Probing the role of RNA-DNA hybrids in instigating trinucleotide repeat instability and their interaction with RNase H1

CH. NIMILITHA
BO12M1001

A Dissertation Submitted to

Indian Institute of Technology Hyderabad

In Partial Fulfillment of the Requirements for

The Degree of Master of Technology

भारतीय प्रौद्योगिकी संस्थान हैदराबाद

Indian Institute of Technology Hyderabad

Department of Biotechnology
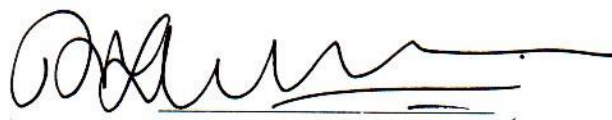
10-07-2014

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

<div align="right">

Ch. Nimilitha

BO12M1001

</div>

Approval Sheet

This thesis entitled "Probing the role of RNA-DNA hybrids in instigating trinucleotide repeat instability and their interaction with RNase H1" by Ch.Nimilitha is approved for the degree of Master of Technology from IIT Hyderabad.
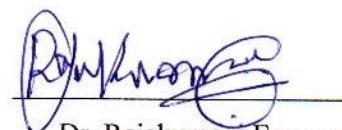
Dr. Thenmalarchelvi Rathinavelan
(Thesis advisor)
Department of Biotechnology
IIT Hyderabad

Dr. D.S. Sharada
Department of Chemistry
IIT Hyderabad

Dr. Anindya Roy
Department of Biotechnology
IIT Hyderabad

Dr. Rajakumara Eerappa
Department of Biotechnology
IIT Hyderabad

Acknowledgements

Abstract

R-loops are transient intermediates that are formed during the transcription and consist of an RNA-DNA hybrid that is formed between the nascent RNA strand and the template DNA strand and a displaced non-template DNA strand. Formation of R-loops have been detected in organisms from bacteria to humans. If R-loops form more frequently, they impact transcription affecting genome stability, genome integrity and cause a number of diseases. Recently, genome instability caused due to stable RNA-DNA hybrid in R-loop was shown to predominantly associate with expansion of trinucleotide repeats, leading to many incurable neurological and neuromuscular genetic disorders like Myotonic dystrophy1, Fragile X syndrome, Huntington's disease, Friedreich's ataxia, Spinocerebellar ataxias etc. Till date, the structural information about RNA-DNA hybrids formed by trinucleotide repeat expansions (TREs) are unknown to elucidate the mechanism behind RNA-DNA hybrid in instigating TREs. In this context, we aim here to study the structures of RNA-DNA hybrids consisting of trinucleotide repeats such as dGAA-rUUC, rGAA-dTTC, rCAG-dCTG, rCUG-dCAG, rCGG-dCCG and rCCG-dCGG by employing molecular dynamics (MD) simulation technique. It's noteworthy that trinucleotide repeat expansion disorders (TREDs) can also be treated at mRNA level, wherein, a short complimentary DNA oligonucleotide is targeted against the mRNA that is followed by the cleavage of the mRNA strand by RNase H1 enzyme. Thus, to understand the efficacy of RNase H1 against TRE containing RNA-DNA hybrid, the interaction of TRE containing RNA-DNA hybrid with RNase H1 is also probed. Such scenario comes into picture during replication, wherein a transient RNA-DNA hybrid is formed and subsequently, the RNA strand is cleaved by RNase H1. Thus, this investigation is expected to shed light on structural properties of TRE containing RNA-DNA hybrids as well as its complex with RNase H1 to understand the TRE mechanism at replication and transcription levels and their treatment. The study on structure and dynamics of RNA-DNA hybrids consisting of various TREs suggest that these hybrids have both A & B characteristics irrespectively of sequence. The study on HsRNase H1 complexed with TREs containing RNA-DNA hybrids suggest that the protein follow the same cleavage mechanism for all the hybrids irrespective of sequence and thus, antisense strategy can be utilized to treat TREDs at RNA level.

Abstract

Abbreviations

Antisense oligonucleotides	- ASOs
Double strand DNA breakage	- DSB
Molecular dynamics	- MD
Oligonucleotide	- ON
Periodic Boundary Conditions	- PBC
Trinucleotide repeats	- TNR
Trinucleotide repeats expansions	- TREs
Trinucleotide repeats expansion disorders	-TREDs
Wild-type	- wt

Preface

Chapter 1 - R-loop is formed during transcription and consists of a three stranded structure RNA-DNA hybrid along with a displaced DNA strand. Increase in the frequency of R-loops were found to impact genomic stability and cause a number of diseases. The association of these R-loop with trinucelotide repeat expansions (TREs) incorporates enhanced stability to the RNA-DNA hybrid and cause ~40 neurological and neuromuscular genetic disorders that can be life threatening. The structural information about the various TREs containing RNA-DNA hybrids is very little to understand the instability caused by them in presence of TREs. We here, survey to characterize the structure of RNA-DNA hybrids from all the PDBs available and study the dynamics of RNA-DNA hybrids containing various TREs like dGAA-rUUC, rGAA-dTTC, rCAG-dCTG, rCUG-dCAG, rCGG-dCCG and rCCG-dCGG by employing molecular dynamics (MD) simulation technique. As it was known from earlier that the RNA strand of RNA-DNA hybrid is cleaved by RNase H1 that is formed during replication as okazaki fragments. Thus, we carried out studies on Human RNase H1 complexed with above mentioned TREs containing RNA-DNA hybrids. This study can address an antisense strategy to treat trinucleotide repeat expansion disorders (TREDs) by using a short DNA oligonucleotide as an antisense strand for the mRNA produced from TREs containing RNA-DNA hybrids.

Chapter 2 is on methodology applied.

Chapter 3 includes systematic structural investigation carried out on RNA-DNA hybrids and characterization of structural features of these hybrids. Survey conducted revealed that RNA-DNA hybrids are in AB- intermediate form and the same were considered for modelling various TREs containing RNA-DNA hybrids.

Chapter 4 is on structure and dynamics of RNA-DNA hybrids with various TREs. The constructed models were simulated and results confirm that these hybrids have both A & B characters equally thus forming an AB-intermediate form. However, as mentioned earlier, TREs enhance the stability of RNA-DNA hybrids, we investigated stacking interactions for these hybrids containing TREs and was seen to have enhance stacking observed between neighboring steps. The repetitive stacking interactions was seen in RNA-DNA hybrids containing various TREs.

Chapter 5 comprises the study conducted on wild-type HsRNase H1 complexed with various TRE containing RNA-DNA hybrids. Docking of RNA-DNA hybrids with various TREs to wild-type HsRNase H1 was done. This study reveals that two metal ions are necessary for the catalysis to take place. Arg-179 is involved in bending the hybrid towards the protein to favor the catalysis. Cleavage mechanism was observed to be similar in all the models with various TREs. The present study suggests antisense therapy for treating TREDs at the RNA level. Few suggestions on the modification of DNA oligonucleotide that binds to the sense mRNA strand to favor the cleavage of RNA strand were given.

Poster presentation

"Probing the role of RNA-DNA hybrids in instigating trinucleotide repeat instability and their interaction with RNase H1" in *"The first annual TCIS summer research symposium"*,TIFR-TCIS Hyderabad, June 26, 2014.

Contents

Probing the role of RNA-DNA hybrids in instigating trinucleotide repeat instability and their interaction with RNase H1

1. Introduction

All types of life forms, starting from bacteria to human, store their genetic information in molecules of double-stranded DNA. The flow of this genetic information is well explained by the central dogma of molecular biology, which explains that DNA makes RNA makes protein. DNA synthesizes RNA by transcription which is translated into protein. During this transcription, RNA hybridizes with DNA to form RNA-DNA hybrid. RNA-DNA hybrids can also be observed during replication and reverse transcription. Thus, RNA-DNA hybrid has vital role in replication and transcription (Xiong and Sundaralingam, 1998). During replication, unwinding of duplex DNA by DNA primase generates ~11 nucleotide long RNA primers along the lagging strand allowing the synthesis of Okazaki fragment. These Okazaki fragments base pair with lagging DNA strand and generates RNA-DNA hybrids every 100-200 nucleotides. The reverse happens during reverse transcription of retrovirus (enveloped virus like HIV), wherein, a single stranded RNA generates a complimentary DNA strand that is catalyzed by the enzyme reverse transcriptase. Thus, the formation of RNA-DNA hybrid is an intermediate observed during the reverse transcription. However, the RNA-DNA hybrid formed during transcription is different from the above, as it consists of a RNA-DNA hybrid and a displaced DNA strand. During transcription, RNA polymerase binds to the duplex DNA, unwinds a few base pairs and incorporates ribonucleotides to the 3$'$ end of the growing RNA strand (Xiong and Sundaralingam, 1998). This RNA strand base pairs with the template DNA strand forming RNA-DNA hybrid. This, along with the displaced non-template DNA strand forms the R-loop.

1.1 R-loop

RNA polymerase binds to the initiation site of DNA duplex and unwinds ~17 base pairs of the DNA, forming the transcription bubble (Gnatt et al., 2001, Kettenberger et al., 2004). Within the transcription bubble, ribonucleotides base pair to the template DNA strand, which subsequently forms a RNA-DNA hybrid duplex leaving the displaced DNA strand as a single-strand (Li and Manley, 2006). This RNA-DNA hybrid duplex along with the displaced non-template DNA strand forms the R-loop (Figure 1.1) (Kogoma, 1997, Gowrishankar et al., 2013). RNA-DNA hybrids

formed at the R-loops usually, are of 8-9 base pair length and may be of significantly longer at times. For example, in the R-loop formed at the col E1 (an *E. coli* plasmid) origin of replication seems to be about 140 base pairs (Yu et al., 2006). Besides transcription, R-loops are also observed *in vivo* at the origins of prokaryotic and mitochondrial replications and mammalian immunoglobulin (Ig) class switch sequences encoding G-rich transcripts in activated B lymphocytes (Camps and Loeb, 2005, Clayton, 2000).
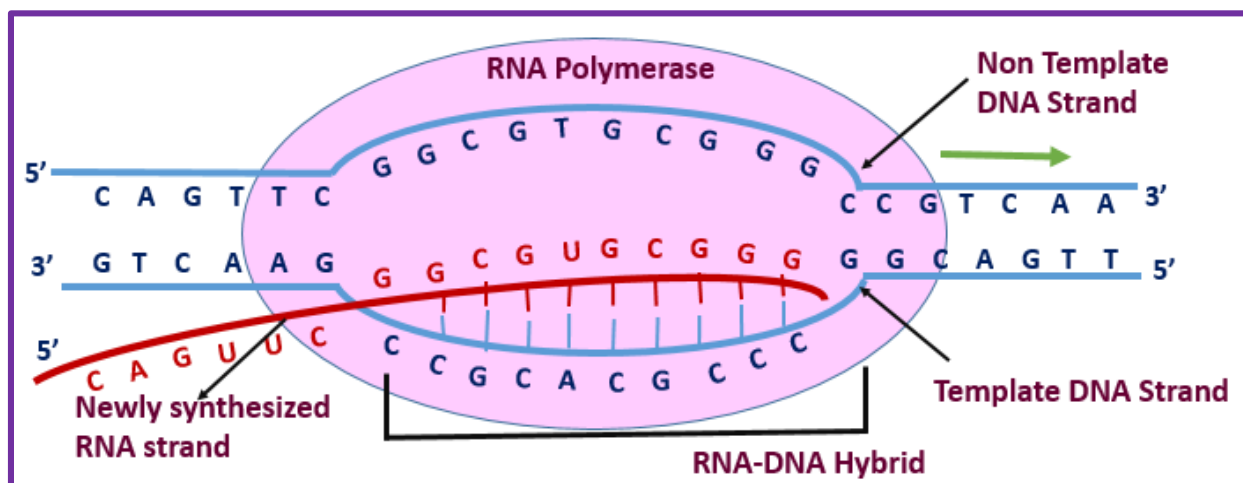
Figure 1.1 Schematic representation of an R-Loop formed during transcription. RNA polymerase (lavender ellipse) binds to the DNA duplex (colored blue) and initiates the transcription process. It unwinds the DNA duplex with the help of sigma factor (not shown). RNA polymerase reads the template DNA strand and adds the complementary RNA nucleotides to the 3'end of the newly synthesized RNA transcript (colored maroon), forming RNA-DNA hybrid. RNA-DNA hybrid along with non-template DNA strand is referred to as R-loop. Green arrow indicates the direction of the RNA polymerase movement. Concept from (McIvor et al., 2010, Lin et al., 2010)

R-loops are discovered recently and research is just underway to understand their mechanism of formation and function. Till date the most accepted mechanisms are 'thread back' or 're-annealing' model (Figure 1.2A) and 'extended hybrid' or 'bubble extension' mechanisms (Figure 1.2B) (Westover et al., 2004, Roy and Lieber, 2009, Roy et al., 2008). In thread back mechanism, the RNA remains single stranded for short duration and then it re-anneals with the template DNA strand, as where in the extended hybrid mechanism, after transcription, the RNA synthesized after transcription still remains associated with template strand as a persistent RNA-DNA hybrid within the transcription elongation complex even after the complex continues to translocate along DNA (Gowrishankar et al., 2013).

A)

'Thread back' mechanism

B)

'Extended hybrid' mechanism
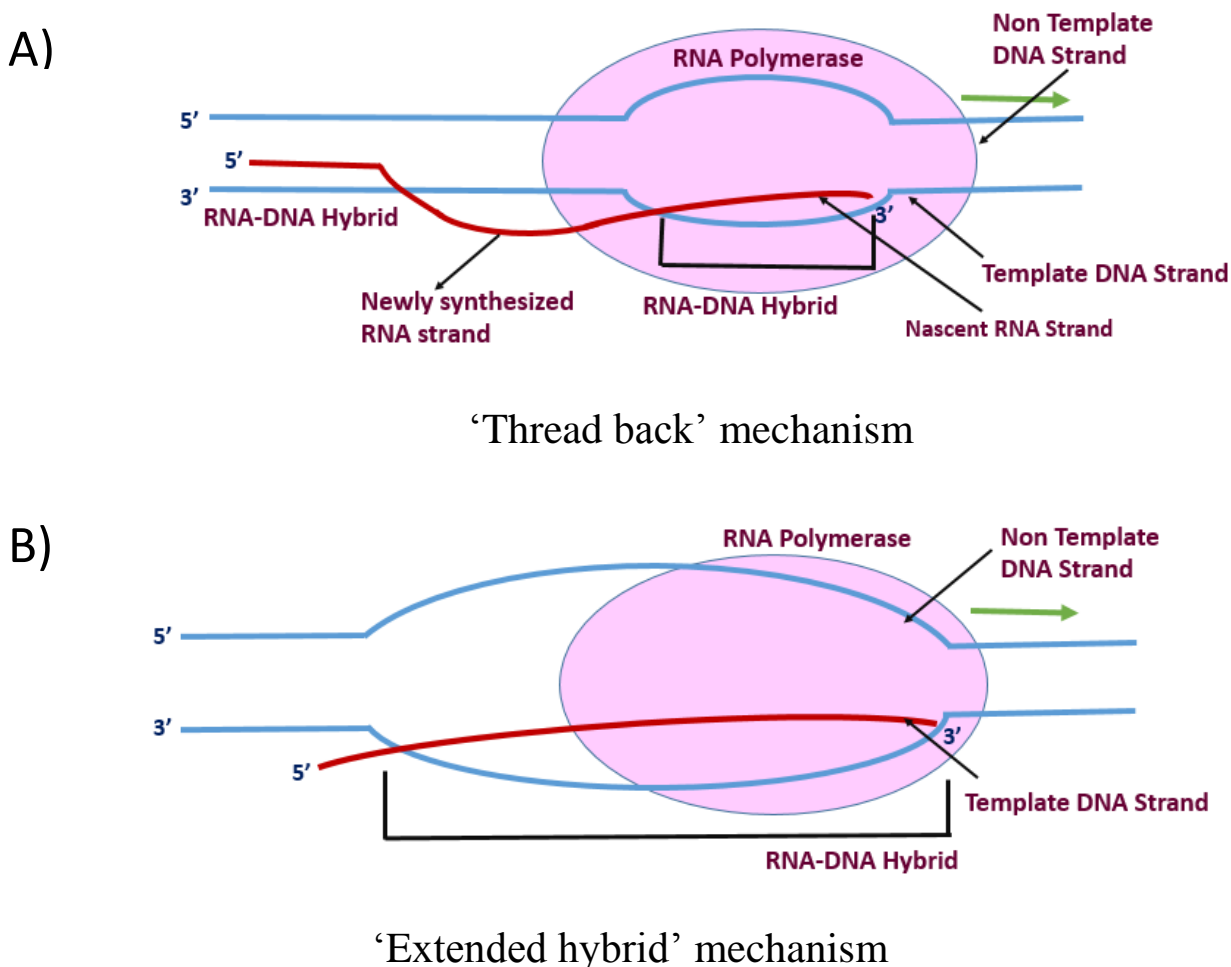
Figure 1.2. Mechanisms for R-Loop formation during transcription. Two models of R-loop formation are schematically depicted, as explained in the text: (**A**) 'Thread back' mechanism where re-annealing of nascent or free transcripts takes place and (**B**) 'extended hybrid' mechanism, the RNA synthesized within the transcription elongation complex remains associated with template strand. DNA, RNA and RNAP are shown in blue, red and lavender respectively. Concept from (Gowrishankar et al., 2013).

Although R-loops are transient in nature, stable R-loops can also form when transcription arises at higher rates (Drolet, 2006). Not surprisingly, many studies revealed that stability of the RNA-DNA hybrid is also influenced by the length of the hybrid and the sequence composition (Shaw and Arya, 2008). For example, RNA-DNA hybrid formed by G-rich RNA strand and C-rich DNA strand (Figure 1) are much more stable than the isosequential DNA duplex (Roberts and Crothers, 1992, Sugimoto et al., 1995). Transcription-induced negative supercoiling can also lead to the formation of R-loops (Drolet, 2006). However, stable R-loops can be typically toxic when the stability of RNA-DNA hybrid is higher than the normal situation. For instance, chromosome-

associated proteins like single-strand DNA binding proteins may contribute to stable R-loop formation during transcription by stabilizing the single-stranded region on the non-template strand causing genomic instability like mutation and chromosomal rearrangement (Huertas and Aguilera, 2003, Li and Manley, 2006) and cells deficient of either RNase H1 or RecG (a helicase that is able to unwind R-loops), generate R-loops which intervene the replication fork's formation for chromosomal DNA replication (Gowrishankar et al., 2013). Similarly, high stability of R-loops have been observed in the overexpansion of trinucleotide repeats, which eventually leads to the repeat instability (insertion or deletion). In addition, high frequency of R-loop formation leads to replication fork stalling and collapse as well as threatens gene expression, genome integrity and causes genomic instability (Aguilera and Garcia-Muse, 2012).

How do transcriptional R-loops induce genome instability?

Genomic instability by transcriptionally induced R-loops interfere with the progression of DNA replication fork and transcription elongation from bacteria to humans (Gan et al., 2011). In eukaryotes, formation of transcriptional R-loop is highly correlated with genome stability indicating that R-loops have an inherent impact on the integrity of the genome. R-loops can be formed at various regions of the genome like promoter regions, exonic regions and untranslated regions (Camps and Loeb, 2005). So far, very little is known about the molecular mechanism(s), by which, transcriptional R-loops influence genome stability as discussed below. For instance, it is assumed that, due to the presence of extensive R-loops, transcribed regions become more vulnerable to DNA-damaging agents by increasing the regularity of single-stranded regions. R-loops may be precisely recognized by few Protein factor(s), like activation induced deaminase (AID), and may induce the formation of mutagenic/recombinant DNA lesions (Li and Manley, 2006). The exceptional stability of the rG/dC base pairs (Sugimoto et al. 1995) favors RNA-DNA over DNA-DNA hybrids in regions having G-rich nontemplate strand thus, facilitates R-loop formation (Gritzmacher 1989). Yet another, R-loop-mediated DNA damage possibly involves DNA replication. For specificity, transcription elongation complexes halted or slowed by R-loop structures might block the progress of replication forks and lead to replication blockage. Replication blockage is lethal if left unresolved and is expected to lead to DNA recombination and/or DNA double strand breaks (DSBs) in the newly replicated DNA (Li and Manley, 2006). RNA- DNA hybrids can be eliminated by ribonuclease H treatment. Mutation in the rnhA1 gene

decreases the activity of ribonuclease H1 stimulating the formation of R-loops (Lin and Wilson, 2007). Very recent study reveals that transcription-induced RNA-DNA hybrids would extend the lifetime of single-stranded repeat DNA, thereby permitting increased formation of non-B DNA secondary structures and at the trinucleotide overexpansion site promotes repeat instability of trinucleotide repeats sequences (Lin et al., 2010).

Thus, The current study aims to understand the role of RNA-DNA hybrid that encompasses various trinucleotide repeats in triggering repeat instability.

1.2 Trinucleotide repeat expansion (TRE)

Instability of simple repeating sequences or microsatellites are responsible for a number of human diseases either because of expansion or contraction. Simple repeating sequences are observed at several genomic sites like promoter regions, exonic regions and untranslated regions of all organisms, and are considered as important source of genetic variation. Simple repeats are classified into mono, di, tri, tetra, penta and hexa depending on the length of the repeating unit. Changes in the number of repeat units in a particular gene can alter its expression and the function of the RNA and/or protein it encodes. The most common unstable disease associated DNA repeats are trinucleotide repeats, which include CAG, CTG, CGG and GAA triplets. For instance, a repeated trinucleotide sequence (GTC) adds a series of the amino acid aspartic acid to the resulting protein.

