

TAR-RNA Recognition by a Novel Cyclic Aminoglycoside Analogue

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von

Devanathan Raghunathan

aus New Delhi, India

Göttingen 2007

Referent: Prof. Dr. Christian Griesinger

Korreferent: Prof. Dr. Ulf Diederichsen

Tag der mündlichen Prüfung:

Zusammenfassung

Die Bildung des Tat-Protein/TAR RNA-Komplexes ist ein entscheidender Schritt in der Regulation der Expression des HI-Virus (*Human Immunodeficiency Virus*, HIV). Für eine vollständige Transkription des viralen Gens ist die Interaktion des Tat/TAR-Komplexes mit dem positiven Transkriptionsfaktor-Komplex P-EFTb (*Positive Transcription Elongation Factor*) über dessen Cyclin T1-Komponente (*CycT1*) notwendig. Durch Mutagenesestudien wurde die Hexanukleotid-Schleife der TAR RNA als Kontaktstelle für die Wechselwirkung mit CycT1 identifiziert.

Zur Entwicklung neuer Arzneimittel gegen das HIV stellt die Störung des Zusammenspiels zwischen dem Tat/CycT1-Komplex und der TAR RNA ein lohnendes Ziel dar. Positiv geladene Verbindungen wie Aminoglycoside oder Peptidmimetika binden an die TAR RNA und brechen so den Tat/TAR-Komplex auf.

In dieser Arbeit wird die Bestimmung der dreidimensionalen Struktur des Komplexes zwischen der HIV-2 TAR RNA und einem Neooligoaminodeoxysaccharid mit Hilfe der NMR-Spektroskopie beschrieben. Im Gegensatz zu anderen Aminoglycosiden wechselwirkt diese neuartige Verbindung gleichzeitig mit den für die Bindung des Tat-Proteins verantwortlichen Resten des Bulges wie auch mit dem Adenosin 35 der Hexanukleotid-Schleife der TAR RNA. Diese Schleifenregion erfährt bei der Bildung des Komplexes mit dem Aminoglycosid eine große konformationelle Änderung.

Dieser neue Bindungsmodus eröffnet zusammen mit der einfachen synthetischen Zugänglichkeit von Neooligoaminodeoxysaccharid-Derivaten die Möglichkeit, eine neue Klasse von TAR RNA bindenden Molekülen zu entwerfen. Diese könnten gleichzeitig die Bildung des binären Tat/TAR- wie auch des ternären Tat/TAR/CycT1-Komplexes durch Störung der Schleifen- und Bulge-Region der RNA verhindern.

Stichwörter: Aminoglycoside, TAR RNA, NMR, Bindungsmodus

Abstract

The formation of the Tat-protein/TAR RNA complex is a crucial step in the regulation of Human Immunodeficiency Virus (HIV)-gene expression. To obtain full-length viral transcripts the Tat/TAR complex has to recruit the positive transcription elongation factor complex (P-EFTb), which interacts with TAR through its CyclinT1 (CycT1) component. Mutational studies identified the TAR hexanucleotide loop as a crucial region for contacting CyclinT1. Interfering with the interaction between the Tat/CycT1 complex and the TAR RNA is an attractive strategy for the design of anti-HIV drugs. Positively charged molecules, like aminoglycosides or peptidomimetics, bind the TAR RNA, disrupting the Tat/TAR complex. Here, we investigate the complex between the HIV-2 TAR RNA and a neoligoaminodeoxysaccharide by NMR spectroscopy. In contrast to other aminoglycosides, this novel aminoglycoside analogue contacts simultaneously the bulge residues required for Tat binding and the A35 residue of the hexanucleotide loop.

Upon complex formation, the loop region undergoes profound conformational changes. The novel binding mode, together with the easy accessibility of derivatives for the neoligoaminodeoxysaccharide, could open the way to the design of a new class of TAR RNA binders, which simultaneously inhibit the formation of both the Tat/TAR binary complex and the Tat/TAR/CyclinT1 ternary complex by obstructing both the bulge and loop regions of the RNA.

Keywords: Aminoglycoside, TAR RNA, NMR, binding mode

Acknowledgements

This thesis is the culmination of three and a half years of work whereby I have been accompanied and ably supported by many people, who have further strengthened my faith that interdependence is as valuable a virtue as independence is. Looking back, the PhD. has been a challenging yet gratifying and an “eye opening” experience. It is indeed an honor and a pleasure to express my gratitude to those who have been instrumental in my scientific growth and more importantly, in the successful completion of my Ph.D. dissertation.

The first person I would like to thank is my supervisor Teresa Carlomagno. I owe a lot of gratitude for her guidance, insights, constructive criticism and encouragement. I have had the opportunity to learn a lot about the subject under her able supervision.

I am also very grateful to Claudia Schwiegk, who introduced me to the nuances of synthesizing and labeling nucleic acids. Her technical expertise and skill contributed immensely to the success of my research. Sincere thanks to Victor Sanchez Pedregal for having been an excellent coworker and mentor in the initial days of my PhD.

A thanks goes out to Jochen Junker for his generous help with the computational aspects of my work. I shall remain indebted to him for having introduced me to the Apple Mac systems and having helped me broaden my horizons from just the PC to the Mac and Linux based machines. A heartfelt thanks to my colleagues Christophe Fares, Marcel Reese, Jörg Fohrer, Dirk Bockelmann, Marco Mukrasch, Peter Haberz, Ping Li and Volker Klaukien who provided a wonderful environment at work.

Jegannath, Raghav, Madhu, Venkatesh and Kiran added the touch of “Indian”ness to my stay in Göttingen and provided me with the much needed “home away from home” environment. They are friends who’s association I shall cherish all my life.

To Kalpana, I owe thanks for being a tremendous source of encouragement and a very special friend who stood by me through thick and thin.

The time spent at the Max Planck has been a knowledge enhancing one and one that allowed me to interact with people from various countries and disciplines. I thank Professor Griesinger who provided me with the financial support that enables me to leave the Max Planck not only enhanced with knowledge, but also enriched culturally. The Max Planck Society, DFG (Deutsche Forschungsgemeinschaft) and SFB (Sonderforschungsbereich) have been gracious with their financial assistance at

various points of my PhD. I owe them thanks for supporting my research and my stay here in Germany.

None of this work would have been possible had it not been for the support, love and encouragement I received from my family. I thank my mother and father for their faith, affection and prayers that fuelled me on towards my goal at times when research seemed too daunting a task. I especially thank my sisters Jaishree and Latha for their loving support. I owe them all a special gratitude for being patient and allowing me to take on this opportunity to do research abroad, while still bearing the angst of separation for extended periods.

Finally, I would like to dedicate this work to my late grandparents who would have been immensely proud to see me get the coveted title of “Dr.”.

So long, and thanks for the PhD...

Devanathan Raghunathan

NMR Based Structural Biology

Max Planck Institute for Biophysical Chemistry

Am Faßberg 11, Göttingen D-37077, Germany

Table of Contents

List of Tables and Figures	viii
Abbreviations	x
1. Introduction	1
1.1 The TAR RNA	3
1.2 The Tat protein	6
1.3 The Tat-TAR complex	7
1.4 Positive Transcription Elongation Factor b (P-TEFb)	10
1.5 The Tat-TAR_cyclinT1 complex interactions	10
1.6 Inhibitors of the Tat-TAR RNA interaction	11
1.6.1 Antisense oligonucleotides	13
1.6.2 Cationic peptides	13
1.6.3 Intercalators	14
1.6.4 Aminoglycosides	14
1.7 Scope of the work and questions we address	16
2. Methods	18
2.1 HIV-2 TAR RNA synthesis	18
2.2 Cyclic Aminoglycoside AMG	20
2.3 RNase footprinting / Digestion	22
2.4 PolyAcrylamide Co Electrophoresis	23
2.5 NMR spectroscopy	26
2.5.1 Resonance assignment of the TAR RNA	27
2.5.2 Resonance assignment of the Aminoglycoside	31
2.5.3 Titrations and chemical shift mapping	32
2.5.4 T1 relaxation rate measurement	33
2.5.5 Field induced residual dipolar couplings	35

2.5.6 NOE distance restraints	37
2.5.7 Dihedral angle restraints	40
2.5.8 Planarity and hydrogen bonding	43
2.5.9 Structure calculations	44
3. Results	50
3.1 Resonance assignment	50
3.2 Chemical shift mapping	55
3.3 Sugar pucker and dihedral angles	57
3.4 Identifying the binding sites of AMG to TAR RNA	58
3.5 Field induced residual dipolar couplings	59
3.6 NOE distance restraints	62
3.7 Structure calculations	63
3.8 The upper stem binding site	66
3.9 Definition of the second binding site and of the dimer geometry	68
3.10 RNase digestion and PACE gel analysis	72
3.11 Structure statistics	79
4. Discussions	80
4.1 The binding of AMG to multiple sites on the upper stem	80
4.2 Cyclic scaffold of AMG	85
4.3 The comparison of the binding of AMG to the TAR RNA with the binding of CyclinT1 to the TAR RNA loop	86
4.4 The lower stem binding site and the dimerization model	88
4.5 Significance for drug design	90

5. Conclusion and Perspectives	91
6. Bibliography	92
7. Appendix	101
7.1 Appendix A: Structural restraints for NMR structure calculation	101
7.2 Appendix B: Xplor-NIH Input protocols for structure determination	110
7.3 Appendix C: Scripts	129
Curriculum Vitae	147
Lebenslauf	148

List of Tables and Figures

- F1.1 (a) HIV-1 and HIV-2 TAR RNA sequence comparison
- F1.1 (b) Motional amplitude of stems with respect to each other in free TAR RNA
- F1.1 (c) Lowest energy free TAR RNA structure
- T1 Viral and cellular proteins involved in TAR RNA functions
- F1.2 Tat protein genome
- F1.3 (a) TAR-argininamide binding site
- F1.3 (b) Tat-TAR interactions
- F1.6 Interfering with the HIV transcription apparatus
- F1.6.2 Cationic peptide CGP 64222
- F1.6.4 2-Deoxystreptamine
- F2.1 HIV-2 TAR RNA sequence
- T2.1 *In-vitro* T7 transcription reaction components for TAR RNA synthesis
- F2.2 Synthesis of AMG
- T2.2 Equipment and Suppliers
- F2.5.1(a) 3D-HCCH-COSY-TOCSY pulse sequence
- F2.5.1(b) 3D-Base Selective NOESY pulse sequence
- F2.5.2 The 1D region of AMG
- F2.5.3(a) Titration AMG vs. TAR RNA followed on the 1D Watergate
- F2.5.3(b) ¹³C HSQC of Bound TAR vs. Unbound TAR RNA
- F2.5.4 Pulse sequence to measure T1 relaxation rate in the TAR/AMG complex
- F2.5.5 Coupled and decoupled peaks on a ¹³C-HSQC spectra to measure *fi*RDCs
- T2.5.6 Intermolecular NOEs between AMG and TAR RNA
- F2.5.6(a) Labeling and scheme of AMG
- F2.5.6(b) Numbering scheme of AMG
- F2.5.6(c) Intermolecular NOE data sets scheme.
- F2.5.7(a) Ribose sugar puckers
- F2.5.7(b) Dihedral angles associated with a nucleotide
- F2.5.7(c) HCCH-E.COSY pulse sequence
- F2.5.7(d) HCCH-E.COSY splitting pattern
- F2.5.7(e) NOEs indicative of A-form helix formation in RNA
- F3.1(a)(b) 2D skewer of HCCH-COSY-TOCSY spectrum
- T3.1(a) Resonance assignment of ribose moieties of HIV-2 TAR RNA

- T3.1(b) Resonance assignment of the base moieties of HIV-2 TAR RNA
- F3.1(c) The C1'-H1' region of HIV-2 TAR RNA in the ¹³C-HSQC spectrum
- F3.1(d) Base region of HIV-2 TAR RNA in the ¹³C-HSQC spectrum
- F3.1(e) Resonance assignment of AMG
- F3.2 Chemical shift mapping of TAR binding sites
- T3.3 Sugar pucker restraints for the nucleotides of TAR RNA
- F3.5(a) Sensitivity of *f*iRDC to RNA stoichiometry
- T3.5 Field induced dipolar couplings determined for the TAR RNA with errors
- F3.5(b) Bar graph with errors of *f*iRDCs against each residue of the TAR RNA
- T3.6 r.m.s.d. values for the core and all* residues for the different intermolecular data sets evaluated for the TAR/AMG complex.
- T3.7 Restraint statistics
- F3.7(a) Superposition of 14 best structures of the TAR/AMG complex
- F3.7(b) Effect of *f*iRDCs on the global and local structure of TAR/AMG complex
- F3.9 (i) Possible mode of association of TAR RNA and AMG during dimerization
- F3.9(ii) Ensemble of 10 structures of the dimer of HIV-2 TAR RNA/AMG
- F3.10(a) RNase footprinting
- F3.10(b) Autoradiogram of PACE gel
- T3.10(a) Experimental values of ϕ as measured from PACE gel.
- F3.10(c) Binding stoichiometry analyzed with PACE gel
- T3.10(b) r.m.s.d values of theoretical models fitted against experimental data
- T3.11 Structure statistics
- F4.1(a) Brownian dynamics prediction of multiple cation binding sites in TAR
- F4.1(b) Interaction of sugar moieties of AMG with residues of TAR RNA
- F4.1(c) Comparison of the bend of TAR RNA in TAR/argininamide and TAR/AMG complex.
- F4.1(d) Stacking interaction of G34 and G36 in TAR RNA loop.
- F4.2 Cyclically restrained analogues targeting TAR RNA
- F4.3 Two views of the dimer interface between the TAR RNAs.

Abbreviations

ϵ	Molar Extinction Coefficient
1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
AIDS	Acquired Immuno Deficiency Syndrome
ADE/A	Adenine
BIV	Bovine Immunodeficiency Virus
CDK9	Cyclin Dependent Kinase 9
COSY	Correlation Spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CT	Continuous Time
CTD	Carboxy Terminal Domain
CycT1	Cyclin T1
CYT/C	Cytosine
DOS	2-Deoxystreptamine
DTT	Dithiothreitol
E. Coli	Escherichia Coli
E. COSY	Enhanced Correlation Spectroscopy
EDTA	EthyleneDiamineTetraacetic Acid
FID	Free Induction Decay
f_1 RDC	Field Induced Residual Dipolar Coupling
GUA/G	Guanine
HAART	Highly Active Antiretroviral Therapies
HIV	Human Immunodeficiency Virus
HSQC	Heteronuclear Single Quantum Coherence
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IPTG	IsoPropyl- β -D-ThioGalactopyranoside
k_d	Dissociation Constant
KDa	kilo Dalton
LJ	Lennard-Jones potential
LTR	Long Terminal Repeat

NELF	Negative Elongation Factor
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NTP	Nucleotide Tri Phosphate
PACE	PolyAcrylamide Co Electrophoresis
PAGE	PolyAcrylamide Gel Electrophoresis
P-EFTb	Positive transcription elongation factor complex
RDC	Residual Dipolar Coupling
rMD	Restrained Molecular dynamics
r.m.s.d	Root Mean Square Deviation
RNase	Ribonuclease
RT	Room temperature
SA	Simulated Annealing
Sani	Susceptibility Anisotropy
TAR	Trans Activation Responsive element
TFA	Trifluoroacetamide
THF	Tetrahydrofuran
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
TOCSY	Total Correlation Spectroscopy
TRIS	Trishydroxymethylaminomethane
URI/U	Uridine
vdw	van der Waal

1 Introduction

The ultimate goal of therapy for HIV infection is to eradicate the virus. To date, therapeutic agents have only been able to push the virus towards a state of latency. Selected drugs can induce rapid, substantial and sustained viral suppression in some patients. However, recent evidence suggests that even in patients with undetectable levels of HIV RNA in plasma, latent reservoirs of virus exist that may require at least 60 to 70 years of therapy to completely eradicate the virus [1; 2]. Furthermore, current therapies are limited by problems such as drug resistance, poor drug compliance and long-term side effects. Thus, the search for novel therapeutic options and approaches continues.

Developments in the field of NMR, exemplified by the attainment of higher field strengths, introduction of the cryo probe technology, new alignment media, residual dipolar couplings as a sensitive probe of molecular structure and dynamics and novel labeling strategies have aided immensely in the search for new drugs [3; 4; 5; 6; 7; 8]. In parallel, rapid advances in related fields such as crystallography, microscopy, combinatorial chemistry and computational methods have provided greater momentum to the ongoing search for novel therapeutics. It is now routinely possible to obtain high-resolution structures of essential proteins and nucleic acids that play a key role in the interactions of important processes such as transcription and design new scaffolds that can specifically and effectively target such interactions. The resolution of structures of reverse transcriptase, protease, integrase and gp41 [9; 10] has already resulted in new drugs that are either approved for clinical use or in the trial phase [11]. Other HIV specific structures such as regulatory protein-RNA interactions such as the Tat-TAR and the Rev, RRE have also simulated rational drug development [12; 13]. Just thirty years after the identification of the HIV [14; 15], several antiviral drugs have been licensed and several more are in the process of development to combat this deadly disease.

Administering a concoction of drugs is generally more efficient than treating the disease based on monotherapy. Constant mutation and increasing drug resistance makes monotherapy inefficient. **Highly Active Antiretroviral Therapies (HAART)** consists of drugs aimed at multiple targets [16; 17]. Research has proved that HAART can reduce viral replication in the plasma to undetectable levels (< 50 RNA

copies/ml) sufficient to make the emergence of resistant variants highly unlikely [¹⁶; ¹⁷]. Further, multi drug resistance has not been detected and hence different drugs can be combined that, though effect the same molecular targets, show a sufficiently different spectrum of escape mutations to prevent cross-resistance [¹⁸].

Therapeutic techniques based on targeting RNA-protein interactions have attracted focus recently [¹⁹; ²⁰; ²¹; ²²; ²³; ²⁴; ²⁵]. RNA is involved in many biologically important processes such as control of gene expression and translation processes. The wide range of conformations that can be adopted by RNA, containing double stranded duplexes, hairpins, loops, pseudo knots, bulges etc. contribute to the structural diversity and functional specificity of secondary structures involved in RNA-protein interactions and cellular functions. Targeting the interaction between the viral Tat protein and the Trans Activation Response element (TAR RNA) is an example of targeting conserved regions and conserved structures in the viral genome [¹³; ²¹; ²⁶]. The virus encodes for a positive regulatory gene which is required for replication and increases gene expression directly by the HIV Long Terminal Repeat (LTR) promoter [⁹; ²⁷; ²⁸; ²⁹]. The TAR element of the HIV-RNA is a structured RNA that binds to a viral nuclear regulatory protein called Tat. Disrupting the binding of Tat to the TAR RNA will thus block viral replication [¹⁰; ²²; ³⁰; ³¹; ³²; ³³; ³⁴; ³⁵; ³⁶; ³⁷; ³⁸; ³⁹]. The 5'-TAR region of HIV mRNA is highly conserved amongst different HIV isolates and the TAR RNA is a highly conserved structure. The correct Tat/TAR RNA interaction is essential for viral gene expression. Tat also plays an essential role in AIDS pathogenesis and associated malignancies, both as an extra cellular protein [²⁸; ⁴⁰; ⁴¹; ⁴²; ⁴³] and as an activator of cellular gene expression encoding cytokines.

The following sections take a closer look at the structure and function of the TAR RNA, Tat protein and associated bio molecules that govern HIV transcription and replication.

1.1 The TAR RNA

The 5' end of all HIV mRNAs, whether they are spliced or not, starts with the formation of an identical stem-bulge-loop structure called the Trans Activation Responsive (TAR) element located from nucleotide position +1 to +59. TAR was originally identified as the target for the trans-activator of HIV, the Tat protein, which is essential for efficient transcription of viral genes and for viral replication [44; 45]. The studies that have examined the structure and the sequence of TAR have shown that the stem structure, the sequence in the bulge and in the loop, as well as the distance between the bulge and the loop are all required for Tat-mediated trans-activation [9; 44; 46; 47; 48; 49]. The influence of mutations, especially in the bulge and loop region of the TAR RNA has emphasized its functional importance in the context of HIV replication. It has been observed that viral kinetics is affected when mutations causing changes in structure and sequence are inflicted upon the native TAR RNA [50]. It is known that there are two varieties of TAR RNA, each one arising from one of the two variants of HIV. TAR-1 differs from TAR-2 only in the presence of an extra nucleotide, C24 in the bulge of TAR-1 while this is absent in the sequence of TAR-2 RNA. Fig 1.4 shows a comparison between TAR-1 and TAR-2 of HIV-1 and HIV-2 respectively.

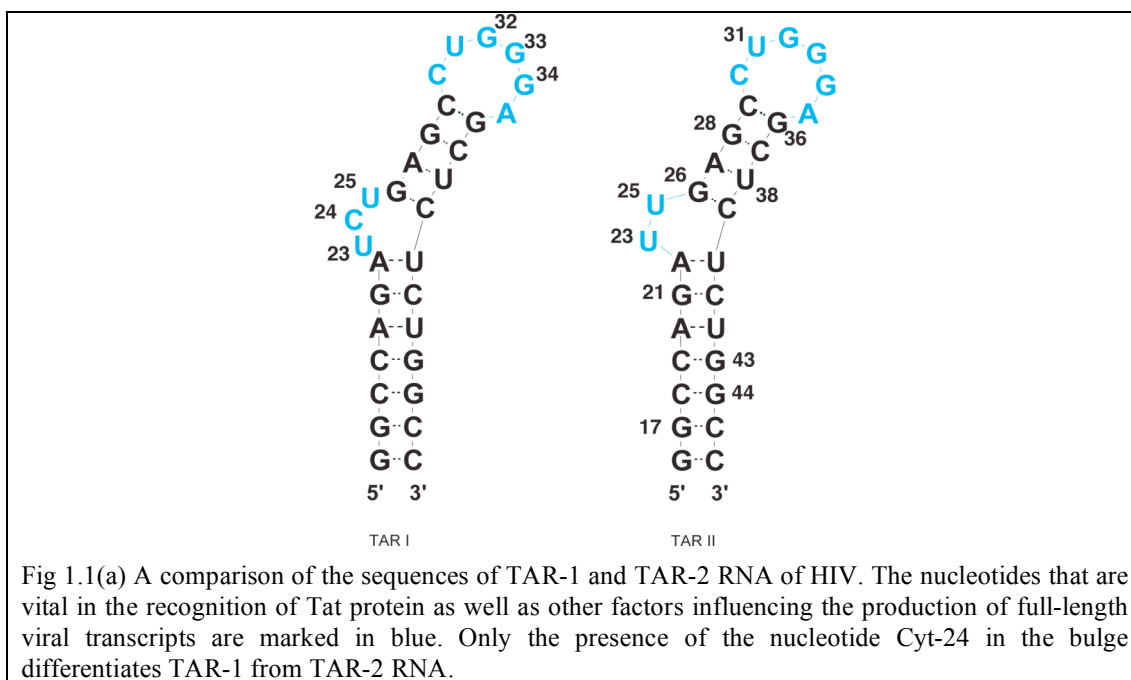
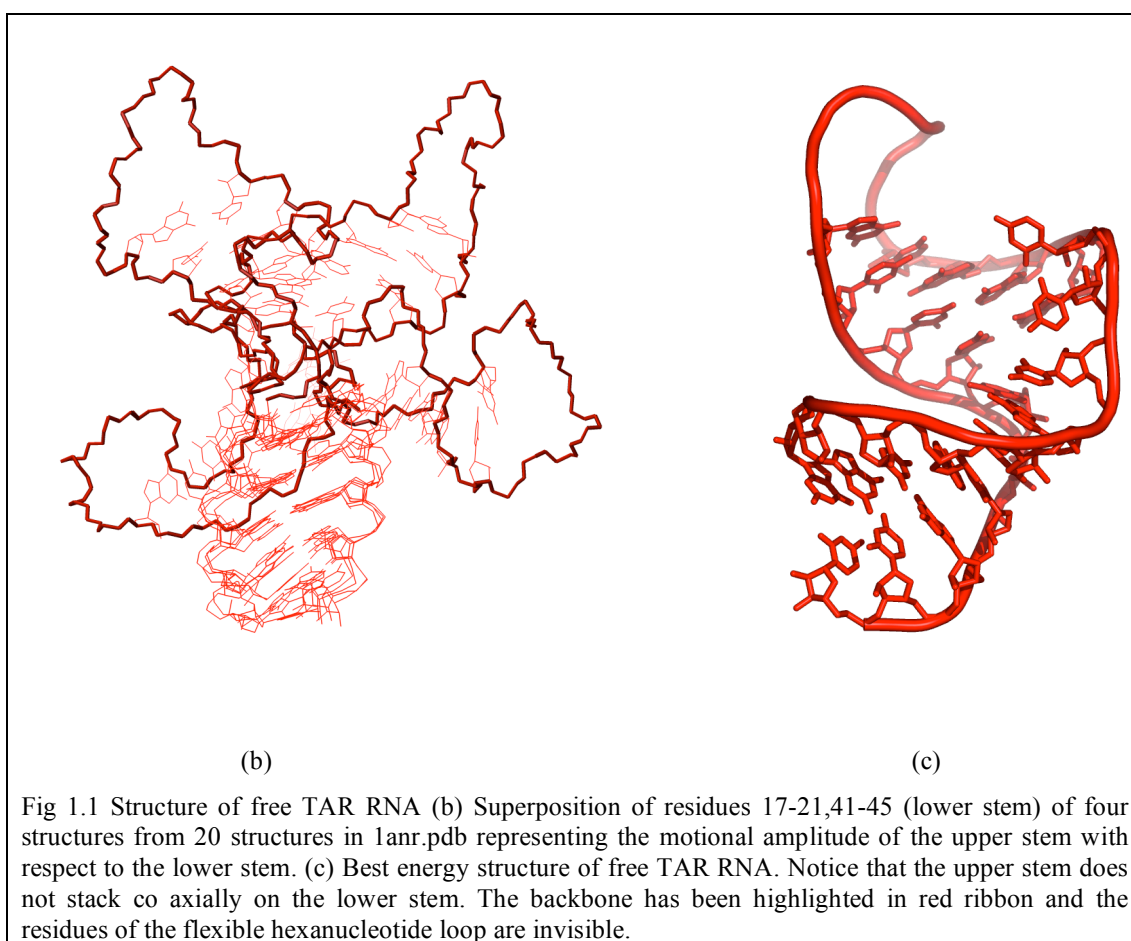


Fig 1.1(a) A comparison of the sequences of TAR-1 and TAR-2 RNA of HIV. The nucleotides that are vital in the recognition of Tat protein as well as other factors influencing the production of full-length viral transcripts are marked in blue. Only the presence of the nucleotide Cyt-24 in the bulge differentiates TAR-1 from TAR-2 RNA.

The structure of the free TAR RNA has been well characterized by NMR spectroscopy. In summary, the TAR RNA consists of a lower stem region spanning residues 16 to 21 and 41 to 46, an upper stem region spanning residues 26 to 29 and 36 to 39, a Uridine rich bulge region that connects the upper stem to the lower (residues 22 to 25) and finally a disordered and dynamic hexanucleotide loop atop the upper stem region. The NMR study [³¹; ⁵¹] reveals that both the stems adopt a conformation close to that of an A-form helix. The bulge residues U23 and C24 (HIV-1 TAR RNA) remain stacked within the helix while U25 is looped out, resulting in a distortion of the phosphate backbone between C24 and G26. Overall, this results in an accessible and wide major groove making it ready for binding to the important residues of the Tat protein. The residues of the hexanucleotide loop are disordered in the NMR structure of the free TAR RNA. The use of relaxation rates and residual dipolar couplings has further contributed towards the understanding of structure and dynamics of the RNA.



Al-Hashimi et al. [⁵²; ⁵³; ⁵⁴] have extensively analyzed residual dipolar couplings in partially oriented systems and concluded that inter helical amplitude of $46 \pm 4^\circ$ is

possible for the TAR RNA in its ground state. A pictorial representation of the motional amplitude in free TAR RNA is shown in Fig 1.1(b) where the lower stems of 4 structures from 1ANR.pdb [31] have been superimposed. The resulting divergence in the upper helix relative to the lower helix is generally attributed to the flexibility present in the bulge residues 22 to 25 and 40.

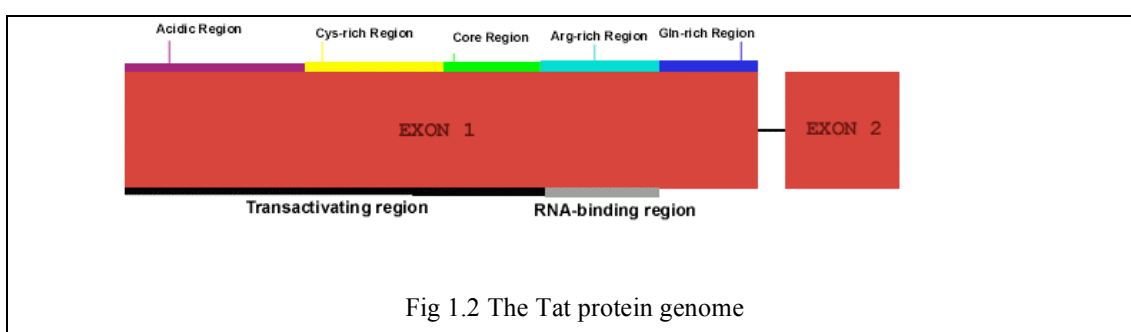
The TAR RNA has been extensively studied in the context of Tat trans-activation, but recent data have shed light on its role in translation and that it is a target for a number of cellular RNA binding proteins that influence the rate of HIV translation (details in Table1). Considering its vital role in the virus life cycle, the HIV TAR RNA has been the target of drug development that has led to small molecules and RNA-based strategies for gene therapy.

	Viral Function	Cellular Function
Tat	HIV trans-activator, binds TAR RNA	
hCycT1	Binds Tat and increases its affinity for TAR RNA	Cyclin
CDK9	Binds CycT1 and phosphorylates RNAP II CTD	Kinase, Phosphorylates RNAP II CTD
P300/CBP	Acetylates Tat at K50 and K51	Histone Acetyl Transferase
PCAF	Acetylates Tat at K28	Histone Acetyl Transferase
hGCN5	Acetylates Tat at K50 and K51	Histone Acetyl Transferase
PKR	Binds TAR RNA, inhibits translation	Kinase, inhibits translation, controls cell growth
TRBP	Binds TAR RNA and enhances translation	Inhibits PKR, increases cell growth, controls spermatogenesis
La autoantigen	Binds TAR RNA and enhances translation	RNAP III transcription

Table 1. Viral and Cellular Proteins Involved in HIV TAR RNA functions [44]

1.2 The Tat protein

Lentiviral Tat proteins are small (~101 amino acid long), arginine rich RNA binding proteins encoded by two exons. The amino acids encoded by the first exons are both necessary and functional for TAR RNA binding and *in vivo* transcriptional activation [28; 40]. Though an 86-amino-acid shortened version of Tat protein is found in the genome of a few laboratory passaged HIV's, naturally occurring HIV's have a 101-amino-acid Tat protein. An atypical transcriptional activator, Tat protein binds not to DNA but to the specific promoter proximal nascent leader the TAR RNA [28; 42; 45; 55; 56; 57]. The Tat protein is composed of several functional regions. A cysteine-rich region (amino-acids 22-37) contains seven cysteine residues; a “core” sequence (amino acids 37-48) contains six arginines and two lysines, and a glutamine rich region at the carboxyl terminus of the first exon. Only the basic and core regions are conserved in lentiviruses and although the integrity of the cysteine rich region is essential for transactivation, this region does not appear to be directly involved in TAR RNA recognition. Based on mutational analysis, Tat can be divided into two domains. The first domain is the activation domain (amino acids 1-47) or co-factor binding domain, which is functionally autonomous and is active when recruited to the HIV long terminal repeat (LTR) via heterologous RNA-binding protein. The second functional domain contains the basic region required for both RNA binding and nuclear localization activities of Tat [58].



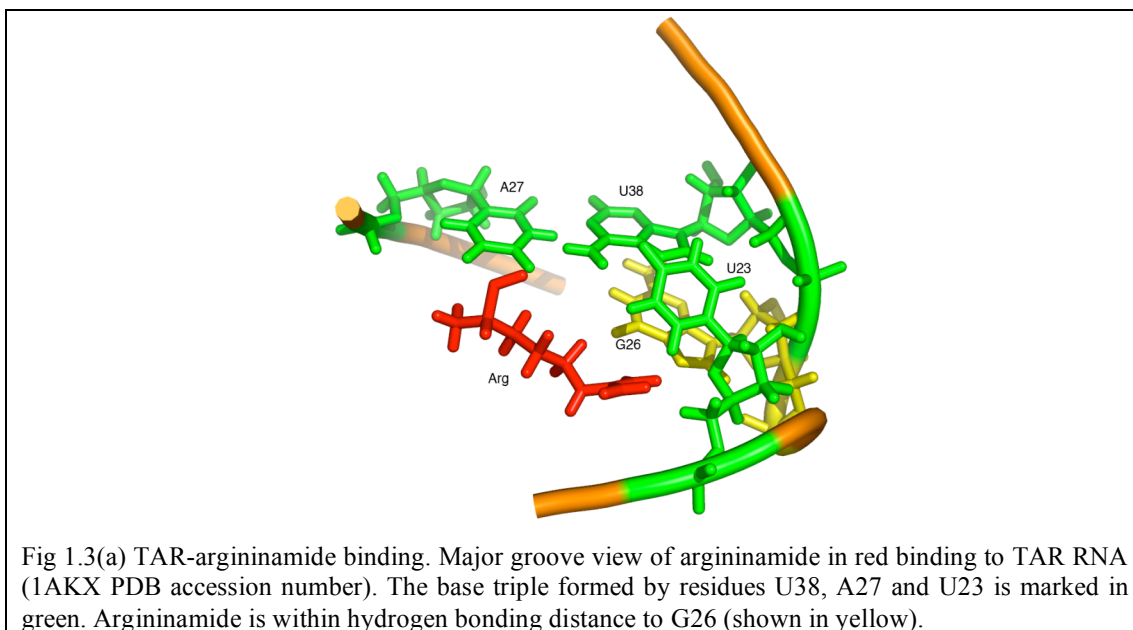
1.3 The Tat-TAR complex

The TAR RNA was originally localized to nucleotides +1 to +80 within the viral LTR but subsequent deletion studies have established that the region from +19 to +42 incorporates the minimal domain that is both necessary and sufficient for *in vivo* Tat protein responsiveness. As described in the previous section, TAR RNA consists of a hexanucleotide loop, an upper stem and a lower stem close to an A-form helix separated by a three (or two) nucleotide pyrimidine bulge.

Interactions between Tat protein and cellular co-factor CyclinT1 are necessary for the recognition of TAR RNA. Fluorescence Energy Transfer studies using dye labeled Tat and TAR shows that the affinity of Tat protein to TAR RNA is increased when the human CyclinT1 interacts with the activation domain of Tat protein [59]. Biochemical and genetic studies show that Tat and human CyclinT1 interaction requires an essential cysteine (C261 on the CyclinT1 protein) and zinc [60], indicating that these proteins may form a metal-dependant heterodimer [61]. Structural studies involving NMR, site specific photo cross linking, site specific modification and affinity cleavage analysis have helped to understand the interaction in the Tat-TAR complex and the ternary CyclinT1-Tat-TAR complex. The following points summarize the salient features of Tat-TAR interaction.

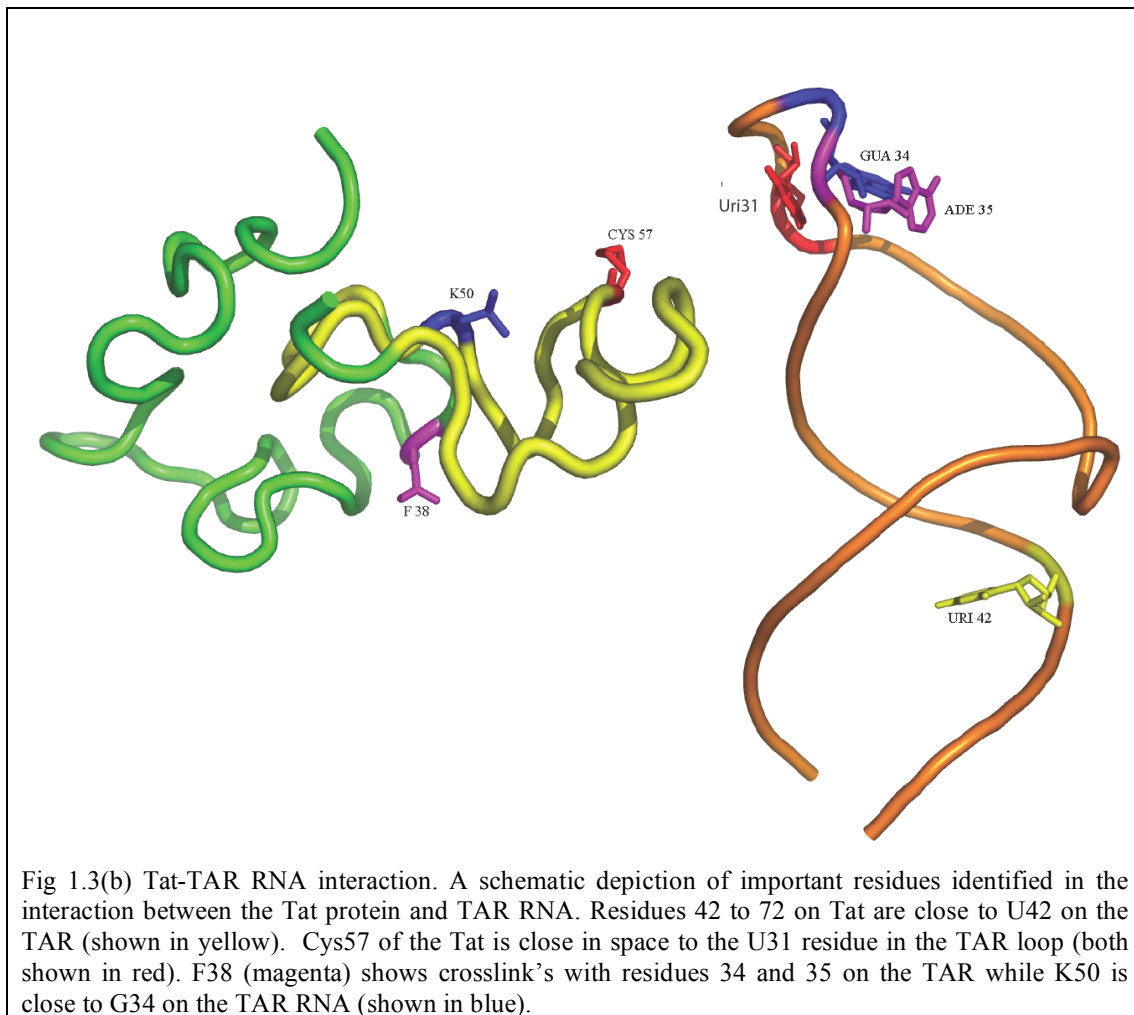
1) The TAR RNA element forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function. Site-specific modification reveals that argininamide can act as an effective Tat mimetic and may reflect on the mechanism that allows Tat protein to specifically bind to the TAR RNA. NMR investigation reveals that in the TAR–argininamide complex, the guanidium group of argininamide forms hydrogen bonds to the major groove face of G26 and is stacked between U23 and A22. This is accompanied by the formation of a base triple between U38, A27 and U23 [46; 62].

2) Unlike the free TAR RNA that exhibits dynamics of the upper stem with respect to the lower stem as depicted in Fig 1.1 (b), in the argininamide bound form of the TAR RNA, the upper stem co axially stacks on top of the lower stem.



3) Site specific cross-linking and affinity cleaving analysis suggests that in the bound form the residues 42 to 72 of Tat are close to the U42 of TAR RNA [63]. Further, Cys 57 of Tat is close to the residue U31 in the loop of TAR RNA. Affinity cleaving experiments also show that F38 of the Tat protein is located in the proximity of nucleotide 34 and 35 of the TAR RNA. Also, the residue K50 of the Tat protein has been found to cross-link with the TAR nucleotide 34 and the binding of CyclinT1 to the Tat/TAR complex considerably enhances this interaction [177]. Fig 1.3 (b) schematically depicts the contacts between residues on the Tat protein and TAR RNA. The formation of the Tat-TAR complex is important because of the following reasons:

- a) Tat protein promotes formation of an activated transcription elongation complex. Transcription reactions performed using immobilized template DNA carrying wild-type TAR RNA elements in the absence and presence of recombinant Tat protein revealed that the processivity of transcription complexes prepared in the presence of Tat is significantly greater than that of complexes prepared in the absence of Tat [64; 65].
- b) Activation of the transcription complex requires recruitment of Tat by a functional TAR RNA element. Tat activated transcription is strictly dependent upon the presence of a functional TAR RNA element [66].



c) Release of TAR RNA from the transcription complex.

The binding of Tat protein is known to induce a conformational change in the TAR RNA. [22; 49; 67] and this in turn may create a recognition site for a cellular co-factor recognizing the TAR apical loop. Karn *et al.* [64] propose that loop recognition factors could help dissociate Tat from TAR RNA, displacing the 5' end of the nascent chain and permitting Tat to bind tightly to RNA polymerase II [68]. Thus, the apical loop is an integral part of the process of initial recruitment of Tat protein and subsequent release of the Tat protein to aid in the formation of a Tat-RNA polymerase II complex.

1.4 Positive Transcription Elongation Factor b

The positive transcription elongation factor b (P-TEFb), a general elongation factor was first identified and purified from *Drosophila* extracts. It acts to prevent RNA polymerase arrest and contains an associated kinase activity capable of hyperphosphorylating the C-terminal domain (CTD) of Rpol II [69; 70; 71]. P-TEFb is composed of two subunits: the catalytic subunit Cyclin-dependent kinase (CDK9) and the regulatory subunit CyclinT1 [59; 72; 73]. Complexes containing CDK9 and CyclinT1-related proteins, CyclinT2a or CyclinT2b, are also active for P-TEFb function [74]. Tat interacts with CyclinT1 subunit of P-TEFb and recruits the kinase complex to the TAR RNA. Recruitment of P-TEFb to TAR has been proposed to be both necessary and sufficient for activation of transcription elongation from the HIV-1 LTR promoter [27; 75]. The first 272 residues of human CyclinT1 are sufficient to bind Tat and TAR *in vitro* and support Tat transactivation *in vivo* [60].

1.5 The Tat-TAR-CyclinT1 complex interactions

The formation of the Tat-TAR-CyclinT1 complex is a highly concerted and cooperative event. Rana *et. al.* [76] investigated the interaction of the TAR RNA and Tat protein in the absence and presence of CyclinT1 (with Zn^{2+}) by monitoring the distance and efficiency of energy transfer in a complex formed between chemically synthesized TAR-Fluorescein and Tat-Rhodamine. The efficiency of energy transfer between the Tat protein and TAR RNA was reduced in the presence of CyclinT1 thus suggesting that the Tat protein may go through a structural reorganization upon CyclinT1 binding. Further, in the absence of Zn^{2+} , no energy transfer was detected thus pointing towards the fact that interactions may be via a metal linked heterodimer between the Tat protein and CyclinT1.

The Tat protein binds TAR RNA with a K_D of 8.2 nM, and this affinity is enhanced ten-fold in the presence of CyclinT1. Additionally, a Tat peptide containing only the arginine rich motif binds TAR RNA with a K_D of 1nM. This increase in binding affinity and the structural rearrangement as suggested by FRET data discussed above help conclude that CyclinT1 enhances the affinity and stability of the Tat-TAR

complex. The TAR RNA loop is of particular significance though the mechanism by which Tat-TAR complex favors the interaction of CyclinT1 with the hexanucleotide loop of the TAR RNA is not known. Photo crosslinking experiments demonstrate that CyclinT1 directly interacts with TAR loop residues 31 and 33 and the U31 side of the loop is the major interaction site. As pointed out in section 1.3, the Tat protein also interacts with the residue 34 side of the TAR RNA loop. Thus, it can be concluded that structural reorganization of the TAR RNA loop may play an important role in the formation and stabilizing of the Tat-TAR-CyclinT1 complex.

1.6 Inhibitors of the Tat-TAR interaction

The interaction between Tat and TAR RNA is critical for virus replication. Altering the RNA binding site results in defective viruses and furthermore, the viral replication is strongly inhibited by the over expression of TAR RNA, which acts as a competitive inhibitor of regulatory protein binding. [77]. A Tat agonist may inhibit HIV replication from integrated proviral DNA at an early stage of viral replication. On the other hand, transcriptional regulators, expressed early in the replication cycle (Tat) to control transcription of other viral genes are attractive targets for antiviral agents directed against chronic viral infection [20]. A compound that inhibits the transactivation of the HIV LTR promoter by the viral Tat protein may force the virus to a state of latency. Such compounds may inhibit both acute and chronic HIV infections and they may lead to less resistance. Mutations in the regulators that antagonize the interaction could be potentially lethal to the viruses. Since the viral transcriptional regulators do not have any cellular counterparts, they represent a potentially selective antiviral targets. Furthermore, since the Tat-TAR interaction is highly conserved between virus isolates, anti Tat-TAR drugs are less likely to be affected by HIV variability than other drugs.

The ability to exploit Tat-TAR function for antiviral therapy was first demonstrated by Lisziewicz et al. [78; 79; 80; 81] where a Tat inhibitory gene (anti-Tat) was used to decrease the amount of Tat protein by inhibiting the translation of the Tat mRNA, thus proving its value as an effective long-term suppressor of HIV and also suggesting that anti-Tat gene therapy may be beneficial to block HIV replication and reconstitute the immune system of late phase AIDS patients.

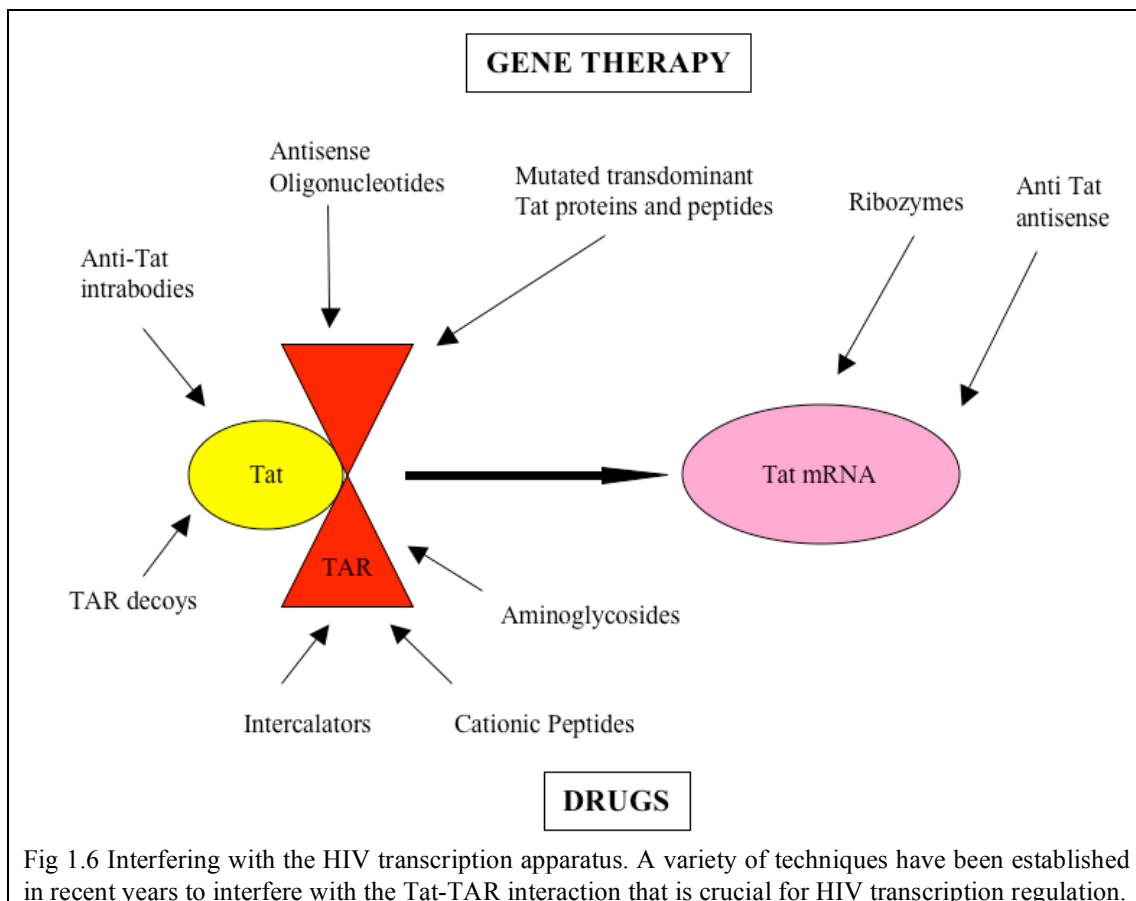


Fig 1.6 Interfering with the HIV transcription apparatus. A variety of techniques have been established in recent years to interfere with the Tat-TAR interaction that is crucial for HIV transcription regulation.

