Structure Determination of an HIV-1 RRE RNA-Rev Peptide Complex by NMR Spectroscopy

by

John L. Battiste

B.A. in Chemistry
Carleton College, 1990

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY
IN BIOCHEMISTRY

at the

Massachusetts Institute of Technology

September, 1996

Signature of Author.....
Department of Chemistry
July 25, 1996

Certified by.....
James R. Williamson
Associate Professor of Chemistry
Thesis Supervisor

Accepted by.....
Dietmar Seyferth
Chair, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Robert G. Griffin.....
Chair

Professor James R. Williamson.....
Thesis Supervisor

Professor Carl O. Pabo.....

ABSTRACT

Structure Determination of an HIV Rev Peptide-RRE RNA Complex by NMR Spectroscopy

by

John L. Battiste

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

The Rev Response Element (RRE) RNA-Rev protein interaction is important for regulation of gene expression in the human immunodeficiency virus. A model system for this interaction, which includes stem IIB of the RRE RNA and an arginine-rich peptide from the RNA-binding domain of Rev, was studied using nuclear magnetic resonance (NMR) spectroscopy. Analysis of exchangeable proton NMR spectra identified a conformational change in the RRE RNA upon binding of the Rev peptide with formation of four additional base pairs, including two non-Watson-Crick pairs. The hydrogen bonding geometry of a G:G and G:A base pair was determined and matched predictions based on biochemical experiments. Complete assignment of the RNA and peptide proton resonances was achieved using heteronuclear NMR with isotopically-labeled RNA and peptide samples. In particular, an approach utilizing selective ^{13}C-labeling of the RNA and isotopic filtering experiments was performed to obtain unambiguous assignments of unusual nonsequential NOE patterns present in the internal loop of the RRE. A three dimensional structure of the RNA-peptide complex was obtained from restrained molecular dynamics using distance and dihedral restraints from NMR NOE and J-coupling data. The two purine-purine base pairs of the RRE stack with adjacent A-form regions to form a continuous helix with the peptide binding in the major groove in an α-helical conformation. The Rev peptide contains an N-cap structure with a threonine side-chain making a hydrogen bond to the peptide backbone. Several arginines make hydrogen bond contacts to Watson-Crick base pairs in the major groove, while an asparagine residue contacts the G:A base pair. The formation of the G:G base pair requires an unusual geometry of the phosphate backbone in the internal loop that appears to open the major groove of the RNA to permit binding of the Rev α-helix. Overall, the structure formed by the two purine-purine base pairs of the RRE creates a distinctive binding pocket that the peptide can utilize for specific recognition.

Thesis Supervisor: James R. Williamson

Title: Associate Professor of Chemistry

ACKNOWLEDGMENTS

My fellow first year Williamson lab compatriot, Pat Zarrinkar, told a story in his thesis acknowledgments of his initial meeting with our advisor, Jamie Williamson. I remember a similar meeting with Jamie before joining the lab, where I too asked "Do I have to do NMR if I join the lab?" Jamie swallowed hard and said "no", yet the rest is history. I think my initial apprehension was based on fear of the unknown. Jamie, however, has always provided an atmosphere of ambitious goals, and an attitude that anything is possible and attainable. I think that has helped me achieve great things during my graduate school career, and I would like to thank him for that.

I would like to thank many people for assistance with my thesis project. Hongyuan Mao for extensive help with many aspects of the project over the past few years, as well as being a pleasant person to work with. I will miss the Tosci runs. Chris Turner for extensive help acquiring NMR spectra and an equally strong liking for Tosci's coffee. Christopher Cilley for mucho help and trouble-shooting with computer problems (Can I pay off my debt by the year 2000?). Alex Brodsky for many discussions on NMR experiments and molecular modeling. Jody Puglisi for providing my initial tutelage on RNA NMR spectroscopy and on the finer arts of giving a tirade. Tom Tolbert for much help with my misguided foray into organic synthesis.

I would like to acknowledge the initial members of the lab, Pat Zarrinkar, Robert Batey, and Heidi Erlacher for being around through thick and thin. Ah, the good old days at the CBC, and who can forget the Hong Kong!. Additional thanks to my baymate and lifting partner Heidi for somehow managing to not get annoyed by my sloppiness (you Go Girl!). Many thanks to the rest of the Williamson lab (Dan Treiber, Martha Rook, Jeff Orr, Feng Tao, Sharon Parker) for providing a stimulating and enjoyable work environment (despite all the carping and whining!). The lab really has been a home away from home, which is a scary thought. Additional thanks to the Governor's Jeff and Dan for companionship while Susy was in England. Similar thanks to Jay, Betsy, Nick and Michelle for helping me keep my sanity that year.

I would like to especially thank my parents who made this all possible. Finally, I would like to thank my wife Susy, who has provided an incredible amount of love and support and was a constant source of inspiration.

ABSTRACT.....3
ACKNOWLEDGMENTS.....4
CHAPTER 1: INTRODUCTION
1.1 RNA-Protein Interactions
1.1.1 Significance of RNA-protein interactions.....7
1.1.2 RNA structure.....10
1.1.3 Structural examples.....12
1.2 Structure Determination of RNA by NMR
1.2.1 General characteristics of RNA NMR spectra.....15
1.2.2 Standard NOESY assignment procedure.....19
1.2.3 Isotopic labeling of RNA and advanced NMR methods.....24
1.3 RRE-Rev Interaction Background
1.3.1 HIV gene expression and the life-cycle.....27
1.3.2 Biochemical analysis of high-affinity binding site.....27
CHAPTER 2: MATERIALS AND METHODS
2.1 Preparation of NMR Samples
2.1.1 Unlabeled RNA.....33
2.1.2 Labeled RNA.....35
2.1.3 Unlabeled and labeled peptides.....38
2.1.4 Formation of 1:1 complexes.....41
2.2 NMR Experiments
2.2.1 Homonuclear NMR.....43
2.2.2 Heteronuclear RNA NMR.....46
2.2.3 Heteronuclear peptide NMR.....49
2.3 Molecular Modeling
2.3.1 Assignment of restraint bounds.....50
2.3.2 Modeling protocols.....52
2.4 Other Methods
2.4.1 Size exclusion chromatography.....55
2.4.2 Circular dichroism.....55
CHAPTER 3: SECONDARY STRUCTURE OF RNA FREE AND BOUND
3.1 Construction of Minimal Complex for NMR
3.1.1 Design of RNA.....57
3.1.2 Design of peptides.....59
3.2 Analysis of Exchangeable NMR Spectra
3.2.1 Free RNA.....65

3.2.2 Bound RNA.....67
3.3 Analysis of Nonexchangeable NMR Spectra.....72
CHAPTER 4: STRUCTURE DETERMINATION OF RNA-PEPTIDE COMPLEX
4.1 Assignment Procedure for RNA
4.1.1 Through-bond correlation experiments.....80
4.1.2 NOESY experiments.....83
4.1.3 Summary of NMR Data.....92
4.2 Assignment Procedure For Peptid
4.2.1 Through-bond experiments.....95
4.2.2 NOESY experiments.....96
4.3 Identification of RNA-Peptide NOEs
4.3.1 Assignment by unambiguous chemical shift ranges.....102
4.3.2 Filtered NOESY experiments.....104
4.4 Molecular Modeling of the Complex.....106
CHAPTER 5: STRUCTURE OF RNA-PEPTIDE COMPLEX
5.1 Global Architecture of the RNA-Peptide Binding Site
5.1.1 RNA structure.....115
5.1.2 Peptide structure.....121
5.1.3 Positioning of α-helix in major groove.....126
5.2 Specific RNA-Peptide Contacts
5.2.1 Nucleotide base contacts.....130
5.2.2 Phosphate backbone contacts.....134
5.2.3 Van der Waals contacts.....136
5.2.4 Summary of contacts.....138
5.3 Comparison to Other DNA- and RNA-Protein Interaction
5.3.1 α-helix/major groove recognition.....141
5.3.2 The arginine-rich motif.....142
5.3.3 Induced fit of RNA-protein interactions.....143
5.3.4 Contribution of RNA structure to binding specificity.....146
5.4 Future Research
5.4.1 Larger fragments of RRE and Rev.....148
5.4.2 Drug binding.....150
5.4.3 Site directed mutants of RRE and Rev.....151
APPENDIX A: NMR PULSE SEQUENCES.....155
APPENDIX B: MOLECULAR MODELING DETAILS.....169
REFERENCES.....188

CHAPTER 1: INTRODUCTION

1.1 RNA-Protein Interactions

1.1.1 Significance of RNA-protein interactions

Large ribonucleoprotein complexes Large macromolecular RNA-protein complexes, such as the ribosome, spliceosome, signal recognition particle (SRP), and telomerase, are responsible for mediating many fundamental biological processes. Perhaps the most well studied complex is the ribosome, which catalyzes protein synthesis, one of the most fundamental functions of life. The prokaryotic ribosome has a molecular weight of approximately 2,500 kilo daltons (kDa) and is composed of ~55 proteins and 3 large RNA molecules. Seminal work in the 1960's and 70's purified all of the components of the ribosome of *E. coli* and elucidated a precise assembly pathway for functional ribosomes that could be replicated in vitro (Nomura, 1973). Interactions of "primary"-binding proteins with the ribosomal RNAs initiate a cascade of subsequent binding by the remainder of the ribosomal proteins in an orderly process to produce the small and large ribosomal subunits. This assembly is almost certainly mediated by a precise set of RNA-protein and protein-protein interactions. Early dogma regarding the ribosome simply viewed the RNA as a "scaffold" for binding of the ribosomal proteins. However, recent work has provided very suggestive evidence that the catalytic function of the ribosome (peptide bond formation) is achieved by the RNA rather than the protein (Noller et al., 1992). Therefore, it is evident that precise three-dimensional structures of the RNA will be critical for ribosomal function.

The spliceosome is another large ribonucleoprotein assembly in eukaroytes that removes introns from messenger RNA (mRNA). The spliceosome is composed of 5 small nuclear RNA (snRNA) molecules (U1, U2, U4, U5, U6) and a growing list of several

dozen protein molecules that are either stable, "core" elements of the spliceosome or transient regulatory factors (Guthrie, 1991). While less has been elucidated about the mechanism of function than the ribosome, it is still clear that precise RNA-RNA and RNA-protein interactions are important for assembly of the snRNP particles on the mRNA and for catalytic function. The mechanism of RNA cleavage and ligation in the spliceosome is very similar to the self-splicing group-II introns, suggesting, in lieu of direct evidence, that the RNA is again the catayltically active molecule (Guthrie, 1991). It is intriguing these important ribonucleoprotein complexes are vestiges of an "RNA-world" (Gesteland and Atkins, 1993), when RNA was responsible for catalytic function in all of life's processes, and the proteins were later incorporated to augment catalytic function and provide more sophisticated regulatory control.

Regulation of gene expression In addition to the general role of RNA-protein interactions in gene expression (ribosome, spliceosome, etc.), formation of specific RNA-protein complexes has been implicated in a variety of novel regulatory mechanisms at all levels of gene expression, including transcription, mRNA-processing, and translation (Aziz and Munro, 1987; Hentze et al., 1987; Owen and Kuhn, 1987; Bell et al., 1988; Feng and Holland, 1988; Berkhout et al., 1989; Lazinski et al., 1989; Malim et al., 1989; Boelens et al., 1993; Klausner et al., 1993; Murata and Wharton, 1995).

There are two well-established mechanisms of transcriptional regulation involving interactions with a growing mRNA transcript. The Tat protein of the immunodeficiency viruses binds to a small RNA element in the 5'-Long Terminal Repeat (LTR) of nascent viral RNA transcripts and interacts with other cellular factors and the RNA polymerase complex to increase the processivity of transcription (Feng and Holland, 1988; Berkhout et al., 1989). The N protein of many bacteriophages also binds to a small RNA element in a nascent RNA transcript and interacts with other cellular proteins and RNA polymerase to produce a termination resistant complex (Lazinski et al., 1989).

Many other RNA-protein complexes are involved in the regulation of pre-mRNA processing. The Rev protein of the human immunodeficiency virus (HIV) interacts with the Rev Response Element RNA to inhibit splicing and/or facilitate export of partially or unspliced HIV mRNA (Malim et al., 1989). There are many factors that regulate the choice of splice sites in pre-mRNA to produce different gene products. One example is protein Sex-lethal, which is the master regulator of sex determination in *Drosophila.* Sex-lethal causes an alternative 3'-splice site to be utilized in *tra* pre-mRNA by interacting with pyrimidine-rich sequences near the normal splice site (Bell et al., 1988). Normal processing of mRNA in eukaryotes involves capping of the 5'-end with methylated guanosine triphosphate and addition of a poly-adenosine tail to the 3'-end before export to the cytoplasm. The U1A protein regulates its own expression by binding to a RNA sequence in the 3'-Untranslated Region (UTR) of the mRNA, preventing poly-adenylation and subsequent nuclear export.

Once in the cytoplasm, RNA-protein interactions can regulate gene expression by affecting the stability of the mRNA or modulating interactions with the ribosome. A classic example of both of these mechanisms is the regulation of iron metabolism genes in mammalian cells through the Iron Response Element (IRE) in mRNA for the transferrin receptor and ferritin (Aziz and Munro, 1987; Hentze et al., 1987; Owen and Kuhn, 1987; Klausner et al., 1993). The Iron Response Element-Binding Protein (IRE-BP) binds a small hairpin loop (IRE) in the 5'-UTR of ferritin mRNA and inhibits initiation of translation. IRE-BP also binds to multiple IREs in the 3'-UTR of transferrin receptor mRNA, inhibiting degradation. IRE-BP is postulated to be allosterically regulated by iron and only binds RNA in the iron-free form, resulting in decreased expression of ferritin and increased expression of transferrin receptor. The IRE system is a fascinating example of post-transcriptional regulation of gene expression with binding of the same protein to similar RNA elements resulting in two different and opposite mechanisms of regulation.

The IRE is unlikely to be a isolated example and there have been many postulated, yet not well defined mechanisms for regulation of translation through protein interactions with elements in mRNA (most notably in Drosophila development) (Curtis et al., 1995).

1.1.2 RNA structure Phylogenetic studies have shown that conserved regions of biologically important RNAs often reside in loops, bulges, and internal loops (Figure 1.1) (Gutell et al., 1985). Naive assumptions about RNA structure might view these regions as "single-stranded" nucleotides linking together the Watson-Crick helices of RNA. Structural studies on a variety of RNAs, however, have revealed that these regions are often stabilized by non-Watson-Crick (mismatch) base pairs or other non-standard base-base or base-phosphate backbone interactions (Cheong et al., 1990; Heus and Pardi, 1991; Szewczak et al., 1993; Wimberly et al., 1993). The earliest and still most elegant example of the intricacies of RNA structure is tRNA, which forms a complex L-shape stabilized by many long-range tertiary interactions between the "loops" in the clover-leaf secondary structure representation (Jack et al., 1976; Rich and Kim, 1978). In addition to providing a role in RNA structure, non-Watson-Crick base interactions may also be important for RNA-protein recognition. Biochemical and genetic identification of protein binding sites in RNAs has shown that they often include important non-Watson-Crick regions (Gutell et al., 1985; Roy et al., 1990; Cook et al., 1991). The complex three-dimensional structures formed by these non-standard regions of RNA can potentially provide a unique environment for specific recognition by proteins.