Figure 1.3 Repeat expansion. (**A**) Normal process of protein synthesis Trinucleotide and the expansion of poly(GTC) are depicted and (**B**) a GTC trinucleotide sequence adds poly(aspartic acid) to the resulting protein.

Abnormal expansion of these trinucleotide repeats (TNR) leads to the formation of unusual secondary structures like hairpins (Figure 3). Such overexpansion results in more than 40 neurological and neuromuscular disorders (Pearson et al., 2005, McIvor et al., 2010, Groh et al., 2014), commonly known as trinucleotide repeat expansion disorders (TREDs). Some of the examples of disorders caused due to TREs are myotonic dystrophy (DM1), spinocerebellar ataxia (SCA), fragile X syndrome (FRAXA), fragile XE syndrome (FRAXE), Friedreich ataxia (FRDA) & Huntington's disease (HD).

Broadly, TREDs are classified into two distinct classes. Firstly, there are the polyglutamine repeat disorders, in which the expanded repeat is translated into an expanded polyglutamine (polyQ) tract and is thought to result in a toxic gain of function of the mutant protein. The second class, in which the trinucleotide repeat is translated into an expanded non-polyglutamine (non-polyQ) tract. The non-polyQ subset of trinucleotide repeats are further divided into two classes, one with repeats containing polyalanine repeats, polyalanine repeat disorders and the other with either glutamine or alanine repeats named non-polyglutamine repeat disorders. Summary of different types of TREDs can be illustrated in Table 1.1, 1.2 and 1.3

These disorders differ from each other by molecular mechanisms and pathogenicity. The mechanism of these diseases depends on expandable repeats and their precise location in the genome (Cummings and Zoghbi, 2000). TNR tracts are short and stable and doesn't show any length changes in case of nonaffected individuals while in affected individuals, the tracts are longer and unstable, have a strong tendency to expand rather than to contract. These length changes are dynamic and are seen between individuals and also between and within tissues of the same individual (Pearson et al., 2005). Disease symptoms can become more severe if these repeat tracts are transmitted through generations.

The repeats are transcribed in either one or both directions and may result in bidirectional transcription. For example, both strands of the expanded (CAG)-(CTG) tract of the myotonic dystrophy (DM1) disease locus are transcribed. A similar bidirectional transcription of both strands of the expanded spinocerebellar ataxia type 1 [(SCA8) (CAG)-(CTG) tract], Huntington's disease [(CTG)-(CAG)], [(CGG)-(CCG)]s for fragile X type E (FRAXE), has also been reported. To be much specific, it is the CCG strand in the FMR2 gene that is transcribed, while in fragile X type A (FRAXA) it is the CGG strand in the FMR1 gene that is transcribed.

1.2.1 Trinucleotide repeat expansion mechanisms

Overexpansion of trinucleotide repeats results in the formation of unusual secondary structures such as hairpin, triplex, quadruplex and i-motif. Such unusual structures play vital role in causing repeat instability at various biological processes like replication, repair, recombination and transcription (Voineagu et al., 2009).

1.2.1.1 Replication

During the process of replication, presence of trinucleotide repeat expansions (TREs) in the nascent lagging strand or its template strand, forces the strand to fold back and form unusual non-B-DNA structures such as hairpins, slipped-structure, triplexes and quadruplexes. These non-B-DNA structures obstruct the progression of the replication fork by stalling the movement of the DNA polymerase. This lead to either expansion or contraction depending on their existence in the template or nascent strand (Mirkin, 2007). Presence of TREs in the template strand results in repeat contractions (Mirkin, 2007). In contrast, expansions can be observed in nascent lagging strand (Figure 1.4).
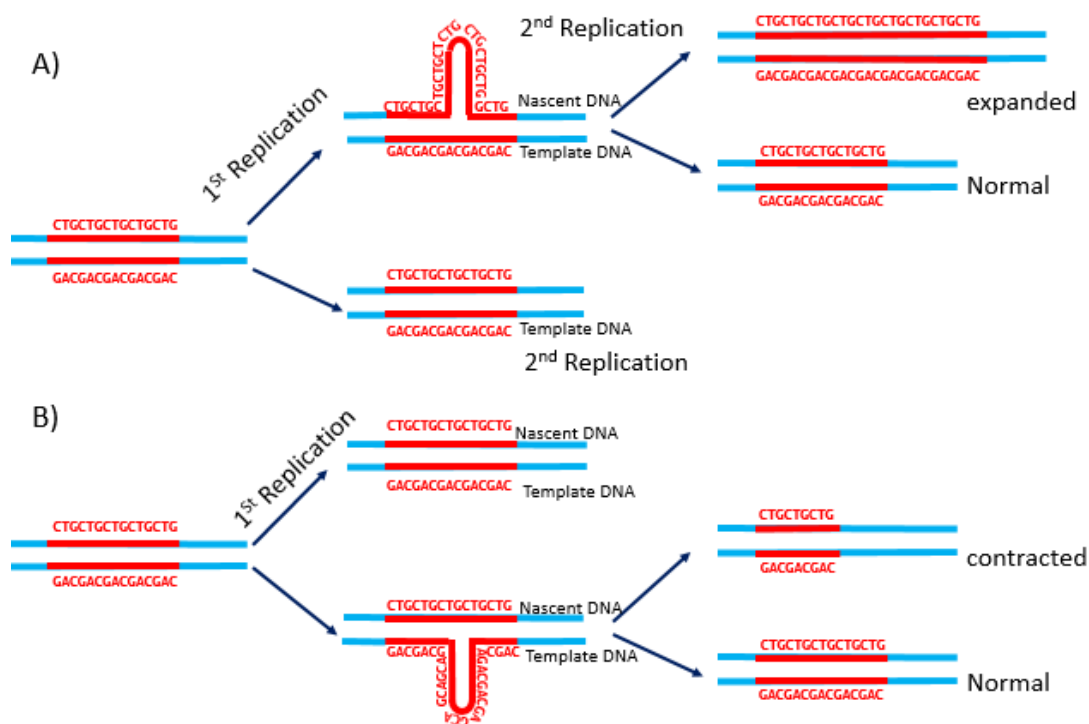

Figure 1.4 Hairpin-induced trinucleotide repeat instability during replication. The TRE is indicated by red, flanking DNA by blue lines. (a) Hairpin formation nascent-strand results in over-replication and in second round of replication, the hairpin strand fixes the expanded allele in the genome. (b) Template-strand hairpin formation results in under-replication and in second round of replication of the non-hairpin strand fixes the contracted allele in the

1.2.1.2 Recombination

Double strand DNA breakage (DSB) occurs due to the formation of non-B-DNA structures of TNRs containing strands. These DSBs are potential instigators for genetic recombination leading to repeat expansion (Wells et al., 2005, Kim and Mirkin, 2013). DSB facilitates strand invasion and gap filling processes. During gap filling, formation of non-B-DNA structures like slippage takes place resulting in the triplet repeat expansion.

1.2.1.3 Repair process

Repeat instability also occurs due to the DNA repair mediated abnormal expansion. The presence of hairpin structure in the strands manipulates the mismatch repair proteins (MMRP) like Mut-S homologue 2 (MSH2) and Mut-S homologue 3 (MSH3) (Mirkin, 2007, Sinden, 2001). DNA polymerase introduces the complementary base to the unfolded hairpin through an 'error-prone' mechanism. Thus, the overexpansion takes place in the nascent strand and thus, the overexpansions happens for generations. Mutation in the genes coding for MSH2 or MSH3 significantly reduces the trinucleotide repeat expansion.

Pathogenesis

1.2.1.4 Polyglutamine (PolyQ) Diseases

Increase in the number of CAG repeats is supposed to cause nine neurologic disorders. The expanded CAG repeats are translated into a series of uninterrupted glutamine residues forming a polyglutamine tract (polyQ). Such polyglutamine tracts may be subject to increased aggregation causing polyglutamine repeat diseases. Table 1.1 gives a brief discription about polyQ disorders that are related to the present study are only described.

Table 1.1. Various polyglutamine disorders and their general information				
Disease code	Disease name	Gene code	Normal repeats	Pathogenic repeats
DRPLA	Dentatorubropallidoluysian atrophy	ATN	1 6 - 35	49 - 88
HD	Huntington's disease	HTT	10 - 35	35+
SBMA	Kennedy disease (Spinobulbar muscular	HS-AR	9 - 36	38 - 62

SCA1	Spinocerebellar ataxia Type 1	ATXN1	6 - 35	49 - 88
SCA2	Spinocerebellar ataxia Type 2	ATXN2	14 - 32	33 - 77
SCA6	Spinocerebellar ataxia Type 6	CACNA1	4 - 18	21 - 30
SCA7	Spinocerebellar ataxia Type 7	ATXN7	7 - 17	38 - 120
SCA17	Spinocerebellar ataxia Type 17	TBP	25 - 42	47 - 63
SCA3	Machado-Joseph disease (Spinocerebellar ataxia Type 3)	ATXN3	12 - 40	55 - 86

The classic example of polyQ repeat disorder is Huntington's disease. Others include the spinocerebellar ataxias (SCAs) 1, 2, 3, 6, 7 and 17, dentatorubral pallidoluysian atrophy (DRPLA) and Kennedy's disease. These diseases are all characterized by progressive neurodegeneration and causes increasing neuronal dysfunction and eventual neuronal loss 10–20 years after onset of symptoms and the presence of protein aggregates or nuclear inclusions (NIs). The polyglutamine repeats cause a gain of function (Cummings and Zoghbi, 2000).Extended polyglutamine tract is a transglutaminase (tTG) substrate that can become cross-linked via ″ε-(γ -glutamine) lysine isopeptide bonds. The cross-linked products along with other copolymers lead to the formation of aggregates.

A) Huntington's disease

Huntington's disease (HD) was described in 1872 by George Huntington. The HD locus was localized to chromosome 4p16 in 1983 (Gusella et al., 1983). In unaffected chromosomes, the CAG repeat in the first exon of the *HD* gene varies from 6 to 35 repeats while in the disease state from 36 to 121 repeats. HD patients have Slight and sporadic motor disability and accompanied by memory deficits, cognitive decline, affective disturbance and/or changes in personality, rigidity and dystonia may also be present. Huntington may play a role in a number of nuclear and cellular events (Cummings and Zoghbi, 2000, Orr and Zoghbi, 2007, Everett and Wood, 2004) .

B) Spinobulbar muscular atrophy (SBMA or Kennedy Disease)

The expanded CAG repeat tract in SBMA patients is located in the first coding exon of the androgen receptor (*AR*) gene. The AR is a steroid hormone-activated transcription factor composed of hormone and DNA-binding domains. It has X-linked recessive pattern of inheritance and is

characterized by degeneration of the anterior horn, bulbar region, and dorsal root ganglion. Male patients often have muscle cramping that leads to proximal muscle weakness and atrophy. They also have difficulties in speech, articulation, and swallowing. Mild androgen insensitivity are typically seen at adolescence, including gynecomastia and testicular atrophy.

C) Spinocerebellar ataxia Type 1 (SCA1)

CAG repeat expansion in the coding region of the *SCA1* gene results in the synthesis of a polyQ track located in the amino-terminal half of ataxin-1. Normal alleles contain 6–44 repeats while uninterrupted 39 to 82 CAG repeats are seen in diseased condition. The primary clinical features include progressive cerebellar ataxia, dysarthria, eventual bulbar dysfunction, pyramidal signs and peripheral neuropathy. SCA1 is characterized by cerebellar atrophy with severe loss of Purkinje cells, dentate nucleus neurons, and neurons in the inferior olive and cranial nerve nuclei III, IV, IX, X, and XII.

D) Spinocerebellar ataxia type 2 (SCA2)

CAG repeats in SCA2 disease are located in the N-terminal coding region of *SCA2*. Normal alleles may contain 15–31 repeats and are normally interrupted by two CAA sequences. Disease alleles contain a perfectly uninterrupted 36 to 63 CAG repeat tract. Slow saccadic eye movements, hypo- or areflexia and ophthalmoparesis are features of SCA2. Degeneration of cerebellum and brain stem is shown with loss of purkinje and granule cells.

E) Machado-Joseph disease: spinocerebellar ataxia type 3 (SCA3)

The polyglutamine repeat is seen near the C terminus of *SCA3/MJD1* gene. Normal individuals have between 12 and 40 glutamines, and in disease the repeat expands to a length of 55–84 glutamines. Bulging eyes, facial & lingual fasciculation, rigidity and progressive ataxia are symptoms. Degeneration is most prominent in the basal ganglia, brain stem, spinal cord and dentate neurons of the cerebellum.

F) Spinocerebellar ataxia type 6

CAG repeats in SCA6 are range from 21 to 33 repeats in the disease state and <18 repeats on normal chromosomes. Clinical features in SCA6 consist predominantly of cerebellar dysfunction,

very slow progressive or episodic ataxia. Neuropathological findings in SCA6 include marked cerebellar atrophy with loss of cerebellar Purkinje cells and only moderate loss of cerebellar granule cells, dentate nucleus neurons and neurons of the inferior olive.

G) Spinocerebellar ataxia type 7

The SCA7 mutation is the most unstable in the polyglutamine disease family. The polyglutamine tract is observed near the amino terminus of the *SCA7* gene product, ataxin-7. All normal alleles contain pure 4–35 CAG repeats tracts whereas disease occurs when the repeats exceed 37. SCA7 patients have cerebellar ataxia and/or visual problems, including pigmentary macular degeneration. Primary neuronal loss is in the cerebellum, inferior olive, and some cranial nerve nuclei, Hypomyelination of the optic tract, gliosis of the lateral geniculate body and visual cortex are all observed.

H) Dentatorubropallidoluysian atrophy (DRPLA)

The Q repeat is found near the C terminus of the *DRPLA* gene product, atrophin-1 in case of DRPLA. Healthy individual contains 6–35 copies of CAG repeats whereas diseased individuals have 49–88 repeats. Characteristics include choreoathetosis (occurrence of involuntary movements), myoclonic epilepsy, dementia, and progressive intellectual deterioration in juvenile-onset cases. Neuronal loss in the cerebral cortex, cerebellar cortex, globus pallidus, striatum, and the dentate, subthalamic, and red nuclei are noticed.

1.2.1.5 Non-polyglutamine diseases

The other subset of trinucleotide repeat disorders are those which do not involve the CAG triplet in the coding region of the gene are referred to as the non-polyglutamine disorders. Polyalanine diseases are also included in this category. Table 1.2 gives a brief description about non-polyQ disorders.

Table 1.2 Various non-polyglutamine disorders and their general information						
Disease code	Disease name	Gene code	Repeat	Normal repeats	Pathogenic repeats	location
FRAXA	Fragile X syndrome	FMR1	CGG	6–53	>230	5' UTR

FRAXE	Fragile XE syndrome	FMR2	CCG	6-35	>200	5'-UTR
Frataxin	Friedreich ataxia	X25	GAA	7-34	>100	Intron1
DM	Myotonic dystrophy	DMPK	CTG	5-37	50-thousands	3'-UTR
SCA8	Spinocerebellar ataxia Type 8	IOSCA	CTG	16-37	110 to <250	3'-Terminal exon

A) Fragile X syndrome (FRAXA)

FRAXA is associated with a folate-sensitive fragile site at Xq27.3 (CGG)n repeat present in the 5'-UTR of the fragile X MR gene (*FMR1*). Normal alleles contain 6 to 53 CGG repeats interrupted by one or more AGGs. Expansion of the CGG repeat beyond 230 trinucleotides cause disease. It is an X-linked disorder and is typically seen in males. Common features include Mitral regurgitation (heart disorder, in which, the mitral valve does not close properly when the heart pumps out blood) post pubescent macroorchidism, long & prominent ears and jaws, high-pitched speech, hyperactivity, poor eye contact, and stereotypic hand movements

B) Fragile XE mental retardation (FRAXE)

Expansion of GCC repeat in a folate-sensitive fragile site immediately adjacent to a CpG island causes FRAXE. The repeat size varies from 6 to 35 copies on normal alleles and may expand upto >200 copies in diseased allele. These GCC repeats reside in the promoter region of *FMR2* gene. Patients have mild mental retardation, have learning deficits, and may be developmentally delayed.

C) Friedreich ataxia (FRDA)

FRDA ooccur due to GAA repeat expansion in the *X25* gene (*frataxin*) and is autosomal recessive disorder. The repeat expansions occur in intronic regio. Affected individuals carry two mutant alleles. Normal alleles contain 7–34 repeats, whereas disease-causing alleles contain >100 repeats. The disease is characterized by unremitting gait, limb & truncal ataxia, diminished tendon reflexes, loss of position and vibratory senses, dysarthria, cardiomyopathy, and diabetes mellitus; less commonly nystagmus, optic atrophy, scoliosis, and/or skeletal abnormalities are evident. FRDA is a multisystem disorder. Disease onset is early childhood, and the combination of neurological impairment, cardiomyopathy and diabetes shortens life span.

D) Myotonic dystrophy (DM)

DM is an autosomal dominant, multisystem disorder with expanded CTG repeat tract in the 3'UTR of a protein kinase gene, *DMPK*. Wild-type chromosomes have 5–37 CTGs while DM patients have 50 to several thousand repeats. Muscle weakness, progressive muscle wasting, facial dysmorphology, presenile cataracts, testicular atrophy, premature balding in males, kidney failure, hyperinsulin secretion, and cardiac conduction abnormalities primarily characterize DM.

E) Spinocerebellar ataxia type 8 (SCA8)

SCA8 occur due to an expansion of a noncoding CTG expansion and is present in the 3' terminal exon of a processed transcript. SCA8 is a dominantly inherited disease. Characterized by variable degrees of degeneration of the cerebellum, spinal tracts, and brain stem. SCA8 is a slowly progressive ataxia with cerebellar atrophy, decreased vibration sense, and sometimes brisk reflexes.

Polyalanine diseases

Beside the well-known polyglutamine expansions, recent findings pointed to the expansion of polyalanine stretches such as GCU, GCC, GCA & GCG as a disease mechanism. Table 1.3 gives a brief description about polyalanine disorders.

Table 1.3 Various polyalanine disorders and their general information				
Disease code	Disease name	Gene code	Normal repeats	Pathogenic repeats
BPES	Blepharophimosis-ptosis-epicanthus inversus syndactyly	FOXL2	14	19-24
HPE5	Holoprosencephaly 5	ZIC2	15	25
CCHS	Congenital failure of autonomic control	PHOX2B	20	25-33
ISSX	X-linked infantile spasm syndrome	ARX	16	27
MRGH	X-linked mental retardation with isolated growth hormone deficiency	SOX3	15	22-26
CCD	Cleidocranial dysplasia	RUNX2	17	27

HFGS	Hand-foot-genital syndrome	HOXA13	18	24-26
SPD1	Synpolydactyly 1	HOXD13	15	22-29
OPMD	Oculopharyngeal muscular dystrophy	PABPN1	10	11-17

The present study would deal with Polyglutamine diseases and in association of RNA-DNA hybrids.