The examples summarized in the following sections demonstrate that inhibiting the Tat-TAR interaction might be an interesting approach to inhibit HIV replication and that the arginine-binding site on the TAR RNA (described previously) could be an important interaction site for potential inhibitors.

1.6.3 Intercalators

A third series of compounds targeting the Tat-TAR interaction are intercalating agents, often combined with cationic groups. Most commonly, an ethidium derivative composed of arginine, a linker and an ethidium is employed to inhibit the Tat-TAR interaction. Molecular modeling studies show that the ethidium seems to fit near an unpaired residue at position 17 while the arginine side chain occupies the pyrimidine bulge at position 23-25 [84; 85; 86].

1.6.4 Aminoglycosides

The aminoglycosides are a clinically important group of antibiotics that have a broad-spectrum of activity and that are bactericidal in action. The family includes streptomycin, gentamicin, neomycin tobramycin, kanamycin, amikacin and netilmicin. Since the discovery of streptomycin, the first aminoglycoside to be isolated by Waksman and co-workers [87; 88]; the role of aminoglycoside antibiotics as a family of RNA binders has been considerably accentuated by reports from Schroeder *et al.* and Green *et al.* who demonstrated the capability of these antibiotics to inhibit splicing in group I introns and interaction between the HIV-1 Rev protein and its RNA target respectively [89; 90]. Consequently, aminoglycosides have become a vital tool in targeting and inhibiting the interaction of Tat protein and TAR RNA in the HIV virus. They also offer a powerful tool to study the structural basis of RNA recognition and inhibition of function by cationic organic molecules in general.

Aminoglycosides have a variety of effects within the bacterial cell but principally they inhibit protein synthesis by binding to the 30S ribosomal subunit to prevent the formation of an initiation complex with messenger RNA. They also cause misreading of the messenger RNA message, leading to the production of nonsense peptides. Another important function of the aminoglycosides is that they increase membrane leakage.

The common core of most aminoglycosides is 2-Deoxystreptamine (2-DOS), a highly functional aminoglycositol. Glycosylation of the 2-DOS core (Figure 1.6.4), typically at the 4- and 5-, or 4- and 6- positions characterizes most aminoglycosides. Examination of binding affinities reveals some rudimentary structure activity

relationships in aminoglycosides. The importance of amino groups for RNA binding and inhibition is apparent.

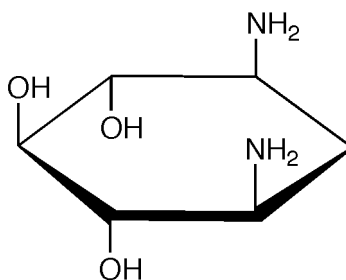


Fig 1.6.4 2- Deoxystreptamine

Also evident from affinity data is that, hydroxyl groups influence RNA affinity. Changing an amino group to a hydroxyl group in kanamycin A abolishes inhibitory activity in self-splicing and Rev-RRE assays [24]. Thus, it can be concluded that electrostatics plays an important role in RNA affinity. Despite apparent disadvantages like unspecific binding, high toxicity and rapid development of bacterial resistance, novel aminoglycosides are a valuable instrument in combating diseases especially related to Gram-positive bacteria, interfering with crucial RNA-protein interaction and in understanding the fundamental characteristics of RNA recognition. The main theme of research undertaken in this thesis is concerned with the interaction of the TAR RNA with a novel macro cyclic 1,4-butanediol-linked aminodeoxyglucoside; henceforth referred to as **AMG** and the preparation of which has been described by Kirschning *et al.* [91].

1.7 Scope of the work and the questions we address

RNA can adopt a plethora of conformations, depending on factors such as pH, temperature and salinity and hence, monitoring and studying the changes in local and global conformations of RNA has become increasingly important to understand principles of RNA recognition and RNA-protein interactions. These changes in conformation are usually triggered by cofactors that are required for biological functions.

As described in previous sections, small positively charged molecules such as polyamines, aminocyclitols and aminoglycosidic scaffolds can specifically recognize the three dimensional motifs of RNA and thus arrest the RNA in a biologically inactive conformation or compete directly with the binding of proteins and cofactors that are essential for biological function. A deeper insight into the principle of recognition of RNA with cationic antibiotics represents a major step towards RNA based drug design.

The Tat protein-TAR RNA complex offers a unique model system to investigate the role of such antibiotics on the interaction of RNA-protein systems that are of vital importance for the proliferation and pathogenicity of viral entities such as HIV. Further, the interaction of Tat protein with CyclinT1 unit of the P-ETFb complex is a key element in the control of HIV viral replication and constitutes attractive targets for anti viral therapy. The structural investigation carried out as part of this thesis work, of a novel aminoglycoside analogue interacting with the TAR RNA, offers unique insights particularly for the hexanucleotide loop of the RNA and how a better understanding of the multiple possibilities of RNA recognition by a cationic antibiotic can lead to the design of improved TAR binders and pave the way for alternative therapeutic approaches that inhibit the interaction of the TAR RNA with the Tat protein and subsequently the P-ETFb complex.

Employing NMR and a variety of biochemical assays to characterize the structure of this novel aminoglycoside (AMG) and the TAR RNA, we aim to shed light on the following key issues:

- 1) Where does the aminoglycoside bind to the TAR RNA?
- 2) What effect does the binding of AMG have on the bulge and the disordered hexanucleotide loop of the TAR RNA?
- 3) Does the binding of AMG to the hexanucleotide loop suggest a possible principle of recognition of CyclinT1 of the same loop?
- 4) What is the basis of RNA recognition by AMG? Do hydrogen bonds and specific contacts to the phosphate backbone of the TAR RNA influence the recognition of TAR RNA by AMG?

2 Methods

2.1 HIV-2 TAR RNA synthesis

The HIV-2 TAR RNA sequence is shown in Figure 2.1. The presence of two Guanine nucleotides at the 5' end opposed by two Cytosine's at the 3' end promotes T7 transcription and prevents aggregation. The RNA was transcribed from DNA templates with T7 RNA polymerase [⁹²] expressed and produced in-house. Labeled ¹⁵N/¹³C NTPs were purchased from Spectra Stable Isotopes. Table 2.1 shows the components of the reaction.

Test reactions to optimize MgCl₂ and NTP concentrations required for the transcription were carried out in 10µl aliquots for both the labeled and unlabeled HIV-2 TAR RNAs. Preparative reactions were subsequently done at the 20ml scale. All reactions were done in disposable Falcon tubes and extreme care was taken to avoid RNase contamination. The components of Table 2.1 were mixed in the appropriate proportions and the reaction was incubated for 5-6 hours at 37°C. An amount of EDTA equivalent to the concentration of MgCl₂ used was employed to arrest the reaction once the transcription was over. The reaction mixture was extracted once with phenol: CHCl₃ and then precipitated overnight with ethanol at -20°C. The precipitant was resuspended in loading buffer (8mM Urea + 0.1% bromophenol blue + 0.1% xylene blue) and separated on denaturing 20% polyacrylamide gels, 40cm x 50cm, run at 50W for ≈ 15 hours till bromophenol blue had migrated most of the gel. The RNA band was visualized by UV shadowing and cut from the gel. Appropriately sized gel slices were electro eluted in an Elutrap apparatus. The resulting elutant was lyophilized and resuspended in appropriate buffer (50mM phosphate buffer and pH 6.4) and the concentration of RNA was estimated by measuring the absorbance at 260 nm.

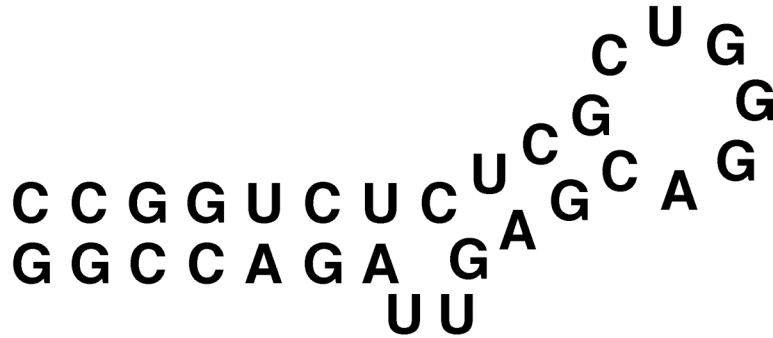


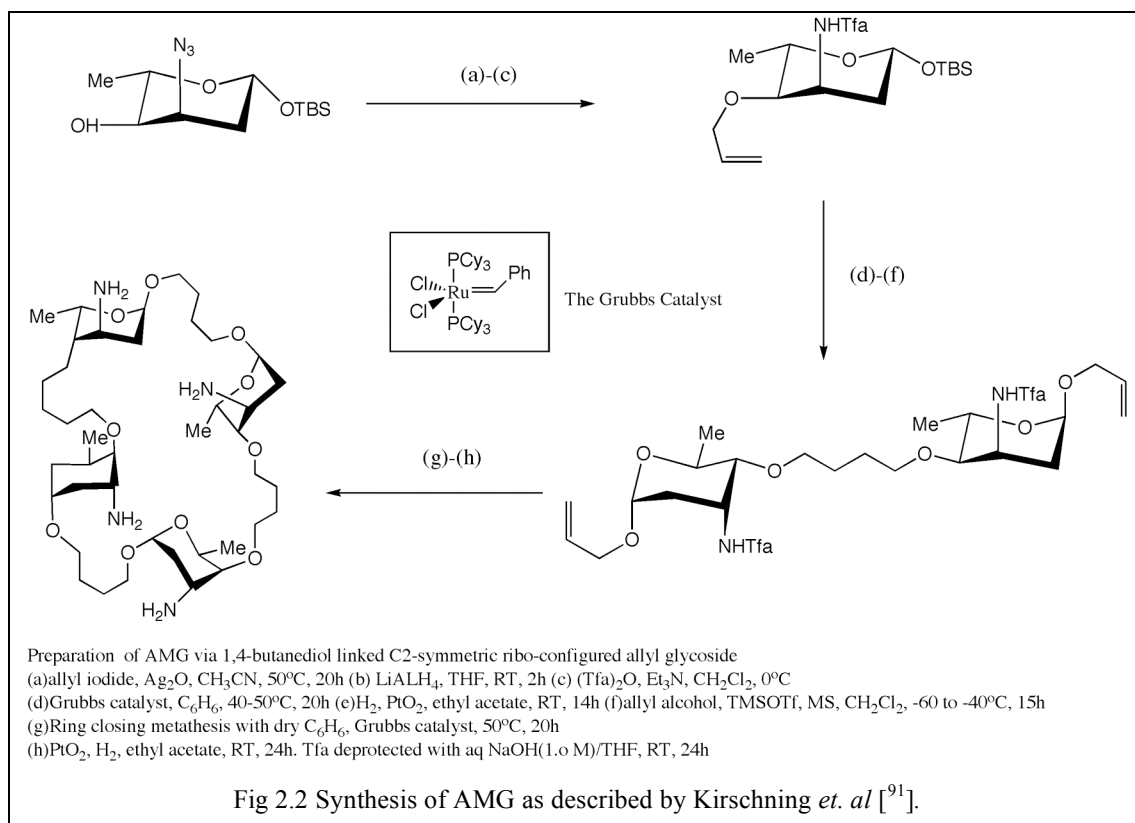
Figure 2.1 The HIV-2 TAR RNA sequence

Component	
NTP (Spectra Stable Isotopes)	15mM (3.75mM each) 20mM (5.00mM each)
MgCl ₂ (Fluka Biochemika)	35mM 30mM
T7 polymerase (produced in-house)	3μl
PEG	0.16μl
DTT (100mM)	0.5μl

Table 2.1 *In-vitro* T7 transcription reaction components for producing labeled and unlabeled HIV-2 TAR RNA. Optimized NTP and MgCl₂ concentrations for unlabeled RNA are shown in red font. The reaction mixture was made up to 10μl by the addition of 10X buffer which consisted of 400mM Tris/HCl at pH 8.0, 10mM Spermidine, 50mM DTT and 0.1% Triton-X 100. Preparative reactions were scaled to 20ml using similar conditions.

2.2 Cyclic Aminoglycoside AMG

A short scheme illustrating the summary of the synthesis of AMG as described by Kirschning *et. al.* [91] is detailed below:



The synthetic scheme of the AMG from a 1,4-butanediol linked C₂ symmetric ribose-configured allyl glycoside is shown above. The ¹⁵N-labeled and unlabeled cyclic amino glycoside (AMG) was obtained from the laboratory of Prof. Dr. Andreas Kirschning, Institute of Organic Chemistry, University of Hannover in lyophilized form. A stock solution of AMG was prepared for titration to monitor chemical shift changes of the TAR RNA residues with increasing concentration of AMG. 5 mg of AMG (molecular mass 805,051 g/mol) was dissolved in 5µl, 50mM sodium phosphate buffer at pH 6.5 and 95µl distilled water. The pH was adjusted to 6.4 - 6.5 by the careful addition of 1M HCl and 1M NaOH solutions. The electrode was allowed to equilibrate after each addition for maximum accuracy. A final stock solution of 49.686 (≅ 50mM), 125µl AMG was obtained.

The important materials used in the course of the production of the samples were affiliated as follows:

Deuterium Oxide D ₂ O	Deutero GmbH (99.9% purity)
Vacuum pumps	Vakkubrand GmbH
Eppendorf cups	Abimed
Lyophilizer Loc-1M Alpha	Christ
Water	Millipore
Heating blocks	neoLab
NMR sample tubes	Shigemi Co. LTD
Vortex Genie	Scientific Industries

Table 2.2 Equipment and suppliers

2.3 RNase footprinting / Digestion

The Ribonuclease T1 (RNase T1) is an endoribonuclease that specifically degrades single-stranded RNA at guanine residues. It cleaves the phosphodiester bond between 3'-guanylic residues and the 5'-OH residues of adjacent nucleotides with the formation of corresponding intermediate 2', 3'-cyclic phosphates [⁹⁷]. The reaction products are 3'-GMP and oligonucleotides with a terminal 3'-GMP. The Ribonuclease T1 does not require metal ions for activity. RNase A is sequence specific for single stranded RNAs. It degrades 3'end of unpaired CYT and URI residues. It cleaves the phosphodiester bond between the 5'-ribose of a nucleotide and the phosphate group attached to the 3'-ribose of an adjacent pyrimidine nucleotide. The resulting 2', 3'-cyclic phosphate is hydrolyzed to the corresponding 3'-nucleoside phosphate [⁹⁸]. This experiment is a sensitive probe of residues that are protected and shielded by the presence of AMG and those that are more solvent exposed and susceptible to cleavage/digestion by the respective RNase enzymes.

In summary, residues that are less shielded by the ligand would be more prone to digestion than those residues that are close to the binding site of the ligand and hence protected from digestion.

10 μ M of 5'-³²P labeled HIV-2 TAR RNA was digested with either GUA selective RNase T or URI and CYT selective RNase A [⁹⁹]. The reaction was performed in autoclaved Eppendorf reaction vessels at 4°C. The concentration of AMG was varied from 0.1 μ M to 10mM. Appropriate controls were performed and extreme precautions were taken to avoid RNase contamination. The reaction was stopped by the addition of 2mM EDTA and 0.1% SDS and the reaction mixture was loaded onto a 20% denaturing polyacrylamide gel. The radioactive complexes were visualized by autoradiography.

2.4 PACE (PolyAcrylamide Co Electrophoresis)

The PACE experiment was originally designed to study RNA-protein (peptides) interactions, particularly those complexes with low affinities but this technique can be conveniently extended to investigate the binding of non-protein based ligands to RNA systems. We prepared PACE gels as described by Cilley and Williamson [100]. 10 μ M of 5'-³²P labeled HIV-2 TAR RNA and a series of lanes containing different concentrations of AMG (0, 1, 2, 5, 7, 10, 20, 70, 100, 200 and 500 μ M) were poured with the native gel (15% acryl amide, 1X TBE and 0.02% ammonium persulphate) rotated by 90⁰.

The gels were run and subsequently visualized by autoradiography and analyzed.

We adapted the PACE experiment to investigate the binding stoichiometry between the TAR RNA and AMG [101]. The principle behind this technique warrants a deeper discussion here.

In the analysis of the PACE gel, the RNA is assumed to exist in two distinct states, one is the free unbound state with mobility equal to that of the RNA in the absence of AMG. The mobility of the bound form is a function of the size, shape and charge of the TAR RNA/ AMG complex formed under the given concentration of AMG. Thus, the maximal retardation observed would be system dependent. Also, it is assumed that the bound form and the free form of the TAR RNA are in rapid equilibrium with each other i.e. in the fast exchange regime relative to the electrophoresis time.

The total distance traveled (D) for each TAR RNA under the influence of each AMG concentration considered is measured from the autoradiogram. As a control, the migration distance of TAR RNA in the absence of AMG (D_F) is also measured to allow comparison of migration distances of different lanes.

Assuming that a simple binding equilibrium exists between the TAR RNA and AMG leads to



Where $K_d = [\text{TAR}] [\text{AMG}] / [\text{TAR/AMG}]$, [TAR] is the concentration of TAR RNA and [AMG] is the concentration of the amino glycoside. Further, the fraction of TAR RNA bound to AMG (ϕ) is given by

$$\phi = [\text{TAR}]_{\text{Bound}} / [\text{TAR}]_{\text{Total}} \quad (\text{ii})$$

The binding constant between the TAR RNA and AMG is only moderate; hence it is safe to assume that the concentration of free AMG is equal to the concentration of total AMG. This leads to

$$\phi = [\text{AMG}]_{\text{Total}} / \{K_{d, \text{app}} + [\text{AMG}]_{\text{Total}}\} \quad (\text{iii})$$

Where $K_{d, \text{app}}$ is the apparent dissociation constant for the TAR/AMG complex under PACE gel conditions. In the absence of AMG, the TAR RNA will travel a distance D_F that corresponds to the migration of the free TAR RNA. When the total distance traveled, D equals D_F then we can conclude that there is no significant interaction between the TAR RNA and AMG complex. A sufficiently high concentration of AMG will lead to maximal retardation of the TAR RNA and this distance D_B is indicative of the migration of the fully bound form of the TAR RNA / AMG complex. The fraction of TAR RNA bound at any given AMG concentration is given by

$$\phi = \{D - D_F\} / \{D_B - D_F\} \quad (\text{iv})$$

Where D is the retardation distance in each lane, D_F is the retardation distance of the TAR RNA in the absence of AMG and D_B is the retardation distance of the TAR RNA completely bound to AMG (in our case at a concentration of 500 μ M).

Continuing the analysis, substituting (iii) in (iv), we get

$$D = \{[\text{AMG}]_{\text{Total}} (D - D_F) / K_{d, \text{app}} + [\text{AMG}]_{\text{Total}}\} + D_F$$

$$\text{Leading to } \boxed{(D - D_F) / (D_B - D_F) = [\text{AMG}] / K_{d, \text{app}} + [\text{AMG}]}$$

For our investigation of binding stoichiometry in the TAR RNA/AMG complex, we assumed different binding models with different stoichiometries of the TAR RNA in complex with AMG and the theoretical dependence of ϕ on the concentration of AMG was calculated in each case from the molar fractions assuming constant concentration of AMG in each lane. The experimental data were fitted to the theoretical function using MATLAB.

The following six different models were considered:

- (i) One molecule of the TAR RNA binds to one molecule of AMG
- (ii) One molecule of the TAR RNA binds to two molecules of AMG with the same K_d
- (iii) Two molecules of the TAR RNA bind to one molecule of AMG.
- (iv) One molecule of TAR RNA binds to two molecules of AMG with different K_d
- (v) One molecule of the TAR RNA binds one molecule of AMG with K_{d1} followed by dimerization of TAR RNA/AMG complex with K_{d2} .
- (vi) Two molecules of the TAR RNA bind one molecule of AMG (K_{d1}). Subsequently two molecules of AMG are recruited by the RNA dimer (K_{d2}) for a total binding of two TAR RNA molecules and three AMG molecules.

The values of D , D_B and D_F is elucidated from the PACE gel and the concentration of AMG is specific to each lane, hence the apparent K_d can be calculated by plotting $(D-D_F)/(D_B-D_F)$ vs. $[AMG]$. Thus, the theoretical dependence of $(D-D_F)/(D_B-D_F)$ on the concentration of AMG can be predicted for each of the above mentioned model cases assuming that the concentration of AMG remains constant in each lane.

2.5 NMR spectroscopy

The labeled and unlabeled HIV-2 TAR RNA produced by T7 transcription was dialyzed extensively against NMR buffer [⁹²].

For the titration of the TAR RNA with AMG, 2.317mg of unlabelled HIV-2 TAR RNA was dissolved in 250µl of H₂O and 20µl of D₂O. AMG was titrated in varying concentrations steps of molar ratio (TAR RNA: AMG) 1:0, 1:1, 1:2, 1:5, 1:10, 1:20, 1:40, 1:50, 1:70 and 1:100. The effect was monitored on the imino region of the TAR RNA employing a 1D Watergate experiment [¹⁰²] on the 700 MHz Avance spectrometer at 298K. For the chemical shift mapping, AMG was titrated in concentrations varying from 0.3mM to 30mM against a sample of 0.3mM sample of ¹⁵N/¹³C-labeled HIV-2 TAR RNA, dissolved in H₂O /D₂O, containing 50mM phosphate buffer at pH 6.4.

The chemical shift changes of both the base (C5-H5, C6-H6, C8-H8) and the sugar (C1'-H1') resonances of the RNA were monitored in ¹³C-HSQC spectra [¹⁰³] during the titrations on a 600 MHz Avance spectrometer at 298K. Each spectrum, on the base and sugar was acquired for a total of 5 hours with the acquisition dimensions being 128 points (¹³C) and 2048 points (¹H). For the final experiments 0.772 mg of ¹⁵N/¹³C-labeled HIV-2 TAR RNA (0.3mM) and 0.724 mg of AMG (3mM) were dissolved in either D₂O or H₂O/ D₂O (90/10%) containing 50mM phosphate buffer at pH 6.4. All experiments were acquired in-house on Bruker 600 DRX, 600 Avance, 700 Avance, 800 DRX and 900 Avance spectrometers. All spectrometers except the 600 DRX were equipped with a cryoprobe. The carrier frequency was set to 4.7 ppm (H₂O/ HDO) for proton and 148 or 79 ppm for ¹³C depending on the chemical shifts to be observed. All spectra were collected at 298K unless otherwise specified. Spectra acquired were transferred and processed using XwinNMR (Bruker, Karlsruhe, Germany) or Felix (MSI).

The COSY step is used to identify coupled spin systems. The J_{H-C} coupling evolves for a period of $2\Delta_1$. Δ_1 is set to $1/8J_{CH}$ where J_{CH} is the one bond J coupling constant of 1H - ^{13}C (≈ 160 Hz). As a consequence, in phase terms of both CH (namely $C1'/H1'$, $C2'/H2'$, $C3'/H3'$ and $C4'/H4'$) and CH_2 (in RNA only $C5'/H5'$ and $H5''$) are refocused. Importantly, during the COSY step, the $J_{C'-C''}$ couplings evolve for a period of $2\Delta_3$. Here magnetization is transferred from C' to the neighboring C'' yielding terms of the form:

$$C'_x \cos(\omega_H t_1) \cos(2\Delta_3 \pi J_{C-C}) \text{ and } C'_y C''_z \cos(\omega_H t_1) \sin(2\Delta_3 \pi J_{C-C})$$

During the CT (Constant Time), the chemical shift of carbon ($\omega_{C'}$ and $\omega_{C''}$) is modulated by the time period t_2 while the J_{CC} is modulated by the time period $2T_C$.

The detectable terms after the CT are

$$C'_x \cos(\omega_H t_1) \cos^n(2\Delta_3 \pi J_{C-C}) \cos(\omega_{C'} t_2) \cos^n(2T_C \pi J_{C-C}) \dots (a)$$

$$C''_x \cos(\omega_H t_1) \cos^{n-1}(2\Delta_3 \pi J_{C-C}) \sin(2\Delta_3 \pi J_{C-C}) \cos(\omega_{C'} t_2) \dots (b)$$

$$\cos^{m-1}(2T_C \pi J_{C-C}) \sin(2T_C \pi J_{C-C})$$

Where n is the number of carbons attached to C' and m is the number of carbons attached to C'' . In case of RNA, it is to be borne in mind that “ m ” and “ n ” can take values of either 1 or 2 depending on the number of neighbors the C atom has. $C1'$ and $C5'$ have only one neighboring C atom namely $C2'$ and $C4'$ respectively, while $C2'$, $C3'$ and $C4'$ have 2 neighbors each. From the above two terms (a) and (b) we can conclude that the proton (H') is correlated to the carbon it is directly attached to (C') and the carbon two bonds away (C''). This magnetization is then transferred to all the attached carbons within the ribose spin system in the TOCSY step by a DIPSI-3 mixing scheme [113]. The reverse inept after the TOCSY step transfers the magnetization from C'_x to H'_x leading to detection on the proton dimension.

In the HCCH-TOCSY experiment, all the 1H or ^{13}C in the ribose system are correlated through the ^{13}C -mixing scheme whereas in the HCCH-COSY experiment only the two neighboring 1H , 1H or ^{13}C , ^{13}C are correlated. Contrarily, the 3D HCCH-COSY TOCSY experiment combines both the transfer schemes into one experiment and offers a powerful technique to resolve whole 1H - ^{13}C COSY patterns for the assignment of poorly dispersed RNA ribose spins.

Further resonance assignment of fully $^{15}N/^{13}C$ labeled RNA is mainly based on a set of heteronuclear through-bond transfer experiments correlating the nuclear spins within the same sugar, the same base, or between the sugar and base of the same

nucleotide [¹⁰⁴; ¹¹⁴]. For large RNAs (30 nucleotides and higher), especially at lower sample concentration, the inter-nucleotide through-bond HCP correlation experiment [¹¹⁵] becomes very inefficient and thus, the ¹³C edited 3D NOESY becomes very important to detect ¹H-¹H NOEs between the sugar protons H1', H2' and the aromatic base protons, H6 or H8 of sequential nucleotides.

A 3D ¹³C-HSQC NOESY [¹¹⁶] (mixing time of 100 ms, acquisition dimensions of 2048 points on F3/¹H, 88 points on F2/¹H and 168 points on F1/¹³C, with 8 scans at 298K on 800MHz) and a base resolved/selective 3D ¹³C-HSQC NOESY [¹¹⁷] (mixing time of 100 ms, acquisition dimensions of 2048 points on F3/¹H, 96 points on F2/¹H and 256 points on F1/¹³C, with 8 scans at 298K on 800MHz) was used for sequential resonance assignment of the TAR RNA. The pulse sequence of the base resolved/selective 3D ¹³C-HSQC NOESY is shown in Fig 2.5.1 (b) and the effect of the pulses on the RNA can be analyzed as follows:

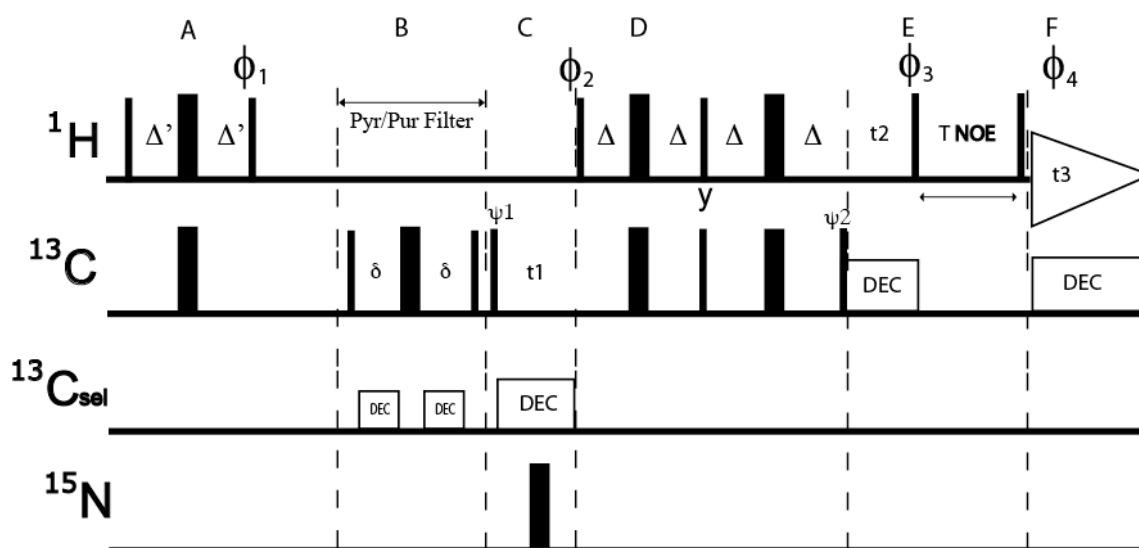


Figure 2.5.1 (b). The 3D Base Selective NOESY pulse scheme for the sequential resonance assignment of nucleic acids [¹¹⁷].

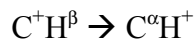
The experiment begins with an initial INEPT step at A, which allows the transfer of ¹H magnetization into two-spin order.

The step at B functions as a base type selective C-C filter sequence to separate the resonances of C8 in purines and C6 in pyrimidines where the filter delay δ was set to $1/2J_{C5C6}$.

This is followed by the carbon evolution period (t_1) when the ¹³C spins are frequency labeled while decoupling homonuclear C-C scalar coupling. No decoupling is applied

on the ^1H dimension so that H^α and H^β states are not interconverted. Applying an 180° pulse on the ^{15}N dimension removes scalar coupling evolution between base carbons and nitrogen at time $t_1/2$.

The block D performs spin-state-selective coherence transfer from ^{13}C to ^1H



During t_2 in block E, the ^1H spins are frequency labeled followed by a NOESY mixing time of 100ms. The carbon spins are decoupled during t_2 .

The ^1H FID is detected while simultaneously decoupling ^{13}C during F.

This experiment helped correlate H6/H8-H1' and H6/H8-H2' connectivity's on the TAR RNA residues.

To measure hydrogen bond mediated J_{NP} couplings, ^{15}N labeled AMG was mixed with unlabeled TAR RNA. This was done to exclude or confirm the presence of hydrogen bonding between the amino groups of AMG and the phosphate backbone of the TAR RNA. The J-quantitative method [118] was employed to measure the J_{NP} couplings in an H-N 2D correlation. The ^{15}N chemical shift of the amino group confirmed that all four groups of AMG are protonated in the NMR buffer. Since the ligand is in fast exchange with the solvent, the resonances of the amino protons could not be observed.

2.5.2 Resonance assignment of the Aminoglycoside

To obtain the resonance assignment of the ligand AMG [119], we relied on the following experiments measured on the 600 MHz Avance.

- (a) **DQF-COSY**: 512 points on F1 and 2048 points on F2 with 108 scans
- (b) **TOCSY**: 512 points on F1 and 2048 points on F2, mixing time 20 ms, 48 scans
- (c) **TOCSY**: 512 points on F1 and 2048 points on F2, mixing time 60 ms, 48 scans
- (d) **NOESY**: 512 points on F1 and 2048 points on F2, mixing time 130 ms, 80 scans
- (e) **NOESY**: 384 points on F1 and 2048 points on F2, mixing time 60 ms, 56 scans

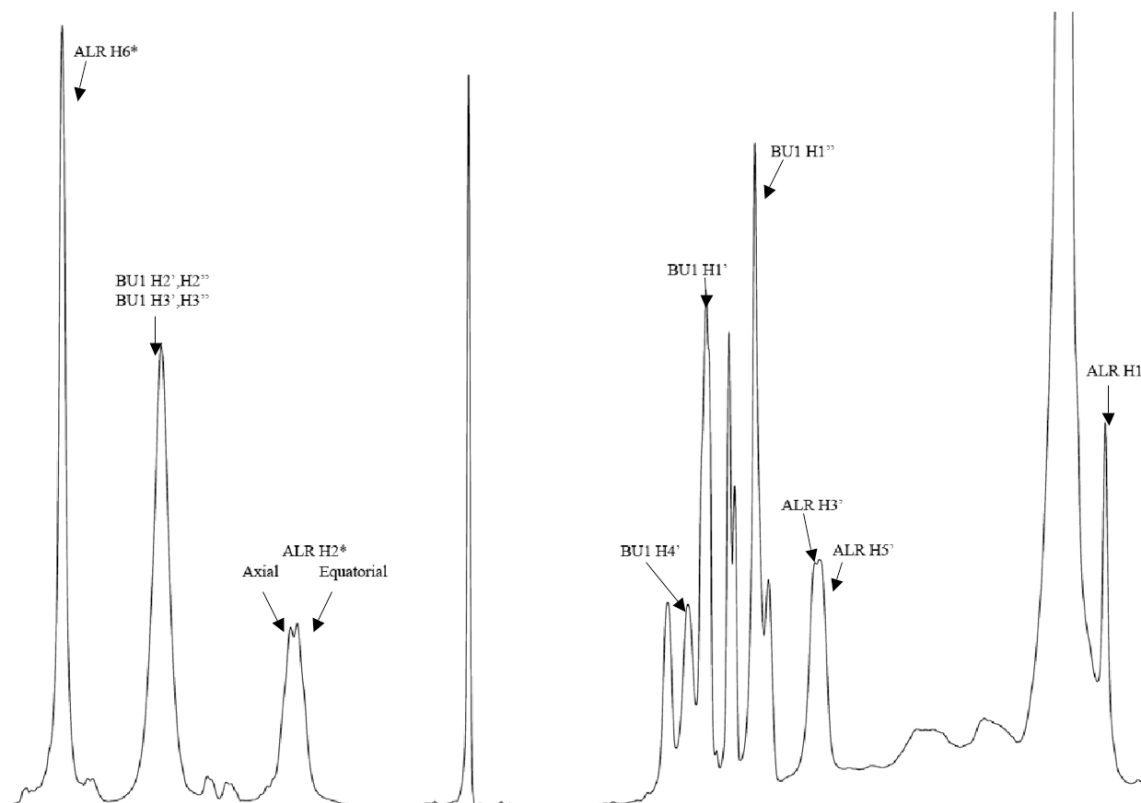


Fig2.5.2 The 1D region of AMG.

2.5.3 Titrations and chemical shift mapping

An easy and rapid way to gain qualitative information about the interaction between two molecules is to study the chemical shift values. Chemical shifts depend on the local environment. The chemical shifts of ^1H and ^{13}C are particularly sensitive to changes of the environment. Therefore ^1H and ^{13}C chemical shift perturbations recorded via ^{13}C -HSQC spectra throughout an NMR titration are commonly combined in order to map a binding site of the ligand on an RNA. The residues exhibiting the largest ^1H and ^{13}C chemical shift changes upon titration of the ligand define the binding interface.

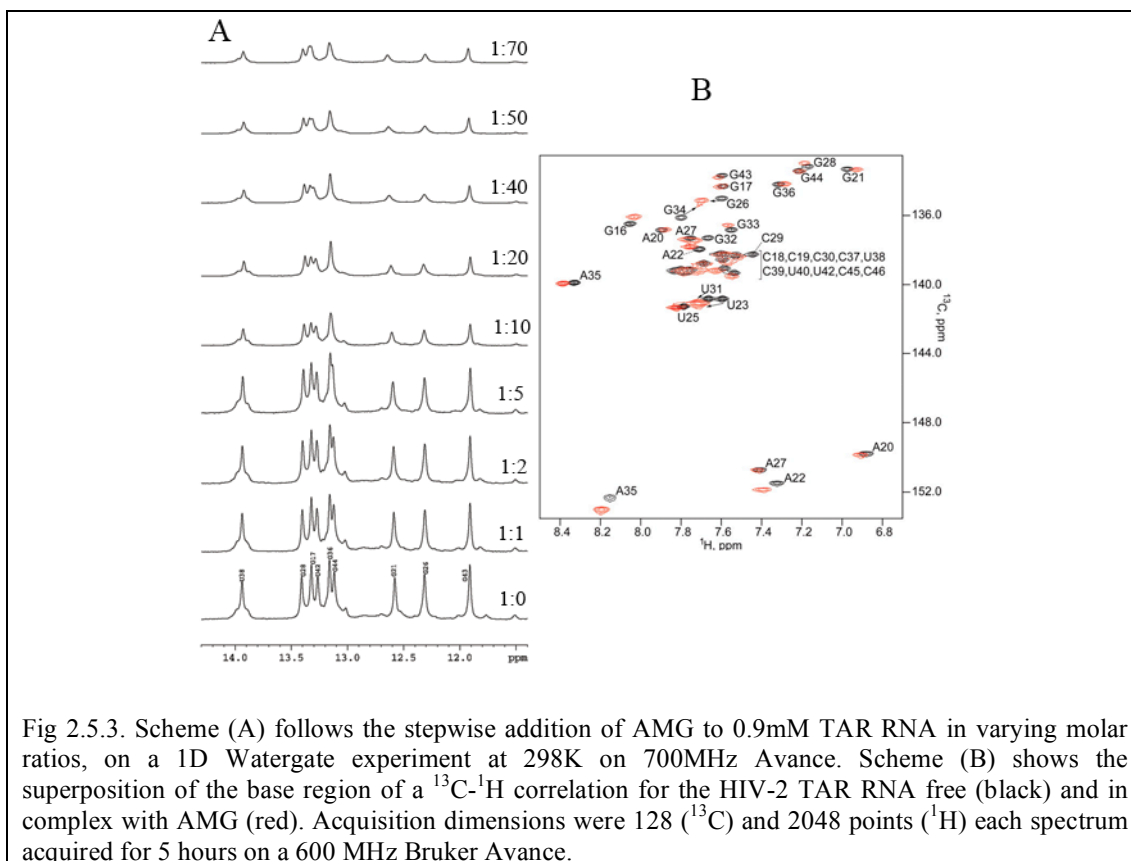


Fig 2.5.3. Scheme (A) follows the stepwise addition of AMG to 0.9mM TAR RNA in varying molar ratios, on a 1D Watergate experiment at 298K on 700MHz Avance. Scheme (B) shows the superposition of the base region of a ^{13}C - ^1H correlation for the HIV-2 TAR RNA free (black) and in complex with AMG (red). Acquisition dimensions were 128 (^{13}C) and 2048 points (^1H) each spectrum acquired for 5 hours on a 600 MHz Bruker Avance.

As described in Section 2.5, the chemical shifts of the TAR were monitored on the 1D Watergate experiment and the ^{13}C -HSQC during the titration against AMG.

2.5.4 T1 relaxation rate measurement

NMR spectroscopic methods can be used to probe motions over a variety of timescales. Autocorrelation (T_1 , T_2 , $T_{1\rho}$, NOE) and cross-correlated relaxation measurements can be used to monitor motions on the ns-ps timescale [¹²⁰].

Relaxation is induced by field fluctuations that occur due to molecular motion, i.e. changes in the local fields as the molecule moves. Thus, NMR relaxation experiments monitor the magnetic fluctuations arising from these molecular motions.

A variety of parameters are at the disposal of the NMR spectroscopist that can be measured using relaxation experiments. It is common to measure the following relaxation parameters for heteronuclear HX spin systems: T_1 –the spin-lattice/longitudinal relaxation time, or the corresponding relaxation rate $R_1=1/T_1$; T_2 – the spin-spin/transverse relaxation time, or the relaxation rate $R_2=1/T_2$. The spin-lattice relaxation rate describes the recovery of the longitudinal magnetization to thermal equilibrium—the return of the populations of the energy level of the spin system to the Boltzmann distribution. T_2 measures the loss of coherence as individual spins experience slightly different resonance frequency due to inhomogeneities in the local magnetic field.

Relaxation rates can also provide useful insights into the molecular weight and size of the molecule. As molecular weights increase the line broadening that is due to slow molecular tumbling rate increases. Consequently, T_1 and T_2 relaxation times become shorter, though T_1 grows again after reaching a minimum.

In order to measure longitudinal relaxation rate, an inversion recovery experiment is performed, while transverse relaxation rates are measured by CPMG (Carr-Purcell-Meiboom-Gill) experiments, which are essentially a train of 180° pulses [¹²¹; ¹²²]. The relaxation times T_1 and T_2 directly related to spectral densities $J(\omega)$ that describes the reorientation motions of the bond vector to be studied. The spectral densities are related to the correlation function that defines the motion by a simple Fourier transformation; this correspondence makes it possible to probe bimolecular motions from relaxation measurements. Unlike proteins, where the ^{15}N atoms of amide residues provide a ready tool for the measurement of relaxation rates and thus give a deeper insight into residue specific dynamics, nucleic acids do not contain as many ^{15}N atoms. The ones that are present in the bases, the water-exchange properties of the GUA, CYT and ADE NH_2 are not conducive to their analysis [¹²³]. While the

measurement of ^{15}N relaxation rates certainly shed light on the motions of base paired GUA and URI residues, they are inadvertently limited and thus it becomes necessary to measure ^{13}C relaxation parameters to gain a deeper insight into the structure and dynamics of the nucleic acid in question.

There is an increasing probability of the formation of multiple binding of ligand or multimerization of the RNA in the presence of high concentration of the ligand. To determine if the TAR RNA retains its monomeric status or whether the presence of the excess ligand encourages dimerization (multimerization), we measured T_1 relaxation times [124] of the C8-H8/C6-H6 in a ^{13}C - ^1H correlation experiment (Fig 2.5.4) using relaxation delays (d8) of 5, 10, 20, 30, 40, 60, 80, 100, 120, 160, 200, 300, 400, 600, 800 and 1200ms.

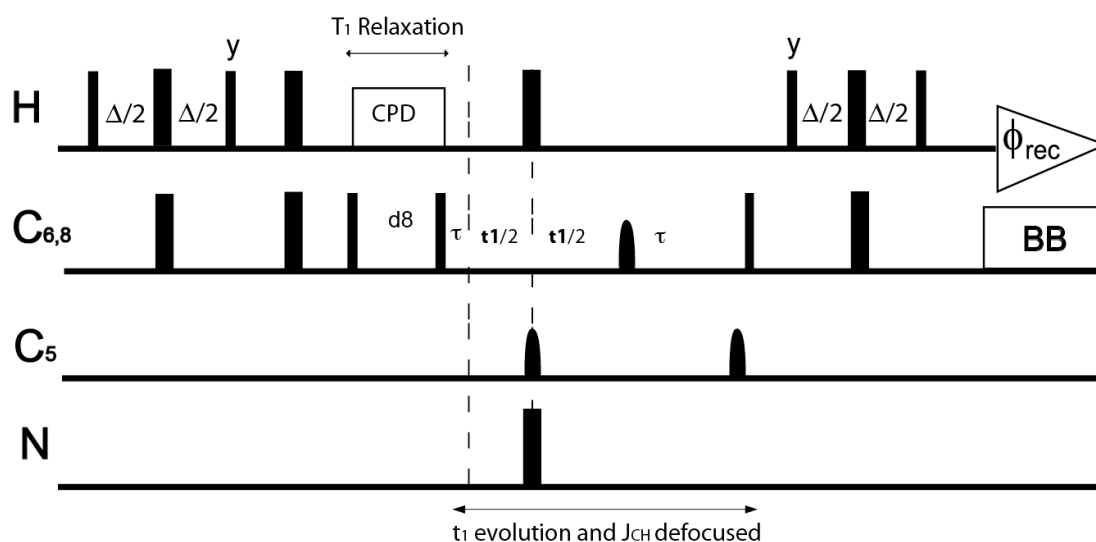


Fig2.5.4 Pulse sequence to measure T_1 relaxation rate for the TAR/AMG complex

After the INEPT step ($H_z C_z$), a reverse INEPT step ensures that we end up with $-C_z$ term thus allowing us to measure only the ^{13}C relaxation rates. The delay d8 is varied systematically in different experiments such that the inverted magnetization on ^{13}C is allowed to undergo longitudinal relaxation. The subsequent steps allow chemical shift evolution of ^{13}C in time period t_1 and coupling modulated by the time period 2τ after which the magnetization is transferred from ^{13}C to ^1H and detected. The delays are set such that $\tau = 1/4J_{\text{C-H}}$.

2.5.5 Field induced residual dipolar couplings

NMR structures of RNA molecules are usually less well defined than proteins with similar size and weight. This is because only short range interactions, such as sequential NOEs (intermolecular) and intra base pair NOEs, and dihedral angles as elucidated from J-coupling constants are utilized while long range distance information remain sparse or absent. In such circumstances, residual dipolar couplings become an invaluable source of long-range distance and angle information. Under conditions of isotropic tumbling of molecules, the RDCs average to zero and hence this valuable information is lost. To prevent this, several methods to induce partial alignment of the biomolecule in question (in our case RNA) have been introduced, prominent among them filamentous Pf1 bacteriophages, bicelles, Otting phase [125] and more recently polyacrylamide gels [126].

However, the simultaneous presence of the negatively charged TAR RNA and an excess of positively charged AMG deterred us from using phages [127; 128; 129] or bicelles [130; 131] to attempt the alignment of TAR RNA/AMG complex. Instead, we exploit the natural anisotropic magnetic susceptibility inherent in most molecules, which can lead to partial alignment in a magnetic field [132; 133]. This alignment is substantial for experimental investigation only at high field strength and even then is extremely small (between 0 to 10Hz). We exploit this property of dipolar couplings scaling linearly with the square of the field strength [134; 135] and measure the splitting (J+D) of the C2-H2, C6-H6, C8-H8 and C1'-H1' at 600, 700 and 800 MHz. The field-induced residual dipolar couplings (*f*RDCs) at 800 MHz were derived by linear fitting of the three values against the square of field strength and extrapolated to zero magnetic field to obtain the value of J (Igor Pro 5 Wavemetrics, Inc). The couplings were extracted from coupled and decoupled version of the ¹³C-HSQC experiments. At higher field strengths, the anti TROSY component of the coupled spectrum becomes much weaker (broad) because of the fast decay of the ¹³C component; therefore the J+D coupling as given by the difference between the TROSY and anti TROSY peak of the coupled spectrum cannot be accurately measured.

By measuring the difference between the distance of the coupled and the corresponding decoupled peaks, an estimate of half the J+D coupling at particular field strength can be made within appropriate error limits. A macro written in Felix

assisted in shifting and overlaying 1D slices of the coupled and decoupled spectra (Appendix C).