In general the "problem" of RNA-recognition by proteins should share many similarities to DNA-recognition, since the chemical composition of the base and phosphate backbone are very similar. Therefore, the principles of DNA recognition ascertained from many structural studies of DNA-protein complexes may be applicable to RNA recognition in many cases. However, the singled-stranded nature of RNA in vivo allows it to form

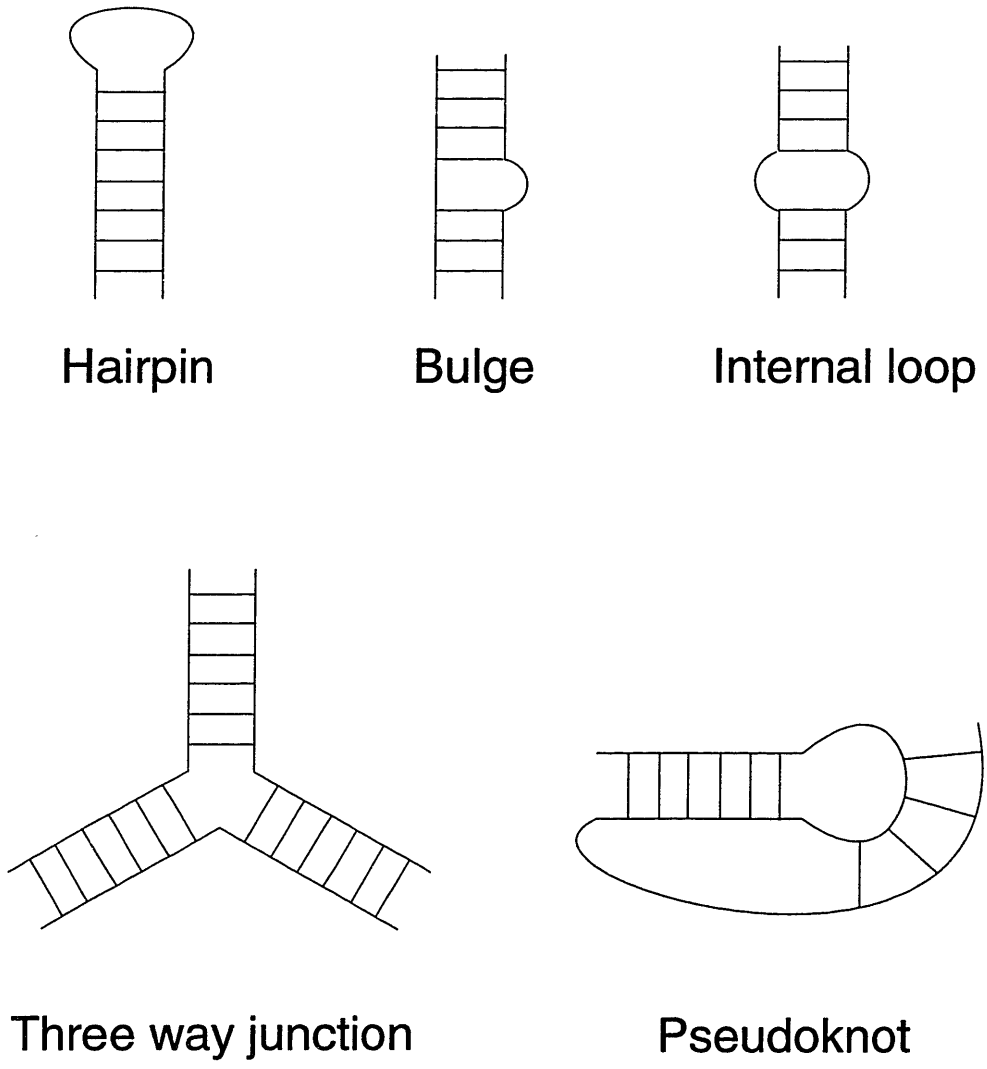


Figure 1.1: RNA secondary structure elements. Lines denote Watson-Crick base pairs. Other regions represent non-Watson-Crick base pairing nucleotides that may or may not be involved in base pairing interactions.

unusual structures with distingushing characteristics that do not exist with DNA. The recognition of these unusual RNA structures is likely to be a key component. The major goal of this thesis will be to provide structural information on a RNA-protein interaction, and in particular understand the contribution of RNA structure to the recognition process.

Even the structural features that RNA and DNA share in common have differences that impact upon protein-recognition. The Watson-Crick base pairing regions of DNA and RNA generally form double helices of distinct morphology (B-form vs. A-form, respectively). The recognition of hydrogen bond donors and acceptors on the major groove edge of nucleotides in B-form helices is a common feature in DNA-protein recognition (Pabo and Sauer, 1992). A similar interaction with RNA is sterically hindered by the narrow and deep major groove of A-form helices. However, biochemical analysis of some RNA-protein interactions by chemical protection or modification interference has suggested major groove recognition of Watson-Crick base pairs (Calnan et al., 1991; Weeks and Crothers, 1991; Kjems et al., 1992; Tiley et al., 1992). It has been recognized that adjacent bulges or other non-standard regions must be important for widening the major groove. In addition, it has been shown experimentally that the major groove of A-form RNA becomes accessible to small chemical probes in context of adjacent bulge structures (Weeks and Crothers, 1993) .

1.1.3 Structural examples Despite their importance, relatively few high resolution structures of RNA-protein complexes have been solved to date. Three tRNA-synthetase cocrystal structures have provided the most detailed examples, showing that proteins can recognize both the global structural features and base sequence of RNA (Rould et al., 1989; Ruff et al., 1991; Biou et al., 1994). Two of the synthetases utilize the L-shape of the tRNA for recognition making contacts from the acceptor stem to the anticodon loop. Seryl-tRNA synthetase, on the other hand, utilizes the long variable arm extension characteristic

of $tRNA^{Ser}$ instead of the anti-codon stem. In addition to the tRNAs, the cocrystal structures of U1A and MS2 coat protein bound to small RNA elements have provided additional insights into protein recognition of RNA through contacts with nucleotides in hairpin loops (Oubridge et al., 1994; Valegard et al., 1994). In both cases, sequence specific recognition was achieved through interactions with splayed bases in the loops, and a unique RNA conformation did not appear to contribute to recognition.

In addition, several solution structures of RNA-protein or RNA-peptide complexes have been solved by nuclear magnetic resonances (NMR) spectroscopy. The structure of U1A protein bound to an asymmetric internal loop in its own mRNA identified similar sequence specific interactions with splayed nucleotides as the co-crystal structure with U1 snRNA (Allain et al., 1996). These two in vivo RNA binding sites of the U1A protein have identical sequences that interact with the protein, but they are presented in a slightly different secondary structure context (hairpin loop and asymmetric internal loop). The HIV regulatory proteins Tat and Rev, which bind to TAR and RRE RNAs, are prototypes of a class of proteins containing what is known as an arginine-rich motif (ARM) (Lazinski et al., 1989). One characteristic feature of many proteins containing ARMs is that short arginine-rich peptides are capable of binding specifically to their respective RNA elements in the absence of a surrounding protein framework (Calnan et al., 1991; Weeks and Crothers, 1991; Kjems et al., 1992; Tan et al., 1993; Chen and Frankel, 1994; Tan and Frankel, 1995). The small size of these minimal RNA-peptide elements has made them suitable for structural analysis by nuclear magnetic resonance (NMR) spectroscopy. The structures of HIV and bovine immunodeficiency virus (BIV) TAR bound to argininamide and BIV tat peptide, respectively, have been by solved by NMR (Puglisi et al., 1992; Aboul-ela et al., 1995; Puglisi et al., 1995; Ye et al., 1995). Both complexes contain arginine-guanine interactions with Watson-Crick base pairs in the context of a widened major groove of the RNA. In HIV TAR, a base triple is proposed to widen the major

groove and create a binding site for the arginine to interact with a guanosine. For the BIV complex, TAR RNA recognition is mediated by a β-hairpin conformation of the Tat peptide. An isoleucine residue interacts with a bulged uracil creating a structured context for arginine and threonine contacts with other bases and phosphates in the major groove. While a lot has been learned from these examples, additional structural studies are clearly needed to more fully understand how proteins recognize small, tertiary RNA elements such as bulges, hairpin loops, internal loops, and junctions. The Rev-RRE complex to be presented in this thesis will provide another example of peptide-recognition of a small RNA element.

1.2 Structure Determination of RNA by NMR

There are two experimental variables that are used for determination of three-dimensional structures by NMR. They are the nuclear Overhauser effect (NOE), whose intensity has a $1/r^6$ dependence on the distance between two protons, and the coupling constant (J), whose value is a trigonometric function of the torsion angle of two protons separated by three bonds (note that in principle, NOEs or coupling constants involving atoms other than protons can be used as structural restraints, yet these values are more difficult to obtain). All of the NOE and J-coupling data can be potentially obtained from the 2-dimensional (2D) NOESY (Nuclear Overhauser Effect SpectroscopY) and COSY (COrrelation SpectroscopY) NMR experiments. For larger biological macromolecules, however, spectral overlap can prevent resonance assignment (the process of linking the chemical shift of a proton resonance to a particular proton in the molecule) or NOE/J-coupling assignment to an unique pair of protons. Much research over the past 10-15 years has focused on new experiments designed to circumvent these problems. Many of these experiments are general in nature (applicable to all types of molecules), while an increasing

number are tailored to the NMR properties of a particular molecule (protein, nucleic acid, carbohydrate, lipid, etc.). In general, research on "protein NMR" is farther advanced than [illegible] macromolecules, in part due to the earlier availability of isotopically labeled samples. Therefore, while the complex studied in this thesis contains both an RNA and a protein component, more emphasis will be given to the description of the process of structure determination for the RNA. In addition to the greater novelty of structure determination of RNA, many of the experiments with the peptide were done in collaboration with Lewis Kay of the University of Toronto, who has extensive experience with specific protein NMR methods and was able to solve a problem of overlap of the arginine side chain resonances. As such, only the results, rather than the entire process of obtaining the NMR results for the peptide will be described.

1.2.1 General characteristics of RNA NMR spectra

Exchangeable protons Nucleic acids (RNA and DNA) contain "exchangeable" (protons that readily exchange with the solvent) imino protons (Figure 1.2A) on guanosine and uracil bases that have unique properties making them useful for structural analysis. Imino protons resonate far downfield (~11-14 ppm) and do not overlap with other proton resonances of the RNA. More importantly, imino protons exchange rapidly with water, such that they are typically not observed unless they are involved in hydrogen bonds. Imino protons are involved in hydrogen bonds in Watson-Crick and most non-Watson-Crick base pairs; therefore, the observation of imino proton resonances is a good indicator of secondary structure. Imino protons in G:C and A:U base pairs give characteristic intrabase pair NOEs to C-amino and A-H2 protons, respectively. Furthermore imino-imino NOEs between stacked base pairs provide a straight forward method for identification of helical regions of RNA.

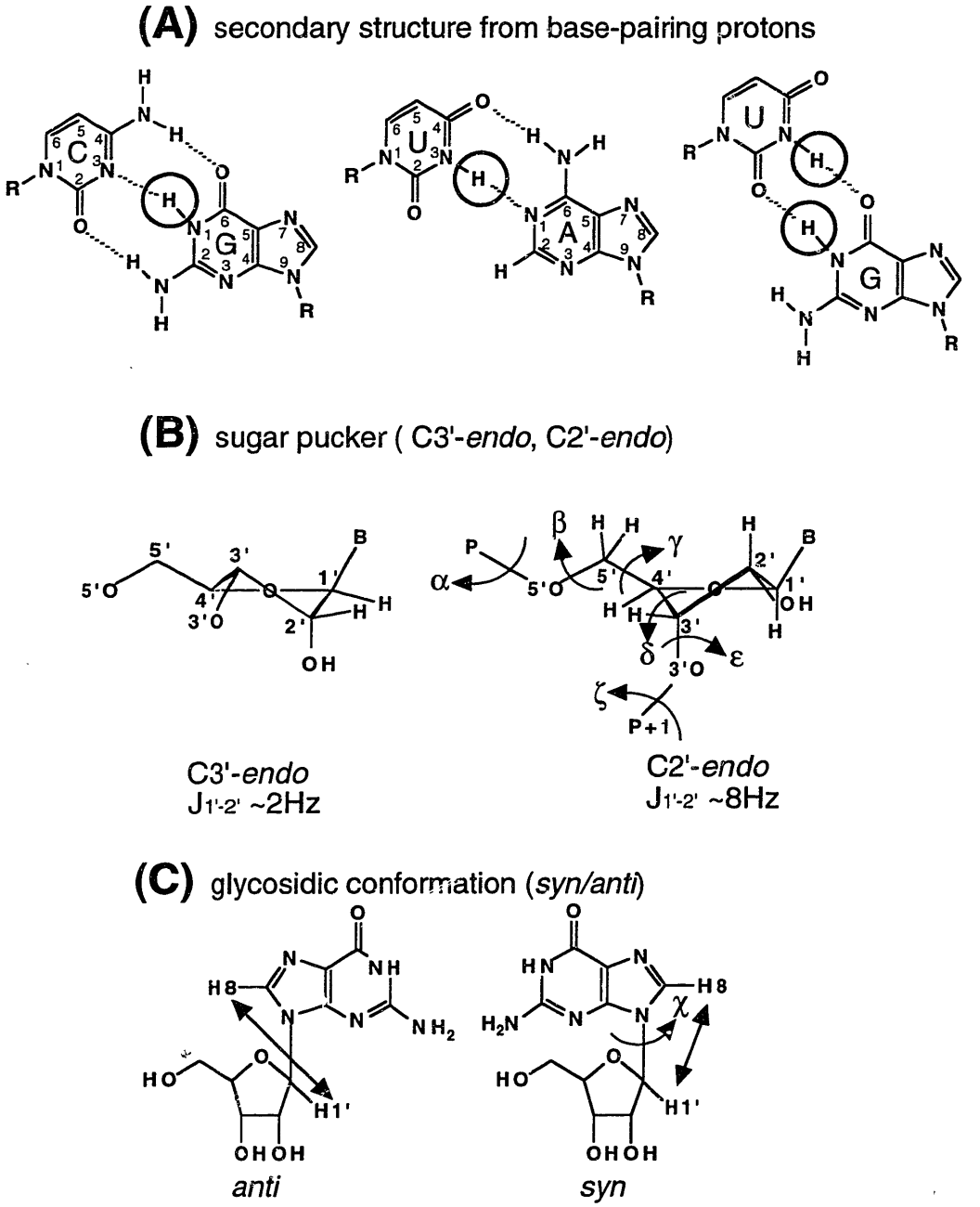


Figure 1.2: Basic RNA structural features that are easily characterized from NMR analysis. For reference, numbering and torsion angle nomenclature are also included in the figure. **(A)** Imino proton resonances (circled proton) are only observed when involved in base pairing. **(B)** The sugar pucker of a ribose can be determined from the coupling constants of the ribose protons. **(C)** The glycosidic conformation can be determined from base to ribose proton NOE intensity.

Sugar pucker and glycosidic torsion These two local conformational features that are important to nucleic acid structure can be qualitatively determined from simple analysis of nonexchangeable NMR spectra. The sugar pucker can be determined from the $^3J_{H1'-H2'}$ coupling constant (Figure 1.2B). The H1'-H2' cross peak resonates in a well-resolved region of the COSY spectrum. C3'-endo sugar pucker, which is the conformation of nucleotides in an A-form helix, has a $^3J_{H1'-H2'}$ of <3Hz that is unobservable for medium to large RNA molecules. C2'-endo sugar pucker, which is often the conformation of unstructured or non-standard regions of RNA structure, has a $^3J_{H1'-H2'}$ of >8Hz.

The glycosidic torsion angle (χ) of a nucleotide can be classified into the two crude conformations, *syn* and *anti*, from analysis of the intranucleotide base to ribose NOEs (Figure 1.2C). *Anti* is the energetically preferred conformation, yet *syn* bases are observed in many types of non-Watson-Crick base pairs. The base to H1' distance is particularly sensitive to the torsion angle χ (~2.5 Å for *syn* vs. ~3.8 Å for *anti*). Therefore, the presence of strong intranucleotide H8-H1' NOE is evidence for a *syn* nucleotide. In addition, the intranucleotide NOE intensities to other ribose protons can be used to determine χ; however, their spectral identification is not as straightforward as the H8/H1' NOE.

Non-exchangeable protons The 1D spectrum of the RRE RNA in 99.996% D_2O (Figure 1.3) shows the chemical shift ranges of the non-exchangeable base and ribose protons of RNA. The most notable feature is that the majority of the ribose protons resonate within a narrow 1 ppm range (~4-5 ppm). This severe crowding of resonances has made structure determination of RNA by homonuclear NMR methods very difficult compared to protein and DNA molecules. For this reason, heteronuclear NMR methods have been particularly essential for the progress of structure determination of RNA by NMR (see section 1.2.3 and 4.1).

