the using formula.[40] In addition, HEK cell were plated to check the cell viability in CGGx₉₉ plasmid transfected cell. Cell were transfected with pcDNA-EGFP plasmid (control), CGGx₂₀ normal repeat plasmid and CGGx₉₉ diseased plasmid using lipofectamine 3000 reagent from Thermo Fisher Scientific as per manufacturer protocol. Curcumin treatment was given after 4hrs incubation of transfection for 24 hrs and similar MTT protocol were followed to calculate cell viability[41].

$$\% \text{ inhibition} = \frac{\text{Control absorbance} - \text{sample absorbance}}{\text{Control absorbance}} \times 100$$

5.3.12 Quantification of RAN translation by western blot

HEK293 cells were seeded in 6 well plate in amonolayer form and transfected with 2.5μg of plasmid coding CGGx99-GFP using lipofectamine 3000 as per manufacturerer protocol from Thermo Fisher Scientific. The transfection cocktail

was removed after 5 hrs incubation and added fresh Compound containing media and the cell were incubated for 24hrs at 37 °C. After incubation, cells were lysed using RIPA buffer 200 µL/well containing 1 µL of Halt Protease Inhibitor Cocktail from Sigma. Cellular proteins concentration were calculated using Bradford assay. Same amount of protein were loaded in SDS-PAGE for separation and transferred to a PVDF membrane. Blot were incubated with antiFMR1-polyG antibody from Merck Millipore as a primary antibodies. After primary incubation anti-IgG-horseradish peroxidase conjugate used as the secondary antibody. A chemiluminescent signal was detected using Luminata Crescendo Western HRP substrate from Merck Millipore in ImageQuant LAS 4000 (GE Healthcare).

5.3.13 Quantification of canonical GFP and non canonical FMRpolyG-GFP protein Inclusions

In order to determine the effect of Curcumin on polyG-GFP inclusions two different constructs were used that contain r(CGG)_{x99}-GFP and pcDNA-GFP. The above mentioned construct was designed and validated by Nicolas group earlier [14]. Briefly, HEK cells were seeded in six well plate in 10% DMEM media. After cells were reached 80-90 percent confluency transfection was done with above mentioned plasmid using lipofectamine 3000 reagent from Thermo Fisher Scientific as per manufacturer standard protocol. The transfection cocktail was removed after 4hrs incubation and fresh media were applied which contain respective Curcumin concentration and the cells were incubated for 18-24 hrs at 37°C. Cells after washing with PBS, fixed with 4% paraformaldehyde for 15 minutes. Images were captured and protein aggregates visualized at higher magnification using fluorescence confocal microscopy and processed to remove background signals. 100 positive transfected cells from each well were selected and manually graded as “with protein aggregates” and “no aggregates”. The number of GFP protein inclusions were manually counted from three independent experiments and standard deviation was calculated [41].

5.4 Conclusion

There is a strong need of effective and safe therapeutic modalities for FXTAS. Previous studies have reported that stabilisation of CGG repeat RNA via small molecule ameliorates FXTAS associated pathogenic defects [35]. Recently, our group has reported a small molecule that thermodynamically stabilise 5'CAG/3'GAC repeats RNA hairpin structure which prevent polyQ aggregation in Huntington disorder [41] [42]. Herein, we have demonstrated that Curcumin binds to r(CG_G)^{exp} and potently reverses FXTAS-associated pathogenic defects. Biophysical characterization using CD spectroscopy, CD melting and gel retardation illustrate the specific binding of Curcumin with CGG repeats RNA. Importantly, Curcumin improves pre-mRNA *SMN2* and *Bcl-x* alternative splicing defect caused by expression of CGG expansions. Similarly, Curcumin also prevents translation of the CGG repeats into the toxic FMRpolyG protein. Curcumin may inhibit translation through two potential mechanisms. First, it has been reported that longer CGG expansions (thermodynamic more stable) or small molecule enhancing thermal stability of r(CG_G)^{exp} correlates with their inhibitory effect on RAN translation [41-43] as RNA hairpin stabilization impedes ribosomal complex *read* through to mRNA. A second model is that binding of compounds to CGG expansions can simply occupy the RNA and prevent ribosomal machinery to start translation. Similarly, Curcumin binding to CGG repeats can prevent further association and thus titration of RNA binding proteins that leads to FXTAS-associated splicing defects. In conclusion, our data suggest that Curcumin could be used as a potential candidate for the therapeutics of FXTAS. However, further studies involving animal models are required to reach any conclusion.

5.5 Reference

1. Sharp, P. A. (2009), The Centrality of RNA, *Cell*, 136, 577-580 (10.1016/j.cell.2009.02.007)
2. Calin, G. A. and C. M. Croce. (2006), MicroRNAs and chromosomal abnormalities in cancer cells, *Oncogene*, 25, 6202-6210 (10.1038/sj.onc.1209910)
3. Wilton, S. D. and S. Fletcher. (2005), RNA splicing manipulation: strategies to modify gene expression for a variety of therapeutic outcomes, *Curr Gene Ther*, 5, 467-483
4. Groh, M., M. M. P. Lufino, R. Wade-Martins and N. Gromak. (2014), R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome, *PLOS Genetics*, 10, e1004318 (10.1371/journal.pgen.1004318)
5. (1993), A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group, *Cell*, 72, 971-983
6. Nelson, D. L., H. T. Orr and S. T. Warren. (2013), The unstable repeats--three evolving faces of neurological disease, *Neuron*, 77, 825-843 (10.1016/j.neuron.2013.02.022)
7. Cooper, T. A., L. Wan and G. Dreyfuss. (2009), RNA and Disease, *Cell*, 136, 777-793 (10.1016/j.cell.2009.02.011)
8. Klesert, T. R., A. D. Otten, T. D. Bird and S. J. Tapscott. (1997), Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP, *Nat Genet*, 16, 402-406 (10.1038/ng0897-402)
9. Orr, H. T. and H. Y. Zoghbi. (2007), Trinucleotide repeat disorders, *Annu Rev Neurosci*, 30, 575-621 (10.1146/annurev.neuro.29.051605.113042)
10. Hagerman, P. J. and R. J. Hagerman. (2015), Fragile X-associated tremor/ataxia syndrome, *Ann N Y Acad Sci*, 1338, 58-70 (10.1111/nyas.12693)

11. Hagerman, R. J. and P. Hagerman. (2016), Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management, *Nat Rev Neurol*, 12, 403-412 (10.1038/nrneurol.2016.82)
12. Coffey, S. M., K. Cook, N. Tartaglia, F. Tassone, D. V. Nguyen, R. Pan, H. E. Bronsky, J. Yuhas, M. Borodyanskaya, J. Grigsby *et al.* (2008), Expanded Clinical Phenotype of Women With the FMR1 Premutation, *American journal of medical genetics. Part A*, 146A, 1009-1016 (10.1002/ajmg.a.32060)
13. Berry-Kravis, E., L. Abrams, S. M. Coffey, D. A. Hall, C. Greco, L. W. Gane, J. Grigsby, J. A. Bourgeois, B. Finucane, S. Jacquemont *et al.* (2007), Fragile X-associated tremor/ataxia syndrome: clinical features, genetics, and testing guidelines, *Mov Disord*, 22, 2018-2030, quiz 2140 (10.1002/mds.21493)
14. Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott *et al.* (2010), Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *Embo j*, 29, 1248-1261 (10.1038/emboj.2010.21)
15. He, F., A. Krans, B. D. Freibaum, J. P. Taylor and P. K. Todd. (2014), TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1, *Hum Mol Genet*, 23, 5036-5051 (10.1093/hmg/ddu216)
16. Sofola, O. A., P. Jin, Y. Qin, R. Duan, H. Liu, M. de Haro, D. L. Nelson and J. Botas. (2007), RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS, *Neuron*, 55, 565-571 (10.1016/j.neuron.2007.07.021)
17. Sellier, C., F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page *et al.* (2013), Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA

- processing in fragile X-associated tremor/ataxia syndrome, *Cell Rep*, 3, 869-880 (10.1016/j.celrep.2013.02.004)
18. Iwahashi, C. K., D. H. Yasui, H. J. An, C. M. Greco, F. Tassone, K. Nannen, B. Babineau, C. B. Lebrilla, R. J. Hagerman and P. J. Hagerman. (2006), Protein composition of the intranuclear inclusions of FXTAS, *Brain*, 129, 256-271 (10.1093/brain/awh650)
 19. Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur *et al.* (2013), CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome, *Neuron*, 78, 440-455 (10.1016/j.neuron.2013.03.026)
 20. Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent *et al.* (2017), Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome, *Neuron*, 93, 331-347 (<https://doi.org/10.1016/j.neuron.2016.12.016>)
 21. Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm and P. K. Todd. (2016), CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins, *Mol Cell*, 62, 314-322 (10.1016/j.molcel.2016.02.034)
 22. Buijsen, R. A. M., C. Sellier, L.-A. W. F. M. Severijnen, M. Oulad-Abdelghani, R. F. M. Verhagen, R. F. Berman, N. Charlet-Berguerand, R. Willemsen and R. K. Hukema. (2014), FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome, *Acta Neuropathologica Communications*, 2, 162 (10.1186/s40478-014-0162-2)
 23. Warner, K. D., C. E. Hajdin and K. M. Weeks. (2018), Principles for targeting RNA with drug-like small molecules, *Nature Reviews Drug Discovery*, 17, 547 (10.1038/nrd.2018.93)

24. Nafisi, S., M. Adelzadeh, Z. Norouzi and M. N. Sarbolouki. (2009), Curcumin binding to DNA and RNA, *DNA Cell Biol*, 28, 201-208 (10.1089/dna.2008.0840)
25. Green, K. M., A. E. Linsalata and P. K. Todd. (2016), RAN translation- What makes it run?, *Brain Res*, 1647, 30-42 (10.1016/j.brainres.2016.04.003)
26. Falconer, R. J. and B. M. Collins. (2011), Survey of the year 2009: applications of isothermal titration calorimetry, *J Mol Recognit*, 24, 1-16 (10.1002/jmr.1073)
27. Freyer, M. W. and E. A. Lewis. (2008), Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions, *Methods Cell Biol*, 84, 79-113 (10.1016/s0091-679x(07)84004-0)
28. Boehr, D. D., A. R. Farley, G. D. Wright and J. R. Cox. (2002), Analysis of the π - π Stacking Interactions between the Aminoglycoside Antibiotic Kinase APH(3')-IIIa and Its Nucleotide Ligands, *Chemistry & Biology*, 9, 1209-1217 ([https://doi.org/10.1016/S1074-5521\(02\)00245-4](https://doi.org/10.1016/S1074-5521(02)00245-4))
29. Das, A., K. Bhadra and G. Suresh Kumar. (2011), Targeting RNA by small molecules: comparative structural and thermodynamic aspects of aristololactam-beta-D-glucoside and daunomycin binding to tRNA(phe), *PLoS One*, 6, e23186 (10.1371/journal.pone.0023186)
30. Disney, M. D., B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney. (2012), A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome, *ACS Chem Biol*, 7, 1711-1718 (10.1021/cb300135h)
31. Tran, T., J. L. Childs-Disney, B. Liu, L. Guan, S. Rzuczek and M. D. Disney. (2014), Targeting the r(CGG) Repeats That Cause FXTAS with Modularly Assembled Small Molecules and Oligonucleotides, *ACS Chemical Biology*, 9, 904-912 (10.1021/cb400875u)
32. Yang, W.-Y., R. Gao, M. Southern, P. S. Sarkar and M. D. Disney. (2016), Design of a bioactive small molecule that targets r(AUUCU)

repeats in spinocerebellar ataxia 10, *Nature Communications*, 7, 11647 (10.1038/ncomms11647<https://www.nature.com/articles/ncomms11647#supplementary-information>)

33. Kypr, J., I. Kejnovská, D. Renciuik and M. Vorlíčková. (2009), Circular dichroism and conformational polymorphism of DNA, *Nucleic acids research*, 37, 1713-1725 (10.1093/nar/gkp026)
34. Giri, P. and G. S. Kumar. (2008), Self-structure induction in single stranded poly(A) by small molecules: Studies on DNA intercalators, partial intercalators and groove binding molecules, *Arch Biochem Biophys*, 474, 183-192 (10.1016/j.abb.2008.03.013)
35. Yang, W. Y., F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney. (2016), Small Molecule Recognition and Tools to Study Modulation of r(CGG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chem Biol*, 11, 2456-2465 (10.1021/acscchembio.6b00147)
36. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), *ACS Chemical Biology*, 13, 180-188 (10.1021/acscchembio.7b00699)
37. Mishra, S. K., N. Jain, U. Shankar, A. Tawani, T. K. Sharma and A. Kumar. (2019), Characterization of highly conserved G-quadruplex motifs as potential drug targets in *Streptococcus pneumoniae*, *Scientific Reports*, 9, 1791 (10.1038/s41598-018-38400-x)
38. Kumar, A., P. Fang, H. Park, M. Guo, K. W. Nettles and M. D. Disney. (2011), A Crystal Structure of a Model of the Repeating r(CGG) Transcript Found in Fragile × Syndrome, *Chembiochem*, 12, 2140-2142 (10.1002/cbic.201100337)
39. M., M. G., G. D. S., H. R. S., H. Ruth, H. W. E., B. R. K. and O. A. J. (1998), Automated docking using a Lamarckian genetic algorithm and an

- empirical binding free energy function, *Journal of Computational Chemistry*, 19, 1639-1662 (doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B)
40. Tawani, A., A. Amanullah, A. Mishra and A. Kumar. (2016), Evidences for Piperine inhibiting cancer by targeting human G-quadruplex DNA sequences, *Scientific Reports*, 6, 39239 (10.1038/srep39239 <https://www.nature.com/articles/srep39239#supplementary-information>)
41. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), 13, 180-188 (10.1021/acscchembio.7b00699)
42. Khan, E., S. Biswas, S. K. Mishra, R. Mishra, S. Samanta, A. Mishra, A. Tawani and A. Kumar. (2019), Rationally designed small molecules targeting toxic CAG repeat RNA that causes Huntington's disease (HD) and spinocerebellar ataxia (SCAs), *Biochimie*, 163, 21-32 (10.1016/j.biochi.2019.05.001)
43. Yang, W. Y., H. D. Wilson, S. P. Velagapudi and M. D. Disney. (2015), Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA, *J Am Chem Soc*, 137, 5336-5345 (10.1021/ja507448y)

Chapter 6

Chemical screening identifies compounds that mitigate r(CGG)_{exp} RNA mediated toxicity in Fragile X-associated tremor ataxia models

6.1 Introduction

Many human diseases are caused by a certain class of biomolecules. If, selectively targeted those biomolecules, could provide a new direction for the precision lead therapeutics. Usually, drug-like small molecules are screened against biomolecules to modulate their structure and functional role in disease pathology [1]. There is emerging evidence that RNA is the unsearched and unexploited target for therapeutic intervention. RNA is the unique biopolymer that has a diverse functional role in the cellular and diseased biology. However, instead of huge interest in this field, very less amount of non-ribosomal RNA targets have been explored for therapeutic development [2]. The small growth in this field is due to the lack of fundamental information and knowledge about the folded RNA secondary structures. Recent studies have been identified small molecules that target repetitive RNAs &, micro RNAs and modulate their functions [3,4]. Nucleotide repeat expansion is the new type of mutation caused by the unstable repeats that leads to various neurological and neuromuscular diseases[5]. Trinucleotide repeats (TNRs) expansion is recognized as the most common cause of neurological disorder and accounts for more than two dozen of diseases like Huntington Syndrome (HD), Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS), Fragile X-Associated Primary Ovarian Syndrome (FXPOI), Muscular Dystrophy type 1 (MD1), Several Spinocerebellar Ataxia Type (1, 2, 3, 6, 7, 8, 12 and 17) [6].

FXTAS is the late-onset (over age 50) neurodegenerative disorder caused due to the CGG nucleotide expansion at 5'UTR of fragile X mental retardation 1 (*FMR1*) gene located on the X chromosome. CGG repeat expansion at 5'UTR varies in normal individual in the range of 4-54 times. If this number is in the range of 55-200 that individual is fragile X permutation carrier, caused FXTAS. Further increase in the number of CGG repeat, above 200 leads to FXS, caused due hypermethylation on the CGG nucleotide results to gene silencing [7]. The most common symptoms of FXTAS are intention tremor, action ataxia, memory loss, dementia, Parkinsonism, autoimmune disease, seizures, and multisystemic atrophy [8]. Most of the neurological disorders, including FXTAS, possess a similar pathogenic mechanism despite different nucleotide expansion. Various pathogenic mechanisms have been reported for expanded CGG repeats diseases. Evidence of RNA gain-of-function mechanism was also studied in model cell lines, animal models and patient brain tissues sample [9]. In cellular models, r(CGG)^{exp} transfected cells form positive intranuclear inclusion bodies and size of inclusions scales up with the length of CGG repeat and time course of the disease condition [10]. The mutant RNA contributes in disease phenotype via sequestration of RNA biogenesis proteins like DiGeorge syndrome critical region 8 (DGCR8), DROSHA [11], Src associated in mitosis of 68 kDa (Sam68) [12], Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNPA2/B1) [13], CUGBP1 and Pura α [14]. Thus, sequestrations of these proteins result in the formation of nuclear RNA foci and also affect splicing regulation in cells [12]. Protein-mediated neuronal toxicity is another hallmark of FXTAS disease in which CGG expansion promotes repeat-associated non-ATG (RAN) translation without the use of canonical start codon (AUG)[15]. In FXTAS, RAN translation produced toxic homopolymeric protein aggregates like polyglycine (FMRpolyG) and polyalanine (FMRpolyA)[15,16].

Currently, therapeutic for FXTAS is symptomatic, aiming particular symptoms like tremor and gait ataxia. However, targeting the actual trigger of FXTAS pathogenesis rather than the treatment of specific symptoms could

provide promising results. The small molecule based therapeutics is taken advantages (high blood barrier absorption, non-immunogenic, target specific) over other methods like antisense oligonucleotide and RNA interference [17]. In this study, we have explored the shape similarity approach, a key concept in drug discovery to discover small molecule with enhanced pharmacodynamics and pharmacokinetic properties [18]. Bisantrene hydrochloride, an anthracene derivative, is a small molecule reported for the effective treatment of metastatic breast cancer [19]. Bisantrene molecule have been identified as a strong binder against CGG RNA motifs. Generally, structurally similar compounds possess similar chemical and physical properties. Thus, we used Bisantrene as a query molecule for shape-based similarity searching in the National Cancer Institute (NCI) chemical database. Compounds with high Tanimoto-score were used for further in-vitro screening using fluorescence enhancement binding assay. Moreover, binding affinity of identified lead molecules was confirmed by isothermal calorimetry (ITC) titration. Additionally, the interaction between lead molecules (B1, B4 and, B11) and multiple repeats CGG RNAs were examined using other biophysical techniques like circular Dichroism (CD) spectroscopy, UV thermal denaturation assay, Nuclear Magnetic Resonance (NMR) spectroscopy titration, Gel retardation assay, and polymerase chain reaction stop (PCR) assay. Next, therapeutic efficacy of these lead small molecules were evaluated in cellular models as well FXTAS cell lines.

6.2 Results and Discussion

6.2.1 Virtual screening for Bisantrene similar molecules

Bisantrene was used as a molecular query molecule in the chemical similarity search to collect compounds from the NCI database. It is a commercially available medium-sized database with more than 250,000 compounds entries, easily accessible and also provides the complete stereochemistry specifications [20]. The database comprises a complete collection of different compounds that showed broad chemical space and are freely available for academic research

purpose only. Rapid Overlay of Chemical Structures (ROCS) alignment tool used three-dimensional shape algorithm for the 3D shape comparison of a virtually screened molecule [21]. ROCS serves every atom as 3D Gaussian function and measure similarity as a function of overlapping volume between alignments of pre-defined conformers. The Gaussian function of ROCS aids to boost the shape-based measurement, as it involves measuring maximal intersection of the volumes between two molecules [21]. Hundreds of conformers of Bisantrene molecule were generated and compared using Omega 2.3.2 (v 2.02) software from OpenEye Scientific Software in NCI database [22]. Chemical similarity search comparison was done using two different methods: Shape Tanimoto coefficient [23] and 3D chemistry alignment and/or color score [24]. Chemical alignment of the compounds marked by the color force field which helps in positioning of anions, cations, hydrogen donor, hydrogen acceptor, ring structures and hydrophobic groups. The chemical similarity was based on either shape or color tanimoto score that lie in the range of 1 (100 % similar) to 0 (no similarity). Highest tanimoto scored Bisantrene similar compound selected and procured from NCI and purity of compounds were checked *via* mass spectrometry further purified compounds used for fluorescence based *in-vitro* screening (Table 6.1 and appendix A).