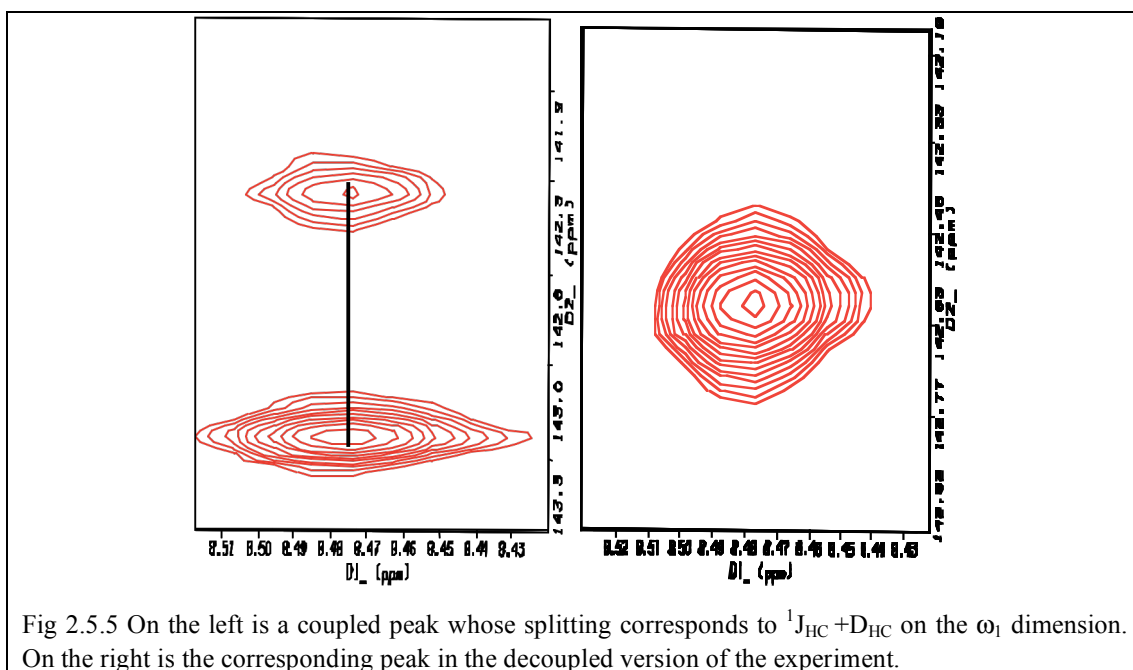


Fig 2.5.5 On the left is a coupled peak whose splitting corresponds to $^1J_{\text{HC}} + D_{\text{HC}}$ on the ω_1 dimension. On the right is the corresponding peak in the decoupled version of the experiment.

Only those f_1 RDCs that showed a linear relationship to B_0^2 within the experimentally determined error limits were employed in a subsequent refinement procedure.

2.5.6 NOE distance restraints

Inter and intra nucleotide distance constraints were extracted from a 3D ^{13}C -HSQC NOESY (mixing time 100 ms) and a base resolved/selective 3D ^{13}C -HSQC NOESY (mixing time 100 ms). Peak intensities were scored as strong, medium and weak using contour levels to maintain consistency. Intermolecular distance restraints were obtained from a ^{13}C -edited/ ^{12}C -filtered HSQC-NOESY [$^{136}; ^{137}$] spectrum acquired in D_2O with a mixing time of 200ms where the NOEs between the ^{13}C -labeled TAR RNA and unlabeled AMG could be selectively observed.

The NOE peak intensities were converted into appropriate NOE distance restraints with upper bounds of 3 Å for strong intense peaks, 4.0 Å for medium intense peaks and 5.0 Å for weakly intense peaks. Very weak NOEs to exchangeable protons or very weak intermolecular NOEs were allowed an upper bound of 6.0 Å.

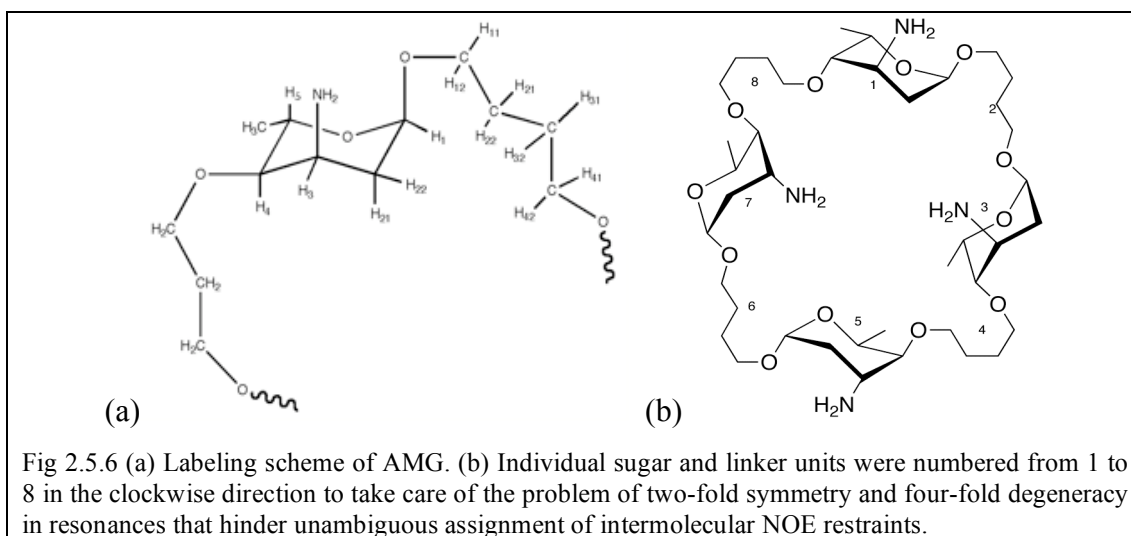
RNA residue/atom	AMG residue/atom	RNA residue/atom	AMG residue/atom
ADE 35 H1'	A H5, B H4*, B H2*/H3*	URI 23 H5	A H5/H3, B H1*/H4*, BH2*/H3*
ADE 35 H2'/H3'	A H3, B H4*, B H2*/H3*, A H1	URI 25 H1'	A H1, A H5/H3, B H1*/H4*, BH2*/H3*, A H6*
ADE 35 H4'	A H1, A H3/H5, B H4*, A H4, B H2*/H3*	URI 25 H5	B H1*/H4*, B H2*/H3*, A H6*
ADE 35 H2	A H1, A H3/H5, B H4*	URI 25 H6	A H6*, B H2*/H3*
ADE 35 H8	A H3, BH4*, AH4, A H2*, AH6*	GUA 26 H5'/H5''	A H6*
ADE 22 H8	A H1, A H3, B H4*, B H2*/H3*	GUA 36 H8	B H1*/H4*, B H2*/H3*
URI 23 H1'	B H2*/H3*	GUA 43 H1'	B H2*/H3*, A H1, A H2*
URI 23 H2'	B H1*/H4*, B H2*/H3*	GUA 44 H4'	A H3/H5, B H2*/H3*, A H6*

Table 2.5.6 Intermolecular connectivity's of RNA residues to the ligand as observed in a ^{13}C -edited/ ^{12}C -filtered HSQC-NOESY spectrum. The sugar moiety of the ligand is referred to as "A" and the linker moiety is referred to as "B" in this table.

The intermolecular NOE data has been summarized in Table 2.5.6. The nomenclature of the protons can be understood from Figure 2.5.6(a).

Given that the ligand AMG has a two-fold symmetry and a four-fold degeneracy of its resonances, a preliminary intermolecular data set with full ambiguity among the four possibilities was evaluated. To arrive at a more appropriate intermolecular data set, a rigorous and exhaustive procedure was adopted to define intermolecular NOEs spanning all reasonable possibilities that are not equivalent by symmetry.

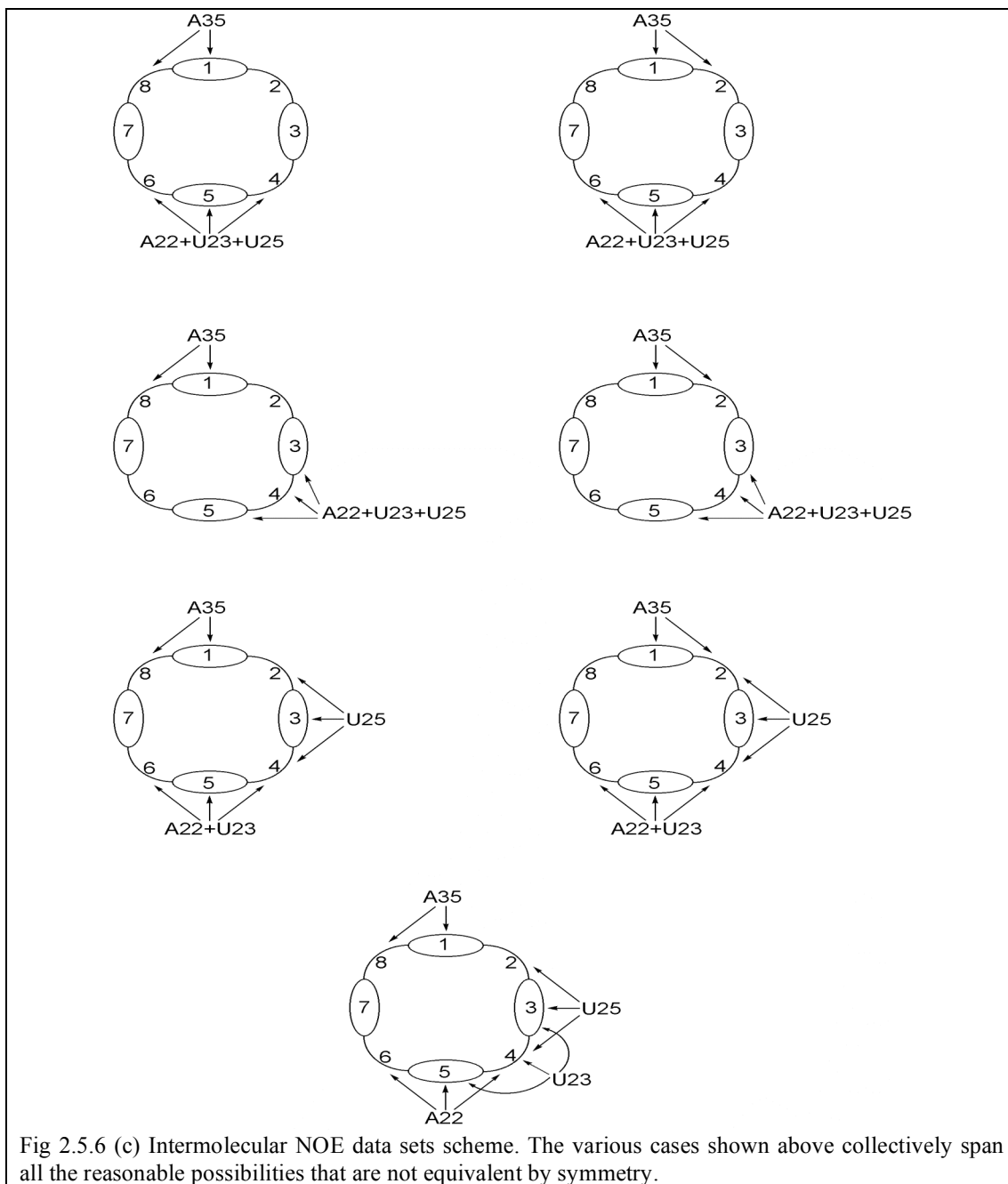
The procedure began by numbering the sugars and linkers of AMG from 1 to 8 in the clockwise direction to define a ring consisting of sugar 1, linker 2, sugar 3, linker 4, sugar 5, linker 6, sugar 7 and linker 8 (Fig 2.5.5 b).



Seven data sets spanning all reasonable possibilities equivalent by symmetry were made as depicted in Fig 2.5.6 (c).

Two assignments were considered for NOEs stemming from residue A35. In the first possibility A35 NOEs were directed towards sugar 1 and linker 2 (column 2 Fig 2.5.5 (b)) while in the second possibility A35 NOEs were directed towards sugar 1 and linker 8 as shown in column 1. For the NOEs stemming from the bulge residues A22, U23 and U25, three alternatives were devised. a) A22, U23 and U25 contact sugar 5 and linkers 4 or 6 (first row); b) A22, U23 and U25 contact sugars 3 or 5 and linker 4 (second row); c) A22 and U23 contact sugar 5 and linkers 4 or 6, while U25 contacts sugar 3 and linkers 2 or 4 (third row). Combination of the two possibilities for NOEs from A35 and with the three possibilities arising from the bulge residues yields a total of 6 intermolecular NOE data sets. In the seventh set (bottom), NOEs stemming from A35 were assigned to sugar 1 and linker 8; those stemming from U25, to sugar 3 or linkers 2 or 4; those stemming from U23, to sugars 3 or 5 and linker 4; those stemming from A22, to sugar 5 and linker 4 or 6. When a residue contacts two AMG moieties equivalent by symmetry (e.g. linkers 4 and 6), those restraints were treated as ambiguous in the structure calculation. Taking into account the two-fold symmetry of AMG, the initial seven sets of intermolecular NOE data sets were duplicated by shifting the sugar and linker numbering by one unit (sugar1 -> sugar3, linker 2 ->

linker 4 etc.). Finally, fourteen unique combinations were used in the restrained Molecular Dynamics protocol.



2.5.7 Dihedral angle restraints

The furanose rings in a nucleic acid are twisted out of plane to minimize non-bonded interactions between their substituents. This phenomenon is called puckering [138].

The sugar pucker determines the shape of the A-helix, whether the helix will exist in the A-form or in the B-form. The five membered rings are for steric reasons not planar (Pitzer tension and Newman strain) because then all bonds would be in the eclipse conformation hence one atom or two are turned out of the plane [139]. In the ribo furanose, the plane C1'-O4'-C4' is fixed. Endo-pucker means that C2' or C3' are turned out of this plane into the direction of C5'. Exo-pucker describes a shift in the opposite direction. In RNA we find predominantly the C3'-endo conformation. The sugar pucker was defined as either C2'-endo or C3'-endo depending on the J coupling constants between H1' and H2' derived from a 2D-HCCH-E.COSY spectrum (140). As shown in Fig 2.5.6 (a), if a ribose sugar adopts a C3'-endo conformation, as is usually the case with A-form backbone geometry, then vicinal couplings between ribose H1' and H2' are typically around 2 Hz and smaller. In contrast, the C2'-endo conformation results in larger coupling values of around 6 to 8 Hz. Intermediate values are indicative of exchange between the two conformations and residues exhibiting such coupling values are best left unrestrained.

The H3'-H4' couplings can also be used as a diagnostic tool to determine the pucker. A small value of J (H3', H4') is indicative of a C2'-endo conformation while a large J (H3', H4') is indicative of a C3'-endo conformation adopted by the sugar pucker.

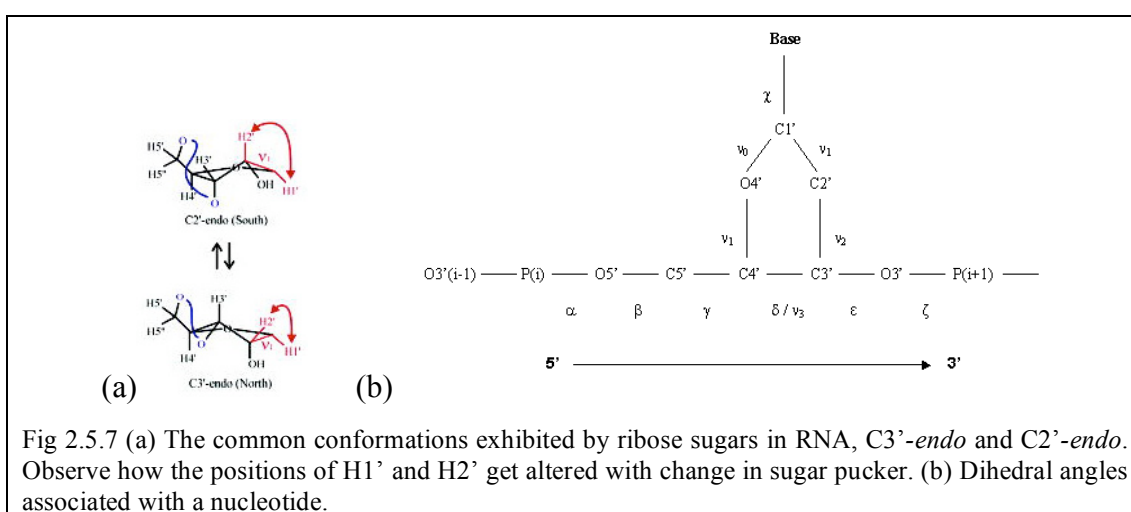


Fig 2.5.7 (a) The common conformations exhibited by ribose sugars in RNA, C3'-endo and C2'-endo. Observe how the positions of H1' and H2' get altered with change in sugar pucker. (b) Dihedral angles associated with a nucleotide.

The HCCH-E.COSY experiment [¹⁴⁰] enabled us to simultaneously detect the ³J_{HH} homonuclear coupling along the ω₂ dimension and the ¹J_{CH} heteronuclear coupling along the ω₁ dimension Fig 2.5.6 (d). For the determination of the sugar pucker of each residue, we are interested in the ³J_{H1'H2'}. The pulse sequence employed is shown in Fig 2.5.6 (c).

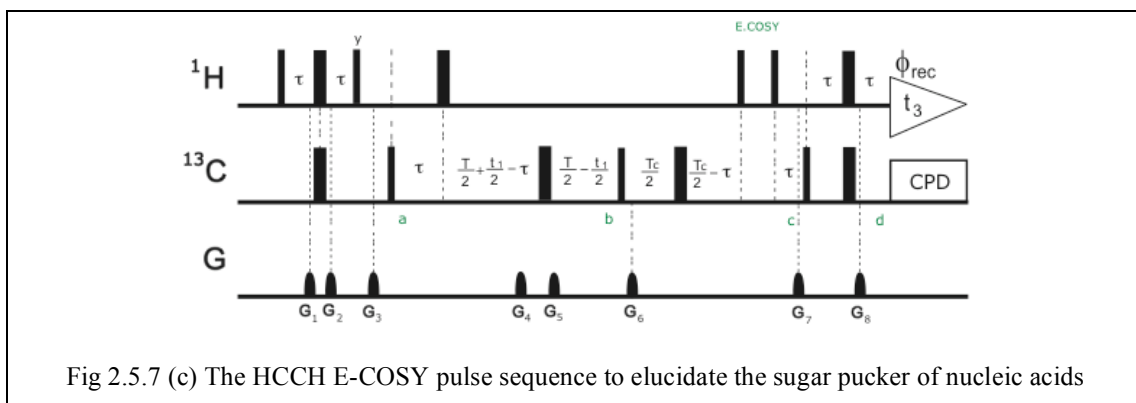


Fig 2.5.7 (c) The HCCH E-COSY pulse sequence to elucidate the sugar pucker of nucleic acids

This experiment relies on the initial transfer of magnetization from ¹H to ¹³C using an INEPT step at which point the magnetization can be expressed in the form 2H_zC_y (a). This is followed by an evolution period wherein the chemical shift of ¹³C evolves in the duration t₁/2 + T/2 – (T/2 – t₁/2) = t₁. The heteronuclear coupling between ¹H and ¹³C evolves for τ - T/2 + τ - t₁/2 + (T/2 – t₁/2) = -t₁ + 2τ. In the evolution time t₁ we get an in phase splitting of the resonances in the ω₁ dimension. By adjusting the time interval 2τ = 1/2J_{HC} the evolution of the J_{HC} coupling can be described in the operator form

$$2H_z C_y \rightarrow -C_x$$

The homonuclear coupling between ¹³C spins evolves in time T as described by

$$-C_x \rightarrow 2C_{1y}C_{2z}\sin(\pi J_{C1C2}T)$$

At step (b), the 90° pulse flips the magnetization of ¹³C to the transverse plane

$$-2C_{1y}C_{2z} \rightarrow -2C_{1z}C_{2y}$$

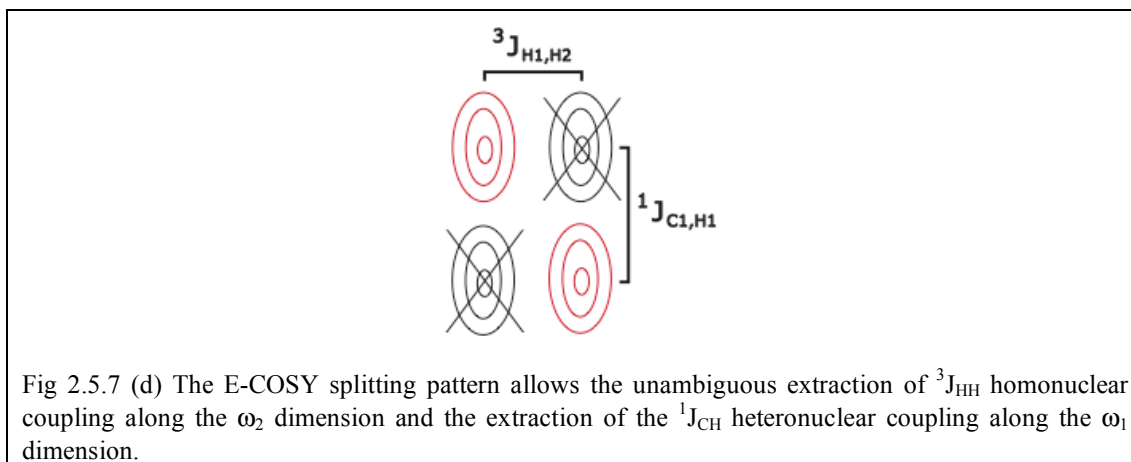
When 2τ is set to 1/2J_{HC} the heteronuclear coupling evolves as

$$2C_{1z}C_{2y} \rightarrow 4C_{1z}C_{2x}H_{2z}$$

And the homonuclear coupling J_{CC} under the influence of time T_c

$$4C_{1z}C_{2x}H_{2z} \rightarrow 2C_{2y}H_{2z} \sin(\pi J_{C1C2}T_c)$$

T_c is optimally set to $1/4J_{cc}$. The E-COSY element ensures that the α and β spin states of ^1H after the evolution time t_1 don't mix. In a final INEPT step (d), the magnetization is transferred back to ^1H and then detected (e).



The presence of typical NOEs associated with an A-form helix (^{104; 105; 141; 142}) led us to impose 41 non-experimental dihedral angle restraints defining the backbone angles α , β , ζ and ϵ , as shown in Fig 2.5.6 (b), within a range of $\pm 60^\circ$ for the lower stem residues 16 to 21 and 41 to 46. Figure 2.5.6 (e) schematically illustrates typical NOEs (exchangeable and non-exchangeable) observable in a stretch of RNA adopting an A-form helix.

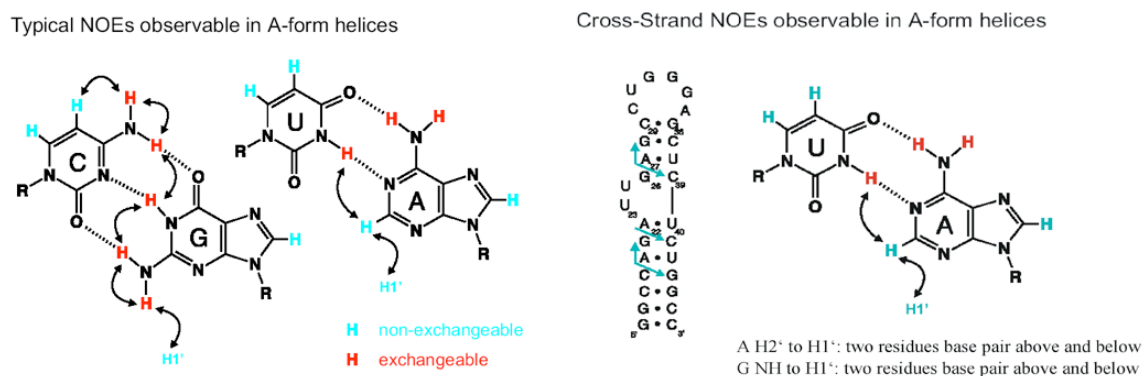


Figure 2.5.7 (e) NOEs indicative of A-form helix formation in RNA

2.5.8 Planarity and hydrogen bonding

Although X-PLOR NIH provides improper angles to maintain planarity of groups of atoms, the specification of these energy terms is sometimes cumbersome, especially when many atoms are involved. The restraints planarity statement allows one to define an effective energy term that penalizes out-of-plane conformations of the selected atoms. The planarity restraints energy term is defined as

$$E_{plan} = \sum_{g \in \text{groups}} \omega_{plan} \sum_{i \in g} g_i^2$$

Where the first sum is carried out over all defined groups of planar atoms, the second sum is carried out for all atoms i within each group, and g_i is the orthogonal distance of i from the least-squares plane defined by all atoms of the group.

Hydrogen bonds are of fundamental importance in the formation of RNA structures.

A version of the 2D ^{15}N imino-NOESY for observing the exchangeable protons was used to elucidate the presence of hydrogen bonds between Watson Crick base pairs in the TAR RNA. The base pairs were constrained in two ways: 1) NOE style distance restraints between each of the two (for A: U base pairs) or three (for the G: C base pairs) donor-acceptor pairs. 2) A term defining planar pseudo energy terms defined for each base pair.

The failure to observe an imino resonance indicative of hydrogen-bonding scheme between A22 and U40, even at 4°C suggested that the A22-U40 hydrogen-bond pattern was sheared. However, several NOEs between U40 to C39/C41, and A22 to G21/G26, which are indicative of residues U40 and A22 remaining stacked between C39/C41 and G21/G26 respectively, lead us to treat the planarity and hydrogen bond restraints between these two restraints with more caution. Hence, while the planarity and hydrogen bond restraints between these two residues were present during the initial high temperature dynamics step of the molecular dynamics protocol to ensure that A22 and U40 remained stacked, these restraints were removed for subsequent stages of the rMD protocol, thus allowing both the residues to sample conformational space in the absence of planarity and hydrogen bond restraints between the two residues but only in the presence of NOE distance restraints indicative of U40 and A22 remaining stacked between C39/C41 and G21/G26 respectively.

2.5.9 Structure calculations

Structures were calculated using restrained Molecular Dynamics (rMD) and energy minimization in XPLOR-NIH [^{143; 144}] to obtain the final ensemble of structures of the TAR RNA/AMG complex.

XPLOR-NIH [^{143; 144}] relies on two files for its knowledge of the molecular structure and energies defined during the restrained Molecular Dynamics run. The topology file defines atom types, bonds, angles and improper terms, while the parameter file contains details regarding force constants for bonds, angles, impropers, dihedrals and non-bonded parameters. It is essential to use the topology and parameter files in tandem because the parameter file uses the information present in the topology file to calculate the molecular structure file (extension .psf).

For the TAR RNA, the latest DNA/RNA Amber [¹⁴⁵] parameter and topology files were downloaded from the Moore Lab Homepage (<http://proton.chem.yale.edu/>).

Since the amino glycoside was a novel structure, not previously studied, the topology and parameter files for AMG were assembled by putting together information from the individual monomer units of N-Acetyl- β -d-Glucosaminidase (NAG) and other similar amino sugars and a butadiene monomer unit, available with the distributions of XPLOR-NIH and CNS (Crystallography and NMR System) [¹⁴⁶].

A starting template and psf file of AMG was created using the customized parameter and topology files.

For the HIV-2 TAR RNA, the RNA coordinates of HIV-2 TAR-Argininamide structure (1AKX.pdb) were used as a starting structure. These two PDB's served as starting templates for subsequent structure calculations using the restrained Molecular Dynamics protocol outlined in the following sections.

In its simplest form, the rMD protocol incorporates distance restraint terms of the form $k(d-d_0)^2$ where d is the distance between the atoms in the current conformation and d_0 is the desired distance dynamics approach derived from NMR spectrum. k is a force constant, the value of which determines how tightly the restraint should be applied. The information derived from the 2D HCCH E.COSY experiment on the sugar puckers and the data on other torsion angle restraints can also be incorporated into the molecular dynamics energy function in addition to the distance restraints as derived from the NOE data. The force constants for the NOE and the Dihedral energy terms were set to 50-kcal/mol Å and 300-kcal/mol ° respectively. Another way to

understand the significance of these force constants is that structures in which the distance between particular atoms or dihedral angles (between planes) exceed the target NOE derived distance or E.COSY derived dihedral angle restraints, then a penalty that is directly proportional to the force constant attributed to these restraints is imposed which contributes to the total energy of the structure. This can be mathematically written as

$$\begin{aligned} V(d) &= k_l (d-d_l)^2 & d < d_l \\ V(d) &= 0 & d_l \leq d \leq d_u \\ V(d) &= k_u (d-d_u)^2 & d_u < d \end{aligned}$$

Where d_l and d_u are the lower and upper distances that are consistent with the experimental data. Thus a distance between d_l and d_u incurs no penalty. Outside this region, the restraint force and the penalty is dictated by the form of the potential function viz. square, harmonic, soft etc. A similar treatment is conferred upon dihedral angles and dipolar coupling restraints (discussed in detail later).

During the simulation, the system develops in a potential field under the influence of various forces, in which all information about the complex is summarized. Two classes of energy terms are distinguished: $E_{\text{empirical}}$ and $E_{\text{experimental}}$.

$$\begin{aligned} V &= E_{\text{empirical}} + E_{\text{experimental}} \\ E_{\text{experimental}} &= E_{\text{NOE}} + E_{\text{torsion}} + E_{\text{sani}} \\ E_{\text{empirical}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{vdw}} + E_{\text{electr}} \end{aligned}$$

$E_{\text{empirical}}$ contains all information about the primary structure of the RNA and AMG and also data about topology and bonds in nucleic acids and AMG in general. The contributions of covalent bonds, bond-angles and dihedral angles towards $E_{\text{empirical}}$ are approximated by a harmonic function. In contrast, non-covalent van-der-Waals forces and electrostatic interactions are simulated by an inharmonic Lennard-Jones potential and Coulomb potential, respectively. $E_{\text{experimental}}$ takes the experimentally determined constraints into account. Angle constraints are introduced by a harmonic function analogous to that for the dihedral angles. For distance constraints, the energy potential will be set to zero, if the corresponding distance is within the given limits. If it is

outside these limits, a harmonic energy potential is used, which tries to push the value of the distance into the limits.

The restrained Molecular Dynamics protocol consisted of four stages.

The first stage consisted of a preliminary High Temperature Molecular Dynamics run at 1000K to generate 150 starting templates of the TAR RNA/AMG complex. The use of different seeds for each structure generation leads to different structures. The starting structure (1AKX.pdb, already folded RNA with appropriate base pairing in place) is allowed to undergo high temperature dynamics at 1000K. Restraints to preserve the A-form helix structure in the lower stem were imposed to maintain helical structure during the high temperature randomization step. The force constants are allowed to increase very slowly (in step factors of 5 till maximum value of the force constants is reached; this can be followed in the XPLOR protocol in the Appendix) thus ensuring that maximum conformational space is sampled during the high temperature dynamics. The goal of a NMR structure calculation is to ensure that the structures generated are defined by the experimental restraints imposed. The total energy and violation energy of NOEs thus provide a tool that can be used to sort structures, which are refining against the restraints to define the structure. The slow increase of the force constants along with the dynamics at high temperature results in structures very different from each other. An r.m.s.d. to the mean of 3.4 Å angstroms (excluding the hexanucleotide loop of the TAR RNA) for an ensemble of chosen structures confirms that the initial high temperature dynamics succeeds in sampling a wide area of conformational space. After the high temperature phase, those structures with total energy less than 600 kcal/mol and NOE violation energy less than 100 kcal/mol were chosen. Only structures meeting the appropriate cut offs in energy values were allowed to pass on to the subsequent stage.

The next stage of the protocol is the Simulated Annealing (SA) wherein structures with reasonable energies generated in the High Temperature phase are subjected to 3 rounds of 3ps dynamics at 1000K, 600K and 400K with 700 Powell minimization steps between the dynamics [¹⁴⁷]. All force constants were ramped to their full value and structures were sorted against their total energies and violation energies. The hydrogen-bonding restraints deserve special attention at this stage. The presence of several NOEs between U40 to C39/C41, and A22 to G21/G26 are indicative of residues U40 and A22 remaining stacked between C39/C41 and G21/G26 respectively. However, the failure to observe an imino resonance even at 4°C implies

that the hydrogen-bonding scheme between A22 –U40 is disrupted. Consequently, hydrogen-bonding restraints between A22 and U40 were employed only during the high temperature phase of the rMD and were removed for subsequent stages of the protocol. Structures were chosen based on total energy less than 400 kcal/mol and NOE violation energy less than 50 kcal/mol after the simulated annealing.

In the refinement protocol that follows the simulated annealing, chosen structures were subjected to 2ps dynamics at 500, 300, 150, 50 and 10K with 1000 steps of Powell minimization between the dynamics. The dihedrals and planarity terms were switched off to allow for sampling all conformational space. At the end of the refinement, structures with NOE violation energy lower than 25 kcal/mol were chosen for subsequent refinement with the *fi*RDCs. Structures at this stage had 0 to 4 NOE violations greater than 0.2Å and no angular violations greater than 5°. The best fourteen structures at this stage converged to 1.6Å r.m.s.d.

The final stage of calculations consisted of refinement with the incorporation of *fi*RDCs. While only repulsive Van der Waals non-bonded energy terms were employed in the earlier stages of the protocol to speed up the calculations, at this stage of refinement, the attractive Lennard Jones potential [143; 144] is switched on. The field induced dipolar couplings were included as restraints for refinement using the “Susceptibility Anisotropy (*sani*)” algorithm with a square potential well in Xplor-NIH [148; 149]. One of the prerequisites to compute the potential arising from “*sani*” is the presence of external orienting axis which enables the calculation of Θ and Φ which are the angles measured between the dipolar vector and the alignment coordinate system, which is represented by four pseudo atoms OO, X, Y and Z in Xplor-NIH. OO represents the origin of the coordinate system while X, Y and Z represent the Cartesian axis system. The projection of the dipolar vector on the OO-Z axis results in $\cos\Theta$ while the projection onto OO-X and OO-Y result in $\sin\Theta\cos\Phi$ and $\sin\Theta\sin\Phi$, respectively. The implementation of Susceptibility Anisotropy [148; 149] can be better understood in the following way. The dipolar coupling D arising due to the partial orientation of macromolecules (in our case the TAR RNA / AMG complex) is a function of the relative orientation of the dipolar vectors with respect to the alignment coordinate frame (Θ and Φ), the magnitude of the alignment tensor (\mathbf{A}_a , \mathbf{A}_r or \mathbf{D}_a , \mathbf{R}), the gyro magnetic ratios of the interacting nuclei (γ_p and γ_q), the

distance between the nuclei (\mathbf{r}_{pq}) and the generalized order parameter of the dipolar vector (\mathbf{S}).

$$\mathbf{D}^{pq} = -S\gamma_p\gamma_q [A_a (3\cos^2\Theta - 1) + (3/2)*A_r (\sin^2\Theta \cos 2\Phi)] / r_{pq}^3$$

The above equation reduces to

$$\mathbf{D}^{pq} = \mathbf{D}^{pq}_{axial} (3\cos^2\Theta - 1) + (3/2)*R (\sin^2\Theta \cos 2\Phi)$$

Where Θ is the angle between the internuclear vector and the z -axis of the tensor, and Φ is the angle between the angle between the projection of the internuclear vector on the x - y plane of the tensor and the x -axis.

In Xplor-NIH, the function that is minimized is the difference between the calculated and measured dipolar couplings.

$$\mathbf{E}_{sani} = k (\mathbf{D}_{measured} - \mathbf{D}_{calculated})^2$$

Where k is the force constant assigned to the “sani” term during the calculation.

To obtain an accurate estimate for D_a and R , the measured dipolar couplings are plotted as a histogram [¹⁴⁸], the extremes of which correspond to the alignment tensor components A_{xx} , A_{yy} , and A_{zz} where $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$. Also, the alignment tensor is traceless leading to $A_{xx} + A_{yy} + A_{zz} = 0$. D_a and R can be computed using the following equations:

$$A_{zz} = 2A$$

$$A_{yy} = -D_a (1+3/2R)$$

$$A_{xx} = -D_a (1- 3/2R)$$

The accuracy of the estimate in the histogram method increases with the increase in the number and types of measured dipolar coupling data. In our case, we rely exclusively on 37 ^{13}C - ^1H field induced dipolar coupling data, which makes the accurate estimation of the magnitude of D_{zz} and R difficult via the histogram method. Hence, a grid search [¹⁴⁹] was performed where we evaluated the “sani” energy term as a function of the alignment tensor. A systematic grid search was performed where we sampled the values of D_{zz} between 5 and 10 in 1 Hz steps and values of R between 0.1 and 0.66 in increments of 0.05. The *sani* energy term (\mathbf{E}_{sani}) associated with each grid value was analyzed and the optimized grid value was chosen such that *sani* energy term was minimized. The optimized grid value, in conjunction with the experimentally determined error limits (square well potential) was used in the final refinement of the TAR RNA / AMG complex.

Correlation coefficients of the experimentally derived *f*iRDCs versus the calculated *f*iRDCs and the Q factors [¹⁵⁰] of the final structures were calculated using DipoCoup [¹⁵¹]. Final structures were viewed and analyzed using MOLMOL [¹⁵²].

Scripts written in Awk, Python and Perl languages aided in data conversion, analysis and preparation of input and output scripts.

The general scheme of the calculation and the Xplor scripts used to arrive at the final structures are presented in APPENDIX B. Further, scripts used for the analysis of violations, energies and tensors can be found in APPENDIX C.

3 Results

3.1 Resonance assignment

The 3D HCCH-COSY-TOCSY experiment was used to get a complete spin assignment of the ribose spin system of the TAR RNA (C1'-H1', C2'-H2', C3'-H3', C4'-H4' and C5'-H5'/H5'' correlations). Fig 3.1 (a) and (b) trace the ribose spin connectivity of the U42 and A35 residues of the TAR RNA in the direct dimension (D1 in Felix) respectively. Though the resonances of the C2'/H2' and C3'/H3' are overlapped as shown in the figures, the HCCH-COSY-TOCSY experiment allows unambiguous assignment of the ribose spin system. Sequential resonance assignment employing a 3D ^{13}C -HSQC NOESY and a base resolved/selective 3D ^{13}C -HSQC NOESY helped correlate the peaks arising from the H1' or H2' and the H6/H8 (i+1) of the TAR RNA nucleotides. The DQF-COSY, TOCSY and NOESY experiments were employed to obtain the resonance assignment of the ligand AMG. Tables 3.1(a, b) summarize the resonance assignment of the TAR RNA and AMG (Fig 3.1(e)) as obtained from the experiments previously described.

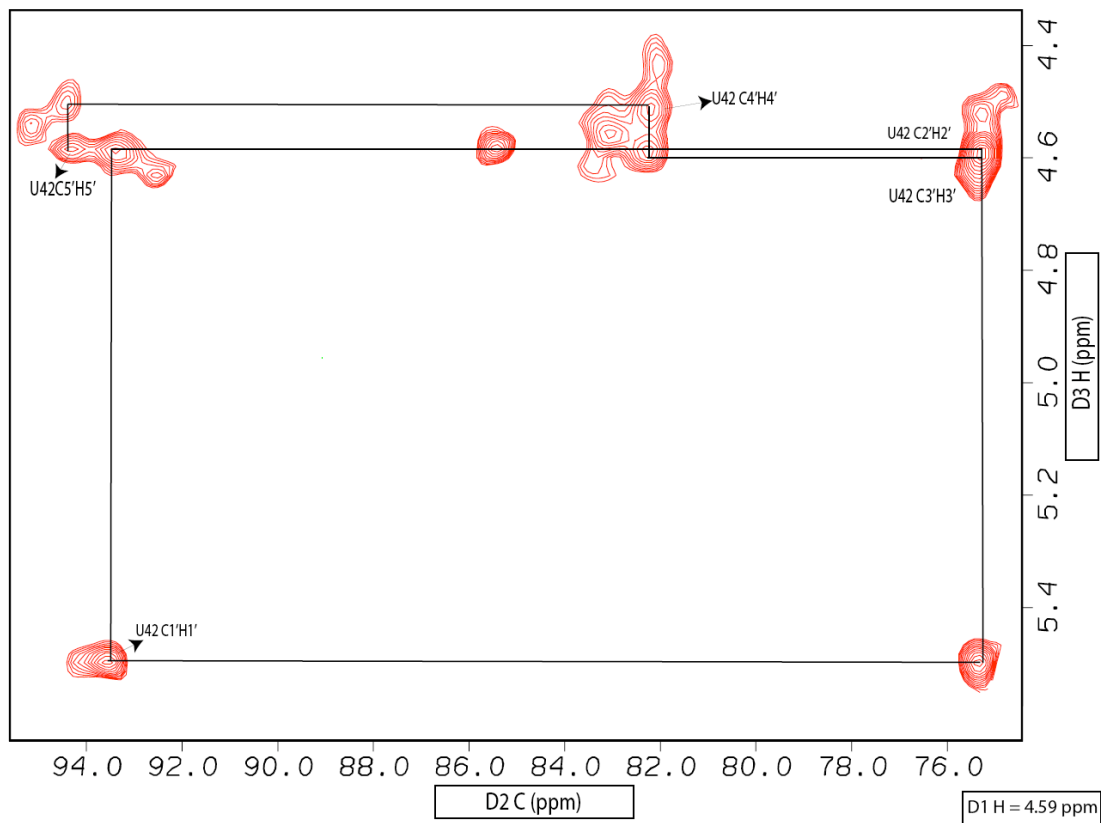


Fig 3.1 (a) 2D skewer of the 3D HCCH-COSY-TOCSY experiment in the direct dimension at 4.59ppm (H2' resonance of the U42) showing the assignment of U42 ribose system.