1.3 RNA-DNA hybrid and repeat instability

Aforementioned, the RNA–DNA hybrid formed via R-loop mechanism has strong association in causing repeat instability in TRE containing sequences due to the stronger affinity of nascent RNA towards the template DNA strand (Figure 4) (Wongsurawat et al., 2012, Reddy et al., 2011). Thus, reducing the stability of RNA-DNA hybrid can be one of the therapeutic approaches to treat the neurological disorders caused by TREDs. However, structural information of TRE containing RNA-DNA hybrid duplexes are poorly understood. In this context, we aim here to study the conformational dynamics of various RNA-DNA hybrids associated with myotonic dystrophy (DM1), fragile X syndrome and friedreich ataxia etc. by employing molecular dynamics simulation technique.
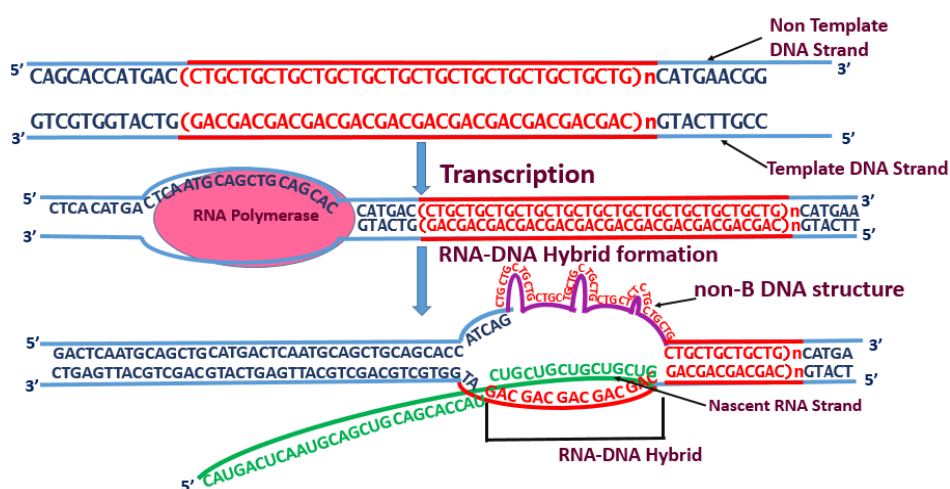
Figure 5. R-loop mediated trinucleotide repeat instability. (Top) Parent DNA strands that contain $(CTG)_n.(CAG)_n$ trinucleotide repeats (shown in red), where 'n' indicates trinucleotide repeat length. (**Middle**) Formation of transcription bubble mediated by RNA polymerase (pink ellipse). **(Bottom)** R-loop that encompasses a stable RNA-DNA hybrid between the template DNA (colored red) and the nascent RNA (colored green) is shown. The non-template DNA strand that adopts non-B DNA structure is indicated in purple. Concept from (Lin et al., 2010)

1.4 <u>Structure of RNA-DNA hybrid</u>

On the basis of earlier X-ray fiber diffraction studies, RNA-DNA hybrids were initially assumed to be similar to RNA in conformation. Later it has been revealed from NMR, fiber diffraction and X-ray diffraction studies that RNA-DNA hybrid conformation is intermediate between those of A form RNA (Figure 6A) and B form DNA (Figure 6B), the so called AB intermediate (Figure 6C) (Shaw and Arya, 2008) as the helical twist (31°-35°) and the X-displacement (-3.2Å to -4.1Å) of the hybrid mimic the RNA duplex. On the other hand, the helical rise (3.4Å) and the major groove width of hybrid mimic the DNA duplex.
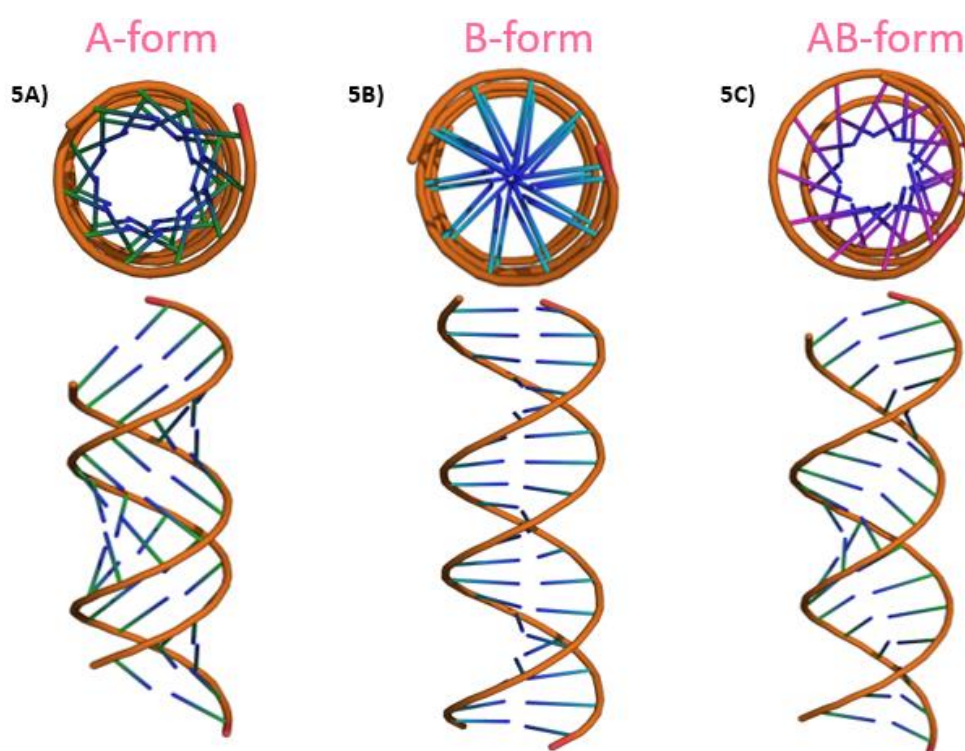
Figure 5. Cartoon representation of (Left) A, (Middle) B and (Right) AB intermediate forms of nucleic acids. View (**Top**) along and (**Bottom**) perpendicular to the helical axis are shown. Note that for the illustration of the figure **(Left)** RNA, **(Middle)** DNA and **(Right)** RNA-DNA hybrid molecules are considered as they favor A, B and AB-intermediate forms.

In an RNA-DNA hybrid, the DNA strand of the hybrid is anticipated to be in very flexible conformation compared to the RNA strand (Hantz et al., 2001). This can be clearly visualized from the sugar pucker preference of the DNA and RNA strands of the hybrid. While the former takes up variety of sugar pucker conformation like $C_3'endo$, $O_4'endo$, $C_1'exo$ and

C2'*endo* sugar puckers, the latter has strong preference for C_3'-*endo* sugar pucker (Sanghani and Lavery, 1994). Various tpes of sugar puckerings are shown in Figure 5.

1.5 TRE containing RNA-DNA hybrid and therapeutics of TREDs

Reducing the stability of RNA-DNA hybrid in the R-loop can be one of the therapeutic approaches to treat the neurological disorders caused by TREs. Besides targeting the RNA-DNA hybrid containing TREs to treat TREDs, an alternate idea would be an allele selective therapeutics where expression of the disease allele is selectively prevented while maintaining the expression of the wild-type (wt) variant same (Ostergaard et al., 2013). Oligonucleotide (ON)-based therapeutics, the well-known antisense strategy of gene regulation is uniquely suited for targeting autosomal diseases, as they can suppress production of the mutant protein or RNA by targeting the mRNA directly through Watson–Crick interactions (O'Connor and Crystal, 2006). The antisense strategy involves hybridizing complementary DNA sequence with the target mRNA forming RNA-DNA hybrid and thus, inhibiting the translation of the target protein. Subsequently, the RNA-DNA hybrid formed through antisense strategy is susceptible to RNase H digestion leading to the degradation of RNA strand (Lee et al., 2012). RNase H ASOs are also chemically modified for improved drug-like properties. The injected ASOs distribute throughout the brain and spinal cord and silence gene expression in most parts of the CNS. Moreover, these chemically modified ASOs have successfully completed early stage human clinical trials for the treatment of familial amyotrophic lateral sclerosis and spinal muscular atrophy. Thus, this warrants detailed structural information about the interaction between various TREs containing RNA-DNA hybrids and RNase H, which would facilitate the design of antisense DNA to target the TRE containing mRNA. Such information would further be useful to understand the sequence specific interaction of various TRE containing RNA-DNA hybrid with RNase H in the context of replication process, wherein, the RNase H binds to the hybrid and cleaves the RNA strand during the lagging strand synthesis. Similarly, during transcription, RNase H process R loops to modulate replication initiation and restore DNA topology. Since, overall study involves TREs in R-loops in humans, it is more relevant to study the role of human RNase H (Hs RNase H) in degrading the RNA strand of the hybrid. The structures of several other RNases H have been determined so far, including those from Escherichia coli, human immunodeficiency virus type 1, Moloney murine leukemia virus,

Thermus thermophiles HB8, archaeal RNase HII, and Bacillus halodurans RNase H bound to an RNA/DNA hybrid.

1.6 Structure of human RNase H1

As mentioned above, information about RNA-DNA hybrid containing various TREs and RNase H would enable the understanding of antisense strategy for treating TREDs, lagging strand replication process involving TREs and instability caused by transcriptional R-loop with TREs containing RNA-DNA hybrid. Since abnormal expansion of trinucleotide repeats leads to several neurological and neuromuscular disorders, its more appropriate here to discuss about human RNase H (Hs RNase H) and is discussed below.

RNase H can be classified into 2 major classes as RNase H1 and RNase H2. Both share overall similar fold but differ in the active site configuration and substrate preference. The endonuclease activity of RNase HII on RNA-DNA hybrids are minimal (Itaya, 1990). Hence, the current focus is on RNase H1 and is discussed here. Hs RNase H1 contains an N-terminal domain of ~50 amino acids that binds with RNA and RNA-DNA hybrid, known as RNA, RNA-DNA hybrid binding domain (RHBD) (Figure 6) (Cerritelli and Crouch, 1995, Gaidamakov et al., 2005). N-terminal RHBD is connected to the C-terminal catalytic domain (consists of ~150 amino acid) through a flexible linker that encompasses ~60 amino acids. Crystal structure of Hs RNase H catalytic domain (Hs RNase HC) indicates that the central element of Hs RNase HC is a mixed ß sheet comprising five α helices (Figure 5) (Nowotny et al., 2007). Human RNase H contains a ''basic protrusion,'' which is present in the E. coli enzyme but absent in Bh-RNase H and HIV RT, leading to different substrate recognition patterns between human and viral RNases H.
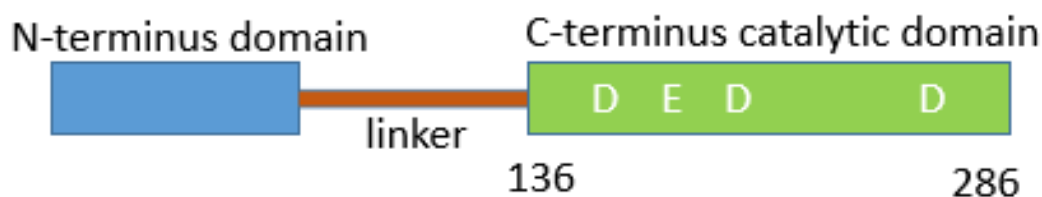

Figure 6. Schematic representation of the domain structure of human RNase H1. The N-terminus domain enhances the processivity of RNase H1. The C-terminus catalytic domain (residues 136-286) contains active sites D145, E186, D210, D274 (shown in white)

Substrate Recognition by Hs RNase H1C

Crystal structure of human RNase H1 catalytic domain in complex with r(AGUGCGACACCUGAUUCC)-d(GGAATCAGGTGTCGCACT) (PDB Id 2QK9) (Figure 7) indicates that the catalytic domain interacts with 11 base pair of the 18mer RNA-DNA hybrid (Nowotny et al., 2007). Majority of the interactions occur along the minor groove of the hybrid. The active site is in proximity with the RNA strand of the hybrid, as the RNA strand is to be get cleaved. RNA strand of the hybrid is recognized by the protein through extensive contacts with 2'-OH groups while the protein recognizes DNA strand by the ability of DNA to adopt a B form conformation. RNase H has no specific activity in cleavage of RNA-DNA hybrids as the active site of the enzyme is formed during initial folding. The enzyme interacts with seven nucleotides of the DNA strand and the first DNA binding pocket is formed by R179, T181, and N240 while second by W221, W225, and S233. The interaction between the DNA strand of the hybrid and the enzyme is driven by van der Waals and hydrogen bonding interaction. It's worth mentioning that backbone atom of the DNA is involved in interaction with the enzyme. RNase H containing four acidic residues D145, E186, D210 and D274 hydrolyzes the phosphate backbone of RNA strand in RNA-DNA hybrids with the help of divalent metal ions, Mg^{2+} or Mn^{2+}, for catalysis (Nowotny et al., 2007). Mostly, two magnesium ions, termed A and B, are observed. Mg^{2+} at A site activates the attacking nucleophile, while the B site Mg2+ destabilizes the substrate, and both A and B together stabilize the transition state and facilitate product release.
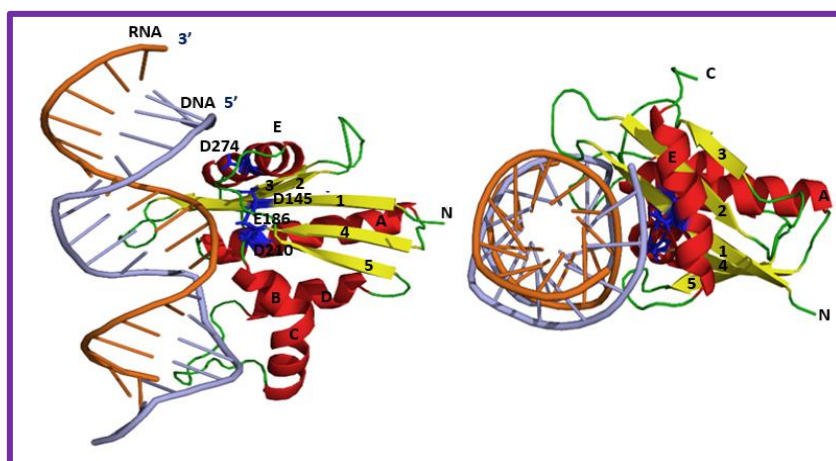

Figure 7. Crystal Structures of Hs RNase H1C interacting with 18-mer RNA-DNA hybrid (PDB ID 2QK9). (Left) View perpendicular to the helical axis of the hybrid: RNase H1 is colored according to the secondary structure elements: red (α helices), yellow (ß sheet) and green (loop). The ß sheets are numbered from 1-5 and the α helices are labeled from A-E. The substrate is shown in orange (RNA) and light blue (DNA) color. The active site residues (D145, E186, D210 and D274) are shown in blue ball-and-stick model. **(Right)** View along the helical axis of the hybrid.

1.7 Scope of the study

Understanding RNA, DNA and protein dynamics would facilitate in increasing our knowledge of nucleic acids and protein structure and their interactions with each other. Studying RNA-DNA hybrid characteristics and interactions with proteins is crucial to develop the knowledge. RNA-DNA hybrids that consist of trinucleotide repeat expansions lead to repeat instability. However, structural information about RNA-DNA hybrid containing trinucleotide repeat expansion (TRE) and their complexes with RNase H1 are very limited. Here, we aim to understand the structure and dynamics of various TRE containing RNA-DNA duplexes and their mode of binding with RNase H1 using molecular dynamics simulation technique. Complex model of TRE containing RNA-DNA hybrid and RNase H1 will further shed light on understanding the role of TREs in replication of lagging strand, antisense strategy of TREDs and instability caused by transcriptional R-loop with TREs containing RNA-DNA hybrid as well as to target RNA-DNA hybrid in the R-loop. Altogether, this study would facilitate in understanding the mechanism, pathogenesis and drug discovery of TREDs.

1.8 REFERENCES

AGUILERA, A. & GARCIA-MUSE, T. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell,* 46**,** 115-24.

CAMPS, M. & LOEB, L. A. 2005. Critical role of R-loops in processing replication blocks. *Front Biosci,* 10**,** 689-98.

CERRITELLI, S. M. & CROUCH, R. J. 1995. The non-RNase H domain of Saccharomyces cerevisiae RNase H1 binds double-stranded RNA: magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA,* 1**,** 246-59.

CLAYTON, D. A. 2000. Transcription and replication of mitochondrial DNA. *Hum Reprod,* 15 Suppl 2**,** 11-7.

CUMMINGS, C. J. & ZOGHBI, H. Y. 2000. Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet,* 1**,** 281-328.

DROLET, M. 2006. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol Microbiol,* 59**,** 723-30.

EVERETT, C. M. & WOOD, N. W. 2004. Trinucleotide repeats and neurodegenerative disease. *Brain,* 127**,** 2385-405.

GAIDAMAKOV, S. A., GORSHKOVA, II, SCHUCK, P., STEINBACH, P. J., YAMADA, H., CROUCH, R. J. & CERRITELLI, S. M. 2005. Eukaryotic RNases H1 act processively by interactions through the duplex RNA-binding domain. *Nucleic Acids Res,* 33**,** 2166-75.

GAN, W., GUAN, Z., LIU, J., GUI, T., SHEN, K., MANLEY, J. L. & LI, X. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev,* 25**,** 2041-56.

GNATT, A. L., CRAMER, P., FU, J., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science,* 292**,** 1876-82.

GOWRISHANKAR, J., LEELA, J. K. & ANUPAMA, K. 2013. R-loops in bacterial transcription: Their causes and consequences. *Transcription,* 4.

GROH, M., LUFINO, M. M., WADE-MARTINS, R. & GROMAK, N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet,* 10**,** e1004318.

HANTZ, E., LARUE, V., LADAM, P., LE MOYEC, L., GOUYETTE, C. & HUYNH DINH, T. 2001. Solution conformation of an RNA--DNA hybrid duplex containing a pyrimidine RNA strand and a purine DNA strand. *Int J Biol Macromol,* 28**,** 273-84.

HUERTAS, P. & AGUILERA, A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell,* 12**,** 711-21.

ITAYA, M. 1990. Isolation and characterization of a second RNase H (RNase HII) of Escherichia coli K-12 encoded by the rnhB gene. *Proc Natl Acad Sci U S A,* 87**,** 8587-91.

KETTENBERGER, H., ARMACHE, K. J. & CRAMER, P. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell,* 16**,** 955-65.

KOGOMA, T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev,* 61**,** 212-38.

LEE, J. E., BENNETT, C. F. & COOPER, T. A. 2012. RNase H-mediated degradation of toxic RNA in myotonic dystrophy type 1. *Proc Natl Acad Sci U S A,* 109**,** 4221-6.

LI, X. & MANLEY, J. L. 2006. Cotranscriptional processes and their influence on genome stability. *Genes Dev,* 20**,** 1838-47.

LIN, Y., DENT, S. Y., WILSON, J. H., WELLS, R. D. & NAPIERALA, M. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A,* 107**,** 692-7.

LIN, Y. & WILSON, J. H. 2007. Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol Cell Biol,* 27**,** 6209-17.

MCIVOR, E. I., POLAK, U. & NAPIERALA, M. 2010. New insights into repeat instability: role of RNA*DNA hybrids. *RNA Biol,* 7**,** 551-8.

NOWOTNY, M., GAIDAMAKOV, S. A., GHIRLANDO, R., CERRITELLI, S. M., CROUCH, R. J. & YANG, W. 2007. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell,* 28**,** 264-76.

O'CONNOR, T. P. & CRYSTAL, R. G. 2006. Genetic medicines: treatment strategies for hereditary disorders. *Nat Rev Genet,* 7**,** 261-76.

ORR, H. T. & ZOGHBI, H. Y. 2007. Trinucleotide repeat disorders. *Annu Rev Neurosci,* 30**,** 575-621.

OSTERGAARD, M. E., SOUTHWELL, A. L., KORDASIEWICZ, H., WATT, A. T., SKOTTE, N. H., DOTY, C. N., VAID, K., VILLANUEVA, E. B., SWAYZE, E. E., FRANK BENNETT, C., HAYDEN, M. R. & SETH, P. P. 2013. Rational design of antisense oligonucleotides targeting single nucleotide polymorphisms for potent and allele selective suppression of mutant Huntingtin in the CNS. *Nucleic Acids Res,* 41**,** 9634-50.

PEARSON, C. E., NICHOL EDAMURA, K. & CLEARY, J. D. 2005. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet,* 6**,** 729-42.

REDDY, K., TAM, M., BOWATER, R. P., BARBER, M., TOMLINSON, M., NICHOL EDAMURA, K., WANG, Y. H. & PEARSON, C. E. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res,* 39**,** 1749-62.

ROBERTS, R. W. & CROTHERS, D. M. 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science,* 258**,** 1463-6.

ROY, D. & LIEBER, M. R. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol,* 29**,** 3124-33.

ROY, D., YU, K. & LIEBER, M. R. 2008. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol,* 28**,** 50-60.

SANGHANI, S. R. & LAVERY, R. 1994. Theoretical studies of DNA-RNA hybrid conformations. *Nucleic Acids Res,* 22**,** 1444-9.

SHAW, N. N. & ARYA, D. P. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie,* 90**,** 1026-39.

SUGIMOTO, N., NAKANO, S., KATOH, M., MATSUMURA, A., NAKAMUTA, H., OHMICHI, T., YONEYAMA, M. & SASAKI, M. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry,* 34**,** 11211-6.

VOINEAGU, I., FREUDENREICH, C. H. & MIRKIN, S. M. 2009. Checkpoint responses to unusual structures formed by DNA repeats. *Mol Carcinog,* 48**,** 309-18.

WESTOVER, K. D., BUSHNELL, D. A. & KORNBERG, R. D. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science,* 303**,** 1014-6.

WONGSURAWAT, T., JENJAROENPUN, P., KWOH, C. K. & KUZNETSOV, V. 2012. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res,* 40**,** e16.

XIONG, Y. & SUNDARALINGAM, M. 1998. Crystal structure and conformation of a DNA-RNA hybrid duplex with a polypurine RNA strand: d(TTCTTBr5CTTC)-r(GAAGAAGAA). *Structure,* 6**,** 1493-501.

YU, K., ROY, D., HUANG, F. T. & LIEBER, M. R. 2006. Detection and structural analysis of R-loops. *Methods Enzymol,* 409**,** 316-29.

Chapter 2

Experimental Methods

The aim of the current investigation is to pinpoint the role of the RNA-DNA hybrid & its complex with RNase H1, in instigating the trinucleotide repeat overexpansion at the structural level. In the absence any such investigation, we choose MD simulation to address this issue as it is an appropriate tool to pinpoint the sequence dependent structural features (Helical parameters, angle of bending, base pair parameters, base step parameters, torsion angles and sugar pucker) of a RNA-DNA hybrid in the context of various trinucleotide repeat expansions. Further as the interaction between a nucleic acid and a protein molecule warrants huge conformational change such as bending in the nucleic acid, the current molecular dynamics (MD) simulations will be useful in understanding the conformational changes that happens in the both protein and RNA-DNA hybrid.