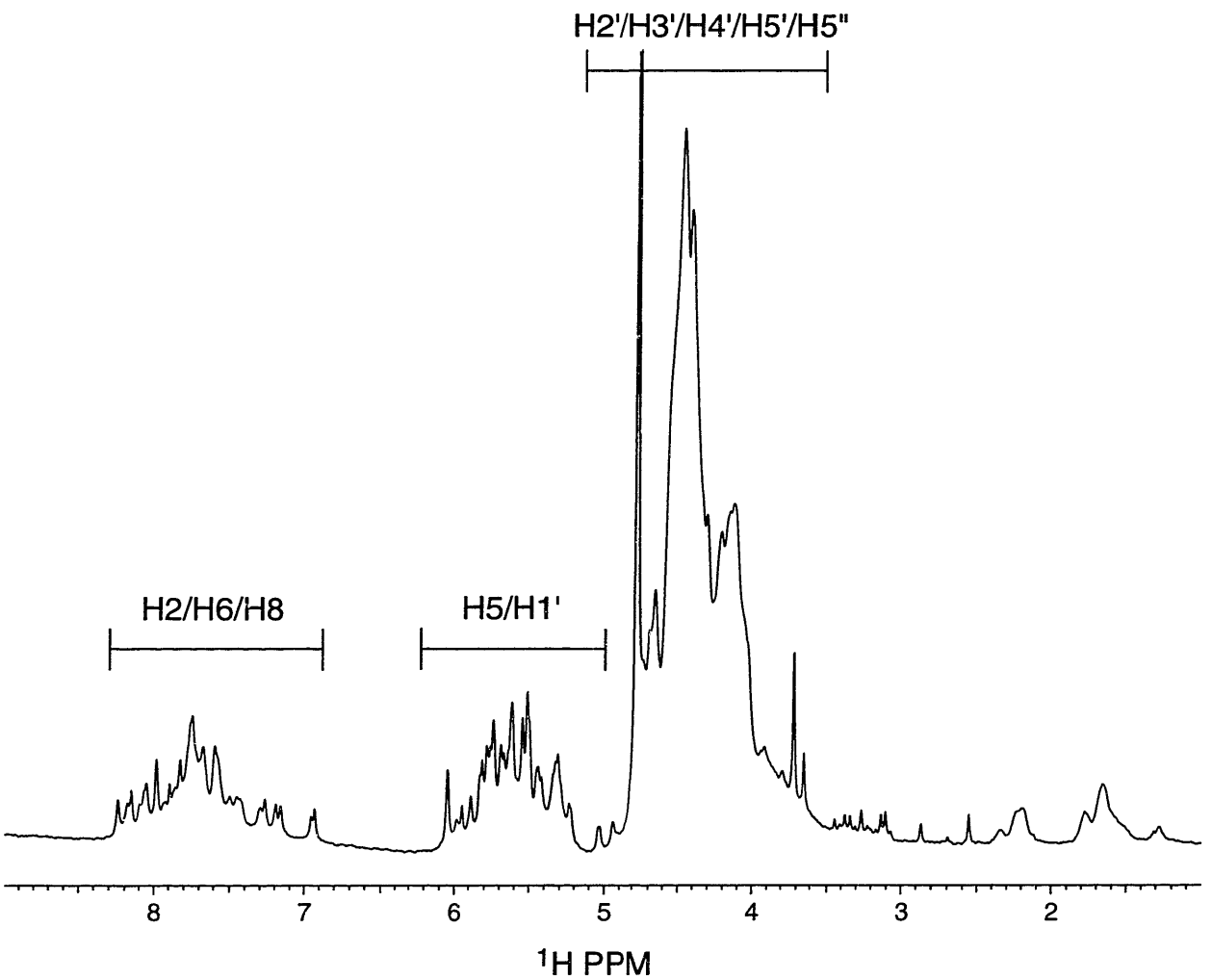


Figure 1.3: One dimensional spectrum of nonexchangeable RRE RNA protons in 99.996% D_2O. Bars delineate the standard chemical shift ranges for RNA protons. Note that half the resonances are between 3.7-5.0 ppm.

Exchange in NMR spectroscopy The manifestation of NMR resonances oscillating between distinct states (i.e. different chemical shifts) depends on the rate of "exchange" between these states. This includes both "chemical exchange" (e.g. with solvent) and "conformational" exchange (e.g. a ribose switching pucker). Discussion of exchange processes occurs several times in the thesis; therefore, the definition of the three definable exchange states in NMR (fast, slow, and intermediate) should be elaborated (Wuthrich, 1986). Note that the terminology "nonexchangeable" or "exchangeable" used in the previous sections only refers to the exchange properties of a proton with solvent, and all protons may still undergo conformational exchange. Figure 1.4 shows hypothetical spectra for a proton in exchange between two distinct conformations with chemical shifts δ_A and δ_B. "Fast" exchange occurs when the rate of exchange (k_{AB}) between conformations is much greater than the chemical shift difference between the protons in each state ($\Delta\delta = \delta_A - \delta_B$). For fast exchange, one resonance is observed at a chemical shift that is a weighted average of the relative populations of the two states. The weighted average is particular significant for exchange processes with solvent, since the concentration of H_2O is 55M. As result, a proton in fast exchange with water resonates with a chemical shift essentially at the water resonance and is therefore unobservable. "Slow" exchange occurs when $k_{AB} << \Delta\delta$, resulting in observation of distinct resonances for each state "Intermediate" exchange occurs when the rates are matched at $k_{AB} \sim \Delta\delta$, resulting in broadening or complete disappearance of the resonances. For typical chemical shift differences between protons in different structural environments (0.1 - 1.0 ppm), the threshold rate between slow and fast exchange is about 300-3000 Hz on a 500 MHz instrument. In other words, the lifetimes of the conformational states are on the 0.3-3 millisecond time scale (1/rate).

1.2.2 Standard NOESY assignment procedure The standard procedure for assigning RNA molecules is a sequential NOESY "walk" from base protons of one

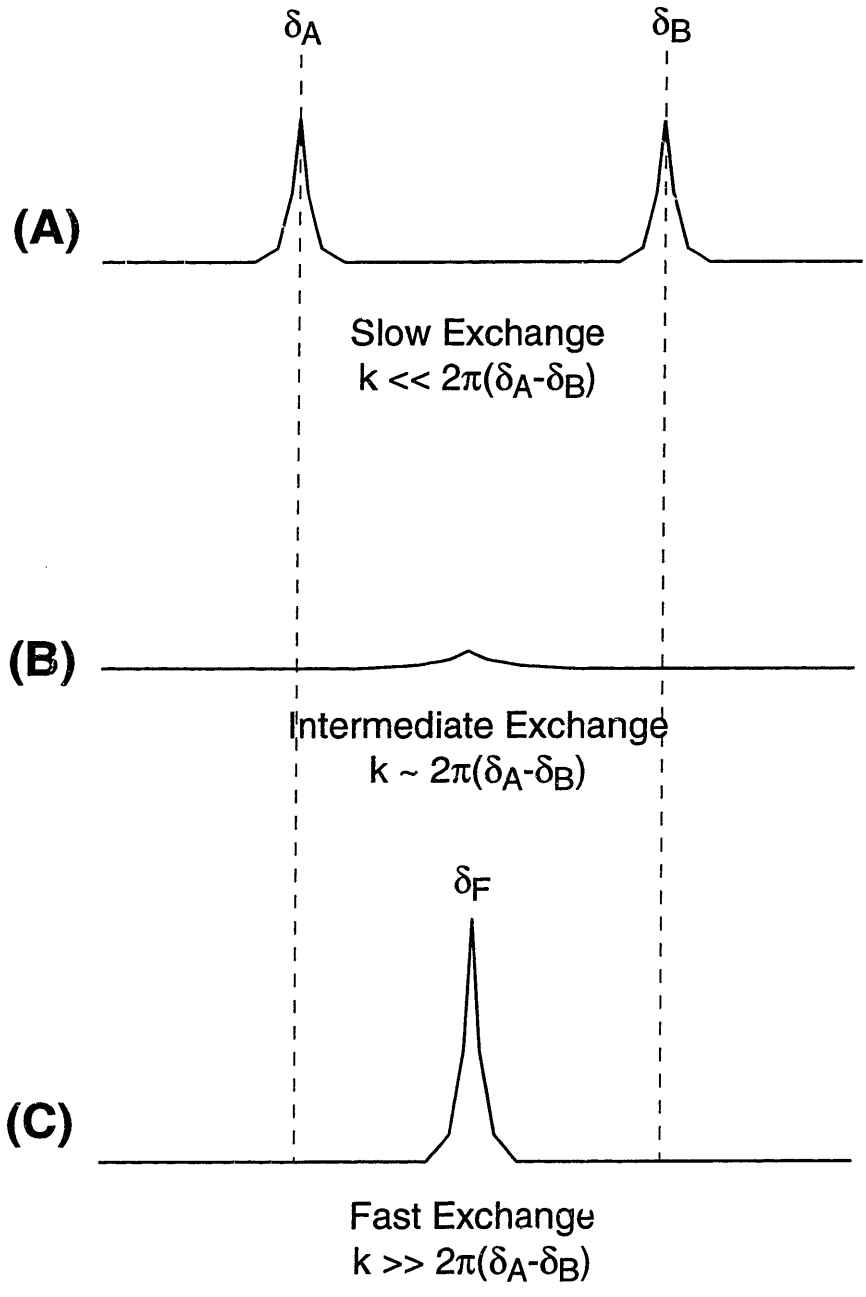


Figure 1.4: Schematic of different exchange regimes in NMR spectroscopy. δ_A and δ_B are the chemical shifts (in Hz) of two distinct states in chemical or conformational exchange. k is the rate of exchange between the two states. **(A)** Slow exchange. A separate resonance is observed for each state. **(B)** Intermediate exchange. When the rate is perfectly matched with δ_A-δ_B, no resonance is observed. When slightly greater or less than δ_A-δ_B, broadened resonances are observed. **(C)** Fast exchange. One time averaged resonance is observed with a chemical shift (δ_F) that is a weighted average of the relative populations of state A and B.

nucleotide to the ribose protons of the 5'-nucleotide (Figure 1.5). This sequential walk is typical of helical nucleotides and may or may not be observed in non-standard regions of RNA structure. For this reason, assignment of RNA resonances by through-bond methods, which are only dependent on primary structure, are being heavily investigated (see section 1.2.3). Nevertheless, as long as significant portions of A-form helical regions are present, an RNA can be assigned using the NOESY-based procedure.

Since H1' are the only well-resolved ribose protons, the region of the NOESY spectrum containing cross peaks between base and H1' protons is used for sequential assignment of RNA. The nonexchangeable base protons (H8/H6/H2) resonate between ~6.5-9.0 ppm, while the H1' resonances are typically between ~4.8-6.5 ppm. The 4.8-6.5 ppm region also contains the pyrimidine base H5 protons, as well as other downfield shifted ribose protons. Therefore, care needs to be taken in identifying H1' resonances. Figure 1.6 shows hypothetical spectra for a short double helical segment of RNA. For A-form RNA, there is a characteristic pattern of NOEs involving the adenosine H2 proton to the H1' protons of nucleotides 3' and 5'-cross strand to the adenosine. The adenosine H2 protons can be assigned from NOEs to the exchangeable U imino protons in Watson Crick base pairs and then be used as a starting point for sequential assignment. Since H1' resonances cannot be distinguished by nucleotide type, the base protons are used to distinguish the type of nucleotide. Even then, only pyrimidine and purine H8/H6 protons can be distinguished by the H5/H6 crosspeaks of pyrimidines in both NOESY and COSY spectra. To assign the RNA, a unique pattern of purine to pyrimidine connectivities needs to be established. For instance, the hypothetical pattern of NOEs in Figure 1.6 gives two strand connectivities, 5'-RYR-3' and 5'-YRY-3', where R and Y stand for purine and pyrimidine, respectively. Knowing the primary sequence of the RNA, it is straightforward to convert these assignments into the respective 5'-GUG-3' and 5'-CAC-3' strands.

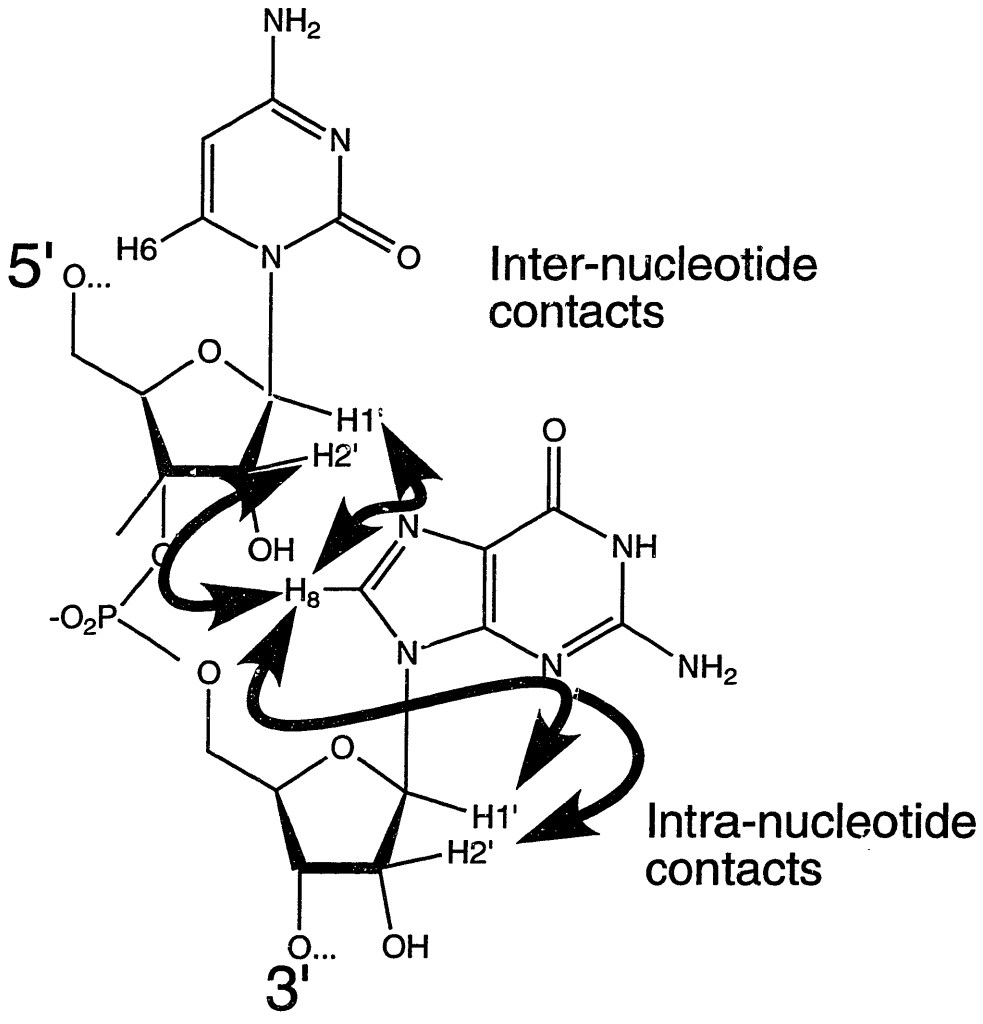


Figure 1.5: Sequential assignment pathway of nucleic acids. Arrows depict NOEs that are observed in regular A-form or B-form helices.