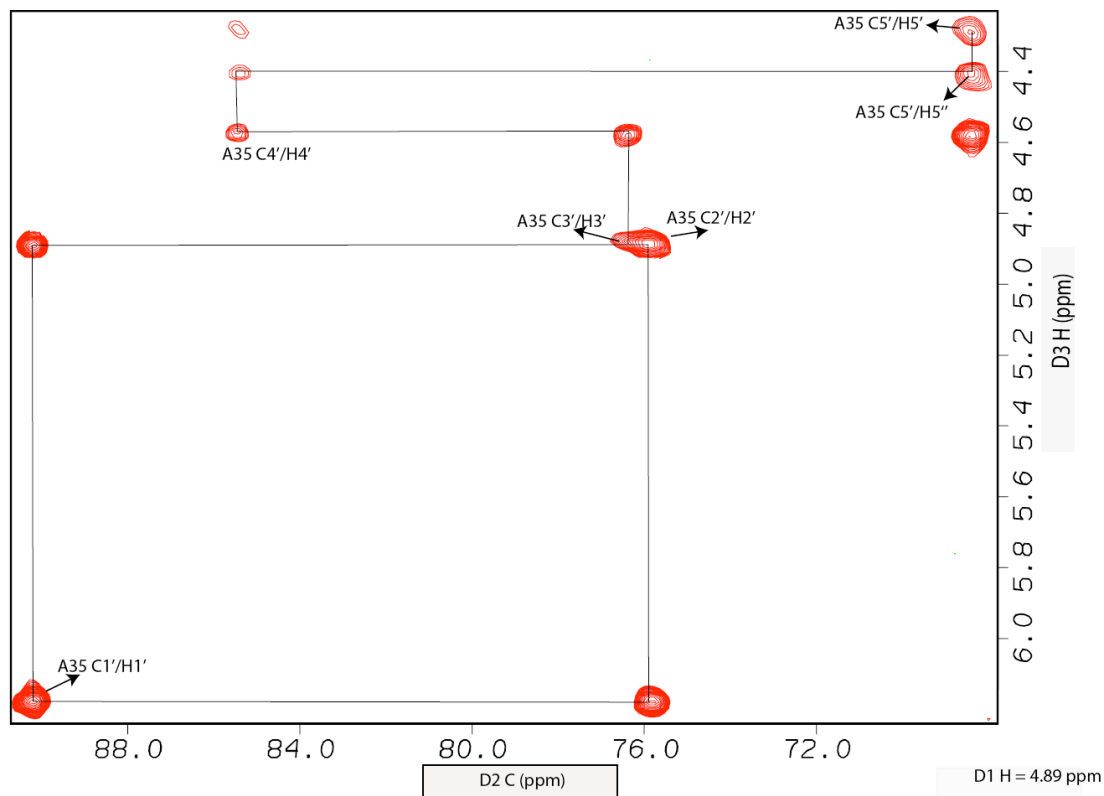


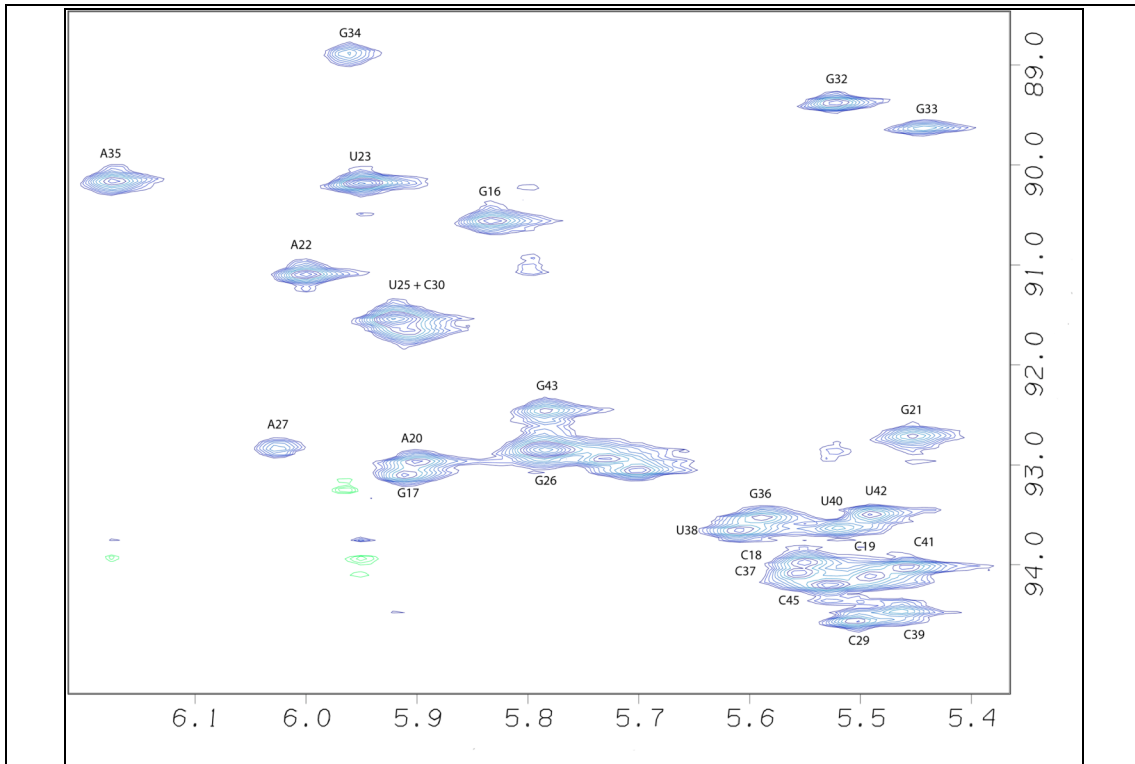
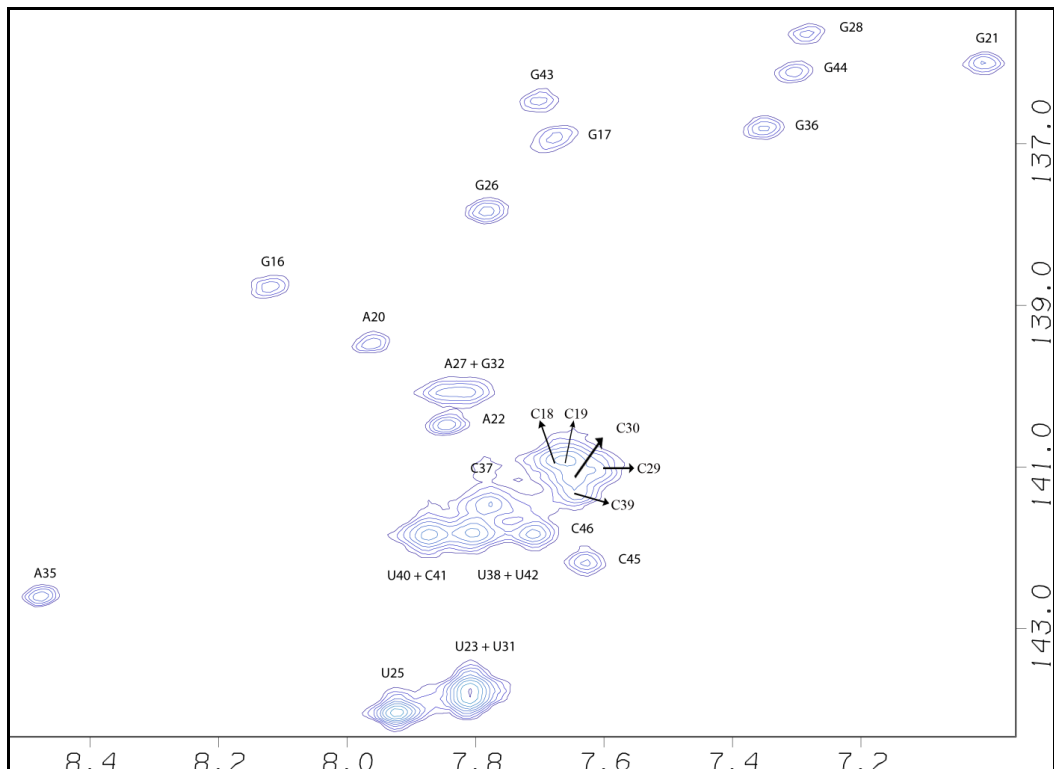
Fig 3.1 (b) 2D skewer of the 3D HCCH-COSY-TOCSY experiment in the direct dimension at 4.89ppm (H2' resonance of the A35) showing the assignment of A35 ribose system.

residue	h1'	c1'	h2'	c2'	h3'	c3'	h4'	c4'	h5'	h5''	c5'
g16	5,792	90,169	4,900	75,750	4,750	76,860	4,560	85,180	4,370	4,260	67,540
g17	5,900	92,930	4,590	75,420	4,520	72,650	4,410	82,080	4,520	4,080	
c18	5,551	93,940	4,440	75,530	4,440	72,207	4,400	81,960	4,520	4,110	64,900
c19	5,486	94,082		75,420	4,540						
a20	5,911	93,057	4,720	75,750	4,640	73,200	4,490		4,490	4,130	66,430
g21	5,459	92,671	4,410	75,530	4,430					4,030	
a22	5,997	91,103	4,410	77,420	4,640	75,530	4,490	84,630	4,490	4,130	
u23	5,937	90,175	4,400	75,310	4,690	77,090	4,440	85,410	4,200	4,240	67,800
u25	5,912	91,536	4,490	75,530	4,670	75,640	4,470	84,630	4,250	4,160	67,540
g26	5,792	92,807	4,850	74,900	4,620	74,600	4,570	83,200	4,400	4,200	68,000
a27	6,029	92,806	4,840	75,640	4,690	73,600	4,390	82,400	4,500	4,260	66,400
g28	5,728	92,905	4,520	75,310	4,430	72,980	4,500	82,080	4,130	4,130	66,100
c29	5,506	94,526	4,180	75,420	4,410	72,540	4,390	81,960	4,040	4,560	64,000
c30	5,912	91,536	4,490	75,530	4,640	75,531			4,140	4,440	66,500
u31	5,835	90,534	4,350	75,750	4,550	76,860	4,250	85,180	3,970		67,540
g32	5,521	89,387	4,650	75,400	4,650	78,000	4,140	85,400	4,010		68,000
g33	5,449	89,616	4,390	75,530	4,650	77,300	3,960	84,500	3,890	3,940	67,900
g34	5,964	88,886	4,870	75,980	4,980	75,980	4,470	84,200	4,230	4,060	67,400
a35	6,176	90,184	4,890	75,870	4,880	76,310	4,580	85,520	4,270	4,410	68,320
g36	5,584	93,498	4,590	74,750	4,490	73,310	4,550	82,080	4,490	4,200	66,880
c37	5,562	94,060	4,440	75,530	4,520	72,090	4,500	82,190	4,580	4,100	64,660
u38	5,610	93,634	4,640	75,530	4,500	72,540	4,420	82,080	4,560	4,150	64,500
c39	5,481	94,380	4,410	75,460	4,250	73,200	4,480		4,100	4,440	66,200
u40	5,519	93,588	4,390	75,310	4,470	72,200	4,390	81,960		4,180	
c41	5,480	94,000	4,250	75,420	4,490	72,760	4,410	81,960	4,070	4,530	64,770
u42	5,490	93,476	4,590	75,310	4,630	75,530	4,500	82,080	4,580	4,150	64,800
g43	5,782	92,428	4,630	75,530	4,460	73,090	4,490	82,190	4,590	4,140	66,320
g44	5,696	93,032	4,470	75,200	4,470	72,760	4,470	82,080	4,070	4,500	65,400
c45	5,529	94,195	4,270	75,420	4,460	71,980	4,390	82,080	4,560	4,070	64,660
c46	5,788	92,798	3,990	77,640	4,180	69,870	4,160	83,410	4,490	4,010	65,200

Table 3.1 (a) Resonance assignment of the ribose moieties of the HIV-2 TAR RNA

residue	h2	c2	h5	c5	h6	c6	h8
g16							8,125
g17							7,701
c18			5,258	97,151	7,670	140,700	
c19			5,553	98,253	7,660	140,700	
a20	7,000	150,506					7,973
g21							7,022
a22	7,485	154,536					7,821
u23			5,790	105,447	7,805	143,771	
u25			5,865	105,093	7,914	144,005	
g26							7,790
a27	7,512	153,409					7,862
g28							7,279
c29			5,119	97,097	7,610	141,064	
c30			5,479	98,678	7,650	141,133	
u31			5,852	105,290	7,812	143,747	
g32							7,846
g33							7,657
g34							7,807
a35	8,287	156,694					8,477
g36							7,377
c37			5,211	96,688	7,789	141,486	
u38			5,410	103,403	7,810	141,834	
c39			5,642	98,157	7,660	141,366	
u40			5,553	103,746	7,870	141,500	
c41			5,752	97,627	7,879	141,889	
u42			5,347	103,571	7,810	141,834	
g43							7,707
g44							7,303
c45			5,152	96,877	7,638	141,150	
c46			5,553	98,253	7,708	141,863	

Table 3.1 (b) Resonance assignment of the base moieties of the HIV-2 TAR RNA

Fig 3.1 (c) The C1'-H1' region of the ^{13}C -HSQCFig 3.1 (d) The C6-H6 and C8-H8 region of the ^{13}C -HSQC

The amino glycoside analogue was defined with 2 constituent monomer units; the sugar moiety **ALR** and the linker unit called **BU1**. The assignment of the protons for the respective units is as follows:

ALR	Chemical Shift
H1	4.936 ppm
H21	2.080 ppm
H22	2.110 ppm
H3	3.910 ppm
H4	3.404 ppm
H5	3.940 ppm
H61, H62, H63	1.285 ppm
BU1	
H11, H41	3.480 ppm
H12, H42	3.711 ppm
H21, H22	1.630 ppm
H31, H32	

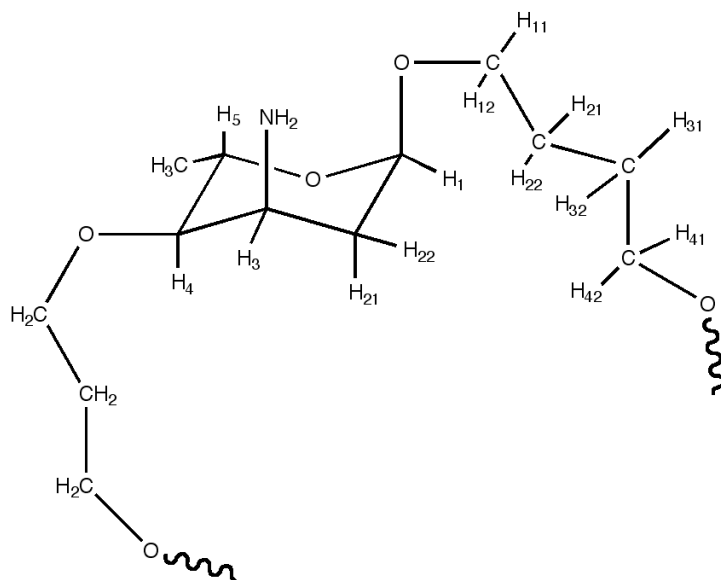


Fig 3.1 (e) Resonance assignment of the protons in the linker (BU1) and ribose (ALR) moiety of the novel aminoglycoside ligand AMG.

3.2 Chemical shift mapping

Chemical shift changes are immediate evidence that binding has occurred. This change is a result from a change in the environment of the nucleotides. A change in line width (or more directly the relaxation rate $1/T_1$ or $1/T_2$) may result from a change in the motional characteristics of the RNA [5; 6].

The chemical shift mapping identifies A22, U23 and the loop nucleotides G32, G34 and A35 as the sites being most affected by the binding of the amino glycoside analog AMG. Large chemical shifts changes are observed upon titration of AMG for the C2/H2 and C1'/H1' resonances of A35. In particular, the line width of the A35–C2 decreases from 80 to 48 Hz at 900 MHz, which suggests that in free TAR RNA, this nucleotide assumes different conformations interchanging on a μs to ms time scale, while in the complex the base occupies a well-defined position. Notable is the chemical shift change experienced by the base of G34, which varies the most upon formation of the complex without showing any direct contact to the ligand. This is indicative of a profound change in the electronic environment of G34 and, possibly, of the overall conformation of the loop, triggered by the interaction of A35 with AMG. The chemical shift change of G34-C8/H8 is accompanied by a considerable decrease in the transverse relaxation time T_2 of G34-C8, indicating that G34, which is disordered in the free TAR and in the TAR/argininamide complex [31; 46; 51; 153], undergoes a slow (μs –ms time scale) conformational transition between two states in complex with AMG [154].

Considerable chemical shifts changes are observed also for the ribose of U31, G32 and G33. In the bulge, the largest chemical shift changes are observed for A22 and U23. A medium-sized chemical shift change of G43 C1'-H1' resonance suggests additional binding effects in the minor groove of the TAR RNA lower stem.

As shown in Fig 2.5.3, significant chemical shift changes were observed up to a 10-fold excess of the ligand AMG over the TAR RNA, indicating that the equilibrium dissociation constant K_d is in the micro molar range [155]. Further addition of AMG doesn't lead to further changes in the peaks thus leading us to conclude that a stable complex of the HIV-2 TAR RNA with AMG is formed under sample conditions of 10 fold excess of AMG with TAR RNA.

In summary, the most significant chemical shifts changes upon binding of the TAR RNA to AMG are observed either for the base or for the ribose ^{13}C and ^1H resonances

of A22, U23, G26, G32, G34 and A35 (rendered in red). Less prominent chemical shift changes are observed for G21, U25, G28, C29, U31, G33, G36 and G43 (rendered in light blue) in Fig 3.2.

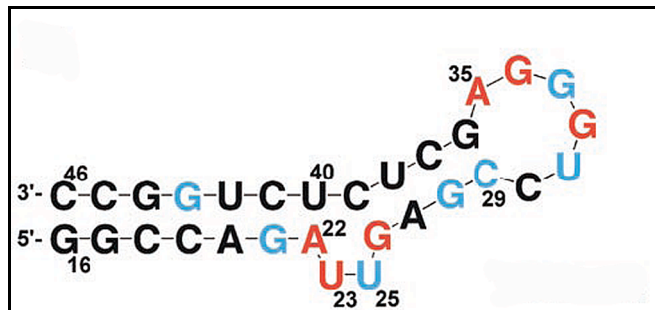


Fig 3.2 Chemical shift mapping of TAR-binding sites.

3.3 Sugar pucker and dihedral angles

Depending on the J coupling constants between H1' and H2' derived from a 2D-HCCH-E.COSY spectrum, residues of the HIV-2 TAR RNA were classified as being in either C3'-endo or C2'-endo conformation. Table 3.3 summarizes the sugar pucker restraints for the residues of the TAR RNA.

Residue	Sugar Pucker	Residue	Sugar Pucker
16	C3'-endo	32	Unrestrained
17	C3'-endo	33	C2'-endo
18	C3'-endo	34	Unrestrained
19	C3'-endo	35	Unrestrained
20	C3'-endo	36	C3'-endo
21	C3'-endo	37	C3'-endo
22	C3'-endo	38	C3'-endo
23	C2'-endo	39	C3'-endo
25	Unrestrained	40	C3'-endo
26	C3'-endo	41	C3'-endo
27	Unrestrained	42	C3'-endo
28	C3'-endo	43	C3'-endo
29	Unrestrained	44	C3'-endo
30	Unrestrained	45	C3'-endo
31	Unrestrained	46	C3'-endo

Table 3.3 Sugar pucker restraints imposed on the residues of the TAR RNA

41 non-experimental dihedral angle restraints defining the backbone angles α , β , ζ and ϵ within a range of $\pm 60^\circ$ for the lower stem residues 16 to 21 and 41 to 46 were imposed based on typical NOEs associated with an A-form helix.

3.4 Identifying the binding sites of AMG to TAR RNA

The intermolecular NOESY spectrum reveals intimate contacts of AMG with the residues A22, U23, U25 of the bulge and A35 of the loop of the HIV-2 TAR RNA. Table 2.5.6 summarized the intermolecular NOE contacts revealed by the NOESY spectra. This result is in agreement with the chemical shift mapping, which identifies A22, U23, U25 and the loop nucleotides G32, G34 and A35 as the residues being most affected by the binding of AMG to the TAR RNA. The atypical C8/H8 chemical shifts of A35 in the free TAR RNA implies that the adenosine base is partially projected out of the loop, thus occupying an optimal position for interacting with a ligand bound to the bulge on the side of the upper stem major groove. In the bulge, the largest chemical shift changes are observed for A22 and U23. U23 stacks on A22 in the free TAR RNA, while in this complex it is involved in several contacts with AMG. The A22–U40 base pair, which is stably formed at low temperature in the free HIV-2 TAR RNA, is disrupted in the complex, as confirmed by the failure to observe the U40 imino resonance even at 4°C. However, several NOEs indicate that both A22 and U40 remain positioned between G21/G26 and C39/C41, respectively.

A35 is the only loop residue that directly contacts the ligand, suggesting that the interaction of AMG with the loop must be part of a more complex network of interactions holding the complex together. Cooperative binding of AMG to both the loop and the bulge of the TAR RNA is supported by the chemical shift data, which indicate a K_d of comparable magnitude for both the bulge and the loop sites.

Moreover, while AMG inhibits the binding of Tat-derived peptides to the TAR RNA, linear analogues or smaller cyclic analogues of AMG do not show any biological activity [⁹¹], underlining the importance of a cyclic constrained structure of appropriate size for the cooperative binding to the bulge and loop regions of the RNA. Therefore, it is safe to conclude that one molecule of AMG binds simultaneously to the loop and the bulge of the RNA.

A medium-sized chemical shift change of G43–C1' upon titration of AMG, together with NOEs between the aminoglycoside analogue and both G43–H1' and G44–H4', reveal an additional binding site of AMG in the minor groove of the RNA lower stem. This finding is not surprising, as structural studies of the TAR/neomycin B complex [¹⁵⁶] had identified the lower stem minor groove as a binding site for positively charged aminoglycosides.

3.5 Field induced residual dipolar couplings

Under a given high magnetic field (B), the TAR RNA assumes a sufficient level of molecular alignment to allow measurement of f_i RDCs (D) as contributions to normally observed scalar couplings ($J+D$). The magnitudes of observed f_i RDCs depend quadratically on the magnetic field strength (B^2), and on the principal values of the magnetic susceptibility tensor (χ_{ii} $i = \{x, y, z\}$). The magnetic susceptibility tensor for the TAR RNA is dominated by the diamagnetic susceptibilities of aromatic base groups ($\chi_{ii}^{(base)}$). Corresponding χ -tensors tend to be close to axially symmetric ($\chi_{yy} \sim \chi_{xx}$) with principal values (χ_{ii}) that increase approximately linearly with the total number of bases due to constructive addition of base susceptibilities arising due to the fact that stacking interactions favor coplanar arrangement of base planes in nucleic acids.

Comparing experimental χ_{ii} values determined for an unknown system (in our case the TAR RNA of the TAR/AMG complex) with the corresponding values expected for a monomer TAR RNA can provide new insights into the stoichiometry of the RNA component of the TAR RNA/AMG complex. The principal value of χ_{ii} for the TAR RNA/AMG system is derived from the values of the f_i RDCs. The following points need to be borne in mind when elucidating the χ_{ii} in nucleic acids in general.

- a) Base interaction vectors are perpendicular to their own principal anisotropy ($\chi_{zz}^{(base)}$) and therefore the vector is preferentially positioned within the $\chi_{yy} - \chi_{xx}$ plane of the principal axis system.
- b) In nucleic acids, the axially symmetric tensors are very close in value i.e. ($\chi_{yy} \sim \chi_{xx}$) thus, many RDC values measured for interaction vectors in the $\chi_{yy} - \chi_{xx}$ plane will provide a good estimate for the value of D_{yy} .

We also assume that the presence of the AMG in ten-fold excess concentration as the TAR RNA, does not contribute to the alignment of the TAR RNA in the magnetic field.

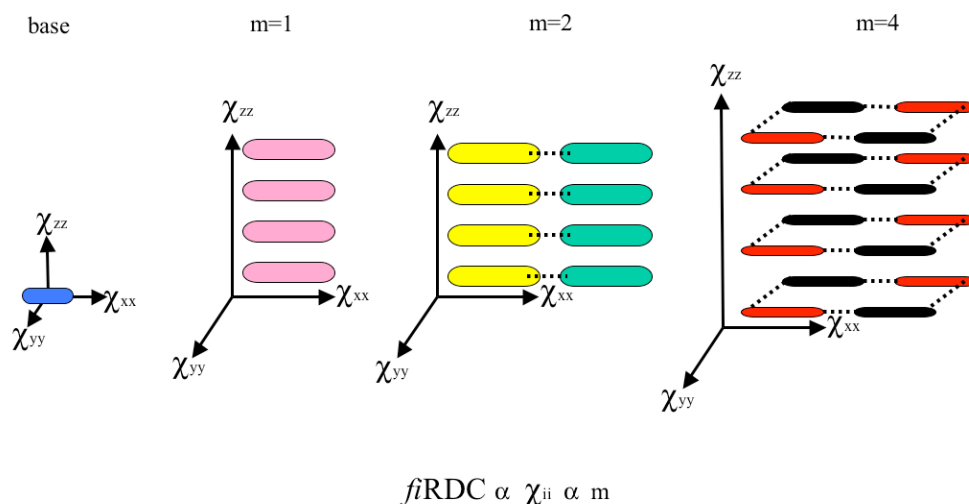


Fig 3.5. (a) Sensitivity of $fIRDC$ to RNA stoichiometry “ m ”. Base pairs are shown as dotted lines between colored rectangles in multimers [134].

An *awk* script written in-house* (Appendix C), allowed us to determine the theoretical values arising from a given model and compare this with the experimental value determined for the system.

The experimental values of the field induced residual dipolar couplings and the associated experimental error for each coupling are summarized in Table 3.4. The same values are graphically depicted with error bars in Fig 3.5 (b). These $fIRDC$ s were employed in the final round of refinement during structure calculations of the TAR RNA/AMG complex.

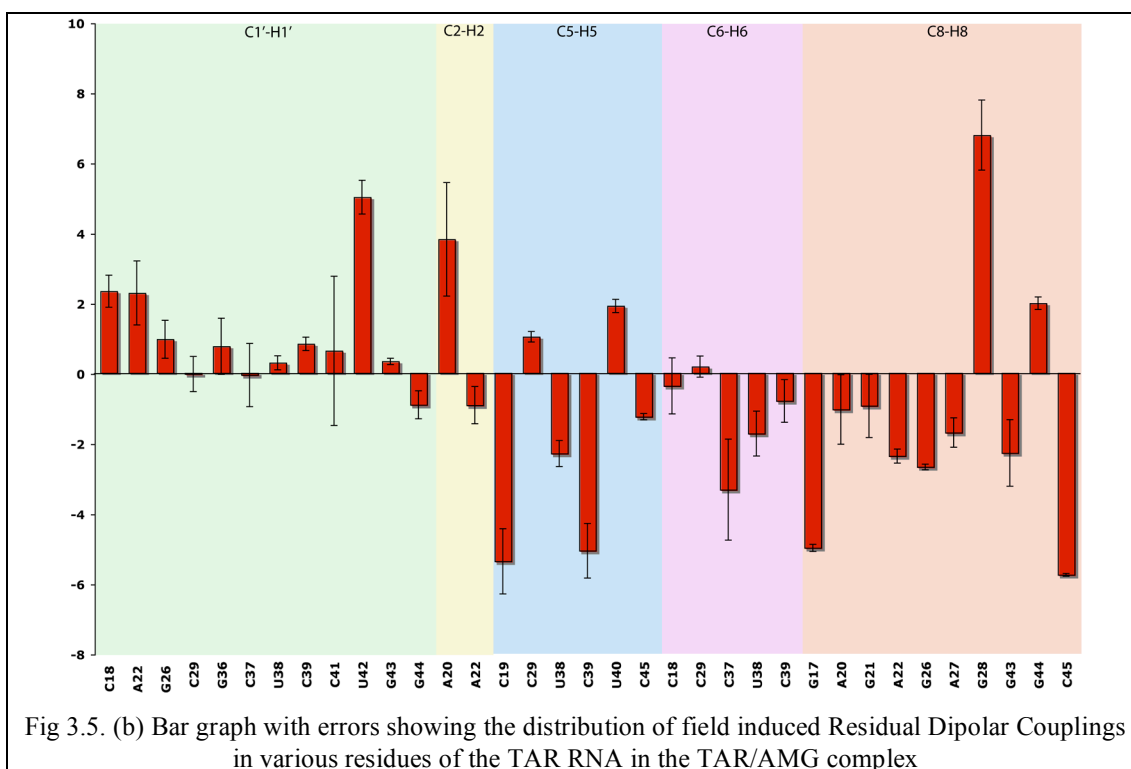
A few dipolar couplings of the base C–H atom pairs measured at 800 MHz assume values between 5 and 7 Hz. These figures exceed by far those expected for the TAR RNA aligned in the magnetic field due to the magnetic susceptibility anisotropy of the bases [135]. The largest dipolar coupling tensor is expected for coaxial stacking of the upper and lower stems and would correspond to a D_{zz} of 8 Hz and a rhombicity R between 0 and 0.1 in an 800 MHz field. For such a tensor, the largest base C–H dipolar coupling should not exceed 4 Hz. The high values found for the TAR RNA/AMG complex suggest that the TAR RNA dimerizes in the presence of the ligand AMG. This conclusion is also in agreement with the T_1 relaxation measurements.

*Script written by Dr. Laurent Verdier, now at Laboratoire de Pharmacognosie, UMR 6013, Moulin de la Housse, Cedex, France

*Script modified by Dr. Victor M. Sanchez Pedregal now at Universidad de Santiago de Compostela, Facultad de Quimica Organica, Campus Universitario Sur, Spain

Residue	D (C1'-H1')	D (C2-H2)	D (C5-H5)	D (C6-H6)	D (C8-H8)
G17					-4,99 ± 0,10
C18	2,33 ± 0,46			-0,37 ± 0,80	
C19			-5,37 ± 0,93		
A20		3,81 ± 1,62			-1,04 ± 0,99
G21					-0,94 ± 0,90
A22	2,28 ± 0,91	-0,92 ± 0,53			-2,37 ± 0,20
G26	0,96 ± 0,54				-2,68 ± 0,08
A27		0.13 ± 0.47			-1,70 ± 0,42
G28					6,78 ± 1,00
C29	-0,03 ± 0,50		1,03 ± 0,15	0,18 ± 0,30	
G36	0,76 ± 0,80				
C37	-0,06 ± 0,90			-3,33 ± 1,44	
U38	0,29 ± 0,20		-2,30 ± 0,37	-1,73 ± 0,64	
C39	0,83 ± 0,19		-5,07 ± 0,78	-0,80 ± 0,61	
U40			1,91 ± 0,19		
C41	0,63 ± 2,13				
U42	5,01 ± 0,48				
G43	0,33 ± 0,09				-2,28 ± 0,95
G44	-0,91 ± 0,40				1,99 ± 0,18
C45			-1,25 ± 0,09		-5,76 ± 0,04

Table 3.5 Field Induced Residual Dipolar Couplings and associated Errors for the TAR RNA/AMG system.



3.6 NOE distance restraints

A total of 127 internucleotide and 320 intranucleotide distance restraints were extracted from a 3D ^{13}C -HSQC NOESY (mixing time 100 ms) and a base resolved/selective 3D ^{13}C -HSQC NOESY (mixing time 100 ms).

As described in the Materials and Methods section (Fig 2.5.5 (b)), an iterative assignment process for the intermolecular NOEs was adopted. The r.m.s.d of the upper stem-binding site, formed by AMG and the TAR RNA residues A22, U23, U25 and A35, and the r.m.s.d of the whole complex, excluding the loop residues C30 to A35, were evaluated for 20 energetically most favorable structures in all the cases discussed. These results are summarized in Table 3.6.

Run	R.M.S.D (Å)	
	Core	All*
1	3.17	2.01
2	3.63	2.31
3	3.04	2.00
4	3.66	2.44
5	2.99	1.91
6	3.55	2.45
7	2.54	1.85

Table 3.6 r.m.s.d values of different intermolecular NOE schemes considered in Fig 2.5.5 (b)

The upper stem binding site of the complex converged best for the seventh set of intermolecular NOEs. The average r.m.s.d value for the remaining calculations was ≈ 3 angstroms for the upper stem binding core and ≈ 2.25 Å for the whole complex excluding the loop, while in the favored seventh set, the average r.m.s.d value ≈ 2.5 Å for the upper stem binding core and ≈ 1.85 Å for the whole complex excluding the loop. An iterative assignment process further released some of the ambiguities in the seventh intermolecular NOEs data set during structure calculation to improve the convergence profile (≈ 2.15 Å for the upper stem binding core and ≈ 1.75 Å for the whole complex excluding the loop). A total of 55 intermolecular distance restraints were obtained from a ^{13}C -edited/ ^{12}C -filtered HSQC-NOESY spectrum acquired in D_2O with a mixing time of 200ms.

3.7 Structure calculations

The restraints used to arrive at the final structures are summarized in Table 3.7.

The XPLOR protocols (Appendix) were used to arrive at the ensemble of final structures of the TAR RNA/AMG complex. The fourteen best structures (energetically favorable), after the refinement with *f*iRDCs are superimposed in Figure 3.7 (a). The effect of the *f*iRDCs on the structures can be better appreciated by comparing the structures before and after the refinement (Figure 3.7 (b)).

Restraint statistics	
Distance restraints	Number of restraints
Internucleotide	127
Intranucleotide	320
Intermolecular	55
Total number of NOE restraints	502
Hydrogen bonds	72
Total distance restraints	574
Dihedral restraints	
Sugar puckers	
C3'-endo	40
C2'-endo	4
Non-experimental A-form	
α	10
β	11
ζ	10
ϵ	10
Total dihedral restraints	85
Field induced RDCs	37
Total number of restraints used	696

Table 3.7 Restraints used for structure determination of the TAR/AMG complex

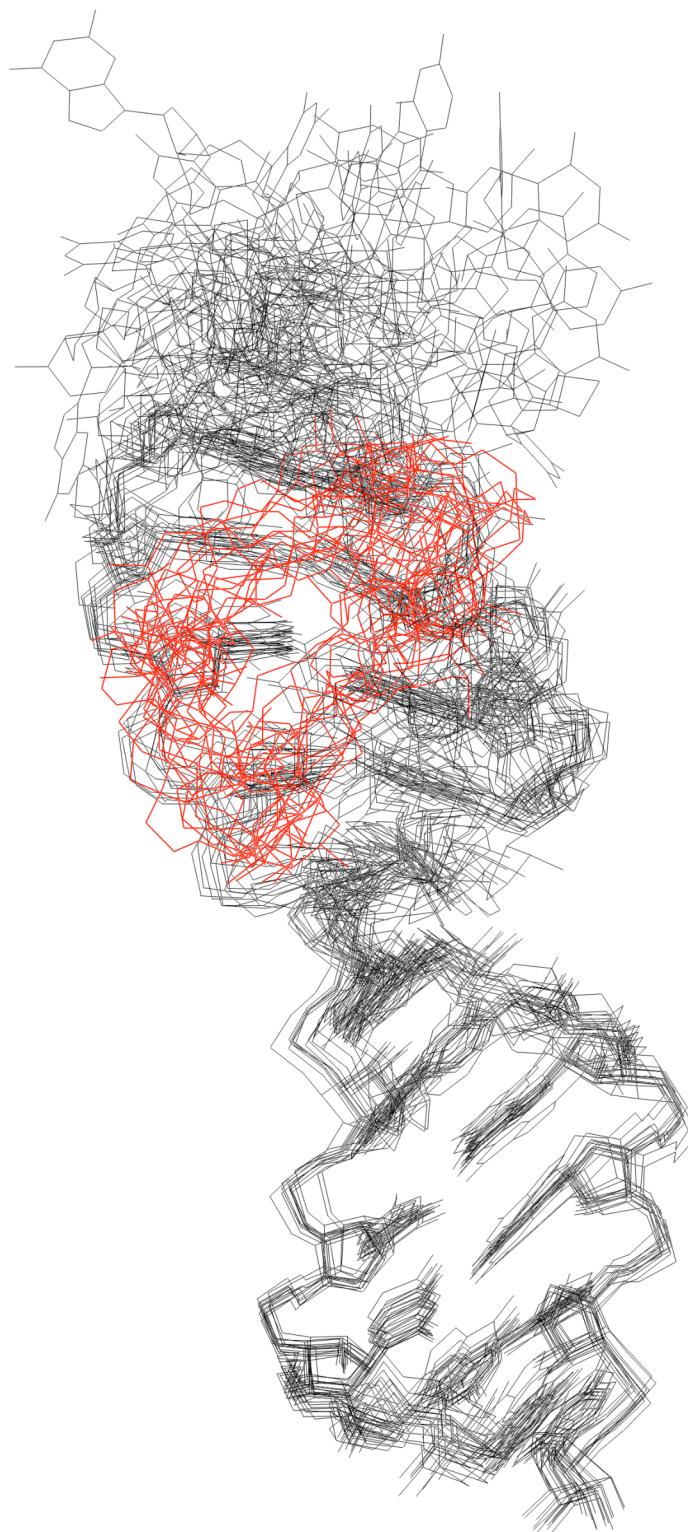


Figure 3.7 (a) The best-fit superposition to the average structure of the 14 best structures of all the residues except the hexanucleotide loop. The AMG is in red and the TAR RNA is colored black. Notice that the upper stem is not coaxially stacked atop the lower stem as it does in the TAR/argininamide structure.

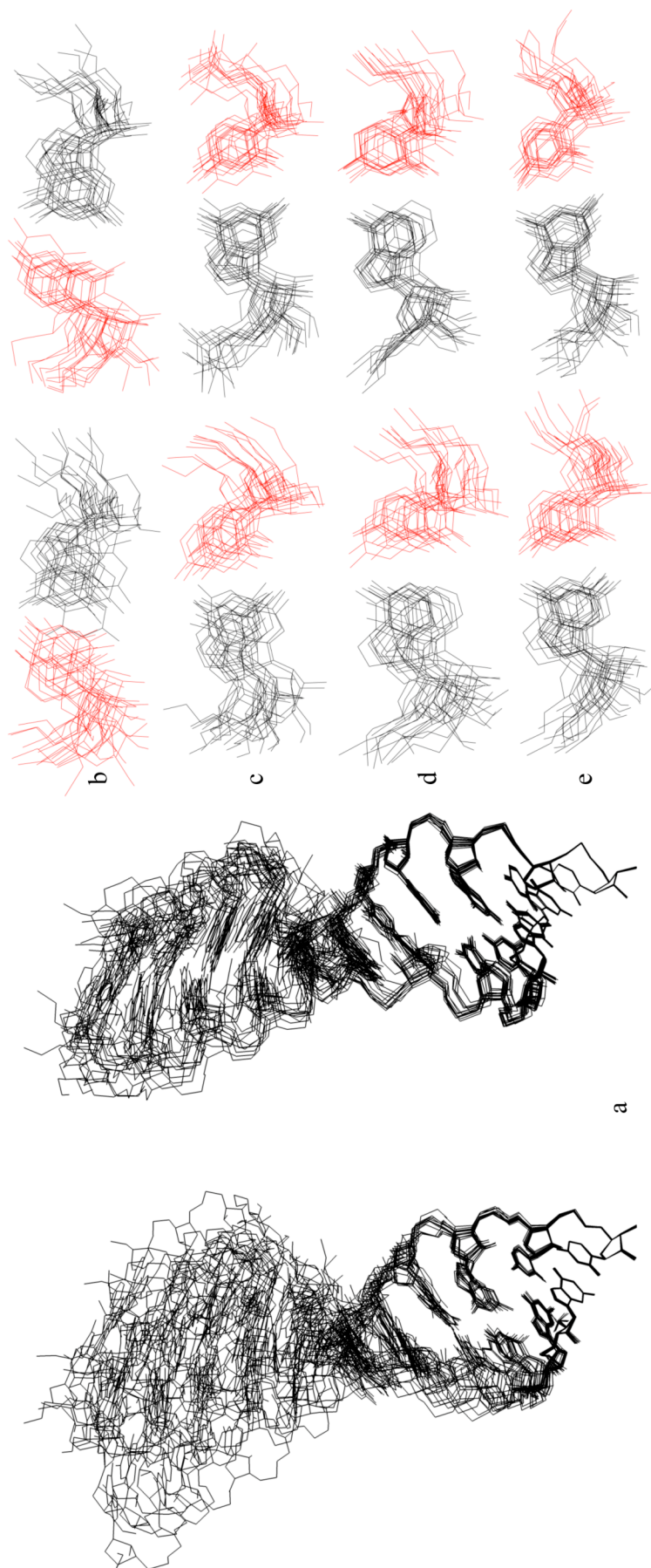


Fig. Results of refinement using field induced dipolar couplings of the RNA on the global and local structure of the TAR RNA - amino glycoside complex. A comparison of the ensembles in which only noe and dihedral angle restraints data were included during refinement (left) and an ensemble which was generated with the additional use of field induced dipolar couplings (right) shows a clear improvement in the global structure of the ensemble in spite of local structures being defined precisely only by employing distance and dihedral restraints (a). For clarity, the residues of the ligand, bulge and disordered loop have been omitted. Only heavy atoms of the terminal base pairs (residues 16 and 46) were superimposed to highlight the significance of RDC data in addition to classical NMR restraints for defining the global bend. In b, c, d and e an attempt has been made to compare the precision of base – pair residues in the upper stem of the TAR RNA (29-36, 28-37, 27-38, 26-39 respectively) upon the use of *f*RDCs. It is clear that RDCs also help in improving the local precision in addition to determining the global fold of the RNA.

3.8 The upper stem binding site

The structure calculation employing NOEs and dihedral angle restraints, conducted on a monomeric unit containing one TAR and one ligand molecule binding to the upper stem, resulted in a well-defined structure in both stems but in a poor definition of the relative orientation of the two stems [root mean square deviation (r.m.s.d.) for the lower stem, 0.5; r.m.s.d. for the upper stem, 0.8]. The structure refinement with the C-H \hat{r} RDCs positioned the two stems at a relative angle of 60°. The 14 lowest-energy structures of the monomeric unit are shown in Figure 3.7 (a). The RNA is well defined with a total heavy atom r.m.s.d. (excluding the loop residues) of 0.92. The A-form helix structure of the lower (nucleotide 16–21, 41–46) and upper (nucleotide 26–29, 36–39) stems is conserved. The upper stem helix is, however, slightly distorted, with the α and γ angles varying from the typical A-form gauche⁻ and gauche⁺ conformations to the trans conformation for residues A27–C29 and C39. AMG binds to the bulge and the loop residues on the major groove side of the upper stem. Owing to the 4-fold symmetry of its resonances and the fast kinetics of complex dissociation, the structure of AMG in the complex cannot be determined. However, the salient features of the interaction between the aminoglycoside analogue and the TAR RNA are well defined and the r.m.s.d. of the whole complex excluding the loop residues 30–34 is 1.2 Å. The backbone of the 22–25 stretch opens up to allow AMG to reach the major groove of the RNA upper stem. The last base pair of the lower stem is disrupted: U40 stacks below C39; A22 stacks below G26 but tilts away from U40 to widen the major groove of the upper stem and make place for the ligand. Thus, in the TAR/AMG complex, A22 and U40 belong to the upper stem, while in free TAR they stack on the top of the lower stem G21–C41 base pair. The planarity of the G26–C39 base pair is slightly distorted, as indicated by the weakening of the G26 imino resonance upon complex formation. The π -orbital of the U23 and U25 bases entertain hydrophobic interaction with linker-4 and sugar-3/linker-2 of AMG, respectively. Such hydrophobic contacts resemble a well-known interaction motif in complexes of proteins with oligosaccharides, where the sugars often stack below aromatic side chains [¹⁵⁷].

The conformation induced by AMG at the bulge site of the TAR RNA, as well as the relative orientation of the two stems, differ from those found in the Tat/TAR complex.

Binding of AMG does not induce either the formation of the U23–A27–U38 base triple or the subsequent coaxial stacking of the upper and lower stems, as it is the case for the binding of the Tat mimetic argininamide [^{62; 82; 158; 159}]. Instead, the bulge nucleotides are all exposed to the solvent in the TAR RNA/AMG complex and the axes of the two stems form an angle of 60°. The relative position of the two domains is probably a consequence of the disruption of the A22–U40 base pair, of the tilting of A22 to accommodate the large ligand and of the electrostatic interactions of the ligand amino group with the phosphate of A22. A35 stacks on the top of the sugar-1 moiety of the ligand and is involved in a π -cation interaction with the positively charged amino group. This contact pulls A35 out of the loop, causing G34 to move on the top of G36 in all structures.

In 60% of the structures G34 stacks on the top of G36, while in the other 40% the two bases interact face-to-edge. The presence of an equilibrium between two conformations is compatible with the large line width of the G34–C8 resonance, suggesting that the exchange between the face-to-face and face-to-edge orientations of G34 and G36 occurs on an μs – ms time scale [^{154; 160}].

The positively charged amino groups of the four sugars of AMG interact with various functional groups of the RNA. The amino group of sugar-1 is involved, as already mentioned, in a π -cation interaction with A35 and in long-range electrostatic interactions with the O2 of U25. The amino group of sugar-3 is directed towards the major groove carbonyls of the A27–U38 and G28–C39 base pairs in all structures but two, while the amino group of sugar-5 is close to the phosphate backbone of A22 and/or to the O4 of U23. Finally, no consistent contact is found for the amino group of sugar-7. These results are in agreement with the failure to detect H-bonding of the amino groups of AMG to the phosphate backbone of the RNA by the J_{NP} scalar coupling measurement, as most of the amino groups of the ligand contact base functional groups rather than backbone phosphates of the RNA.

3.9 Definition of the second binding site and of the dimer geometry

Similar to neomycin [¹⁵⁶], the AMG binds also to the minor groove of the lower stem of the TAR RNA. The chemical shift change of G43–C1' is maximum at a 1:1 ratio of TAR RNA: AMG, as opposed to the chemical shift changes in the bulge and loop region, which increase up to a 1:10 ratio. Thus, the binding constant of AMG to the minor groove of the TAR RNA is higher than that to the bulge and loop regions. This finding is confirmed by the analysis of the PACE data (discussed later). Field induced Residual Dipolar Couplings and relaxation data suggest that TAR RNA dimerizes during the complex formation.

The presence of only one set of resonances in the complex and the absence of NOEs between two TAR RNA moieties further suggests that the dimer is symmetric and that the TAR RNA dimerizes via the presence of one or more AMG molecules between the two dimerizing TAR RNAs.

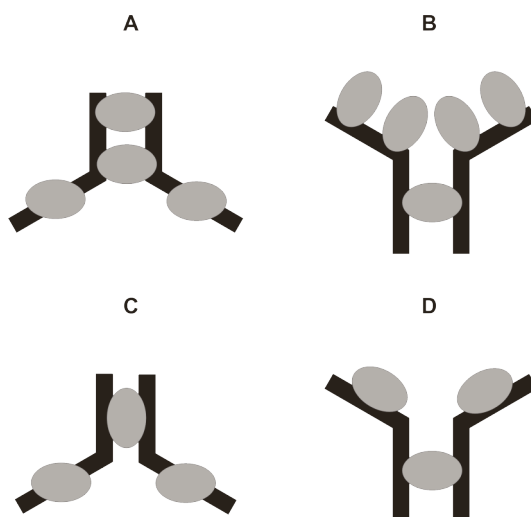


Figure 3.9 (i) Schematic representations of the association of TAR RNA with the aminoglycoside analogue AMG and of dimer formation.

The *f*_iRDC data suggests association of two TAR/AMG units with two stems of the two RNA molecules in a roughly parallel orientation and the other two stems diverging from each other. Further, the data is compatible with a dimer structure where lower stems are parallel to each other and upper stems diverge at an angle of about 60°. This geometry is compatible with the tensor derived from the systematic grid search which revealed a minimum at $D_{zz} = 7$ Hz and $R = 0.55$. The alignment tensor of the given model can be back calculated using the awk script provided in

APPENDIX C. These constraints on the dimer can be satisfied in four different scenarios each of them depicted in Figure 3.9 (i). In the first case (a), three molecules of AMG bind to the loop, the bulge and lower stem of the TAR RNA and the RNA dimerizes with the two upper stems parallel to each other. Case (b) considers two molecules of AMG binding to the loop and the bulge of the TAR RNA and the RNA dimerizes with the two lower stems roughly parallel to each other with one molecule of AMG intervening between the two dimerizing RNAs thus giving rise to complex formation with 2 TAR RNAs and 5 molecules of AMG. In case (c), one molecule of AMG binds to the upper stem major groove simultaneously contacting both the bulge and the loop of the TAR RNA, while a second molecule of AMG binds to the lower stem minor groove and the RNA dimerizes with the two upper stems parallel to each other. The final case (d) considers the binding of one molecule of AMG to the upper stem major groove simultaneously contacting the bulge and the loop of each of the TAR RNAs, while the second molecule of AMG contacts the lower stem minor groove of the TAR RNA. The RNA dimerizes with the two lower stems oriented roughly parallel to each other with one molecule of AMG between the parallel lower stem minor grooves of the RNAs (2 TAR RNAs: 3 AMG). Structure calculations of all the four cases investigated, resulted in very high energy structures due to the distortion of AMG in the dimer complex in case (a) and (c). Only case (b) and (d) were found to be consistent with the observed intermolecular NOEs and did not result in distortion of AMG and high-energy structures. However, case (b), which includes 5 molecules of AMG binding to two molecules of TAR RNA can be excluded based on the following reasons:

- (i) Residue A35 is the only nucleotide of the hexanucleotide loop showing intermolecular NOEs to AMG indicating that the binding of AMG to this residue must be a part of a more complex network of interactions
 - (ii) Chemical shift data indicate that AMG binds to the bulge and the loop of the RNA with the same K_d , suggesting cooperative binding to both the upper stem binding site and the lower stem binding site.
 - (iii) Smaller cyclic analogues or linear analogues of AMG do not show any detectable binding to the TAR RNA [⁹¹], emphasizing on the relevance of cooperative binding to multiple sites.
 - (iv) Entropic contribution strongly disfavors case (b) with respect to case (d).
- Thus, case (b) was excluded from further consideration and case (d) was identified as

the preferred model for dimer formation through the aminoglycoside bound to the lower stem minor groove as the only one consistent with the NOE data.

To arrive at an ensemble of dimers that is compatible with our interpretation of the residual dipolar coupling data, relaxation and biochemical results i.e. a dimer model where the TAR RNA undergoes dimerization through one molecule of AMG bridging the lower stems of two molecules of TAR RNA, the following approach was adopted. The monomer entity (TAR/AMG complex) was predisposed by rigid body dynamics (see protocol in APPENDIX for details) in an orientation that was compatible with all the experimental data (**case d** as described above). Then, a third molecule of AMG was docked to the minor groove of the two TAR/AMG molecules, guided by NOE restraints of AMG to G43-H1' and G44-H4'. The docking was performed by MD simulations in explicit water solvent [¹⁶¹]. A brief description of the MD simulation is summarized below: Simulations were performed with X-PLOR [^{143; 144}] by using the same force fields for the RNA and AMG as described above with an electrostatic cutoff of 12 Å. The complex is surrounded by a layer of TIP3 water molecules [¹⁶²], and the solvent is briefly equilibrated. Before the dynamics began, the system was equilibrated by 500 steps of Powell minimization of only the protons of the complex while the heavy atoms are kept fixed, another 100 steps of minimization to equilibrate the water molecules while the complex is kept fixed, 500 steps of minimization while the complex is harmonically restrained and finally 500 steps of free minimization of the TAR/AMG in water. This ensured that the TAR RNA/AMG complex was well equilibrated in the water layer (solvent) generated by TIP3 protocol. The MD protocol consisted of an initial heating stage (200 steps at 100K, 200K, 300K, 500K and 600K with a time step of 0.002 ps) followed by refinement at 600K (2000 steps, 0.003 ps). Finally, the system was cooled from 600K to 25K with the temperature decreasing in steps of 25K (200 steps, 0.003 ps at each step). The MD protocol ended with 200 steps of Powell minimization.

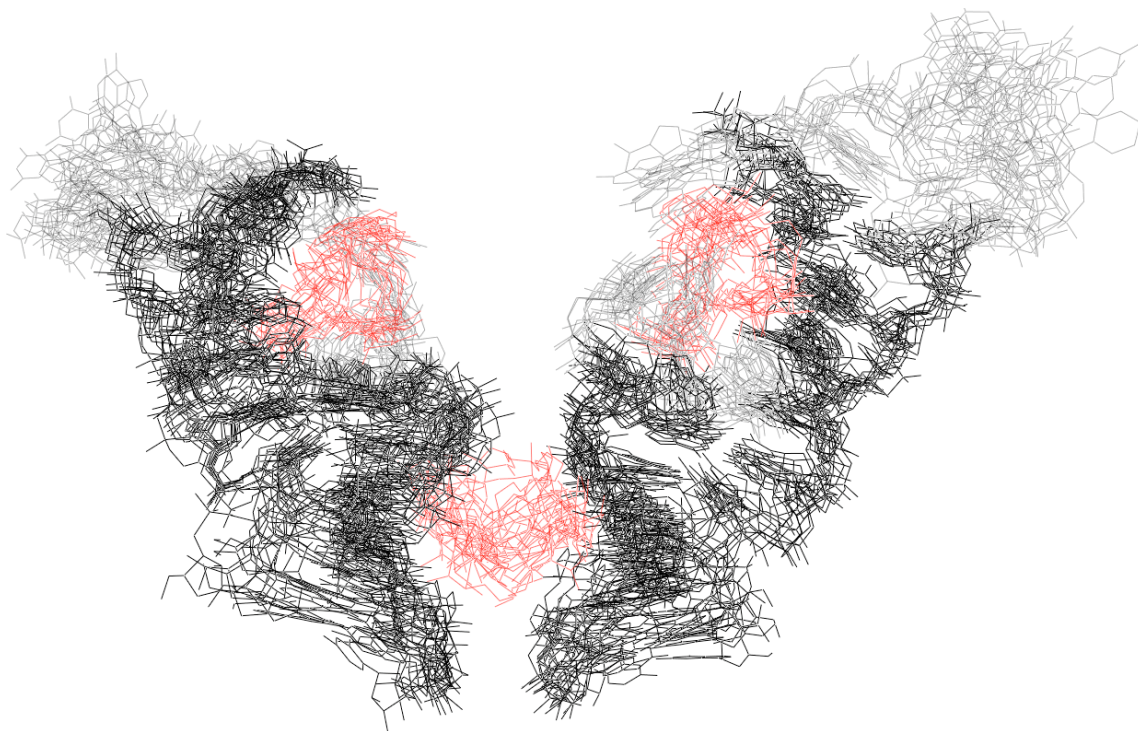


Figure 3.9 (ii) Ensemble of 10 structures of the dimer of HIV-2 TAR RNA (black) in conjunction with 3 molecules of AMG (red). The heavy atoms of residues 16 to 29, 36 to 46 and AMG have been superimposed. The bulge and loop residues of the TAR are shown in grey.