2.1 Structural characterization of RNA-DNA hybrid

Although several modeling studies have been carried out on RNA-DNA hybrids, sofar, no systematic investigation has been performed to characterize their structural features such as base-step, base-pair, helical and sugar-phosphate conformational angles. However, to model the RNA-DNA hybrid that consist of the trinucleotide repeat expansions to understand the enhanced stability, it is necessary to characterize their structural features. For that, we have considered 6 base-pair parameters: 3 rotational (propeller, buckle & opening) & 3 translational (stagger, stretch & shear) to define the relative orientation of bases in a base pair (Figure 2.1 Top left). Similarly, 3 rotational (twist, roll & tilt) & 3 translational (slide, shift & rise) (Figure 2.1 Top right) base step parameters along the X, Y & Z direction are considered to define the relative orientation of the base-pairs. Helical parameters with 2 rotational (tip & inclination) & 2 translational (X & Y displacements) bases & base pair orientation w.r.t each other &/or the global helical axis. In addition, the sugar-phosphate backbone torsion angles (Figure 2.2) are also considered. In this context, all the above parameters have been investigated for RNA-DNA hybrid deposited in the protein data bank using 3DNA program ((Lu and Olson, 2003).
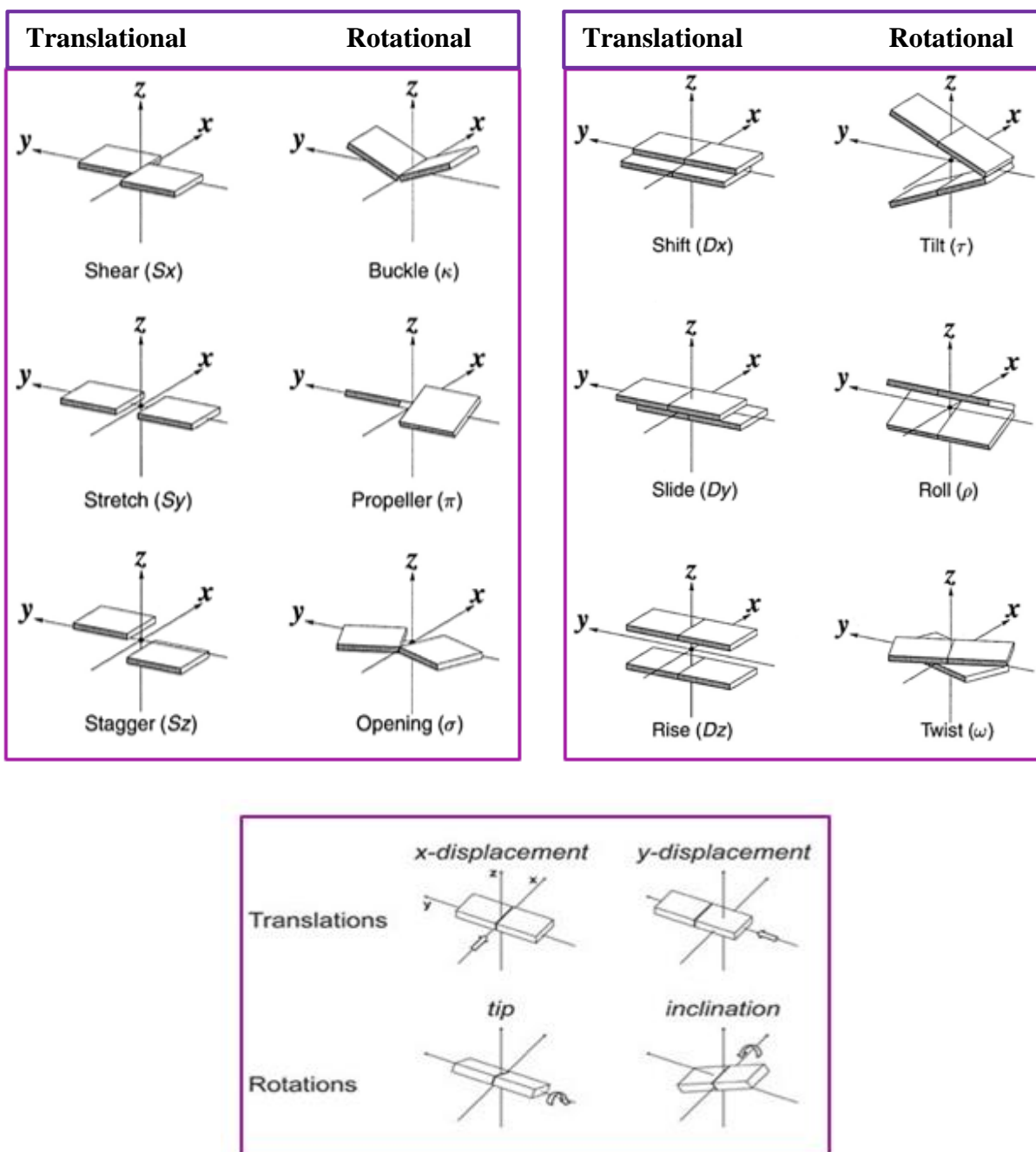
Translational	Rotational	Translational	Rotational
Shear (Sx)	Buckle (κ)	Shift (Dx)	Tilt (τ)
Stretch (Sy)	Propeller (π)	Slide (Dy)	Roll (ρ)
Stagger (Sz)	Opening (σ)	Rise (Dz)	Twist (ω)

Translations: x-displacement, y-displacement
Rotations: tip, inclination

Figure 2.1 Schematic illustration of (a) base pair (b) base-step and (C) helical parameters. Base pair parameters shear, buckle, stretch, propeller, stagger and opening are defined by the orientation of one base with respect to the other in terms of local X, Y & Z axis (Top left). Base-step parameters tilt, roll, twist, shift, slide and rise are defined by the relative orientation of the 2 adjacent base-pairs in terms of local X, Y & Z axis (Top right). Helical parameters X, Y-displacements, tip & inclination w.r.t the helical axis is also shown (Bottom).
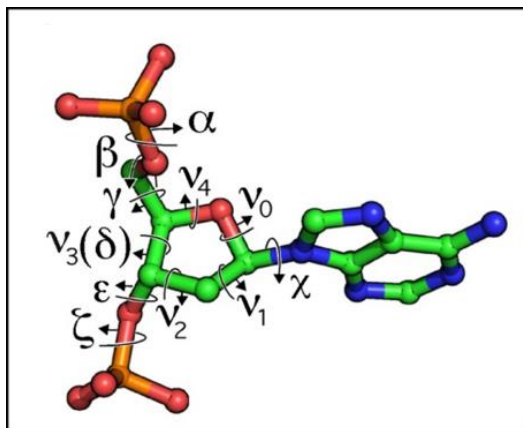
Figure 2.2 Schematic representation of backbone torsions. Six sugar-phosphate back bone torsion angles (α, β, γ, δ, ε, and ζ), five endocyclic torsion angles & the glycosydic torsion (φ) define the relative orientation of the nucleotide w.r.t the sugar are sugar ring (v_0 to v_4) is depicted

2.2 Modelling of RNA-DNA hybrid & its complex with Human RNase H1

18mer RNA-DNA hybrids are modeled using in-house programs and web-3DNA server (http://w3dna.rutgers.edu/) with ideal AB-intermediate form by using appropriate, helical twist, rise & X-displacement derived from the data base analysis. Thus, it's noteworthy that other parameters are not considered as they are negligible. Xplor-NIH was used to refine the model and Pymol is used for visualization. Models were restrained by planarity, dihedral angle and hydrogen bond distance restraints & proper internal geometry is restored & steric hindrance was removed by verlet algorithm. The Human HS RNase H1 complexed with RNA-DNA hybrid was constructed using the recent crystal structure (Nowotny et al., 2005) (Protein Data Bank entry 2QK9). Docking of HS RNase H1 and various RNA-DNA hybrids containing TNRs were done manually using Pymol.

2.3 Molecular dynamics (MD) simulation

Computer simulations provide the behaviour of a molecular system behavior w.r.t the time and act as a bridge between microscopic length and time scales and the macroscopic world of the laboratory. One can not only make direct comparisons with experimental measurements but also can simulate the system at extreme conditions like high temperature and pressure which may not be possible at laboratory. Detailed information on the fluctuations and conformational changes of proteins and nucleic acids can be studied using MD simulations. These MD

simulations are used to investigate the structural properties, dynamics and thermodynamics of biological molecules and their complexes.

Alder and Wainwright in the late 1950's first introduced molecular dynamics method to study the interactions of hard spheres. Rahman in 1964 carried out the first simulation using a realistic potential for liquid argon (Rahman, 1964). In 1974, Rahman and Stillinger did the first molecular dynamics simulation of a realistic liquid water system (Stillinger and Rahman, 1974). The simulation of the bovine pancreatic trypsin inhibitor (BPTI) was the first protein simulations (McCammon, et al, 1977). From then molecular dynamics simulations are being carried out for solvated proteins, protein-DNA complexes, nucleic acids and lipid systems to address a variety of biological phenomenon. Molecular dynamics simulation techniques are also widely used in experimental procedures NMR structure determination.

TRE containing RNA-DNA hybrids and complexes generated previously are subjected to molecular dynamics (MD) simulations. MD can capture the conformational dynamics of a molecule at the atomistic level using Newton's second law of motion. (Hansson et al., 2002).

According to Newton's second law of motion,

$$F_i = m_i \, a_i \qquad \longrightarrow \quad 1$$

where F_i= force exerted on the particle ' i ', m_i is mass and a_i is acceleration of the particles. Force can be further related to the potential energy of the system as below:

$$F_i = -\nabla V_i \qquad \longrightarrow \quad 2$$

where, $-\nabla V_i$ is gradient of the potential energy.

A set of mathematical functions and parameters which describe the potential energy of the system, the so called 'force field' takes care of ideal bond length, bond angle and torsion angle geometries along with non-bonded Vander Waals and electrostatic interactions during the simulation. With appropriate force fields, MD simulation technique can provide useful insights into the structural dynamics of DNA, RNA and protein molecules, as well as their molecular complexes. Several force fields have been developed such as AMBER, GROMOS and CHARMM. In the current MD simulation, AMBER12 (Assisted Model Building with Energy Refinement) suit is used. "AMBER" is the collective name for a set of programs that allows users to carry out and analyze molecular dynamics simulations, particularly for proteins, nucleic acids and carbohydrates. Here, we used the AMBER99 force field. AMBER99 has

been refined from its original ff94 variant. The potential energy function in AMBER99 force field (Cornell, 1995) is defined as:

$$V = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \frac{1}{2} \sum_{dihedrals} V_n (1 + \cos[n\varphi - \gamma]) + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}}\right]$$

The terms in the above equations are bond length (term 1), bond angle (term 2), dihedral angles (term 3) and non-bonded interactions vander Waals and electrostatic interactions (term 4). r_0 and θ_0 are ideal bond length and bond angles. Force constant for bond stretching and bond bending are represented as K_r and K_θ are. φ is the dihedral angle, where V_n is the rotational barrier height, n is the periodicity of rotation and γ is the phase. A_{ij} is the coefficient of the repulsive term, B_{ij} is the coefficient of the attractive term. q_i and q_j are partial charges on the atoms i and j respectively, separated by the distance R_{ij}. ε is the dielectric constant.

2.3.1 Preparation

Antechamber (assembles force fields for residues or organic molecules that are not part of the standard libraries) and *LEaP* (constructs biopolymers from the component residues, solvates the system, and prepares lists of force field terms and their associated parameters) are the main preparation program (Krepl et al., 2012). In the present MD simulations, LEaP is being used. The result of this preparation phase is contained in two text files: a coordinate *prmcrd* file containing the Cartesian coordinates of all atoms in the system, and a parameter-topology *prmtop* file that contains all other information needed to compute energies and forces; this includes atom names and masses, force field parameters, lists of bonds, angles, and dihedrals (Case et al., 2005).

Most users use tLEaP (LEaP in a text mode) to get the system from a PDB format file that contains acceptable starting coordinates (Case et al., 2005, Krepl et al., 2012). The appropriate forcefield is chosen using tleap. For, instance in the current simulation, ff99SB was used. Neutralization of the system is done by adding Na+ counter ions. The final step in preparatory phase is to create solvated model with explicit water. This is done with the command "solvatebox". For the present models, we used a 10 Å (non-bond cutoffs) of TIP3P water around the model in X, Y & Z directions to circumvent the problem of surface effect (molecule on the surface will experience a quite different force from molecules in the bulk) Periodic Boundary Conditions (PBC) is used. In this method, a box is created to enclose the particles to be simulated and is then imitated in all three directions to give a periodic array. A two-dimensional representation of PBC is shown in Figure 2.3.
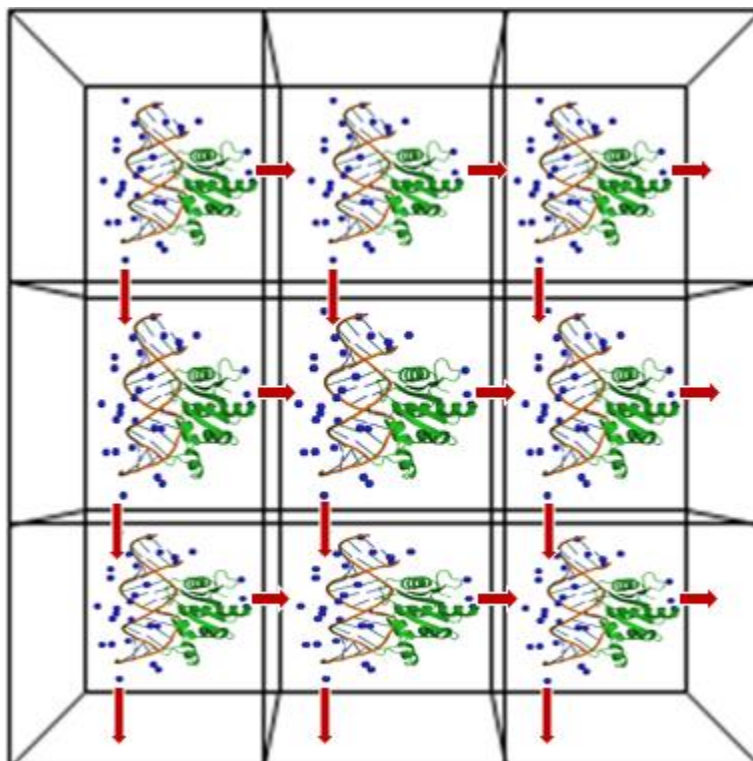
Figure 2.3. Representation of the principle of the periodic boundary conditions used in MD simulations. A particle which goes out from the simulation box by one side is reintroduced in the box by the opposite side.

Total number of particles in the central box in the figure remains constant by re-entry of particles that leave one side of the box from the opposite side as their image.

Equilibration (minimization) simulation is carried out to relax solvent molecules around the solute. This is a way to improve our *in silico* simulations to mimic the real situation.

2.3.2 Simulation protocol

Equilibration run is carried out, during which, the solvent molecule is relaxed around the solute. This is to "fix up" the positions of the atoms in order to remove any bad contacts that may lead to unstable molecular dynamics. This will take us towards the closest local minima. Minimization with *sander* will only ever take us to the nearest minima, it cannot cross transition states to reach lower minima. This will be done in several steps. During the first round the solute will be held with the force constant of 500 kcal/mol $Å^2$ and relaxing water molecules around the solute. In the subsequent runs, the force constant for holding the solute is slowly reduced to 300, 100, 50, 20, 10 and 5 kcal/mol $Å^2$. Any steric hindrance in the molecule will be removed during the equilibration run. Equilibration will be carried out using NVE condition

(Microcanonical ensemble). Production run will be carried out following the equilibration run. Production run will be carried out using NPT condition i.e. isobaric (constant pressure) of 1 atm and isothermal (constant temperature) of (300 K) with 1fs time step. The simulation is carried out to 100ns.
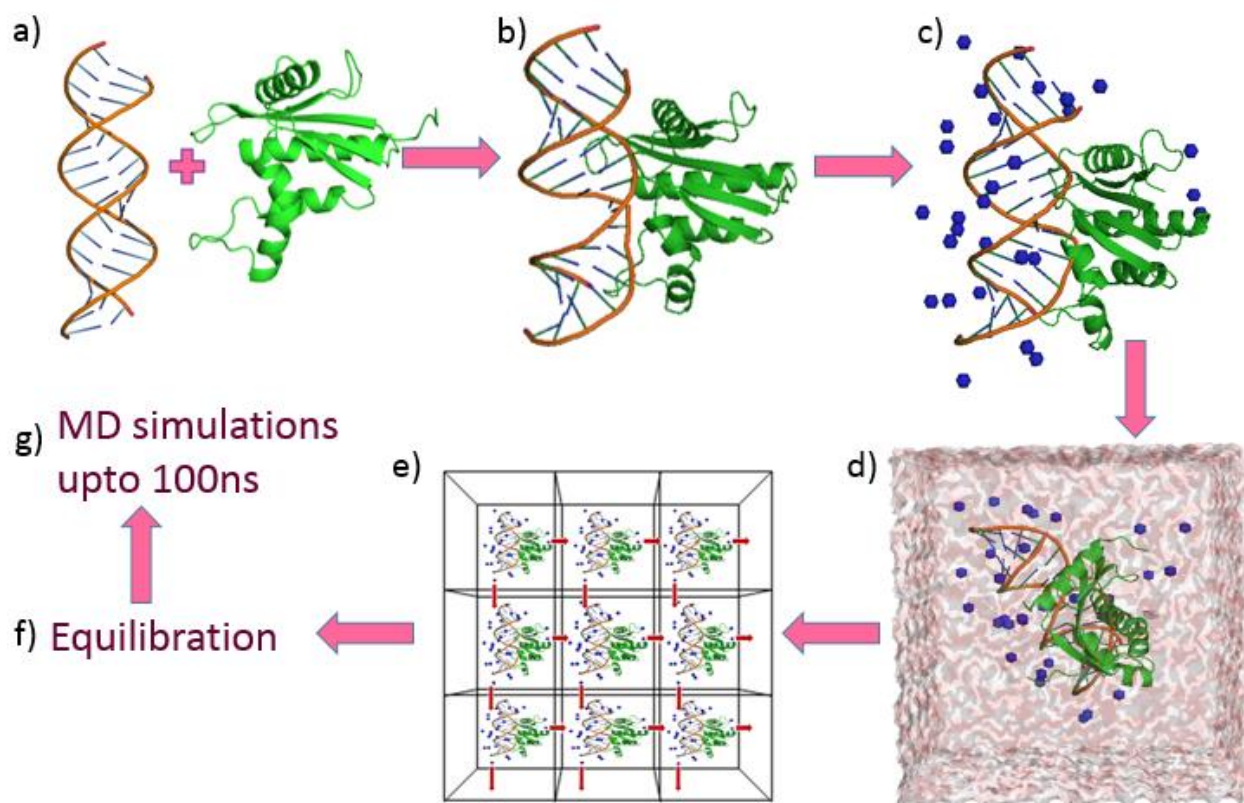
Figure 5. Outline of the molecular dynamic simulations of the RNA-DNA hybrid & RNase H1 complex. a) Modelling and docking, b) Initial model c) Neutralizing the system addition of Na+ counter ions, d) Addition of TIP3P pre-equilibrated water box, e) by the maintenance of periodic boundary conditions, f) Equilibration simulation is carried out to relax solvent molecules around the solute and g) production run simulations is carried out.

2.3.3 Analysis

Once we have equilibrated the system it is essential to check if equilibrium has been successfully achieved or not before continuing the production run simulations. Properties like potential, kinetic and total energies, the temperature, pressure, volume and density are extracted from the output files and stability of the simulations are verified. Subsequently, production run

simulation is continued similar to the equilibrium run simulation upon verified by checking the temperature & pressure.

Root-mean-square deviation (RMSD) of the backbone atoms has to be calculated to see what is happening to the structure. It is comparison of one structure to another might be initial structure or experimental structure. If the RMSd remains more or less constant at the end of 100ns of simulation, it indicates that our starting structure changed slowly into a more stable structure.

3DNA program was used for analyzing the base-step, base-pair & helical parameters.

2.4 REFERENCES

CASE, D. A., CHEATHAM, T. E., 3RD, DARDEN, T., GOHLKE, H., LUO, R., MERZ, K. M., JR., ONUFRIEV, A., SIMMERLING, C., WANG, B. & WOODS, R. J. 2005. The Amber biomolecular simulation programs. *J Comput Chem,* 26**,** 1668-88.

HANSSON, T., OOSTENBRINK, C. & VAN GUNSTEREN, W. 2002. Molecular dynamics simulations. *Curr Opin Struct Biol,* 12**,** 190-6.

KREPL, M., ZGARBOVA, M., STADLBAUER, P., OTYEPKA, M., BANAS, P., KOCA, J., CHEATHAM, T. E., 3RD, JURECKA, P. & SPONER, J. 2012. Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J Chem Theory Comput,* 8**,** 2506-2520.

LU, X. J. & OLSON, W. K. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res,* 31**,** 5108-21.

NOWOTNY, M., GAIDAMAKOV, S. A., CROUCH, R. J. & YANG, W. 2005. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell,* 121**,** 1005-16.