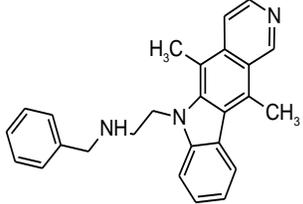
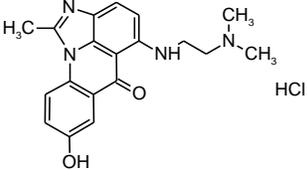
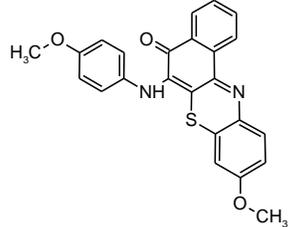
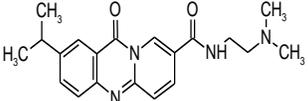
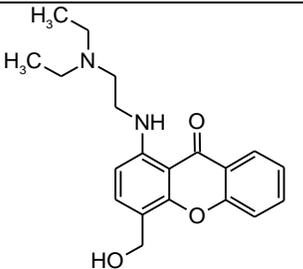
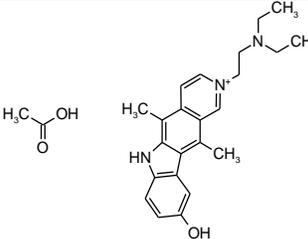
6.2.2 In-vitro selectivity of Bisantrene similar compound against different trinucleotide repeats RNAs

To find out the binding affinity and selectivity of Bisantrene & Bisantrene similar molecules, we employed fluorescence-based binding assay of Bisantrene similar molecules against single 5'CNG/3'GNCx1 internal nucleotide loop RNAs (Figure 6.1). We took advantage of natural fluorescence of small molecules at their emission maxima. With the gradual addition of RNA in small molecules solution, fluorescence enhancement was detected which was plotted against concentration of RNA (Figure 6.2a & 6.2b, Figure 6.3 – 6.4 and Table 6.2 – 6.3). Dissociation constant values obtained from two-mode fitting analysis showed the strong interaction of B1, B4, B6, B7 and B11 with single (5'CGG/3'GGCx1) loop RNA.

Furthermore, to check the selectivity of lead compounds as a function of CGG repeats length, we performed the binding study with higher repeat RNAs (5'CGG/3'GGC_x6), out of five lead molecules three were showed consistent binding (Figure 6.2c, Figure 6.5 and Table 6.4). This may happen due to different modes of binding of compounds. Further, higher number of CGG repeats RNA contributes in FXTAS disease pathogenesis.

Thus, we performed same binding assay of the lead molecules with 5'CGG/3'GGC_x20/40/60 repeats RNAs. The binding affinities of the lead molecules with higher repeats RNA were further improved as the length of CGG RNA increased while no significant binding was detected with 5'CAG/3'GUC_x6 duplex RNA (Figure 6.2d, Figure 6.6 and Table 6.5). Our data evidenced that increasing numbers of CGG RNA motifs improve the binding ability of B1, B4 and B11 molecules. Additionally, as controls different G-quadruplex DNAs (c-myc, c-kit, bcl2), yeast t-RNA, and calf thymus (CT) duplex DNA were used to analyze the untargeted binding. B1, B4, and B11 not showed any considerable binding (Figure 6.2e and Figure 6.7 and Table 6.5).

Table 6.1 Bisantrene similar molecules color and shape tanimoto score values and structure of best similar molecules.

S. no.	Manuscript ID	NSC ID	Color tanimoto score	Shape tanimoto score	Structure
1.	B1	NSC 164016	0.2650	0.7530	
2.	B2	NSC 637991	0.1710	0.7270	
3.	B3	NSC 374977	0.1210	0.7270	
4.	B4	NSC 600305	0.1220	0.7310	
5.	B5	NSC 166368	0.0550	0.7240	
6.	B6	NSC 311152	0.1350	0.7240	

7.	B7	NSC 401302	0.1320	0.7270	
8.	B8	NSC 90847	0.2100	0.7270	
9.	B9	NSC 150838	0.0870	0.7500	
10.	B10	NSC 36879	0.1590	0.7330	
11.	B11	NSC 196403	0.1750	0.7260	
12.	B12	NSC 43592	0.1000	0.7160	

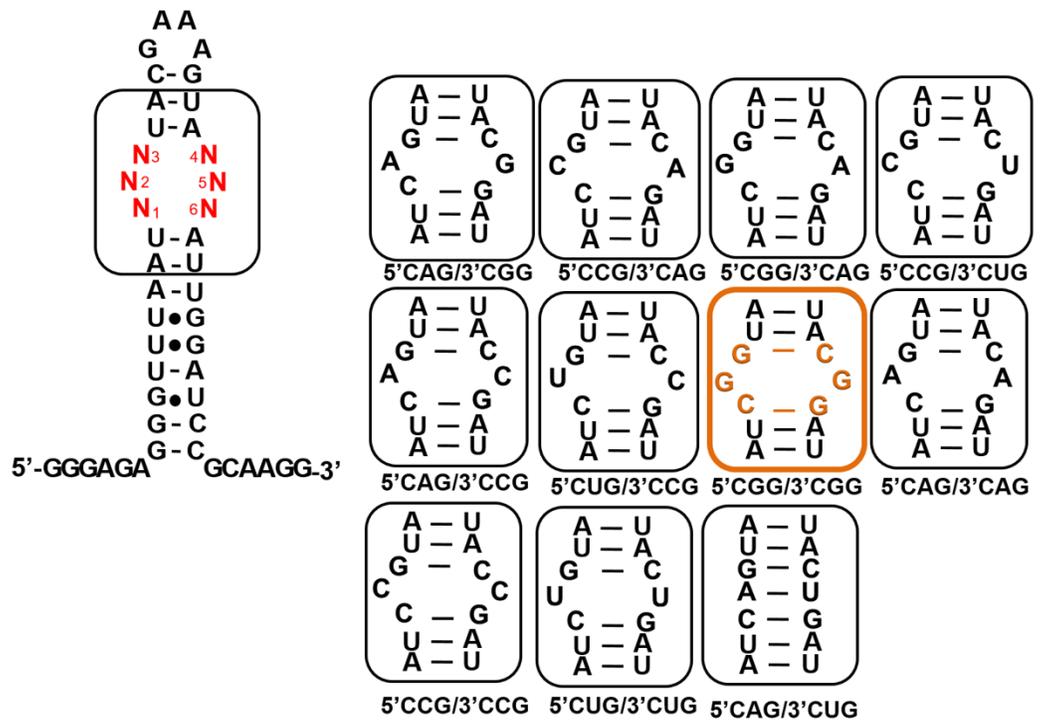


Figure 6.1 Secondary structure of different 1x1 internal nucleotide motif (5'CNG/3'GNC) RNAs sequence.

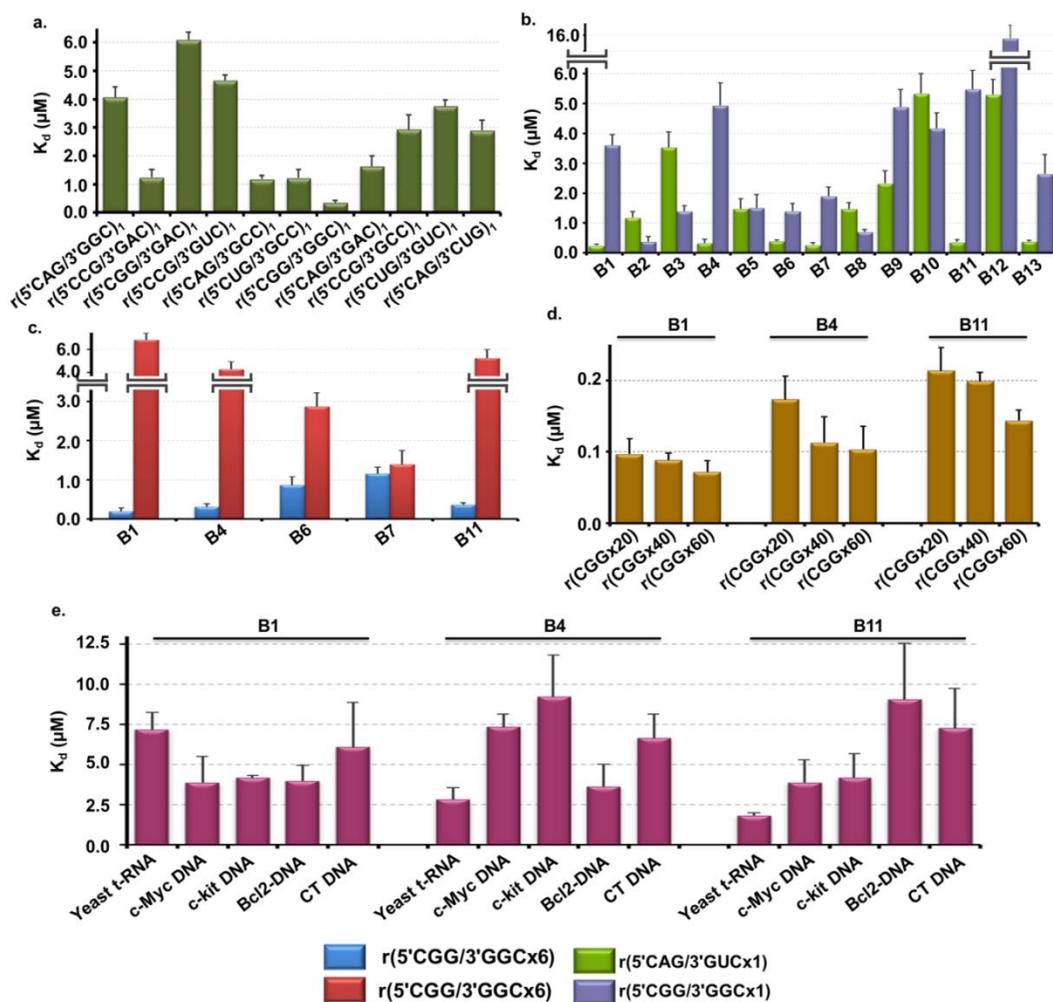


Figure 6.2 Fluorescence-based screening of small molecules against single GG mismatched RNA and different RNA and DNA controls. *a)* Bargraph depicts dissociation constant (K_d) value analysis of Bisantrene with different mismatch single repeat RNA internal loop. *b)* Bargraph represents K_d values comparison of Bisantrene similar molecules with different mismatch single repeat RNA motif. *c)* Dissociated constant value plot of five lead molecules with 5'CGG/3'GGCx6. *d)* Bargraph representing K_d values of final lead molecules with 5'CGG/3'GGCx20/40/60. *e)* Representation of K_d values plot of final lead compounds with different RNA and DNA controls.

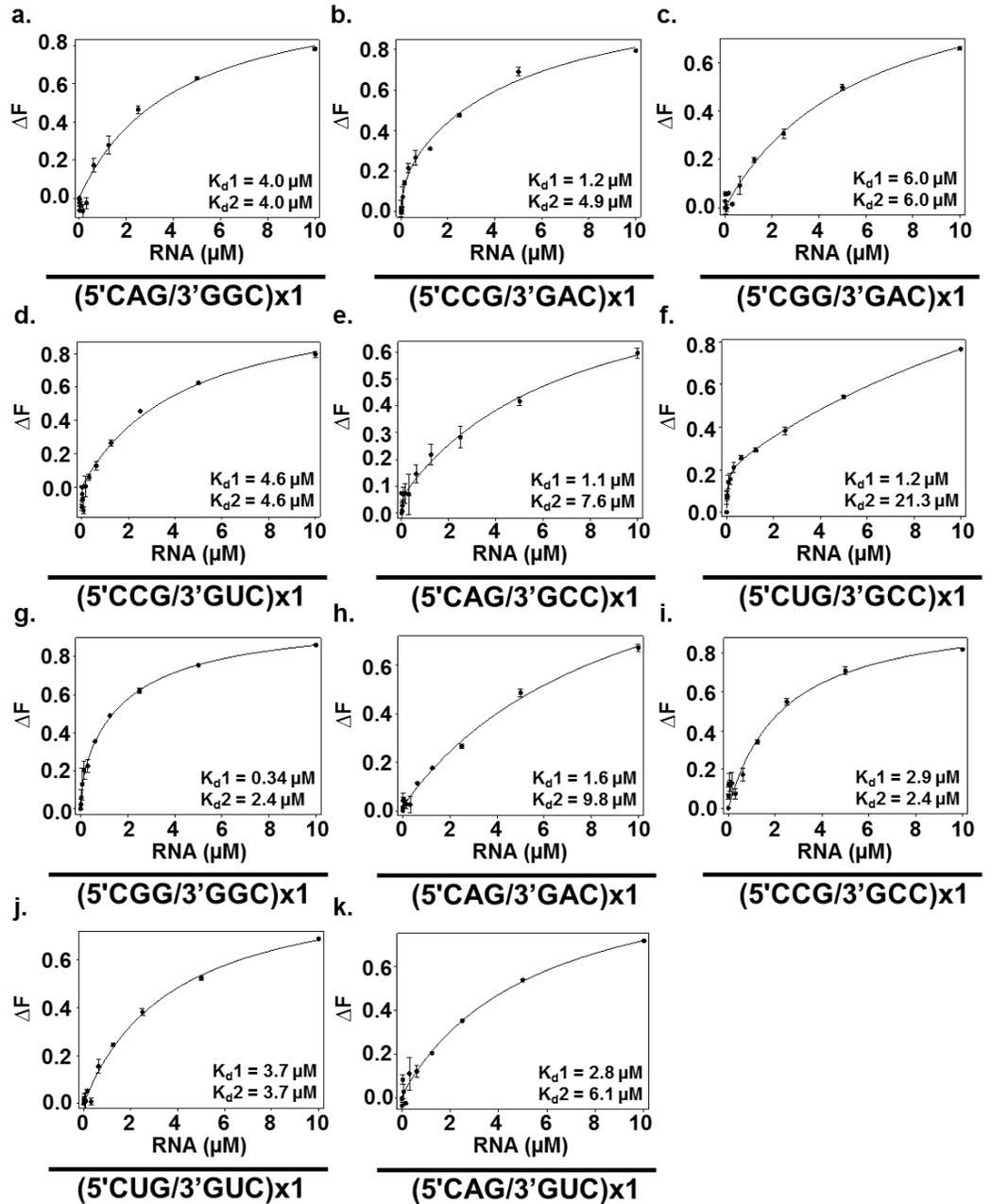


Figure 6.3 Plots show the fluorescence titration assay of different 1 x 1 5'(CNG/3'GNC) RNAs motif with Bisantrène. The dissociation constant (K_d) values are also mentioned below.

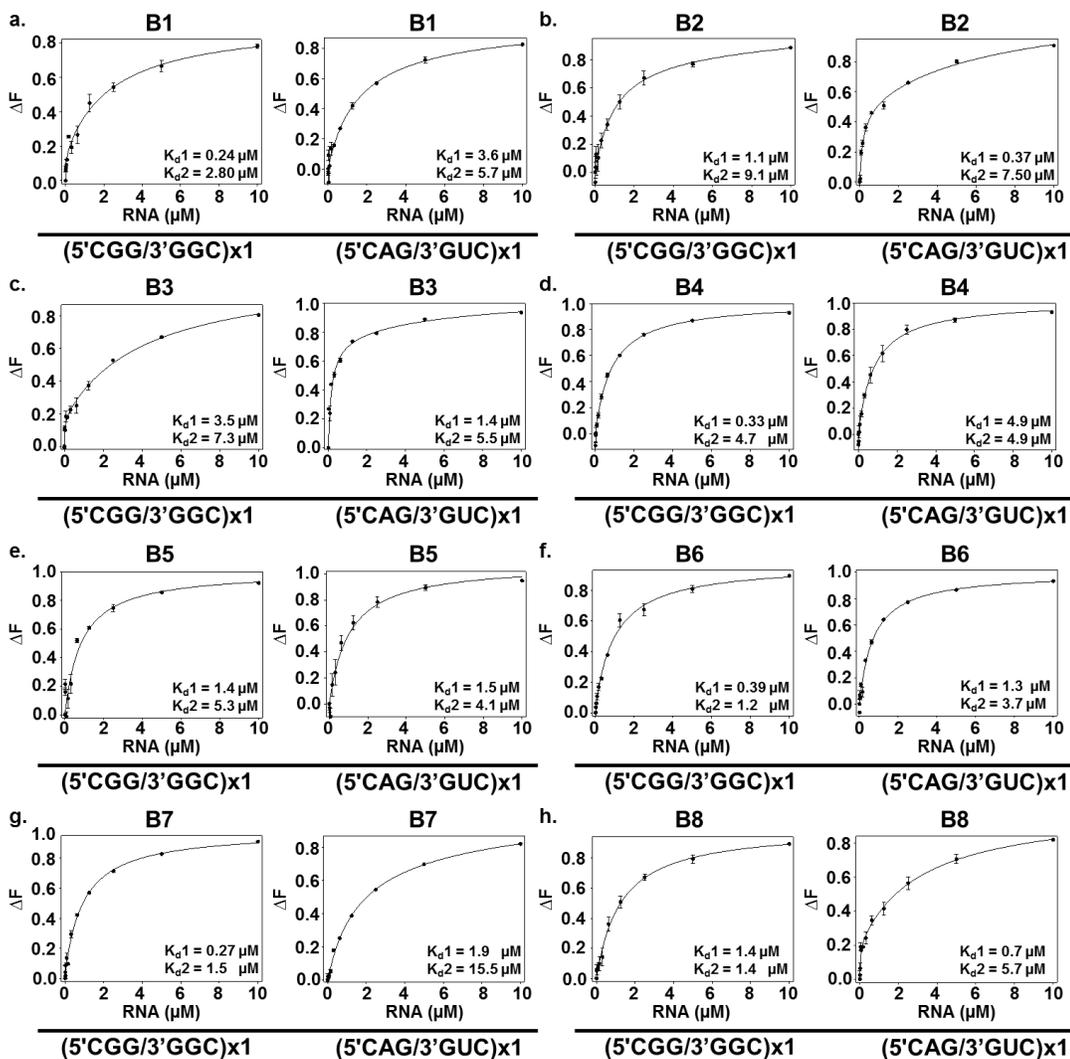


Figure 6.4 Diagram shows the fluorescence based titration assay plots of Bisantrene similar molecules with single repeat CGG RNA and AU duplex RNA. The dissociation constant (K_d) values are also mentioned below.

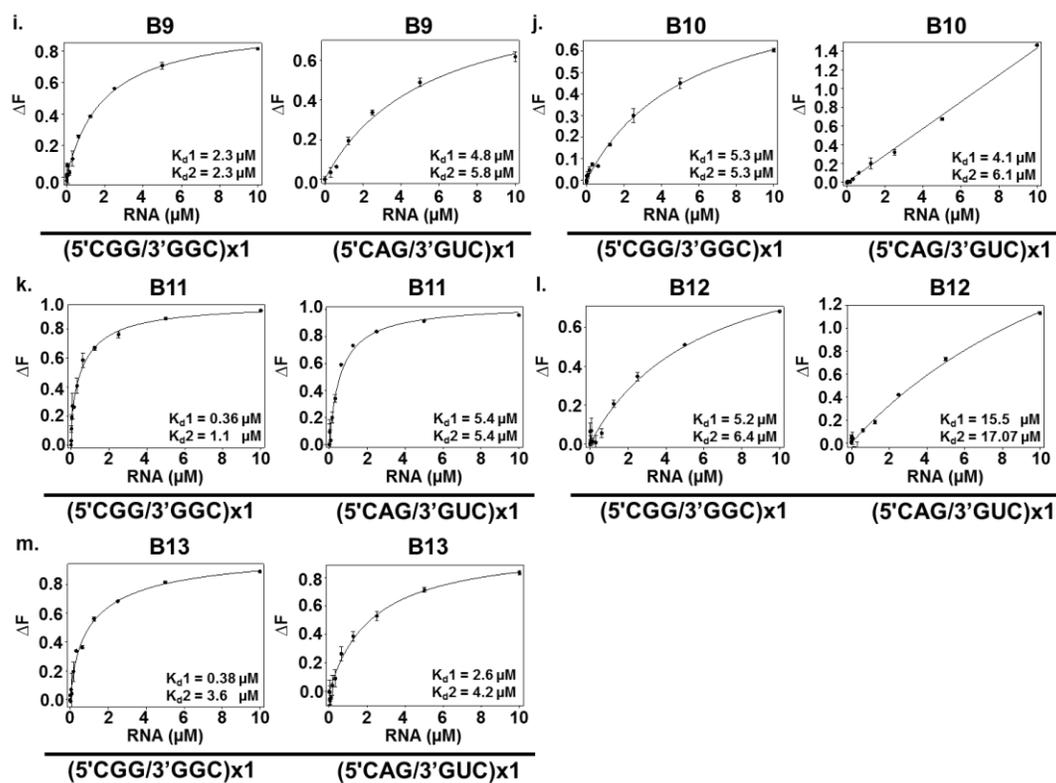


Figure 6.4 conti... Diagram shows the fluorescence based titration assay plots of Bisantrene similar molecules with single repeat CGG RNA and AU duplex RNA. The dissociation constant (K_d) values are also mentioned below.

Table 6.2 The dissociation constant (K_d) values was determined from flourescene based titration assay of Bisantrone with different (5'CNG/3'CNG)x1 RNAs motif.