The ensemble of models shown in Figure 3.9 (ii) are compatible with all experimental data, including NOEs, relaxation rates, *f*_iRDCs and PACE gel electrophoresis. In this model, the aminoglycoside analogue AMG binds almost symmetrically to the minor groove of the lower stems of the two TAR RNA units. One amino group enjoys electrostatic interactions with the negative spot created by the phosphate groups of G21, A22 and U23, while another amino group closely contacts the 2'-hydroxyl of C45. Two sugars and the intervening linker of AMG penetrate the minor groove of each RNA molecule at the G43–C45 site.

3.10 RNase digestion and PACE gel analysis

In agreement with the identification of A22, U23, U25 in the bulge and G32, G34 and A35 as the sites being most affected by the binding of AMG to TAR RNA by Chemical Shift mapping, digestion with RNase T1 and with RNase A suggests that loop residues G32 to G34 and residue C30 respectively, are protected by the binding of AMG to TAR RNA. The RNase foot printing analysis can be followed on Figure 3.10(a). These protection assays of the TAR RNA / AMG complex were carried out with 10 μM 5'- ^{32}P -labeled HIV-2 TAR RNA at 277 K. The digestion experiments were analyzed by gel electrophoresis.

As is evident from the digestion experiments, the ligand AMG protects residues G32, G34 and C30 from digestion by the RNase. This is in agreement with the chemical shift mapping, which identifies G32, G34 and A35 as the sites being most affected by the binding of AMG to the TAR RNA.

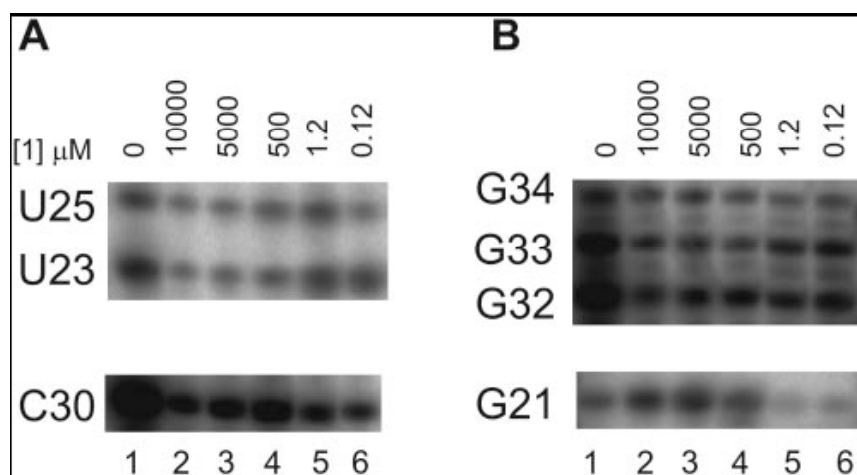


Figure 3.10 (a) RNase footprinting shows digestion with (A) RNase A (selective for CYT and URI residues). Lane 1, control without RNase, Lane 2, AMG (10 mM); Lane 3, AMG (5 mM); Lane 4, AMG (500 μM); Lane 5, AMG (1.2 μM); Lane 6, AMG (0.12 μM). (B) Digestion with RNase T1 (selective for GUA residues). Lane 1, control without RNase; Lane 2, AMG (10 mM); Lane 3, AMG (5 mM); Lane 4, AMG (500 μM); Lane 5, AMG (1.2 μM); Lane 6, AMG (0.12 μM).

The PACE gel analysis allowed us to test different binding models of the TAR RNA and AMG. We considered a total of 6 different models as described in the Methods (Section 2.4). The retardation distance of the TAR RNA binding to AMG at various concentrations (0, 2, 5, 7, 10, 20, 70, 100, 200 and 500 μM) can be elucidated from the autoradiogram as shown in Figure 3.10(b).

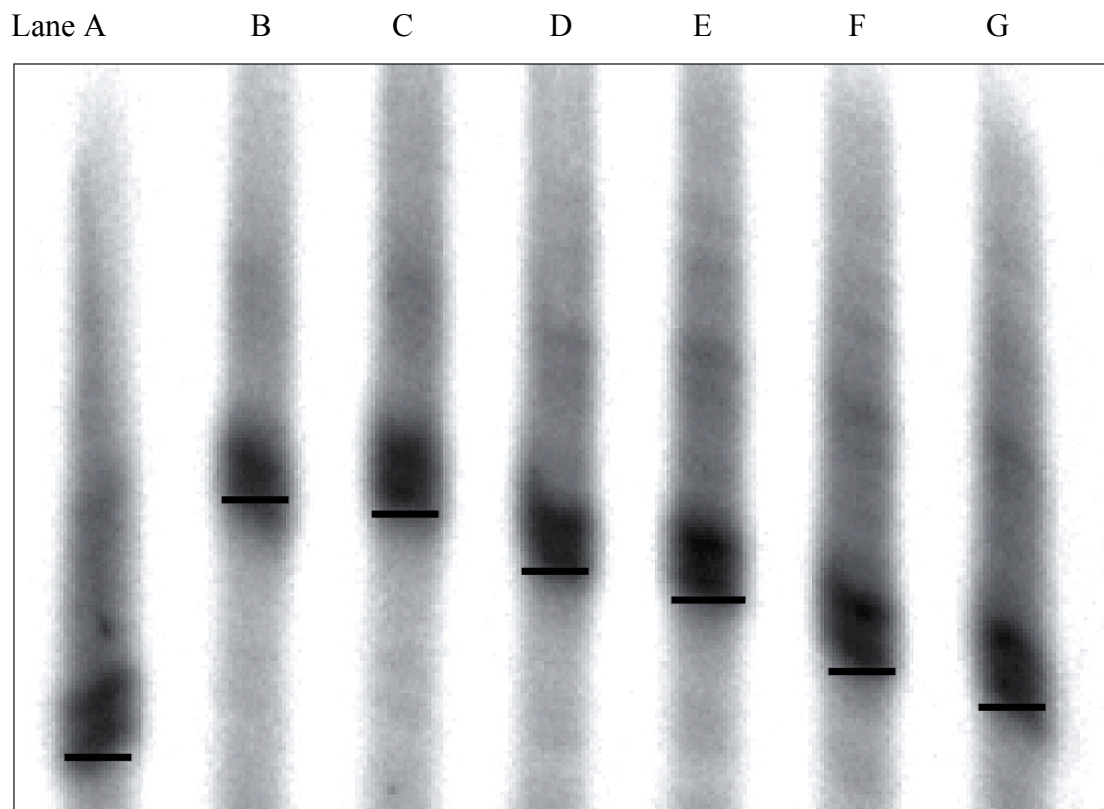


Figure 3.10 (b) Autoradiogram of 5'- ^{32}P labeled HIV-2 TAR RNA binding to different concentrations of AMG. Lane A depicts unbound RNA while lanes B, C, D, E, F and G show TAR RNA in the presence of 500, 200, 100, 70, 20 and 10 μM AMG respectively. The bars enable us to elucidate the retardation distance in each lane, data that is subsequently used for the calculation of binding stoichiometry.

The experimentally determined fraction of the RNA bound (ϕ) are summarized in Table 3.10.

Concentration [AMG]	ϕ
500 μM	1.00
200 μM	0.96
100 μM	0.73
70 μM	0.63
20 μM	0.36
10 μM	0.20
7 μM	0.18
2 μM	0.08
0 μM	0.00

Table 3.10 Experimental values of ϕ as measured from PACE gel.

The six theoretical binding models can be analyzed as follows:

For the sake of convenience, the following conventions were adopted.

The AMG concentration was designated as [A], the TAR RNA concentration is [T]. [C₁] and [C₂] are the concentrations of the complexes formed, K_{d1} and K_{d2} are the constants.

Case (i)

One molecule of the TAR RNA binds to one molecule of AMG.



$$\text{Where } K_d = [T][A]/[C_1]$$

$$\text{Rearranging, we have } [C_1] = [T][A]/K_d \dots (a)$$

The fraction of bound RNA

$$\phi = \text{Concentration of complex/Total concentration of RNA} = [C_1]/\{[T] + [C_1]\}$$

Substituting for [C₁] from (a) we obtain

$$\phi = [A]/\{K_d + [A]\} \dots \text{Equation 1}$$

Case (ii)

One molecule of the TAR RNA binds to two molecules of AMG



$$\text{Where } K_d = [T][A]^2/[C_1]$$

$$\text{Rearranging, we have } [C_1] = [T][A]^2/K_d \dots (b)$$

The fraction of bound RNA

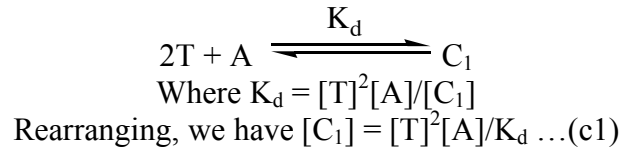
$$\phi = \text{Concentration of complex/Total concentration of RNA} = [C_1]/\{[T] + [C_1]\}$$

Substituting for [C₁] from (b) we obtain

$$\phi = [A]^2/\{K_d + [A]^2\} \dots \text{Equation 2}$$

Case (iii)

Two molecules of the TAR RNA binds to one molecule of AMG



The total concentration of RNA (C) is given by $[C] = 2[C_1] + [T] \dots(c2)$

Keeping in mind that each $[C_1]$ contributes 2 TAR moieties.

Combining (c1) and (c2), we obtain a quadratic equation as follows:

$$\{2[T]^2[A]/K_d\} + [T] - [C] = 0$$

which can be solved for $[T]$. One of the two solutions can be discarded, as concentration cannot be negative, thus leading to

$$[T] = (-K_d + \sqrt{K_d^2 + 8[C][A]K_d})/4[A] \dots(c3)$$

The equation for the fraction of bound RNA (ϕ) can be arrived at using the above equations and the final equation after simplification yields:

$$\phi = (4[C][A] + K_d - \sqrt{K_d^2 + 8[C][A]K_d}) / (4[C][A] - K_d + \sqrt{K_d^2 + 8[C][A]K_d}) \dots \text{Equation 3}$$

Case (iv)

One molecule of the TAR RNA binds to two molecules of AMG with different K_d



Where $K_{d1} = [A][T]/[C_1]$

Rearranging, we have $[C_1] = [T][A]/K_{d1} \dots(d1)$

Where $K_{d2} = [A][C_1]/[C_2]$

Substituting for $[C_1]$ from (d1) and rearranging, we have

$$[C_2] = [T][A]^2 / K_{d1}K_{d2} \dots(d2)$$

Further, it has to be borne in mind that the total distance migrated of the bound form of the RNA is a combination of the distance migrated by C_1 and C_2 . Though the individual distances migrated by the two complexes cannot be determined accurately,

it is safe to say that bound distance of C_1 is less than the bound distance C_2 and a factor Z (value between 0 and 1) can be introduced that is defined as $Z = \frac{D_{C1} - D_F}{D_{C2} - D_F}$

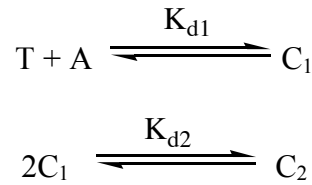
where D_{C1} and D_{C2} are the distance traveled by the complexes C_1 and C_2 respectively and D_F is the distance traveled by the free TAR RNA. The value of Z is defined within the boundary condition of 0 and 1 during the curve fitting/ optimization procedure.

The fraction of bound RNA ϕ is given by

$$\phi = \frac{[A]^2 + [A]Z K_{d2}}{\{K_{d1}K_{d2} + K_{d2}[A] + [A]^2\}} \dots \text{Equation 4}$$

Case (v)

One molecule of the TAR RNA binds one molecule of AMG to give complex C_1 with constant K_{d1} . Subsequently C_1 dimerizes to give C_2 with constant K_{d2} .



$$\text{Where } K_{d1} = \frac{[A][T]}{[C_1]}$$

$$\text{Rearranging, we have } [C_1] = \frac{[T][A]}{K_{d1}} \dots (e1)$$

$$\text{Where } K_{d2} = \frac{[C_1]^2}{[C_2]}$$

$$\text{Substituting for } [C_1] \text{ from (e1) and rearranging, we have } [C_2] = \frac{[T]^2[A]^2}{K_{d1}^2 K_{d2}} \dots (e2)$$

The total concentration of RNA (C) is given by $[C] = [T] + [C_1] + 2[C_2]$ keeping in mind that each $[C_2]$ contributes 2 TAR RNA moieties.

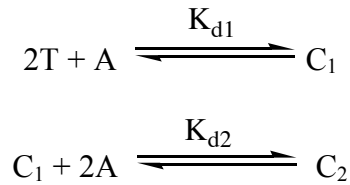
Also, like in case (iv), we define a constant Z , that takes into account the differential migration rates of C_1 and C_2 .

The equation for the fraction of bound RNA (ϕ) can be arrived at using the above results and the final equation after simplification yields:

$$\phi = \frac{((K_{d2}*(K_{d1}+[A])^2+4*C*[A]^2-(K_{d1}+[A])*sqrt(K_{d2}^2*(K_{d1}+[A])^2+8*C*K_{d2}*[A]^2))+Z*(2*[A]*sqrt(K_{d2}^2*(K_{d1}+[A])^2+8*C*K_{d2}*[A]^2)-(K_{d1}+[A])^2*K_{d2}*[A]))}{([A]+K_{d1})*sqrt(K_{d2}^2*(K_{d1}+[A])^2+8*C*K_{d2}*[A]^2)-K_{d2}*(([A]+K_{d1})^2+4*C*[A]^2)} \dots \text{Equation 5}$$

Case (vi)

Two molecules of the TAR RNA bind one molecule of AMG to give complex C_1 with constant K_{d1} . Subsequently C_1 recruits two molecules of AMG to give C_2 with a constant K_{d2} . C_2 is a complex of 2 TAR RNA molecules and three AMG.



$$\text{Where } K_{d1} = [A][T]^2/[C_1]$$

$$\text{Rearranging, we have } [C_1] = [T]^2[A]/K_{d1} \dots (f1)$$

$$\text{Where } K_{d2} = [C_1][A]^2/[C_2]$$

$$\text{Substituting for } [C_1] \text{ from (f1) and rearranging, we have } [C_2] = [T]^2[A]^3/K_{d1}K_{d2} \dots (f2)$$

The total concentration of RNA (C) is given by $[C] = [T] + 2[C_1] + 2[C_2]$ keeping in mind that $[C_1]$ and $[C_2]$ contribute 2 TAR RNA moieties each. As in case (iv) and (v), we define a constant Z, which takes into account the differential migration rates of C_1 and C_2 . The equation for the fraction of bound RNA (ϕ) can be arrived at using the above results and the final equation after simplification yields:

$$\phi = \frac{(4 * C * [A]^3 * ([A]^2 + K_{d2}) - [A]^2 * \sqrt{(K_{d1}^2 * K_{d2}^2 + 8 * C * [A] * K_{d1} * K_{d2} * ([A]^2 + K_{d2}))} + K_{d1} * K_{d2} * [A]^2) + Z * (4 * C * K_{d2} * [A] * ([A]^2 + K_{d2}) - K_{d2} * \sqrt{(K_{d1}^2 * K_{d2}^2 + 8 * C * [A] * K_{d1} * K_{d2} * ([A]^2 + K_{d2}))} + K_{d1} * K_{d2} * [A]^2))}{(4 * C * [A] * ([A]^2 + K_{d2})^2 - K_{d1} * K_{d2} * ([A]^2 + K_{d2}) + (K_{d2} + [A]^2) * \sqrt{(K_{d1}^2 * K_{d2}^2 + 8 * C * [A] * K_{d1} * K_{d2} * ([A]^2 + K_{d2}))})} \dots \text{Equation 6}$$

The experimental data can be fitted to each of the equations derived above (Equation 1, 2, 3, 4, 5 and 6) to assess the quality of fit of the assumed model with respect to the experimental data.

To test how each of the equations derived for a particular model fits with the experimentally derived data as summarized in Table 3.10 (a), ϕ was defined as a function of the concentration of the ligand AMG in accordance with the equations discussed above. These were then fit with a linear least square fitting procedure in MATLAB and the results are discussed in Figure 3.10(c). Scripts used to simulate and plot the functions are available in APPENDIX C.

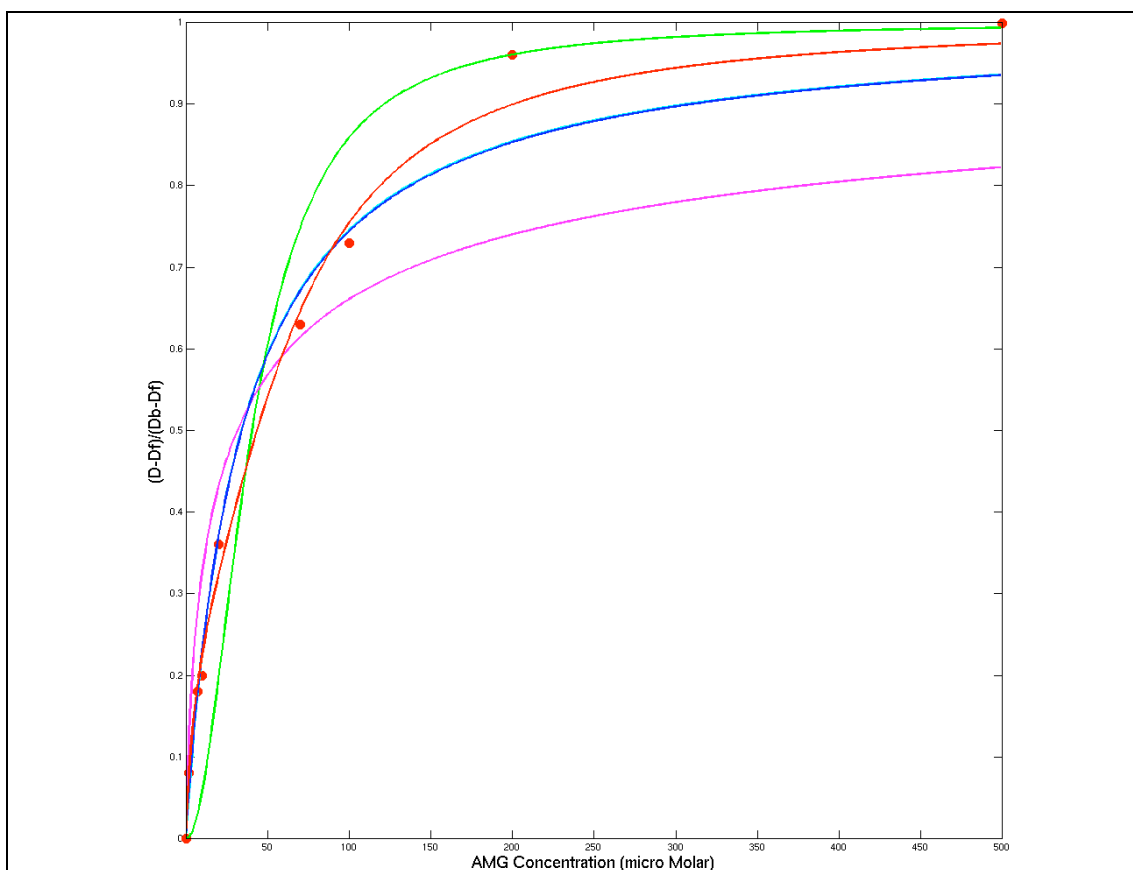


Figure 3.10 (c) Binding stoichiometry. The function $(D-D_F)/(D_B-D_F)$ is plotted versus the concentration of the aminoglycoside analogue AMG in each lane of the PACE gel. D is the retardation distance in each lane, D_F is the retardation distance of the TAR RNA without AMG and D_B is the retardation distance of the TAR RNA completely bound to AMG (concentration of AMG = 500 μ M). The function $(D-D_F)/(D_B-D_F)$ can be expressed in terms of molar fractions of bound and free RNA, according to the procedure described by Cilley and Williamson [100]. Different binding models were assumed and the theoretical dependence of $(D-D_F)/(D_B-D_F)$ on the concentration of AMG was calculated in each case from the molar fractions assuming constant concentration of AMG in each lane. The experimental data were fitted to the theoretical functions using MATLAB to derive the binding stoichiometry for the TAR/AMG complex. Six different models were considered: (i) one molecule of the TAR RNA binds one molecule of AMG (cyan line); (ii) one molecule of the TAR RNA binds two molecules of AMG with the same K_d (green line); (iii) two molecules of the RNA bind one molecule of AMG (magenta); (iv) one molecule of the TAR RNA binds two molecules of AMG with different K_d (blue line very close to cyan line); (v) one molecule of the TAR RNA binds one molecule of AMG ($\text{TAR} + \text{AMG} \xleftarrow{K_{d1}} \text{C}_1$); subsequently the TAR/AMG complex dimerizes ($2\text{C}_1 \xleftarrow{K_{d2}} \text{C}_2$) (close to the cyan line, not shown); (vi) two molecules of the RNA bind one molecule of AMG ($2 \text{TAR} + \text{AMG} \xleftarrow{K_{d1}} \text{C}_1$); subsequently two molecules of AMG are recruited by the RNA dimer for a total of two RNA molecules and three molecules of AMG ($\text{C}_1 + 2\text{AMG} \xleftarrow{K_{d2}} \text{C}_2$). (red line; $K_{d1} = 20 \mu\text{M}^2$; $K_{d2} = 3 \text{mM}^2$) The best fit is obtained for case (vi), which supports the stoichiometry inferred from the NMR data (r.m.s.d 0.084).

3.11 Structure statistics

Energy profile over best 14 structures (monomer)	
Average total energy (kcal/mol)	-323.26 ± 12.12
Average distance violation energy (kcal/mol)	16.80 ± 1.74
Average <i>sani</i> violation energy (kcal/mol)	33.71 ± 3.09
r.m.s.d profile over best 14 structures (monomer)	
Experimental restraints	
All distance / NOE restraints (Å)	0.024 ± 0.0012
RDC / Susceptibility Anisotropy restraints (Hz)	1.29 ± 0.094
Non-experimental/idealized restraints (monomer)	
Bonds (Å)	0.001 ± 0.0007
Angles (°)	0.54 ± 0.02
Improper (°)	0.46 ± 0.02
Superposition statistics over best 14 structures (monomer)	
All residues (TAR +AMG) **	1.16
Lower stem of TAR (residues 16-21, 41-46)	0.40
Upper stem pf TAR (residues 26-29, 36-39)	0.86
Energy profile over best 10 structures (dimer)	
Average total energy (kcal/mol)	-2401.83 ± 32.23
Average distance violation energy (kcal/mol)	23.62 ± 2.54
Average <i>sani</i> violation energy (kcal/mol)	17.41 ± 3.05
Experimental restraints	
All distance / NOE restraints (Å)	0.020 ± 0.0011
RDC / Susceptibility Anisotropy restraints (Hz)	1.20 ± 0.071
*For each value, the mean is given with the standard deviation.	
**Excluding loop (residues 30-35)	

Table 3.11

4 Discussions

4.1 Binding of AMG to multiple spots on the upper stem

It is widely recognized that the intermolecular surface in RNA/protein complexes is rather large and therefore small molecules cannot efficiently compete with proteins for RNA binding. Despite many efforts in the search for competitive inhibitors of the Tat/TAR interaction, no drug lead has been identified to date. The major difficulty in the development of TAR-binders as potential therapeutic agents is their low specificity.

Brownian dynamics simulations of cation diffusion performed on the NMR structures of free TAR RNA [²⁵; 39; 163; 164] revealed three electro negative pockets in the RNA hairpin.

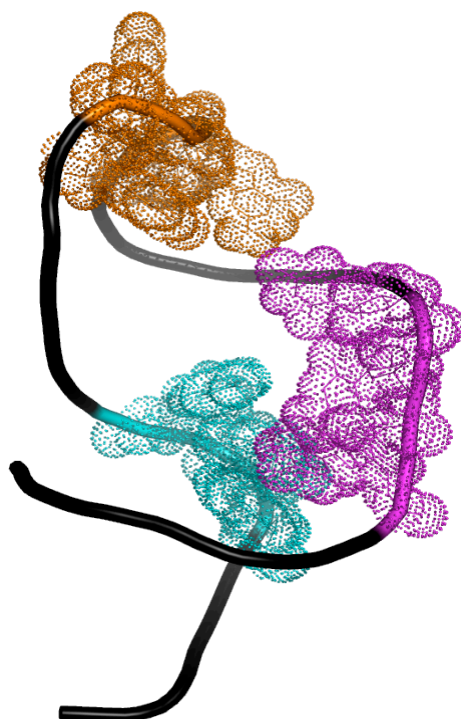


Fig 4.1 (a) Multiple cation binding pockets predicted by Brownian dynamics simulation in TAR RNA

The predicted cation binding sites are located on top of the U31-GG-G34 loop (orange), in the U23-C-U25 bulge (magenta), and in the deep groove of the lower stem (cyan) as shown in Figure 4.1(a). The “hot-spot” in the lower stem is particularly extended, involving the backbone of nucleotides in the upper stem, which connects the bulge and the loop. Multifunctional drugs, such as amino glycosides (AMG)

carrying several cationic groups could be expected to interact simultaneously with the lower and upper stem regions, thereby stabilizing the TAR RNA in its free conformation [30; 31; 164]. An extended effort in identifying the minimal number of interactions necessary to stabilize the Tat-bound conformation of the TAR RNA resulted in the detection of two binding sites for positively charged groups at the bulge site and in the upper stem major groove close to A27 and G28 [37].

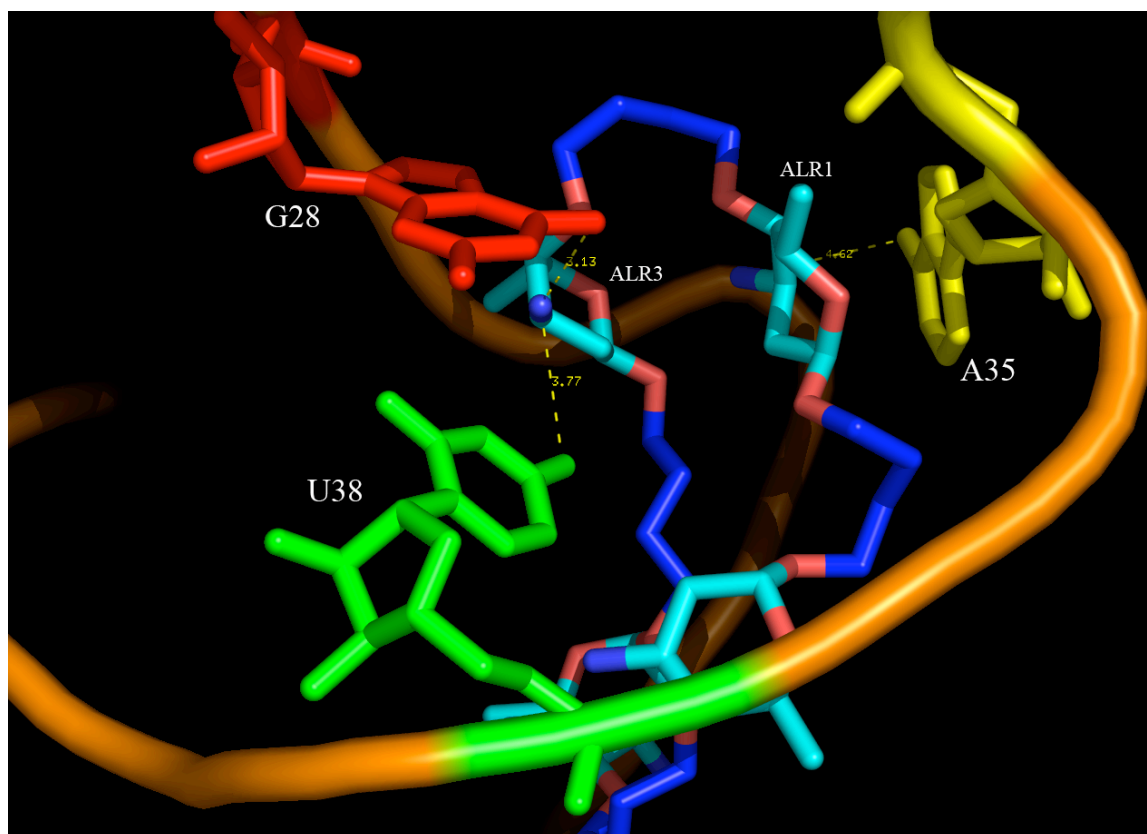


Fig 4.1 (b) Multiple interaction of amino group of the sugar moieties. Sugar 1 and 3 of AMG simultaneously interact with G28, A35 and U38 of HIV-2 TAR RNA.

Synthetic ligands exploiting both binding sites showed the highest activity and induced an RNA conformation similar to that binding to the Tat-protein [13; 32; 37; 38; 57; 66; 165; 166; 167; 168; 169; 170]. The amino groups of ligand AMG contact both the bulge site and the major groove functional groups at the G28 site (Figure 4.1(b)). Although the interaction with the bulge is suboptimal and could be improved by introducing a guanidinium group at the C1 or C5 of sugar-5, the amino group of sugar-3 nicely interacts with the electron-rich spot at the G28 and U38 site as shown in Figure 4.1(b). Additionally, the amino group of sugar-1 interacts with the loop residue A35. Also, the intermolecular NOEs to G36 from AMG to H8 and the partial

rearrangement of the hexanucleotide loop as suggested by the stacking of G34 on G36 suggests that this large but flexible scaffold of AMG allows reaching multiple binding sites on the TAR RNA, and therefore represents a valuable starting structure for the design of binders that explore a large surface of the TAR upper stem.

The conformation induced by AMG at the bulge site of the TAR RNA, as well as the relative orientation of the two stems is markedly different from that found in the TAR/argininamide complex which is accompanied by the formation of the U23-A27-U38 base triple and subsequent coaxial stacking of the upper stem on the lower stem. In contrast, the bulge nucleotides are solvent exposed in the TAR RNA/AMG complex and the upper stem is oriented at an angle of 60° with respect to the lower stem (Fig 4.1 (c)). The disruption of the A22-U40 base pair leads to the relative position of the two stems, thus allowing the tilting of A22 to accommodate the large ligand and of the electrostatic interactions of AMG amino groups with the phosphate of A22. The sugars and linkers of AMG show extensive hydrophobic contacts with U23 and U25 rings which are characteristic of a well-known interaction motif in complexes of proteins with oligosaccharides, where the sugars usually stack below aromatic side chains [^{157; 163}].

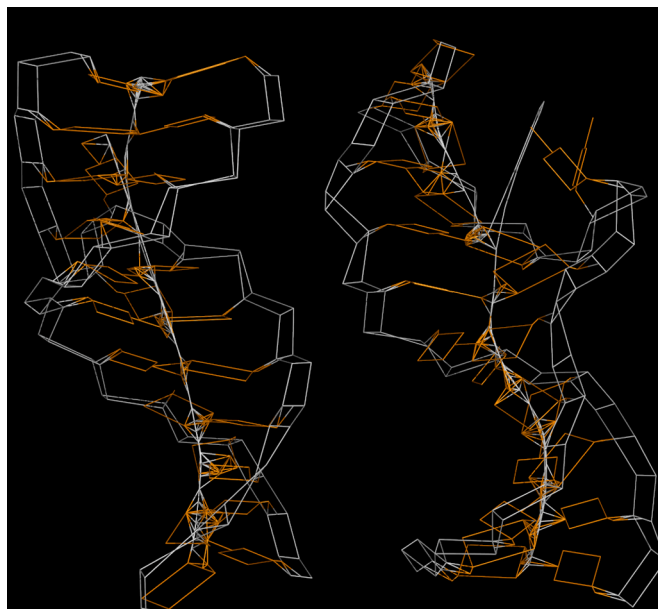


Fig 4.1 (c) The TAR RNA argininamide structure on the left reveals the co-axial stacking of the upper stem on the lower stem. In contrast, the upper stem is tilted at an angle of approximately 60° with respect to the lower stem in the TAR RNA/AMG complex on the right. (Figures generated with Pymol [DeLano Scientific] and CURVES [¹⁷¹])

The hexanucleotide apical loop (residues 30 to 35) has been found to be quiet flexible in previous NMR studies on the TAR RNA, with the ribose of C30, G33 and G34

averaging between C2' endo and the C3' endo conformations. Further, studies on a 14 and 19 nucleotide hairpin containing the TAR hexanucleotide loop revealed the stacking of A35 on G36 and of G34 on G36 [172]. On the basis of such structural information, it was proposed that the apical loop is stabilized by base stacking on both sides and possibly by the formation of base pairs between C30 and A35 or C30 and G34 [173]. In a more recent work, biochemical analysis of native and mutant TAR RNAs was used together with MD simulations to show that the apical loop of the TAR RNA is stabilized by a cross-loop base pair between C30 and G34, with A35 projected out of the loop [174]. This observation does not necessarily contradict the NMR studies hypothesizing the stacking of A35 on G36. The large line broadening of the A35–C2 resonance of the free TAR RNA and the NOEs between the bases of G34, A35 and G36 are consistent with A35 being in equilibrium between the stacked and looped-out conformations. The binding of AMG to the TAR RNA however eliminates this equilibrium. This is supported by the disappearance of NOE contacts between the base of A35 to those of G34 and G36, while several NOEs appear between the bases of G34 and G36. This suggests the stacking of G34 over G36. While G34 stacks on G36; the formation of C30-G34 base pair becomes possible even though no evidence of such a base pair is found in the TAR RNA/AMG complex. In the stacked conformation, the N7 and O6 functional groups of G34 assume well-defined positions in the major groove and the formation of the C30-G34 base pair becomes possible. MD simulations also suggest that G34 adopts a C3' endo conformation when G34 stacks upon G36. This is in agreement with our observation wherein we find that in all the structures of the TAR RNA/AMG complex, the ribose of G34 adopts the C3' endo conformation, in spite of the sugar pucker remaining unrestrained during the molecular dynamics protocol.

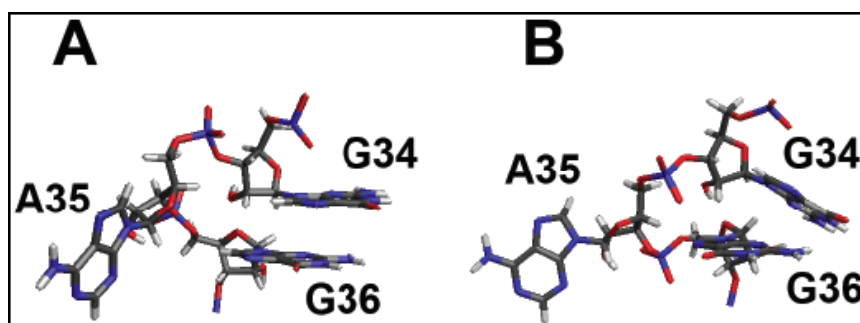


Fig 4.1 (d) G34 is positioned on top of G36. A35 interacts with the ligand and therefore moves out of the loop. As a consequence, G34 moves on top of G36. The two bases interact either face to face (A) or face to edge (B). This is indicative of two conformations in intermediate exchange in the microsecond to millisecond time scale. Further, the line width of C8 resonance of G34 also confirms this.

Modeling studies involving artificial restraints between C30-G34 enforcing the formation of a base pair in the hexanucleotide loop in the TAR RNA/AMG complex lead to formation of structures, which satisfy such constraints and do not lead to the violation of other experimental restraints and high energy. In summary, our data reveal that binding of a cationic ligand of appropriate size, simultaneously to the bulge and the loop of the TAR RNA on the major groove side of the upper stem leads to the stabilization of the looped out conformation of A35 and subsequently to the stacking of G34 on top of G36 [101]. Given the very large intermolecular surface in RNA/protein complexes, small molecules cannot efficiently act as competitive inhibitors to proteins for RNA binding. Despite many efforts in this direction, the search for a competitive inhibitor of the TAT/TAR RNA interaction has not met with much success because of toxicity or low specificity. The scaffold of AMG thus represents a promising class of aminoglycosides for the development of new TAR RNA binders that maybe used to address this problem, due to the following reasons:

- 1) Cyclic structure and flexibility of AMG allows functional groups to bind to multiple sites on the RNA.
- 2) AMG interacts with A35 of the hexanucleotide loop, leading to partial rigidification and restructuring of the previously unstructured loop, in a manner that may be suggestive of that required for the binding of CyclinT1 to the TAR RNA.

4.2 Cyclic Scaffold of AMG

The value of cyclic structures as a conformationally constrained backbone for TAR binders has been proven by the design of peptidomimetics based on a cyclic, beta-hairpin-like scaffold [170], binding the TAR RNA with a much higher affinity than the linear analogs [91; 170]. The structure of the BIV (Bovine Immunodeficiency Virus) TAR RNA bound to one of these peptidomimetics revealed extensive interactions with the upper stem major groove, including a conserved hydrophobic contact of an Isoleucine residue with the bulged-out Uridine of the base triple as well as electrostatic contacts of Arginine side chains with the tetra loop capping the upper stem and with the second and third base pairs of the upper stem [170]. Intermolecular interactions of similar nature are found for the TAR/AMG complex. Analogous to the BIV TAR RNA/peptide complex, AMG uses hydrophobic contacts to bind to the bulge pyrimidines and electrostatic contacts to bind to the major groove face of the upper stem base pairs and to the loop residues. As observed for the peptides, linear neooligoaminodeoxysaccharides show a much lower affinity for the TAR RNA than their cyclic analog AMG [91].

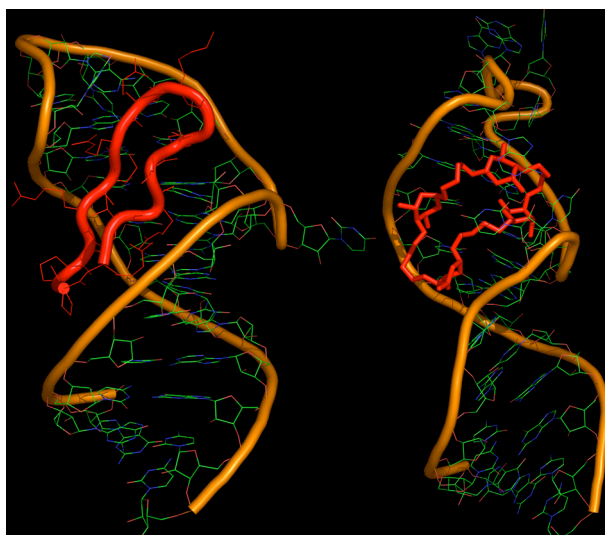


Figure 4.2 Cyclically restrained analogues as a tool to target the bulge and the loop of the TAR RNA simultaneously. The BIV TAR RNA/peptide complex on the right and the TAR RNA/AMG complex on the left offer analogous examples of cyclically restrained ligands targeting the bulge and the loop of the TAR RNA simultaneously. The ligands are colored red.

4.3 Comparison of the binding of AMG to TAR RNA complex with the binding of CyclinT1 to the TAR loop

It was believed for many years that sequences in the apical loop of TAR RNA are not required for Tat binding even though this region of the RNA is essential for efficient *trans*-activation [175]. Genetic studies implied that the loop acts as a binding site for cellular co-factors of Tat, however the identity of the loop factor remained mysterious and controversial for many years. The first evidence identifying the cellular co-factor required for loop recognition was obtained by Wei *et al.* [59] who reported that CyclinT1, a subunit of P-TEFb complex, can form a stable ternary complex with Tat and TAR RNA that could be detected by gel electrophoresis. Ternary complex formation requires both the Tat binding site on TAR as well as the loop sequence. Further, the formation of the Tat/TAR complex is required for the phosphorylation and subsequent activation of the RNA polymerase stalled at the TAR RNA site. The interaction site of CyclinT1 with the TAR RNA is localized in the apical loop. The two atoms O6 and N7 of G32 and G34 and the presence, but not the nature, of a nucleotide in position 35 are essential for the interaction. This was confirmed by monitoring the effect of base substitution on the affinity of the TAR RNA for the Tat-CyclinT1 complex and observing the efficiency of binding in a deletion mutant for the residue 35 [176]. Results indicated a dramatic reduction in the binding efficiency of the TAR RNA with the Tat-CyclinT1 complex when the normal TAR RNA sequence with the hexanucleotide loop was replaced with a deletion mutant at residue 35 (pentanucleotide loop). On substituting the guanine residues at positions 32 and 34 with adenine, a loss in binding affinity was observed. Further, if these guanine residues were substituted with a cytosine, then a dramatic loss in binding efficiency ensued. This could be reasoned out based on the chemical structure of the nucleotides in question. While the guanine moiety is not only capable of functioning as a hydrogen bond acceptor at the O6 site of the carbonyl group, it has also a electron rich Nitrogen at position 7 capable of functioning as an electron donor in the purine ring. Compared with guanine, adenine contains only the electron rich N7 but lacks the O6 hydrogen bond acceptor. In stark contrast, the cytosine nucleotide contains an amino group (hydrogen bond donor) in place of the carbonyl group (hydrogen bond acceptor) and also N7 is absent. Thus, the importance of the N7 and O6 at positions

32 and 34 can be rationalized based on the efficiency of RNA-protein complex formation in different mutants and the chemical structure of the nucleotides.

The mechanism by which the Tat protein favors the interaction of CyclinT1 with the apical loop is not known, although it is reasonable to hypothesize that Tat structurally pre-organizes the RNA for the binding of CyclinT1. On the basis of the structure of the complex of TAR with the aminoglycoside analogue AMG, we propose an explanation for the dependence of the TAR/CyclinT1 interaction on the Tat/TAR complex formation, which fits well the biochemical data summarized above. We suggest that the binding of Tat to TAR pre-organizes the hexanucleotide loop in a favorable conformation for CyclinT1 binding. It is conceivable that the arginine-rich stretch of the Tat protein, which is responsible for TAR binding at the bulge site, also contacts the A35 residue of the hexanucleotide loop and blocks it in its looped-out conformation, in a similar way as our aminoglycoside. The major driving force for this contact could be the interaction of an aromatic π orbital with a cation, similarly to that observed for the TAR RNA/AMG complex. A generic π -cation interaction provides a rationale for the insensitivity of the TAR/CyclinT1 interaction to the nature of the nucleotide at position 35. A small negative effect on complex formation is observed exclusively for the A35G mutant, where a possible cross-loop C30–G35 base pair would destabilize the looped-out conformation of the nucleotide in position 35. Supporting our hypothesis, affinity-cleaving experiments show that F38 of Tat is located in the proximity of nucleotide 34 and 35 of the TAR RNA [177]. Furthermore, the residue K50 of the Tat-protein has been found to cross-link with the TAR nucleotide 34, and this interaction is considerably enhanced by the binding of CyclinT1 to the Tat/TAR complex [176]. In the TAR RNA/AMG complex, the looped-out conformation of A35 is accompanied by the relocation of G34 upon G36. Analogously, in the Tat/TAR complex stacking of G34 upon G36 would place the O6 and N7 atoms of G34, which are essential for the binding of CyclinT1, in a well-defined position in the major groove of the TAR RNA. The formation of a base pair between C30 and G34 would further stabilize the position of the G34 functional groups. In fact, such base pair has been shown to favor the interaction of the TAR RNA with CyclinT1 without being directly involved in the complex formation [176]. Based on the irrelevance of the nucleotide nature at position 35, it has been hypothesized that this nucleotide functions as a spacer with no direct contact to the

proteins. Contrarily, we suggest that A35 could be the basis of a regulation mechanism for the formation of the Tat/TAR/CyclinT1 complex based on the conformational stability of the hexanucleotide loop.

4.4 The lower stem binding site and dimerization

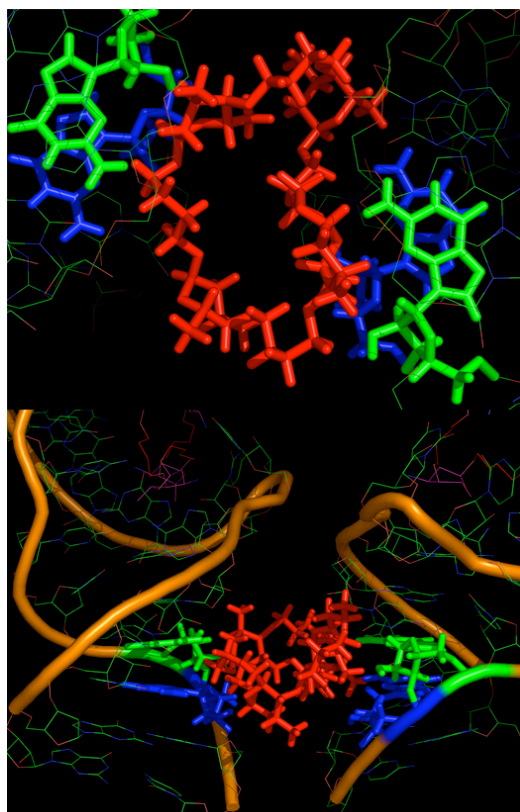


Figure 4.4 Two views of the dimer interface between the TAR RNA. Above is the top view of the dimer. The residues 43 and 44 are marked in green and blue respectively. The picture below offers a side view of the ligand in red at the dimer-forming interface.

As described in section 4.1, the TAR RNA has multiple “hot spots” capable of acting as attractive targets of cationic moieties. Previous studies involving neomycin B are proof of the minor groove of the lower stem of the TAR RNA forming a negatively charged cleft that forms an attractive binding site for aminoglycosides. Thus, it isn’t surprising that the aminoglycoside analogue AMG also displays similar behavior, binding to the minor groove of the TAR RNA, sandwiched between the parallel, lower stems of two adjacent TAR RNA molecules. Furthermore the TAR/AMG complex dimerizes. A model invoking oligomerization of the TAR RNA/AMG complex through the molecule of AMG bound to the lower stem minor groove is compatible with all experimental data. This process is likely to be favored by the

appropriate size of the constrained cyclic structure of AMG that allows it to simultaneously contact both the TAR RNAs.

4.5 Significance for drug design

In principle, targeting the interaction between the Tat protein and TAR RNA should be a sufficient condition to induce an inhibition of the transcription mechanism and lead to a decrease of full-length viral transcripts essential for the activity and pathogenicity of the Human Immunodeficiency virus. However, considering that biological systems are not static, and that proteins and nucleic acids present a certain degree of plasticity due to the pre existence of conformational populations, the traditional single drug-single target approach maybe (should be) replaced by single drug-multiple target approach [¹⁷⁸]. However, it is essential to counterbalance this affinity of the drug towards multiple targets with specific recognition and a high degree of affinity towards these specific targets. The novel cyclic aminoglycoside, AMG, introduced as an attractive scaffold that can serve as the basis for the design of promising future leads to target the Tat-TAR interaction, is an example of how a single drug-multiple target moiety may function.