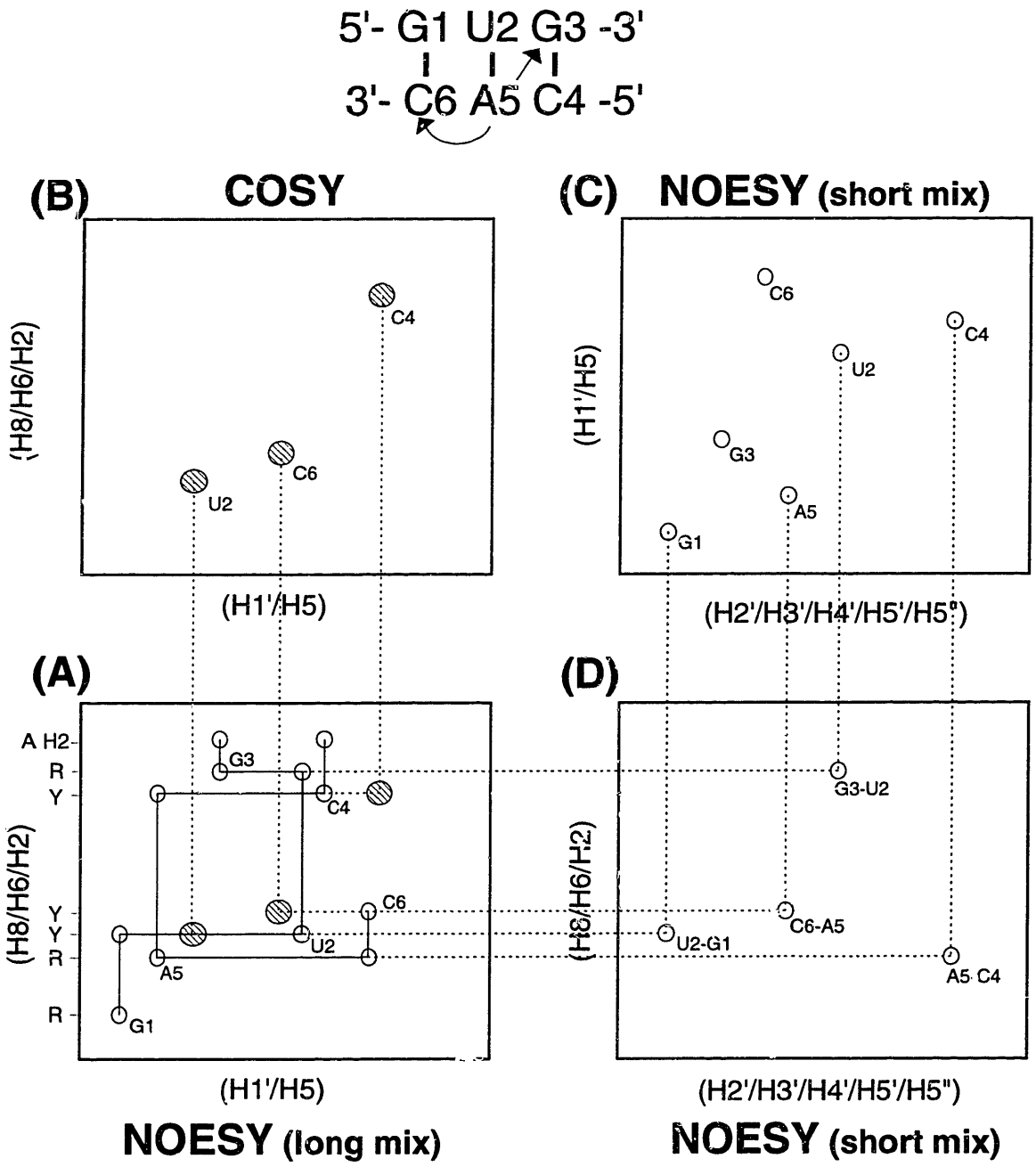


Figure 1.6: Schematic of NOESY-based assignment procedure for a model A-form duplex (above with arrows showing canonical A-form adenosine H2-H1' NOEs). **(A)** Base to H1' region of NOESY (mixing time 300-400ms). R and Y indicate purine and pyrmidine H8/H6 resonances identified by the absense or presence, respectively, of H5-H6 cross-peaks (hatched). The labels indicate intranucleotide NOEs. **(B)** Region of COSY spectrum identifying pyrimidine H5-H6 cross peaks. **(C)** H1'-ribose region of a NOESY spectrum (mixing time ~50ms) identifying intranucleotide H1'-H2' crospeaks. **(D)** Base-ribose region of a NOESY spectrum (mixing time ~50ms) identifying strong internucleotide base-H2' NOEs characteristic of A-form RNA.

1.2.3 Isotopic labeling of RNA and advanced NMR methods The homonuclear NMR sequential assignment strategy detailed above only works for RNA molecules <20-30 nucleotides. In addition, identification of NOEs to ribose protons other than the H1' becomes increasing difficult for larger RNA molecules, decreasing the number of NOEs and quality of structures calculated. The availability of ^{13}C-labeled RNA samples (Batey et al., 1992; Nikonowicz et al., 1992) revolutionized the field of RNA spectroscopy, allowing a multitude of heteronuclear NMR experiments to performed. Initial work in this field was a recapitulation of general heteronuclear experiments that been successful in the protein NMR field. In particular, this included three dimensional (3D) versions of the standard 2D NOESY, COSY, and TOCSY experiments (e.g. NOESY-HSQC) (Pardi and Nikonowicz, 1992; Nikonowicz and Pardi, 1993). These experiments spread the two dimensional spectra into a third dimension according to the chemical shift of the heteronucleus attached to one of the protons involved in the 2D crosspeak (Figure 1.7). This reduces spectral overlap and allows for both easier sequential assignment and identification of more NOEs for structure determination.

While not utilized in this thesis, a plethora of new heteronuclear experiments tailored specifically for RNA molecules have been described in the last few years. A large number of these experiments aim towards the goal of complete assignment of RNA proton resonances by through-bond rather then NOE based transfer, which is more prone to error in interpretation. These developments mirror the protein NMR field, which now has a large repertoire of triple resonance (^{1}H, ^{13}C, ^{15}N) correlation experiments for sequential assignment of protein side chains by transfer of magnetization through the protein backbone (Bax, 1994). The new RNA sequential assignment experiments can be broken down into two classes. One class of triple resonance (^{1}H, ^{13}C, ^{15}N) experiments correlate the isolated proton spin systems on nucleotide bases to one another and to the ribose H1' protons through the large $^{1}J_{CC}$ and $^{1}J_{CN}$ couplings (Marino et al., 1993; Sklenar et al.,

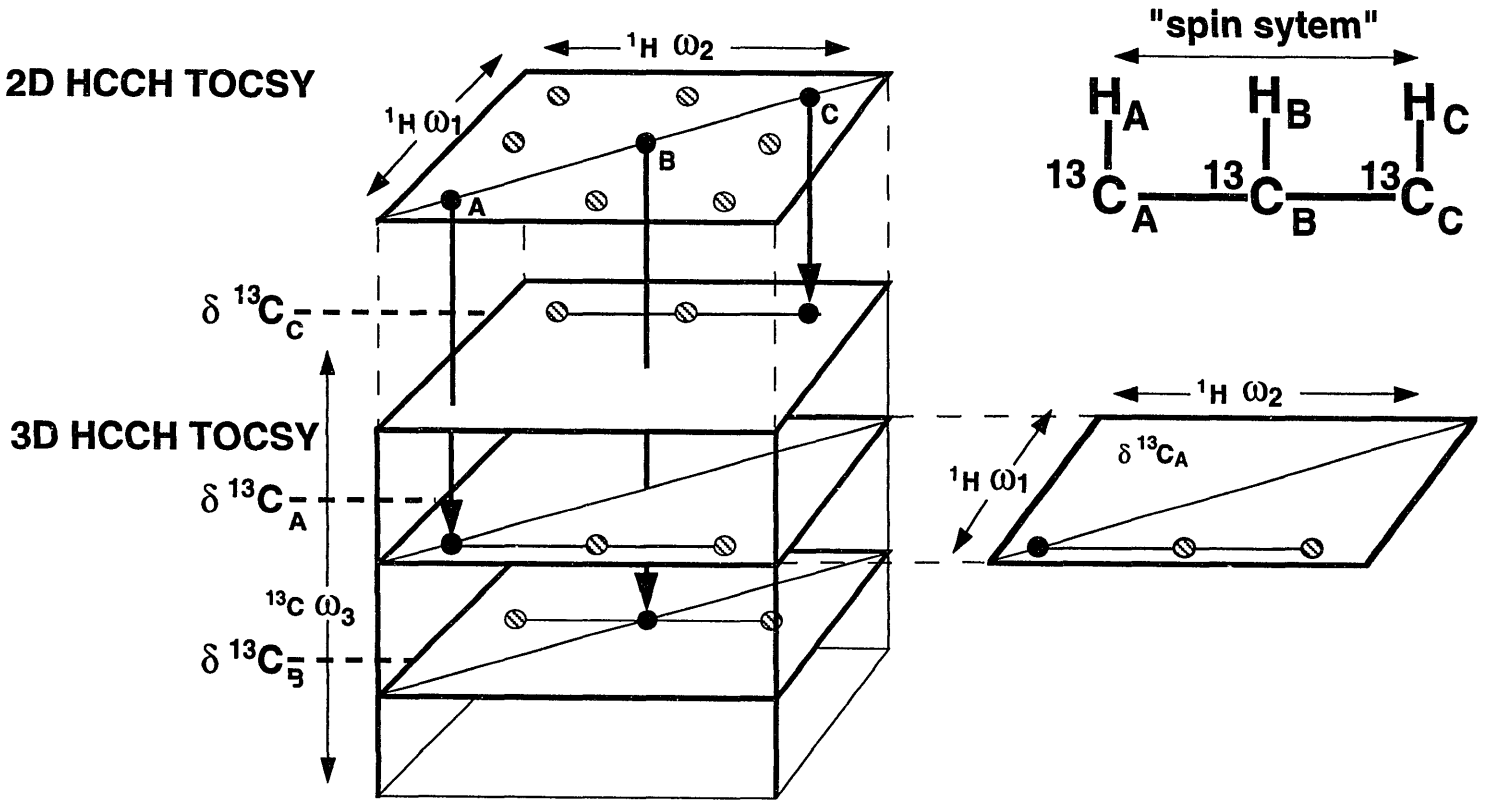


Figure 1.7: Schematic of a three dimensional (3D) NMR spectrum. The top plane shows a 2D TOCSY spectrum for a hypothetical molecule containing three protons A, B, and C. The bottom planes are a 3D TOCSY spectrum, where the 2D 1H-1H frequencies (ω1 and ω2) are spread into the third dimension (ω3) by the chemical shift of the carbon attached by one bond to the proton in ω1. "HCCH" refers to the mechanism of magnetization transfer through large one-bond coupling constants (see text).

1993a; Legault et al., 1994; Simorre et al., 1995; Fiala et al., 1996; Sklenar et al., 1996). The other class of triple resonance (^{1}H, ^{13}C, ^{31}P) experiments transfers magnetization through the phosphate backbone via heteronuclear $^{2}J_{CP}$ and $^{3}J_{CP}$ couplings (Heus et al., 1994; Marino et al., 1994; Marino et al., 1995; Varani et al., 1995; Wijmenga et al., 1995). The HCP experiments provide an NOE-independent method to sequentially assign RNA ribose protons. The base protons can then be assigned in an NOE-independent manner by through-bond correlation to the ribose protons with the HCN experiments.

One significant limitation for NMR structural determination of large oligonucleotides at this time is the difficulty in obtaining torsion restraints to define the conformation of the phosphate backbone. In the standard 2D experiments for determination of coupling constants, there is a large amount of cancellation of the antiphase ^{1}H-^{1}H and ^{1}H-^{31}P COSY correlations due to the increased linewidth of larger molecules such as studied here (~15 kDa). While 3D HCCH-COSY and HCCH-TOCSY experiments can aid in assignment of ribose protons, there is no coupling constant information in these spectra to define any of the torsions α-ζ in nucleic acids. Some new methods have recently been developed utilizing ECOSY type correlations which are less sensitive to linewidth problems (Hines et al., 1994; Schwalbe et al., 1994; Legault et al., 1995; Schwalbe et al., 1995). In addition, a semi-quantitative method for determining crude ranges of backbone torsions through analysis of peak intensity in HCP experiments has been proposed (Varani et al., 1995). This experiment has the caveat that relaxation of the nuclei can affect the intensity of the crosspeak besides just the coupling constant; therefore, very careful analysis is necessary to prevent misinterpretation. Overall, these methods still suffer from serious overlap problems for larger RNAs, particularly in A-form regions. None of these methods were used for structure determination of the RRE RNA, since the linewidths of the ^{13}C- and ^{31}P-resonances in the ~15 kDa bound RRE were too broad for obtaining high quality spectra.

1.3 RRE-Rev Interaction Background

1.3.1 HIV gene expression and the life-cycle HIV contains ~10-20 viral proteins that are produced from overlapping exons on a 10 kilobase (kB) RNA transcript (Figure 1.8). Early in the life cycle of the virus the small regulatory proteins (Rev, Tat, Vif, Nef, etc.) are produced from mRNAs generated by multiple splicing events on the 10kB transcript. Later in the life cycle, singly-spliced and unspliced mRNAs appear in the cytoplasm for expression of the structural proteins Gag, Pol, and Env, as well as for packing of the genomic RNA into virus particles. The Rev protein is essential for regulation of this process, and functional Rev protein is required for virus replication. Rev functions by binding to the Rev Response Element (RRE) RNA located within the *env* gene and inhibits mRNA splicing (Kjems et al., 1991; Malim and Cullen, 1993) and/or facilitates nuclear export of partially processed mRNAs (Ruhl et al., 1993; Bogerd et al., 1995; Fischer et al., 1995; Stutz et al., 1995). Two domains in Rev that are essential for these functions have been identified: an arginine-rich domain that mediates binding to the RRE (Olsen et al., 1990; Zapp et al., 1991), and a leucine-rich domain that is responsible for targeting Rev and the associated mRNA for nuclear export (Ruhl et al., 1993; Bogerd et al., 1995; Fischer et al., 1995; Stutz et al., 1995).

1.3.2 Biochemical analysis of high-affinity binding site Biochemical analyses led to the identification of a small RNA-peptide complex that is the model system for the Rev-RRE interaction used in this thesis. The Rev protein is a tetramer at concentrations >μM, and from two to eight Rev protein molecules are thought to bind to the entire RRE. However, mutational studies have identified a high affinity binding site for a single Rev protein, which is localized to a relatively small stem-loop (IIB) structure of the RRE (Figure 1.9) (Cook et al., 1991; Kjems et al., 1991; Tiley et al., 1992). In addition,

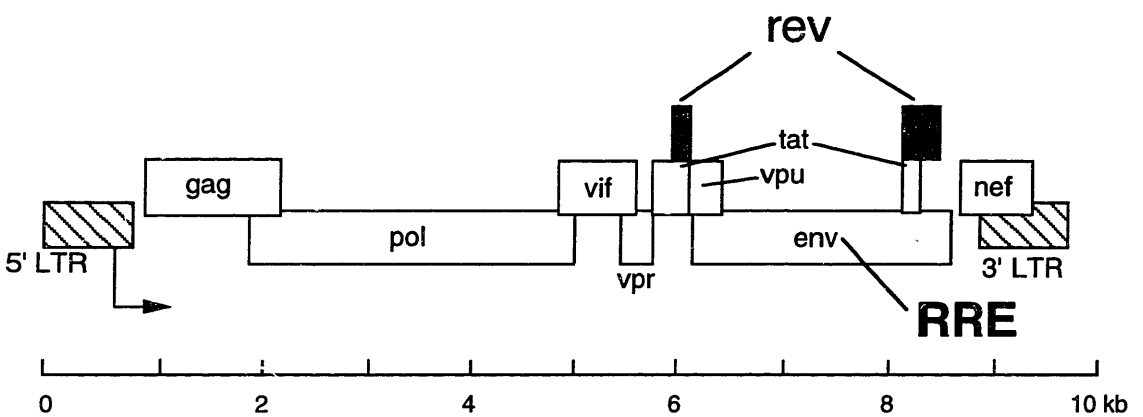


Figure 1.8: Open reading frames in HIV. The boxes represent the positions of genes on the genomic RNA. Single and multiple splicing events occur to produce all mRNAs shown above for expression of the gene products. The Rev Respone Element (RRE) RNA is within the opening reading frame of the *env* gene. The open reading frame for the Rev protein is highlighted in black.