Chapter 3

Structural characterization of RNA-DNA hybrid

3.1 Introduction

Earlier X-ray fiber diffraction studies revealed that RNA-DNA hybrids have conformation similar to RNA (Milman et al., 1967). For instance, the structures have 11 base pairs per helical turn, highly tilted with C3'-*endo* sugar puckering. However, later fiber diffraction studies proposed C2'-*endo* sugar puckering for the DNA strand and a C3'-*endo* conformation for the RNA strand (Arnott and Selsing, 1974, Arnott et al., 1986, Zimmerman and Pheiffer, 1981). Depending upon the relative humidity, fibers of the homopolymeric hybrid poly(dT)-poly(rA) were shown to adopt two distinct conformations (Zimmerman and Pheiffer, 1981): while C3'-*endo* sugar pucker being favored at low relative humidity, the C2'-*endo* sugars within the DNA strand and C3'-*endo* sugars within the RNA strand at the high relative humidity. The results obtained from CD spectra (Kypr et al., 2009, Gray and Ratliff, 1975) also supports the above puckering preference. NMR studies revealed that, while the RNA strand retained the C3'-*endo* with DNA strands favoring either C3'-*endo* or C2'-*endo* conformation and supported that the DNA sugars may adopt a global O4'-*endo* conformation (Fedoroff et al., 1993, Chou et al., 1989). From the above study, it can be concluded that the structures of RNA-DNA hybrids are polymorphous and their conformation may depend on various parameters, such as solvent conditions and base compositions. However, it has been proposed through modeling that O4'-*endo* sugar pucker of DNA is important for recognition by RNase H1 (Priyakumar and Mackerell, 2008). Till date, no systematic investigations has been done to characterize the structural features of RNA-DNA hybrids. Such information would be very much useful to understand the enhanced stability of RNA-DNA hybrids observed in the overexpansion of trinucleotide repeats (Wongsurawat et al., 2012, Reddy et al., 2011) and its association with pathogenesis of the diseases. This warrants more systemic investigation on the structural characteristics of RNA-DNA hybrids.

To characterize the RNA-DNA hybrid structure, all the known structures of RNA-DNA hybrids from the protein database (PDB) are downloaded and structural characteristics like helical

parameters, base-pair parameters, base-step parameters, torsion angles and sugar pucker (Figure 2.1) are analyzed.

3.2 Results

Structural characterization of RNA-DNA hybrid has been carried out on 31 available hybrid structures (14 uncomplexed hybrids and 17 hybrids complexed with various enzymes) based on features like base-pair, base-step & helical parameters and sugar-phosphate backbone conformational angles. Survey performed on 31 available RNA-DNA hybrid structures confirms that RNA-DNA hybrids have intermediate conformation between A and B-forms. The helical twist (31°-35°) and the X-displacement (-3.2Å to -4.1Å) of the hybrid mimic the A-form, while the helical rise (3.4Å) resembles the B-form. Sugar puckering of RNA is C3'-*endo* and DNA is favored in south type i.e. C1'-*exo* and C2'-*endo*. Detailed summary of analysis is given in Table 3.1 and 3.2 of hybrid and complex structures respectively and comparison of RNA-DNA hybrid with A-RNA & B-DNA is given in Table 3.3.

Table 3.1 Structural characterization of uncomplexed RNA-DNA hybrids

Note that the table reports the average values over the entire survey along with standard deviation. Sol. NMR- solution NMR.

PDB ID	LENGTH OF SEQUENCE	EXP. METHOD	HELICAL TWIST (º)	X – DISP (Å)	ROLL (º)	TILT (º)	RISE (Å)	SUGAR PUCKER DNA	SUGAR PUCKER RNA
124D	8mer	Sol NMR	33.5(3.3)	-3.7(1.06)	7.08(6)	-0.1(0.3)	3(0.3)	50 % C1' exo/ 50 % O4' endo	87.5 %C3' endo/ 12.5 % C2' exo
1DRR	10mer	Sol NMR	30.23(4)	-4.8(1.4)	4.24(2.9)	-2.7(2)	3.1(0.2)	60 % C1' exo/ 40 % C3' endo	80 % C3' endo/ 20 % C2' exo
1EFO	9mer	X-Ray diffraction	31.1(4)	-4.1(1.1)	4.6(4.5)	-1.1(2.5)	3.3(0.2)	77 % C3' endo/ 23 % C4' exo	100-% C3' endo
1EFS	13mer	Sol NMR	32.7(3)	-3.4(1.3)	3.9(3.8)	-3.2(2.8)	3.8(0.2)	77 % C2' endo/ 7.6% O4' endo/ 7.6% C1' exo/ 7.6% C3' endo	100 % C3' endo
1FIX	10mer	X-Ray diffraction	33.1(3.2)	-3.8(1.5)	8(3.2)	0.6(3.2)	3.3(0.2)	70 % C3' endo/ 20% C2' endo/ 10 % O4' endo	100 % C3' endo
1G4Q	10mer	X-Ray diffraction	32.5(5.2)	-3.5(1.7)	7.3(5.2)	-1(4)	3.2(0.2)	90 % C3' endo/ 10% C1' exo	90 % C3' endo/ 10 % C2' endo
1HG9	9mer	Sol NMR	33.7(6)	-3.6(1.5)	10.3(7.7)	-2.6(2.8)	3.2(0.3)	44% O4' endo/ 33% C1' exo/ 11% C2'endo/	55 % C3' endo/ 44 % C2' exo

Continued…

PDB ID	SIZE	METHOD	HELICAL TWIST	X-DISP	ROLL	TILT	RISE	SUGAR PUCKER DNA	SUGAR PUCKER RNA
								11% C4' exo	
1JB8	10mer	X-Ray diffraction	29.7(7.3)	-3.8(1.4)	4.4(5.8)	-0.1(4.2)	3.6(1.05)	100 % C3' endo	90 % C3' endo/ 10 % C2' endo
1NXR	8mer	Sol NMR	32.1(4.2)	-4.4(1)	7.8(6.7)	-0.5(3.5)	3.2(0.1)	37.5 % C1' exo/ 37.5%O4' endo/ 12.5% C4' exo/ 12.5% C3' endo	75 % C3' endo/ 12.5% C4' exo/ 12.5% O4' endo
1PJG	10mer	X-Ray diffraction	32.5(5.4)	-3.5(1.7)	7.3(5.4)	-0.9(3.7)	3.2(0.2)	90 % C3' endo/ 10 % C2' endo	90 % C3' endo/ 10 % C2' endo
1PJO	10mer	X-Ray diffraction	32.6(5.9)	-3.6(1.6)	7.9(5.2)	-1.2(4.2)	3.2(0.2)	80 % C3' endo/ 10 % C2' endo/ 10 % O4' endo	80 % C3' endo/ 10 % C2' exo/ 10 % C2' endo
1RRD	10mer	Sol NMR	32(3.8)	-4.1(0.7)	4(2.4)	1.7(1.4)	3.1(0.1)	60 % C1'exo/ 30 % C3'endo/ 10 % O4'endo/	90 % C3' endo/ 10 % C4' exo
404D	10mer	X-Ray diffraction	30.6(3.7)	-4.2(1.3)	5.5(3.4)	0.7(1.9)	3.3(0.2)	90 % C3' endo/ 10 % C2' endo	70 % C3' endo/ 30 % C2'exo
8DRH	8mer	Sol NMR	35(2.9)	-3.2(1.3)	8.1(4.1)	-2.5(2.1)	3.1(0.2)	50 % C1'exo/ 37.5 % O4' endo/ 12.5 % C4' exo	100 % C3' endo

Table 3.2 Structural characterization of RNA-DNA hybrid in complexed with various enzymes.

Note: E.C represents elongation complex, C.D represents catalytic domain and Pol II represents polymerase II. All the models are from X-ray diffraction method.

PDB ID	DESCRIPTION (ORGANISM)	SIZE	HELICAL TWIST (º)	X – DISP (Å)	ROLL (º)	TILT (º)	RISE (Å)	SUGAR PUCKER DNA	SUGAR PUCKER RNA
1HYS	HIV1 RT (Mouse)	21mer	35.1(9.4)	-2.1(1.8)	-0.3(9.4)	1.7(5.7)	3.4(0.7)	62 % C2' endo/ 33 % C3' endo/ 5 % C1' exo	100 % C3' endo
1I6H	RNA Pol II E.C (YEAST) 1MISMATCH	9mer	29.5(7.3)	-4.7(2.3)	6.3(4.1)	0.8(5.2)	3.2(0.3)	89 % C3' endo/ 11 % C4' exo	89 % C3' endo/ 11 % C2' exo
1SFO	RNA Pol II E.C (YEAST)	7mer	29.6(7.6)	-4.8(2.2)	5.1(5.1)	-0.4(3.3)	3.4(0.6)	57 % C4'exo/ 43 % 04'endo	100 % C3' endo
1ZJL	RNase H C.D mutant D192N (B.halodurans)	12mer	30.3(6.6)	-3.7(1.1)	4.7(5.3)	3.3(3.5)	3.3(0.2)	93 % C2'endo/ 7 % C3' exo	100 % C3' endo
2G8F	RNase H C.D mutant E118A	5mer	33.3(5.2)	-3.5(1.4)	7(2.2)	3.4(5)	3.3(0.2)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2G8H	RNase H C.D mutant D192N Mg+2 (B.halodurans)	5mer	33.1(4.6)	-3.5(1.2)	6.4(1.6)	2.3(4.4)	3.4(0.3)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo

2G8I	RNase H C.D mutant D192N Mn+2 (B.halodurans)	5mer	33.3(4.9)	-3.5(1.2)	6.8(2.2)	3(5)	3.4(0.2)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2G8K	RNase H C.D mutant D192N Ca+2 (B.halodurans)	5mer	33.4(5.5)	-3.5(1.2)	6.8(2.5)	3.2(5)	3.4(0.3)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2G8U	RNase H C.D mutant D132N Mg+2 (B.halodurans)	5mer	33.6(6.8)	-3.6(2)	7.5(4)	2.7(4)	3.3(0.3)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2G8V	RNase H C.D mutant E188A Mg+2 (B.halodurans)	5mer	32.2(4)	-3.7(0.9)	7.3(3.4)	3.2(5)	3.4(0.2)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2G8W	RNase H C.D mutant E188A Ca+2 (B.halodurans)	5mer	30.6(5.2)	-3.7(1.2)	5.6(5.4)	2.2(2.4)	3.2(0.2)	80 % C2' endo/ 20 % C1' exo	100 % C3' endo
2QK9	Human RNase H C.D mutant D210N	18mer	32.5(4.3)	-3.8(1.6)	6.1(4)	1.2(2.5)	3.3(0.2)	50 % C3'endo/ 22 % C2'endo/ 11 % C2'exo/ 5.5 % C3' exo / 5.5 % O4' endo/ 5.5 % C4' exo	94.5 % C3' endo/ 5.5 % C2' exo
2QKB	Human RNase H C.D mutant D210N	18mer	32(4.5)	-3.6(1.6)	6.2(5.5)	2.5(4.4)	3.3(0.2)	27.8 % C1'exo/ 27.8 % C3'endo/ 16.6 % C2'endo/ 16.6 % C4'exo/ 5.5 % O4' endo/ 5.5 % C2' exo/ 5.5 % C3' exo	89 % C3' endo/ 11 % C2'exo
2QKK	Human RNase H C.D mutant D210N	18mer	33(30.6)	-3.4(4.5)	3.6(21)	2(19.9)	3.3(0.6)	64.3 % C1'exo/ 28.6 % O4'endo/ 7.14 % C2'endo/ 7.14 % C4' endo	85.7 % C3'endo/ 14.3 % C2'exo
3GTQ	RNA Pol II complex	9mer	31(8.2)	-3.6(4.1)	5.1(12)	-0.2(8)	3.4(0.3)	56 % C3'endo/ 11 % C1'exo/ 11%C2'endo/ 11 % O 4'endo/ 11 % C4'exo	78 % C3' endo/ 22 % C2'exo
3S14	RNA Pol II complex	6mer	32.5(4.2)	-3.7(1.5)	6.3(5.5)	-0.6(1.6)	3.2(0.2)	83 % C3'endo/ 17 % C4'exo	100 % C3' endo
3S17	RNA Pol II complex	9mer	30.5(3.4)	-4(1)	6.5(2.8)	0.25(3.1)	3.3(0.2)	56 % C3'endo/ 44 % C1'exo	89 % C3' endo/ 11 % C4'exo

Conformational parameters	A-RNA	B-DNA	RNA-DNA hybrid
Type of helix	Right handed	Right handed	Right handed
Number of base pairs per turn	11	10	11
Helical twist (t)	33° (~11°)	36° (~9°)	32.7° (~2°)
Helical rise per base pair (Å)	2.9 Å	3.4 Å	3.4 Å
Base step roll	6°	-1°	7°
X- displacement	-4.4 to -4.9 Å	-0.2 to -1.8 Å	-3.0 to -3.8 Å
Base pair inclination	13-20°	-2°	12-14°
Sugar pucker	C_3' Endo	C_2' Endo	RNA- C_3' Endo DNA-C_2'Endo C_1' Exo

Table 3.3 Summary of RNA-DNA hybrid analysis and its comparison with A-RNA and B-DNA.

Note that instead of local tilt,the inclination that is calculated w.r.t global helical axis is given in the table.

Based on the analysis of RNA-DNA hybrids, following helical and base step parameters are considered to model RNA-DNA hybrids. Helical parameters used are helical twist [32.72º (0.63)] & helical rise [3.4Å (0.02)], local base step parameters used are roll [7.08º (2.8)] & X-displacement [-3.4 Å (0.14)] and C_3'endo Sugar puckering for RNA and C_1'exo for DNA are considered for modeling. Table 3.4 briefly gives the list of parameters used for modeling RNA-DNA hybrids containing various TREs (Figure 3.5)

Table 3.4 Parameters considered for hybrid modeling.	
Conformational Parameter	**Value considered**
Twist	32.7°
X-displacement	-3.44 Å
Roll	7.08
Tilt	-1.1°
Rise	3.4 Å
DNA sugar puckering	C1'-*exo*
RNA sugar puckering	C3'-*endo*

3.3 Modeling of RNA-DNA hybrids

Table 3.5 Initial models have been built using web-3DNA server (http://w3dna.rutgers.edu/) and in-house FORTRAN programs. Subsequently Xplor was used to refine the models by maintaining the AB intermediate geometry. Figure 3.1 illustrates the various steps involved in modeling of RNA-DNA hybrids.

Table 3.5 Various TREs containing RNA-DNA hybrids considered for the current investigation.		
Model No.	**Sequence**	**Disease asssociated**
1	d(GAA)$_6$-r(UUC)$_6$	Friedreich ataxia
2	r(GAA)$_6$-d(TTC)$_6$	Friedreich ataxia
3	r(CAG)$_6$-d(CTG)$_6$	Myotonic dystrophy 1, Huntington's disease, spinocerebellar ataxia (SCA 1,2,3,6,7 &17)
4	r(CUG)$_6$-d(CAG)$_6$	Myotonic dystrophy 1, Huntington's disease, Huntington's disease like 2, spinocerebellar ataxia (SCA 1,2,3,6,7 &17).
5	r(CGG)$_6$-d(CCG)$_6$	Fragile X syndrome A
6	r(CCG)$_6$-d(CGG)$_6$	Fragile X syndrome E
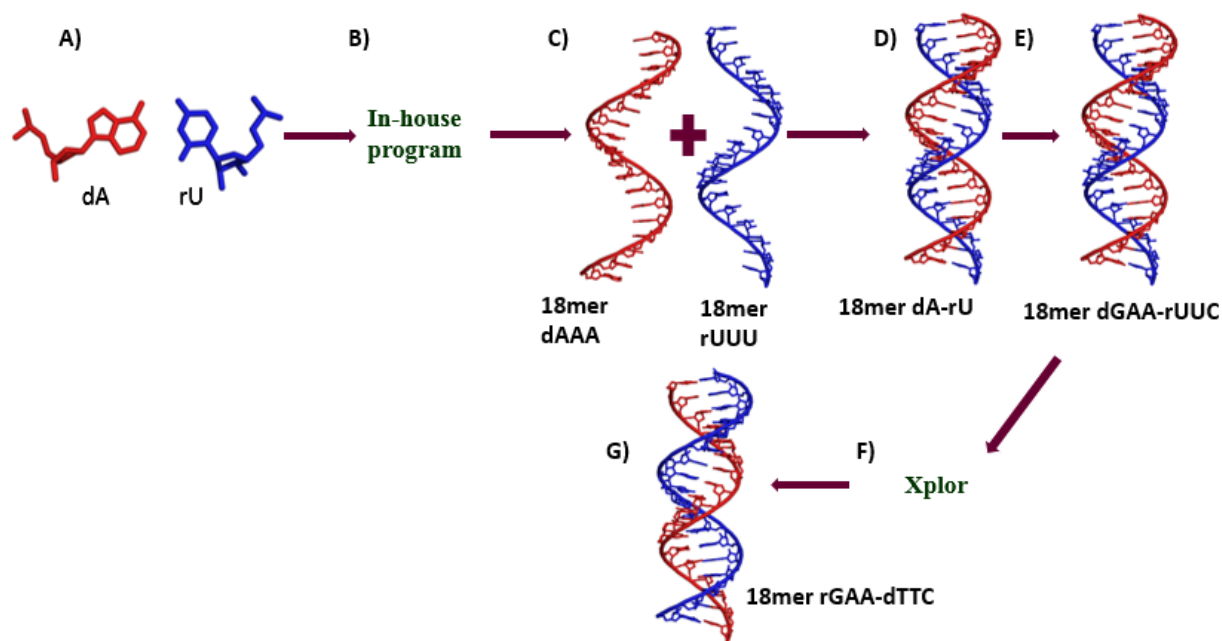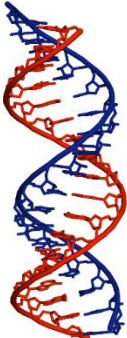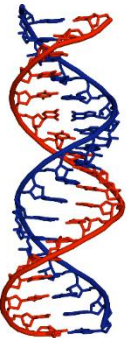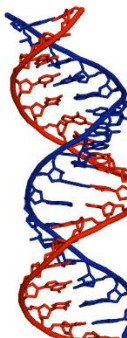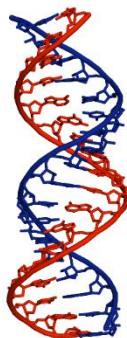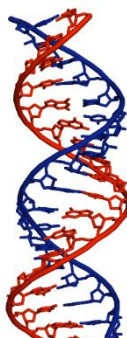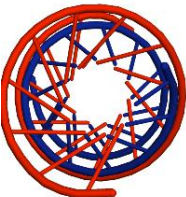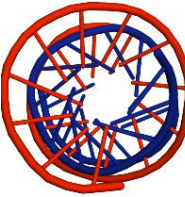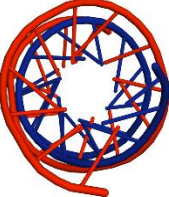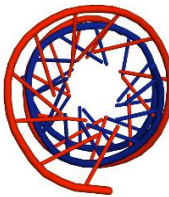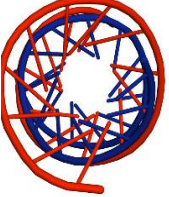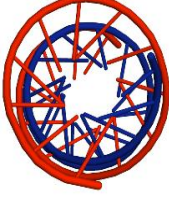
Figure 3.1 Schematic representation of various steps involved in the models of RNA-DNA hybrids. The entire procedure is explained by considering dGAA-rUUC and rGAA-dTTC as examples. A) dA-rU base-pair are constructed using 3DNA with the parameters given in Table 3.4. B & C) In-house program is used to generate 18mer d(A) & r(U). D) 18mer d(A) & r(U) are combined to generate 18mer dA-rU. E) Subsequent changes are done to 18mer dA-rU to build 18mer dGAA-rUUC. F & G) Xplor was used to restain the model and 18mer rGAA-dTTC is built by making necessary changes to 18mer dGAA-rUUC. RNA strand is depicted in blue and DNA in red.

Various TRE's containing RNA-DNA hybrids were successfully modelled (Table 3.6 top) and verified for various parameters (Table 3.6 bottom) mentioned in Table 3.5. Subsequently, the models were submitted to MD simulation.

Table 3.6 Models of various TREs containing RNA-DNA hybrids and the parameters considered.

Note that base-pair & helical parameters & the sugar puckering corresponding to each model is given in the bottom row.

	d(GAA)$_6$ - r(UUC)$_6$	r(GAA)$_6$ - d(TTC)$_6$	d(CTG)$_6$ - r(CAG)$_6$	r(CUG)$_6$ - d(CAG)$_6$	r(CCG)$_6$ - d(CGG)$_6$	d(CCG)$_6$ - r(CGG)$_6$
Initial Model						
Helical view						
X-disp	-3.4(0.14)	-3.4(0.16)	-3.4(0.14)	-2.9(1.48)	-2.88(1.5)	-2.88(1.5)
Twist	32.72(0.63)	32.73(0.72)	33.58(4.1)	33.29(1.29)	33.29(1.66)	33.29(1.66)
Rise	3.40(0.02)	3.40(0.02)	3.31(0.31)	3.46(0.44)	3.46(0.43)	3.46(0.43)
Roll	7.08(0.01)	7.08(0.01)	-5.78(3.18)	7.42(1.48)	7.44(1.47)	7.44(1.47)
Tilt	1.13(0.02)	1.12(0.03)	1.28(9.14)	1. 05(8.08)	1.02(8.06)	1.02(8.06)
Sugar (RNA)	C3'endo	C3'endo	C3'endo	C3'endo	C3'endo	C3'endo
(DNA)	C1'exo	C1'exo	C1'exo	C1'exo	C1'exo	C1'exo

3.4 Conclusion

RNA-DNA hybrids with various trinucloetide repeats are constructed successfully using helical, base-pair & base-step parameters & sugar-phosphate backbone conformational angles derived from the detailed analysis done on RNA-DNA hybrids. Thus, the built models are confined to AB intermediate geometry. These models will subsequently be subjected to MD simulation to understand the role of RNA-DNA hybrids in the pathogenesis of TREDs.