It maybe possible to exploit the “promiscuous” behavior of an AMG based drug to target different sites, each of vital essence for the binding of different proteins required for biological activity, in this instance, simultaneously inhibit the interaction of the Tat protein with the bulge of the TAR RNA and the interaction of CyclinT1 with the hexanucleotide loop of the TAR RNA.

5. Conclusion and Perspectives

This thesis makes a contribution towards the understanding of the mechanism underlying the interaction of the Tat protein with the TAR RNA and how aminoglycosides based on a cyclic scaffold can afford a tool to explore RNA recognition by cationic entities. State of the art NMR methodology complimented with biochemical techniques was employed to arrive at the structure of the RNA in complex with the aminoglycoside analogue.

The structure reveals two ligand-binding sites. The first spans from the bulge to the loop of the TAR RNA on the major groove side of the upper stem. In particular, AMG interacts with the residue A35 of the RNA capping loop, thereby inducing partial structuring, while simultaneously binding to the bulge residues resulting in obstruction and inhibition of the Tat protein/TAR RNA and Tat/TAR RNA/CycT1 complex formation. The minor groove in the lower stem of the TAR RNA forms the second site capable of binding AMG.

The structural rearrangement of the loop may resemble and be part of a necessary conformational change of the TAR RNA induced by the Tat protein for the binding of CyclinT1. Biological processes are usually concerted in nature and interfering with one process may necessarily cause an affect in subsequent interactions necessary for biological activity. The interaction of this novel aminoglycoside is an example of how an organic moiety can interfere with a necessary interaction (i.e. that of Tat protein and TAR RNA), provide insights into how further activities that precede the Tat/TAR interaction are regulated and how this mechanism can be exploited to design better scaffolds based on the structure of this organic moiety (AMG) for a new generation of TAR binders that can simultaneously inhibit both the Tat/TAR RNA and the Tat/TAR/CycT1 complex formation.

Aminoglycosides have since long, formed an integral part of the armory at the disposal of chemists, to not only better explore the realm of RNA recognition by cations, but also an extremely valuable source of drugs used to target a wide variety of diseases prevailing in the world. However, lack of specificity and the quick development of resistance against such aminoglycosides by pathogenic organisms fuel the urgent need to design and synthesize new and novel scaffolds of antibiotics that would be more effective in curbing diseases. The author believes that the work

carried out in this thesis paves the way for the design of a promising generation of lead compounds that exhibit specific affinity towards the TAR RNA, while simultaneously contacting two sites which are of paramount importance for the binding of Tat protein and the CyclinT1 complex. These compounds would be novel in the fact that they would facilitate interference with the TAR-Tat and TAR-Tat-CyclinT1 interaction in a cooperative way.

It is also the author's hope, that the conclusions reached in this work will be a fruitful source of further experiments that may eventually lead to a deeper understanding of HIV pathogenicity and the underlying mechanism.

Bibliography

- [1] Finzi, D., Blankson, J., Siliciano, J. D., Margolick, J. B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C., Quinn, T. C., Chaisson, R. E., Rosenberg, E., Walker, B., Gange, S., Gallant, J. & Siliciano, R. F. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med* **5**, 512-7.
- [2] Zhang, L., Ramratnam, B., Tenner-Racz, K., He, Y., Vesanen, M., Lewin, S., Talal, A., Racz, P., Perelson, A. S., Korber, B. T., Markowitz, M. & Ho, D. D. (1999). Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med* **340**, 1605-13.
- [3] Billeter, M. (1995). NMR for structural studies in drug discovery. *Perspectives in Drug Discovery and Design* **3**, 151-167.
- [4] Diercks, T., Coles, M. & Kessler, H. (2001). Applications of NMR in drug discovery. *Current Opinion in Chemical Biology* **5**, 285-291.
- [5] Fesik, S. W. (1999). NMR as a tool in drug research. *Faseb Journal* **13**, A1422-A1422.
- [6] Fesik, S. W., Petros, A. M., Logan, T. M., Yu, L., Xu, R., Theriault, Y., Nettesheim, D., Meadows, R. P., Hansen, A. P., Liang, H. & Olejniczak, E. T. (1993). NMR-Studies of Biomacromolecules as an Aid in Drug Research. *Abstracts of Papers of the American Chemical Society* **206**, 119.
- [7] Fesik, S. W., Shuker, S. B., Hajduk, P. J. & Meadows, R. P. (1997). SAR by NMR: An NMR-based approach for drug discovery. *Protein Engineering* **10**, 73-73.
- [8] Moore, J. M. (1999). NMR screening in drug discovery. *Current Opinion in Biotechnology* **10**, 54-58.
- [9] Gait, M. J. & Karn, J. (1993). RNA recognition by the human immunodeficiency virus Tat and Rev proteins. *Trends Biochem Sci* **18**, 255-9.
- [10] Gait, M. J. & Karn, J. (1995). Progress in Anti-HIV Structure-Based Drug Design. *Trends in Biotechnology* **13**, 430-438.
- [11] Montembault, M., Vo-Thanh, G., Deyine, A., Fargeas, V., Villieras, M., Adjou, A., Dubreuil, D., Esquieu, D., Gregoire, C., Opi, S., Peloponese, J. M., Campbell, G., Watkins, J., de Mareuil, J., Aubertin, A. M., Bailly, C., Loret, E. & Lebreton, J. (2004). A possible improvement for structure-based drug design illustrated by the discovery of a Tat HIV-1 inhibitor. *Bioorganic & Medicinal Chemistry Letters* **14**, 1543-1546.
- [12] Frankel, A. D. (1997). RNA-protein interactions in the HIV life cycle. *Faseb Journal* **11**, A867-A867.
- [13] Froeyen, M. & Herdewijn, P. (2002). RNA as a target for drug design, the example of Tat-TAR interaction. *Curr Top Med Chem* **2**, 1123-45.
- [14] Barresinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dautuet, C., Axlerblin, C., Vezinetbrun, F., Rouzioux, C., Rozenbaum, W. & Montagnier, L. (1983). Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune-Deficiency Syndrome (AIDS). *Science* **220**, 868-871.
- [15] Gallo, R. (1997). AIDS: The Process of Discovery. *Science* **275**, 141-141.
- [16] Dybul, M., Fauci, A. S., Bartlett, J. G., Kaplan, J. E. & Pau, A. K. (2002). Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. *Ann Intern Med* **137**, 381-433.
- [17] Dybul, M., Fauci, A. S., Bartlett, J. G., Kaplan, J. E. & Pau, A. K. (2002). Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. Recommendations of the Panel on Clinical Practices for Treatment of HIV. *MMWR Recomm Rep* **51**, 1-55.
- [18] Hirsch, M. S., Conway, B., D'Aquila, R. T., Johnson, V. A., Brun-Vezinet, F., Clotet, B., Demeter, L. M., Hammer, S. M., Jacobsen, D. M., Kuritzkes, D. R., Loveday, C., Mellors, J. W., Vella, S., Richman, D. D., Panel, I. A. S. U. (1998). Antiretroviral drug resistance testing in adults with HIV infection - Implications for clinical management. *JAMA-Journal of the American Medical Association* **279**, 1984-1991.
- [19] Hermann, T. & Westhof, E. (2000). Rational drug design and high-throughput techniques for RNA targets. *Comb Chem High Throughput Screen* **3**, 219-34.
- [20] Hsu, M. C., Dhingra, U., Earley, J. V., Holly, M., Keith, D., Nalin, C. M., Richou, A. R., Schutt, A. D., Tam, S. Y., Potash, M. J. & et al. (1993). Inhibition of type 1 Human Immunodeficiency virus replication by a Tat antagonist to which the virus remains sensitive after prolonged exposure in vitro. *Proc Natl Acad Sci U S A* **90**, 6395-9.

- [21] Murchie, A. I. H., Davis, B., Isel, C., Afshar, M., Drysdale, M. J., Bower, J., Potter, A. J., Starkey, I. D., Swarbrick, T. M., Mirza, S., Prescott, C. D., Vaglio, P., Aboul-ela, F. & Karn, J. (2004). Structure-based drug design targeting an inactive RNA conformation: Exploiting the flexibility of HIV-1 TAR RNA. *Journal of Molecular Biology* **336**, 625-638.
- [22] Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1993). RNA-Protein Interactions in Model Systems for HIV TAR/Tat Recognition. *Biophysical Journal* **64**, A1-A1.
- [23] Tamilarasu, N., Huq, I. & Rana, T. M. (2000). Design, synthesis, and biological activity of a cyclic peptide: An inhibitor of HIV-1 Tat-TAR interactions in human cells. *Bioorganic & Medicinal Chemistry Letters* **10**, 971-974.
- [24] Tor, Y. (2003). Targeting RNA with small molecules. *Chembiochem* **4**, 998-1007.
- [25] Vicens, Q. & Westhof, E. (2003). RNA as a drug target: the case of aminoglycosides. *Chembiochem* **4**, 1018-23.
- [26] Renner, S., Ludwig, V., Boden, O., Scheffer, U., Gobel, M. & Schneider, G. (2005). New inhibitors of the Tat-TAR RNA interaction found with a "fuzzy" pharmacophore model. *Chembiochem* **6**, 1119-1125.
- [27] Bieniasz, P. D., Grdina, T. A., Bogerd, H. P. & Cullen, B. R. (1999). Recruitment of cyclin T1/P-TEFb to an HIV type I long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7791-7796.
- [28] Feinberg, M. B., Baltimore, D. & Frankel, A. D. (1991). The Role of Tat in the Human-Immunodeficiency-Virus Life-Cycle Indicates a Primary Effect on Transcriptional Elongation. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 4045-4049.
- [29] Graham, G. J. & Maio, J. J. (1990). RNA transcripts of the human immunodeficiency virus transactivation response element can inhibit action of the viral transactivator. *Proc Natl Acad Sci U S A* **87**, 5817-21.
- [30] Aboul-ela, F. & Varani, G. (1998). Recognition of HIV-1 TAR RNA by Tat protein and Tat-derived peptides. *Theochem-Journal of Molecular Structure* **423**, 29-39.
- [31] Aboulela, F., Karn, J. & Varani, G. (1995). The Structure of the Human-Immunodeficiency-Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *Journal of Molecular Biology* **253**, 313-332.
- [32] Afshar, M., Prescott, C. D. & Varani, G. (1999). Structure-based and combinatorial search for new RNA-binding drugs. *Current Opinion in Biotechnology* **10**, 59-63.
- [33] Arzumanov, A., Stetsenko, D. A., Malakhov, A. D., Reichelt, S., Sorensen, M. D., Babu, B. R., Wengel, J. & Gait, M. J. (2003). A structure-activity study of the inhibition of HIV-1 Tat-dependent trans-activation by mixmer 2 '-O-methyl oligoribonucleotides containing locked nucleic acid (LNA), alpha-L-LNA, or 2 '-thio-LNA residues. *Oligonucleotides* **13**, 435-453.
- [34] Arzumanov, A., Walsh, A. P., Liu, X. H., Rajwanshi, V. K., Wengel, J. & Gait, M. J. (2001). Oligonucleotide analogue interference with the HIV-1 Tat protein-TAR RNA interaction. *Nucleosides Nucleotides & Nucleic Acids* **20**, 471-480.
- [35] Arzumanov, A., Walsh, A. P., Rajwanshi, V. K., Kumar, R., Wengel, J. & Gait, M. J. (2001). Inhibition of HIV-1 Tat-dependent trans activation by steric block chimeric 2 '-O-methyl/LNA oligoribonucleotides. *Biochemistry* **40**, 14645-14654.
- [36] Athanassiou, Z., Dias, R. L. A., Leeper, T. C., Moehle, K., Robinson, J. A. & Varani, G. (2005). Structure-based design of novel Tat inhibitors with increased specificity and enhanced pharmacological potential. *Biopolymers* **80**, 567-567.
- [37] Davis, B., Afshar, M., Varani, G., Murchie, A. I. H., Karn, J., Lentzen, G., Drysdale, M., Bower, J., Potter, A. J., Starkey, I. D., Swarbrick, T. & Aboul-ela, F. (2004). Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic "hot spots". *Journal of Molecular Biology* **336**, 343-356.
- [38] Frankel, A. D. (2000). Fitting peptides into the RNA world. *Current Opinion in Structural Biology* **10**, 332-340.
- [39] Hermann, T. & Westhof, E. (1998). RNA as a drug target: chemical, modelling, and evolutionary tools. *Curr Opin Biotechnol* **9**, 66-73.
- [40] Karn, J. (1999). Tackling Tat. *Journal of Molecular Biology* **293**, 235-254.
- [41] Keen, N. J., Churcher, M. J. & Karn, J. (1997). Transfer of Tat and release of TAR RNA during the activation of the human immunodeficiency virus type-1 transcription elongation complex. *EMBO Journal* **16**, 5260-5272.

- [42] Marciniak, R. A., Calnan, B. J., Frankel, A. D. & Sharp, P. A. (1990). HIV-1 Tat Protein Trans-Activates Transcription In Vitro. *Cell* **63**, 791-802.
- [43] Tan, R. Y., Brodsky, A., Williamson, J. R. & Frankel, A. D. (1997). RNA recognition by HIV-1 Tat and Rev. *Seminars in Virology* **8**, 186-193.
- [44] Bannwarth, S. & Gatignol, A. (2005). HIV-1 TAR RNA: The target of molecular interactions between the virus and its host. *Current HIV Research* **3**, 61-71.
- [45] Chaloin, O., Peter, J. C., Briand, J. P., Masquida, B., Desgranges, C., Muller, S. & Hoebeker, J. (2005). The N-terminus of HIV-1 Tat protein is essential for Tat-TAR RNA interaction. *Cellular and Molecular Life Sciences* **62**, 355-361.
- [46] Brodsky, A. S. & Williamson, J. R. (1997). Solution structure of the HIV-2 TAR-argininamide complex. *Journal of Molecular Biology* **267**, 624-639.
- [47] Chen, L. & Frankel, A. D. (1994). TAR-Tat Recognition in Bovine Immunodeficiency Virus - an Interesting System to Study Transcriptional Regulation. *Journal of Cellular Biochemistry*, 57-57.
- [48] Gomes, N., Garber, M. E. & Jones, K. A. (2003). Techniques to analyze the HIV-1 Tat and TAR RNA-dependent recruitment and activation of the cyclin T1 : CDK9 (P-TEFb) transcription elongation factor. In *RNA Polymerases and Associated Factors, Pt D*, Vol. 371, pp. 324-+. Academic Press Inc, San Diego.
- [49] Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. (1993). Role of Rna Structure in Arginine Recognition of TAR RNA. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3680-3684.
- [50] Harrich, D., Hsu, C., Race, E. & Gaynor, R. B. (1994). Differential Growth-Kinetics Are Exhibited by Human-Immunodeficiency-Virus Type-1 TAR Mutants. *Journal of Virology* **68**, 5899-5910.
- [51] AboulEla, F., Karn, J. & Varani, G. (1996). Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Research* **24**, 3974-3981.
- [52] Al-Hashimi, H. M. (2005). Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *ChemBiochem* **6**, 1506-1519.
- [53] Al-Hashimi, H. M., Gossler, Y., Gorin, A., Hu, W. D., Majumdar, A. & Patel, D. J. (2002). Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *Journal of Molecular Biology* **315**, 95-102.
- [54] Al-Hashimi, H. M. & Patel, D. J. (2002). Residual dipolar couplings: Synergy between NMR and structural genomics. *Journal of Biomolecular NMR* **22**, 1-8.
- [55] Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M. & Skinner, M. A. (1990). HIV-1 Tat Protein Stimulates Transcription by Binding to a U-Rich Bulge in the Stem of the TAR RNA Structure. *EMBO Journal* **9**, 4145-4153.
- [56] Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A. & Valerio, R. (1989). Human Immunodeficiency Virus-1 Tat Protein Binds Trans-Activation-Responsive Region (TAR) RNA In Vitro. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6925-6929.
- [57] Frankel, A. D. (1992). Peptide Models of the Tat-TAR Protein-RNA Interaction. *Protein Science* **1**, 1539-1542.
- [58] Chirivi, R. G. S., Taraboletti, G., Bani, M. R., Barra, L., Piccinini, G., Giacca, M., Bussolino, F. & Giavazzi, R. (1999). Human immunodeficiency virus-1 (HIV-1)-Tat protein promotes migration of acquired immunodeficiency syndrome-related lymphoma cells and enhances their adhesion to endothelial cells. *Blood* **94**, 1747-1754.
- [59] Wei, P., Garber, M. E., Fang, S. M., Fischer, W. H. & Jones, K. A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* **92**, 451-462.
- [60] Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R. & Jones, K. A. (1998). The interaction between HIV-1 Tat and human Cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes & Development* **12**, 3512-3527.
- [61] Frankel, A. D., Brecht, D. S. & Pabo, C. O. (1988). Tat Protein from Human Immunodeficiency Virus Forms a Metal-Linked Dimer. *Science* **240**, 70-73.
- [62] Brodsky, A. S., Erlacher, H. A. & Williamson, J. R. (1998). NMR evidence for a base triple in the HIV-2 TAR C-G center dot C+ mutant argininamide complex. *Nucleic Acids Research* **26**, 1991-1995.

- [63] Wang, Z. Y., Shah, K. & Rana, T. M. (2001). Probing Tat peptide-TAR RNA interactions by psoralen photo-cross-linking. *Biochemistry* **40**, 6458-6464.
- [64] Karn, J., Dingwall, C., Finch, J. T., Heaphy, S. & Gait, M. J. (1991). Rna-Binding by the Tat and Rev Proteins of HIV-1. *Biochimie* **73**, 9-16.
- [65] Keen, N. J., Gait, M. J. & Karn, J. (1996). Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 2505-2510.
- [66] Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. G., Gait, M. J. & Karn, J. (1993). High-Affinity Binding of TAR RNA by the Human-Immunodeficiency-Virus Type-1 Tat Protein Requires Base-Pairs in the RNA Stem and Amino-Acid-Residues Flanking the Basic Region. *Journal of Molecular Biology* **230**, 90-110.
- [67] Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992). Conformation of the TAR RNA-Arginine Complex by NMR-Spectroscopy. *Science* **257**, 76-80.
- [68] Gu, W. G., Powell, W., Mote, J. & Reines, D. (1993). Nascent Rna Cleavage by Arrested Rna Polymerase-Ii Does Not Require Upstream Translocation of the Elongation Complex on DNA. *Journal of Biological Chemistry* **268**, 25604-25616.
- [69] Marshall, N. F., Dahmus, G. K. & Dahmus, M. E. (1998). Regulation of carboxyl-terminal domain phosphatase by HIV-1 Tat protein. *Journal of Biological Chemistry* **273**, 31726-31730.
- [70] Marshall, N. F. & Dahmus, M. E. (2000). C-terminal domain phosphatase sensitivity of RNA polymerase II in early elongation complexes on the HIV-1 and adenovirus 2 major late templates. *Journal of Biological Chemistry* **275**, 32430-32437.
- [71] Marshall, N. F., Peng, J. M., Xie, Z. & Price, D. H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *Journal of Biological Chemistry* **271**, 27176-27183.
- [72] Yang, R., Morosetti, R. & Koeffler, H. P. (1997). Characterization of a second human cyclin a that is highly expressed in testis and in several leukemic cell lines. *Cancer Research* **57**, 913-920.
- [73] Yang, X. Z., Gold, M. O., Tang, D. N., Lewis, D. E., AguilarCordova, E., Rice, A. P. & Herrmann, C. H. (1997). TAK, an HIV Tat-associated kinase, is a member of the cyclin-dependent family of protein kinases and is induced by activation of peripheral blood lymphocytes and differentiation of promonocytic cell lines. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12331-12336.
- [74] Peng, J. M., Marshall, N. F. & Price, D. H. (1998). Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *Journal of Biological Chemistry* **273**, 13855-13860.
- [75] Bieniasz, P. D., Grdina, T. A., Bogerd, H. P. & Cullen, B. R. (1999). Analysis of the effect of natural sequence variation in Tat and in cyclin T on the formation and RNA binding properties of Tat-cyclin T complexes. *Journal of Virology* **73**, 5777-5786.
- [76] Zhang, J., Tamilarasu, N., Hwang, S. W., Garber, M. E., Huq, I., Jones, K. A. & Rana, T. M. (2000). HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. *Journal of Biological Chemistry* **275**, 34314-34319.
- [77] Graham, G. J. & Maio, J. J. (1990). Rna Transcripts of the Human-Immunodeficiency-Virus Transactivation Response Element Can Inhibit Action of the Viral Transactivator. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 5817-5821.
- [78] Lisziewicz, J., Smythe, J., Sun, D., Lusso, P., Hall, L., Gallo, R. C. & Reitz, M. S. (1993). Potential Application of Gene-Therapy to HIV-1 Infection. Inhibition of Tat Function. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **6**, 670-670.
- [79] Lisziewicz, J., Sun, D., Klotman, M., Agrawal, S., Zamecnik, P. & Gallo, R. C. (1993). Oligonucleotides Phosphorothioates for HIV-1 Therapy - Invitro Model for Treatment. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **6**, 672-672.
- [80] Lisziewicz, J., Sun, D., Metelev, V., Zamecnik, P., Gallo, R. C. & Agrawal, S. (1993). Long-Term Treatment of Human Immunodeficiency Virus-Infected Cells with Antisense Oligonucleotide Phosphorothioates. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 3860-3864.
- [81] Lisziewicz, J., Sun, D., Smythe, J., Lusso, P., Lori, F., Louie, A., Markham, P., Rossi, J., Reitz, M. & Gallo, R. C. (1993). Inhibition of Human-Immunodeficiency-Virus Type-1 Replication by Regulated Expression of a Polymeric Tat Activation Response RNA Decoy as

-
- a Strategy for Gene-Therapy in Aids. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8000-8004.
- [82] Tao, J. S. & Frankel, A. D. (1992). Specific Binding of Arginine to TAR RNA. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 2723-2726.
- [83] Tao, J. S. & Frankel, A. D. (1996). Arginine-binding RNAs resembling TAR identified by in vitro selection. *Biochemistry* **35**, 2229-2238.
- [84] Patino, N., Di Giorgio, C., Dan-Covalciuc, C., Peytou, V., Terreux, R., Cabrol-Bass, D., Bailly, C. & Condom, R. (2002). Modelling, synthesis and biological evaluation of an ethidium-arginine conjugate linked to a ribonuclease mimic directed against TAR RNA of HIV-1. *European Journal of Medicinal Chemistry* **37**, 573-584.
- [85] Peytou, V., Condom, R., Patino, N., Guedj, R., Aubertin, A. M., Gelus, N., Bailly, C., Terreux, R. & Cabrol-Bass, D. (1999). Synthesis and antiviral activity of ethidium-arginine conjugates directed against the TAR RNA of HIV-1. *Journal of Medicinal Chemistry* **42**, 4042-4053.
- [86] Terreux, R., Cabrol-Bass, D., Peytou, V., Condom, R. & Guedj, R. (1999). Modeling of the interaction between new ethidium derivatives and TAR RNA of HIV-1. *Journal of Chemical Information and Computer Sciences* **39**, 413-419.
- [87] Schatz, A., Bugie, E. & Waksman, S. A. (1944). Streptomycin, a substance exhibiting antibiotic activity against gram positive and gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine* **55**, 66-69.
- [88] Waksman, S. A. & Schatz, A. (1945). Streptomycin - Origin, Nature, and Properties. *Journal of the American Pharmaceutical Association-Scientific Edition* **34**, 273-291.
- [89] Schroeder, R., Waldsich, C. & Wank, H. (2000). Modulation of RNA function by aminoglycoside antibiotics. *EMBO Journal* **19**, 1-9.
- [90] Waldsich, C., Semrad, K. & Schroeder, R. (1998). Neomycin B inhibits splicing of the td intron indirectly by interfering with translation and enhances missplicing in vivo. *RNA-a Publication of the RNA Society* **4**, 1653-1663.
- [91] Kirschning, A., Chen, G. W., Jaunzems, J., Jesberger, M., Kalesse, M. & Lindner, M. (2004). Synthesis of extended spacer-linked neooligodeoxysaccharides by metathesis olefination and evaluation of their RNA-binding properties. *Tetrahedron* **60**, 3505-3521.
- [92] Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995). Preparation of isotopically enriched RNAs for heteronuclear NMR. *Nuclear Magnetic Resonance and Nucleic Acids* **261**, 300-322.
- [93] Funk, T. W., Berlin, J. M. & Grubbs, R. H. (2006). Highly active chiral ruthenium catalysts for asymmetric ring-closing olefin metathesis. *Journal of the American Chemical Society* **128**, 1840-1846.
- [94] Grubbs, R. H. (2004). Olefin metathesis. *Tetrahedron* **60**, 7117-7140.
- [95] Trnka, T. M. & Grubbs, R. H. (2001). The development of L2X2Ru = CHR olefin metathesis catalysts: An organometallic success story. *Accounts of Chemical Research* **34**, 18-29.
- [96] Ward, D. W., Seiders, T. J. & Grubbs, R. H. (2001). Ruthenium-catalyzed asymmetric ring-closing metathesis. *Abstracts of Papers of the American Chemical Society* **222**, U137-U137.
- [97] Doniskeller, H. (1979). Site Specific Enzymatic Cleavage of Rna. *Nucleic Acids Research* **7**, 179-192.
- [98] Raines, R. T. (1998). Ribonuclease A. *Chemical Reviews* **98**, 1045-1065.
- [99] Doniskeller, H., Maxam, A. M. & Gilbert, W. (1977). Mapping Adenines, Guanines, and Pyrimidines in Rna. *Nucleic Acids Research* **4**, 2527-2537.
- [100] Cilley, C. D. & Williamson, J. R. (1997). Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA* **3**, 57-67.
- [101] Raghunathan, D., Sanchez-Pedregal, V. M., Junker, J., Schwiegk, C., Kalesse, M., Kirschning, A. & Carlomagno, T. (2006). TAR-RNA recognition by a novel cyclic aminoglycoside analogue. *Nucleic Acids Research* **34**, 3599-3608.
- [102] Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular NMR* **2**, 661-665.
- [103] Bodenhausen, G. & Ruben, D. J. (1980). Natural Abundance N-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chemical Physics Letters* **69**, 185-189.
- [104] Furtig, B., Richter, C., Wohnert, J. & Schwalbe, H. (2003). NMR Spectroscopy of RNA. *Chembiochem* **4**, 936-962.
- [105] Varani, G., Aboulela, F. & Allain, F. H. T. (1996). NMR investigation of RNA structure, pp. 51-127.

- [106] Wijmenga, S. S. & van Buuren, B. N. M. (1998). The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 287-387.
- [107] Hu, W. D., Kakalis, L. T., Jiang, L. C., Jiang, F., Ye, X. M. & Majumdar, A. (1998). 3D HCCH-COSY-TOCSY experiment for the assignment of ribose and amino acid side chains in C-13 labeled RNA and protein. *Journal of Biomolecular NMR* **12**, 559-564.
- [108] BAX, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M. & Kay, L. E. (1990). Practical Aspects of Proton Carbon Carbon Proton 3-Dimensional Correlation Spectroscopy of C-13-Labeled Proteins. *Journal of Magnetic Resonance* **87**, 620-627.
- [109] Kay, L. E., Ikura, M. & Bax, A. (1990). Proton Proton Correlation Via Carbon Carbon Couplings - a 3-Dimensional NMR Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with C-13. *Journal of the American Chemical Society* **112**, 888-889.
- [110] Fesik, S. W., Eaton, H. L., Olejniczak, E. T., Zuiderweg, E. R. P., McIntosh, L. P. & Dahlquist, F. W. (1990). 2d and 3d NMR-Spectroscopy Employing C-13-C-13 Magnetization Transfer by Isotropic Mixing - Spin System-Identification in Large Proteins. *Journal of the American Chemical Society* **112**, 886-888.
- [111] Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1992). A 4d-Hcch-Tocsy Experiment for Assigning the Side-Chain H-1-Resonance and C-13-Resonance of Proteins. *Journal of Biomolecular NMR* **2**, 655-659.
- [112] Sorensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G. & Ernst, R. R. (1983). Product Operator-Formalism for the Description of NMR Pulse Experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy* **16**, 163-192.
- [113] Shaka, A. J., Lee, C. J. & Pines, A. (1988). Iterative Schemes for Bilinear Operators - Application to Spin Decoupling. *Journal of Magnetic Resonance* **77**, 274-293.
- [114] Furtig, B., Richter, C., Bermel, W. & Schwalbe, H. (2004). New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *Journal of Biomolecular NMR* **28**, 69-79.
- [115] Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. & Griesinger, C. (1994). A 3-Dimensional Triple-Resonance H-1,C-13,P-31 Experiment - Sequential through-Bond Correlation of Ribose Protons and Intervening Phosphorus Along the RNA Oligonucleotide Backbone. *Journal of the American Chemical Society* **116**, 6472-6473.
- [116] Talluri, S. & Wagner, G. (1996). An optimized 3D NOESY-HSQC. *Journal of Magnetic Resonance Series B* **112**, 200-205.
- [117] Brutscher, B., Boisbouvier, J., Kupce, E., Tisne, C., Dardel, F., Marion, D. & Simorre, J. P. (2001). Base-type-selective high-resolution C-13 edited NOESY for sequential assignment of large RNAs. *Journal of Biomolecular NMR* **19**, 141-151.
- [118] Vuister, G. W. & BAX, A. (1993). Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins. *Journal of the American Chemical Society* **115**, 7772-7777.
- [119] Cox, J. R. & Serpersu, E. H. (1995). The complete 1H NMR assignments of aminoglycoside antibiotics and conformational studies of butirosin A through the use of 2D NMR spectroscopy. *Carbohydr Res* **271**, 55-63.
- [120] Palmer, A. G., Williams, J. & McDermott, A. (1996). Nuclear magnetic resonance studies of biopolymer dynamics. *Journal of Physical Chemistry* **100**, 13293-13310.
- [121] Carr, H. Y. & Purcell, E. M. (1954). Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Physical Review* **94**, 630-638.
- [122] Meiboom, S. & Gill, D. (1958). Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **29**, 688-691.
- [123] Akke, M., Fiala, R., Jiang, F., Patel, D. & Palmer, A. G. (1997). Base dynamics in a UUCG tetraloop RNA hairpin characterized by N-15 spin relaxation: Correlations with structure and stability. *RNA-a Publication of the RNA Society* **3**, 702-709.
- [124] Shajani, Z. & Varani, G. (2005). C-13 NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *Journal of Molecular Biology* **349**, 699-715.
- [125] Ruckert, M. & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *Journal of the American Chemical Society* **122**, 7793-7797.
- [126] Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by

-
- mechanical stress or embedding of oriented purple membranes. *Journal of Biomolecular NMR* **18**, 303-309.
- [127] Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Structural Biology* **5**, 1065-1074.
- [128] Hansen, M. R., Rance, M. & Pardi, A. (1998). Observation of long-range H-1-H-1 distances in solution by dipolar coupling interactions. *Journal of the American Chemical Society* **120**, 11210-11211.
- [129] Zweckstetter, M. & Bax, A. (2001). Characterization of molecular alignment in aqueous suspensions of Pfl bacteriophage. *Journal of Biomolecular NMR* **20**, 365-377.
- [130] Bax, A., Tjandra, N., Ottiger, M., Marquardt, J., Cornilescu, G. & Hu, J. S. (1998). A new approach to the study of biomolecular structure by NMR. *Biophysical Journal* **74**, A137-A137.
- [131] Tjandra, N., Bax, Ad. (1997). Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* **278**, 1111-1114.
- [132] Losonczi, J. A. & Prestegard, J. H. (1998). Nuclear magnetic resonance characterization of the myristoylated, N-terminal fragment of ADP-ribosylation factor 1 in a magnetically oriented membrane array. *Biochemistry* **37**, 706-716.
- [133] Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 9279-9283.
- [134] Al-Hashimi, H. M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E. & Patel, D. J. (2001). Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex, relative orientation of G center dot(C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. *Journal of the American Chemical Society* **123**, 633-640.
- [135] Al-Hashimi, H. M., Tolman, J. R., Majumdar, A., Gorin, A. & Patel, D. J. (2001). Determining stoichiometry in homomultimeric nucleic acid complexes using magnetic field induced residual dipolar couplings. *Journal of the American Chemical Society* **123**, 5806-5807.
- [136] Leupin, W., Otting, G., Amacker, H. & Wuthrich, K. (1990). Application of C-13(Omega-1)-Half-Filtered [H-1,H-1]-Noesy for Studies of a Complex Formed between DNA and a C-13-Labeled Minor-Groove-Binding Drug. *Febs Letters* **263**, 313-316.
- [137] Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R. & Kay, L. E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex. *Journal of the American Chemical Society* **119**, 6711-6721.
- [138] Yathindra, N. & Sundaralingam, M. (1975). Conformational Correlation between Sugar Pucker and Internucleotide P-O Bonds in Polynucleotides. *Biophysical Journal* **15**, A87-A87.
- [139] Weinhold, F. (2001). Chemistry - A new twist on molecular shape. *Nature* **411**, 539-541.
- [140] Schwalbe, H., Marino, J. P., King, G. C., Wechselberger, R., Bermel, W. & Griesinger, C. (1994). Determination of a Complete Set of Coupling-Constants in C-13-Labeled Oligonucleotides. *Journal of Biomolecular NMR* **4**, 631-644.
- [141] Chang, K. Y. & Varani, G. (1997). Nucleic acids structure and recognition. *Nature Structural Biology* **4**, 854-858.
- [142] Varani, G. (1993). RNA Secondary Structure Elements from NMR-Spectroscopy. *Biophysical Journal* **64**, A1-A1.
- [143] Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. (2006). Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 47-62.
- [144] Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160**, 65-73.
- [145] Cheatham, T. E. & Young, M. A. (2000). Molecular dynamics simulation of nucleic acids: Successes, limitations, and promise. *Biopolymers* **56**, 232-256.
- [146] Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica Section D-Biological Crystallography* **54**, 905-921.

- [147] Jucker, F. M. & Pardi, A. (1995). Solution Structure of the Cuug Hairpin Loop - a Novel RNA Tetraloop Motif. *Biochemistry* **34**, 14416-14427.
- [148] Clore, G. M., Gronenborn, A. M. & Bax, A. (1998). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *Journal of Magnetic Resonance* **133**, 216-221.
- [149] Clore, G. M., Gronenborn, A. M. & Tjandra, N. (1998). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *Journal of Magnetic Resonance* **131**, 159-162.
- [150] Zhang Q, T. R., Pitt S. W., Serganov A, Al.Hashimi H. M. (2003). Probing Motions between Equivalent RNA Domains Using Magnetic Field Induced Residual Dipolar Couplings: Accounting for Correlations between Motions and Allignment. *Journal of American Chemical Society* **125**, 10530-10531.
- [151] Meiler, J., Peti, W. & Griesinger, C. (2000). DipoCoup: A versatile program for 3D-structure homology comparison based on residual dipolar couplings and pseudocontact shifts. *Journal of Biomolecular NMR* **17**, 283-294.
- [152] Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* **14**, 51-&.
- [153] Dayie, K. T., Brodsky, A. S. & Williamson, J. R. (2002). Base flexibility in HIV-2 TAR RNA mapped by solution N-15,C-13 NMR relaxation. *Journal of Molecular Biology* **317**, 263-278.
- [154] Latham, M. R., Brown, D. J., McCallum, S. A. & Pardi, A. (2005). NMR methods for studying the structure and dynamics of RNA. *Chembiochem* **6**, 1492-1505.
- [155] Fielding, L. (2003). NMR methods for the determination of protein-ligand dissociation constants. *Current Topics in Medicinal Chemistry* **3**, 39-53.
- [156] Faber, C., Sticht, H., Schweimer, K. & Rosch, P. (2000). Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of Neomycin B. *Journal of Biological Chemistry* **275**, 20660-20666.
- [157] Weis, W. I. & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry* **65**, 441-473.
- [158] Weeks, K. M. & Crothers, D. M. (1991). RNA Recognition by Tat-Derived Peptides - Interaction in the Major Groove. *Cell* **66**, 577-588.
- [159] Huthoff, H., Girard, F., Wijmenga, S. S. & Berkhout, B. (2004). Evidence for a base triple in the free HIV-1 TAR RNA. *Rna-a Publication of the Rna Society* **10**, 412-423.
- [160] Shajani, Z., Deka, P. & Varani, G. (2006). Decoding RNA motional codes. *Trends in Biochemical Sciences* **31**, 421-424.
- [161] Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J. & Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins-Structure Function and Genetics* **50**, 496-506.
- [162] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **79**, 926-935.
- [163] Hermann, T. & Westhof, E. (1998). Saccharide-RNA recognition. *Biopolymers* **48**, 155-65.
- [164] Hermann, T. & Westhof, E. (1999). Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J Med Chem* **42**, 1250-61.
- [165] Chang, H. K., Lisziewicz, J., Gallo, R. C. & Ensoli, B. (1993). Block of HIV-1 Gene-Expression by Combining Antisense Tat RNA and Poly-TAR Decoy. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **6**, 706-706.
- [166] Cheng, A. C., Calabro, V. & Frankel, A. D. (2001). Design of RNA-binding proteins and ligands. *Current Opinion in Structural Biology* **11**, 478-484.
- [167] Gelman, M. A., Richter, S., Cao, H., Umezawa, N., Gellman, S. H. & Rana, T. M. (2003). Selective binding of TAR RNA by a tat-derived beta-peptide. *Organic Letters* **5**, 3563-3565.
- [168] Hamy, F., Brondani, V., Florsheimer, A., Stark, W., Blommers, M. J. J. & Klimkait, T. (1998). A new class of HIV-1 Tat antagonist acting through Tat-TAR inhibition. *Biochemistry* **37**, 5086-5095.
- [169] Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., AboulEla, F., Varani, G., Karn, J. & Klimkait, T. (1997). An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3548-3553.
- [170] Leeper, T. C., Athanassiou, Z., Dias, R. L. A., Robinson, J. A. & Varani, G. (2005). TAR RNA recognition by a cyclic peptidomimetic of Tat protein. *Biochemistry* **44**, 12362-12372.

-
- [171] Lavery, R. & Sklenar, H. (1988). The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic-Acids. *Journal of Biomolecular Structure & Dynamics* **6**, 63-91.
- [172] Colvin, R. A., White, S. W., Garciblanco, M. A. & Hoffman, D. W. (1993). Structural Features of an Rna Containing the Cuggga Loop of the Human-Immunodeficiency-Virus Type-1 Transactivation Response Element. *Biochemistry* **32**, 1105-1112.
- [173] Jaeger, J. A. & Tinoco, I. (1993). An NMR-Study of the HIV-1 TAR Element Hairpin. *Biochemistry* **32**, 12522-12530.
- [174] Kulinski, T., Olejniczak, M., Huthoff, H., Bielecki, L., Pachulska-Wieczorek, K., Das, A. T., Berkhout, B. & Adamiak, R. W. (2003). The apical loop of the HIV-1 TAR RNA hairpin is stabilized by a cross-loop base pair. *Journal of Biological Chemistry* **278**, 38892-38901.
- [175] Feng, S. & Holland, E. C. (1988). HIV-1 Tat Trans-Activation Requires the Loop Sequence within TAR. *Nature* **334**, 165-167.
- [176] Richter, S., Cao, H. & Rana, T. M. (2002). Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. *Biochemistry* **41**, 6391-6397.
- [177] Huq, I. & Rana, T. M. (1997). Probing the proximity of the core domain of an HIV-1 Tat fragment in a Tat-TAR complex by affinity cleaving. *Biochemistry* **36**, 12592-12599.
- [178] Espinoza-Fonseca, L. M. (2006). The benefits of the multi-target approach in drug design and discovery. *Bioorganic & Medicinal Chemistry* **14**, 896-897.

assign (resi 17 and name O6)(resi 45 and name N4)	2.91	0.01	0.01	assign (resi 35 and name H1')(resi 2 and name H2*)			
assign (resi 17 and name H1)(resi 45 and name N3)	1.89	0.20	0.20	or (resi 2 and name H3*)	3.50	1.00	1.50
assign (resi 17 and name N1)(resi 45 and name N3)	2.95	0.01	0.01	assign ((resi 35 and name H2')			
assign (resi 17 and name HN")(resi 45 and name O2)	2.08	0.20	0.20	or (resi 35 and name H3'))(resi 3 and name H3)	3.50	0.50	3.50
assign (resi 17 and name N2)(resi 45 and name O2)	2.86	0.01	0.01	assign ((resi 35 and name H2')			
assign (resi 17 and name N2)(resi 45 and name N3)	3.65	0.01	0.01	or (resi 35 and name H3'))(resi 2 and name H4*)	3.50	0.50	2.00
assign (resi 17 and name O6)(resi 45 and name O2)	5.42	0.01	0.01	assign ((resi 35 and name H2')			
assign (resi 44 and name O6)(resi 18 and name HN")	1.71	0.20	0.20	or (resi 35 and name H3'))((resi 2 and name H2*)			
assign (resi 44 and name O6)(resi 18 and name N4)	2.91	0.01	0.01	or (resi 2 and name H3*))	3.50	0.50	2.00
assign (resi 44 and name H1)(resi 18 and name N3)	1.89	0.20	0.20	assign ((resi 35 and name H2')			
assign (resi 44 and name N1)(resi 18 and name N3)	2.95	0.01	0.01	or (resi 35 and name H3'))(resi 3 and name H1)	3.50	0.50	3.50
assign (resi 44 and name HN")(resi 18 and name O2)	2.08	0.20	0.20	assign (resi 35 and name H4')(resi 3 and name H1)	3.50	0.50	3.50
assign (resi 44 and name N2)(resi 18 and name O2)	2.86	0.01	0.01	assign (resi 35 and name H4')((resi 3 and name H3)			
assign (resi 44 and name N2)(resi 18 and name N3)	3.65	0.01	0.01	or (resi 3 and name H5))	3.50	0.50	2.00
assign (resi 44 and name O6)(resi 18 and name O2)	5.42	0.01	0.01	assign (resi 35 and name H4')(resi 2 and name H4*)	3.50	0.50	2.00
assign (resi 43 and name O6)(resi 19 and name HN")	1.71	0.20	0.20	assign (resi 35 and name H4')(resi 3 and name H4)	3.50	0.50	3.50
assign (resi 43 and name N2)(resi 19 and name N4)	2.91	0.01	0.01	assign (resi 35 and name H4')((resi 2 and name H2*)			
assign (resi 43 and name H1)(resi 19 and name N3)	1.89	0.20	0.20	or (resi 2 and name H3*))	3.50	0.50	2.00
assign (resi 43 and name N1)(resi 19 and name N3)	2.95	0.01	0.01	assign (resi 35 and name H2)(resi 3 and name H1)	3.50	0.50	2.00
assign (resi 43 and name HN")(resi 19 and name O2)	2.08	0.20	0.20	assign (resi 35 and name H2)((resi 3 and name H3)			
assign (resi 43 and name N2)(resi 19 and name O2)	2.86	0.01	0.01	or (resi 3 and name H5))	3.50	0.50	2.00
assign (resi 43 and name N2)(resi 19 and name N3)	3.65	0.01	0.01	assign (resi 35 and name H2)(resi 2 and name H4*)	3.50	0.50	2.00
assign (resi 43 and name O6)(resi 19 and name O2)	5.42	0.01	0.01	assign (resi 35 and name H8)(resi 3 and name H3)	3.50	0.50	2.00
assign (resi 21 and name O6)(resi 41 and name HN")	1.71	0.20	0.20	assign (resi 35 and name H8)(resi 2 and name H4*)	3.50	0.50	3.50
assign (resi 21 and name O6)(resi 41 and name N4)	2.91	0.01	0.01	assign (resi 35 and name H8)(resi 3 and name H4)	3.50	0.50	3.50
assign (resi 21 and name H1)(resi 41 and name N3)	1.89	0.20	0.20	assign (resi 35 and name H8)(resi 3 and name H2*)	3.50	0.50	3.50
assign (resi 21 and name N1)(resi 41 and name N3)	2.95	0.01	0.01	assign (resi 35 and name H8)(resi 3 and name H6*)	3.50	0.50	2.00
assign (resi 21 and name HN")(resi 41 and name O2)	2.08	0.20	0.20	assign (resi 22 and name H8)(resi 7 and name H1)	3.50	0.50	2.00
assign (resi 21 and name N2)(resi 41 and name O2)	2.86	0.01	0.01	assign (resi 22 and name H8)(resi 7 and name H3)	3.50	0.50	2.00
assign (resi 21 and name N2)(resi 41 and name N3)	3.65	0.01	0.01	assign (resi 22 and name H8)(resi 6 and name H4*)	3.50	0.50	2.00
assign (resi 21 and name O6)(resi 41 and name O2)	5.42	0.01	0.01	assign (resi 22 and name H8)((resi 6 and name H2*)			
assign (resi 26 and name O6)(resi 39 and name HN")	1.71	0.20	0.20	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 26 and name O6)(resi 39 and name N4)	2.91	0.01	0.01	assign (resi 23 and name H1')(resi 6 and name H2*)			
assign (resi 26 and name H1)(resi 39 and name N3)	1.89	0.20	0.20	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 26 and name N1)(resi 39 and name N3)	2.95	0.01	0.01	assign (resi 23 and name H2')(resi 6 and name H1*)			
assign (resi 26 and name HN")(resi 39 and name O2)	2.08	0.20	0.20	or (resi 6 and name H4*))	3.50	0.50	2.00
assign (resi 26 and name N2)(resi 39 and name O2)	2.86	0.01	0.01	assign (resi 23 and name H2)((resi 6 and name H2*)			
assign (resi 26 and name N2)(resi 39 and name N3)	3.65	0.01	0.01	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 26 and name O6)(resi 39 and name O2)	5.42	0.01	0.01	assign (resi 23 and name H5)((resi 5 and name H5)			
assign (resi 28 and name O6)(resi 37 and name HN")	1.71	0.20	0.20	or (resi 7 and name H3))	3.50	0.50	2.00
assign (resi 28 and name O6)(resi 37 and name N4)	2.91	0.01	0.01	assign (resi 23 and name H5)((resi 6 and name H1*)			
assign (resi 28 and name H1)(resi 37 and name N3)	1.89	0.20	0.20	or (resi 6 and name H4*))	3.50	0.50	3.50
assign (resi 28 and name N1)(resi 37 and name N3)	2.95	0.01	0.01	assign (resi 23 and name H5)((resi 6 and name H2*)			
assign (resi 28 and name HN")(resi 37 and name O2)	2.08	0.20	0.20	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 28 and name N2)(resi 37 and name O2)	2.86	0.01	0.01	assign (resi 25 and name H1')(resi 5 and name H1)	3.50	0.50	2.00
assign (resi 28 and name N2)(resi 37 and name N3)	3.65	0.01	0.01	assign (resi 25 and name H1')(resi 5 and name H3)			
assign (resi 28 and name O6)(resi 37 and name O2)	5.42	0.01	0.01	or (resi 5 and name H5))	3.50	0.50	2.00
assign (resi 36 and name O6)(resi 29 and name HN")	1.71	0.20	0.20	assign (resi 25 and name H1')(resi 4 and name H1*)			
assign (resi 36 and name O6)(resi 29 and name N4)	2.91	0.01	0.01	or (resi 6 and name H4*))	3.50	0.50	2.00
assign (resi 36 and name H1)(resi 29 and name N3)	1.89	0.20	0.20	assign (resi 25 and name H1)((resi 4 and name H2*)			
assign (resi 36 and name N1)(resi 29 and name N3)	2.95	0.01	0.01	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 36 and name HN")(resi 29 and name O2)	2.08	0.20	0.20	assign (resi 25 and name H1')(resi 5 and name H6*)	3.50	0.50	2.00
assign (resi 36 and name N2)(resi 29 and name O2)	2.86	0.01	0.01	assign (resi 25 and name H5)(resi 4 and name H1*)			
assign (resi 36 and name N2)(resi 29 and name N3)	3.65	0.01	0.01	or (resi 6 and name H4*))	3.50	0.50	2.00
assign (resi 36 and name O6)(resi 29 and name O2)	5.42	0.01	0.01	assign (resi 25 and name H5)((resi 4 and name H2*)			
assign (resi 20 and name N1)(resi 42 and name H3)	1.93	0.20	0.20	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 20 and name N1)(resi 42 and name N3)	2.82	0.01	0.01	assign (resi 25 and name H5)(resi 5 and name H6*)	3.50	0.50	2.00
assign (resi 20 and name HN")(resi 42 and name O4)	1.82	0.20	0.20	assign (resi 25 and name H6)(resi 5 and name H6*)	3.50	0.50	2.00
assign (resi 20 and name N6)(resi 42 and name O4)	2.95	0.01	0.01	assign (resi 25 and name H6)((resi 4 and name H2*)			
assign (resi 27 and name N1)(resi 38 and name H3)	1.93	0.20	0.20	or (resi 6 and name H3*))	3.50	0.50	2.00
assign (resi 27 and name N1)(resi 38 and name N3)	2.82	0.01	0.01	assign ((resi 26 and name H5')			
assign (resi 27 and name HN")(resi 38 and name O4)	1.82	0.20	0.20	or (resi 26 and name H5'))(resi 5 and name H6*)	3.50	0.50	2.00
assign (resi 27 and name N6)(resi 38 and name O4)	2.95	0.01	0.01	assign (resi 36 and name H8)((resi 2 and name H4*)			
assign (resi 22 and name N1)(resi 40 and name H3)	1.93	0.20	0.20	or (resi 4 and name H1*))	5.50	3.00	0.00
assign (resi 22 and name N1)(resi 40 and name N3)	2.82	0.01	0.01	assign (resi 36 and name H8)((resi 2 and name H3*)			
assign (resi 22 and name HN")(resi 40 and name O4)	1.82	0.20	0.20	or (resi 4 and name H2*))	5.50	1.50	0.00
assign (resi 22 and name N6)(resi 40 and name O4)	2.95	0.01	0.01	assign (resi 43 and name H1)((resi 6 and name H3*)			
*****Intermolecular NOE Restraints*****				or (resi 8 and name H2*))	5.00	4.00	0.00
assign (resi 35 and name H1')(resi 3 and name H5)	3.50	0.50	2.00	assign (resi 43 and name H1')(resi 7 and name H1)	5.00	4.00	0.00
assign (resi 35 and name H1')(resi 2 and name H4*)	3.50	0.50	3.50	assign (resi 43 and name H1')(resi 7 and name H2*)	5.00	3.00	0.00

```

assign (resi 44 and name H4*)((resi 7 and name H3)
      or (resi 7 and name H5)) 5.00 4.00 0.00
assign (resi 44 and name H4*)((resi 6 and name H3*)
      or (resi 8 and name H2*)) 5.00 4.00 0.00
assign (resi 44 and name H4')(resi 7 and name H6*) 5.00 4.00 0.00
assign (resi 43 and name H1')(resi 2 and name H3*)
      or (resi 4 and name H2*)) 5.00 4.00 0.00
assign (resi 43 and name H1')(resi 3 and name H1) 5.00 4.00 0.00
assign (resi 43 and name H1')(resi 3 and name H2*) 5.00 3.00 0.00
assign (resi 44 and name H4')(resi 3 and name H3)
      or (resi 3 and name H5)) 5.00 4.00 0.00
assign (resi 44 and name H4')(resi 2 and name H2*)
      or (resi 4 and name H2*)) 5.00 4.00 0.00
assign (resi 44 and name H4')(resi 3 and name H6*) 5.00 4.00 0.00