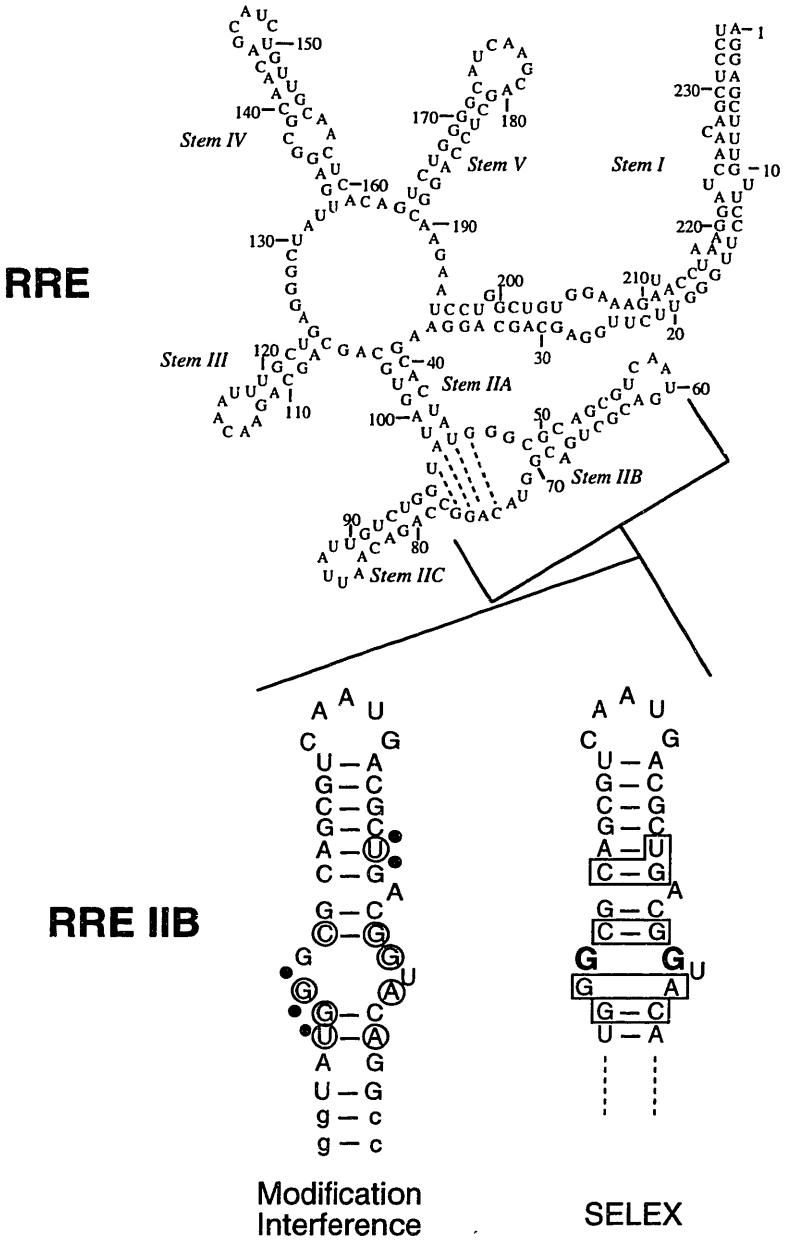


Figure 1.9: High affinity Rev binding site in the RRE. On top is the secondary structure prediction of the entire RRE RNA defined by mutational analysis. Potential alternative base pairs in stem II are highlighted by dashed lines. Below is the high-affinity binding site for a single Rev protein molecule in stem IIB of the RRE that was identified by several independent groups. The results of two different biochemical experiments are summarized showing the importance of the purine-rich internal loop for Rev binding. Nucleotides (open circles) or phosphates (closed circles) that interfere with Rev binding upon chemical modification are highlighted (Kjems *et al.*, 1992). Nucleotides that were invariant in RNA molecules selected for Rev binding in vitro are boxed (Bartel *et al.*, 1991). In bold are two guanosine nucleotides that covary to adenosine.

an in vitro selection study of Rev binding to the entire stem II portion of the RRE found a cluster of invariant nucleotides around a putative purine-rich internal loop in stem IIB (Figure 1.9) (Bartel et al., 1991). This and other selection studies (Giver et al., 1993; Jensen et al., 1994) identified a covariation of the G48 and G71 positions of the RRE to A48:A71 or C48:A71 which suggested the formation of a mismatch base pair that was critical for Rev recognition. Chemical modification interference studies of an isolated RRE stem IIB construct also identified nucleotides around the purine-rich internal loop to be important for Rev binding and also showed that the minimal stem IIB RNA was sufficient for a specific high-affinity interaction (~ 1-10 nM) (Kjems et al., 1992).

Mutational analyses of the Rev protein have identified three regions of the protein with four distinct functions (Figure 1.10) (Malim and Cullen, 1991). The RNA-binding and activation domains were mentioned previously, but Rev contains two additional "domains" for nuclear localization and multimerization (Olsen et al., 1990; Malim and Cullen, 1991; Zapp et al., 1991). Nuclear localization function is in the arginine-rich domain and overlaps with important RNA-binding amino acids. Sequences flanking the arginine-rich domain are important for multimerization of the Rev protein. Many mutants have been isolated that are multimerization deficient and do not function in vivo. Whether Rev initially binds the RRE RNA as a monomer or a tetramer, however, is somewhat of a contentious point. While it is clear that binding of multiple Rev molecules is necessary for complete in vivo function, several studies have identified specific monomeric binding to the RRE (Malim and Cullen, 1991; Cole et al., 1993). The argument may ultimately be a mechanistic one; however, identification of 1:1 Rev-RRE stem IIB complexes suggests that initial binding of Rev to the high-affinity site on the RRE nucleates formation of the multimeric ribonucleoprotein complex (Cook et al., 1991; Heaphy et al., 1991; Tiley et al., 1992). Lastly, a rigorous thermodynamic analysis of Rev multimerization by equilibrium analytical centrifugation showed that Rev was predominately monomeric at concentrations

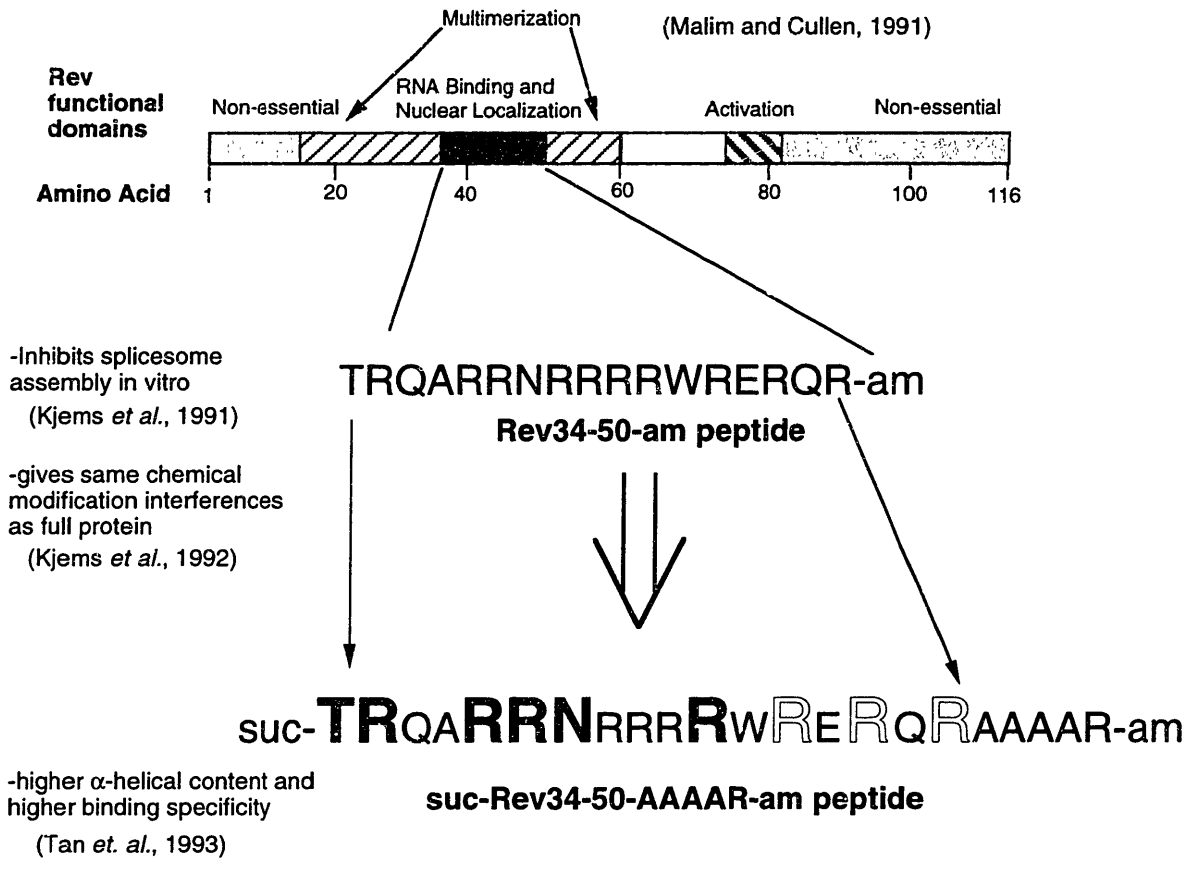


Figure 1.10: Minimal fragment of Rev protein necessary for RNA-binding. The top boxes show the functional domains of Rev identified by deletion and mutational analysis. Below are the peptide sequences for which biochemical analysis of the RRE-Rev interaction has been performed. suc and am represent succinyl and amide modification of the N- and C- terminus of the peptide, respectively. Bold and outlined letters indicate amino acids that give >10- and >2-fold, respectively, decrease in peptide binding specificity upon mutation to alanine.

(<10 nM) where specific binding to full length RRE was near saturation (Cole et al., 1993).

Further analysis showed that basic peptides from the RNA-binding domain of Rev could bind specifically to the stem IIB RNA (Kjems et al., 1992). The peptide gave the same chemical modification interferences as the intact protein, providing convincing evidence that the peptide was binding in a specific manner similar to the protein. Additional studies with Rev peptides modified at their termini showed a correlation between α-helicity and binding specificity, suggesting that the Rev protein recognizes the RRE with an α-helix (Tan et al., 1993). Due to molecular weight limitations of the NMR technique and oligomerization/aggregation of the full length protein at high concentrations, identification of the minimal RNA and protein elements was essential for structural analysis of the RRE-Rev interaction by NMR.

CHAPTER 2: MATERIALS AND METHODS

2.1 Preparation of NMR Samples

2.1.1 Unlabeled RNA The sequence of the RNA used for NMR analysis (see Figure 3.1) was 5'-GGUCUGGGCGCAGCGCAAGCUGACGGUACAGGCC-3'. This RNA was transcribed from synthetic DNA templates containing a double stranded promoter region and single stranded anti-sense template using T7 RNA polymerase (Wyatt et al., 1991). The sequences of the DNA template and promoter strand used to make this RNA were 5'-GGCCTGTACCGTCAGCTTGCGCTGCGCCCAGACCTATAGTGAGTCG-TATTA-3' and 5'-TAATACGACTCACTATAG-3', respectively. The milligram quantities of T7 RNA polymerase needed for transcription of enough RNA for NMR analysis were overexpressed and purified as described (Grodberg and Dunn, 1988). Transcription yields were optimized by varying reagent concentrations ±~50% around "standard" values (Wyatt et al., 1991). Optimization trials were done in 20 µl reactions with trace amounts of ^{32}P-labeled α-GTP, and product bands were detected by autoradiography. Conditions that gave the most intense product band (quantitated by eye) were chosen. Trials for each reagent were serially run in the following order: nucleotide triphosphate (NTP), magnesium chloride ($MgCl_2$), template/promoter DNA, and T7 RNA polymerase. The final buffer conditions used for RRE RNA transcriptions are given in Table 2.1.

The RNA from large scale (~50-100 ml) transcriptions was purified on preparative (45 x 35 x .3cm) 20% (19:1) polyacrylamide gels containing 8M Urea and 1X TBE buffer. Gels were typically run at 55 Watts until the xylene cyanol dye approached the bottom of the gel (~20-24 hours). The product band was visualized by UV absorption, cut from the gel, and recovered by electroelution in an Elutrap (S&S). Electroelution was run at 250 Volts for 2-3 hours, changing the product trap once roughly half way through. Completion

Table 2.1: Reagent concentrations for all RNAs made by transcription with T7 RNA polymerase. All reactions had a constant buffer of 80.0 mM HEPES, pH 8.1, 5.0 mM DTT, 1.0 mM Spermidine, 0.01% (v/v) Triton X-100, and 80 mg/ml PEG-8000. Note that for RRE, HEPES gave preferable yields to Tris buffer, which is more commonly used for T7 transcriptions.

Reagent	Unlabeled RNA	Uniformly ^{13}C-labeled RNA	Selectively ^{13}C-labeled RNA (guanosines)	Selectively ^{13}C-labeled RNA (cytosines)
Nucleotide Triphosphates	4 mM each NTP (16 mM total)	10 mM total	2 mM ATP 2 mM CTP 2 mM UTP 4 mM ^{13}C-GTP (10 mM total)	4 mM ATP 4 mM GTP 4 mM UTP 2 mM ^{13}C-CTP (14 mM total)
Magnesium Chloride	28 mM (1.75)[a]	16 mM (1.6)[a]	16 mM (1.6)[a]	19.6 mM (1.4)[a]
Template/ Promoter DNA	300 nM each	400 nM each	400 nM each	400 nM each
T7 RNA Polymerase (mg/ml)	0.03 mg/ml	0.06 mg/ml	0.075 mg/ml	0.05 mg/ml

[a]value in parantheses is the ratio of $[Mg^{2+}]$/[total NTP], which is the value that was optimized.

of elution was monitored by UV shadowing of gel slices. Pooled RNA was ethanol precipitated and resuspended in 250-400 µl of H_2O and dialyzed 24-48 hours in a microdialysis chamber (BRL) against 500-1000 ml of NMR buffer (10mM sodium phosphate, pH 6.5, 50mM sodium chloride, and 0.1 mM EDTA), which was usually changed once during the dialysis. An 80-ml transcription reaction yielded 0.5 µmole of purified RNA (~1mM sample in 500 µl NMR buffer).

2.1.2 Labeled RNA

Preparation and separation of ^{13}C-labeled nucleotides Uniformly ^{13}C-enriched NTPs can be used as reagents for T7 RNA transcription to make isotopically labeled RNA samples for heteronuclear NMR analysis. The methodology of Batey et al. (Batey et al., 1992) was used to obtain ^{13}C-labeled NTPs, which are isolated from *Methylophilus methylotrophus* bacteria grown on 99.9% ^{13}C-methanol. A detailed account of the procedures for obtaining ^{13}C-RNA from this approach is given in (Batey et al., 1995). Briefly, cells harvested from a 10-L fermentor were lysed with 0.5% SDS, extracted with phenol/chloroform to separate proteins and nucleic acids, which were then digested to nucleotide monophosphates with P1 nuclease. The ribonucleotide monophosphates (NMPs) and deoxyribonucleotide monophosphates (dNMPs) were separated by boronate affinity chromatography, which selectively binds the cis-diols of the ribose ring. The NMPs were enzymatically converted to NTPs and purified for transcriptions with another boronate affinity column.