S. No.	RNA (5' <u>C</u> <u>N</u> G/3' <u>G</u> <u>N</u> C) x 1	K_d^1 (μ M)	K_d^2 (μ M)
1.	(5'CAG/3'GGC) x 1	4.0 \pm 0.36	4.06 \pm 0.36
2.	(5'CCG/3'GAC) x 1	1.2 \pm 0.28	4.9 \pm 0.57
3.	(5'CGG/3'GAC) x 1	6.0 \pm 0.29	6.08 \pm 0.28
4.	(5'CCG/3'GUC) x 1	4.6 \pm 0.19	4.6 \pm 0.19
5.	(5'CAG/3'GCC) x 1	1.1 \pm 0.13	7.6 \pm 0.30
6.	(5'CUG/3'GCC) x 1	1.2 \pm 0.30	21.3 \pm 3.0
7.	(5'CGG/3'GCC) x 1	0.34 \pm 0.08	2.4 \pm 0.28
8.	(5'CAG/3'GAC) x 1	1.6 \pm 0.36	9.8 \pm 0.68
9.	(5'CCG/3'GGC) x 1	2.9 \pm 0.51	2.4 \pm 0.12
10.	(5'CUG/3'GUC) x 1	3.7 \pm 0.22	3.7 \pm 0.22
11.	(5'CAG/3'CUG) x 1	2.8 \pm 0.36	6.1 \pm 0.65

Table 6.3 The dissociation constant (*K_d*) values of Bisantrene similar molecules with (5'CGG/3'GGC)x1 and (5'CAG/3'GUC)x1 RNAs was obtained from flourescene based titration experiment.

S. No.	Compounds	(5'CGG/3'GGC) x1 RNA		(5'CAG/3'GUC) x1 RNA	
		<i>K_d</i> ¹ (μM)	<i>K_d</i> ² (μM)	<i>K_d</i> ¹ (μM)	<i>K_d</i> ² (μM)
1.	B1	0.24 ± 0.05	2.8 ± 0.36	3.6 ± 0.35	5.7 ± 1.05
2.	B2	1.1 ± 0.19	9.1 ± 0.12	0.37 ± 0.17	7.5 ± 0.98
3.	B3	3.5 ± 0.51	7.3 ± 0.21	1.4 ± 0.17	5.5 ± 0.93
4.	B4	0.33 ± 0.12	4.7 ± 2.3	4.9 ± 0.75	4.9 ± 1.3
5.	B5	1.4 ± 0.33	5.3 ± 0.13	1.5 ± 0.44	4.1 ± 1.0
6.	B6	0.39 ± 0.04	1.2 ± 0.24	1.3 ± 0.25	3.7 ± 1.02
7.	B7	0.27 ± 0.06	1.5 ± 0.30	1.9 ± 0.3	15.5 ± 1.1
8.	B8	1.4 ± 0.21	1.4 ± 0.16	0.70 ± 0.08	5.7 ± 0.85
9.	B9	2.3 ± 0.41	2.3 ± 0.39	4.8 ± 0.57	5.8 ± 0.93
10.	B10	5.3 ± 0.66	5.3 ± 0.66	4.1 ± 0.52	6.1 ± 1.87
11.	B11	0.36 ± 0.07	1.1 ± 0.25	5.4 ± 0.62	5.4 ± 0.65
12.	B12	5.2 ± 0.51	6.4 ± 0.73	15.5 ± 1.6	17.07 ± 0.71
13.	B13	0.38 ± 0.04	3.6 ± 1.02	2.6 ± 0.65	4.2 ± 1.7

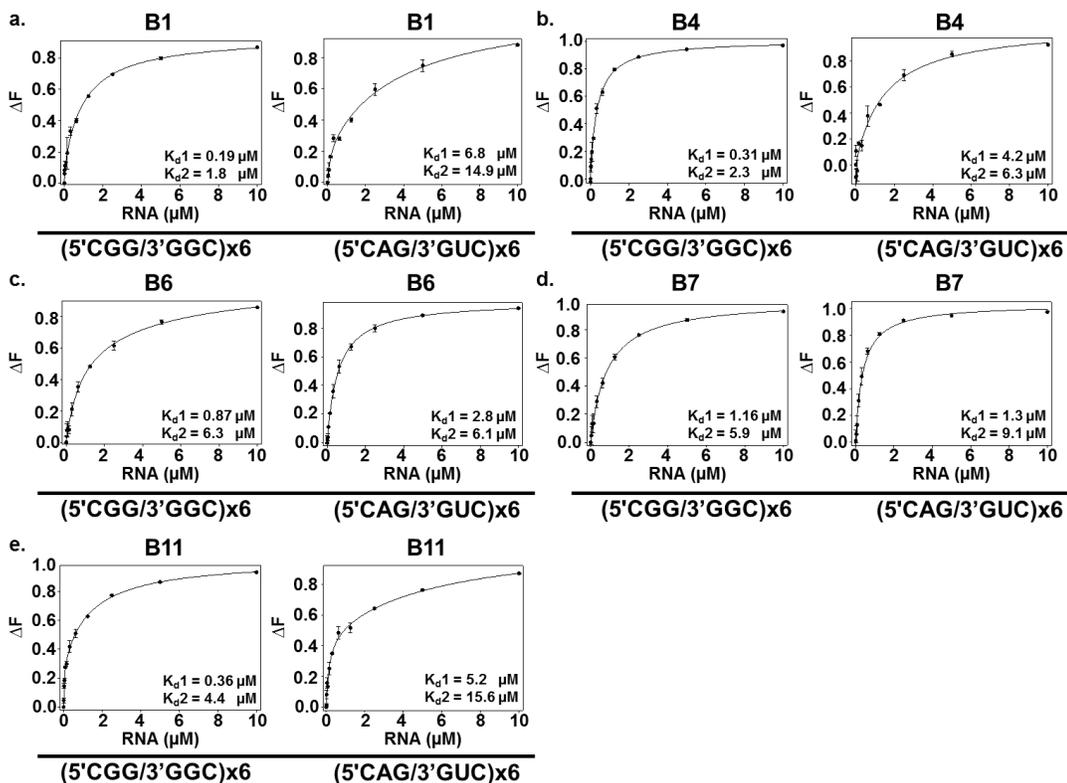


Figure 6.5 Diagram shows the fluorescence based titration assay plots of lead molecules with CGGx6 repeats and AUCx6 repeats duplex RNAs. The dissociation constant (K_d) values are also mentioned below.

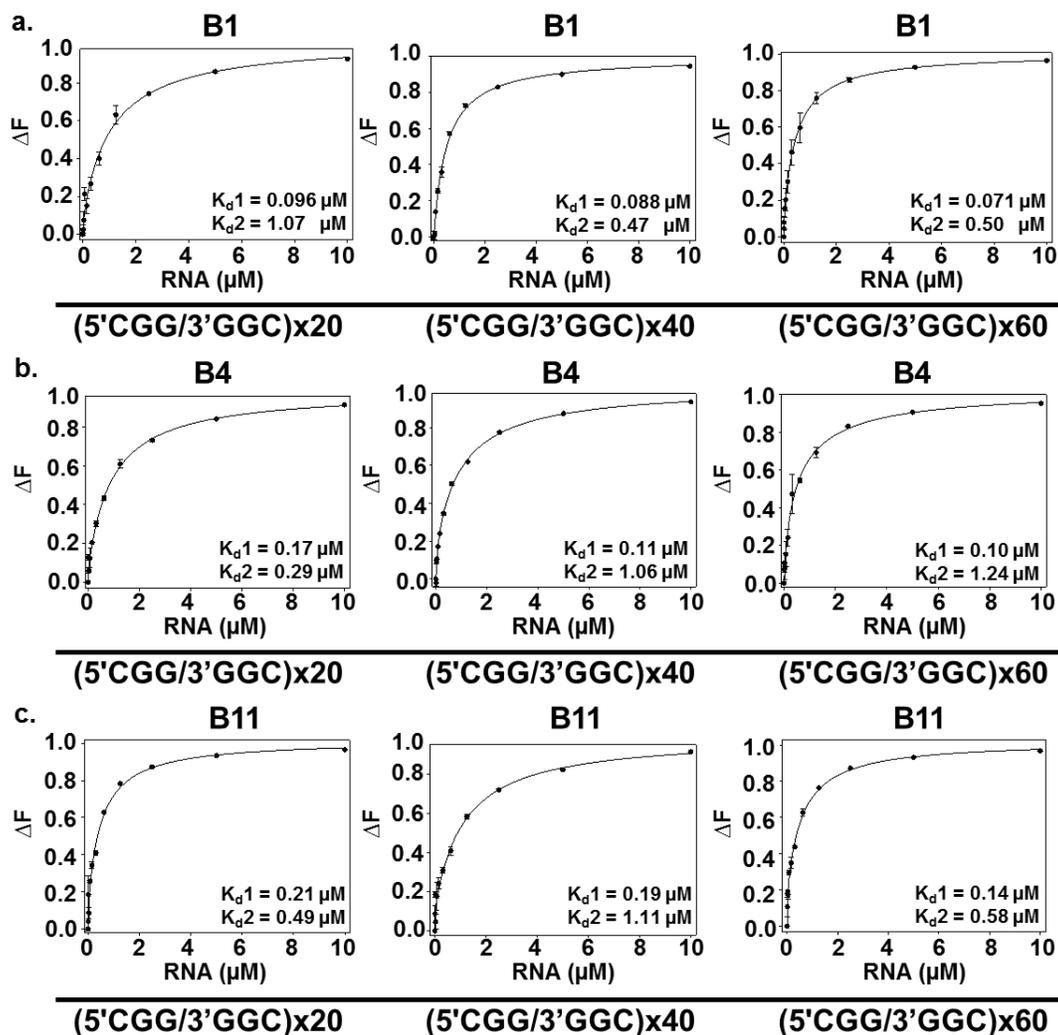


Figure 6.6 Diagram shows the fluorescence based titration assay plots of lead molecules with higher repeats (CGGx20, CGGx40, and CGGx60) RNAs. The dissociation constant (K_d) values are also mentioned below.

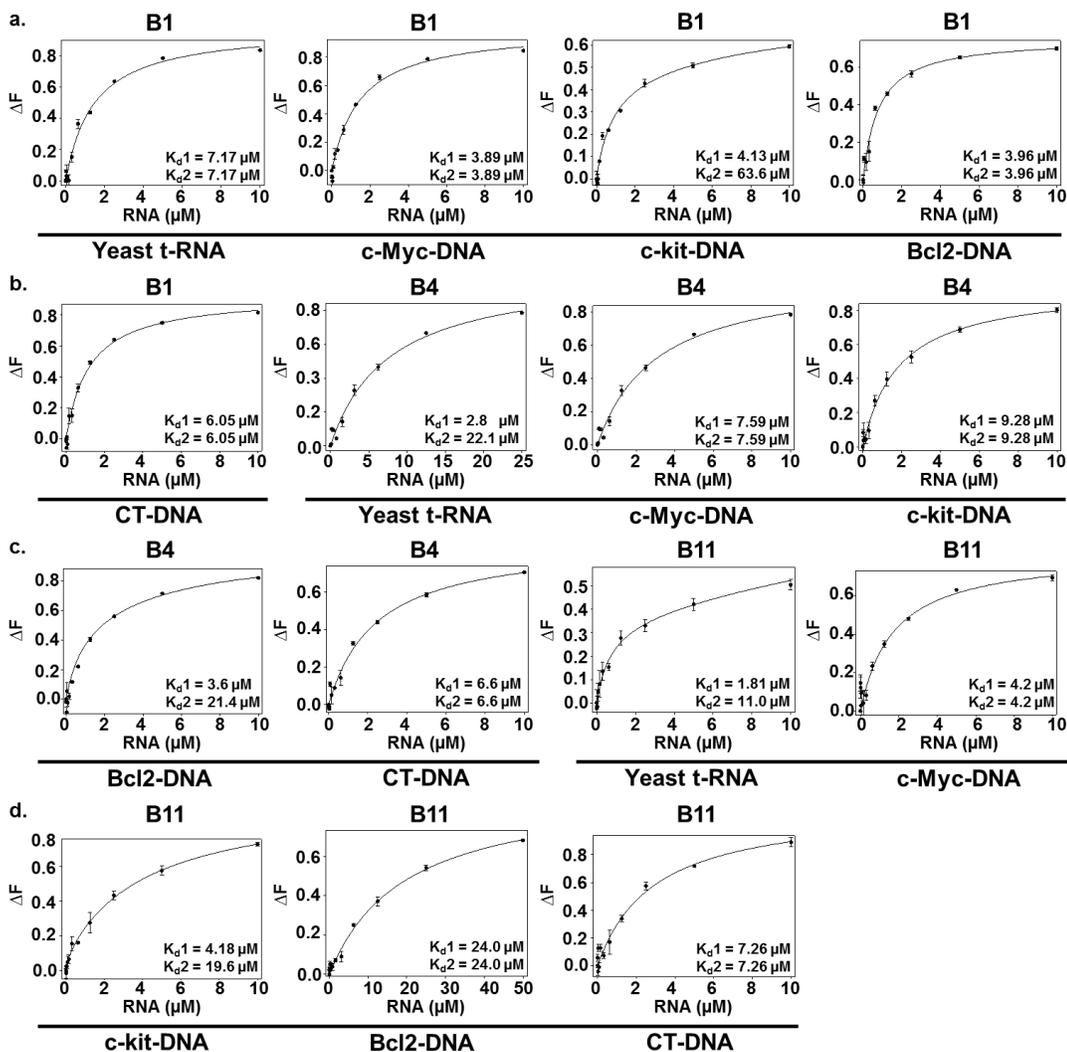


Figure 6.7 Plots represents the fluorescence based titration assay plots of lead compounds with different RNA (yeast t-RNA) and DNA controls (C-myc DNA, C-kit DNA, Bcl2 DNA, & CT DNA). The dissociation constant (K_d) values are also mentioned below.

Table 6.4 The dissociation constant (K_d) values of lead molecules with higher repeat target and control RNA, (5'CGG/3'GGC) \times 6 and (5'CAG/3'GUC) \times 6 respectively. The values was obtained from fluorescence based titration experiment.

S. No.	Compounds	(5'CGG/3'GGC) \times 6 RNA		(5'CAG/3'GUC) \times 6 RNA	
		K_d^1 (μ M)	K_d^2 (μ M)	K_d^1 (μ M)	K_d^2 (μ M)
1.	B1	0.19 ± 0.09	1.8 ± 1.3	6.8 ± 0.57	14.9 ± 9.2
2.	B4	0.31 ± 0.07	2.3 ± 0.9	4.2 ± 0.65	6.3 ± 3.5
3.	B6	0.87 ± 0.2	6.3 ± 3.5	2.8 ± 0.35	6.1 ± 4.1
4.	B7	1.16 ± 0.16	5.9 ± 1.5	1.3 ± 0.35	9.1 ± 3.04
5.	B11	0.36 ± 0.05	4.4 ± 2.8	5.2 ± 0.76	15.6 ± 1.1

Table 6.5 The dissociation constant (K_d) values of lead molecules with higher repeat (5'CGG/3'GGC)x20, (5'CGG/3'GGC)x40, (5'CGG/3'GGC)x60 and with RNA and DNA controls. The values was obtained from fluorescence based titration experiment.

		B1 compound	
S. No.	DNA/RNA	K_d^1 (μM)	K_d^2 (μM)
1.	(5'CGG/3'GGC) x20	0.096 \pm 0.021	1.07 \pm 0.07
2.	(5'CGG/3'GGC) x40	0.088 \pm 0.010	0.47 \pm 0.15
3.	(5'CGG/3'GGC) x60	0.071 \pm 0.015	0.50 \pm 0.24
4.	Yeast t-RNA	7.17 \pm 1.07	7.17 \pm 1.07
5.	C-myc DNA	3.89 \pm 1.63	3.89 \pm 1.63
6.	C-kit DNA	4.13 \pm 0.14	63.6 \pm 7.07
7.	Bcl2 DNA	3.69 \pm 0.98	3.69 \pm 0.98
8.	CT DNA	6.05 \pm 2.8	6.05 \pm 2.8
		B4 compound	
9.	(5'CGG/3'GGC) x20	0.17 \pm 0.032	0.29 \pm 0.06
10.	(5'CGG/3'GGC) x40	0.11 \pm 0.036	1.06 \pm 0.36
11.	(5'CGG/3'GGC) x60	0.10 \pm 0.032	1.24 \pm 0.22
12.	Yeast t-RNA	2.8 \pm 0.75	22.1 \pm 7.47
13.	C-myc DNA	7.59 \pm 0.81	7.59 \pm 0.81
14.	C-kit DNA	9.28 \pm 3.68	9.28 \pm 2.68
15.	Bcl2 DNA	3.6 \pm 1.37	21.4 \pm 7.1
16.	CT DNA	6.6 \pm 1.5	6.6 \pm 1.5
		B11 compound	
17.	(5'CGG/3'GGC) x20	0.21 \pm 0.032	0.49 \pm 0.21
18.	(5'CGG/3'GGC) x40	0.19 \pm 0.01	1.11 \pm 0.35
19.	(5'CGG/3'GGC) x60	0.14 \pm 0.015	0.58 \pm 0.11
20.	Yeast t-RNA	1.81 \pm 0.16	11.0 \pm 3.4
21.	C-myc DNA	4.2 \pm 1.41	4.2 \pm 1.41
22.	C-kit DNA	4.18 \pm 1.47	19.6 \pm 6.0
23.	Bcl2 DNA	24.0 \pm 7.5	24.0 \pm 7.5
24.	CT DNA	7.26 \pm 2.5	7.26 \pm 2.5

6.2.3 Isothermal calorimetry analysis of B1, B4 and, B11 with CGG repeats RNAs

To confirm the selective affinity of B1, B4 and B11 for CGG repeat RNAs, we performed isothermal calorimetry (ITC) titration. ITC is the well-developed technique that directly measures the thermodynamics associated during an interaction between small molecule and biological macromolecule [25]. The heat evolved or absorbed during automated titration reveals the interaction between two entities. Generally, negative enthalpy change suggested the favorable contribution of energy during interactions [26]. The plots were fitted using two-mode binding model and the thermodynamic parameter such as association constant (K_a), enthalpy change (ΔH), entropy change (ΔS), and stoichiometry (N) were calculated by Origin 7.0 software (Table 6.6 a-c). The exothermic peaks appear during the titration showed the negative enthalpy change that affirmed favorable contribution of energy during ligand interaction. The association constant values of B1 for highest affinity binding site of r(CG_Gx20), r(CG_Gx40), and r(CG_Gx60) RNAs are $4.63 \times 10^6 \text{ M}^{-1}$, $1.89 \times 10^7 \text{ M}^{-1}$, and $3.35 \times 10^7 \text{ M}^{-1}$ respectively (Figure 6.8 a-d & Table 6.6a). Similarly, the K_a value of B4 for highest affinity binding site of r(CG_Gx20), r(CG_Gx40), and r(CG_Gx60) RNAs are $1.16 \times 10^6 \text{ M}^{-1}$, $2.04 \times 10^7 \text{ M}^{-1}$, and $2.19 \times 10^7 \text{ M}^{-1}$ respectively (Figure 6.8 e-h & Table 6.6b). Additionally, the K_a of B11 for r(CG_G)20/40/60 RNAs is $1.54 \times 10^6 \text{ M}^{-1}$, $2.32 \times 10^6 \text{ M}^{-1}$, and $1.63 \times 10^7 \text{ M}^{-1}$ respectively (Figure 6.8 i-l & Table 6.6c). Our data suggests that binding affinity of lead molecules increased with the increase in length of CGG RNA. In control, lead molecules showed lower association constant with AUx6 duplex RNA, that gives an idea about the selectivity of B1, B4, and B11 molecules for FXTAS motifs (Figure 6.8 d, h & l and Table 6.6 a-c). Interestingly, B1, B4, and B11 bind approximately 188, 128, 1100 fold tighter with r(CG_Gx60) RNA motifs as compared to AU fully paired RNA. Furthermore, the association constant values of lead molecules with CGG repeat RNAs is in the order of B1>B4>B11, evidenced that B1 is the better binder as compared to B4 and B11 (Figure 6.8 a - l). Importantly, observed exothermic

peaks and negative enthalpy inference the mode of binding of lead molecule with CGG RAN motifs as it could be π - π stacking interaction between aromatic structure and bases of RNA [27,28]. Overall, our ITC data were in-line of fluorescence results that showed the highest affinity and selectivity of B1, B4, and B11 molecules with expanded CGG RNAs.