```

!*****Planarity Restraints *****

!Format and syntax taken from XPLOR nmr/xplor/htmlman/node186.html

!It forces the base pairs to be planar.

restraints plane

!G28 - C37

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 28 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 37 )
)
weight=0.5
end

```

!G17 - C45

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 17 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 45 )
)
weight=0.5
end

```

!G36 - C29

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 36 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 29 )
)
weight=0.5
end

```

!G43 - C19

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 43 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 19 )
)
weight=0.5
end

```

!G44 - C18

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 44 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 18 )
)
weight=0.5
end

```

!G16 - C46

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 16 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 46 )
)
weight=0.5
end

```

!G26 - C39

```

group
selection=(
  (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 26 )
  or (not( name P or name O1P or name O2P or name O5' or name H5T or
        name H1' or name H2' or name H3' or name H4' or name H5' or
        name H5'' or name O2' or name HO2' or
        name O3' or name C1' or name C2' or name C3' or name C4' or
        name C5' or name O4' or name H3T) and residue 39 )
)
weight=0.5
end

```

!G21 - C41

group


```

selection=(
  (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 21 )
  or (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 41 )
)
weight=0.5
end

```

```

!A20 - U42
group
selection=(
  (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 20 )
  or (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 42 )
)
weight=0.5
end

```

```

!A27 - U38
group
selection=(
  (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 27 )
  or (not( name P  or name O1P or name O2P or name O5' or name H5T or
    name H1' or name H2' or name H3' or name H4' or name H5' or
    name H5" or name O2' or name HO2' or
    name O3' or name C1' or name C2' or name C3' or name C4' or
    name C5' or name O4' or name H3T) and residue 38 )
)
weight=0.5
end
end

```

{* The planarity energy term needs to be turned on.*}

```

flags include plan end
!*****Dihedrals and Sugar Pucker*****
!apply sugar pucker for TAR-RNA
!definitions
!O4' C1' C2' C3' nu1
!C1' C2' C3' C4' nu2
!C3' C4' O4' C1' nu4
!C2' C3' C4' O4' nu3
!C5' C4' C3' O3' delta (delta = nu3 + 125)
!C4' O4' C1' C2' nu0 (nu0 = 0 for C3' endo conformations)
!nu(x) = phi(m).cos[P + 144(x-2)]
!where x = 0,1,2,3,4
!phi(m) = pucker amplitude between 34 and 42 degrees
!P = pseudo rotation phase angle
!P = 18 for pure C3' endo
!P = 162 for pure C2' endo
!*****C2' endo*****
assign (resid 23 and name C1')
      (resid 23 and name C2')
      (resid 23 and name C3')

```

```

      (resid 23 and name C4') 10.0 -40.0 5.0 2 !nu2
assign (resid 23 and name C2')
      (resid 23 and name C3')
      (resid 23 and name C4')
      (resid 23 and name O4') 10.0 23.0 10.0 2 !nu3
assign (resid 33 and name C1')
      (resid 33 and name C2')
      (resid 33 and name C3')
      (resid 33 and name C4') 10.0 -40.0 5.0 2 !nu2
assign (resid 33 and name C2')
      (resid 33 and name C3')
      (resid 33 and name C4')
      (resid 33 and name O4') 10.0 23.0 10.0 2 !nu3
!*****c3' endo*****
!The values given below for each suger pucker angle have been taken
!from standard xplor formats.
!nu0,nu1,nu2,nu3 and nu4 (delta) have been restrained below
!A form helix values were taken from the arnarst.inp file.
assign (resid 16 and name C1')
      (resid 16 and name C2')
      (resid 16 and name C3')
      (resid 16 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 16 and name C5')
      (resid 16 and name C4')
      (resid 16 and name C3')
      (resid 16 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 17 and name C1')
      (resid 17 and name C2')
      (resid 17 and name C3')
      (resid 17 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 17 and name C5')
      (resid 17 and name C4')
      (resid 17 and name C3')
      (resid 17 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 18 and name C1')
      (resid 18 and name C2')
      (resid 18 and name C3')
      (resid 18 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 18 and name C5')
      (resid 18 and name C4')
      (resid 18 and name C3')
      (resid 18 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 19 and name C1')
      (resid 19 and name C2')
      (resid 19 and name C3')
      (resid 19 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 19 and name C5')
      (resid 19 and name C4')
      (resid 19 and name C3')
      (resid 19 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 20 and name C1')
      (resid 20 and name C2')
      (resid 20 and name C3')
      (resid 20 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 20 and name C5')
      (resid 20 and name C4')
      (resid 20 and name C3')
      (resid 20 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 21 and name C1')
      (resid 21 and name C2')
      (resid 21 and name C3')
      (resid 21 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 21 and name C5')
      (resid 21 and name C4')
      (resid 21 and name C3')
      (resid 21 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 22 and name C1')
      (resid 22 and name C2')
      (resid 22 and name C3')
      (resid 22 and name C4') 10.0 38.0 10.0 2 !nu2

```

```

assign (resid 22 and name C5')
  (resid 22 and name C4')
  (resid 22 and name C3')
  (resid 22 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 26 and name C1')
  (resid 26 and name C2')
  (resid 26 and name C3')
  (resid 26 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 26 and name C5')
  (resid 26 and name C4')
  (resid 26 and name C3')
  (resid 26 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 28 and name C1')
  (resid 28 and name C2')
  (resid 28 and name C3')
  (resid 28 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 28 and name C5')
  (resid 28 and name C4')
  (resid 28 and name C3')
  (resid 28 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 36 and name C1')
  (resid 36 and name C2')
  (resid 36 and name C3')
  (resid 36 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 36 and name C5')
  (resid 36 and name C4')
  (resid 36 and name C3')
  (resid 36 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 37 and name C1')
  (resid 37 and name C2')
  (resid 37 and name C3')
  (resid 37 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 37 and name C5')
  (resid 37 and name C4')
  (resid 37 and name C3')
  (resid 37 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 38 and name C1')
  (resid 38 and name C2')
  (resid 38 and name C3')
  (resid 38 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 38 and name C5')
  (resid 38 and name C4')
  (resid 38 and name C3')
  (resid 38 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 39 and name C1')
  (resid 39 and name C2')
  (resid 39 and name C3')
  (resid 39 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 39 and name C5')
  (resid 39 and name C4')
  (resid 39 and name C3')
  (resid 39 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 40 and name C1')
  (resid 40 and name C2')
  (resid 40 and name C3')
  (resid 40 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 40 and name C5')
  (resid 40 and name C4')
  (resid 40 and name C3')
  (resid 40 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 41 and name C1')
  (resid 41 and name C2')
  (resid 41 and name C3')
  (resid 41 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 41 and name C5')
  (resid 41 and name C4')
  (resid 41 and name C3')
  (resid 41 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 42 and name C1')
  (resid 42 and name C2')
  (resid 42 and name C3')
  (resid 42 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 43 and name C1')
  (resid 43 and name C2')
  (resid 43 and name C3')
  (resid 43 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 43 and name C5')
  (resid 43 and name C4')
  (resid 43 and name C3')
  (resid 43 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 44 and name C1')
  (resid 44 and name C2')
  (resid 44 and name C3')
  (resid 44 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 44 and name C5')
  (resid 44 and name C4')
  (resid 44 and name C3')
  (resid 44 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 45 and name C1')
  (resid 45 and name C2')
  (resid 45 and name C3')
  (resid 45 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 45 and name C5')
  (resid 45 and name C4')
  (resid 45 and name C3')
  (resid 45 and name O3') 10.0 82.0 5.0 2 !delta
assign (resid 46 and name C1')
  (resid 46 and name C2')
  (resid 46 and name C3')
  (resid 46 and name C4') 10.0 38.0 10.0 2 !nu2
assign (resid 46 and name C5')
  (resid 46 and name C4')
  (resid 46 and name C3')
  (resid 46 and name O3') 10.0 82.0 5.0 2 !delta
!*****
!below I have the backbone angles for the stretch between 16 and 21.
assign (resid 16 and name C3')
  (resid 16 and name O3')
  (resid 17 and name P)
  (resid 17 and name O5') 10.0 -60.0 60.0 2 {zeta}
assign (resid 16 and name O3')
  (resid 17 and name P)
  (resid 17 and name O5')
  (resid 17 and name C5') 10.0 -60.0 60.0 2 {alpha}
assign (resid 16 and name C4')
  (resid 16 and name C3')
  (resid 16 and name O3')
  (resid 17 and name P) 10.0 240.0 60.0 2 {epsilon}
assign (resid 17 and name P)
  (resid 17 and name O5')
  (resid 17 and name C5')
  (resid 17 and name C4') 10.0 180.0 60.0 2 {beta}
assign (resid 17 and name C3')
  (resid 17 and name O3')
  (resid 18 and name P)
  (resid 18 and name O5') 10.0 -60.0 60.0 2 {zeta}
assign (resid 17 and name O3')
  (resid 18 and name P)
  (resid 18 and name O5')
  (resid 18 and name C5') 10.0 -60.0 60.0 2 {alpha}
assign (resid 17 and name C4')
  (resid 17 and name C3')
  (resid 17 and name O3')
  (resid 18 and name P) 10.0 240.0 60.0 2 {epsilon}
assign (resid 18 and name P)
  (resid 18 and name O5')

```

```

( resid 18 and name C5'
( resid 18 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 18 and name C3'
( resid 18 and name O3' )
( resid 19 and name P )
( resid 19 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 18 and name O3' )
( resid 19 and name P )
( resid 19 and name O5' )
( resid 19 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 18 and name C4' )
( resid 18 and name C3' )
( resid 18 and name O3' )
( resid 19 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 19 and name P )
( resid 19 and name O5' )
( resid 19 and name C5' )
( resid 19 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 19 and name C3' )
( resid 19 and name O3' )
( resid 20 and name P )
( resid 20 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 19 and name O3' )
( resid 20 and name P )
( resid 20 and name O5' )
( resid 20 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 19 and name C4' )
( resid 19 and name C3' )
( resid 19 and name O3' )
( resid 20 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 20 and name P )
( resid 20 and name O5' )
( resid 20 and name C5' )
( resid 20 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 20 and name C3' )
( resid 20 and name O3' )
( resid 21 and name P )
( resid 21 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 20 and name O3' )
( resid 21 and name P )
( resid 21 and name O5' )
( resid 21 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 20 and name C4' )
( resid 20 and name C3' )
( resid 20 and name O3' )
( resid 21 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 21 and name P )
( resid 21 and name O5' )
( resid 21 and name C5' )
( resid 21 and name C4')10.0 180.0 60.0 2 {beta}
!*****
!below i define the backbone angles for the stretch between 41 and 46.

assign ( resid 41 and name P )
( resid 41 and name O5' )
( resid 41 and name C5' )
( resid 41 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 41 and name C3' )
( resid 41 and name O3' )
( resid 42 and name P )
( resid 42 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 41 and name O3' )
( resid 42 and name P )
( resid 42 and name O5' )
( resid 42 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 41 and name C4' )
( resid 41 and name C3' )
( resid 41 and name O3' )
( resid 42 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 42 and name P )

( resid 42 and name O5' )
( resid 42 and name C5' )10.0 180.0 60.0 2 {beta}
assign ( resid 42 and name C3' )
( resid 42 and name O3' )
( resid 43 and name P )
( resid 43 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 42 and name O3' )
( resid 43 and name P )
( resid 43 and name O5' )
( resid 43 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 42 and name C4' )
( resid 42 and name C3' )
( resid 42 and name O3' )
( resid 43 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 43 and name P )
( resid 43 and name O5' )
( resid 43 and name C5' )
( resid 43 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 43 and name C3' )
( resid 43 and name O3' )
( resid 44 and name P )
( resid 44 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 43 and name O3' )
( resid 44 and name P )
( resid 44 and name O5' )
( resid 44 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 43 and name C4' )
( resid 43 and name C3' )
( resid 43 and name O3' )
( resid 44 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 44 and name P )
( resid 44 and name O5' )
( resid 44 and name C5' )
( resid 44 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 44 and name C3' )
( resid 44 and name O3' )
( resid 45 and name P )
( resid 45 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 44 and name O3' )
( resid 45 and name P )
( resid 45 and name O5' )
( resid 45 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 44 and name C4' )
( resid 44 and name C3' )
( resid 44 and name O3' )
( resid 45 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 45 and name P )
( resid 45 and name O5' )
( resid 45 and name C5' )
( resid 45 and name C4')10.0 180.0 60.0 2 {beta}
assign ( resid 45 and name C3' )
( resid 45 and name O3' )
( resid 46 and name P )
( resid 46 and name O5' ) 10.0 -60.0 60.0 2 {zeta}
assign ( resid 45 and name O3' )
( resid 46 and name P )
( resid 46 and name O5' )
( resid 46 and name C5' ) 10.0 -60.0 60.0 2 {alpha}
assign ( resid 45 and name C4' )
( resid 45 and name C3' )
( resid 45 and name O3' )
( resid 46 and name P ) 10.0 240.0 60.0 2 {epsilon}
assign ( resid 46 and name P )
( resid 46 and name O5' )
( resid 46 and name C5' )
( resid 46 and name C4')10.0 180.0 60.0 2 {beta}
!*****End Restraints*****

```

RESTRAINED MOLECULAR DYNAMICS PROTOCOL**1) HIGH TEMPERATURE**

- a) Temperature: 1000K
- b) Number of structures generated: 150
- c) Energy terms: bonds, angles, Vdw, NOE, plan, impr and cdih
- d) Force constants employed:
 - NOE=50.0
 - Cdih=300.0
 - Vdw=4.0
- e) Energy cutoff 600 kcal/mol (total) and 100 kcal/mol (NOE)

2) SIMULATED ANNEALING

100 steps of Powell energy minimization including bonds, vdw, NOE, angl, plan, cdih and impr

- a) 3 rounds of 3ps dynamics at 1000, 600 and 400K with 700 steps of Powell minimization between the dynamics.
- b) 1000 Powell energy minimization in the end.
- c) Hydrogen bonding restraints between A22-U40 removed
- d) Energy cutoff 400 kcal/mol (total) and 50 kcal/mol (NOE)

3) REFINEMENT

- a) 2ps dynamics at 500, 300, 150, 50 and 10K each with 1000 steps of Powell minimization between each step.
- b) Calculations ended with 2000 steps of Powell minimization.
- c) Energy cutoff 25 kcal/mol (NOE)
- d) All chosen structures had only 0 to 4 NOE violations greater than 0.2Å and none greater than 0.5 Å and no angular violations greater than 5°.

4) REFINEMENT WITH *f*IRDCs

- a) Switch on attractive Lennard-Jones non-bonded parameter.
- b) 75 cycles, 1000 steps of molecular dynamics at 300K during which K_{sani} was increased from 0.001 to 1.5 kcal/mol Hz⁻².
- c) 3 rounds of 5ps dynamics at 300, 150, 50 and 10K while all force constants were switched to their maximum value.
- d) 2000 steps of Powell minimization.

Protocols for initial structure generation, simulated annealing, refinement
refinement with field induced Residual Dipolar Couplings used in XPLOR-NIH
for the generation of structures of HIV-2 TAR-RNA/AMG.

```
!*****
!*REMARKS This protocol allows gradual introduction of restraints *
!*REMARKS at High Temperature with Verlet Dynamics and Powell Minimization *
!*REMARKS and generation of Multiple starting structures *
!*****
evaluate ($seed=(264437) + 976876)
set seed $seed end
!*****
! read in the PSF file and initial structure

parameter
  @ (Read Parameter file here)
end

structure
  @ (Read Topology/PSF file here)
end

coordinates @ (Read TAR co-ordinates here)
              (Read AMG co-ordinates here)
              (Read Axis co-ordinates here)
!*****
! set the weights for initial experimental energy terms
!*****Loop D*****
evaluate ($lf_cool_steps = 4000)
evaluate ($lf_init_t = 1500.00)
evaluate ($lf_final_t = 1000.00)
evaluate ($lf_tempstep = 10)
evaluate ($lf_in_ang = 0.5)      evaluate ($lf_fin_ang = 1.0)
evaluate ($lf_in_imp = 0.1)     evaluate ($lf_fin_imp = 1.0)
evaluate ($lf_in_con= 0.002)   evaluate ($lf_fin_con= 4.0)
evaluate ($lf_in_noe = 20.0)   evaluate ($lf_fin_noe = 50.0)
evaluate ($lf_in_cdih = 100.0) evaluate ($lf_fin_cdih = 300.0)
evaluate ($lf_in_rad = 0.9)    evaluate ($lf_fin_rad = 0.80)

!*****Loop C*****
evaluate ($l3_cool_steps = 2000)
evaluate ($l3_init_t = 1750.00)
evaluate ($l3_final_t = 1500.00)
evaluate ($l3_tempstep = 25)
evaluate ($l3_fin_ang = $lf_in_ang)
evaluate ($l3_in_ang = $l3_fin_ang / 5)
evaluate ($l3_fin_imp = $lf_in_imp)
evaluate ($l3_in_imp = $l3_fin_imp / 5)
evaluate ($l3_fin_con = $lf_in_con)
evaluate ($l3_in_con = $l3_fin_con / 5)
evaluate ($l3_fin_noe = $lf_in_noe)
evaluate ($l3_in_noe = $l3_fin_noe / 5)
evaluate ($l3_fin_cdih = $lf_in_cdih)
evaluate ($l3_in_cdih = $l3_fin_cdih / 5)
evaluate ($l3_fin_rad = $lf_in_rad)
evaluate ($l3_in_rad = $l3_fin_rad / 5)
```

```

|*****Loop B*****
evaluate ($l2_cool_steps = 1000)
evaluate ($l2_init_t = 2000.00)
evaluate ($l2_final_t = 1750.00)
evaluate ($l2_tempstep = 50)
evaluate ($l2_fin_ang = $l3_in_ang)
evaluate ($l2_in_ang = $l2_fin_ang / 5)
evaluate ($l2_fin_imp = $l3_in_imp)
evaluate ($l2_in_imp = $l2_fin_imp / 5)
evaluate ($l2_fin_con = $l3_in_con)
evaluate ($l2_in_con = $l2_fin_con / 5)
evaluate ($l2_fin_noe = $l3_in_noe)
evaluate ($l2_in_noe = $l2_fin_noe / 5)
evaluate ($l2_fin_cdiH = $l3_in_cdiH)
evaluate ($l2_in_cdiH = $l2_fin_cdiH / 5)
evaluate ($l2_fin_rad = $l3_in_rad)
evaluate ($l2_in_rad = $l2_fin_rad / 5)

|*****Loop A*****
evaluate ($l1_cool_steps = 500)
evaluate ($l1_init_t = 2500.00)
evaluate ($l1_final_t = 2000.00)
evaluate ($l1_tempstep = 50.0)
evaluate ($l1_fin_ang = $l2_in_ang)
evaluate ($l1_in_ang = $l1_fin_ang / 5)
evaluate ($l1_fin_imp = $l2_in_imp)
evaluate ($l1_in_imp = $l1_fin_imp / 5)
evaluate ($l1_fin_con = $l2_in_con)
evaluate ($l1_in_con = $l1_fin_con / 5)
evaluate ($l1_fin_noe = $l2_in_noe)
evaluate ($l1_in_noe = $l1_fin_noe / 5)
evaluate ($l1_fin_cdiH = $l2_in_cdiH)
evaluate ($l1_in_cdiH = $l1_fin_cdiH / 5)
evaluate ($l1_fin_rad = $l2_in_rad)
evaluate ($l1_in_rad = $l1_fin_rad / 5)
|*****
evaluate ($knoe = 50.0)
evaluate ($asym = 1.0)
|*****
noe
  reset
  nres = 4000
  class all
  @ (Read NOE restraints here)
  @ (Read H-Bonding restraints here)
end

noe
  ceiling 1000
  averaging * sum
  potential * square
  scale * $knoe
  sqconstant * 1.
  sqexponent * 2
end
|*****
restraints dihedral
  nass = 500
  @ (Read Sugar Pucker & Dihedral restraints here)
end

```

```

coor copy end
|*****
@ (Read Planarity restraints here)
|*****
!the input file below creates a form helix for the stems
@ (A-form restraints for RNA-helices)
|*****
set echo on message on end
evaluate ($rcon = 0.003)
|*****
parameters
  nbonds
  atom
  nbxmod 3  wmin = 0.01  cutnb=4.5  ctonnb=2.99  ctofnb=3
  tolerance 0.5  repel= 0.9  rexp = 2  irex = 2
  rcon=$rcon
end
end
|*****
set message on echo on end

vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)

constraints fix ((resid 500 or resid 501) and name OO) end

vector do (mass = 100.0) (not (resid 500 or resid 501))
vector do (mass = 30.0) (resid 500 or resid 501)
vector do (fbeta = 10.0) (all)
|*****

{* Generate Structures 1 ->N}
evaluate ($count =0)
evaluate ($end_count = N)

|*****
while ($count < $end_count) loop main
  evaluate ($count = $count + 1)
  set seed $count end
  coor swap end
  coor copy end

  vector do (refx=x) (all)
  vector do (refy=y) (all)
  vector do (refz=z) (all)

  noe scale * $knoe end
|*****Loop A*****
  evaluate ($ncycle = ($l1_init_t-$l1_final_t)/$l1_tempstep)
  evaluate ($nstep =int($l1_cool_steps/$ncycle))
  evaluate ($bath = $l1_init_t)
  evaluate ($radfact = ($l1_fin_rad/$l1_in_rad)^(1/$ncycle))
  evaluate ($radius= $l1_in_rad)
  evaluate ($ang_fac = ($l1_fin_ang/$l1_in_ang)^(1/$ncycle))
  evaluate ($k_ang = $l1_in_ang)
  evaluate ($simp_fac = ($l1_fin_imp/$l1_in_imp)^(1/$ncycle))
  evaluate ($k_imp = $l1_in_imp)
  evaluate ($k_vdwfact = ($l1_fin_con/$l1_in_con)^(1/$ncycle))
  evaluate ($k_vdw = $l1_in_con)

```

```

evaluate ($noe_fac = ($l1_fin_noe/$l1_in_noe)^(1/$ncycle))
evaluate ($knoe = $l1_in_noe)
evaluate ($cdih_fac = ($l1_fin_cdih/$l1_in_cdih)^(1/$ncycle))
evaluate ($kcdih = $l1_in_cdih)

flags
  exclude *
  include bonds vdw noe angl plan impr cdih
end

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)
  evaluate ($bath = 1000) ! mantaining 1000 K
  evaluate ($k_vdw=min($l1_fin_con,$k_vdw*$k_vdwfact))
  evaluate ($radius=max($l1_fin_rad,$radius*$radfact))
  evaluate ($k_ang = $k_ang*$ang_fac)
  evaluate ($k_imp = $k_imp*$imp_fac)
  evaluate ($knoe = $knoe*$noe_fac)
  evaluate ($kcdih=min($l1_fin_cdih,$kcdih*$cdih_fac))

  constraints interaction (all) (all) weights
    * 1 angles $k_ang improper $k_imp
  end end

  parameter
    nbonds
    cutnb=4.5 rcon=$k_vdw nbxmod=3 repel=$radius
  end
end

noe scale * $knoe end

restraints dihedral
  scale $kcdih
end

dynamics verlet
  nstep=$nstep timestep=0.003 iasvel=current firsttemp= $bath
  tcoupling = true tbath = $bath nprint=500 iprfreq=0
  ntrfr = 99999999
end
end loop cool
!*****Loop B*****
evaluate ($ncycle = ($l2_init_t-$l2_final_t)/$l2_tempstep)
evaluate ($nstep = int($l2_cool_steps/$ncycle))
evaluate ($bath = $l2_init_t)
evaluate ($radfact = ($l2_fin_rad/$l2_in_rad)^(1/$ncycle))
evaluate ($radius= $l2_in_rad)
evaluate ($ang_fac = ($l2_fin_ang/$l2_in_ang)^(1/$ncycle))
evaluate ($k_ang = $l2_in_ang)
evaluate ($imp_fac = ($l2_fin_imp/$l2_in_imp)^(1/$ncycle))
evaluate ($k_imp = $l2_in_imp)
evaluate ($k_vdwfact = ($l2_fin_con/$l2_in_con)^(1/$ncycle))
evaluate ($k_vdw = $l2_in_con)

```

```

evaluate ($noe_fac = ($l2_fin_noe/$l2_in_noe)^(1/$ncycle))
evaluate ($knoe = $l2_in_noe)
evaluate ($cdih_fac = ($l2_fin_cdih/$l2_in_cdih)^(1/$ncycle))
evaluate ($kcdih = $l2_in_cdih)

```

```

flags
  exclude *
  include bonds angl impr vdw noe plan cdih
end

```

```

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)
  evaluate ($bath = 1000) ! maintaining 1000 K
  evaluate ($k_vdw=min($l2_fin_con,$k_vdw*$k_vdwfact))
  evaluate ($radius=max($l2_fin_rad,$radius*$radfact))
  evaluate ($k_ang = $k_ang*$ang_fac)
  evaluate ($k_imp = $k_imp*$imp_fac)
  evaluate ($knoe = $knoe*$noe_fac)
  evaluate ($kcdih=min($l2_fin_cdih,$kcdih*$cdih_fac))

```

```

constraints interaction (all) (all) weights
  * 1 angles $k_ang improper $k_imp
end end

```

```

parameter
  nbonds
  cutnb=4.5 rcon=$k_vdw nbxmod=3 repel=$radius
end
end

```

```

noe scale * $knoe end

```

```

restraints dihedral
  scale $kcdih
end

```

```

dynamics verlet
  nstep=$nstep timestep=0.003 iasvel=current firsttemp= $bath
  tcoupling = true tbath = $bath nprint=500 iprfreq=0
  ntrfr = 99999999
end
end loop cool

```

```

!*****Loop C*****

```

```

evaluate ($ncycle = ($l3_init_t-$l3_final_t)/$l3_tempstep)
evaluate ($nstep = int($l3_cool_steps/$ncycle))
evaluate ($bath = $l3_init_t)
evaluate ($radfact = ($l3_fin_rad/$l3_in_rad)^(1/$ncycle))
evaluate ($radius= $l3_in_rad)
evaluate ($ang_fac = ($l3_fin_ang/$l3_in_ang)^(1/$ncycle))
evaluate ($k_ang = $l3_in_ang)
evaluate ($imp_fac = ($l3_fin_imp/$l3_in_imp)^(1/$ncycle))
evaluate ($k_imp = $l3_in_imp)
evaluate ($k_vdwfact = ($l3_fin_con/$l3_in_con)^(1/$ncycle))
evaluate ($k_vdw = $l3_in_con)
evaluate ($noe_fac = ($l3_fin_noe/$l3_in_noe)^(1/$ncycle))
evaluate ($knoe = $l3_in_noe)
evaluate ($cdih_fac = ($l3_fin_cdih/$l3_in_cdih)^(1/$ncycle))
evaluate ($kcdih = $l3_in_cdih)

```

```

flags
  exclude *
  include bonds angl impr vdw noe plan cdih
end

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)
  evaluate ($bath = 1000) ! maintaining 1000 K
  evaluate ($k_vdw=min($l3_fin_con,$k_vdw*$k_vdwfact))
  evaluate ($radius=max($l3_fin_rad,$radius*$radfact))
  evaluate ($k_ang = $k_ang*$ang_fac)
  evaluate ($k_imp = $k_imp*$imp_fac)
  evaluate ($knoe = $knoe*$noe_fac)
  evaluate ($kcdih=min($l3_fin_cdih,$kcdih*$cdih_fac))

constraints interaction (all) (all) weights
  * 1 angles $k_ang improper $k_imp
end end

parameter
  nbonds
  cutnb=4.5 rcon=$k_vdw nbxmod=3 repel=$radius
end
end

noe scale all $knoe end

restraints dihedral
  scale $kcdih
end

dynamics verlet
  nstep=$nstep timestep=0.003 iasvel=current firsttemp=$bath
  tcoupling = true tbath = $bath nprint=500 iprfreq=0
  ntrfr = 99999999
end
end loop cool
!*****Loop D*****
evaluate ($ncycle = ($lf_init_t-$lf_final_t)/$lf_tempstep)
evaluate ($nstep =int($lf_cool_steps/$ncycle))
evaluate ($bath = $lf_init_t)
evaluate ($radfact = ($lf_fin_rad/$lf_in_rad)^(1/$ncycle))
evaluate ($radius= $lf_in_rad)
evaluate ($ang_fac = ($lf_fin_ang/$lf_in_ang)^(1/$ncycle))
evaluate ($k_ang = $lf_in_ang)
evaluate ($imp_fac = ($lf_fin_imp/$lf_in_imp)^(1/$ncycle))
evaluate ($k_imp = $lf_in_imp)
evaluate ($k_vdwfact = ($lf_fin_con/$lf_in_con)^(1/$ncycle))
evaluate ($k_vdw = $lf_in_con)
evaluate ($noe_fac = ($lf_fin_noe/$lf_in_noe)^(1/$ncycle))
evaluate ($knoe = $lf_in_noe)
evaluate ($cdih_fac = ($lf_fin_cdih/$lf_in_cdih)^(1/$ncycle))
evaluate ($kcdih = $lf_in_cdih)

flags
  exclude *
  include bonds angl impr vdw noe cdih plan
end

```

```

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
evaluate ($i_cool=$i_cool+1)
evaluate ($bath = 1000) ! maintaining 1000 K
evaluate ($k_vdw=min($lf_fin_con,$k_vdw*$k_vdwfact))
evaluate ($radius=max($lf_fin_rad,$radius*$radfact))
evaluate ($k_ang = $k_ang*$ang_fac)
evaluate ($k_imp = $k_imp*$imp_fac)
evaluate ($knoe = $knoe*$noe_fac)
evaluate ($kcdih=min($lf_fin_cdih,$kcdih*$cdih_fac))

constraints interaction (all) (all) weights
* 1 angles $k_ang improper $k_imp
end end

parameter
nbonds
cutnb=4.5 rcon=$k_vdw nbxmod=3 repel=$radius
end
end

noe scale * $knoe end

restraints dihedral
scale $kcdih
end

dynamics verlet
nstep=$nstep timestep=0.003 iasvel=current firsttemp= $bath
tcoupling = true tbath = $bath nprint=500 iprfreq=0
ntrfr = 99999999
end
end loop cool

```

```

!*****Powell Minimization*****

```

```

mini powell nstep= 2000 nprint= 20 end

```

```

!*****

```

```

set message on echo on end

```

```

!lines below this print out threshold and violations

```

```

print threshold=0.2 noe
evaluate ($rms_noe=$result)
evaluate ($viol_noe=$violations)
print threshold=3.0 cdih
evaluate ($rms_cdih=$result)
evaluate ($viol_cdih=$violations)
print threshold=0.03 bonds
evaluate ($rms_bonds=$result)
evaluate ($viol_bonds=$violations)
print threshold=2.0 angles
evaluate ($rms_angles=$result)
evaluate ($viol_angles=$violations)
print threshold=3.0 impropers
evaluate ($rms_impropers=$result)
evaluate ($viol_imp=$violations)
set message on echo on end

```

```

remarks *****

```

```

remarks overall,bonds,angles,improper,vdw,cdih,noe
remarks energies: $ener, $bond, $angl, $impr, $vdw, $cdih, $noe

```

```

remarks *****
remarks      bonds,angles,impropers,cdih,noe
remarks bonds etc: $rms_bonds,$rms_angles,$rms_impropers,$rms_cdih,$rms_noe
remarks *****
remarks      noe, cdih, bonds, angles, impropers
remarks violations:$viol_noe,$viol_cdih,$viol_bonds,$viol_angles,$viol_imp
remarks *****

set message on echo on end
  evaluate ($filename="./(Directory_Name)/(File_Name_)"+"pdb")
  write coordinates output =$filename end
end loop main
stop

!*****
!*REMARKS Annealing protocol does 3 rounds of constant temperature dynamics      *
!*REMARKS at various temperatures decreasing at 1000k 600k and 400k              *
!*REMARKS each 3000 steps at 3 femtosecond with 1000 powell mini between        *
!*****
evaluate ($seed=(264437) + 976876)
set seed $seed end
!*****
parameter
  @ (Read Parameter file here)
end

structure
  @ (Read Topology/PSF file here)
end

coordinates @ (Read Initial Structure Coordinates)
!*****
evaluate ($knoe = 50.0)
evaluate ($asym = 1.0)

noe
  reset
  nres = 4000
  class all
  @ (Read NOE restraints here)
  @ (Read H-Bonding restraints here)
end

noe
  ceiling 1000
  averaging * sum
  potential * square
  scale * $knoe
  sqconstant * 1.
  sqexponent * 2
end
!*****
restraints dihedral
  nass = 500
  @ (Read Sugar Pucker & Dihedral restraints here)
end
coor copy end
!*****
@ (Read Planarity restraints here)
!*****

```

```

set echo on message on end
evaluate ($rcon = 0.003)
!*****
set message on echo on end

vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)

constraints fix ((resid 500 or resid 501) and name OO) end

vector do (mass = 100.0) (not (resid 500 or resid 501))
vector do (fbeta = 10.0) (all)
!*****
{* Generate Structures 1 ->1 *}
evaluate ($count =0)
evaluate ($end_count =1)
!*****
  while ($count < $end_count) loop main
  evaluate ($count = $count + 1)
  set seed $count end
  coor swap end
  coor copy end
  evaluate ($icount =0)
  while ($icount<3) loop imain  !3 rounds like Pardi suggests
  evaluate ($icount = $icount + 1)
!*****
flags
  exclude *
  include bonds vdw noe angl plan cdih impr
end
mini powell nstep= 100 nprint= 50 end
!*****
vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)

noe scale * $knoe end
!*****
evaluate ($nstep=3000)
evaluate ($radius=0.8)
evaluate ($k_ang =1)
evaluate ($k_imp =1)
evaluate ($k_vdw =4)
evaluate ($knoe =50)
evaluate ($kcdih =300.00)

flags
  exclude *
  include bonds vdw noe angl plan cdih impr
end

for $bath in (1000 600 400) loop anneal

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

constraints interaction (all) (all) weights
  * 1 angles $k_ang improper $k_imp

```

```

end end

parameter
  nbonds
  cutnb=4.5 rcon=$k_vdw nbxmod=3 repel=$radius
end
end

noe scale * $knoe end

restraints dihedral
  scale $kcdih
end

dynamics verlet
  nstep=$nstep timestep=0.003 iasvel=current firsttemp= $bath
  tcoupling = true tbath = $bath nprint=100 iprfreq=0
  ntrfr = 99999999
end
mini powell nstep= 700 nprint= 100 end
end loop anneal
|*****
mini powell nstep= 1000 nprint= 100 end
|*****
end loop imain

set message on echo on end

!lines below this print out threshold and violations

print threshold=0.5 noe
evaluate ($rms_noe=$result)
evaluate ($viol_noe=$violations)
print threshold=5.0 cdih
evaluate ($rms_cdih=$result)
evaluate ($viol_cdih=$violations)
print threshold=0.05 bonds
evaluate ($rms_bonds=$result)
evaluate ($viol_bonds=$violations)
print threshold=5.0 angles
evaluate ($rms_angles=$result)
evaluate ($viol_angles=$violations)
print threshold=5.0 impropers
evaluate ($rms_impropers=$result)
evaluate ($viol_imp=$violations)

set message on echo on end

remarks *****
remarks overall,bonds,angles,improper,vdw,cdih,noe
remarks energies: $ener, $bond, $angl, $impr, $vdw, $cdih, $noe
remarks *****
remarks bonds,angles,impropers,cdih,noe
remarks bonds etc: $rms_bonds,$rms_angles,$rms_impropers,$rms_cdih,$rms_noe
remarks *****
remarks noe, cdih, bonds, angles, impropers
remarks violations:$viol_noe,$viol_cdih,$viol_bonds,$viol_angles,$viol_imp
remarks *****

set message on echo on end

```

```

evaluate ($filename="./(Directory_Name)/(File_Name_)"+" .pdb")
write coordinates output =$filename end
end loop main

stop

!*****
!*REMARKS Refine protocol Verlet dynamics at 500K, 300K, 150K, 50K and 10K  *
!*REMARKS 20,000 steps of 1 femtosecond dynamics at each temperature      *
!*REMARKS and 1000 Powell minimization steps between the dynamics        *
!*****
evaluate ($seed=(264437) + 976876)
set seed $seed end
!*****

parameter
  @ (Read Parameter file here)
end

structure
  @ (Read Topology/PSF file here)
end

coordinates @ (Read Annealed Structure Coordinates)
!*****
evaluate ($knoe = 50.0)
evaluate ($asym = 1.0)

noe
  reset
  nres = 4000
  class all
  @ (Read NOE restraints here)
  @ (Read H-Bonding restraints here)
end

noe
  ceiling 1000
  averaging * sum
  potential * square
  scale * $knoe
  sqconstant * 1.
  sqexponent * 2
end
!*****
restraints dihedral
  nass = 500
  @ (Read Sugar Pucker & Dihedral restraints here)
end
coor copy end
!*****
@ (Read Planarity restraints here)
!*****
set echo on message on end
vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)

constraints fix ((resid 500 or resid 501) and name OO) end

vector do (mass = 100.0) (not (resid 500 or resid 501))

```

```

vector do (mass = 30.0) (resid 500 or resid 501)
vector do (fbeta = 10.0) (all)
!*****
{* Generate Structures 1 ->1 *}
evaluate ($count =0)
evaluate ($end_count =1)
!*****
  while ($count < $end_count) loop main
  evaluate ($count = $count + 1)
  set seed $count end
  coor swap end
  coor copy end

vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)
!*****
evaluate ($nstep =20000)
evaluate ($radius=0.8)
evaluate ($k_ang =1)
evaluate ($k_imp =1)
evaluate ($k_vdw =2)
evaluate ($knoe =50)
evaluate ($kcdih =300.00)

flags
  exclude *
  include bonds vdw noe angl impr cdih plan
end

for $bath in (500 300 150 50 10) loop refine

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

constraints interaction (all) (all) weights
  * 1 angles $k_ang improper $k_imp
end end

parameter
  nbonds
  cutnb = 4.5 ctonnb = 2.99 ctofnb = 3
  rcon=2.0 nbxmod=3 repel=$radius
end
end

noe scale * $knoe end

restraints dihedral
  scale $kcdih
end

dynamics verlet
  nstep=$nstep timestep=0.001 iasvel=current firsttemp= $bath
  tcoupling = true tbath = $bath nprint=5000 iprfrq=0
  ntrfr = 99999999
end
mini powell nstep= 1000 nprint= 20 end
end loop refine

```



```

|*****
mini powell nstep= 2000 nprint= 20 end
|*****

set message on echo on end

!lines below this print out threshold and violations

  print threshold=0.5 noe
  evaluate ($rms_noe=$result)
  evaluate ($viol_noe=$violations)
  print threshold=5.0 cdih
  evaluate ($rms_cdih=$result)
  evaluate ($viol_cdih=$violations)
  print threshold=0.05 bonds
  evaluate ($rms_bonds=$result)
  evaluate ($viol_bonds=$violations)
  print threshold=5.0 angles
  evaluate ($rms_angles=$result)
  evaluate ($viol_angles=$violations)
  print threshold=5.0 impropers
  evaluate ($rms_impropers=$result)
  evaluate ($viol_imp=$violations)

set message on echo on end

remarks *****
remarks  overall,bonds,angles,improper,vdw,cdih,noe
remarks energies: $ener, $bond, $angl, $impr, $vdw, $cdih, $noe
remarks *****
remarks  bonds,angles,improppers,cdih,noe
remarks bonds etc: $rms_bonds,$rms_angles,$rms_improppers,$rms_cdih,$rms_noe
remarks *****
remarks  noe, cdih, bonds, angles, improppers
remarks violations:$viol_noe,$viol_cdih,$viol_bonds,$viol_angles,$viol_imp
remarks *****

set message on echo on end
  evaluate ($filename="./(Directory_Name)/(File_Name_)+".pdb")
  write coordinates output =$filename end
end loop main

stop

|*****
!*REMARKS RDC refine protocol. Introduces RDC using sani algorithm. Attractive *
!*REMARKS LJ switched on, scale Sani force constant at 300K very slowly in      *
!*REMARKS 75 cycles, 1000 steps 1 fs dynamics followed by refinement at      *
!*REMARKS 300K, 150K, 50K and 10K, 50,000 steps, 1 fs dynamics and 1000 steps*
!*REMARKS Powell Minimization between the steps                               *
|*****
evaluate ($seed=(264437) + 976876)
set seed $seed end
|*****
parameter
  @ (Read Parameter file here)
end

structure

```

```

@ (Read Topology/PSF file here)
end

coordinates @ (Read Refined Structure Coordinates)
!*****

noe
  reset
  nres = 4000
  class all
  @ (Read NOE restraints here)
  @ (Read H-Bonding restraints here)
end

noe
  ceiling 1000
  averaging * sum
  potential * square
  scale * $knoe
  sqconstant * 1.
  sqexponent * 2
end
!*****

restraints dihedral
  nass = 500
  @ (Read Sugar Pucker & Dihedral restraints here)
end
!*****

@ (Read Planarity restraints here)
!*****read in dipolar coupling restraints and constants*****
evaluate ($ksani =0.001)
  sani
  nres=500
  class JCH
  force $ktenso
  potential square
  coeff 0.0 3.5 0.55 {corresponds to 0.0 Da R}
  @ (Read field induced Residual Dipolar Coupling restraints)
end
!*****

set message on echo on end

vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)

constraints fix ((resid 500 or resid 501) and name OO) end

vector do (mass = 100.0) (not (resid 500 or resid 501))
vector do (mass = 30.0) (resid 500 or resid 501)
vector do (fbeta = 10.0) (all)
!*****
  { * Generate Structures 1 ->1 * }
  evaluate ($count =0)
  evaluate ($end_count =1)
!*****
  while ($count < $end_count) loop main
  evaluate ($count = $count + 1)
  set seed $count end
  coor swap end
  coor copy end

```

```

vector do (refx=x) (all)
vector do (refy=y) (all)
vector do (refz=z) (all)
!*****Introduce firDC with slow increase in force constant*****
evaluate ($ncycle = 75)
evaluate ($nstep =1000)
evaluate ($bath = 300)
evaluate ($radius=0.8)
evaluate ($k_ang =1)
evaluate ($k_imp =1)
evaluate ($k_vdw =2.0)
evaluate ($knoe =50)
evaluate ($kcdih =300.00)
evaluate ($ini_sani = 0.001)
evaluate ($fin_sani = 1.2)
evaluate ($sani_fac = ($fin_tenso/$ini_tenso)^(1/$ncycle))
evaluate ($ksani = $ini_sani)

flags
  exclude *
  include bonds vdw noe angl impr sani cdih plan

end

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop sani
evaluate ($i_cool=$i_cool+1)
evaluate ($ktenso=$ktenso*$tens_fac)

  constraints interaction (all) (all) weights
    * 1 angles $k_ang improper $k_imp
  end end

!*****introducing attractive Lennard Jones*****
parameter
  nbonds
  repel=0
  atom wmin=0.01 tolerance=0.5
  nbxmod=3 cutnb=11.5 ctonnb=9.5 ctofnb=10.5 rcon=2.0
  rdie vswitch switch
end
end
!*****
noe scale * $knoe end

restraints dihedral
  scale $kcdih
end

sani
force $ksani
end

dynamics verlet
  nstep=$nstep timestep=0.0001 iasvel=current firsttemp= $bath

```

```

    tcoupling = true tbath = $bath nprint=10000 iprfrq=0
    ntrfr = 99999999
end
mini powell nstep= 1000 nprint= 20 end
end loop sani
|*****
evaluate ($icount =0)
while ($icount<3) loop imain {3 rounds}
evaluate ($icount = $icount + 1)
|*****
evaluate ($nstep =50000)
evaluate ($bath = 300)
evaluate ($radius=0.8)
evaluate ($k_ang =1)
evaluate ($k_imp =1)
evaluate ($k_vdw =2.0)
evaluate ($knoe =50)
evaluate ($kcdih =300.00)
evaluate ($ksani = 1.2)

flags
  exclude *
  include bonds vdw noe angl impr sani cdih plan
end

for $bath in (300 150 50 10) loop refine

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)

constraints interaction (all) (all) weights
  * 1 angles $k_ang improper $k_imp
end end
|*****introducing attractive Lennard Jones*****
parameter
  nbonds
  repel=0
  atom wmin=0.01 tolerance=0.5
  nbxmod=3 cutnb=11.5 ctonnb=9.5 ctofnb=10.5 rcon=2.0
  rdie vswitch switch
end
end
|*****
noe scale * $knoe end

restraints dihedral
  scale $kcdih
end

sani
force $ksani
end

dynamics verlet
  nstep=$nstep timestep=0.0001 iasvel=current firsttemp= $bath
  tcoupling = true tbath = $bath nprint=10000 iprfrq=0
  ntrfr = 99999999
end