For preparation of ^{13}C-RNA selectively labeled with only one nucleotide, the individual ^{13}C-labeled NMPs were first separated by high performance liquid chromatography (HPLC) on a 7.8 x 100 mm HP-PEI anion exchange column (Interaction Chemicals, Inc.) before conversion to NTPs. The flow rate was 1 ml/min and the nucleotides were monitored by UV absorbance at 268 or 300 nm. Buffer A was 50 mM

ammonium formate pH 3.0 and buffer B was 500 mM ammonium formate pH 2.5. Isocratic elution with 100% buffer A for 20 minutes was followed by a linear gradient to 100% buffer B over 20 minutes. The retention times were 8, 16, 33, and 42 minutes for CMP, AMP, UMP, and GMP, respectively (Figure 2.1). Up to 60 mg of nucleotides could be injected while maintaining baseline resolution. Pooled fractions were lyophilized to dryness. GMP was enzymatically converted to GTP following the standard procedure (Batey et al., 1992), except that nucleoside monophosphate kinase was omitted, the concentration of guanylate kinase was increased to .05 units/ml, and the catalytic amount of unlabeled ATP was increased to 0.3 mM. Similarly, CMP was enzymatically converted to CTP with the standard conditions minus guanylate kinase, increased quantities of nucleoside monophosphate kinase, and 0.3 mM unlabeled ATP. A boronate affinity column was then run to purify and desalt the individually charged ^{13}C-NTPs.

Transcription of labeled samples Uniformly ^{13}C-labeled RNA was transcribed with ^{13}C-NTPs prepared as described above (not separated into the individual NTPs). Transcription trials were performed in an analogous fashion to the unlabeled RNA, except that the criterion for "optimal" conditions was altered slightly. To minimize costs, the yield per unit of ^{13}C-NTP was maximized rather than just total yield. For example, if increasing the concentration of ^{13}C-NTP 4-fold only results in a 2-fold increase in yield, it would be cheaper to keep the ^{13}C-NTP concentration low and increase the volume of the reaction 2-fold (same yield, but half the amount of NTP). This rationale assumes that the cost of the ^{13}C-NTPs is far greater than the cost of other reagents in the reaction. For these transcription trials, the product band was quantitated by phosphoimaging of the acrylamide gel. With the conditions given in Table 2.1, a 150-ml transcription reaction yielded 0.75 μmole of purified (as described above) ^{13}C-RRE RNA (~1.5 mM NMR sample).

To make the selectively ^{13}C-labeled RRE RNA at guanosine or cytosine, ^{13}C-GTP or ^{13}C-CTP was simply mixed with the other three unlabeled NTPs in the transcription

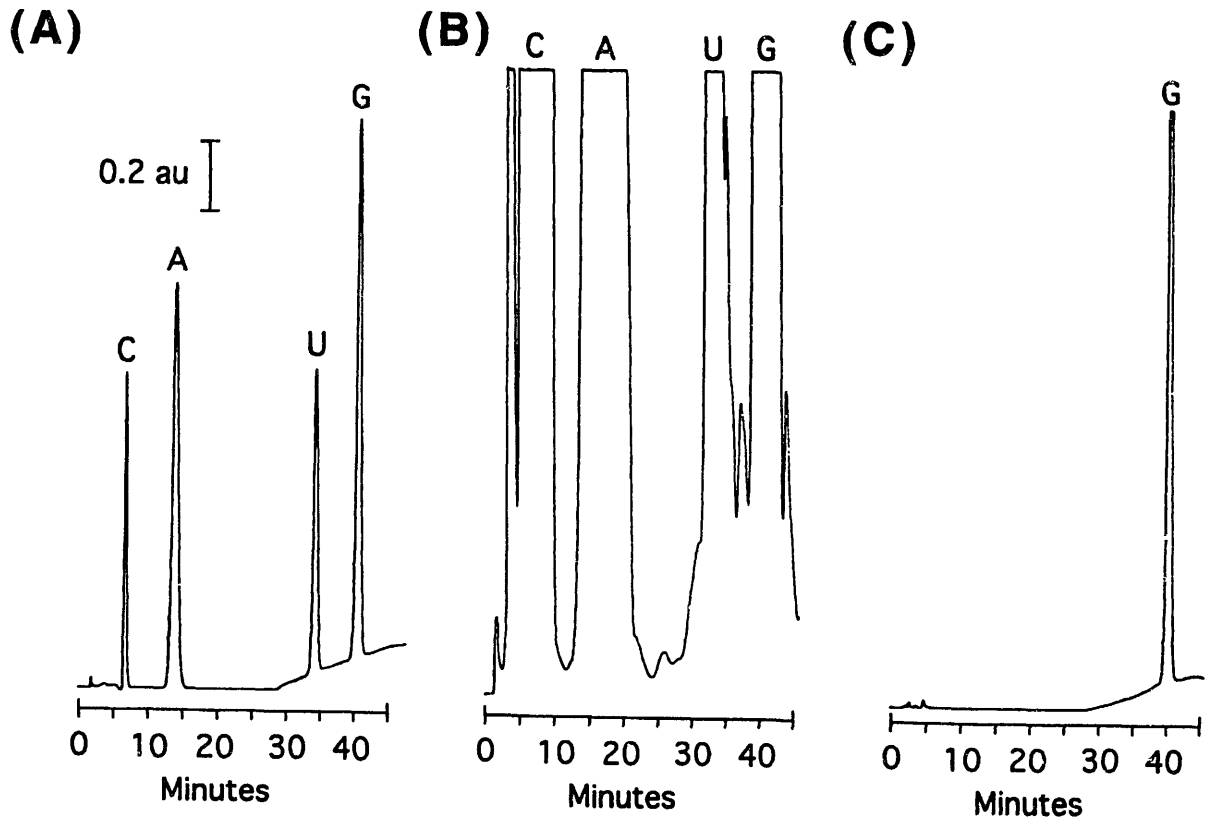


Figure 2.1: separation of NTPs by HPLC. **(A)** Chromatogram of 100 µg of NTPs injected onto a PEI anion-exchange column. Peaks corresponding to four nucleotides are labeled (detected by UV absorption at 268nm). **(B)** Injection of ~60 mg of NTPs showing baseline resolution of the four nucleotides (detected at 300nm). **(C)** GTP isolated from part b reinjected onto column showing purity of isolated nucleotides (detected at 268 nm).

reaction. Again, for transcription trials the yield per unit of ^{13}C-GTP or ^{13}C-CTP was maximized (Table 2.1). With 50 ml and 80 ml transcriptions, 0.6 and 0.9 μmole of purified (as described above) ^{13}C-guanosine or ^{13}C-cytosine RRE RNA, respectively, was obtained. This gave 1.2 and 1.5 mM samples of ^{13}C-guanosine and ^{13}C-cytosine RRE RNA in 500 and 600 μl of NMR buffer, respectively.

2.1.3 Unlabeled and labeled peptides The peptides Rev_{34-50} (TRQARRNR-RRRWRERQR) and suc-Rev_{34-50}-AAAAR-am (TRQARRNRRRRWRERQRAAAAR) were made synthetically with the latter peptide containing succinyl and amide blocking groups at the N- and C-terminus, respectively. These peptides were synthesized and purified by our collaborators Ruoying Tan and Alan Frankel as described (Tan et al., 1993).

To obtain uniform isotopic (^{13}C,^{15}N) enrichment for heteronuclear NMR analysis, the peptide D-Rev_{34-50}-AAAAR (DTRQARRNRRRRWRERQRAAAAAR) was expressed as a C-terminal fusion with a modified, His-tagged TrpLE leader polypeptide (Figure 2.2) (Staley and Kim, 1994; Schumacher et al., 1996). The modified TrpLE protein has several features that aid in easy purification: 1) It contains a stretch of histidines at its N-terminus, which permits affinity chromatography on a Ni^{2+} column. 2) It readily forms inclusion bodies, which increases cellular yields of the protein, and isolation of the insoluble portion of the cell provides clean crude protein. 3) It contains only one methionine immediately N-terminal to the Rev sequence, so that only two easily separable products are produced upon cleavage with cyanogen bromide.

Deoxyoligonucleotides containing the codons for the D-Rev_{34-50}-AAAAR sequence with flanking restriction sites were inserted into the Hind III-Bam HI sites of the pMM vector, and transformed into *E. coli* JM109. The pMM vector contains the TrpLE protein under the control of a T7 RNA polymerase promoter and an ampicillin resistance gene

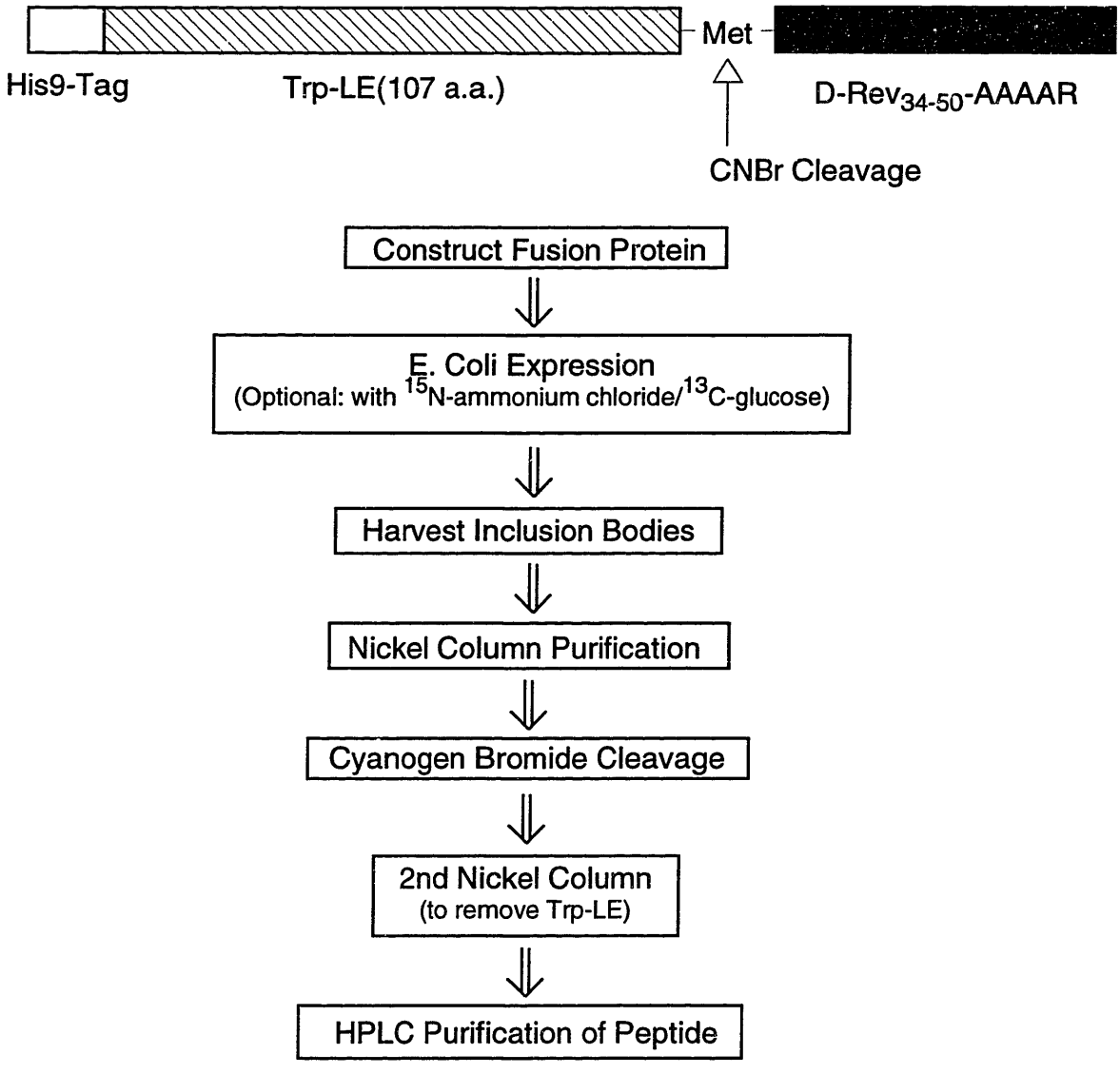


Figure 2.2: Schematic and outline of peptide purification from a Trp-LE protein fusion

marker. Plasmid from transfected colonies was sequenced to ensure proper insertion of the Rev coding sequence. The plasmid was then transfected into *E. coli* BL21, which contains T7 RNA polymerase on the bacterial chromosome, for expression of the protein.

Expression of unlabeled fusion protein in rich media (LB) was induced at 0.4 OD_{600} by addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After an additional two hours of growth (37°C) the cells were harvested by centrifugation at 6000g for 10 minutes. The cells were resuspended in 25% sucrose, 50 mM Tris, pH 8.0, 0.1 mM EDTA (50 ml per 1-L culture) and sonicated for 2 minutes on ice with stirring (30 second intervals every 3 minutes). 50 λ of 10 ng/ml DNase I can be added at this stage to reduce viscosity. The suspension was centrifuged for 20 minutes at 35000g, and the supernatant was discarded. The pellet was resuspended in 50 ml of 20 mM Tris, pH 8.0, 1% Triton X-100, 1 mM EDTA and the sonication was repeated as described above. This procedure solubilizes most of the cellular proteins, except the inclusion bodies containing the fusion protein, resulting in a simple, large-scale purification. The isolated inclusion body pellet was solubilized with 8M urea, 20mM sodium phosphate, pH7.8, 500mM sodium chloride, loaded onto a Ni^{2+}-column (Invitrogen ProBond), and washed with 2-3 column volumes of urea buffer. The protein was then eluted with 8M urea, 20 mM sodium phosphate, pH 4.0, 500 mM sodium chloride. Fractions containing the fusion protein were pooled and dialyzed against 5 liters of 2% acetic acid (1000 molecular weight cut off dialysis tubing). The purified protein was lyophilized to dryness and resuspended in ~5-10 ml of 70% formic acid. The Rev peptide was liberated from the fusion by cleavage at the methionine immediately N-terminal to the Rev peptide with cyanogen bromide (~100 mg/ml) (Staley and Kim, 1994; Schumacher et al., 1996). After two hours, the reaction was partially lyophilized to remove cyanogen bromide. Note that cyanogen bromide is extremely dangerous and all procedures were carefully done in a hood (including lyophilization)!! There are three main products after cleavage: uncleaved fusion protein,

N-terminal Trp-LE, and Rev peptide. The Rev peptide was isolated by neutralization of the formic acid solution to pH 7.0 with ammonium hydroxide, which precipitates the TrpLE cleavage product as well as any uncleaved fusion protein, but not the Rev peptide. The suspension was centrifuged and the peptide (supernatant) was further purified by reverse phase HPLC with 0.1% TFA in H_2O/acetonitrile and concentrated by lyophilization. The peptide composition was checked by amino acid analysis and quantitated by tryptophan absorbance at 280 nm. Typical yields were ~10-15 mg of peptide per liter of culture.

For expression of isotopically labeled peptides, the D-Rev$_{34-50}$-AAAAR sequence was subcloned into the Hind III-Bam HI sites of the pTM vector, which is similar to the pMM vector with substitution of the kanamycin resistance gene for the ampicillin resistance gene (M.A. Milhollen and P.S. Kim, unpublished results). In minimal media, loss of the ampicillin resistance plasmids was significant due to slow growth and secretion of β-lactamase (degrades ampicillin) into the media (Studier et al., 1990). ^{15}N- or ^{13}C,^{15}N-labeled peptides were obtained by expression of the fusion protein in M9 minimal media supplemented with either 0.7 g/L ^{15}N-ammonium sulfate and/or 1.0 g/L ^{13}C-glucose (Cambridge Isotopes). Cells were grown for four hours after induction with IPTG, and the isotopically-labeled peptides were purified as described above, except that the sonication of the cells was done only once with 1X STE buffer. The more exhaustive sonication described above resulted in the solubilization of the smaller inclusion bodies obtained from expression in minimal media. Typical yields were ~3-5 mg of peptide per liter of culture.