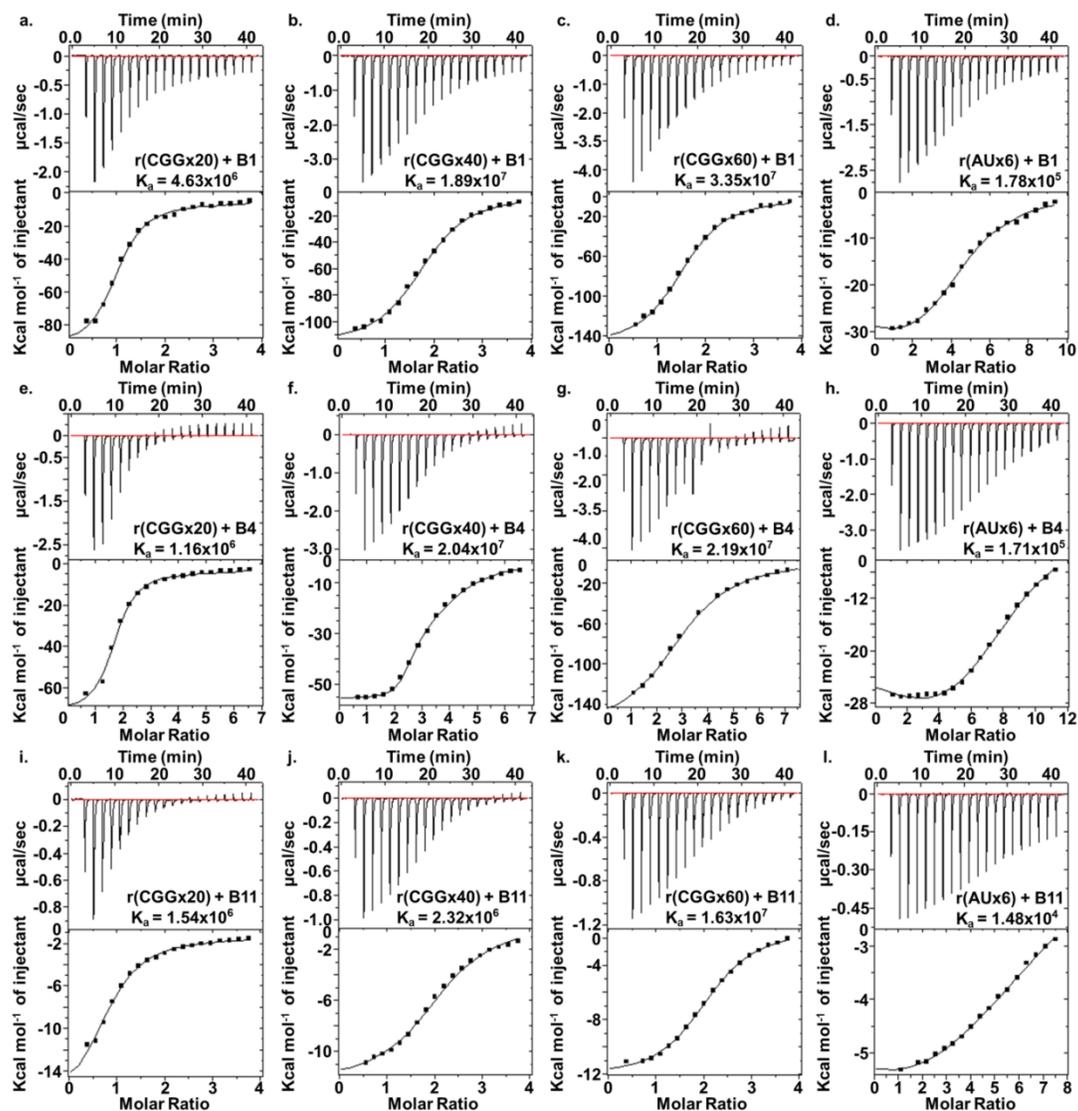


Figure 6.8 Plot represents the isothermal calorimetry titration normalized thermogram of B1, B4, and B11 with multiple CGG repeat RNA. The top panel of each thermogram indicates the raw data for automated sequential injection of small molecules solution in the RNA solution. The bottom panel of each thermogram showed the normalized data point of each injection. a-d), e-f), and i-k) represent the ITC plots of B1, B4, and B11 with different CGG repeat RNAs and AUx6 RNA respectively. The exothermic peaks of all the plots represent the favorable contribution of energy during an interaction between the small molecule and CGG RNAs. The following plots were fitted using two-mode of binding model and solid curve showing is the calculated best fit data.

Table 6.6 a Thermodynamic parameter values of different r(CGG)_n RNAs & r(AU)₆ duplex RNA (control) with B1 compound.

S.no.	Parameters	RNA + B1			
		r(CGG) ₂₀ x	r(CGG) ₄₀ x	r(CGG) ₆₀ x	r(AU) ₆ x
1.	N1 (sites)	1.4 ± 0.84	4.15 ± 1.2	1.09 ± 0.64	1.12 ± 0.21
2.	K1 (M ⁻¹)	1.57E7 ± 0.17	1.89 E7 ± 0.53	3.35E7 ± 2.65	1.78E5 ± 0.19
3.	ΔH1 cal/mol	-7.05E4 ± 1.23	-4141 ± 31.1	-1.37E5 ± 0.14	6.02E4 ± 2.08E4
4.	ΔS1 cal/mol/deg	-177.8 ± 78.0	19.35 ± 0.35	-432 ± 60.8	-837.5 ± 286.3
5.	N2 (sites)	0.697 ± 0.36	1.72 ± 0.014	3.34 ± 0.20	2.66 ± 0.76
6.	K2 (M ⁻¹)	3.58E5 ± 3.85	5.43E6 ± 0.16	1.33E5 ± 0.9E4	1.32E5 ± 0.92E4
7.	ΔH2 (cal/mol)	-1.16 E5 ± 5.7E4	-1.19E5 ± 70.7	-1.35E4 ± 0.87E3	1.92E4 ± 1.5E3
8.	ΔS2 (cal/mol/deg)	-356.5 ± 112.4	-423 ± 91.9	-22.8 ± 61.0	159 ± 8.48

Table 6.6 b Thermodynamic parameter values of different r(CGG)_{exp} RNAs & r(AU)₆ duplex RNA (control) with B4 compound.

S.no.	Parameters	RNA + B4			
		r(CGG) _{20x}	r(CGG) _{40x}	r(CGG) _{60x}	r(AU) _{6x}
1.	N1 (sites)	5.095 ± 0.13	1.88 ± 0.36	1.5 ± 0.83	6.04 ± 1.2
2.	K1 (M ⁻¹)	1.25E6 ± 0.15	2.04E7 ± 1.37	6.77E7 ± 2.19	1.71E5 ± 0.79
3.	ΔH1 cal/mol	-3.57E3 ± 614.4	-3.79E4 ± 3.0E4	-1.31E5 ± 364	2.88E5 ± 1.61
4.	ΔS1 cal/mol/deg	20.25 ± 2.05	-93.8 ± 101.6	-259 ± 32.6	406.6 ± 40.4
5.	N2 (sites)	1.50 ± 0.14	1.53 ± 0.09	1.1 ± 0.72	1.79 ± 0.29
6.	K2 (M ⁻¹)	4.80E5 ± 1.01	3.26E5 ± 3.3	6.45E2 ± 2.1	1.19E5 ± 0.15
7.	ΔH2 (cal/mol)	-8.31E4 ± 1.14	-7.68E4 ± 2.06	-2.67E3 ± 1.94	5.4E5 ± 3.09E3
8.	ΔS2 (cal/mol/deg)	-239 ± 53.7	-288.3 ± 62.5	-368 ± 37	1.76E3 ± 665.3

Table 6.6 c Thermodynamic parameter values of different r(CGG)_{exp} RNAs & r(AU)₆ duplex RNA (control) with B11 compound.

S.no.	Parameters	RNA + B11			
		r(CGG) _{20x}	r(CGG) _{40x}	r(CGG) _{60x}	r(AU) _{6x}
1.	N1 (sites)	6.25 ± 2.4	2.8 ± 1.08	2.5 ± 0.70	0.74 ± 0.2
2.	K1 (M ⁻¹)	5.54E6 ± 0.24	2.32E6 ± 473.7	1.63E7 ± 1.71	1.48E4 ± 0.67
3.	ΔH1 cal/mol	9.22E3 ± 3.79	-1.23E4 ± 1.09	-5.88E3 ± 0.19	-8.39E4 ± 6.08E4
4.	ΔS1 cal/mol/deg	-37.95 ± 12.09	-8.94 ± 1.49	14.78 ± 23.9	-2.63E2 ± 202.9
5.	N2 (sites)	7.88 ± 5.39	6.21 ± 2.52	2.06 ± 0.098	3.76 ± 2.2
6.	K2 (M ⁻¹)	1.09E5 ± 0.09	3.07E5 ± 2.02E4	3.81E5 ± 3.09	2.98E3 ± 0.83
7.	ΔH2 (cal/mol)	-9.77E3 ± 0.25	-529.2 ± 0.05	-7.5E3 ± 6.88E3	-6.80E3 ± 3.5
8.	ΔS2 (cal/mol/deg)	-21.2 ± 4.1	28.39 ± 8.4	9.14 ± 11.9	38.79 ± 17.5

6.2.4 Conformational analysis of CGG RNA in the presence of the lead molecules

To directly analyse the interaction between lead molecules and CGG repeats RNA we employed CD spectroscopy of different length CGG repeats RNA with lead molecules. CD is a very sensitive technique detects even very small conformational changes of nucleic acid caused due to different chemical environment, temperature, and in presence of ligands [29,30]. Expanded CGG RNA form 1x1 arrays of GG hairpin structure which can be reoriented due to interaction of small molecule [31]. The CD spectrum of CGG repeats RNA displayed a positive peak near 265-270 and negative peak around 215-220, which is a typical characteristic of double standard A-type RNA [30]. Interestingly, CD spectrum portrayed the data similar to binding results as observed through screening assay, ligand interacts with (5'CGG/3'GGC) RNA over control RNA. Gradual increasing the concentration of lead molecules (B1, B4, and B11) in RNA solution during titration resulted persistent hypochromic and redshift in the negative peak of CGG repeats RNAs, suggested the B1, B4 and B11 interfere the CGG RNAs structure (Figure 6.9 a-c, 6.9 e-g, & 6.9 i-l and Figure 6.10). Likewise, B1 and B11 also perturbed the positive peaks of CGG repeat RNAs, however, no changes observed in remaining RNA structure. In contrast to CGG RNA no significant perturbation were detected with (5'CAG/3'GUC)x6 duplex RNA control (Figure 6.9 d, h & l). All CGG repeat RNAs showed similar structural variation with B1, B4, and B11 molecules evidenced that binding geometry of these molecules is similar to different GG pair RNAs secondary structure (Figure 6.9). Such a hypochromic and redshift in the GG pair RNAs apart from AU pair RNA suggested the selectivity of lead compounds towards G-rich motifs. Generally, decrease in negative and positive peaks conclude the weak π - π stacking interaction due to π - π stacking between small molecules and their targeted RNAs[32-34]. In conclusion, CD data enforced that B1, B4 and B11 molecules are strongly interacting with CGG RNA motifs and form stable complex.

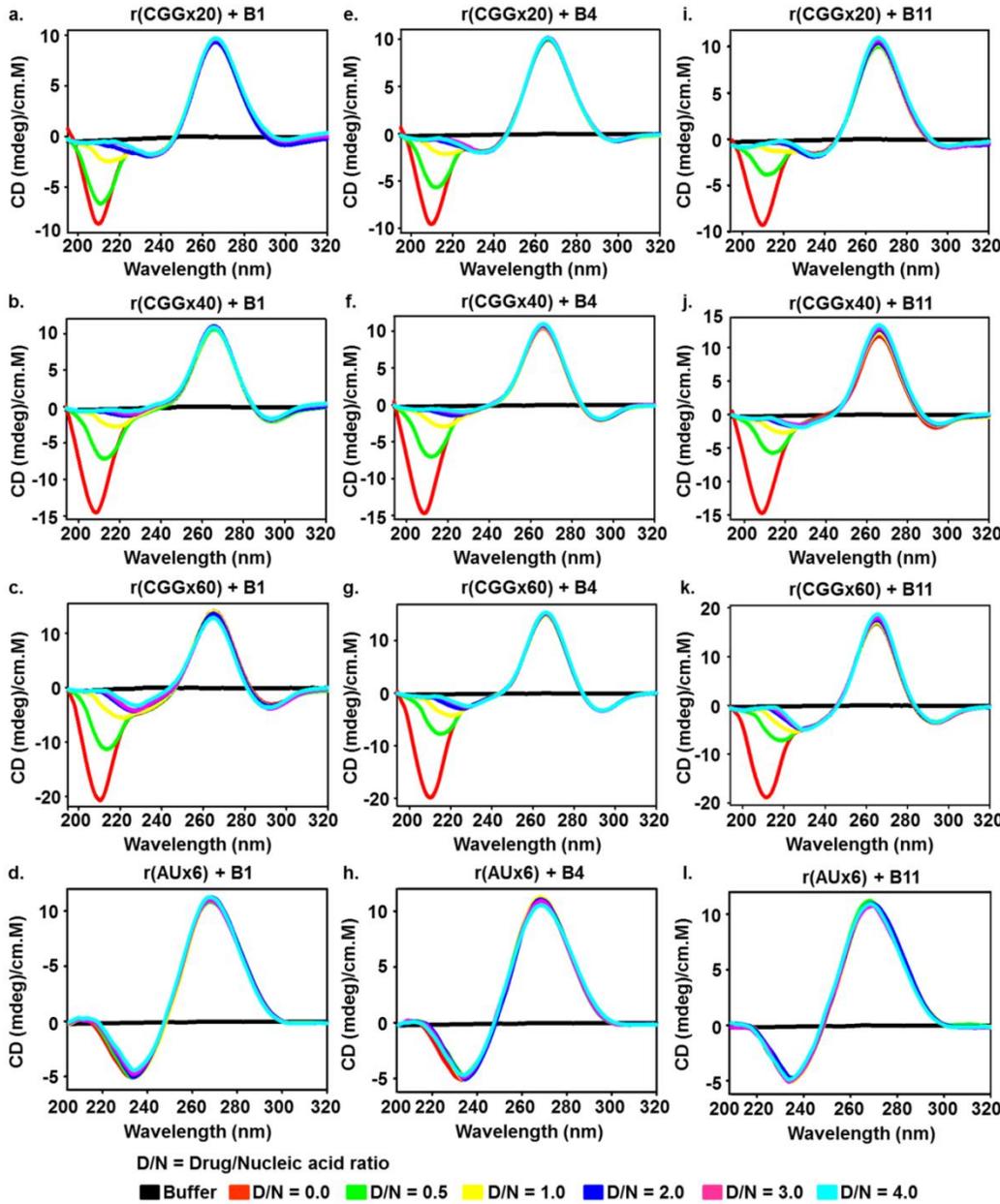


Figure 6.9 Plots represent the Circular Dichroism spectroscopy titrations of B1, B4, and B11 lead molecules with r(CG)exp RNA. Plots a) r(CG)20, b) r(CG)40, c) r(CG)60, and d) r(AU)6 RNAs represent titration spectrums with B1 molecule. Plots e) r(CG)20, f) r(CG)40, g) r(CG)60, and h) r(AU)6 RNAs showed titrated spectrums with B4 molecule. Plots i) r(CG)20, j) r(CG)40, k) r(CG)60, and l) r(AU)6 RNAs represent titration spectrums with B11 molecule. A significant conformation variation observed in CGG repeats RNAs hairpin structure in the presence of B1, B4, and B11. D / N denotes the drug by nucleic acid ratio.

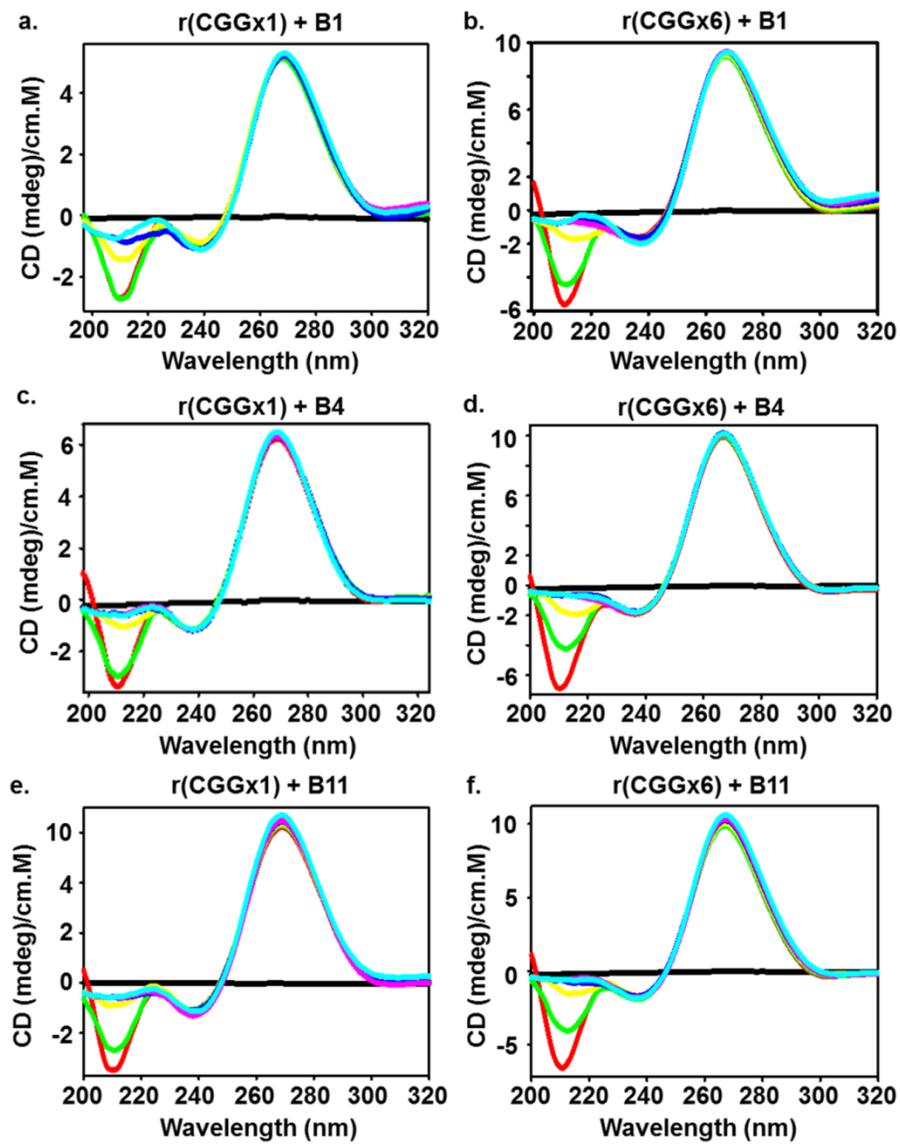


Figure 6.10 Representative CD spectroscopy titration plots of CGG RNA with lead molecules

6.2.5 Thermal denaturation profile of CGG repeats RNAs with B1, B4 and B11 molecules

Next, we studied the thermal denaturation profile of different length CGG RNA to assess the stabilization and destabilization effect of B1, B4, and B11 compounds as a function of concentration. A UV-based thermal denaturation technique is widely used to monitor the interaction of the bioactive compound with biological macromolecules[35]. Note, magnitude of melting temperature (ΔT_m) change depends on binding affinity and stoichiometry RNAs. The UV thermal denaturation assay of all CGG repeat-containing RNA was recorded at 260nm wavelength up to drug by nucleic acid ratio = 2.0. B1 increase the thermal stability curve of CGG repeats RNAs (20/40/60) at the same molar ratio, further enhancement in melting temperature were observed with increased molar ratio (Figure 6.11). The magnitude of ΔT_m of CGGx20, CGGx40 and CGGx60 RNAs with B1 molecule is 9.0, 10.8 and 11.8 °C respectively (Figure 6.11 a-c and Table 6.7). The changes in T_m of CGG repeat RNA in a similar pattern were observed with other lead molecules B4 and B11 (Figure 6.11 e-g & 6.11 i-l and Table 6.7). Interestingly, B1, B4, and B11 do not show significant magnitude change in the T_m with control RNA AUx6, that evidenced the B1, B4, and B11 are selective for G-rich RNA motifs over duplex RNA (Figure 6.11 d, h & l). Additionally, B1 showed more pronounced effect on stability of GG pair RNA compared to B4 and B11. The melting data were consistent with fluorescence binding and ITC data. The molecule depicts the highest affinity have high stabilization effect. In conclusion, thermal stabilization of GG pair RNA motif with small molecules could provide the inhibitory effect of RANT and simultaneously not allowed the sequestration of splicing regulator on RNA motifs. In our lab previously reported the natural compounds to stabilize the RNA motif and rescue the pathological defect of FXTAS and HD diseases [35-37]. Thus, B1, B4, and B11 molecules could provide the neuroprotective effect and could be used for the therapeutic of FXTAS.