3.5 References

ARNOTT, S., CHANDRASEKARAN, R., MILLANE, R. P. & PARK, H. S. 1986. DNA-RNA hybrid secondary structures. *J Mol Biol,* 188**,** 631-40.

ARNOTT, S. & SELSING, E. 1974. Letter: The structure of polydeoxyguanylic acid with polydeoxycytidylic acid. *J Mol Biol,* 88**,** 551-2.

CHOU, S. H., FLYNN, P. & REID, B. 1989. High-resolution NMR study of a synthetic DNA-RNA hybrid dodecamer containing the consensus pribnow promoter sequence: d(CGTTATAATGCG).r(CGCAUUAUAACG). *Biochemistry,* 28**,** 2435-43.

FEDOROFF, O., SALAZAR, M. & REID, B. R. 1993. Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J Mol Biol,* 233**,** 509-23.

GRAY, D. M. & RATLIFF, R. L. 1975. Circular dichroism spectra of poly[d(AC):d(GT)], poly[r(AC):r(GU)], and hybrids poly[d(AC):r(GU)] and poly[r(AC):d(GT)] in the presence of ethanol. *Biopolymers,* 14**,** 487-98.

KYPR, J., KEJNOVSKA, I., RENCIUK, D. & VORLICKOVA, M. 2009. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res,* 37**,** 1713-25.

MILMAN, G., LANGRIDGE, R. & CHAMBERLIN, M. J. 1967. The structure of a DNA-RNA hybrid. *Proc Natl Acad Sci U S A,* 57**,** 1804-10.

PRIYAKUMAR, U. D. & MACKERELL, A. D., JR. 2008. Atomic detail investigation of the structure and dynamics of DNA.RNA hybrids: a molecular dynamics study. *J Phys Chem B,* 112**,** 1515-24.

REDDY, K., TAM, M., BOWATER, R. P., BARBER, M., TOMLINSON, M., NICHOL EDAMURA, K., WANG, Y. H. & PEARSON, C. E. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res,* 39**,** 1749-62.

WONGSURAWAT, T., JENJAROENPUN, P., KWOH, C. K. & KUZNETSOV, V. 2012. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res,* 40**,** e16.

ZIMMERMAN, S. B. & PHEIFFER, B. H. 1981. A RNA.DNA hybrid that can adopt two conformations: an x-ray diffraction study of poly(rA).poly(dT) in concentrated solution or in fibers. *Proc Natl Acad Sci U S A,* 78**,** 78-82.

Chapter 4

Structure and dynamics characterization of RNA-DNA hybrid consisting of a variety of TREs

4.1 Introduction

RNA-DNA hybrids in association with trinucleotide repeat expansions (TREs) cause genetic instability, and leads to more than 20 neurological and neuromuscular disorders which include DM1 $(CTG)_n$, SCA1 $(CAG)_n$, FRAXA $(CGG)_n$, FRAXE $(CCG)_n$ and FRDA $(GAA)_n$ (Table 3.2.1) (Pearson et al., 2005, McIvor et al., 2010, Reddy et al., 2011). The presence of these TREs in RNA-DNA hybrids enhance the stability of these RNA-DNA hybrids (Lin et al., 2010). Earlier it was demonstrate that transcribed CTG-CAG repeat sequences can exist in human cells as R loops (Lin et al., 2010) and $(CTG)_n$-$(CAG)_n$ TREs can alone cause more than a dozen human neurodegenerative diseases (Liu and Leffak, 2012), on the other hand, $(GAA)_n$ containing RNA-DNA hybrid inhibit transcription elongation by arresting RNA polymerase II (Grabczyk et al., 2007).

However, the exact mechanism behind the formation of persistent RNA-DNA hybrid with trinucleotide repeat over expansion & its role in causing the pathogenesis of the disorders is unknown. Thus, to understand the formation of persistent RNA-DNA hybrids under over expansion of trinucleotides, is important to understand their structure at the atomistic level. In this context, we carried out 100ns MD simulation of various TREs containing RNA-DNA hybrids (Table 3.5). In addition, a control simulation is also carried for sequence that is found to complex with human RNase H1 (PDB ID- 2QK9).

4.2 Results

100ns MD simulation was carried out for each RNA-DNA hybrid. Last 5ns i.e. 95 to 100ns were considered for the present study. 2 base-pairs from each terminal are removed to minimize end fraying effect. Thus, models considered for analysis consists only 14mer instead of 18mer. Time averaged structures of first 0.5ns and last 0.5ns are shown in Table 4.1
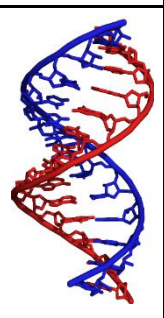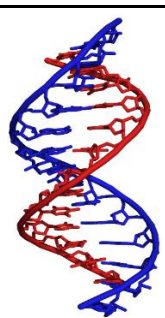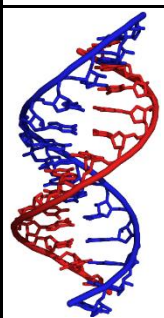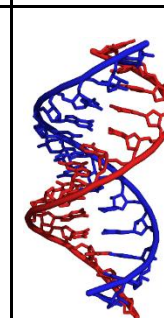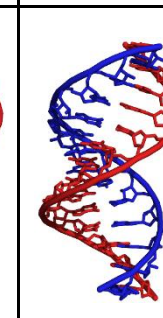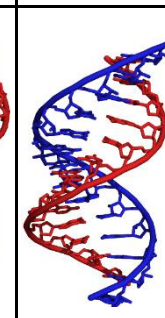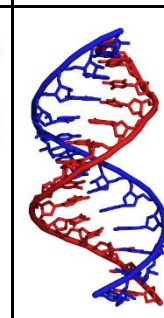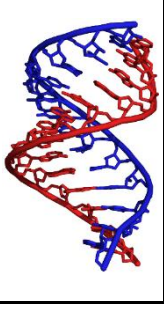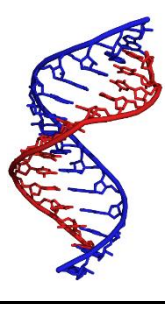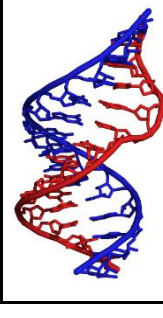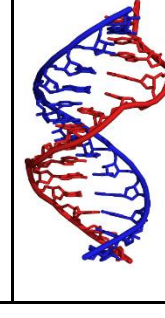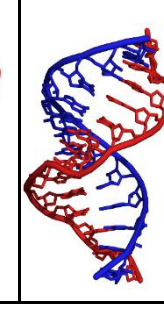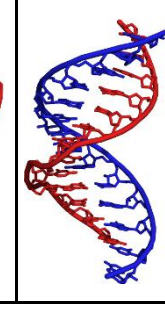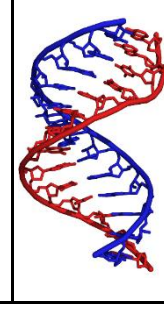
Analysis was begun by considering the overall stability of the conformation of the hybrids. The main aim is to examine the progress of the structure during the trajectory by measuring the root-

mean-square deviation (RMSD) with respect to the initial structure. RMSD calculated over the entire simulation indicates that the hybrid structure doesn't deviate much from the initial structure (Figure 4.1). In fact, the RMSD stays less than 5Å for all the hybrids (Table 4.1). This eventually reflects in the average structure calculated over the 1st & last 500ps, which is eventually very similar.

This can be clearly visualized from the ensembles of the structures shown for 3 different TREs (Figure 4.2). Thus, it is evident from the average structure, RMSD & ensemble of structures throughout the simulation that hybrids were intact over the course of the time.

Fig. 4.1 illustrates RMSD plots of one of the models considered i.e. dGAA-rUUC and the control hybrid were plotted for all other models too and time vs RMSD profile for all the hybrids show Stability similar to control hybrid.

Table 4.1 Average structures corresponding to 1st & last 500ps of the control simulation carried out for various TRE's containing RNA-DNA hybrids that complex with HsRNase H1

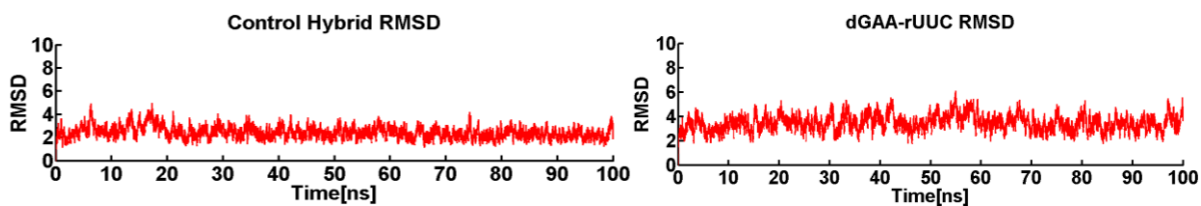|	$d(GAA)_6$-$r(UUC)_6$	$r(GAA)_6$-$d(TTC)_6$	$d(CTG)_6$-$r(CAG)_6$	$r(CUG)_6$-$d(CAG)_6$	$r(CCG)_6$-$d(CGG)_6$	$d(CCG)_6$-$r(CGG)_6$	Control simualtion
1 nano second							
100 nano seconds							

Figure 4.1 Time vs RMSD profile (A°): (Top) for control simulation & (Bottom) for dGAA-rUUC hybrid. Note that the last 2 base-pairs on either ends are not considered due to end fraying effect. Thus, only the central 14 mer is alone considered for analysis.
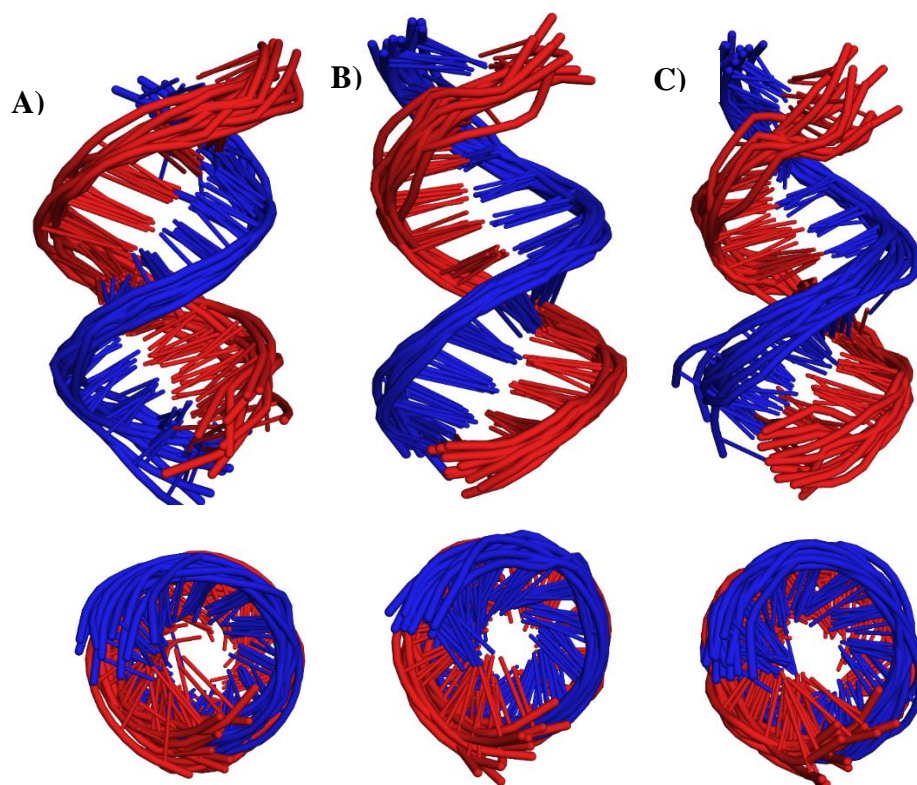

Figure 4.2 Ensemble of (A) dGAA-rUUC, (B) rGAA-dTTC and (C) rCCG-dCGG hybrids observed throughout the simulation.

Note that 11 structures at every 10ns is considered. (Top) View perpendicular to the helical axis (bottom) along the helical axis.

All the hybrids were seen to be intact over the course of the simulation. In overall physical appearance, the MD structures adapt closely to that of an A-form of RNA. Further analysis has

been carried out to validate the above observation. Different simulations require different approaches for analyzing depending on the features of interest. Here various parameters like base-pair, base-step, helical parameters, sugar-phosphate backbone torsion angles, sugar puckering and stacking interactions are investigated and are discussed below.

Sugar puckering preference for RNA-DNA hybrid

Calculation of sugar puckering over the entire simulation indicates that the RNA strand retains C_3'*endo* while DNA strand favor C_1'*exo/* C_2'*endo.* However, a small fraction of C_2'*exo* was also seen in RNA & O_4'*endo* seen for DNA. The results were in favor of NMR studies which revealed that RNA strands have C_3'-*endo* and DNA would be favored by south puckering, C_2'-*endo* (Fedoroff et al., 1993, Chou et al., 1989).

Torsional parameters

Torsional parameters include alpha, beta, gamma, delta, epsilon and zeta. Chi being measured with same backbone atoms as of delta, can be included under torsional parameters. All the Torsional parameters were found to be within the normal range. Alpha and zeta were gauche-, gamma in gauche+, beta and epsilon in trans region and chi in anti-conformation.

Base-pair, base-step and helical parameters

One of the important part of nucleic acid simulation analysis is the examination of Base-pair, base-step and helical parameters. As part of the validation, all the parameters of hybrids containing various TRE's are compared to the control hybrid without any TRE. Parameters were computed using the 3DNA program.

X-displacement, Helical Twist, Tilt, Rise, Roll are further analyzed in detail. Table 4.1 gives a brief summary of the parameters.

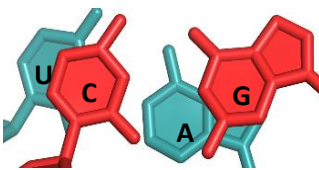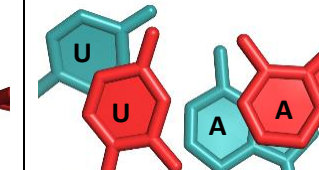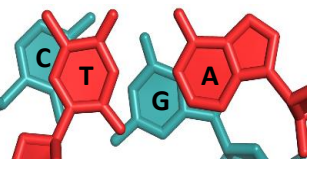|	d(GAA)$_6$. r(UUC)$_6$	r(GAA)$_6$. d(TTC)$_6$	d(CTG)$_6$. r(CAG)$_6$	r(CUG)$_6$. d(CAG)$_6$	r(CCG)$_6$. d(CGG)$_6$	d(CCG)$_6$. r(CGG)$_6$	Hybrid
Table 4.1 Summary of analysis of TRE's containing RNA-DNA hybrid and its comparison with control.							
X-disp	-4.55(1.5)	-4.42(1)	-3.37(0.9)	-3.38(1.5)	-4.29(1.1)	-3.99(1.4)	-4.49(1.4)
Twist	29.47(2.6)	28.4(3.7)	30.41(4.3)	30.17(4.6)	29.16(3.8)	29.32(4.3)	29(3.6)
Rise	3.37(0.12)	3.36(0.2)	3.33(0.2)	3.29(0.2)	3.33(0.1)	3.41(0.2)	3.32(0.2)
Roll	9.05(2.3)	4.26(3.3)	6.69(4.8)	6.69(4.8)	6.5(3.12)	2.52(3.12)	6.78(4.1)
Tilt	2.3(1.5)	-1.47(1.8)	-2.59(3)	3.14(2.2)	2.33(1.3)	-2.65(2.06)	2.81(2.7)
Groove Major	16.6	20.4	18.2	16.3	16.3	20.7	18.8
minor	14.7	14.6	15	14	14	13.1	16.8
Sugar (RNA)	C3'endo	C3'endo	C3'endo	C3'endo	C3'endo	C3'endo/ C2'exo	C3'endo
(DNA)	C1'exo/ C2'endo	C1'exo/ C2'endo	C1'exo	C1'exo/ C2'endo	C1'exo/ O4'endo	C1'exo/ O4'endo/ C2'endo	C1'exo/ O4'endo/ C2'endo
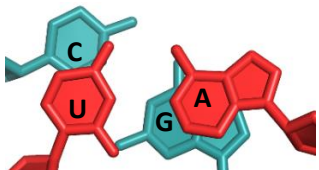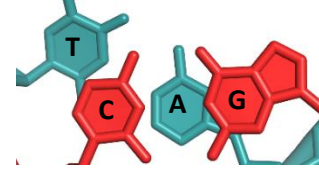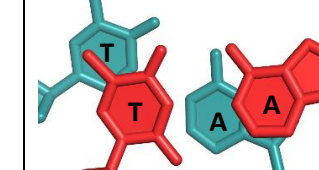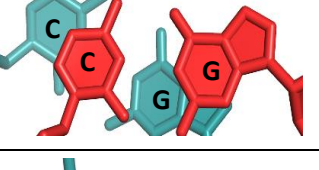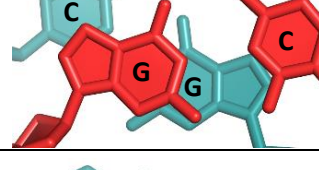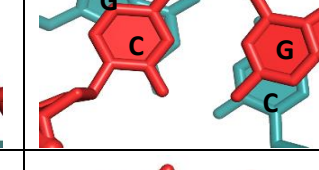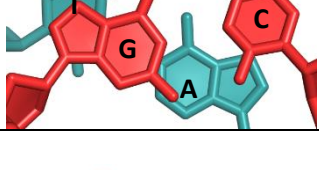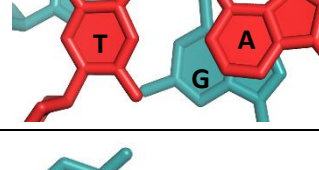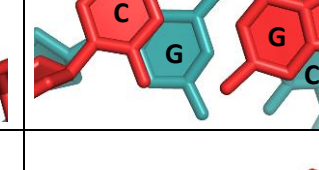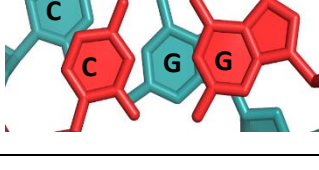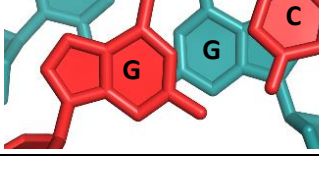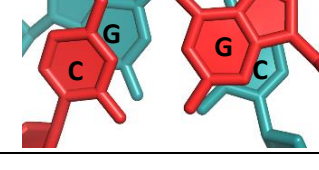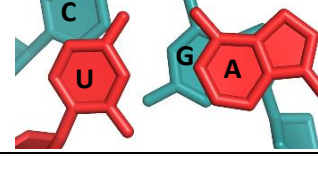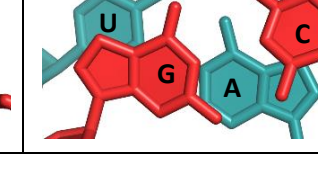
Present study revealed that RNA-DNA hybrids show characteristics intermediate to A and B forms with both A-form characteristics and B-form characteristics. X-displacement, helical twist, roll favor A-RNA while rise and groove widths favor B-DNA.

Reduced Twist values increase the compactness and residues per turn (~12) are observed in all cases including the control hybrid. This study is in support to the diffraction studies conducted by Zimmerman et.al (Zimmerman and Pheiffer, 1981). The above study reveals that characteristics of hybrid containing TRE's are very much similar to the control hybrid without TRE's.

Stacking interactions

The stability of the nucleic acid duplex depends mostly on base pairing between complementary strands i.e. hydrogen bonds between bases and base-stacking interactions between adjacent bases. Strong stacking interactions 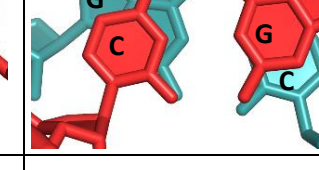seen in individual bases are key contributors to the duplex stability, as base stacking is predominant in duplexes than in single strands (Yakovchuk et al., 2006). These

interactions are hydrophobic in nature. Stacking is dependent on both base composition and base sequence (Marmur and Doty, 1962). The strength of these stacking interactions depends on the polarizability of p electron cloud of a base. Stacking interactions capability is different for different bases depending on the substitutions of the bases. Thus, we investigate the stacking interactions to offer stereochemical rationale for the enhanced stability of RNA-DNA hybrids with various TREs.

Table 4.3 Stacking interactions in various TRE's containing RNA-DNA hybrid		
dGAA-rUUC		
rGAA-dTTC		
dCCG-rCGG		
dCTG-rCAG		
rCCG-dCGG		
rCUG-dCAG		

As can be seen in Figure 4.3, there is an enhanced stability observed between the neighboring steps in all the hybrids considered for the current investigation. For instance, in the case of dGAA, enhanced stability is seen between GA, AG & AA steps as well UC steps. Similarly, in rGAA, CT, GA, AG, AA step possess good stacking interactions. In dCCG, CC, GC & CG steps exhibits good stacking interaction. Likewise, GG, GC & CG steps have better stacking. In the case of dCTG, TG, AG, CT, CG & GC step has good stacking interactions. Similar trend is also observed from rCUG, where GC, GA, UG steps show better stacking between the aromatic ring. Thus, in all the cases there is an uninterrupted stacking between the adjacent steps, especially between the aromatic rings, contributing to enhanced stability. Thus, such repetitive stacking interactions in the TRE containing RNA-DNA hybrids may offer enhanced stacking.