2.1.4 Formation of 1:1 complexes As a typical example, a 1:1 complex of RRE RNA and suc-Rev$_{34-50}$AAAAR-am peptide was formed by adding 25 µl of 20mM peptide solution in 5 µl increments into the 500 µl RNA NMR sample. Complex formation was monitored by 1D NMR of the imino proton resonances. Since the free RNA was in slow

exchange with peptide-bound RNA (see section 3.1.2), two sets of NMR resonances (free and bound) were observed. Disappearance of the free RNA imino proton resonances was used as a more precise determination of 1:1 stoichiometry (±5%), than quantitation of the RNA and peptide concentrations by UV spectroscopy (±20%). This procedure of titrating a concentrated peptide stock into a ~500 µl RNA NMR sample was used for preparation of complexes containing all variations of isotopically labeled samples in this thesis (both RNA and peptide). A total of nine samples (two unlabeled and seven isotopically labeled) were used for NMR analysis: 1) unlabeled RNA (suc-Rev_{34-50}AAAAR-am); 2) unlabeled RNA (D-Rev_{34-50}-AAAAR); 3) ^{15}N-labeled RNA (D-Rev_{34-50}-AAAAR); 4) ^{13}C-labeled RNA (suc-Rev_{34-50}AAAAR-am); 5) ^{13}C-labeled RNA (D-Rev_{34-50}-AAAAR); 6) Specifically ^{13}C-guanosine labeled RNA (suc-Rev_{34-50}AAAAR-am); 7) Specifically ^{13}C-cytosine labeled RNA (D-Rev_{34-50}-AAAAR); 8) ^{15}N-labeled D-Rev_{34-50}-AAAAR peptide (RRE) and 9) ^{13}C/^{15}N-labeled D-Rev_{34-50}-AAAAR peptide (RRE). All unlabeled and RNA-labeled samples were dissolved in or dialyzed against NMR buffer (10mM $NaPO_4$, pH 6.5, 50mM NaCl, and 0.1mM EDTA) . Peptide-labeled samples had identical buffer conditions, except for a pH of 5.5. All experiments with ^{13}C-labeled RNA were performed in 99.996% D_2O, while all experiments with ^{15}N-labeled RNA or labeled peptide were performed in 90/10% H_2O/D_2O. For preparation of samples in 99.996% D_2O, samples were repetitively (3 times) lyophilized to dryness and dissolved in 99.9% D_2O. After the final drying, samples were dissolved in 99.996% D_2O (Isotec) under nitrogen or argon in a glove box.

2.2 NMR Experiments

Almost 100 multi-dimensional NMR experiments have been performed for this thesis project over the course of four years. Not every experiment will be described in this

thesis for conceptual simplicity, yet organization of the experiments that will be described is still a difficult task for a couple of reasons. The instrumentation and methodology has changed dramatically over the course of four years. As a result, experiments that are conceptually similar, yet separated by a large gap in time, may have vastly different intricacies in the details of their execution. Nevertheless, the experiments will still be grouped in a conceptual manner below with an effort to make clear the differences in the instrumentation and pulse sequences used. In addition, two complexes were used with slightly different peptides (suc-Rev_{34-50}AAAAR-am and D-Rev_{34-50}-AAAAR; see section 3.1.2). D-Rev_{34-50}-AAAAR came chronologically later in the project when it was realized that isotopic labeling of the peptide was necessary. To ensure that the RNA structure was the same in each complex, many experiments with labeled RNA were reperformed, which may seem redundant and unnecessary in the manner that the experiments will be described below. To clarify this subject, each section will be segregated into two sub-sections (one for each peptide) where it seems appropriate. Lastly, the biggest change in the execution of NMR experiments during this thesis project was the incorporation of pulse field gradient technology. Application of field gradients to many existing experiments enhances performance and spectral quality. For brevity, little explanation of the advantages of field gradients will be provided in this thesis (for a review of field gradient methodology, see (Kay, 1995)), and experiments that used field gradients will be simply be designated with the prefix G- (e.g. G-HSQC for a version of an HSQC utilizing field gradients). Pulse sequences diagrams of many of the sequences described here, including those with field gradients, are given in Appendix-A with brief descriptions of their function.

2.2.1 Homonuclear NMR All NMR spectra were recorded at 25°C on a Varian VXR-500 MHz instrument with an 5mm Varian inverse detection probe, unless noted otherwise. 1D spectra of RRE RNA in 10/90% D_2O/H_2O NMR buffer were recorded using a 1331

binomial solvent suppression pulse sequence (Hore, 1983) with a sweep width of 12000 Hz, 8K complex data points, and 64 scans. The excitation maximum was set to the imino proton region of the spectrum (3500 Hz). 1D NOE difference spectroscopy was performed on the free RRE sample in a interleaved fashion with 512 scans per imino resonance. Each imino resonance was irradiated for .6s at the lowest power required to achieve 90% saturation to minimize spillover artifacts.

NOESY, DQF-COSY, and TOCSY experiments were also recorded for the free RRE RNA in 99.996% D_2O NMR buffer (Varani and Tinoco, 1991). Data sets with 2048 complex points in t_2 and 512 complex points in t_1 were acquired with 5000 Hz sweep widths in both dimensions and 32 scans per slice. Four NOESY spectra were acquired with mixing times of 50, 100, 200, and 400 ms and a recycle delay of 1.2s. An additional NOESY spectra at 35°C with a mixing time of 400 ms was also acquired. The TOCSY spectrum was recorded with a 125ms MLEV spin lock pulse (Bax and Davis, 1985a), and a recycle delay of 1.8s. The DQF-COSY was recorded with WALTZ decoupling of ^{31}P during acquisition and a recycle delay of 1.9s. All spectra were processed with a combination of exponential and gaussian weighting functions and zero-filled to 4K x 4K data points.

NOESY experiments of the RRE-suc-Rev_{34-50}AAAAR-am complex in 10/90% D_2O/H_2O NMR buffer were acquired using a Z-SPEC MID500-3 indirect detection 3mm probe. For experiments with this probe, the sample was lyophilized, resuspended in 100µl H_2O, redialyzed against 500 ml NMR buffer, and diluted to final volume of 175µl with NMR buffer. The NOESY experiment was acquired using a jump-return echo pulse sequence for water suppression with the excitation maximum set to the imino resonances (~3500 Hz) (Sklenar et al., 1987). The spectrum was 2048 x 256 complex data points with a sweep width of 12000 Hz, a mixing time of 250ms, a recycle delay of 1.7 seconds, and 128 scans per slice. 10ms homospoil pulses were applied during the recycle delay and

mixing time. Spectra were processed with either exponential weighting functions , or a skewed, shifted sinebell function to resolve overlapped imino protons. In addition, 1D NOE difference spectroscopy on the complex was performed as described above for the free RNA with 1536 scans per imino resonance .

2D NMR spectra (NOESY, ^{31}P-decoupled DQF-COSY, and TOCSY experiments) of the RRE-suc-Rev$_{34-50}$AAAAR-am complex in 99.996% D_2O NMR buffer were taken with a 5mm probe as described above for the free RNA. A total of four NOESY experiments were acquired with mixing times of 50, 100, 200, and 400ms to monitor NOE buildups. NOESY spectra (400 ms) at 15°C and 30°C were also acquired to help resolve overlapped crosspeaks. A ROESY experiment was performed to identify potential conformational exchange peaks. The bandwidth of the spin-lock mixing time (100 ms) was 3 kHz, and the flip angle used was 30° to help remove TOCSY artifacts (Kessler et al., 1987). The sweep width in both dimensions of the ROESY was 5000 Hz with 2048 complex points in acquisition and 256 t_1 increments. 48 scans per increment were taken with a relaxation delay of 2 seconds. Additional NOESY experiments were taken on a Varian 750 MHz instrument with mixing times of 100, 200, and 400 ms. These spectra were 4096(t_2) by 1024(t_1) complex data points with a sweep width of 9000 Hz, a recycle delay of 1.5 seconds, and 32 scans per slice.

For the RRE RNA-D-Rev$_{34-50}$-AAAAR peptide complex in 10/90% D_2O/H_2O NMR buffer, NOESY spectra at mixing times of 50 and 100 ms were acquired at 750 MHz using a "3-9-19" WATERGATE solvent suppression scheme (Sklenar et al., 1993b). The spectra were 2048(t_2) by 1024(t_1) complex data points with a sweep width of 15000 Hz, a recycle delay of 1.7 seconds, and 8 scans per slice. NOESY (400 ms) and DQF-COSY experiments in 99.996 D_2O were also taken on the Varian VXR-500 as described above for the RRE-suc-Rev$_{34-50}$AAAAR-am complex.

2.2.2 Heteronuclear RNA NMR Table 2.2 lists the acquisition parameters for the heteronuclear NMR experiments performed on all labeled RRE RNAs complexed with unlabeled Rev peptide. For all experiments on the Varian VXR-500, ^{15}N- or ^{13}C-decoupling was achieved with a WALTZ pulse train (Shaka et al., 1983) during acquisition and a 180° ^{15}N- or ^{13}C-pulse in the middle of the t_1 proton evolution times. On the instruments at the Francis Bitter National Magnet Laboratory (500 and 600 MHz) or the Varian 750 MHz, WALTZ (Shaka et al., 1983) and GARP (Shaka et al., 1985) were used for ^{15}N- and ^{13}C-decoupling, respectively, during acquisition. A 180° ^{15}N- or ^{13}C-pulse in the middle of the t_1 proton evolution times was used for decoupling. The ^{13}C-transmitter frequency was set to ~80 ppm, between the C4' and C2'/C3' resonances, for all experiments focusing on coherence transfer in the ribose ring. For the HSQCs on the aromatic protons, the ^{13}C-transmitter frequency was set to ~130 ppm. For all NOESY experiments the transmitter was set to ~105 ppm, just downfield of the C5 resonances. For ^{15}N-HSQC and NOESY-HSQC experiments, the transmitter was set to ~90 ppm near the C-amino nitrogens. Both of the ^{15}N experiments use a modified version of the WATERGATE solvent suppression scheme that minimizes intensity losses due to saturation of H_2O (Mori et al., 1995). For the ^{13}C-NOESY experiments, the INEPT delay was set using an average of the aromatic and ribose one bond carbon-proton coupling constants (160 Hz). The relaxation delay was between 1.6 and 2.0 seconds for all experiments. The total acquisition times were 3.0-3.5 days for 3D experiments and 18-24 hours for 2D experiments. Phase cycling for all experiments was performed as described in the references. Quadrature detection was achieved using the TPPI-States method for all experiments except the double-half-filtered NOESY which used the States method (States et al., 1982; Marion et al., 1989). In general the 3D experiments were processed with exponential/gaussian weighting functions in first dimension and sinebell functions in other

Table 2.2: Data acquisition parameters for heteronuclear NMR experiments

Molecule	Experiment	F_3[a]	F_2[a]	F_1[a]	j_{ch}[b]	scans	other	Ref
uniformly ^{13}C-RRE/suc-Rev$_{34\text{-}50}$-AAAAR-am	NOESY-HMQC (Varian 500)	512 pts 4000 Hz ^{1}H	64 pts 5000 Hz ^{13}C	64 pts 4000 Hz ^{1}H	160 Hz	8	250ms[c]	(Clore et al., 1990)
	HCCH-DIPSI (Varian 500)	512 pts 2500 Hz ^{1}H	40 pts 5000 Hz ^{13}C	64 pts 2000 Hz ^{1}H	142 Hz	16	18ms[d]	(Clore et al., 1990)
	HCCH-COSY (Varian 500)	512 pts 2500 Hz ^{1}H	32 pts 5000 Hz ^{13}C	64 pts 2500 ^{1}H	142 Hz	16		(Clore et al., 1990)
	HSQC-CT (Varian 500)	--	1024 pts 5500 Hz ^{1}H	112 pts 5000 Hz ^{13}C	142 Hz	32	40 Hz[e]	(Santoro and King, 1992)
uniformly ^{13}C-RRE/D-Rev$_{34\text{-}50}$-AAAAR	G-HCCH-DIPSI (Bitter 600)	512 pts 4000 Hz ^{1}H	32 pts 8000 Hz ^{13}C	64 pts 4000 Hz ^{1}H	150 Hz	8	18ms[d]	(Constantine, et al., 1993)
	G-NOESY-HSQC (Bitter 600)	512 pts 6000 Hz ^{1}H	32 pts 6000 Hz ^{13}C	128 pts 6000 Hz ^{1}H	165 Hz	8	150ms[c]	(Clore et al., 1990; Vuister and Bax, 1993)
	G-double-half-filtered NOESY (Varian 750)	--	1024 pts 7000 Hz ^{1}H	512 pts 7000 Hz ^{1}H	170 Hz	8	200ms[c]	Appendix A
	G-HMQC-NOESY-HSQC (Bitter 500)	15 pts 3200 Hz 13C	64 pts 4500 Hz ^{1}H	16 pts 3200 Hz ^{13}C	160 Hz	2	150ms[c] w4-^{1}H 256 pts 4500 Hz	(Vuister and Bax, 1993)

Table 2.2(cont): Data acquisition parameters for heteronuclear NMR experiments

Molecule	Experiment	F_3[a]	F_2[a]	F_1[a]	j_{ch}[b]	scans	other	Ref
^{13}C-guanosine RRE/suc-Rev$_{34-50}$-AAAAR-am	NOESY-HMQC	512 pts 5000 Hz ^{1}H	32 pts 6000 Hz ^{13}C	128 pts 5000 Hz ^{1}H	160 Hz	8	250ms[c]	(Clore et al., 1990)
	HSQC-CT (Varian 500)	--	1024 pts 5000 Hz ^{1}H	128 pts 6000 Hz ^{13}C	142 Hz	64	40 Hz[e]	(Santoro and King, 1992)
	double-half-filtered NOESY (Varian 500)	--	1024 pts 5500 Hz ^{1}H	256 pts 5500 Hz ^{1}H	160 Hz	16	300ms[c]	(Otting and Wuthrich, 1989; Otting and Wuthrich, 1990)
^{13}C-cytosine RRE/D-Rev$_{34-50}$-AAAAR	G-NOESY-HSQC (Bitter 600)	512 pts 6000 Hz ^{1}H	28 pts 6000 Hz ^{13}C	128 pts 6000 Hz ^{1}H	160	8	200ms[c]	(Clore et al., 1990; Vuister and Bax, 1993)
	G-double-half-filtered NOESY (Varian 750)	--	1024 pts 7000 Hz ^{1}H	512 pts 7000 Hz ^{1}H	160 Hz	4-8	70,130, 200ms[c]	Appendix A
uniformly ^{15}N-RRE/D-Rev$_{34-50}$-AAAAR	G-HSQC (Bitter 600)	--	1024 pts 13000 Hz ^{1}H	64 pts 3000Hz ^{15}N	90 Hz	4	15°C	(Mori et al., 1995)
	G-NOESY-HSQC (Varian 750)	1024 pts 18000 Hz ^{1}H	16pts 3000Hz ^{15}N	64 pts 16502 Hz ^{1}H	90 Hz	32	150ms[c]	(Mori et al., 1995)

(a) listed for each dimension are the number of complex data points acquired, the spectral width, and the nucleus detected (b) given is the value of the one bond carbon-hydrogen coupling constant used for determination of the delay for transfer of magnetization between proton and carbon. For the NOESY experiments the value used was an average of the aromatic and ribose coupling constants. (c) The mixing time of the period for NOE transfer (d) The mixing time of the carbon spin lock pulse for transfer of magnetization through the ribose ring. (e) Value of the constant time interval for evolution of carbon chemical shift in F_1.

two. In addition, linear prediction of 2X number of data points was usually performed in the 2nd and 3rd dimensions.