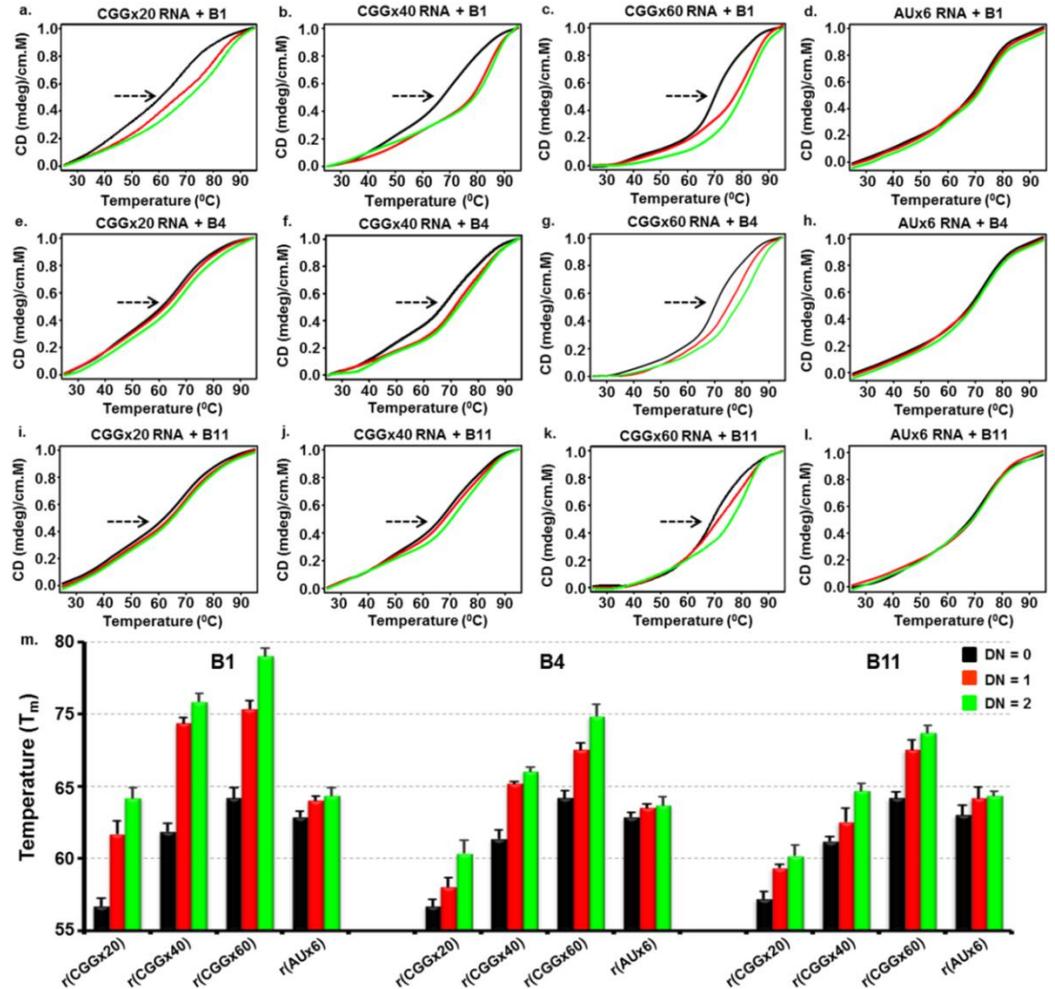


Figure 6.11 A systemic representation of thermal denaturation profile of *r(CG)exp* RNAs with B1, B4, and B11 molecules. Graphs a) *r(CG)20*, b) *r(CG)40*, c) *r(CG)60*, and d) *r(AU)6* RNAs represent titrated melting curve with B1 molecule. Graphs e) *r(CG)20*, f) *r(CG)40*, g) *r(CG)60*, and h) *r(AU)6* RNAs showed thermal denaturation profile in presence of B4 molecule. Graphs i) *r(CG)20*, j) *r(CG)40*, k) *r(CG)60*, and l) *r(AU)6* represent the melting profile with B11 molecule. m) Bargraph showed the ΔT_m values of expanded CGG repeats RNAs with B1, B4, and B11 molecules. The enhancement in the T_m of CGG repeats RNAs was observed with the addition of B1, B4, and B11 molecules whereas no significant change detected in fully paired AUx6 RNA. D / N denote the drug by nucleic acid ratio.

Table 6.7 Thermal melting denaturation assay; T_m values of CGG repeat RNAs in presence and absence of lead compound.

S.no	Compounds	Ratio	RNAs			
			CGGx20	CGGx40	CGGx60	AUx6
1.	B1	D/N = 0	61.0	67.2	70.0	68.4
		D/N = 1	67.0	76.2	77.4	69.8
		D/N = 2	70.0	78.0	81.8	70.2
2.	B4	D/N = 0	61.0	66.6	70.0	68.4
		D/N = 1	62.6	71.2	74.0	69.2
		D/N = 2	65.4	72.2	76.8	69.4
3.	B11	D/N = 0	62	66.4	70.0	68.6
		D/N = 1	64.2	68.0	74.0	70.0
		D/N = 2	65.2	70.6	75.4	70.2

6.2.6 Electrophoretic mobility shift and polymerase chain reaction (PCR) stop assay of r(CGG)^{exp} RNAs with lead compounds

Next, we performed electrophoretic gel mobility shift experiment and PCR stop experiment to affirm the above-described results. With the gradual increasing the concentration of lead molecules the retardation in movement of GG mismatch RNAs observed while no significant changes were found in the AUx6 duplex RNA (Figure 6.12 a-c). This retardation in the movement depends on the RNA sequence and topology that permit to formation of a stable complex. Thus, gel-based assays further signify the selective interaction of lead compound with GG pair RNA as compared to AU pair RNA (Figure 6.12 a-c). In parallel, we checked the binding of lead molecules with CGG template using PCR stop assay. The basic hypothesis of the assay is that lead molecules interact with CGG template it will inhibit the movement of polymerase enzyme. Based on this model, addition of lead compounds decreased the band intensity PCR product as compared to without addition of molecules (Figure 6.13 a-c). We also used the control sequence, AU duplex template, as anticipated no significant change were observed in the band intensity of AU template PCR product (Figure 6.13 a-c). Taken together it is postulated that these lead compounds have selective binding for G-rich sequences.

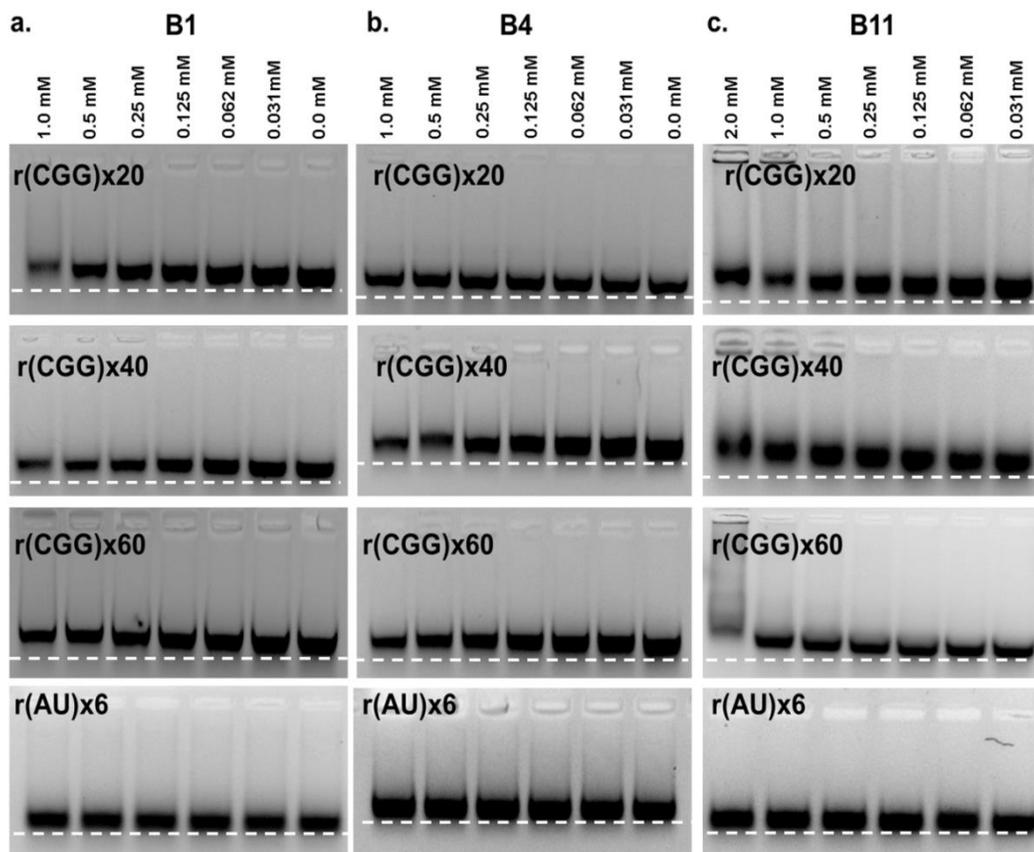


Figure 6.12 Electrophoretic mobility shift assay of expanded CGG RNAs with lead compounds. a) B1, b) B4, and c) B11 images showed the increasing concentration of molecules retard in the mobility of r(CG G)x20/40/60) repeats RNAs and AUx6 duplex RNA

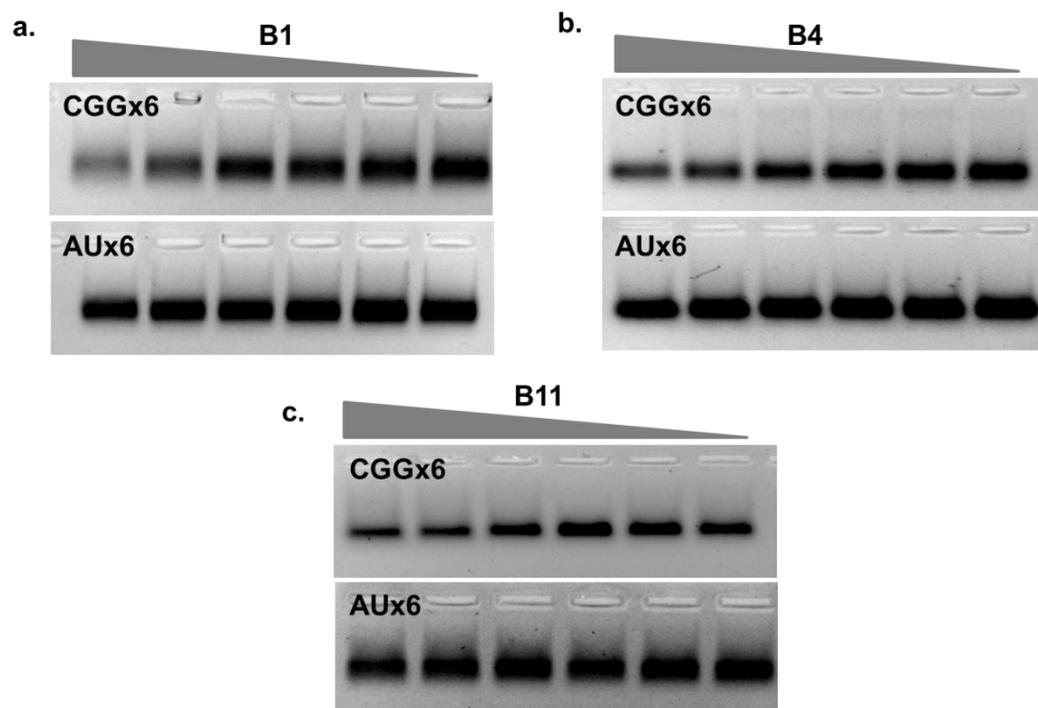


Figure 6.13 a), b), & c) Representing gel image of PCR inhibition assay using CGGx6 and AUx6 as a template with increasing concentration of lead compounds

6.2.7 Structural insight of CGG repeats RNA and B1, B4, and B11 molecule complex using NMR spectroscopy and molecular docking

After deducing the binding affinity of these molecules with CGG RNA we further sought to determine the atomistic detail of these interactions. NMR and molecular docking are the valued tools to study RNA-drug interaction at the molecular level [37,38]. NMR chemical shifting (upshift & downshift) and peak broadening of ligand proton upon addition of nucleic acid could provide insight into ligand proton involved in bound and unbound form [39]. NMR titration assay of CGGx6 RNA was performed with B1, B4, and B11 molecule. The change in ligand proton resonance peak height, broadening, and chemical shift was monitor at each step of titration (Figure 6.14). With the successive addition of CGGx6 RNA into B1 compound solution results to increase in peak height of the B1 protons but no chemical shift was detected at the D/N ratio of 100:1.5 (Figure 6.14a). On the other side, the broadening of B4 and B11 resonance protons were seen at the molar ratio of 100:2.0 and 100:2.5 (Figure 6.14 b-c). Additionally, downfield and upfield chemical shifts of resonating protons were observed in B4 and B11 compound respectively (Figure 6.14 b-c). This result also corroborates the participation of these protons in the interaction of lead compounds with CGGx6 RNA. Furthermore, both aromatic and aliphatic conjugated and non-conjugated protons of compounds were taken part in the interaction. Generally, aromatic ring systems of planar molecules were showed π - π interaction with nucleic acid. Here, most of the interchangeable protons of compound with GG mismatch pair RNA were lie in the aromatic region (7.0-10.0 ppm), that further support our earlier statement of π to π interactions [40]. In conclusion, here, we elucidated the binding of B1, B4, and B11 molecules with GG mismatch pair RNA. According to NMR data interaction binding mode of lead molecules could be one of the best binding ways with GG mismatch pair RNA.

After getting the atomistic detail of lead molecules interaction we performed the molecular docking study to visualize the real-time interaction between CGG RNA and compounds. For docking study we used whole RNA structure (CGGx3 RNA

(PDB ID: 3JS2)) [31,36] is in the grid to allow compound to fit at highest binding position. Interestingly, B1, B4, and B11 bind at the C₄G₅G₆ triad of the RNA as an intercalative with high binding energy of -7.23 kcal/mol, -6.7 kcal/mol, and -6.31 kcal/mol respectively (Figure 6.14 d-f). These compounds also form hydrogen bond that further increases the strength of bonding [41] (Figure 6.14 d-f). Docking data were in-line with other biophysical studies where B1 found as a most effective compound than B4 and B11. Thus, docking and other biophysical data have confirmed the selectivity of these compounds for GG-mismatched pair RNA within C₄G₅G₆ triad.

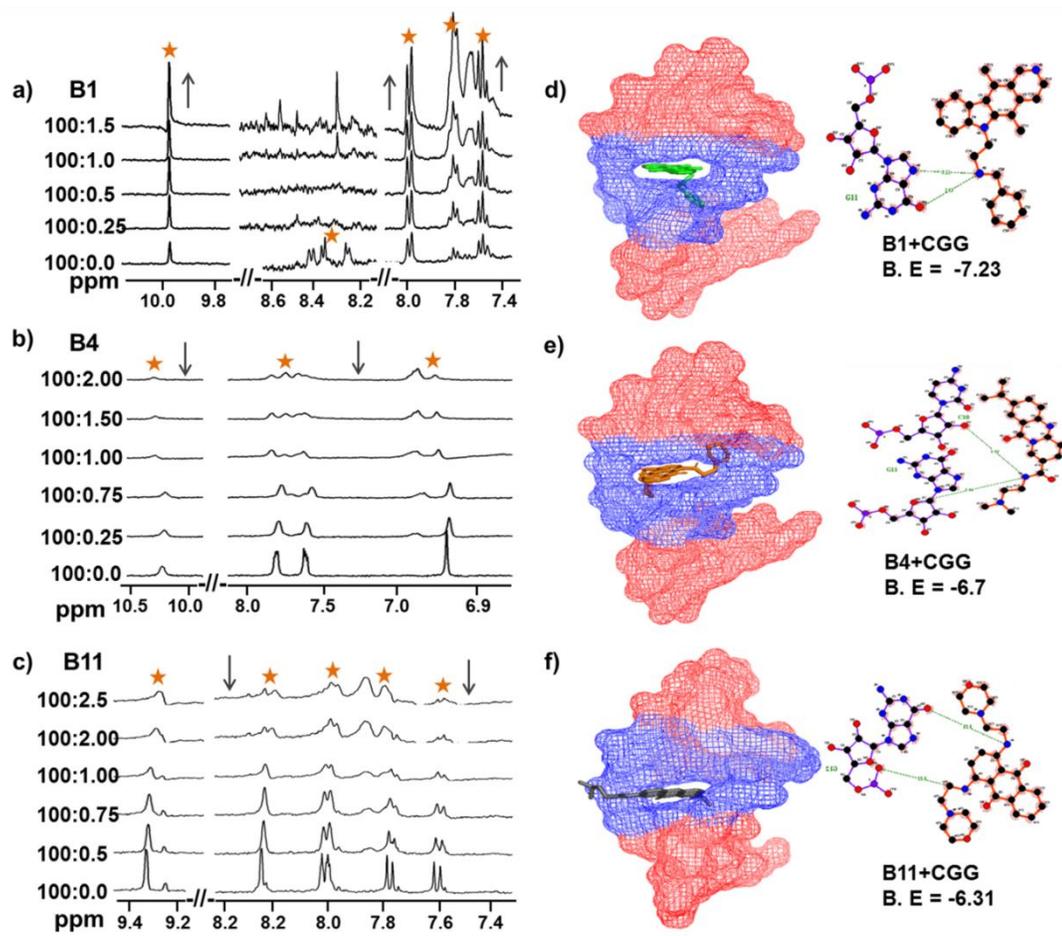


Figure 6.14 One dimension proton NMR spectroscopy titration of B1, B4, and B11 with CGGx6 RNA. a) B1, b) B4, and c) B11 has shown NMR titration with CGGx6 RNA. d) B1, e) B4, and f) B11 images represent the docking study with CGGx3 RNA (PDB ID: 3JS2).

6.2.8 *In-cellulo* potency of lead molecules inhibition of RAN translation in FXTAS developed cell model

RAN translation is another phenotype of FXTAS disorder usually involved in the neuronal toxicity. We further sought to determine the cellular potency of lead molecules for inhibition of canonical and non-canonical translation. For these studies, we developed a cellular model system in which CGG_{x99} repeat are inserted in human 5'UTR of enhanced green fluorescence protein (EGFP) gene. The construct of CGG_{x99}-EGFP designed in such an orientation that mimics the natural open reading frame (ORF) of the *FMR1* gene. The construct expresses the FMRpolyG protein from its natural ACG near-cognate start codon. COS7 cells were transfected with construct and drug-containing media added after post-transfection. Western blot was performed using anti FMRpolyG antibody.

Indeed, the lead molecule reduced the RAN translation product, FMRpolyG, at a different percentage. Treatment of B1 showed statistically significant reduction of FMRpolyG protein with 45% of RAN translation inhibition at 25 μ M. Further, a potent concentration-dependent response is observed (Figure 6.15 b). Although B4 and B11 also inhibited the RAN translation process 56% and 40% at higher concentration, it is mild effective (Figure 6.15 b). Interestingly, the efficacy of RAN translation impairment correlates with the binding affinity of each compound and also modulation of CGG hairpin structure. Furthermore, these molecules do not significantly affect the EGFP protein expression of control plasmid lacking CGG expansion (Figure 6.15 b). As reported, CGG expansion associated RAN product form nuclear inclusion body is repeating length-dependent. We manually measured the numbers of PolyG inclusions body after 24 to 48 hours treatment. B1, B4, and B11 compounds reduced the number of protein aggregates inside the cell, plotted in the form of a bar graph. Maximum inhibition found in the presence of B1 compound (Figure 6.15 a). Taken together, it can be concluded that expanded CGG repeat small molecule modulator could be used for the rescue of RAN translation process without significantly influencing the downstream ORF of the gene. FXTAS is a neurodegenerative disease

characterized by neuronal cell loss. Thus, we tested whether lead compounds can prevent cell toxicity induced by expanded CGG repeats expression. Interestingly, compound B1, B4, and B11 restores normal cell viability of COS7 cells expressing CGGx99 repeats embedded within the 5'UTR of (Figure 6.16). Further, as control we checked compound cytotoxicity using patient derived cell line (GM04026), no significant cytotoxicity found at concentration used in biological assays (Figure 6.17). Moreover, figure 6.18 showing intake of small molecules in treated patient derived cells.

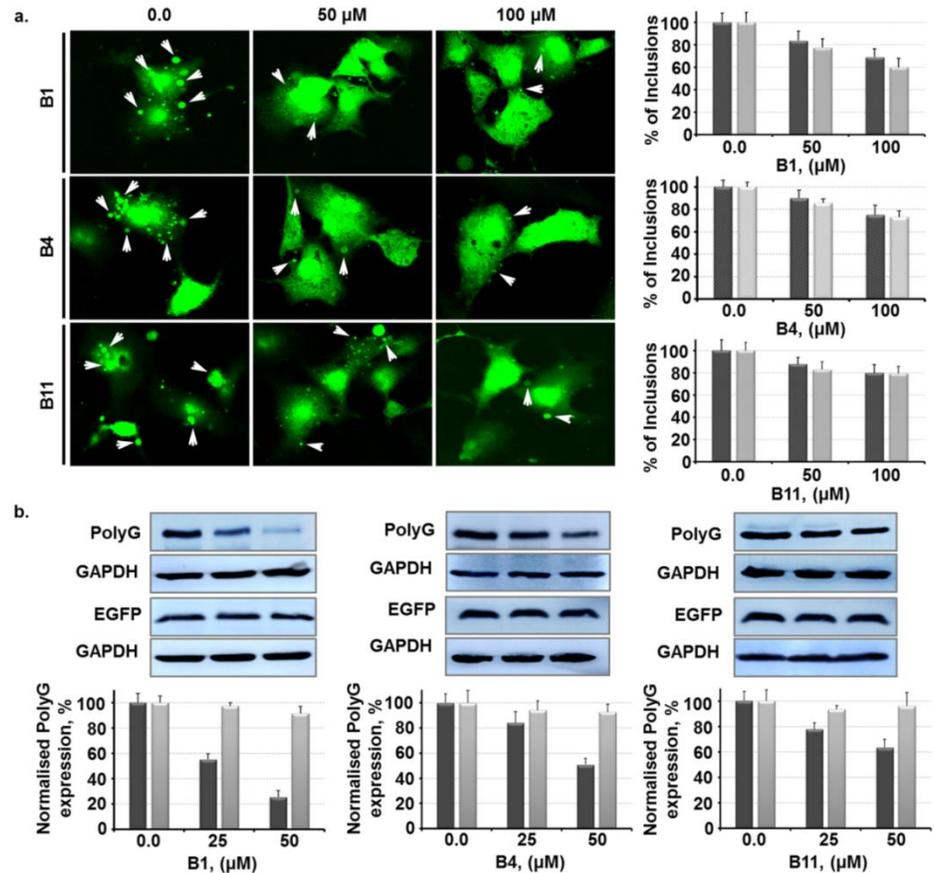


Figure 6.15 *In cellular efficacy of lead compounds in FXTAS developed cell models. In FXTAS, CGG triplet repeats expansion mediates the formation of homopolymeric protein aggregates, FMR1polyG inside the cells. a) B1, B4, and B11 treatment reduced the numbers and aggregation of FMRpolyG-GFP inclusions. Left, images represent the FMRpolyG-GFP inclusions in r(CGG)x99-EGFP expressing COS7 cells with the vehicle (DMSO) and compounds B1, B4, and B11 at the mentioned concentration. Right, bar graph represents percentage reduction of inclusion numbers 24 and 48hrs post-treatment of a vehicle and lead compounds. b) FMRpolyG-GFP protein aggregates reduction was further confirmed using western blots. Blots represent the inhibition of protein inclusions with the treatment of B1, B4, & B11 and vehicle (DMSO) for 24hrs while it not considerably inhibits the canonical translation of EGFP. Normalized bar graphs have shown the quantification of respective blots.*

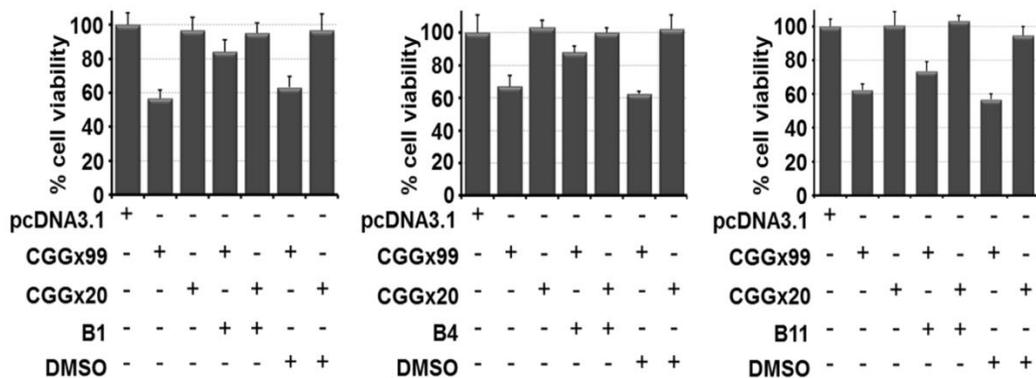


Figure 6.16 Cell viability enhancement of toxic plasmid transfected cells (COS7) in presence of lead compounds.