4.3 REFERENCES

CHOU, S. H., FLYNN, P. & REID, B. 1989. High-resolution NMR study of a synthetic DNA-RNA hybrid dodecamer containing the consensus pribnow promoter sequence: d(CGTTATAATGCG).r(CGCAUUAUAACG). *Biochemistry,* 28**,** 2435-43.

FEDOROFF, O., SALAZAR, M. & REID, B. R. 1993. Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J Mol Biol,* 233**,** 509-23.

GRABCZYK, E., MANCUSO, M. & SAMMARCO, M. C. 2007. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res,* 35**,** 5351-9.

LIN, Y., DENT, S. Y., WILSON, J. H., WELLS, R. D. & NAPIERALA, M. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A,* 107**,** 692-7.

LIU, G. & LEFFAK, M. 2012. Instability of (CTG)n*(CAG)n trinucleotide repeats and DNA synthesis. *Cell Biosci,* 2**,** 7.

MARMUR, J. & DOTY, P. 1962. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J Mol Biol,* 5**,** 109-18.

MCIVOR, E. I., POLAK, U. & NAPIERALA, M. 2010. New insights into repeat instability: role of RNA*DNA hybrids. *RNA Biol,* 7**,** 551-8.

PEARSON, C. E., NICHOL EDAMURA, K. & CLEARY, J. D. 2005. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet,* 6**,** 729-42.

REDDY, K., TAM, M., BOWATER, R. P., BARBER, M., TOMLINSON, M., NICHOL EDAMURA, K., WANG, Y. H. & PEARSON, C. E. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res,* 39**,** 1749-62.

YAKOVCHUK, P., PROTOZANOVA, E. & FRANK-KAMENETSKII, M. D. 2006. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res,* 34**,** 564-74.

ZIMMERMAN, S. B. & PHEIFFER, B. H. 1981. A RNA.DNA hybrid that can adopt two conformations: an x-ray diffraction study of poly(rA).poly(dT) in concentrated solution or in fibers. *Proc Natl Acad Sci U S A,* 78**,** 78-82.

Chapter 5

Structure & dynamics of Human RNase H1 and various TREs containing RNA-DNA hybrid complexes

5.1 Introduction

As mentioned in the chapter 1, structural insight into the complex of RNA-DNA hybrid containing various TREs and RNase H would facilitate the understanding of antisense strategy for treating TREDs at RNA level. RNase H, a ribonuclease, specifically cleaves the RNA strand of the RNA-DNA hybrid duplex without any sequence specificity to produce 5'-phosphate and 3'-hydroxy terminated products (Priyakumar and Mackerell, 2008). Here, it would be more appropriate to discuss about Hs RNase H1 as we are dealing with the abnormal expansion of trinucleotide repeats that lead to several neurological and neuromuscular disorders in humans. In context of the formation of persistent RNA-DNA hybrids by TREs during R-loop formation. It is to explore their complex with RNase H1 to understand their role in disease pathogenesis. To the end, we model the complex of various TREs containing RNA-DNA hybrids with Human RNase H1 & understand the conformational dynamics by MD simulation technique. The RNA-DNA hybrids considered for the modeling are complexed with Hs RNase H1 is (Table 3.5). The outcome of this study will offer the stereochemical rationale for the susceptibility of TRE containing mRNAs to antisense strategy as well as the effect of persistant RNA-DNA hybrids formed by TREs in pathogenesis of the diseases. Exceptionally, this study would not be relevant in the case of GAA repeat over expansion in the perspective of antisense strategy as it does not get transcribed to mRNA.

Results

5.2 Docking of Hs RNase H1 and TRE containing RNA-DNA hybrids

RNA-DNA hybrids with various trinucloetide repeats mentioned in Table 3.2.1 have been docked to HsRNase H1 (PDB ID-2QK9) using PyMOL. The protein with D210N mutation is converted to wild type. Various Interactions between HsRNase H1 and the hybrid were analyzed extensively from the X-ray crystal structure and 17 interactions were considered for modelling the various TREs containing hybrids with HsRNase H1 (Table 5.1). During modeling, we tried to retain most of the interactions given in Table 5.1.

Table 5.1 Various interactions considered for modeling HsRNase H1 and the hybrid		
Catalytic Site	**DNA binding site**	**RNA binding site**
Asp 145-R10 OD2...O1P	Arg 179-D30 NH2...O1P	Cys 148-R10 C=O...O2'
Glu 186-R9 OE2...O2'	Thr 181-D30 OG1...O1P	Asn 182-R9 ND2...N3
Asp 210-R8 C=O...O2'	Asp 240-D30 ND2...O1P	Met 212-R8 N...O2'
Asp 274-R9 OD2...O1P	Asn 151-D28 ND2...O2	His 260-R9 NE2...O1P
	Gln 183-D29 OE1...N3	Gly 263-R9 N...O1P
	Trp 225-D32 NE1...O1P	Arg 278-R11 NH1...O1P
	Ser 233-D32 N...O2P	
	Ile 239-D31 N...O1P	
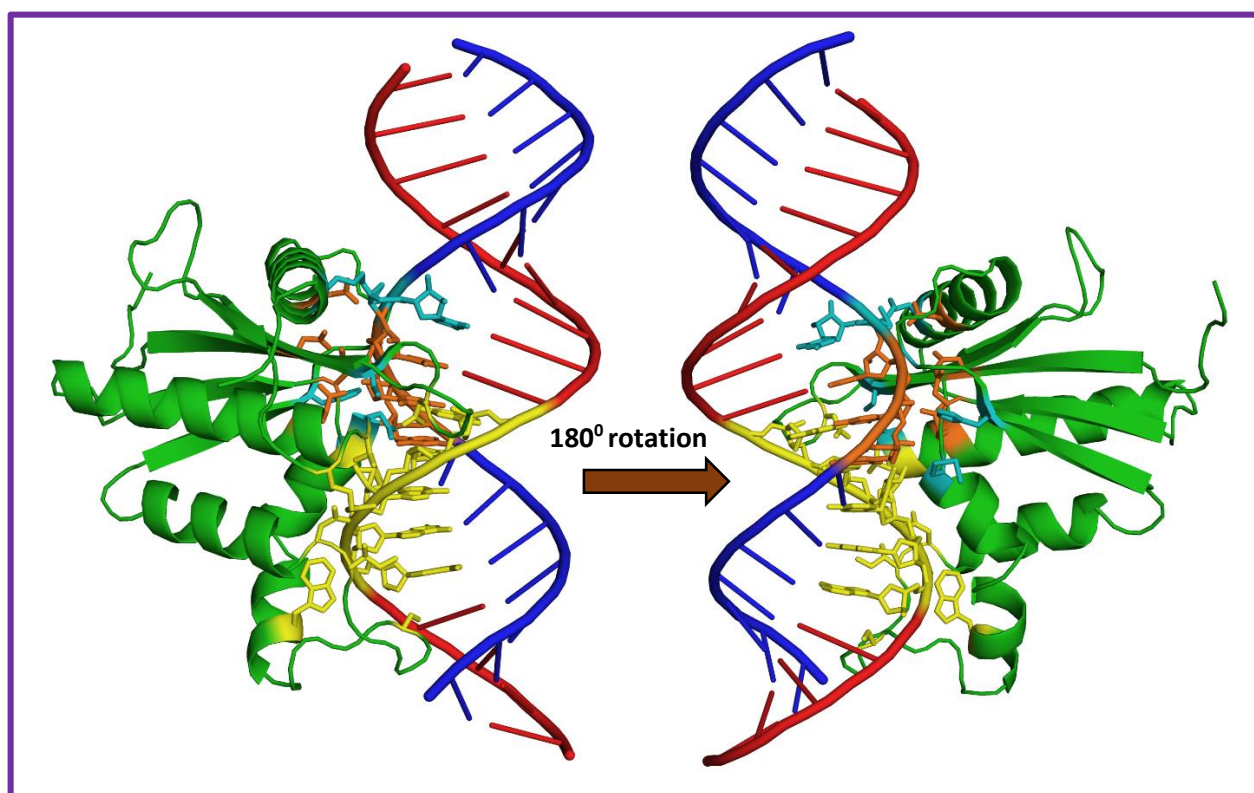
Figure 5.3 Illustration of interactions considered for docking HsRNase H1 and dGAA-rUUC hybrid. RNase H1 is colored green, the substrate is shown in blue (RNA) and red (DNA) color. Interaction between active site residues (D145, E186, D210 and D274) and RNA residues are shown in orange ball-and-stick model, DNA binding sites are shown in yellow while RNA binding sites in cyan. 180° rotation of left is given in the right side.

All the RNA-DNA hybrids with various TRE's constructed above were superimposed on each docked models are similar to each other (Figure 5.4). The docked models were subjected to 100ns MD simulation.
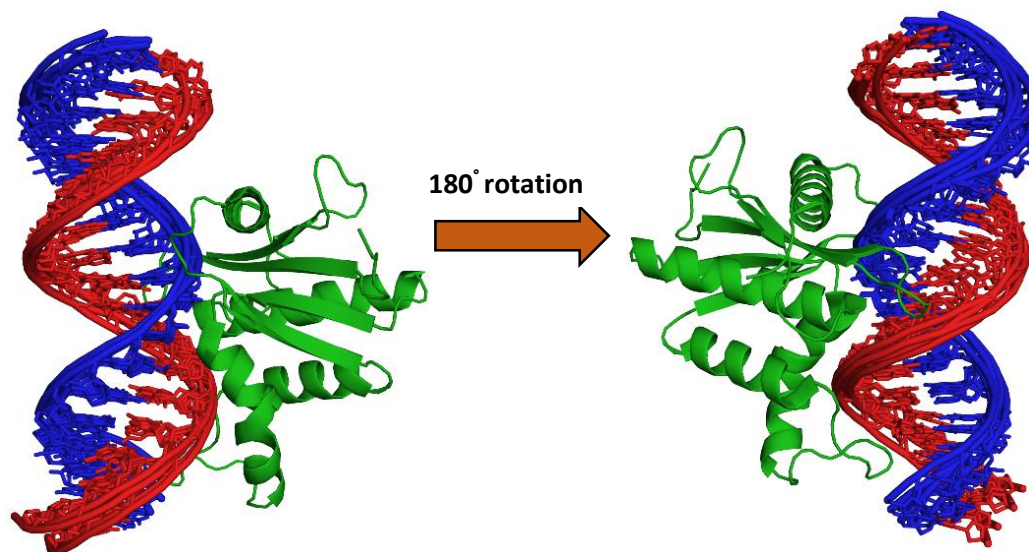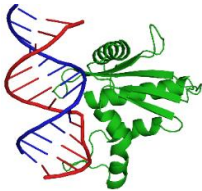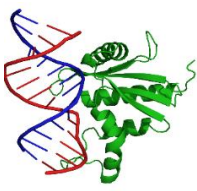
Figure 5.4 Superimposition of all the docked models of TRE's containing hybrids with HsRNase H1. (Left) All the 18mer hybrids with TRE's with HsRNase H1 are superimposed. (Right) 180° rotation of left. Note that 18mer sequence is considered for superimposition.

5.3 MD Simulation

Since, the crystal structure of RNA-DNA hybrid [r(AGUGCGACACCUGAUUCC)-d(GGAATCAGGTGTCGCACT)] HsRNase H1 is solved by mutating the residue D210 of the wild type to N210. This simulationcan also act as the control simualtion fro the HsRNase H1-TRE containing RNA-DNA hybrid system. In addition, we also simulated the D210N mutant of the RNase H1 –RNA-DNA hybrid complex. Table 5.2 represnts the average structure calculated at the 1ns & 100ns of the wild-type (wd) and mutant Hs RNase H1 complex structures. Note that the significant tilt in the initial models which significantly reduced at the end of the simulation.

Table 5.2 Time averaged structures of HsRNase H1 –RNA-DNA hybrid complex.		
	wild-type	**mutant**
1 nano second		
100 nano seconds		

5.3.1 Effect of simple point mutation on the activity of HS RNase H1

Although it can be well understood that change of D to N will change characteistics of whole protein, due to lack of structural evidance for the wd Hs RNase H1-hybrid complex, it is not well understood that how the presence of D210 would facilitate the cleavage of scissile phosphate. To address this, the movement of the loop consisting of 210 residue was considered. Very intrestingly. it was found that wild-type with D210 residue comes close to the scissile phosphate within first 15 ns and stay there while the one with D210N residue moves away from scissile phosphate and remains till end of simualtions (Figure 5.5). This will be further explained in chapter 5.2.6 through electrostatics.
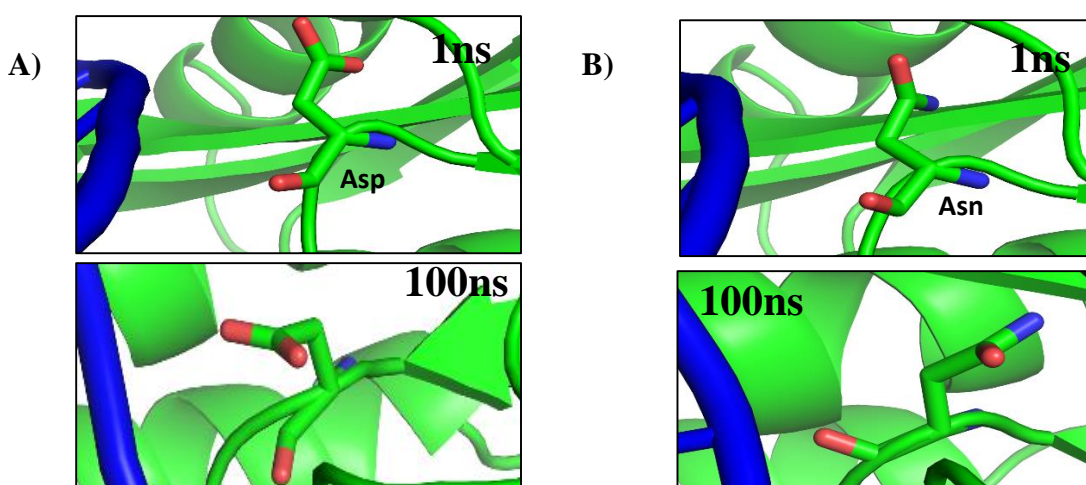

Figure 5.5 Illustration of interaction of wild-type and mutated HsRNase H1 with the hybrid. (A & B) Side chain position of wd (A) and mutant (B) RNase H1 at 1ns (top) & 100ns (bottom). Note the tendency of D210 to move towards the scissile phosphate. 210 residue is shown in green ball and stick model, orange represents oxygens and blue stick represents nitrogen.

Control simulations were studied in depth to understand the cleavage mechanism and inhibition of cleavage by D210N mutation. Main focus was on the substrate recognition pattern of HsRNase H1 and the catalytic site where cleavage takes place.

5.3.2 Substrate Recognition by HsRNase H1

HsRNase H1 profusely interact with DNA and RNA to facilitate substrate recognition. It interacts with only 11 base-pairs of 18mer hybrid. Extensive network of interactions occur between these 11 base-pairs and HsRNase H1. Out of 17 interactions found in the crystal structure, 10 were found to be reatined during simulation.

Table 5.3 Retained interactions between HsRNase H1 and the hybrid at 100ns		
Note that the interactions retained were for both wd & mutant. Hybrid sequence is r(AGUGCGACACCUGAUUCC)-d(GGAATCAGGTGTCGCACT)		
Catalytic Site	**DNA binding site**	**RNA binding site**
Asp 145-R10	Arg 179-D30	Cys 148-R10
	Thr 181-D30	Asn 182-R9
	Asn 151-D28	Met 212-R8
	Trp 225-D32	
	Ser 233-D32	
	Ile 239-D31	

5.3.3 Catalytic site ion pocket

Aforementioned in introduction, HsRNase H1 contain a conserved catalytic site with four acidic residues that coordinate with two divalent Mg^{2+} or Mn^{2+} ions. Mg^{2+} or Mn^{2+} ions were not provided in the present study to facilitate the overall study of the complex from the initial recognition of the hybrid till the cleavage takes place. To facilitate cleavage of RNA strand, four Na^+ ions (Figure 5.6A) occupy the site replacing the activity of 2 Mg^{2+} or Mn^{2+} ions. This study supports two-metal ion catalysis mechanism by HsRNase H1 (Nowotny et al., 2007). These four Na^+ ions are accommodated in the pocket formed at the cleavage site at the minor groove of the RNA. Only 3 Na^+ ions are observed in Mutant HsRNase H1 due to the neutral charge of the amino acid.
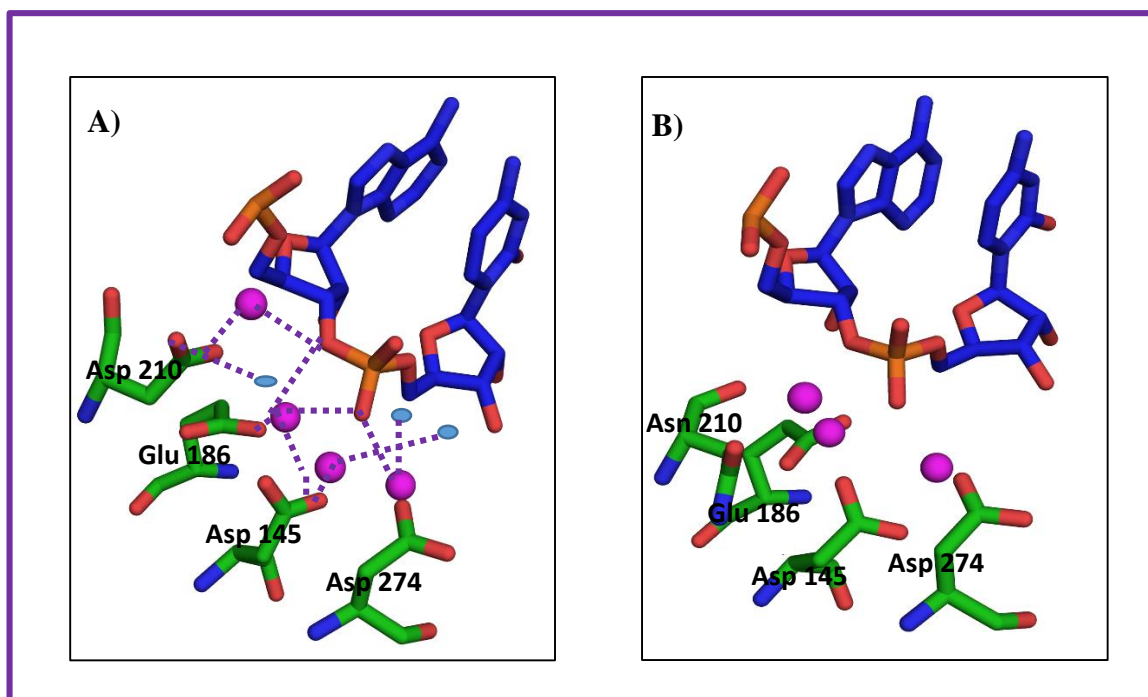
Figure 5.6 The active site of Hs-RNase H1 wild-type and D210N. A) Catalytic site of the wide-type complex. B) Non-productive catalytic site of complex with D210N mutant complex. Pink spheres are Na$^+$ ions and water are indicated in blue spheres.

5.3.4 Loop opening and locking

Although it was shown previously that loop connecting Helix E & sheet 5 is non conserved (Nowotny et al., 2007), our study clearly demonstrates that the specific loop has a major role in catalysis. This is by allowing ions to leave and occupy the catalytic site which is inturn very crucial for cleavage to take place. The loop is found to open and lock to facilitate the movement of ions in and out of the ion pocket (Figure 5.7).
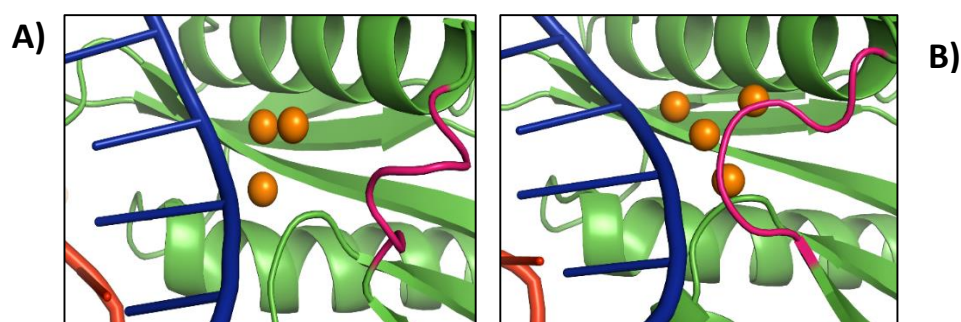
Figure 5.7 The illustration of loop opening and locking at the active site to facilitate movement of Na$^+$ ions in wd HsRNase H1. A) loop opening to facilitate intake of Na$^+$ ions B) Loop comes to locked state once 4 Na$^+$ ions are incorporated into the catalytic site. Loop is shown in hot pink colour and Na$^+$ ions in orange. Note that all the above structures are at 100ns.