```

```

mini powell nstep= 1000 nprint= 20 end
end loop refine
!*****
mini powell nstep= 2000 nprint= 20 end
end loop imain
!*****

set message on echo on end

!lines below this print out threshold and violations

print threshold=0.5 noe
evaluate ($rms_noe=$result)
evaluate ($viol_noe=$violations)
print threshold=5.0 cdih
evaluate ($rms_cdih=$result)
evaluate ($viol_cdih=$violations)
print threshold=0.05 bonds
evaluate ($rms_bonds=$result)
evaluate ($viol_bonds=$violations)
print threshold=5.0 angles
evaluate ($rms_angles=$result)
evaluate ($viol_angles=$violations)
print threshold=5.0 impropers
evaluate ($rms_improppers=$result)
evaluate ($viol_imp=$violations)
sani print threshold=0.5 class JCH end
evaluate ($rms_san_JCH=$result)
evaluate ($viol_san_JCH=$violations)
set message on echo on end
remarks *****
remarks overall,bonds,angles,improper,vdw,cdih,noe,sani
remarks energies: $ener, $bond, $angl, $impr, $vdw, $cdih, $noe, $sani
remarks *****
remarks bonds,angles,improppers,cdih,noe,sani
remarks bonds etc: $rms_bonds,$rms_angles,$rms_improppers,$rms_cdih,
remarks $rms_noe,$rms_san_JCH
remarks *****
remarks noe, cdih, bonds, angles, improppers, sani
remarks violations:$viol_noe,$viol_cdih,$viol_bonds,$viol_angles,$viol_imp,
remarks $viol_san_JCH
remarks *****
set message on echo on end
evaluate ($filename="./(Directory_Name)/(File_Name_)+".pdb")
write coordinates output =$filename end
end loop main
stop

```

Useful scripts for structure calculations, modeling and analysis.

```
*****
*Perl script to generate multiple XPLOR input files each with new structures*
*****
#!/usr/bin/perl
if ( $#ARGV != 2 ) {
    print "Usage: mp_xplor template_script output_script nam_file\n";
    exit;
}
($script_def, $script_out, $nam_name) = @ARGV;
open(I, $nam_name);
@nam=<I>;
close I;
foreach $i (0 .. $#nam)
{
    chomp $nam[$i];
    $nam[$i]=~/ini_(.*)\.pdb/;
    $num=$1;
open( I, $script_def);
@script = <I>;
close I;
foreach $j (1.. $#script )
{
    if ( $script[$j] =~ /ini_/ )
    {
        $script[$j] =~ s/(.*)ini_(.*)/$1ini_$num$2/;
    }
    if ( $script[$j] =~ /ann_/ )
    {
        $script[$j] =~ s/(.*)ann_(.*)/$1ann_$num$2/;
    }
}
    $j = join( "", @script);
    open( I, ">${script_out}_${i}.inp");
    print I "$j";
    close I;
}
exit 0;
```

```

*****
*Perl script to make dihedral angle restraints from a given PDB file*
* for nucleic acids and proteins *
*****
#!/bin/sh -- # perl
eval `exec perl -w -S $0 ${1+"$@"}`
if 0;

open (INFILE, "HH_dihedral.txt");
open (OUTFILE, ">cns_dihedrals.txt");
while ($line = <INFILE>)
{
    chop ($line);
    @line_split = split (' ', $line);

# ensure that only intradomain restraints are generated
    if (
        ((1<=$line_split[0])&&($line_split[0]<=4)) ||
        ((131<=$line_split[0])&&($line_split[0]<=134)) ||
        ((7<=$line_split[0])&&($line_split[0]<=13)) ||
        ((101<=$line_split[0])&&($line_split[0]<=107))
    )
# Reference Nomenclatures from Crothers/Tinoco and Bloomfeld
# Name = Definition
# alpha = O3'(n-1)-P-O5'-C5'    nu0 = C4'-O4'-C1'-C2'
# beta = P-O5'-C5'-C4'        nu1 = O4'-C1'-C2'-C3'
# gamma = O5'-C5'-C4'-C3'     nu2 = C1'-C2'-C3'-C4'
# epsilon = C4'-C3'-O3'-P(n+1) nu3 = C2'-C3'-C4'-O4'
# zeta = C3'-O3'-P(n+1)-O5'(n+1) nu4 = C3'-C4'-O4'-C1'
# chi(pur) = O4'-C1'-N9-C4
# chi(pyr) = O4'-C1'-N1-C2
    {

        if ( ($line_split[2] eq ALPHA))
##&&& ($secondary{$line_split[0]} eq "true") )
        {
            $alpha = $line_split[3];
            $dalpha = 10;
            $resnum = $line_split[0];
            $iminus = $line_split[0] -1;
            $iplus = $line_split[0] +1;
            print OUTFILE "! Residue $resnum ALPHA\n";
            print OUTFILE "assign (resid $iminus and name O3')(resid $resnum and name P )\n";
            print OUTFILE "    (resid $resnum and name O5')(resid $resnum and name C5)";
            printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $alpha, $dalpha;
        }

        if ( ($line_split[2] eq BETA))
## && ($secondary{$line_split[0]} eq "true") )
        {
            $beta = $line_split[3];
            $dbeta = 10;
            $resnum = $line_split[0];
            $iminus = $line_split[0] -1;
            $iplus = $line_split[0] +1;
            print OUTFILE "! Residue $resnum BETA\n";
            print OUTFILE "assign (resid $resnum and name P )(resid $resnum and name O5')\n";
            print OUTFILE "    (resid $resnum and name C5')(resid $resnum and name C4)";
            printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $beta, $dbeta;
        }

        if ( ($line_split[2] eq GAMMA))
##&&& ($secondary{$line_split[0]} eq "true") )
        {
            $gamma = $line_split[3];
            $dgamma = 10;
            $resnum = $line_split[0];
            $iminus = $line_split[0] -1;
            $iplus = $line_split[0] +1;

```

```

print OUTFILE "! Resdiue $resnum GAMMA\n";
print OUTFILE "assign (resid $resnum and name O5' )(resid $resnum and name C5' )\n";
print OUTFILE "      (resid $resnum and name C4' )(resid $resnum and name C3' )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $gamma, $dgamma;
}
if ( ($line_split[2] eq DELTA))
##&&& ($secondary{$line_split[0]} eq "true")
{
    $delta = $line_split[3];
    $ddelta = 10;
    $resnum = $line_split[0];
    $iminus = $line_split[0] -1;
    $iplus = $line_split[0] +1;
print OUTFILE "! Resdiue $resnum DELTA\n";
print OUTFILE "assign (resid $resnum and name C5' )(resid $resnum and name C4' )\n";
print OUTFILE "      (resid $resnum and name C3' )(resid $resnum and name O3' )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $delta, $ddelta;
}
if ( ($line_split[2] eq EPSI))
##&&& ($secondary{$line_split[0]} eq "true")
{
    $epsi = $line_split[3];
    $depsi = 10;
    $resnum = $line_split[0];
    $iminus = $line_split[0] -1;
    $iplus = $line_split[0] +1;
print OUTFILE "! Resdiue $resnum EPSI\n";
print OUTFILE "assign (resid $resnum and name C4' )(resid $resnum and name C3' )\n";
print OUTFILE "      (resid $resnum and name O3' )(resid $iplus and name P )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $epsi, $depsi;
}
if ( ($line_split[2] eq ZETA))
##&&& ($secondary{$line_split[0]} eq "true")
{
    $zeta = $line_split[3];
    $dzeta = 10;
    $resnum = $line_split[0];
    $iminus = $line_split[0] -1;
    $iplus = $line_split[0] +1;
print OUTFILE "! Resdiue $resnum ZETA\n";
print OUTFILE "assign (resid $resnum and name C3' )(resid $resnum and name O3' )\n";
print OUTFILE "      (resid $iplus and name P )(resid $iplus and name O5' )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $zeta, $dzeta;
}
if ( ($line_split[2] eq NU1))
##&&& ($secondary{$line_split[0]} eq "true")
{
    $nu1 = $line_split[3];
    $dnu1 = 10;
    $resnum = $line_split[0];
    $iminus = $line_split[0] -1;
    $iplus = $line_split[0] +1;
print OUTFILE "! Resdiue $resnum NU1\n";
print OUTFILE "assign (resid $resnum and name O4' )(resid $resnum and name C1' )\n";
print OUTFILE "      (resid $resnum and name C2' )(resid $resnum and name C3' )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $nu1, $dnu1;
}
if ( ($line_split[2] eq NU2))
##&&& ($secondary{$line_split[0]} eq "true")
{
    $nu2 = $line_split[3];
    $dnu2 = 10;
    $resnum = $line_split[0];
    $iminus = $line_split[0] -1;
    $iplus = $line_split[0] +1;
print OUTFILE "! Resdiue $resnum NU2\n";
print OUTFILE "assign (resid $resnum and name C1' )(resid $resnum and name C2' )\n";

```

```

    print OUTFILE "      (resid $resnum and name C3' )(resid $resnum and name C4' )";
    printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $nu2, $dnu2;
  }
# Uncomment the lines below in case you are measuring dihedral angles for Proteins
# Then comment out the lines above for alpha, beta, gamma, delta, nu1 and nu2 which are specific
# for nucleic acids.
#   if ( ($line_split[2] eq PHI)
##&& ($secondary{$line_split[0]} eq "true" )
#     {
#       $phi = $line_split[3];
#       $dphi = 0.1;
#       $resnum = $line_split[0];
#       $iminus = $line_split[0] -1;
#       print OUTFILE "! Residue $resnum PHI\n";
#       print OUTFILE "assign (resid $iminus and name C )(resid $resnum and name N )\n";
#       print OUTFILE "      (resid $resnum and name CA)(resid $resnum and name C )";
#       printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $phi, $dphi;
#     }
#   if ( ($line_split[2] eq PSI)
##&& ($secondary{$line_split[0]} eq "true" )
#     {
#       $psi = $line_split[3];
#       $dpsi = 0.1;
#       $resnum = $line_split[0];
#       $iplus = $line_split[0] +1;
#       print OUTFILE "! Residue $resnum PSI\n";
#       print OUTFILE "assign (resid $resnum and name N )(resid $resnum and name CA )\n";
#       print OUTFILE "      (resid $resnum and name C )(resid $iplus and name N )";
#       printf OUTFILE " 25.0 %5.2f %5.2f 2\n", $psi, $dpsi;
#     }
#   }
}
}
close (INFILE);
close (OUTFILE);
#die;

```

```

*****
*Perl script to generate artificial NOE restraints from PDB files*
* that can be used for modeling in XPLOR/CNS format *
*****
#!/usr/bin/perl

# change the file name here
open(INFILE, "/nmr1/deva/HH_calc/perl_scripts/ref_HH_180.pdb");
open (OUTFILE, ">cns_intradomain_noes.txt");
$index = 0 ;
$scutoff = 100;
# below fix the error values you would like to give the noe restraints +/-
$distallowance = 0.01;

while ($line = <INFILE>)
{
    chop ($line);
    @line_split = split(' ', $line);
    print "$line_split[2] $line_split[3]\n";

# fix or statement to you liking....
    if ( ($line_split[0] eq "ATOM") &&
        (
            ($line_split[2] eq "P")
        )
    )
    {
        $atom[$index] = $line_split[2];
        $type[$index] = $line_split[3];
        $resnum[$index] = $line_split[4];
        $x[$index] = $line_split[5];
        $y[$index] = $line_split[6];
        $z[$index] = $line_split[7];
        $index ++;

    } # end if
} # end while

#die;
$index --;
for $i (0..$index)
{
    for $j ($i..$index)
    {
# only intradomain NOEs in rna
# explanation : For instance if you want to fix the 2 domains (helices) of the Hammerhead RNA
# given by the residue numbers (7 to 13 and 101 to 107 stem 1) and ( 134 to 131 and 1 to 4 stem 3)
# you can choose which atom type you want to use for restraints eg. P, H8, C8, H1' etc.
# and specify the residue numbers between which you want to generate model / simulated noe restraints
        if (
            (
                ((1<=$resnum[$i])&&($resnum[$i]<=4)) && ((131<=$resnum[$j])&&($resnum[$j]<=134))
            )||
            (
                ((7<=$resnum[$i])&&($resnum[$i]<=13)) && ((101<=$resnum[$j])&&($resnum[$j]<=107))
            )||
            (
                ((115<=$resnum[$i])&&($resnum[$i]<=118)) && ((123<=$resnum[$j])&&($resnum[$j]<=126))
            )
        )
        {
            $dist = sqrt( (($x[$i] - $x[$j])*( $x[$i] - $x[$j])) +
                (($y[$i] - $y[$j])*( $y[$i] - $y[$j])) +
                (($z[$i] - $z[$j])*( $z[$i] - $z[$j])) ) ;
            if ( ($dist < $scutoff) && ($i != $j) )
            {
                print OUTFILE "Assign (resid $resnum[$i] and name $atom[$i])\n";
                print OUTFILE "    (resid $resnum[$j] and name $atom[$j]) ";
            }
        }
    }
}

```

```
printf OUTFILE "%3.2f %3.2f %3.2f", $dist,
    $distallowence, $distallowence;
print OUTFILE " ! $type[$i]$resnum[$i] and $type[$j]$resnum[$j] \n\n";
}
}
}

close (INFILE);
close (OUTFILE);
```

```

*****
*Awk script to sort PDB structures generated by XPLOR *
*according to total energies, violation energies and      *
* number of violations                                  *
*****

#!/usr/bin/awk -f

# This script gathers the useful data from the ensemble of structures
# and creates several output files with the N best decoys according to
# three criteria: total-E, noe-E and n-violations.
# The data are read from the headings of the PDB of each final decoy.
# This script must be placed in the same directory as the PDBs

BEGIN{
    system("ls -l sa* > devanlist.txt")

    while(getline < "devanlist.txt" >0){
        num++
        filenames[num]=$1
    }close("devanlist.txt")

    for(i=1;i<=num;i++){
        OK_en=0
        OK_vio=0

        while(getline < filenames[i] >0){
            if($2=="energies:"){
                en[i]=$0
                OK_en=1
            }

            if($2=="violations:"){
                vio[i]=$0
                OK_vio=1
            }

            if(OK_vio*OK_en)
                break
        }close(filenames[i])
    }

    for(i=1;i<=num;i++){
        printf "%20s %s  %s\n",filenames[i], en[i], vio[i] > "long.txt"
    }

    #Heading of by_nam_short.txt
    {
    printf "%s", "pdb_name, total energy, noe energy, noe violations\n" > "short.txt"
    }

    # This while loop is only required to remove the unused columns of by_nam_long.txt
    while(getline < "long.txt" >0){
        {
        printf "%20s %s %s %s %s\n", $1, $4, $10, $14, "end" > "short.txt"
        }
    }

    close("long.txt")
    close("short.txt")

    # Generate useful outputs
    system("sort -k2n short.txt > tot_e_sorted.out")
    system("sort -k3n short.txt > noe_e_sorted.out")
    system("sort -k4n short.txt > viol_sorted.out" )
}

```

```

# Clean up unused files
#system("rm by_nam_short.txt by_nam_long.txt vslist.txt")

# Generate lists with 10 best for Molmol
count=0
j=0
violations=0
while(getline < "tot_e_sorted.out" >0){
    count++
    vector[count]=$1
    violat[count]=substr($4,1,length-1) #counts violations
}close("tot_e_sorted.out")

for(j=2; j<=11; j++){
    print vector[j] > "tot_e_10best.nam"

    violations=violations + violat[j]
}

    addme="average violations in the 10 best-total-E decoys = "
    print addme violations/10 > "aver_viol.out"

count=0
j=0
while(getline < "noe_e_sorted.out" >0){
    count++
    vector[count]=$1
}close("noe_e_sorted.out")

for(j=2; j<=11; j++){
    print vector[j] > "noe_e_10best.nam"
}
count=0
j=0
while(getline < "viol_sorted.out" >0){
    count++
    vector[count]=$1
}close("viol_sorted.out")

for(j=2; j<=11; j++){
    print vector[j] > "viol_10best.nam"
}

# List useful files on shell
system("ls -d *out")
system("ls -d *nam")
}
# The file by_nam_long.txt contains all names, energies and violations sorted by filename.
# The file by_nam_short.txt contains only the numbers that I need sorted by filename.

# The last line system("sort -k4n namesorted.txt") uses the UNIX "sort" command:
#     -k stands for column
#     4 is the number of the column used for sorting
#     n means that that column contains a number
#     For sorting strings, remove the 'n', as follows: system("sort -k4 namesorted.txt")

```

```

*****
*Awk script that back calculates RDC and alignment tensor from PDB*
*****
#!/usr/bin/awk -f

BEGIN{
  ### -----
  ### User options
  ### -----
  ### MOLECULE
  molec = "TAR"
  # molec = "1naj"
  # molec = "2gat"
  # molec = "lave"          #test DNA
  print "Molecule =", molec, ARGV[1]

  ### EQUATION
  # To build X tensor from normal vector.
  # eqn = "vsan"          #vsan view of Angew2004 - cartesian coord.
  # eqn = "vsph"          #vsan view of spherical coord -correction of lave's eqn.
  # eqn = "cigr"          #cigr, like in Angew2004; non-traceless X.
  # eqn = "lave"          #as in the original program - two signs wrong.
  # print "Equation =", eqn

  ### PREFACTORS
  # pre = "vsan"          #implementation of Glaser - cartesian lab frame.
  # pre = "cigr"          #Angew2004 - exclude 1/2 as I extrapolate to 0 field
  # pre = "bax"          #as in JACS 2004 - prefactors change.
  #                       #Note: vsan(Glaser) and cigr differ only in 1/2.
  # print "Prefactors =", pre

  ### Xi VALUES
  # xi_val = "Bax"        # Bax, JACS2004.
  # xi_val = "Angew"     # Angew 2004, 187, Table 1.
  # xi_val = "lave"      # Angew 2004, 187, Table 2. Wrong values.
  # xi_val = "zhang"     # JACS2004p10530=bastiaan

  scale = -1 # either -1.0 or -2/3 to scale Xzz and delta-Xi
  # +1 for Laurent equations
  if(eqn == "lave") scale = -1*scale #as lave's eqn invert signs

  ### TAR: include / exclude loop or bulge. Note that r24 is missing.
  # jump = "yes"         # yes / no -- for missing residues
  # bulge = "exclude"   # include / exclude
  # loop = "exclude"    # include / exclude
  # if(molec != "TAR"){jump = "no" ; bulge = "include" ; loop = "include"}

  ### -----
  ##### End of user options
  ### -----
  ### Assign constants
  ### -----
  #initial count of TAR
  if(molec=="TAR" ) iniN=16
  if(molec=="1naj") iniN=1 # iniN=1, jump=no, bulge/loop=include
  if(molec=="2gat") iniN=101 # iniN=101, jump=no, bulge/loop=include
  if(molec=="lave") iniN=1

  #column counter for input PDB - dependent on number of columns
  k=0 #for TAR
  if(molec=="1naj" || molec=="2gat") k=1

  #CIGR Xi values.
  if(xi_val == "Angew"){
    print "Angew2004, Table 1. Xi values as Xzz"
    xade = 13 *scale
    xgua = 13 *scale
    xcyt = 13 *scale
  }
}

```

```

        xthy = 13 *scale
        xuri = xthy
        xlig = 2.70/2.04 *xade      #lave's DNA ligand
    }
#Xi? values in Angew 2004,187,Table 2, HM method.
if(xi_val == "lave"){
    print "Lave's: Angew2004, Table 2 - Xi values? Ring currents!"
    xade = 1.5 *scale
    xgua = 1.51 *scale
    xcyt = 0.9 *scale
    xthy = 0.9 *scale
    xuri = xthy
    xlig = 2.70 *scale #lave's DNA ligand
}
#Bax X values. JACS2004p10820, table 1. Units fixed.
if(xi_val == "Bax"){
    print "Bax Xi values as Xa=Xzz*3/2. JACS2004p10820, table 1."
    pi=3.14159265
    #scale = scale *2/3 #transform Xa into Xzz
    scale = scale *(4*pi)*3/2    #Bax uses different units and eqn.
    xade = 1.304 *scale
    xgua = 0.922 *scale
    xcyt = 0.392 *scale
    xthy = 0.411 *scale
    xuri = 0.386 *scale
}
#Zhang-Hashimi X values. JACS2004p10530, citing Bastiaan chapter.
if(xi_val == "zhang"){
#    print "Zhang + Hashimi Xi values. JACS2004p10530."
    xade = 19.1 *scale
    xgua = 18.3 *scale
    xcyt = 12.0 *scale
    xthy = xcyt *scale
    xuri = 12.0 *scale
}

#names of residues can be ADE, THY ... or A, T, G, C, U
nade = "A" ; ngua = "G" ; ncyt = "C" ; nthy = "T" ; nuri = "U"
if(molec=="TAR" || molec=="lave"){
    nade = "ADE" ; ngua = "GUA"
    ncyt = "CYT" ; nthy = "THY" ; nuri = "URI"
}

### -----
#### Read input data
### -----
while ( getline < ARGV[1] > 0 ){
    if(molec == "TAR" ) if($1 != "TAR1" ) continue      #skip non-TAR1 lines
    if(molec == "lnaj") if($1 == "ENDMDL") break #exit WHILE loop
    if(molec == "2gat") if($1 == "CONNECT") break #exit WHILE loop

# READ BASE PLANE -- Read 4 base atoms to define base plane with two vectors
if($3 == "N1" && $4 != "LIG") for(j=1;j<=3;j++) pN1[$(5+k),j]=$(5+j+k)
if($3 == "C1" && $4 == "LIG") for(j=1;j<=3;j++) pN1[$(5+k),j]=$(5+j+k)
if($3 == "C2" )                for(j=1;j<=3;j++) pC2[$(5+k),j]=$(5+j+k)
if($3 == "C4" )                for(j=1;j<=3;j++) pC4[$(5+k),j]=$(5+j+k)
if($3 == "C6" )                for(j=1;j<=3;j++) pC6[$(5+k),j]=$(5+j+k)
if($3 == "C2" ) endN=$(5+k) # stores the last residue number

# READ BONDS -- Read vectors for RDC of interest
if($3 == "C8" && $4 != "LIG") for(j=1;j<=3;j++) pC8[$(5+k),j]= $(5+j+k)
if($3 == "H8" && $4 != "LIG") for(j=1;j<=3;j++) pH8[$(5+k),j]= $(5+j+k)
if($3 == "H1" && $4 == ngua ) for(j=1;j<=3;j++) pH1[$(5+k),j]= $(5+j+k)
if($3 == "N1" && $4 == ngua ) for(j=1;j<=3;j++) pNa1[$(5+k),j]= $(5+j+k)
if($3 == "H2" && $4 == nade ) for(j=1;j<=3;j++) pH2[$(5+k),j]= $(5+j+k)
if($3 == "C2" && $4 == nade ) for(j=1;j<=3;j++) pCa2[$(5+k),j]= $(5+j+k)
if($4 == nthy || $4 == ncyt || $4 == nuri ){

```

```

if ( $3 == "C6" ) for(j=1;j<=3;j++) pCa6[$(5+k),j] = $(5+j+k)
if ( $3 == "H6" ) for(j=1;j<=3;j++) pH6[$(5+k),j] = $(5+j+k)
}

# Assign a susceptibility value to each base ring.
if ( $4 == ngua ) { xi[$(5+k)]=xgua ; res[$(5+k)]= $4 }
if ( $4 == nade ) { xi[$(5+k)]=xade ; res[$(5+k)]= $4 }
if ( $4 == ncyt ) { xi[$(5+k)]=xcyt ; res[$(5+k)]= $4 }
if ( $4 == nthy ) { xi[$(5+k)]=xthy ; res[$(5+k)]= $4 }
if ( $4 == nuri ) { xi[$(5+k)]=xuri ; res[$(5+k)]= $4 }
if ( $4 == "LIG" ) { xi[$(5+k)]=xlig ; res[$(5+k)]= $4 } #only lave's DNA
}
close(ARGV[1])

### -----
### Calculate individual tensors and resultant tensor.
### -----
for ( i=iniN ; i <= endN ; i++ ) {
if(jump == "yes") if(i==24 ) continue # missing nt in TAR
if(bulge == "exclude") if(i==23 || i==25) continue # exclude bulge
if( loop == "exclude") if(i>=30 && i<=35) continue # exclude loop
if(jump == "yes") if(i==55 ) continue # missing nt in TAR
if(bulge == "exclude") if(i==54 || i==56) continue # exclude bulge
if( loop == "exclude") if(i>=61 && i<=66) continue # exclude loop

#Build two vectors on the plane of each base.
for(j=1;j<=3;j++){
vb1[j] = pN1[i,j] - pC4[i,j]
vb2[j] = pC2[i,j] - pC6[i,j]
}
# Calculate their vector product to get a normal vector.
vn[1] = vb1[2]*vb2[3] - vb1[3]*vb2[2]
vn[2] = vb1[3]*vb2[1] - vb1[1]*vb2[3]
vn[3] = vb1[1]*vb2[2] - vb1[2]*vb2[1]
vnorm = sqrt(vn[1]^2 + vn[2]^2 + vn[3]^2)
for(j=1;j<=3;j++) vn[j] = vn[j]/vnorm

# Laurent's way: spherical coordinates. Original program.
if(eqN == "lave"){
nx[i]=vn[1]; ny[i]=vn[2]; nz[i]=vn[3]
s[i]=sqrt(1-nz[i]^2)
cp[i]=cos(atan2(ny[i],nx[i]))
sp[i]=sqrt(1-cp[i]^2)
mat[i,1,1] = xi[i]*(-(3*s[i]^2-1)*cp[i]^2 + sp[i]^2)
mat[i,1,2] = xi[i]*cp[i]*sp[i]*(3*s[i]^2)
mat[i,1,3] = xi[i]*(-3*nz[i]*s[i]*cp[i])
mat[i,2,3] = xi[i]*(3*nz[i]*s[i]*sp[i])
mat[i,2,2] = xi[i]*(-(3*s[i]^2-1)*sp[i]^2 + cp[i]^2)
mat[i,3,3] =-xi[i]*(3*nz[i]^2-1)
mat[i,2,1] = mat[i,1,2]
mat[i,3,1] = mat[i,1,3]
mat[i,3,2] = mat[i,2,3]
}
# vsan version of spherical coordinates.
if(eqN == "vsph"){
czen = vn[3]
szen = sqrt( 1 - vn[3]^2 ) #zenital theta is in [0,pi]
cphi = vn[1] /sqrt(vn[1]^2 + vn[2]^2) #phi is in [0,2*pi]
sphi = vn[2] /sqrt(vn[1]^2 + vn[2]^2)
mat[i,1,1] = xi[i]/3* ( 3 *szen^2 - 1 ) *cphi^2 - sphi^2 )
mat[i,1,2] = xi[i]/3* ( cphi *sphi *3 *szen^2 )
mat[i,1,3] = xi[i]/3* ( 3 *czen *szen *cphi )
mat[i,2,3] = xi[i]/3* ( 3 *czen *szen *sphi )
mat[i,2,2] = xi[i]/3* ( 3 *szen^2 - 1 ) *sphi^2 - cphi^2 )
mat[i,3,3] = xi[i]/3* ( 3 *czen^2 - 1 )
mat[i,2,1] = mat[i,1,2]
mat[i,3,1] = mat[i,1,3]
}
}

```



```

        mat[i,3,2] = mat[i,2,3]
    }
    # vsan view of Angew2004. X built traceless by vsan.
    if(eqn == "vsan"){
        for(p=1;p<=3;p++){
            for(q=1;q<=3;q++){
                dk = 0.0 ; if(p==q) dk=1.0      # delta de kronecker
                mat[i,p,q]= xi[i]* ( vn[p]*vn[q] -dk/3 )
            }
        }
    }
    #Exactly Angew2004 way. Note that X tensor is NOT traceless here!
    if(eqn == "cigr"){
        for(p=1;p<=3;p++) for(q=1;q<=3;q++) mat[i,p,q]= xi[i]*vn[p]*vn[q]
    }

    # Resultant tensor is the tensor sum of each residue tensor
    for(p=1; p <= 3; p++) for(q=1; q <= 3; q++) Smat[p,q] += mat[i,p,q]
} #end for i

if(eqn=="lave") for(p=1;p<=3;p++) for(q=1;q<=3;q++) Smat[p,q]=Smat[p,q]/3

### -----
# Conversion factors from susceptibility units to Hz.
# See, for instance, Bax, Nat.Str.Biol. 1997, 732.
# Constants in SI units.
# Decomposed in mantissa and exponential.
### -----

m_distn1h1=1.040
m_distc6h6=1.090
m_distc8h8=1.080
m_distc2h2=1.090
e_dist=-10          # factor metre/Angstrom

factor= 10^-34      # Exponential of susceptibility values
e_factor=-34        # in Laurent's units = SI units = m3

B1=800              # MHz
                    #B2=400          # I'd rather use zero field
B2=0                # As we extrapolated to zero field by linear fitting (TAR)
B3=(B1*B1)-(B2*B2)
                    #B=B3*10^12
e_B=12
pi=3.14159265
T=298
                    #R=8.134#J/(K*mol)
                    #N=6.023*10^23; m_N=6.023; e_N=23
h=6.634*10^-34     #J*s
m_h=6.634
e_h=-34
kboltz=1.3807*10^-23      #Boltzmann = R/N in SI units = J/K
m_boltz=1.3807
e_boltz=-23
diff1=0.25           #g-C / g-H, magnetogyric ratio
diff2=0.10          #g-N / g-H,

exp_cnst = e_B + e_h + e_factor - e_boltz -3*e_dist
#print "exp_cnst=", exp_cnst
cnst = -(B3 * m_h)/(15 *m_boltz *T) *(3/(4*pi)) *10^exp_cnst

if(pre == "bax"){
    e_factor=-27      # in Bax's units = J/T2
    e_mu=-7 # mu0/4*pi = 10^-7
    exp_cnst = e_B + e_h + e_factor - e_boltz -3*e_dist + e_mu
    cnst = -(B3 * m_h)/(15 *m_boltz *T) *3 *10^exp_cnst *(4*pi)
}

```

```

cnst1 = cnst * diff1 / m_distc6h6^3
cnst2 = cnst * diff1 / m_distc8h8^3
cnst3 = cnst * diff2 / m_distn1h1^3
cnst4 = cnst1 #cnst * diff1 / m_distc2h2^3
#conve = sprintf("%3.5f %3.5f %3.5f %3.5f", cnst1, cnst2, cnst3, cnst4)
### -----
### Compute some RDC to be compared with experimental ones
### -----
#printf " RDC in Hz -- nt tensors in m^3 (*1e34).\n"
#printf " resid Dc2h2 Dc6h6 Dc8h8 Dn1h1 (1,1) (1,2) (1,3) (2,2) (2,3) (3,3)\n"

for ( i=iniN ; i <= endN ; i++ ){
    if(jump == "yes") if(i==24) continue # missing nt in TAR
    #RDC can be calculated, even if they were not used to build X-tensor
    if(""){
        if(bulge == "exclude") if(i==23 || i==25) continue # exclude bulge
        if(loop == "exclude") if(i>=30 && i<=35) continue # exclude loop
    }
# CALCULATE BOND VECTORS
for(j=1;j<=3;j++){
    vc2[i,j]= pH2[i,j]-pCa2[i,j] ; rc2 += vc2[i,j]^2
    vc6[i,j]= pH6[i,j]-pCa6[i,j] ; rc6 += vc6[i,j]^2
    vc8[i,j]= pH8[i,j]- pC8[i,j] ; rc8 += vc8[i,j]^2
    vn1[i,j]= pH1[i,j]-pNa1[i,j] ; m1 += vn1[i,j]^2
}
rc2 = sqrt(rc2) ; rc6 = sqrt(rc6) ; rc8 = sqrt(rc8) ; m1 = sqrt(m1)

### Calculate each RDC by multiplying v*Tensor*v, as Angew2004
# tensor times column vector
for(p=1; p <= 3; p++){
    ac2[p] = 0 ; ac6[p] = 0 ; ac8[p] = 0 ; an1[p] = 0
    for(r=1; r <= 3; r++){
        ac2[p] += Smat[p,r]*vc2[i,r]
        ac6[p] += Smat[p,r]*vc6[i,r]
        ac8[p] += Smat[p,r]*vc8[i,r]
        an1[p] += Smat[p,r]*vn1[i,r]
    }
}
# vector scalar product
Dc2h2[i] = 0 ; Dc6h6[i] = 0 ; Dc8h8[i] = 0 ; Dn1h1[i] = 0
for(r=1; r <= 3; r++){
    Dc2h2[i] += ac2[r]*vc2[i,r]
    Dc6h6[i] += ac6[r]*vc6[i,r]
    Dc8h8[i] += ac8[r]*vc8[i,r]
    Dn1h1[i] += an1[r]*vn1[i,r]
}
#Rescale D because the eqn requires unitary vectors (Angew2004,187).
if(rc2>0) Dc2h2[i]= cnst4* Dc2h2[i]/rc2^2
if(rc6>0) Dc6h6[i]= cnst1* Dc6h6[i]/rc6^2
if(rc8>0) Dc8h8[i]= cnst2* Dc8h8[i]/rc8^2
if(m1>0) Dn1h1[i]= cnst3* Dn1h1[i]/m1^2
if(eqn == "cigr"){
    tr_x = 0.0
    for(j=1;j<=3;j++){ tr_x += Smat[j,j] # trace of tensor
    print i,"tr_x=",tr_x
    if(rc2>0) Dc2h2[i]= Dc2h2[i] - cnst4 * tr_x / 3
    if(rc6>0) Dc6h6[i]= Dc6h6[i] - cnst1 * tr_x / 3
    if(rc8>0) Dc8h8[i]= Dc8h8[i] - cnst2 * tr_x / 3
    if(m1>0) Dn1h1[i]= Dn1h1[i] - cnst3 * tr_x / 3
}
}
### End of Calculate each RDC by multiplying v*Tensor*v
# prepare output of individual nt tensors
linMat=""
for(p=1;p<=3;p++) for(q=p;q<=3;q++) linMat = sprintf("%s %6.3f",linMat,mat[i,p,q])
#printf "%2s %3s %6.3f %6.3f %6.3f %6.3f %6.3f %s\n",i,res[i],Dc2h2[i],Dc6h6[i],Dc8h8[i],Dn1h1[i],linMat
} #end for i

```

```

### DIAGONALIZE TOTAL TENSORS
### -----
# Solution of Symmetric nxn Eigenvalue problem using Jacobi Method
# Store Smat in new matrix
for(p=1; p <= 3; p++)      for(q=1; q <= 3; q++)      aa[p-1,q-1] = Smat[p,q]
n=3      #dimension of tensor (=3)
eps=10^(-20)
ntot=100      #ntot is # of sweeps
for(count=0; count < ntot; count++){
for(p=0; p < n-1; p++){      #sweep through aa matrix
for(q=p+1; q < n; q++){      #zero elements via Jacobi transformation
if( sqrt(aa[p,q]^2) >= eps ){      #if(aa[p,q] != 0.0){
a=aa[p,p] #from here, its same as 2x2 case as in lecture
b=aa[p,q]
d=aa[q,q]
theta=(d-a)/(2.0*b)
t1=-theta+sqrt(theta*theta+1.0)
t2=-theta-sqrt(theta*theta+1.0)
# The smallest root is more stable ...
t=t2
if( sqrt(t2*t2) > sqrt(t1*t1) ) t=t1
if(t == 0.0) break      #vsan
c=1.0/sqrt(1.0+t*t)      # cos(theta)
st=c*t      # sin(theta)
h=t*b
aa[p,p]= aa[p,p] -h
aa[q,q]= aa[q,q] +h
aa[p,q]= 0.0
aa[q,p]= 0.0
tau= st/(1.0+c)
for(r=0; r < p; r++){
g1=aa[p,r] ; g2=aa[q,r]
aa[p,r]=g1-st*(g2+tau*g1) ; aa[q,r]=g2+st*(g1-tau*g2)
aa[r,p]=aa[p,r] ; aa[r,q]=aa[q,r]
}
for(r=p+1; r < q; r++){
g1=aa[p,r] ; g2=aa[q,r]
aa[p,r]=g1-st*(g2+tau*g1) ; aa[q,r]=g2+st*(g1-tau*g2)
aa[r,p]=aa[p,r] ; aa[r,q]=aa[q,r]
}
for(r=q+1; r < n; r++){
g1=aa[p,r] ; g2=aa[q,r]
aa[p,r]=g1-st*(g2+tau*g1) ; aa[q,r]=g2+st*(g1-tau*g2)
aa[r,p]=aa[p,r] ; aa[r,q]=aa[q,r]
}
#print intermediate tensors
for(p=0; p < n; p++){
line[p]=" "
}
for(p=0; p < n; p++) print line[p]
}      #close IF
}      #close q
}      #close p

}      #close count

### OUTPUT TENSORS
for(p=1; p <= 3; p++){
line[p]=" "
for(q=1; q <= 3; q++) line[p] = sprintf("%s %8.3f",line[p],Smat[p,q])
}
for(p=0; p < n; p++){
line[p]=" "
}
for(p=0; p < n; p++) print line[p]

# rescale Smat to Hz relative to CH

```

```

for(p=1; p <= 3; p++)      for(q=1; q <= 3; q++)      Smat[p,q] *= cst2
for(p=1; p <= 3; p++){
    line[p]=""
    #      for(q=1; q <= 3; q++)      line[p] = sprintf("%s %8.3f",line[p],Smat[p,q])
    }
# vsan: rescale aa[] to Hz relative to CH
for(p=1; p <= 3; p++)      for(q=1; q <= 3; q++)      aa[p-1,q-1] *= cst2
for(p=1; p <= 3; p++){
    line[p]=""
    #      for(q=1; q <= 3; q++)      line[p] = sprintf("%s %8.3f",line[p],aa[p-1,q-1])
    }
print "Diagonalized tensor in Hz relative to CH"
for(p=1; p <= 3; p++) print line[p]

} #close BEGIN

```

```

*****
*MATLAB scripts to analyze PACE gel data and *
*compare theoretical and experimental fits of *
*different models *
*****
#####Curve_fits.m#####
clear;

hold off
conc=[0.000001 2 7 10 20 70 100 200 500];
phi = [0 0.08 0.18 0.2 0.36 0.63 0.73 0.96 0.999 ];
C=0.104e-3;
% Fit all functions
[k1,S1,rmsd1]=fitcurve1(conc,phi);
k1
rmsd1
[k2,S2,rmsd2]=fitcurve2(conc,phi);
k2
rmsd2
[k3,S3,rmsd3]=fitcurve3(conc,phi,C);
k3
rmsd3
[k4,S4,rmsd4]=fitcurve4(conc,phi,C);
k4
rmsd4
[k5,S5,rmsd5]=fitcurve5(conc,phi,C);
k5
rmsd5
[k6,S6,rmsd6]=fitcurve6(conc,phi,C);
k6
rmsd6
% Back-compute the curves
x=(0.00001:1:500);
fit1=x./(k1+x);
fit2=x.^2./(k2+x.^2);
fit3=(k3+x.*4*C-sqrt(k3.^2+x.*8*C*k3))./(-k3+x.*4*C+sqrt(k3.^2+x.*8*C*k3));
fit4=(x.^2+x.*C*k4(2))./(k4(1)*k4(2)+x.*k4(2)+x.^2);
%fit5=x.^2./(k5(1)^2*k5(2)+x.^2+k5(1)*k5(2)*x);
fit6= ((4.*C.*x.^3.*(x.^2+k6(2))-
x.^2.*sqrt(k6(1).^2*k6(2).^2+8.*C.*x.*k6(1)*k6(2).*(x.^2+k6(2)))+k6(1).*k6(2).*x.^2+k6(3).*(4.*C.*k6(2).*x.
*(x.^2+k6(2))-
k6(2).*sqrt(k6(1).^2*k6(2).^2+8.*C.*x.*k6(1).*k6(2).*(x.^2+k6(2)))+k6(1).*k6(2).^2))./(4.*C.*x.*(x.^2+k6(2)).^
2-k6(1).*k6(2).*(x.^2+k6(2)))+(k6(2)+x.^2).*sqrt(k6(1).^2.*k6(2).^2+8.*C.*x.*k6(1)*k6(2).*(x.^2+k6(2))));
%plot 6 curves
plot(conc,phi,'o','Color','r')
hold on
plot(x,fit1,'Color','c','LineWidth',2);
plot(x,fit2,'Color','g','LineWidth',2);
plot(x,fit3,'Color','m','LineWidth',2);
plot(x,fit4,'Color','b','LineWidth',2);
%plot(x,fit5,'Color','y','LineWidth',2);
plot(x,fit6,'Color','r','LineWidth',2);

```

```

#####PACEcurve1.m#####
function [estimates, model] = fitcurve1(xdata, ydata)
start_point = 100*rand(1);
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting1;
[estimates,sse1] = fminsearch(model, start_point,options);
rmsd2=sqrt(mean(sse2));
function [sse, FittedCurve] = fitting1(params)
k1 = params(1);
FittedCurve = xdata./(k1+xdata);
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
#####PACEcurve2.m#####
function [estimates, model,rmsd2] = fitcurve2(xdata, ydata)
start_point = 100*rand(1);
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting2;
[estimates,sse2] = fminsearch(model, start_point,options);
rmsd2=sqrt(mean(sse2));
function [sse, FittedCurve] = fitting2(params)
k2 = params(1);
FittedCurve = xdata.^2./(k2+xdata.^2);
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
#####PACEcurve3.m#####
function [estimates, model,rmsd3] = fitcurve3(xdata, ydata, C)
start_point = 0.01;
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting3;
[estimates,sse3] = fminsearch(model, start_point,options);
rmsd3=sqrt(mean(sse3));
function [sse, FittedCurve] = fitting3(params)
k3 = params(1);
FittedCurve = (k3+xdata.*4*C-sqrt(k3.^2+xdata.*8*C*k3))./(-k3+xdata.*4*C+sqrt(k3.^2+xdata.*8*C*k3));
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
#####PACEcurve4.m#####
function [estimates, model,rmsd4] = fitcurve4(xdata, ydata,C)
start_point = 100*rand(1,2);
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting4;
[estimates,sse4] = fminsearch(model, start_point,options);
rmsd4=sqrt(mean(sse4));
function [sse, FittedCurve] = fitting4(params)
k4a = params(1);
k4b = params(2);
FittedCurve = (xdata.^2+xdata.*C*k4b)./(k4a*k4b+xdata.*k4b+xdata.^2);
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
#####PACEcurve5.m#####
function [estimates, model,rmsd5] = fitcurve5(xdata, ydata,C)
% Call fminsearch with a random starting point.
start_point = rand(1,3);
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting5;
[estimates,sse5] = fminsearch(model, start_point,options);
rmsd5=sqrt(mean(sse5));
function [sse, FittedCurve] = fitting5(params)
k5a = params(1);

```

```

k5b = params(2);
D = params(3);
FittedCurve = ((k5b.*(k5a+xdata).^2.+4*C.*xdata.^2-
(k5a+xdata).*sqrt(k5b.^2*(k5a+xdata).^2))+D*(2*xdata.*sqrt(k5b^2*(k5a+xdata).^2+8*C*k5b*xdata.^2)-
(k5a+xdata).^2.*k5b.*xdata))./((xdata+k5a).*sqrt(k5b.^2*(k5a+xdata).^2+8*C*k5b*xdata.^2)-
k5b*(xdata+k5a).^2+4*C*xdata.^2);
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
#####PACEcurve6.m#####
function [estimates, model,rmsd6] = fitcurve6(xdata, ydata,C)
start_point = rand(1,3);
options = optimset('Display','final','TolX',1e-25, 'TolFun',1e-25,'MaxFunEvals',50000,'MaxIter',50000);
model = @fitting6;
[estimates,sse6] = fminsearch(model, start_point,options);
rmsd6=sqrt(mean(sse6));
function [sse, FittedCurve] = fitting6(params)
k6a = params(1);
k6b = params(2);
D = params(3);
FittedCurve = ((4.*C.*xdata.^3.*(xdata.^2+k6b)-
xdata.^2.*sqrt(k6a.^2*k6b.^2+8.*C.*xdata.*k6a*k6b.*(xdata.^2+k6b))+k6a.*k6b.*xdata.^2)+D.*(4.*C.*k6b.*xdata.*(xdata.^2+k6b)-
k6b.*sqrt(k6a.^2*k6b.^2+8.*C.*xdata.*k6a.*k6b.*(xdata.^2+k6b))+k6a.*k6b.^2))./(4.*C.*xdata.*(xdata.^2+k6b).
^2-k6a.*k6b.*(xdata.^2+k6b)+(k6b+xdata.^2).*sqrt(k6a.^2.*k6b.^2+8.*C.*xdata.*k6a*k6b.*(xdata.^2+k6b)));
ErrorVector = FittedCurve - ydata;
sse = sum(ErrorVector.^2);
end
end
*****
*Shell script to submit multiple jobs to the LSF queue*
*****
#!/bin/ksh
n=-1
nn=30 (number of input protocols to submit to the queue)
while ((n=n+1));((n<$nn))
do
bsub "/progs/bin/xplor <../dipo_optimize/input/sani_lj_ $n.inp >& ../dipo_optimize/output/$n.out"
done
exit
*****

```

Curriculum Vitae

Personal data

Name Devanathan Raghunathan
Date of birth 15th March 1980
Place of birth Secunderabad, India
Citizenship Indian

Education

03/1996 All India Secondary School Examination

03/1998 All India Senior Secondary Certificate Examination

05/2001 Bachelor of Science (with Honors) Chemistry, University of Delhi

05/2002-
07/2002 Internship at Bayer AG Leverkusen, Department of Polymers
Funded by AIESTE and DAAD

08/2003 Master of Science Chemistry, Indian Institute of Technology
Madras, Thesis titled "Application of Floquet theory and van
Vleck transformations to pulse sequences in NMR" under the
supervision of Prof. Dr. Mangala Sunder Krishnan

08/2003-
03/2007 PhD. in the department of NMR based Structural Biology at the
Max Planck Institute for Biophysical Chemistry pursued under
the guidance of Prof. Dr. Christian Griesinger. Thesis titled
"TAR-RNA Recognition by a Novel Cyclic Aminoglycoside
Analogue"

Lebenslauf

Persönliche Daten

Name Devanathan Raghunathan
Geburtstag 15.3.1980
Geburtsort Secunderabad, Indien
Staatsangehörigkeit Indisch

Education

03/1998 Allgemeine Hochschulreife, Delhi

05/1998 Beginn des Studium der Chemie an der Universität Delhi

05/2001 Bachelor of Science in Chemie (mit Auszeichnung),
Universität Delhi

05/2002-
07/2002 Praktikum bei der Bayer AG Leverkusen, Abteilung fuer
Polymerentwicklung (finanziert durch ein Stipendium des
DAAD und AIESTE).

08/2003 Master of Science in Chemie, Indisches Institut für
Technology, Madras, Titel der Arbeit "Anwendung des Floquet
Theorems und van Vleck Transformationen auf Pulssequenzen
in der NMR-Spektroskopie" unter Anleitung von
Prof. Dr. Mangala Sunder Krishnan

08/2003-03/2007 Doktorarbeit am Max-Planck-Institut für Biophysikalische
Chemie, Abteilung NMR-basierte Strukturbiologie (Prof. Dr.
C. Griesinger), Titel der Arbeit: "Erkennung von TAR-RNA
durch ein neues zyklisches Aminoglycosid-Analog".