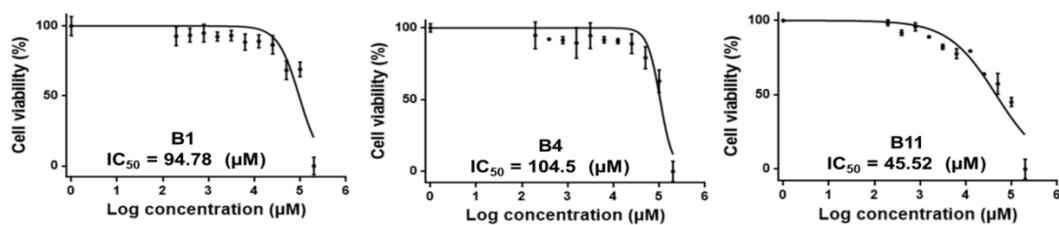


Figure 6.17 Cell cytotoxicity assessment of lead molecules using patient derived (GM04026) cell line.

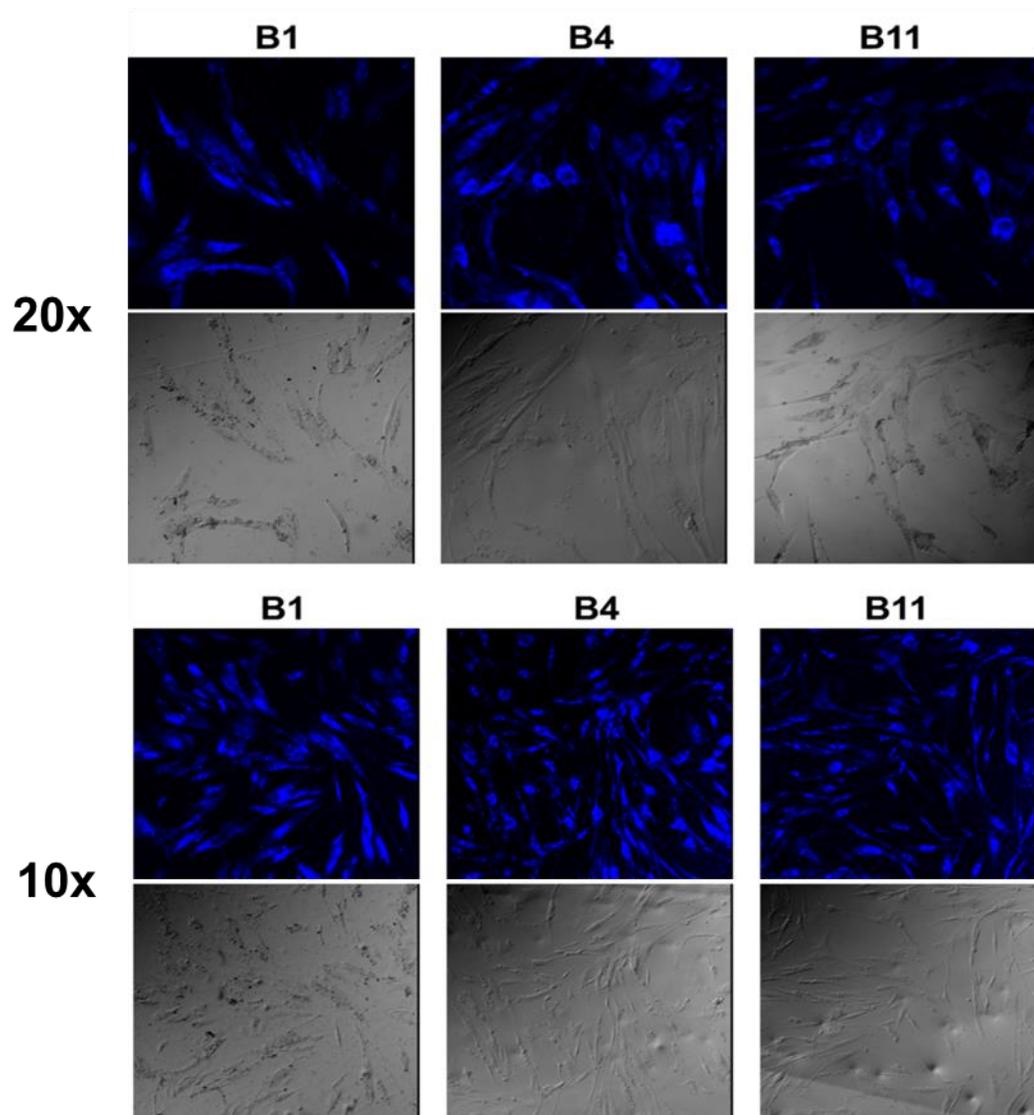


Figure 6.18 Cellular intake of lead compounds in patient derived cells (GM04026). a) 10x magnification b) 20x magnification.

6.2.9 Improvement of a splicing defect in FXTAS cellular model

Next, we examine the cellular efficacy of lead molecules using FXTAS cellular model system. It has been reported that Survival motor neuron 2 (*SMN2*) and B-cell lymphoma (*Bcl-x*) gene are mis-spliced in animal and cellular models [12,42]. In normal cells, Sam68 binds to *SMN2* and *Bcl-x* pre-mRNA and regulate the alternative splicing, a defect occurs due to inactivation of Sam68 and DGCR8 via CGG expansion. Exon 7 of *SMN2* mRNA expressed 40% in healthy cells while its expression enhanced to 70% when cell-expressed CGG repeats expanded (Figure 6.19). In a similar fashion, exon 2 of *Bcl-x* gene expressed 65% of *Bcl-xL* isoform whereas in the presence of CGG repeat it reduced to 35% [12,43] (Figure 6.19). Cos7 cells were co-transfected with toxic CGG repeat-containing plasmid (CGGx99) and *SMN2* minigenes. All three lead compounds were added in the growth media after transfection.

Improvement of mis-splicing defects in *SMN2* and *Bcl-x* was observed with the treatment of lead molecule at the same concentrations. However, the potency of molecules for pre-mRNA splicing correction varied. B1 corrected 40% of mis-splicing defect of *SMN2* at as little as 25 μM and further statistically significant improvement of mis-splicing defect, near to wild type, 71%, were seen at 50 μM (as calculated by comparing the exon inclusion percentage of CGG expressed untreated cells with the exon inclusion percentage of treated cells) (Figure 6.19 a). Similarly, dose responses against *SMN2* were observed with B4 and B11 treatment; however, they are modestly effective at the same concentration. Improvement of mis-splicing defect was also detected in *Bcl-x* minigene as a function of concentration and statistically, significant improvement was found at higher concentration. Here, also B1 was found to be more effective for splicing correction as compared to B4 and B11 compound (Figure 6.19 b). This would be due to the bioactivity and bioavailability of the compound, as previous assays showed B1 has high binding affinity and thermal stability with CGG RNA over B4 and B11.

In contrast, treatment with lead compounds does not significantly affect the transcription level of *SMN2* and *Bcl- x* minigene in absence of r(CGG) expansion (Figure 6.20). This evidenced that lead molecules displaced splicing regulatory protein from (CGG)₉₉ and enhanced normal splicing. These molecules also not substantially affect the mRNA expression level of EGFP gene (Figure 6.19 c). Interestingly B1, B4, and B11 have somewhat similar structural moiety to previously reported molecule which has selectively bind with G-rich motif and improve FXTAS-associated pre-mRNA splicing defect in cellular models [44] [45,46].

Further control experiments were used to address the specificity of the lead molecule. We utilized a cardiac troponin T (*cTNT*) exon 5 which alternate splicing does not modulate in presence of r(CGG) repeats [47] . Importantly, B1, B4, and B11 compounds do not significantly affect splicing outcome of respective transcripts suggested that these compounds do not alter the splicing efficiency of other transcripts globally and specifically involve in mis-splicing defect regulation affected by r(CGG) expansion (Figure 6.21). As previously reported, rescue mis-splicing event correlated to reducing the RNA foci formation inside the cell. According to our data, these molecules would be helpful in reducing RNA foci formation.

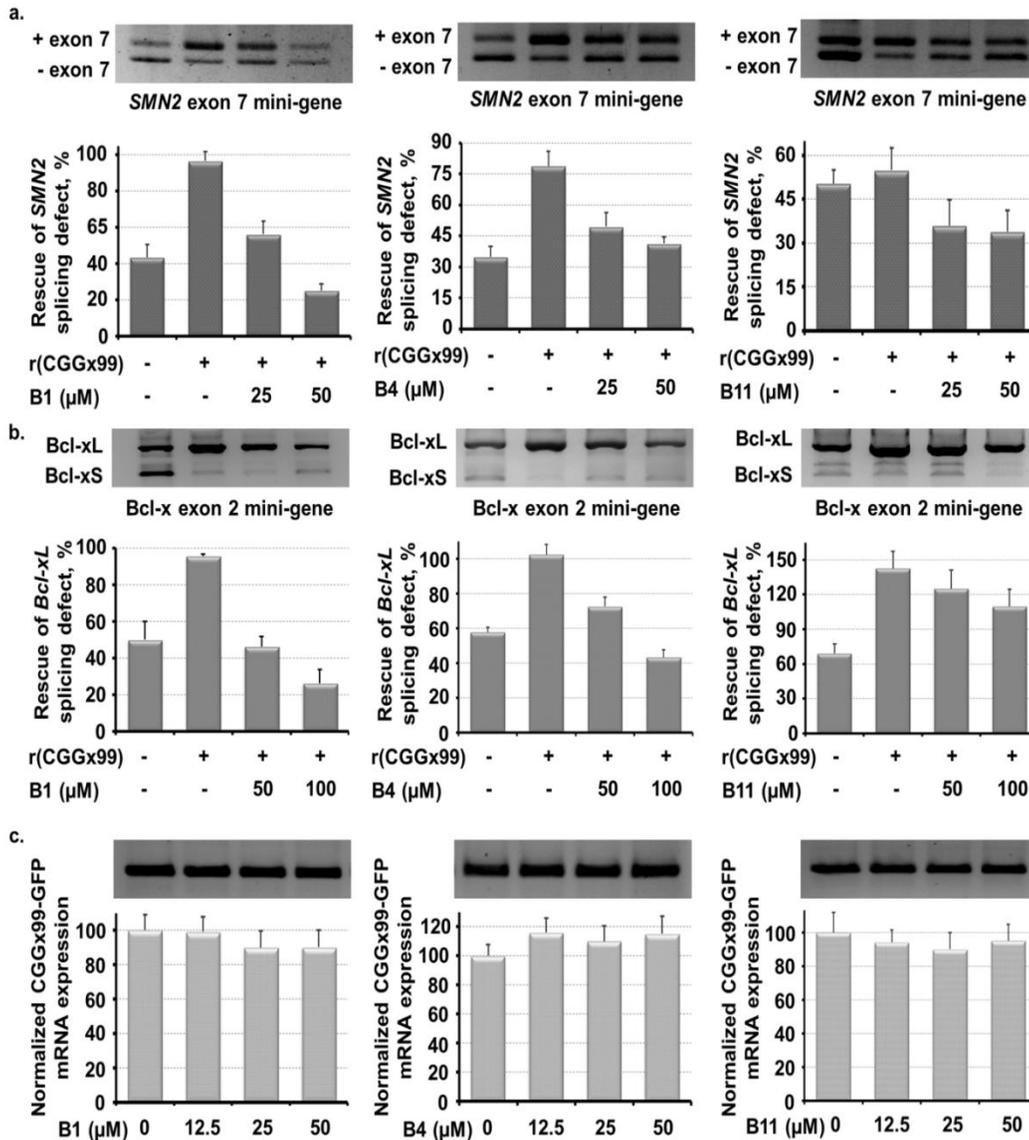


Figure 6.19 In-vitro potency of lead compounds in developed FXTAS cellular models. Bisantrene similar molecules improve pre-mRNA splicing defect of SMN2 and Bcl-x minigenes associated with sequestration of Sam68 and DGCR8 by r(CGg)exp in a concentration-dependent manner. Briefly, COS7 cells were co-transfected with SMN2 & Bcl-x minigenes with CGGx99 repeat-containing plasmid. Splicing event of both SMN2 and Bcl-x genes regulated by DGCR8 and Sam68. a) & b) B1 improves both SMN2 and Bcl-x splicing defect as a function of concentration. Statistically, significant improvement was observed of B1 at a higher concentration. B4 and B11 also improve the SMN2 and Bcl-x splicing defect towards wild-type phenotype in a similar manner as B1. c) Represent the normalized CGGx99-GFP mRNA expression in the presence of lead compounds.

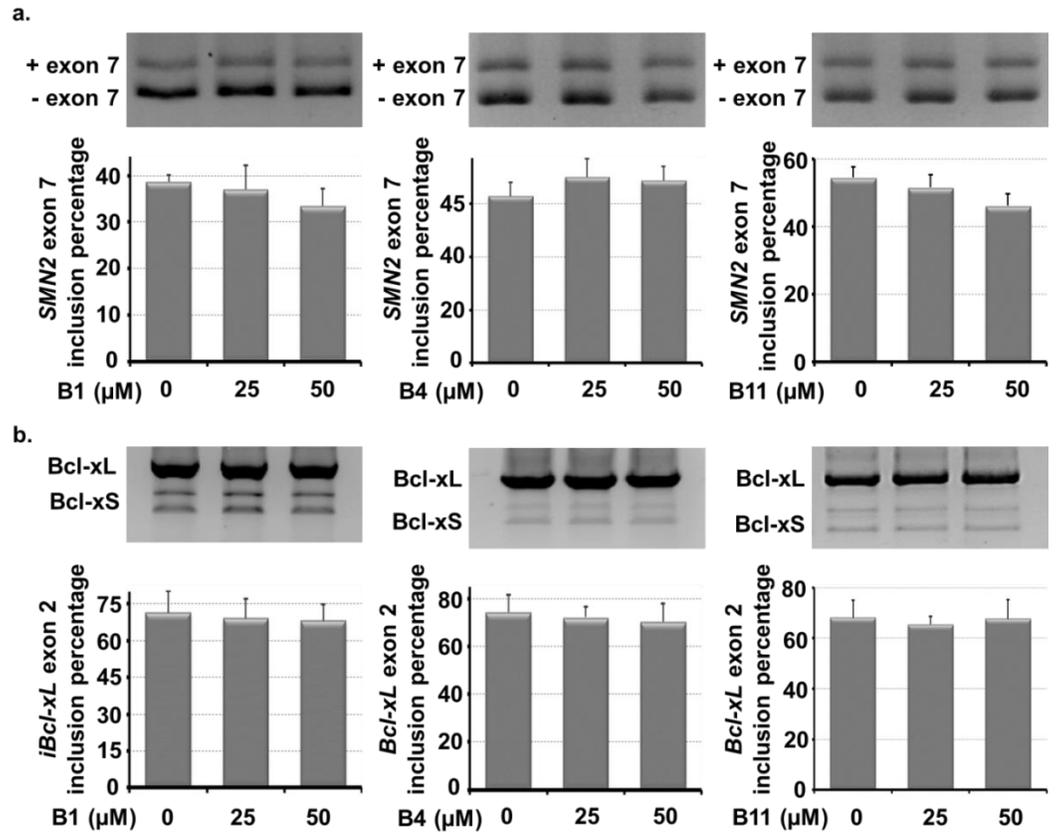


Figure 6.20 Gel image (top) and quantification of pre-mRNA splicing defect of *SMN2* and *Bcl-x* mini-genes with plasmid lacking CGG repeats in transfected and untransfected cells as a function of compound concentration. Briefly, *SMN2* and *Bcl-x* and *EGFP* plasmid (lacking CGG repeat) co-transfected in *COS7* cells. lead compounds does not significantly affect wildtype expression of *SMN2* and *Bcl-x* mini-genes.

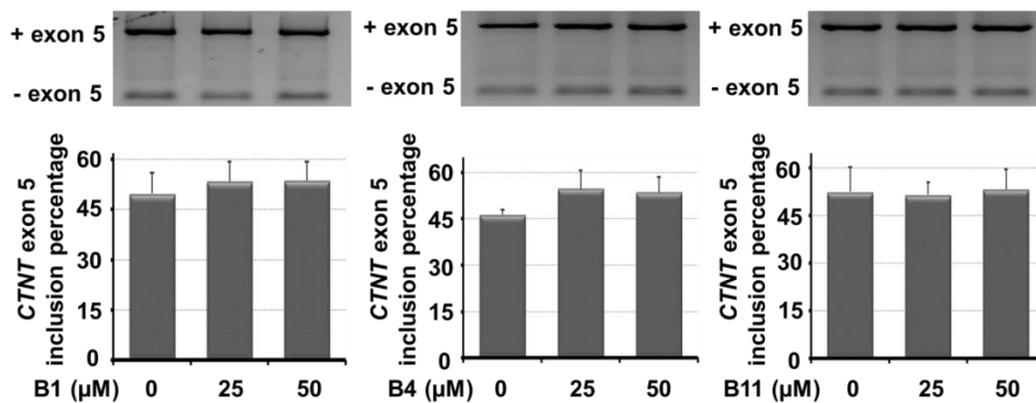


Figure 6.21 Gel image (top) and quantification of alternative pre-mRNA splicing defect of cTNT mini-gene in CGGx99 transfected and healthy cells as a function of lead compounds concentration. Briefly, cTNT and CGGx99 plasmid (toxic) co-transfected in COS7 cells. lead compounds does not affect alternaive splicing defect of cTNT mini-gene.

6.3 Material and Methods

6.3.1 Reagents

Bisantrene similar molecules are procured from National Cancer Institute. Chemical used in this study such as NaH_2PO_4 , Na_2HPO_4 , KH_2PO_4 and K_2HPO_4 , KCl and NaCl were purchased from Sigma Aldrich chemical Pvt. Ltd. The Bisantrene similar molecule was used after checking the purity through ESI mass spectroscopy. HPLC grade organic solvent like dimethyl sulfoxide (DMSO), 2-butanol, ethanol, deuterium oxide, deuterated DMSO were purchased from sigma Aldrich. Plasmid isolation kit, midiprep and miniprep purchased from Himedia. All the PCR reagent including Taq polymerase, dNTP mix, primers, DNA templates (d(CGG)₁₋₆ repeats) were also obtained from sigma Aldrich. Invitro transcription kit (MEGAscript® T7 Kit) were purchased from Thermo Fisher Scientific pvt. Ltd. For CGG hairpin motif formation, CGG repeat RNAs were dissolved in phosphate buffer (10 mM (K^+), pH 7.0) with 50 mM KCl. The CGG RNAs were annealed by heating at 90 °C for 5 - 10 mins, and then allowed to cool down at the room temperature for 1-2 hrs. All the biophysical experiment was carried out using above mentioned buffer composition otherwise mentioned separately. For the cell culture experiment, growth media (Dulbecco's Modified Eagle Medium (DMEM), Fetal bovine serum (FBS), Phosphate buffer saline (PBS), Antibiotic solution (ABAM) were purchased from Gibco. The Human Kidney cell line (HEK293) was purchased from National Centre for Cell Science (NCCS), Pune, India.