5.3.5 Arginine Loop

Significant differences in the complex was observed in the loop lying between sheet 1 and 2 which is found to be Arginine rich with 3 arginine residues together and is basic in nature due to positive charge of the Arg. In the present study, it was observed that the Arg-153 of this loop changes its conformation from compact to extended i.e. rotation around CD atom (Figure 5.8). Earlier it was reported that this Arg residue that maybe responsible for initial recognition of the substrate i.e. hybrid. This was in accordance with its movement towards the nucleic acid where it was found to extend the conformation and interact with the hybrid. These interactions with DNA and RNA residues incorporate kink in the hybrid which inturn facilitates bending of the hybrid towards minor groove and makes the hybrid compact. Thus, overall helping the protein to recognize the hybrid. The same was observed in mutant type thus, conforming the observation.
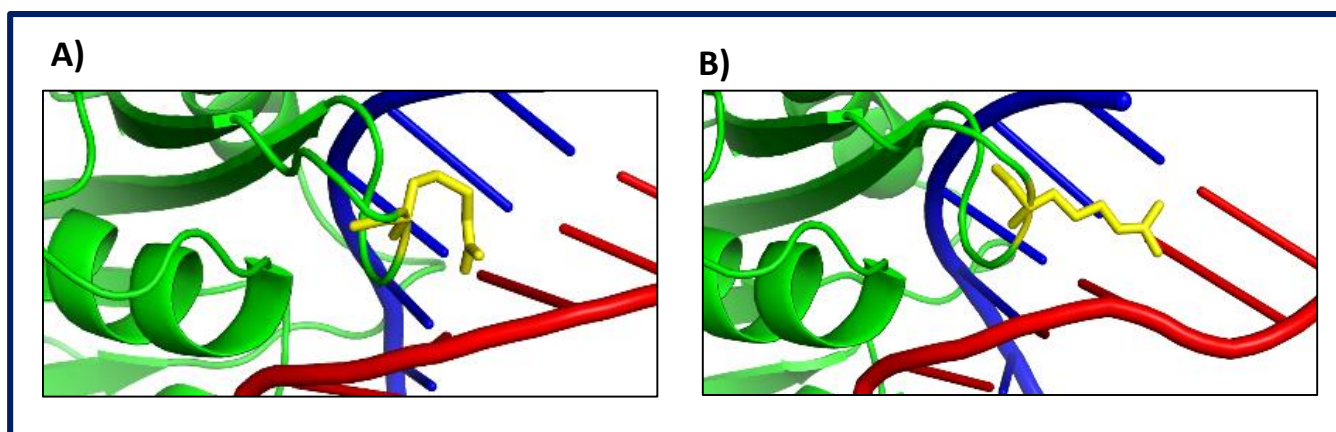
Figure 5.8 Arginine-153 (shown as yellow sticks) staggers over the hybrid to facilitate substrate recognition. (A & B) 2 different side chain conformation of Arg.

5.3.6 Electrostatics

The electrostatic properties of proteins depend on distribution of polar and charged residues which form interactions, like salt-bridges and hydrogen-bonds, and by calculating the over-all electrostatic environment in the protein. Electrostatics define protein-protein complex formation, thermal stabilities, conformational adaptabilities and protein movements. In the present study, with relevance to section 5.2.1, electrostatics properties of HsRNase H1 were studied to know how just a point mutation would affect the activity of the whole protein. HsRNase H1 wild-type with D210

residue at catalytic site was highly electronegative which upon point mutation to D210N reduced the electronegativity of the catalytic site (Figure 5.9). This reduction in charge doesn't facilitate accommodation of Mg^{+2} ions, which is necessary for RNase H activity.
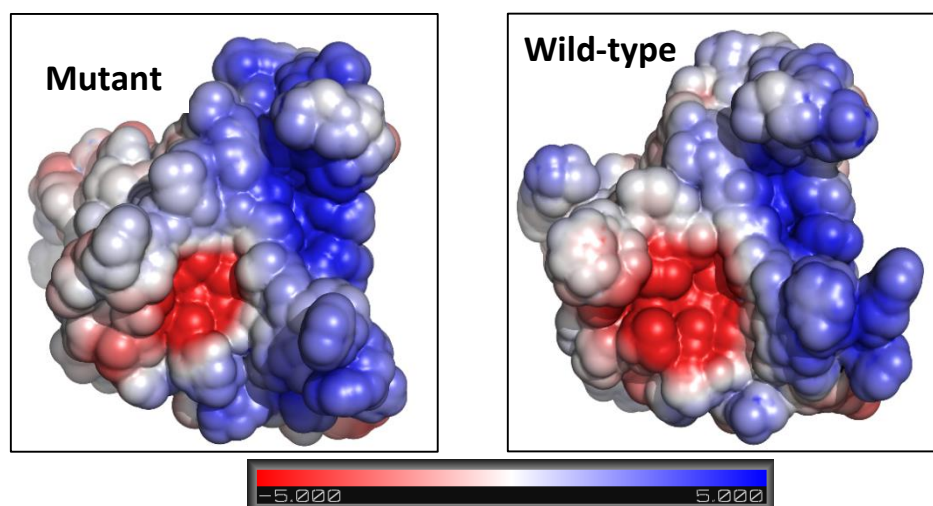
Figure 5.9 Electrostatic surface of Hs RNase H1. (Left) Mutant HsRNase H1 D210N shows less electronegativity when compared to wild type (right). Red and blue regions indicate negative and positive charge, respectively.

5.4 Structure and dynamics of TRE containing RNA-DNA hybrids and HsRNase H1 complexes

As mentioned earlier in the chapter introduction, structural insight into the complex of TREs containing RNA-DNA hybrid and HsRNase H1 would enable us to understand the cleavage mechanism of HsRNase H1 and incorporate the same knowledge in designing an antisense strategy for treating TREDs at RNA level. Structural characterization of various TREs containing hybrids were done.

5.4.1 Hybrid characterization

Analysis was carried out for characterizing the RNA-DNA hybrids with various TREs in complex with HsRNase H1. This study was then compared with the survey conducted on hybrids in complex with various proteins. The summary concludes that the hybrids with TREs complexed with HsRNase H1 still retain AB-intermediate characteristics with X-displacement, helical twist and roll favoring A-RNA and rise & groove widths favoring B-DNA. Negative roll was observed in purine rich RNA stand.

Table 5.4 Time average structures of HsRNase H1 in complex with various TRE's containing RNA-DNA hybrids.

Note: Average structures for every 500ps were considered for analysis.

Time scale	d(GAA)₆. r(UUC)₆	r(GAA)₆. d(TTC)₆	d(CTG)₆. r(CAG)₆	r(CUG)₆. d(CAG)₆	r(CCG)₆. d(CGG)₆	d(CCG)₆. r(CGG)₆
1nano second						
100nano seconds						

Table 5.4 Summary of analysis of TRE's containing RNA-DNA hybrid complexed with HsRNase H1

	d(GAA)₆. r(UUC)₆	r(GAA)₆. d(TTC)₆	d(CTG)₆. r(CAG)₆	r(CUG)₆. d(CAG)₆	r(CCG)₆. d(CGG)₆	d(CCG)₆. r(CGG)₆	Hybrid in wd	Hybrid in mutant
X-disp (Å)	-4.48(2.04)	-4.34(1.64)	-4.34(1.64)	-5.78(5.94)	-4.12(1.71)	-2.01(6.85)	-5(4.32)	-3.73(1.25)
Twist (º)	29.15(6.1)	29.71(4.06)	30.56(3.77)	28.61(4.32)	30.36(4.25)	29.18(11.67)	28. 9(7.26)	29.8(4.5)
Rise (Å)	3.29(0.11)	3.42(3.25)	3.33(0.21)	3.36(0.47)	3.39(0.23)	3.44(0.2)	3.36(0.21)	3.38(0.2)
Roll (º)	6.88(5.03)	4.04(4.38)	7.27(6.16)	6.13(5.29)	6.5(4.01)	4.77(7.64)	5.1(3.87)	5.2 (6.09)
Tilt (º)	2.38(1.94)	-1.9(2.21)	-1.91(2.5)	2.74(2.9)	3.69(1.72)	-3.89(3.29)	3.32(3.3)	2.79(3.21)
Groove Major (Å)	18.7	17.9	18.6	22.2	16.3	16.0	21.3	18
Minor (Å)	13.7	13.3	15	12.1	14.7	14.2	12.7	11.2
Sugar (RNA)	C3'endo	C3'endo/ C2'exo	C3'endo	C3'endo	C3'endo/ C2'exo	C3'endo/ C2'exo	C3'endo/ C2'exo	C3'endo/ C2'exo
(DNA)	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo	C1'exo/ C2'endo/ O4'endo

5.5 Interactions between TREs containing hybrid and HsRNase H1

5.5.1 Ion pocket

The catalytic site in HsRNase H1 complexed with various TREs containing RNA-DNA hybrids were in favor of the observation carried out for wild-type HsRNase H1. The catalytic site was occupied by 4 Na^+ ions in all the complexes (Figure 5.10).
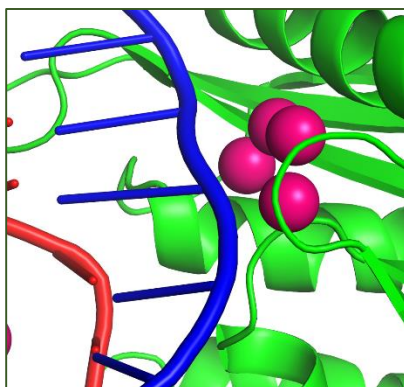

Figure 5.10 Occupancy of ion pocket by 4 Na^+ ions and loop locking the passage of ions out from the catalytic site of rCUG-dCAG. Na^+ ions are depicted as pink spheres. Note that the image is at 100 ns.

5.5.2 Loop opening and locking

Loop opening and closing was seen to be associated with ions. Loop was seen to be open to facilitate the occupancy of catalytic site by Na^+ ions. Same tendency was observed in hybrids containing TREs complexed with HsRNase H1. Loop opens to accommodate the incoming ions, once these 4 Na^+ ions occupy the catalytic site, the loop closes and hinders the movement of ions to outside facilitating catalysis (Figure 5.10). Loop thus, indirectly helps in cleavage of the RNA strand by HsRNase H1. Since, Loop opening and locking was associated with the number of ions present in ion pocket, rGAA-dTTC which has only 3 Na^+, do not close at the end of the 100 ns.

5.5.3 Base pair breathing event

Base pair breathing usually occurs when the Watson-Crick hydrogen bonds are temporarily disrupted as one of the base flips out of the helix. This may happen due to the base being exposed to the solvent or might be recognized by a DNA-binding protein. Base pair breathing was observed in our simulation of dGAA-rUUC-complex. Breathing started around 93ns and remained up to 97ns i.e. for 4ns. Breathing might be due to interaction with water molecules. Finally, Watson-

crick pairing is retained by 97ns and remained further. One might expect such breath in other TRE simulation also when they complex with the enzyme.
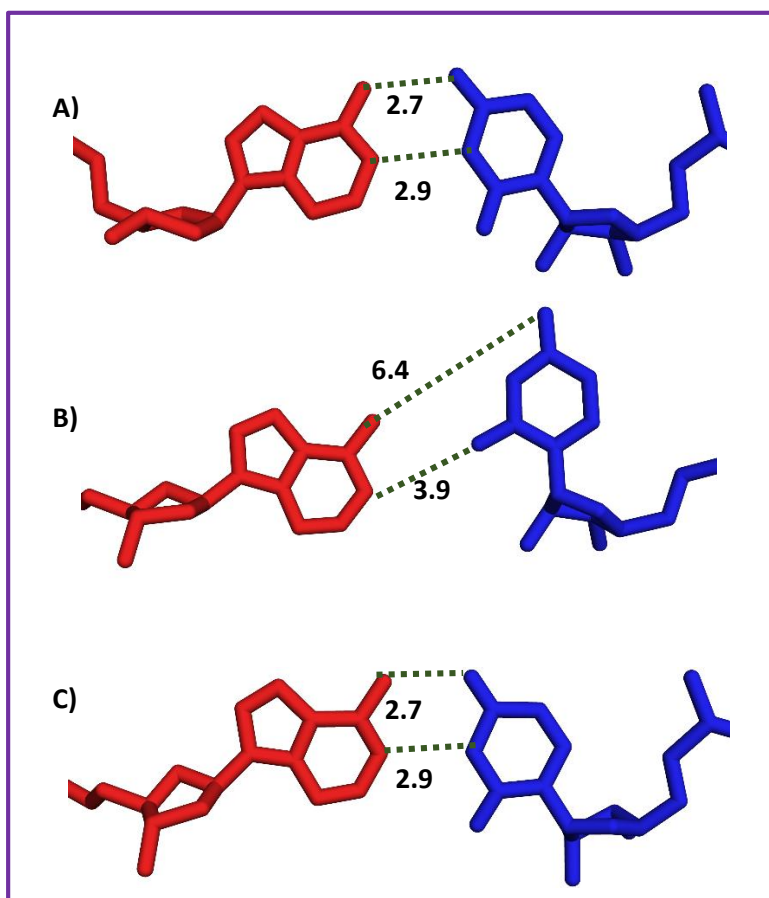

Figure 5.11 Depiction of base-pair breathing in rU-dA in dGAA-rUUC complexed with HsRNase H1. A) Initial Watson-crick base-pairing at 1ps structure. B) Breaking of Watson-crick base-pairing and base rU moves towards the minor groove. C) Retainment of Watson-crick base-pairing. Note that hydrogen bonding is seen in A & C but B with broken lines are not hydrogen bonds and are just for representation.

5.5.4 Arginine interaction with the hybrid

Arginine-153 was observed to change its conformation in order to facilitate the initial interactions with the hybrid. Figure 5.12 illustrates the straggling movement of Arg-153 towards the hybrid. Here it was seen in compact, slightly extended and highly extended conformation based on the initial interactions to be done with the hybrid. In rCUG-dCAG, this residue is found to climb over

the next base-pair to interact with the purine aid in bending of the structure and increase the compactness enhancing the substrate recognition by HsRNase H1.
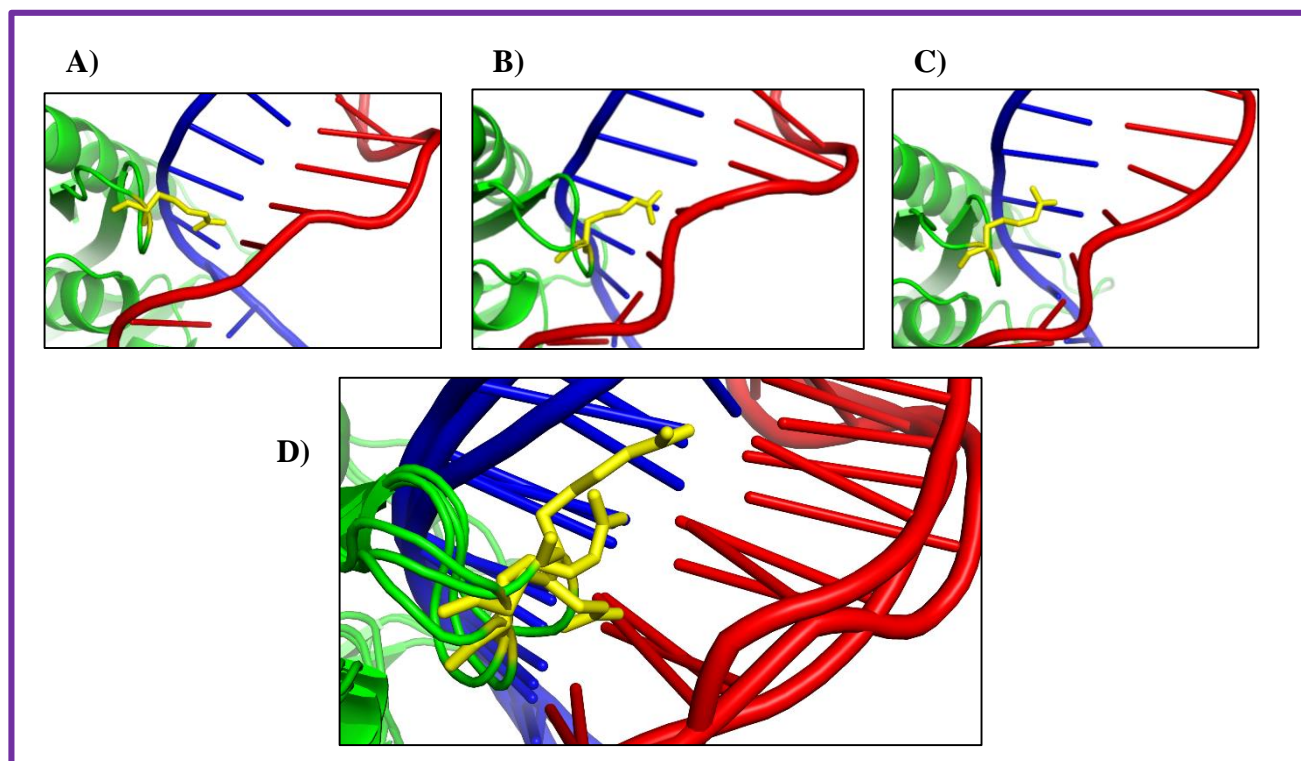

Figure 5.12 Complex of HsRNaseH1 with RNA-DNA hybrids containing A) dGAA-rUUC B) dCCG-rCGG and C) rCCG-dCGG repeat. D) Ensemble of all 3 models to depict the climbing of Arg-179 residue towards purine rich strand to compact the overall structure of the hybrid.

5.6 Conclusion

MD simulations were performed on hybrids with various TREs complexed with HsRNase H1 to understand the structural determinants and cleavage mechanism for RNase H activity. Our 100ns MD simulations of HsRNase H1 with various TREs containing RNA-DNA hybrids show that the cleavage mechanism of HsRNase H1 is same for RNA-DNA hybrids with various TREs indicating that this enzyme can act upon any RNA-DNA hybrid irrespective of the sequence. In the present study, the catalytic site gets activated once the its valancy is satisfied (De Vivo et al., 2008). This is observed in accordance with our simulation which has no Mg^{+2} ions but the neutralizing Na^+

ions and these ions play an important role in the catalysis. Presence of 4 Na^+ ions in the catalytic site and the loop associate gets locked supporting the above observation (Figure 5.6 and 5.10). The catalytic site consisting of four conserved residues interact with metal ions, water and 2'-OH of the nucleotide adjacent to the scissile phosphate and the scissile phosphate (Figure 5.6). water molecules act as a nucleophile.

Further, it was observed that the initial interaction between RNA-DNA hybrids by HsRNase H1 depends on the Arg loop where this residue is found to interact with the hybrid base-pairs and incorporate bending (Figure 5.8), due to which, the compactness of the hybrid increased. The overall study suggest that the present study can be made use of in antisense strategy in treating TREDs. This strategy can be made use for all the hybrids which reach upto the RNA level. For instance, hybrids with $(GAA)_n$ can 't come till RNA level as it forms an R-loop that is stable and obstruct RNA Polymerase II transcription in patient cells (Groh et al., 2014).

An approach for designing an antisense strategy could be challenging but suggestions made are to incorporate LNA at the terminals of the DNA oligonucleotide rather than the whole sequence (Priyakumar and Mackerell, 2008, Suresh and Priyakumar, 2013). The presence of Locked nucleic acid (LNA) (2'-*O*-4'-*C*-methylene bridge that locks the molecule in a C3'-*endo* sugar conformation) in the whole sequence may hinder the sugar backbone and result in poor recognition and activity of HsRNase H1. Peptide nucleic acid (PNA) (normal phosphate linkage found in DNA and RNA is replaced by an neutral peptide-like *N*-(2-aminoethyl)glycine backbone), unlocked nucleic acid (UNA) (C2'-C3' bond cleaved and the whole structure is very flexible unlike LNA), 5'-O-methylphosphonate nucleic acids (Sipova et al., 2014), 3'-O-P-CH2-O-5' phosphonate linkages all can be made use in designing an DNA oligonucleotide that can exhibited a significant increase in RNase H cleavage activity.

5.7 References

DE VIVO, M., DAL PERARO, M. & KLEIN, M. L. 2008. Phosphodiester cleavage in ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J Am Chem Soc,* 130**,** 10955-62.

GROH, M., LUFINO, M. M., WADE-MARTINS, R. & GROMAK, N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet,* 10**,** e1004318.

NOWOTNY, M., GAIDAMAKOV, S. A., GHIRLANDO, R., CERRITELLI, S. M., CROUCH, R. J. & YANG, W. 2007. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell,* 28**,** 264-76.

PRIYAKUMAR, U. D. & MACKERELL, A. D., JR. 2008. Atomic detail investigation of the structure and dynamics of DNA.RNA hybrids: a molecular dynamics study. *J Phys Chem B,* 112**,** 1515-24.

SIPOVA, H., SPRINGER, T., REJMAN, D., SIMAK, O., PETROVA, M., NOVAK, P., ROSENBERGOVA, S., PAV, O., LIBOSKA, R., BARVIK, I., STEPANEK, J., ROSENBERG, I. & HOMOLA, J. 2014. 5'-O-Methylphosphonate nucleic acids--new modified DNAs that increase the Escherichia coli RNase H cleavage rate of hybrid duplexes. *Nucleic Acids Res,* 42**,** 5378-89.

SURESH, G. & PRIYAKUMAR, U. D. 2013. Structures, dynamics, and stabilities of fully modified locked nucleic acid (beta-D-LNA and alpha-L-LNA) duplexes in comparison to pure DNA and RNA duplexes. *J Phys Chem B,* 117**,** 5556-64.