6.3.2 Invitro transcription and RNA Purification

RNA samples used in screening assay and other biophysical assay were prepared through invitro transcription method. Synthetic DNA template and linearized plasmid containing T7 RNA polymerase binding site were utilized. Invitro transcription was done according to manufacturer protocol (thermoscientific) and transcribed products were purified using denaturing 15% polyacrylamide. After

UV shadowing, RNAs were extracted using 300mM NaCl by tumbling down for overnight at 4 °C [48].

6.3.3 Fluorescence binding assay

The natural fluorescence of Bisantrene similar compounds were utilized for screening. The assays were conducted on Synergy™ H1 multi-mode microplate reader using corning 96 well microplate at 25 °C. The excitation and emission wavelength of Bisantrene similar molecule were calculated in 1x potassium phosphate buffer by performing its absorption and fluorescence scan. Each sample was measured in duplicates in 75 µl. The different mismatched RNAs and multiple loop CGG RNAs and AU duplex RNA at the final concentration of 10 µM were serially diluted and last well taken as blank (no RNA). Similarly, RNAs AU duplex pair (10 µM), yeast t-RNA, different Gquadruplex and duplex DNAs like c-ymc, c-kit, and bcl2 (25-50 µM) was served as controls. Bisantrene similar molecules were excited at certain wavelength and emitted fluorescence change detected with RNA addition. The obtained data were analyzed using Sigma Plot 12.0 software (Systat Software, Chicago, USA) according to the following equation.

$$f = \frac{B_{max}^1 \times \text{abs}(x)}{k_d1 \times \text{abs}(x)} + \frac{B_{max}^2 \times \text{abs}(x)}{k_d2 \times \text{abs}(x)}$$

B_{max} = maximum number of binding sites. K_d = equilibrium binding constant

6.3.4 Isothermal calorimetry titration (ITC) assay

The ITC cell was filled with 10-20 µM RNA whereas 400-600 µM lead compound was filled in syringe in the same buffer. 1.87 µL of the compounds added at the each step to the sample with initial 0.4 µL injection having 60 sec initial equilibrium delays. Total 21 injections were used with 120 sec pause to homogenous the heat between injections and reference power kept it 7 µcal/s. The samples were stirred at 750 rpm during experiment. The heats of dilution were also calculated by titrating same concentration of compound within same buffer

only. The binding thermogram of RNA-compound complex was plotted by two site binding model to determine the dissociation constant (K_d) after the subtracting with the heat of dilution. ITC were carried out using a MicroCal™ isothermal calorimeter iTC200 (GE healthcare) at a 25 °C. RNA sample were prepared as describe above. Other thermodynamic parameters were also calculated using MicroCal Origin software[49].

6.3.5 Circular Dichroism spectroscopy experiment

Circular Dichroism spectroscopy was completed as described previously [50,51]. Briefly, 5.0 μ M CGG RNA sample were prepared in 1x potassium phosphate buffer (50.0mM KCl, 10.0mM potassium phosphate buffer at 90 °C for 5-10 min then allowed to slowly cooldown at RT (25°C). The annealed RNAs spectrums were recorded at scan rate 20 nm/min, 0.1 nm intervals from 200 nm to 350 nm with a 1 nm-slit width and averaged over three scans using the quartz cuvette, 0.2 cm path length and titration were done with increasing concentrations of lead compounds. Lastly, CD spectrum of RNA and RNA-compound complex plotted after subtraction with the same buffer. CD experiments were performed using J-815 Spectropolarimeter (JASCO). Peltier junction temperature controller was equipped to maintain the uniform temperature 25°C during entire experiment,. A uniform stream of dry nitrogen with the rate of 5L/min, flushed into the cuvette-holding chamber to prevent water condensation outside the cuvette and formation of ozone around lamp.

6.3.6 RNA thermal denaturation experiment

Melting experiment of r(CG₃)^{exp} repeats RNAs and AU paired RNA were diluted in 1x potassium phosphate buffer and heated at 90-92°C for 5-10 min then allow it to cool down at RT for next 30-60 min for the proper annealing. The Melting curve of RNAs were monitored from 25 °C to 95 °C at the rate of 1 °C/min in absence and presence of compound till drug/nucleic acid ratio 2.0. The normalized absorbance changes at 260 nm against temperature were plotted using

the SigmaPlot 12.0 software. All the melting experiments were carried out using Perkin Elmer Lambda 35 UV-vis spectrophotometer attached with peltier temperature programmer and water peltier system PCB-1500.

6.3.7 Nuclear magnetic resonance (NMR) spectroscopy titration assay

NMR experiment for r(CG₁)₁ and r(CG₆)₆ repeat RNA were performed on the high resolution advance III 400 and 500 MHz BioSpin International AG, Switzerland equipped with a 5/8/10 mm broad band inverse probe able to deliver z-field gradients. Drug-RNA sample were prepared in 1x potassium phosphate buffer (10 mM phosphate buffer, pH 7.2, 0.1 M KCl, and 50 mM EDTA) and 10% D₂O. NMR titration experiments of lead compounds were performed with the successive addition of RNA. H₂O + D₂O solvent at a 9:1 ratio were used to lock of the radio frequency during entire titration studies. The 64K data points were recorded for 1D proton NMR spectra with relaxation delay 8 sec, numbers of scan 64 - 128 at 298K and a digital resolution of 0.15 - 0.3 Hz/point was applied. The obtained NMR spectra were consequently processed for the baseline and phase correction. To avoid the sample handling variations for each sample during successive titrations tuning and matching and shimming were performed. Topspin, versions 3.5 was used to process, integrate and analyzed the data. DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) was used to referenced NMR samples.

6.3.8 Molecular Docking of lead molecule with CGG RNA

The crystal structure of duplex CGG motif (PDB ID: 3JS2) were used for the docking.[31] The addition, replacements of residues and optimization of duplex RNA structure with lead compounds were built in Discovery studio 3.5 (Accelrys Inc.,USA) using CHARMM force field. Lead compounds docking studies were performed on Autodock 4.0 where duplex RNA treated as solid rigid structure, RNA and compounds file prepared in PDBQT format. The other parameters kept as a default values during docking. CGG duplex RNA and lead

milli Q water. lead compounds was serially added from 0.0 to 700 μ M then reaction mixtures were incubated in thermo cycler (Eppendorf) with following conditions: 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 50 °C for 30 s, 72 °C for 1 minute and finally held at 4 °C following completion. The obtained PCR product was mixed with 6X DNA loading dye and resolved on 3 % agarose gel stained with ethidium bromide. Gel image was analyzed on ImageQuant LAS 4000 (GE Healthcare) [50].

6.3.11 Quantification of RANT product by western blot

COS7 cells were plated in six well plates in single layer form and transfected with 2500ng of construct containing r(CGG)_{x99}-EGFP sequence using lipofectamine 3000 according to manufacturer protocol (Thermo Fisher Scientific). After 4 to 5 hrs incubation transfection cocktail was removed and fresh drug containing media added and incubated the cell for next 24 hrs at 37 °C. After incubation, cells were lysed using RIPA buffer 200 μ L/well containing 1 μ L of Halt Protease Inhibitor Cocktail from Sigma. Bradford assay were used to determine proteins concentration. Equal amount of protein were loaded in SDS-PAGE and transferred to a PVDF membrane. Blots were incubated with antiFMR1polyG antibody as a primary antibodies from merck millipore. After primary incubation anti-IgG-horseradish peroxidase conjugate secondary antibody used. A chemiluminescent signals was detected using Luminata Crescendo Western HRP substrate from merck millipore in ImageQuant LAS 4000 (GE Healthcare).

6.3.12 Quantifying of non canonical FMRpolyG-GFP protein aggregates

In order to determine the effect of compounds on polyGGFP protein aggregates two different construct were used that contain r(CGG)_{x99}-GFP and pcDNA-GFP. The above mentioned construct was designed and validated by Nicolas group earlier[12]. Briefly, HEK cell were seeded in six well plate in 10% DMEM media. After cell were reached 80-90 percent confluency Transfection were done with above mentioned plasmid using lipofectamine 3000 reagent from Thermo

Fisher Scientific as per manufacturer standard protocol. The transfection cocktail were removed after 4hrs incubation and fresh media were applied which contain respective compounds concentration and the cells were incubate for 18-24 hrs at 37°C. Cell after washing with PBS, fixed with 4% paraformaldehyde for 15 minutes. Image were captured and protein aggregates visualised at higher magnification using fluorescence confocal microscopy and processed to remove background signals. 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 were manually counted from three independent experiment and standard deviation were calculated[50].

6.3.13 Improvement of percentage cell viability of FXTAS cellular model and calculate cell cytotoxic of lead compound by using MTT assay

In order to determine the cytotoxic effect of compound in vivo, MTT (3-(4,5-dimethylthiazol-2-yl)-2,5 diphenyltetrazolium bromide dye) assay were performed on normal cell line (HEK293 and GM00357) as well as patient derived cell line (GM04026). Both normal and patient derived cell line ($5 - 10 \times 10^3$ cells/well) were plated in the 96 well culture plates as mono layer in triplicate and allowed to grow in the complete media (1X DMEM, 10% FBS), at 37 °C with 5% CO₂. After cell were attached properly and reached to optimum confluence, cell were treated with compounds (200 μM to 0.2 μM) with DMSO taken as a vehicle for 24 hrs. After 24 hrs incubation, 10 μL of 5 mg/mL MTT in PBS was added in each well and further incubated for next four hours at 37 °C. The insoluble yellow coloured MTT into dark coloured formazon crystals by intracellular reduction. 100μL DMSO were added to dissolve the formazon crystal structure and absorbance was taken at 570 nm using a microplate reader (Synergy™ H1 multi-mode microplate reader). Half-maximal inhibitory concentration (IC₅₀) of the compounds was determined by the using formula.[38] In addition, HEK cell were plated to check the cell viability in CGGx₉₉ plasmid transfected cell. Cell were transfected with pcDNA-EGFP plasmid (control), CGGx₂₀ normal repeat plasmid and CGGx₉₉ diseased plasmid using lipofectamine 3000 reagent from Thermo

Fisher Scientific as per manufacturer protocol. compounds treatment was given after 4hrs incubation of transfection for 24 hrs and similar MTT protocol were followed to calculate cell viability[50].

$$\% \text{ inhibition} = \frac{\text{Control absorbance} - \text{sample absorbance}}{\text{Control absorbance}} \times 100$$

6.3.14 Assess splicing defect correction by lead compounds in cell cellular model

In order to determine whether lead compounds improve pre-mRNA alternative splicing defect *in-vitro*, a FXTAS cellular model was used. Briefly, HEK293 cells were grown in 24 well plates as monolayer in the growth medium contain 1X DMEM, 10% fetal bovine serum, 1% antibiotic and antimycotic at 37 °C with 5% CO₂. After cell reached 80-90% confluency, HEK cell were transfected with the equal amount of plasmid contain CGGx99 repeat and targeted mini-genes (*SMN2* and *Bcl-x*) using Lipofectamine 3000 reagent (Thermo Fisher Scientific), according to standard manufacturers protocol. Transfection cocktail was removed after 4-5 hours and lead compounds containing growth medium were added. After 24 hours, cells were lysed in the plate and total RNA were recovered using RNA isolation kit (Invitrogen), according to standard manufacturers protocol.

All different conditioned RNA sample were subjected to reverse transcribe (RT-PCR) using cDNA synthesis kit from Bio-Rad, followed standard protocol. Out of 500 ng reverse transcribed mRNA 100 ng were subjected for semi-quantitative PCR. The PCR protocol used: denaturation for 95 °C for 1 min, annealing 55 °C for 1 min, extension at 72 °C for 2 min and final extension at 72 for 10 min. The PCR protocol was run for 25-30 cycles. The PCR product were analysed on agarose gel electrophoresis, stained with ethidium bromide and image were recorded using ImageQuant LAS 4000 (GE Healthcare). The splicing isoform intensity was quantified using imageJ software. The primer sequence mentioned in the table S4 for each construct[12]. Two control condition were used to study

pre-mRNA splicing defect. 1) Targeted mini-genes (*SMN2* and *Bcl-x*) were co-transfected with plasmid that lack CGG repeat at 5'UTR as describe above. 2) Co-transfection of CGG repeat plasmid with mini-genes whose pre-mRNA splicing not govern by sam68 and DGCR8 (cTNT)[12,44].

Table 6.8 PCR primer sequences for splicing defect.

S no.	Gene	Forwards primer	Reverse primer
1.	(CGG) ⁹ 9-GFP	5'GCACGACTTCTTCA AGTCCGCCATGCC	5'GCGGATCTTGAAGTTC ACCTTGATGCC
2.	β -Actin	5' CCTGGCACCCAGCA CAAT	5' GGGCCGGACTCGTCA TAC
3.	SMN2 mini- gene	5'GGTGTCCACTCCCA GTTCAA	5' GCCTCACCACCGTGC TGG
4.	Bc l-x mini- gene	5'GGAGCTGGTGGTTG ACTTTCT	5' TAGAAGGCACAGTCG AGG
5.	cTNT mini- gene	5'GTTCACAACCATCTA AAGCAAGATG	5' GTTGCATGGCTGGTG CAGG

6.4 Conclusion

A shape-based similarity search techniques were utilized to identify small molecule with improved therapeutic efficacy and bioactivity. Here, we screened Bisantrone similar small molecules that showed the binding affinity in nanomolar and also correct FXTAS associated pathophysiological defects. The binding potential of lead compounds with G-rich RNAs was characterized by using different biophysical techniques such as isothermal calorimetry, CD spectroscopy, UV thermal denaturation assay, NMR spectroscopy. We also used some gel-based to further affirm this interaction. Importantly, B1, B4, and B11 molecules showed profound effect in the FXTAS cellular models. All these small molecules have the potential to improve the pre-mRNA splicing defect and repeat-associated protein aggregates formation. Thus, our results signify that these could be used for the therapeutic intervention of FXTAS in future. Further animal studies are required to reach any conclusion.

6. 5 Reference

1. Warner, K. D., C. E. Hajdin and K. M. Weeks. (2018), Principles for targeting RNA with drug-like small molecules, *Nature Reviews Drug Discovery*, 17, 547 (10.1038/nrd.2018.93)
2. Hopkins, A. L. and C. R. Groom. (2002), The druggable genome, *Nat Rev Drug Discov*, 1, 727-730 (10.1038/nrd892)
3. Matsui, M. and D. R. Corey. (2017), Non-coding RNAs as drug targets, *Nat Rev Drug Discov*, 16, 167-179 (10.1038/nrd.2016.117)
4. Verma, A. K., E. Khan, S. R. Bhagwat and A. Kumar. (2019), Exploring the Potential of Small Molecule-Based Therapeutic Approaches for Targeting Trinucleotide Repeat Disorders, *Molecular neurobiology* (10.1007/s12035-019-01724-4)
5. Orr, H. T. and H. Y. Zoghbi. (2007), Trinucleotide repeat disorders, *Annu Rev Neurosci*, 30, 575-621 (10.1146/annurev.neuro.29.051605.113042)
6. Krzyzosiak, W. J., K. Sobczak, M. Wojciechowska, A. Fiszler, A. Mykowska and P. Kozlowski. (2012), Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target, *Nucleic acids research*, 40, 11-26 (10.1093/nar/gkr729)
7. Hoem, G., C. R. Raske, D. Garcia-Arocena, F. Tassone, E. Sanchez, A. L. Ludwig, C. K. Iwahashi, M. Kumar, J. E. Yang and P. J. Hagerman. (2011), CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS, *Hum Mol Genet*, 20, 2161-2170 (10.1093/hmg/ddr101)
8. Mila, M., M. I. Alvarez-Mora, I. Madrigal and L. Rodriguez-Revenga. (2018), Fragile X syndrome: An overview and update of the FMR1 gene, *Clin Genet*, 93, 197-205 (10.1111/cge.13075)
9. Wenzel, H. J., M. R. Hunsaker, C. M. Greco, R. Willemsen and R. F. Berman. (2010), Ubiquitin-positive intranuclear inclusions in neuronal and glial cells in a mouse model of the fragile X premutation, *Brain research*, 1318, 155-166 (10.1016/j.brainres.2009.12.077)

10. Garcia-Arocena, D. and P. J. Hagerman. (2010), Advances in understanding the molecular basis of FXTAS, *Hum Mol Genet*, 19, R83-89 (10.1093/hmg/ddq166)
11. Sellier, C., F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page *et al.* (2013), Sequestration of DROSHA and DGCR8 by Expanded CGG RNA Repeats Alters MicroRNA Processing in Fragile X-Associated Tremor/Ataxia Syndrome, *Cell reports*, 3, 869-880 (10.1016/j.celrep.2013.02.004)
12. Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott *et al.* (2010), Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients, *Embo j*, 29, 1248-1261 (10.1038/emboj.2010.21)
13. Sofola, O. A., P. Jin, Y. Qin, R. Duan, H. Liu, M. de Haro, D. L. Nelson and J. Botas. (2007), RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a *Drosophila* model of FXTAS, *Neuron*, 55, 565-571 (10.1016/j.neuron.2007.07.021)
14. Jin, P., R. Duan, A. Qurashi, Y. Qin, D. Tian, T. C. Rosser, H. Liu, Y. Feng and S. T. Warren. (2007), Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome, *Neuron*, 55, 556-564 (10.1016/j.neuron.2007.07.020)
15. Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur *et al.* (2013), CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome, *Neuron*, 78, 440-455 (10.1016/j.neuron.2013.03.026)
16. Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent *et al.* (2017), Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May

- Contribute to Fragile X Tremor Ataxia Syndrome, *Neuron*, 93, 331-347 (10.1016/j.neuron.2016.12.016)
17. Evers, M. M., L. J. A. Toonen and W. M. C. van Roon-Mom. (2015), Antisense oligonucleotides in therapy for neurodegenerative disorders, *Advanced Drug Delivery Reviews*, 87, 90-103 (<https://doi.org/10.1016/j.addr.2015.03.008>)
 18. Kumar, A. and K. Y. J. Zhang. (2018), Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery, *Frontiers in chemistry*, 6, 315-315 (10.3389/fchem.2018.00315)
 19. Yap, H. Y., B. S. Yap, G. R. Blumenschein, B. C. Barnes, F. C. Schell and G. P. Bodey. (1983), Bisantrene, an active new drug in the treatment of metastatic breast cancer, *Cancer Res*, 43, 1402-1404
 20. Parkesh, R., J. L. Childs-Disney, M. Nakamori, A. Kumar, E. Wang, T. Wang, J. Hoskins, T. Tran, D. Housman, C. A. Thornton *et al.* (2012), Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database and chemical similarity searching, *J Am Chem Soc*, 134, 4731-4742 (10.1021/ja210088v)
 21. GRANT, J. A., M. A. GALLARDO and B. T. PICKUP. (1996), A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape, *Journal of Computational Chemistry*, 17, 1653-1666 (10.1002/(sici)1096-987x(19961115)17:14<1653::aid-jcc7>3.0.co;2-k)
 22. Bostrom, J., J. R. Greenwood and J. Gottfries. (2003), Assessing the performance of OMEGA with respect to retrieving bioactive conformations, *J Mol Graph Model*, 21, 449-462
 23. Haigh, J. A., B. T. Pickup, J. A. Grant and A. Nicholls. (2005), Small molecule shape-fingerprints, *J Chem Inf Model*, 45, 673-684 (10.1021/ci049651v)
 24. Mills, J. E. and P. M. Dean. (1996), Three-dimensional hydrogen-bond geometry and probability information from a crystal survey, *J Comput Aided Mol Des*, 10, 607-622

25. Salim, N. N. and A. L. Feig. (2009), Isothermal titration calorimetry of RNA, *Methods (San Diego, Calif.)*, 47, 198-205 (10.1016/j.ymeth.2008.09.003)
26. Feig, A. L. (2007), Applications of isothermal titration calorimetry in RNA biochemistry and biophysics, *Biopolymers*, 87, 293-301 (10.1002/bip.20816)
27. Ren, J., T. C. Jenkins and J. B. Chaires. (2000), Energetics of DNA intercalation reactions, *Biochemistry*, 39, 8439-8447 (10.1021/bi000474a)
28. Tse, W. C. and D. L. Boger. (2004), Sequence-selective DNA recognition: natural products and nature's lessons, *Chem Biol*, 11, 1607-1617 (10.1016/j.chembiol.2003.08.012)
29. Chang, Y.-M., C. K. M. Chen and M.-H. Hou. (2012), Conformational changes in DNA upon ligand binding monitored by circular dichroism, *International journal of molecular sciences*, 13, 3394-3413 (10.3390/ijms13033394)
30. Kypr, J., I. Kejnovská, D. Renciuik and M. Vorlícková. (2009), Circular dichroism and conformational polymorphism of DNA, *Nucleic acids research*, 37, 1713-1725 (10.1093/nar/gkp026)
31. Kumar, A., P. Fang, H. Park, M. Guo, K. W. Nettles and M. D. Disney. (2011), A Crystal Structure of a Model of the Repeating r(CGG) Transcript Found in Fragile × Syndrome, *Chembiochem*, 12, 2140-2142 (10.1002/cbic.201100337)
32. Giri, P. and G. S. Kumar. (2008), Self-structure induction in single stranded poly(A) by small molecules: Studies on DNA intercalators, partial intercalators and groove binding molecules, *Arch Biochem Biophys*, 474, 183-192 (10.1016/j.abb.2008.03.013)
33. Das, A., K. Bhadra and G. Suresh Kumar. (2011), Targeting RNA by small molecules: comparative structural and thermodynamic aspects of aristololactam-beta-D-glucoside and daunomycin binding to tRNA(phe), *PLoS One*, 6, e23186 (10.1371/journal.pone.0023186)

34. Hossain, M., P. Giri and G. S. Kumar. (2008), DNA intercalation by quinacrine and methylene blue: a comparative binding and thermodynamic characterization study, *DNA Cell Biol*, 27, 81-90 (10.1089/dna.2007.0652)
35. Khan, E., S. Biswas, S. K. Mishra, R. Mishra, S. Samanta, A. Mishra, A. Tawani and A. Kumar. (2019), Rationally designed small molecules targeting toxic CAG repeat RNA that causes Huntington's disease (HD) and spinocerebellar ataxia (SCAs), *Biochimie* (<https://doi.org/10.1016/j.biochi.2019.05.001>)
36. Verma, A. K., E. Khan, S. K. Mishra, N. Jain and A. Kumar. (2019), Piperine Modulates Protein Mediated Toxicity in Fragile X-Associated Tremor/Ataxia Syndrome through Interacting Expanded CGG Repeat (r(CGG)(exp)) RNA, (10.1021/acscemneuro.9b00282)
37. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), *ACS chemical biology*, 13, 180-188 (10.1021/acscembio.7b00699)
38. Tawani, A., A. Amanullah, A. Mishra and A. Kumar. (2016), Evidences for Piperine inhibiting cancer by targeting human G-quadruplex DNA sequences, *Scientific Reports*, 6, 39239 (10.1038/srep39239 <https://www.nature.com/articles/srep39239#supplementary-information>)
39. Tawani, A. and A. Kumar. (2015), Structural Insight into the interaction of Flavonoids with Human Telomeric Sequence, *Scientific Reports*, 5, 17574 (10.1038/srep17574<https://www.nature.com/articles/srep17574#supplementary-information>)
40. Tawani, A. and A. Kumar. (2015), Structural Insights Reveal the Dynamics of the Repeating r(CAG) Transcript Found in Huntington's Disease (HD) and Spinocerebellar Ataxias (SCAs), *PLoS One*, 10, e0131788 (10.1371/journal.pone.0131788)

41. Oliva, R., L. Cavallo and A. Tramontano. (2006), Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions, *Nucleic Acids Res*, 34, 865-879 (10.1093/nar/gkj491)
42. Yang, W. Y., H. D. Wilson, S. P. Velagapudi and M. D. Disney. (2015), Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA, *J Am Chem Soc*, 137, 5336-5345 (10.1021/ja507448y)
43. Yang, W. Y., F. He, R. L. Strack, S. Y. Oh, M. Frazer, S. R. Jaffrey, P. K. Todd and M. D. Disney. (2016), Small Molecule Recognition and Tools to Study Modulation of r(CGG)(exp) in Fragile X-Associated Tremor Ataxia Syndrome, *ACS Chem Biol*, 11, 2456-2465 (10.1021/acscchembio.6b00147)
44. Disney, M. D., B. Liu, W. Y. Yang, C. Sellier, T. Tran, N. Charlet-Berguerand and J. L. Childs-Disney. (2012), A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome, *ACS Chem Biol*, 7, 1711-1718 (10.1021/cb300135h)
45. Su, Z., Y. Zhang, T. F. Gendron, P. O. Bauer, J. Chew, W. Y. Yang, E. Fostvedt, K. Jansen-West, V. V. Belzil, P. Desaro *et al.* (2014), Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS, *Neuron*, 83, 1043-1050 (10.1016/j.neuron.2014.07.041)
46. Wang, Z. F., A. Ursu, J. L. Childs-Disney, R. Guertler, W. Y. Yang, V. Bernat, S. G. Rzuczek, R. Fuerst, Y. J. Zhang, T. F. Gendron *et al.* (2019), The Hairpin Form of r(G4C2)(exp) in c9ALS/FTD Is Repeat-Associated Non-ATG Translated and a Target for Bioactive Small Molecules, *Cell Chem Biol*, 26, 179-190.e112 (10.1016/j.chembiol.2018.10.018)
47. Tran, T., J. L. Childs-Disney, B. Liu, L. Guan, S. Rzuczek and M. D. Disney. (2014), Targeting the r(CGG) Repeats That Cause FXTAS with Modularly Assembled Small Molecules and Oligonucleotides, *ACS Chemical Biology*, 9, 904-912 (10.1021/cb400875u)

48. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), *ACS Chemical Biology*, 13, 180-188 (10.1021/acscchembio.7b00699)
49. Mishra, S. K., N. Jain, U. Shankar, A. Tawani, T. K. Sharma and A. Kumar. (2019), Characterization of highly conserved G-quadruplex motifs as potential drug targets in *Streptococcus pneumoniae*, *Scientific Reports*, 9, 1791 (10.1038/s41598-018-38400-x)
50. Khan, E., A. Tawani, S. K. Mishra, A. K. Verma, A. Upadhyay, M. Kumar, R. Sandhir, A. Mishra and A. Kumar. (2018), Myricetin Reduces Toxic Level of CAG Repeats RNA in Huntington's Disease (HD) and Spino Cerebellar Ataxia (SCAs), 13, 180-188 (10.1021/acscchembio.7b00699)
51. Verma, A. K., E. Khan, S. K. Mishra, N. Jain and A. Kumar. (2019), Piperine Modulates Protein Mediated Toxicity in Fragile X-Associated Tremor/Ataxia Syndrome through Interacting Expanded CGG Repeat (r(CGG)exp) RNA, *ACS Chemical Neuroscience* (10.1021/acscchemneuro.9b00282)
52. M., M. G., G. D. S., H. R. S., H. Ruth, H. W. E., B. R. K. and O. A. J. (1998), Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *Journal of Computational Chemistry*, 19, 1639-1662 (doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B)

Chapter 7

Conclusion and Future Perspective

7.1 Conclusion of the thesis

Expansion of trinucleotide repeats in human genome caused several neurological and neuromuscular disorders. Generally, repeat containing messenger RNA is the key trigger for the diseased pathogenesis. Targeting repeat containing mRNA *via* small molecule modulators with high affinity and selectivity could open a new door for the therapeutic intervention. The current thesis work demonstrated the structural insight of CGG repeats RNA and discover several potent natural and synthetic small molecule modulators that selectively interact with r(CGG)^{exp} repeats RNA and improve physiological defects of Fragile X associated tremor/ataxia syndrome and Fragile X associated primary ovarian insufficiency. Initially, dynamics of GG mismatched pair of CGG repeat RNA were studied using NMR spectroscopy, conventional molecular dynamic simulation and Umbrella sampling techniques. We evidenced the importance of structural dynamic information of the expanded CGG repeats system to develop therapeutically potential small molecules. Furthermore, we have described the therapeutic potential of naturally available small molecules against trinucleotide disorders. For that a set of RNA motifs library were screened against previously reported neuroprotective binders as many of these RNAs motif cause neurological diseases. Piperine and Curcumin naturally available small molecules showed high affinity and selectivity against r(CGG)^{exp} RNA. The binding affinity was confirmed using other biophysical methods like Isothermal Calorimetry (ITC), Circular dichroism (CD), Melting study, gel based assays. The therapeutic efficacy was assessed in developed FXTAS cellular models and patient-derived cells for improving splicing defects and ameliorating formation of polyG aggregates.

Additionally, we have screened bioactive synthetic compounds with great binding potential and lesser cytotoxicity from National Cancer Institute database. We have used shape based similarity search approach using already known compound, Bisantrone as a molecular query. Further binding potential and selectivity of Bisantrone similar compounds were screened using fluorescence based assay and ITC. Next, only lead compounds interaction was affirmed by other biophysical methods. Lastly, these lead compounds were checked for the therapeutic potency using CGG repeat containing cellular model and FXTAS patient derived cells.

Conclusively, in this current thesis work provide the knowledge about the molecular dynamic of mismatch RNA and discovery of natural and synthetic small molecules using different approaches for screening such as shape similarity search and checking interaction of previously reported nucleic acid binder. Furthermore, similar approach can be explored for small molecule based therapeutic development of other trinucleotide disorders and/or repeat associate disorders.

7.2 Future scope

Trinucleotide repeat expansion causes different neurodegenerative disorder such as Huntington's disease (HD), Fragile X associated tremor/ataxia syndrome (FXTAS), several Spinocerebellar ataxia (SCAs) and Myotonic Dystrophy 1 & 2 (DM1 & DM2). Currently, therapeutics of neurological disorders is based on symptoms of diseased individuals. It would be better if we could target the actual triggers of these diseases. Different types of therapeutics agents were explored to target the nucleotide repeat-containing DNA and RNA such as natural and synthetic small molecules, antisense oligonucleotides (ASOs), RNA interference (RNAi), and CRISPR/Cas9 gene-editing system. Small molecules based therapeutic intervention is generally preferred over others as it is easily assessable, intracellular trafficking, and high blood brain retention. Many of such small molecules neurological drugs are in clinical trial phases.

Currently, researchers are focusing on a combination of small molecules for the therapeutics of neurodegenerative diseases. Targeting multiple selective triggers instead of single component in regulatory circuits could be beneficial for the treatment. Additionally, researchers are also aiming to develop dimer or trimer of similar or different small molecules with easy blood brain penetration to further enhance the selectivity and specificity at nanomolar. Thus, combination drugs therapy for TNRs associated neurological disorders would be one best advantageous approach. Conclusively, considerable efforts are required to reduce the off target effect of small molecules. Screened lead molecules should examine using disease specific cellular & animal models and clinically to transfer translational research from bench side to bed side.

Appendix A

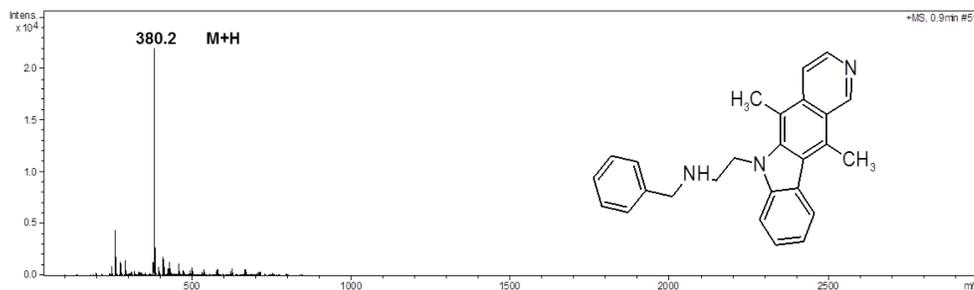
Excitation and emission wavelength of Bisantrene similar molecules used in chapter 6.

Compound ID	Manuscript ID	Excitation wavelength (λ_{\max} abs)	Emission wavelength (λ_{\max} emi)
Bisantrene		412 nm	510 nm
NSC164016	B1	290 nm	520 nm
NSC637991	B2	370 nm	545 nm
NSC374977	B3	370 nm	420 nm
NSC600305	B4	345 nm	465 nm
NSC166368	B5	405 nm	540 nm
NSC311152	B6	346 nm	450 nm
NSC401302	B7	320 nm	395 nm
NSC90847	B8	305 nm	460 nm
NSC150838	B9	300 nm	460 nm
NSC36879	B10	285 nm	335 nm
NSC196403	B11	405 nm	540 nm
NSC43592	B12	280 nm	320 nm

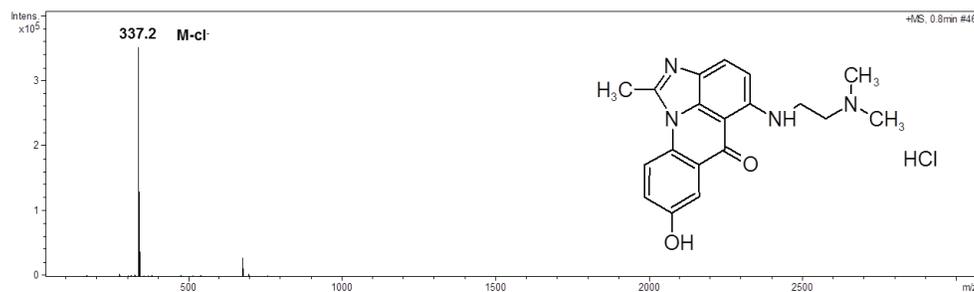
Appendix B

Mass spectrometry data of Bisantrene similar molecules for assesment of purity of compounds.

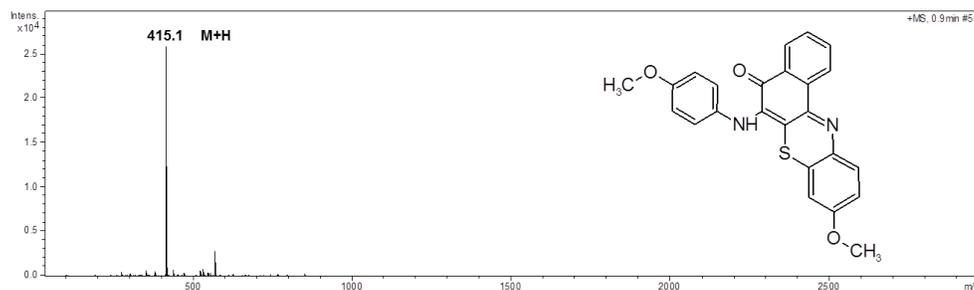
B1 = 164016, MW = 379.49 g/mol



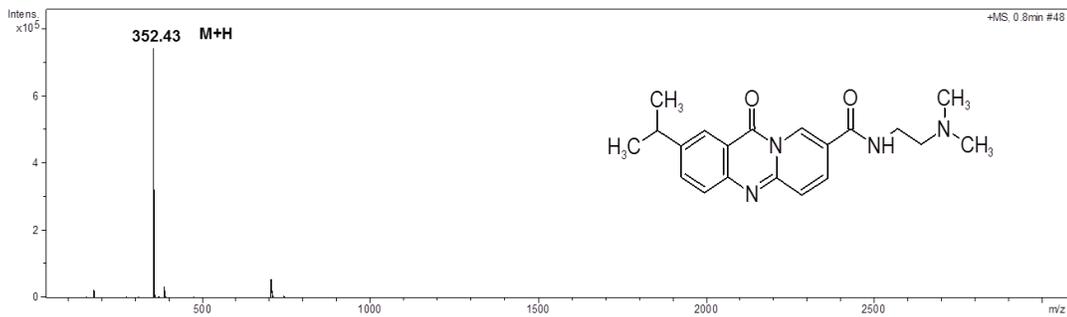
B2 = 637991, MW = 372.84 g/mol



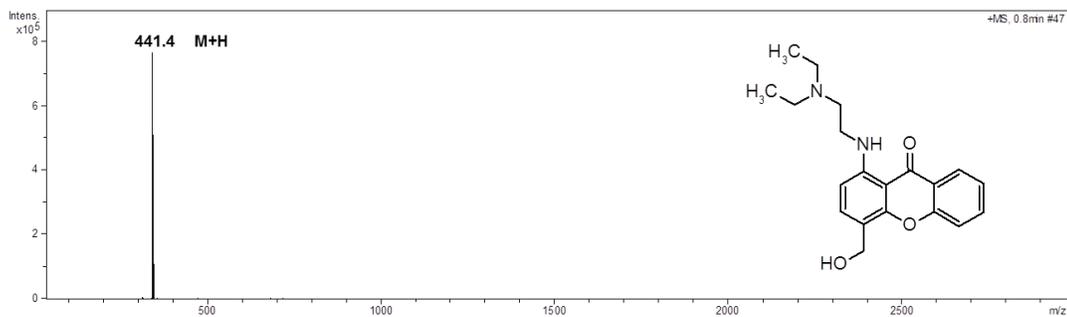
B3 = 374977, MW = 414.47 g/mol



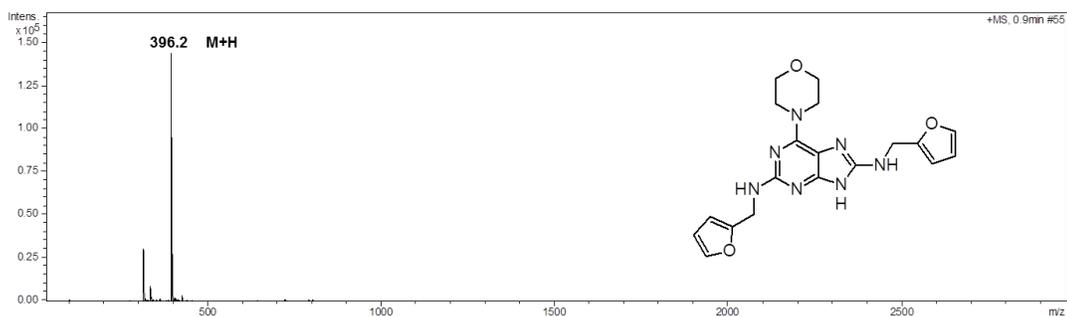
B4 = 600305, MW = 352.4 g/mol



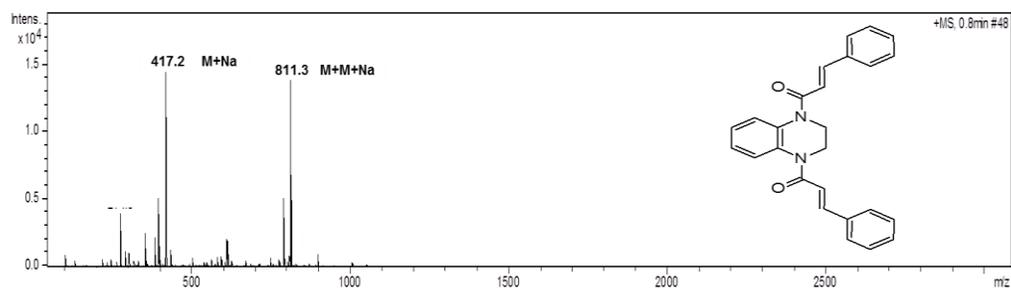
B5 = 166368, MW = 340.4 g/mol



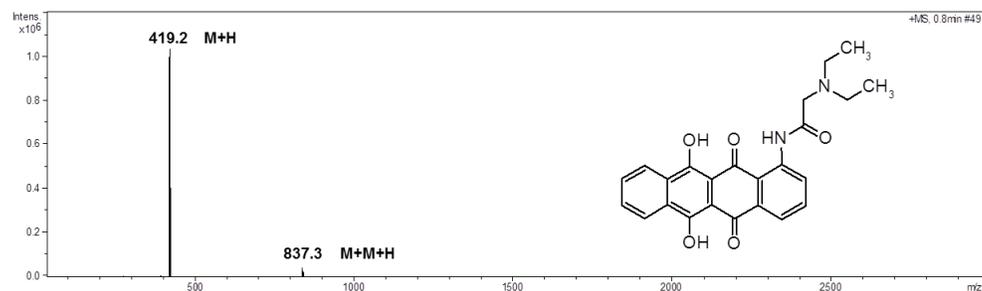
B7 = 401302, MW = 395.4 g/mol



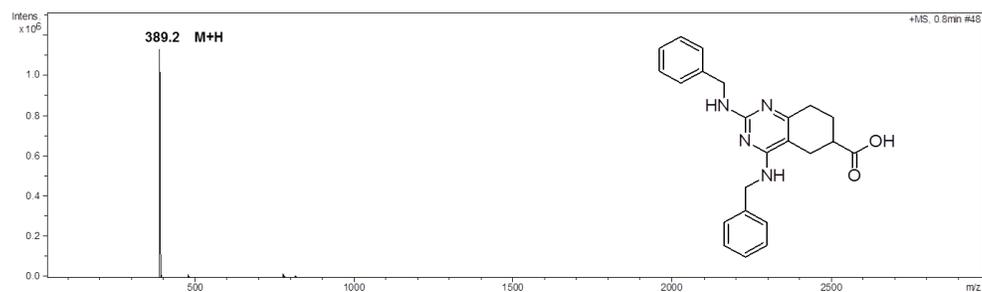
B8 = 90847, MW = 394.5 g/mol



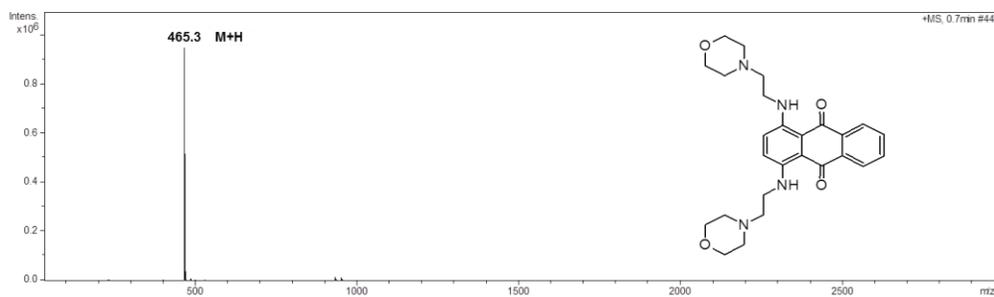
B9 = 150838, MW = 418.4 g/mol



B10 = 36879, MW = 388.5 g/mol



B11 = 196403, MW = 464.4 g/mol



B12 = 43592, Mw = 396.4 g/mol