2.2.3 Heteronuclear peptide NMR While some of the experiments with ^{15}N-labeled peptide were done in our laboratory to assign the backbone resonances of the peptide (H. Mao and J. R. Williamson, unpublished results), the complete assignment of the peptide (particularly the arginines) and final spectra used for NOE analysis were done in collaboration with Sambasiva Rao and Lewis Kay of the University of Toronto. Therefore, only a brief description of the experimental detail will be given. Published H(C)(CO)NH-TOCSY (Montelione et al., 1992), (H)C(CO)NH-TOCSY (Logan et al., 1992), and CBCANH (Wittekind and Mueller, 1993; Muhandiram and Kay, 1994) through-bond correlation experiments, in combination with novel Arg-C(CC)-TOCSY-Nepsilon-Hepsilon and Arg-H(CC)-TOCSY-Nepsilon-Hepsilon experiments (R. Sambasiva & L.E. Kay, unpublished results), were performed on the $^{13}C,^{15}N$-labeled Rev peptide sample to assign all of the resonances. The novel arginine experiments use heteronuclear cross-polarization to transfer magnetization from all protons and carbons of the arginine side-chain to the epsilon protons, which are well resolved in the RRE-Rev complex.

A CN-NOESY-HSQC (Pascal et al., 1994) with a mixing time of 150ms was performed to simultaneously obtain NOEs for all $^{13}C,^{15}N$-attached protons of the peptide. A $^{13}C,^{15}N$ purge sequence (Ikura and Bax, 1992) was added prior to F1 of the CN-NOESY-HSQC experiment (F1-filtered,F3-edited-CN-NOESY-HSQC) to identify RNA-peptide NOEs. To obtain ϕ torsion restraints for the peptide backbone, a HNHA experiment (Vuister and Bax, 1993) was performed on ^{15}N-labeled Rev peptide sample.

2.3 Molecular Modeling

2.3.1 Assignment of restraint bounds

Hydrogen bonds: The hydrogen bonding patterns of the base pairs, including the purine-purine base pairs in the internal loop, were determined from analysis of imino proton spectra in H_2O. The presence of imino-proton resonances protected from exchange with water and downfield shifted (>12 ppm) is convincing evidence of hydrogen bond formation. Single hydrogen bonds were included as two restraints for the proton-hydrogen bond acceptor and heavy atom donor-acceptor distances using upper and lower bounds of 1.70-2.20 and 2.7-3.2 Å, respectively, which is the range of hydrogen bond distances observed in nucleotides (Saenger, 1984). Hydrogen bond restraints for the peptide backbone were also included in modeling. Unlike imino protons, protein backbone amide protons are typically observed without hydrogen bonding. Therefore, the hydrogen bonds for the peptide backbone were only added at a later stage of modeling refinement, when the secondary structure was convincingly established.

NOE distance restraints: Distance restraints involving nonexchangeable RNA protons were derived from visual inspection of crosspeak intensities in 50, 100, and 200ms NOESY experiments in 99.996% D_2O. The following criteria were used for classification of NOEs into three distance bound ranges: 1) strong 1.8 - 3.0 Å, medium to strong intensity NOEs at 50ms 2) medium 1.8 - 4.0 Å, all additional NOEs present at 100ms 3) weak 1.8 - 5.0 Å, all additional NOEs present at 200 ms. Note that lower bounds of 1.8 Å (~van der Waals radii) were used for all distance restraint ranges. This is a conservative approach to avoid inaccuracies from the loss of intensity of cross peaks due to relaxation or conformational exchange broadening. NOEs involving exchangeable proton resonances were classified as strong, medium, or weak based on visual inspection of intensities from a NOESY in H_2O, mixing time 250 ms. NOEs resolved only in the 3D NOESY experiments were classified

into the same ranges by visual inspection of crosspeak intensities at longer mixing times (150-250ms). Similarly, peptide NOEs were quantitated from the 3D CN-NOESY-HSQC experiment at a mixing time of 150ms. Where not resolvable in the standard 2D and 3D experiments, RNA-peptide NOEs were quantitated from the F1-filtered,F3-edited-CN-NOESY-HSQC at a mixing time of 300ms. For weak crosspeaks that could only be identified in 3D experiments at longer mixing times, a fourth distance bound range "very weak" was used (1.8 - 6.0 Å). This is too account for any effects of spin diffusion at the longer mixing times. For NOEs to non-stereospecifically assigned methylene or methyl protons, the distance was restrained to a pseudoatom representing an average of the proton positions. For every pseudoatom in a distance restraint, an approximate geometric correction factor of 1 Å was added to the upper bound.

Torsion restraints: The sugar pucker was estimated from analysis of the H1'-H2' coupling constants in the ^{31}P-decoupled DQF-COSY spectrum. Nucleotides with a H1'-H2' coupling constant of > 8 Hz in the COSY spectrum were classified as C2'-endo. Nucleotides with no COSY and TOCSY crosspeaks between the H1'-H2' protons (J < 3 Hz) were classified as C3'-endo. Some nucleotides had weak H1'-H2' crosspeaks in the TOCSY spectrum, but no COSY crosspeaks. These are most likely nucleotides with some mixed population of C2'/C3'-endo conformations. The ribose puckers for these nucleotides were left unrestrained during molecular dynamics.

The glycosidic torsion angle χ was determined from the intensity of the intranucleotide H8 - H1' NOE. Nucleotides with a *syn* glycosidic conformation have characteristic short distance of ~2.5 Å (vs. 3.8 Å for *anti*) that can be observed in NOESY spectra with short mixing times. Therefore, nucleotides were assigned as *anti* with χ loosely restrained between -125 to -190° if no intranucleotide H8-H1' NOEs were observed at a mixing time of 50ms. Where possible, other intranucleotide base to ribose NOEs where checked to confirm the glycosidic conformation. The presence of a strong base to

H3' NOE for nucleotides with C3'-endo sugar pucker confirm an *anti* conformation (1.8 - 3.5 Å for *anti* vs. 4.3 - 5.2 Å for *syn*). The presence of a strong base to H2' NOE for nucleotides with C2'-endo sugar pucker confirm an *anti* conformation (2.0 - 3.8 Å for *anti* vs. 3.4 - 4.6 Å for *syn*).

Peptide ϕ torsion restraints were obtained from the ratio of cross peak to diagonal intensity in an HNHA experiment (Vuister and Bax, 1993). The coupling constant (J) is obtained from this ratio via the following equation

$$J = (\tan^{-1}(\mathrm{sqrt}(S_c/S_d))/2\pi\zeta)$$

where S_c and S_d are the cross peak and diagonal intensities, respectively, and ζ is delay time for evolution of the J^3_{HH} couplings (13.05 ms). The torsion angle ϕ can then be obtained from the Karplus equation

$$J = A\cos^2(\theta) + B\cos(\theta) + C$$

where A, B, and C are 6.5, -1.8, and 1.60, respectively for HN-Hα couplings, and $\phi = |\theta| + 60°$. There are two or four solutions for ϕ depending on the value of J. Typically the modeling restraint ranges are chosen to include only the stereochemically reasonable solutions of ϕ.

2.3.2 Modeling protocols Three dimensional models of the RRE-Rev complex were generated using a hybrid distance geometry-simulated annealing approach (Nilges et al., 1988) with the DGII and DISCOVER modules of Insight II (Biosym, Inc.) (Figure 2.3). In DISCOVER, NMR-derived distance and dihedral restraints are included in the

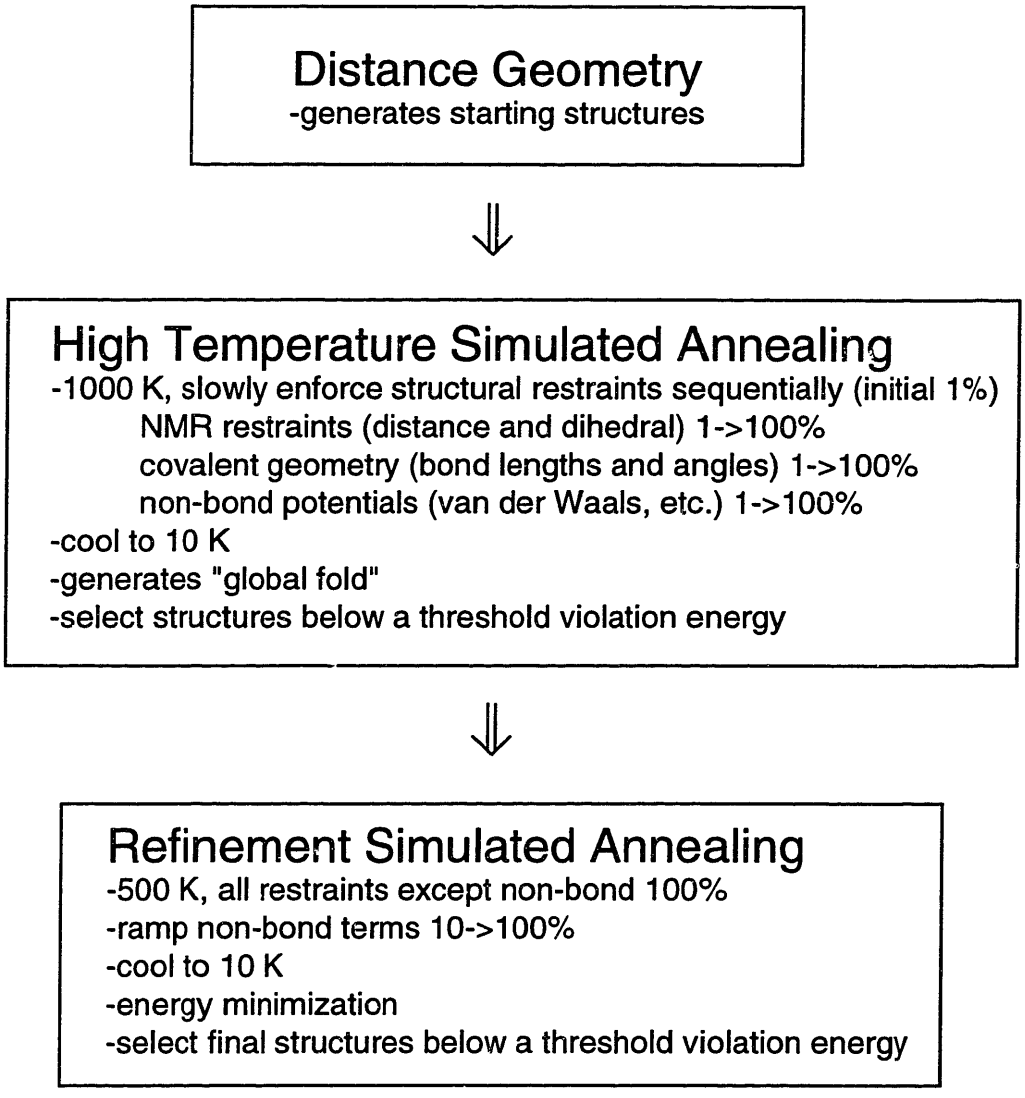


Figure 2.3: Flowchart of molecular modeling protocol for generation of three dimensional structures. % values indicate a scaling constant for the energy terms of a particular class of restraints in the potential energy function.

empirical energy function as flat-well harmonic potentials with force constants of 10 kcal/mol^{-1}-Å^2 and 60 kcal-mol^{-1}-deg^2, respectively.

Distance geometry structures were generated by embedding triangle-smoothed distance-bound ranges with an independent metrization protocol and constant majorization. Left-handed mirror image structures are often observed from distance geometry calculations. The distance geometry structures have poor local covalent geometry and chirality that permit formation of gobally left-handed RNA double helices and protein α-helices. These structures were transformed into right-handed structures by multiplication of the x-coordinates by -1. Several left-handed structures were put through part of the following simulated annealing protocol to ensure that they were not consistent with the NMR data.

The distance geometry structures were input into a high-temperature (1000 K) simulated annealing protocol using the AMBER forcefield with the masses for all atoms set to 100 (Brunger, 1992). For precise details on the parameters at all steps of the following simulated annealing runs see Appendix B. The force constants for the covalent geometry, distance/dihedral, and nonbonded terms of the potential energy function were sequentially scaled from 1% to full value over 36 picoseconds (ps) of dynamics in 3 femtosecond (fs) steps. While the non-bond force constants were at a low value, the van der Waals radii scaling factor was set to 1.25 to prevent the molecule from collapsing. The temperature was then cooled to 10K over 21 ps. The structures with low NMR restraint violation energies (<10 kcal/mol) were chosen for further refinement. Peptide hydrogen bonds were added for the well defined α-helix at this stage. The refinement protocol consisted of 6 ps of dynamics (1 fs steps) at 500 K, scaling the nonbonded terms from 10% to full value. The temperature was then exponentially cooled to 10K over 9 ps, followed by 100 and 500 steps of steepest and conjugate gradient minimization, respectively. The structures with the lowest NMR restraint violation energies (<8 kcal/mol) and no violations > 0.3 Å were

chosen for detailed structural analysis. An average structure (SA) was generated by superimposing all structures, then averaging the coordinates. A minimized average structure (SA_r) with regularized covalent geometry was generated by 100 steps of steepest descent and 500 steps of conjugate gradient minimization on the average structure (SA).

2.4 Other Methods

2.4.1 Size Exclusion Chromatography To monitor dimerization of RNA samples at NMR concentrations (millimolar) , size exclusion chromatography was performed. Experiments were performed with a Bio-Sil SEC 125 HPLC column (Biorad) equilibrated with NMR buffer. 1-3 µl of sample, chased by 10 µl of buffer to ensure complete passage through the injection port, was injected onto the column with a flow rate of 0.3 ml/min and monitored by UV absorption at 260nm. The elution times of the various peaks were compared to molecular weight standards to determined the approximate mobility of the RNA monomers or dimers. For samples with suspected dimers, a series of dilutions were injected to monitor for any concentration dependence of the monomer/dimer ratio.

2.4.2 Circular Dichroism The α-helical content of the free D-Rev_{34-50}AAAAR peptide was measured by circular dichroism on an Aviv 60 DS spectrophotometer. Scans from 190-250nm (0.5 nm steps; 2 second averaging time) were acquired on 200 µl samples (0.1 cm cuvette) of 200 µM peptide in 10mM $NaPO_4$ (pH 6.5), 100mM NaF at 2°C. The output of the instrument, millidegrees (m°), was converted to molar ellipiticity ([θ]) via the following equation

$$[\theta] = 100*m°/(lcN_{res})$$

where l is the path length of the cell (in centimeters), c is the concentration of the sample (in millimolar), and N_{res} is the number of residues in the peptide. Molar ellipiticity was converted to percent helix assuming a $[\theta]_{222}$ value of -35,000 for stable 100% α-helix.

CHAPTER 3: SECONDARY STRUCTURE OF RNA FREE AND BOUND

3.1 Construction of Minimal Complex for NMR

3.1.1 Design of RNA To tailor the minimal RRE RNA fragment for analysis by NMR, the stem IIB RRE RNA used for biochemical analysis of Rev peptide binding (Kjems et al., 1992; Tan et al., 1993) was modified slightly (Figure 3.1). All of the sequence changes were made in regions of the RRE determined to be unimportant for specific binding of Rev and Rev peptides (Bartel et al., 1991; Kjems et al., 1992). The exact base pairing in the lower part of stem IIB in the wild-type RRE is uncertain, since it is near a three-way junction and there are numerous base-pairing possibilities. Constructs used for analysis of the stem IIB interaction vary in the region below the U45:A75 base pair, and the only feature necessary for Rev binding is a stable stem (Bartel et al., 1991; Cook et al., 1991; Heaphy et al., 1991; Kjems et al., 1992; Tiley et al., 1992). Therefore, the G77-A43 mismatch was changed to a Watson-Crick G-C pair to stabilize the lower stem. Potential conformational flexibility could lead to unfavorable exchange properties that would complicate NMR analysis. In addition, the upper stem of RRE IIB was shortened by two base pairs to make the RNA fragment as minimal as possible to reduce spectral overlap in the NMR spectra. Preliminary analysis with RRE RNA containing the wild-type loop sequence indicated dimer formation of the RNA in the presence of Rev peptide at millimolar concentrations, as assayed by size-exclusion chromatography (data not shown). Therefore, the tetraloop sequence was chosen to prevent RNA dimer formation at the high concentrations of RNA necessary for NMR analysis. The GNRA tetraloop is a commonly found motif that stabilizes hairpin formation, and less than 5% dimer was present in millimolar concentrations of the RRE-Rev complex containing the

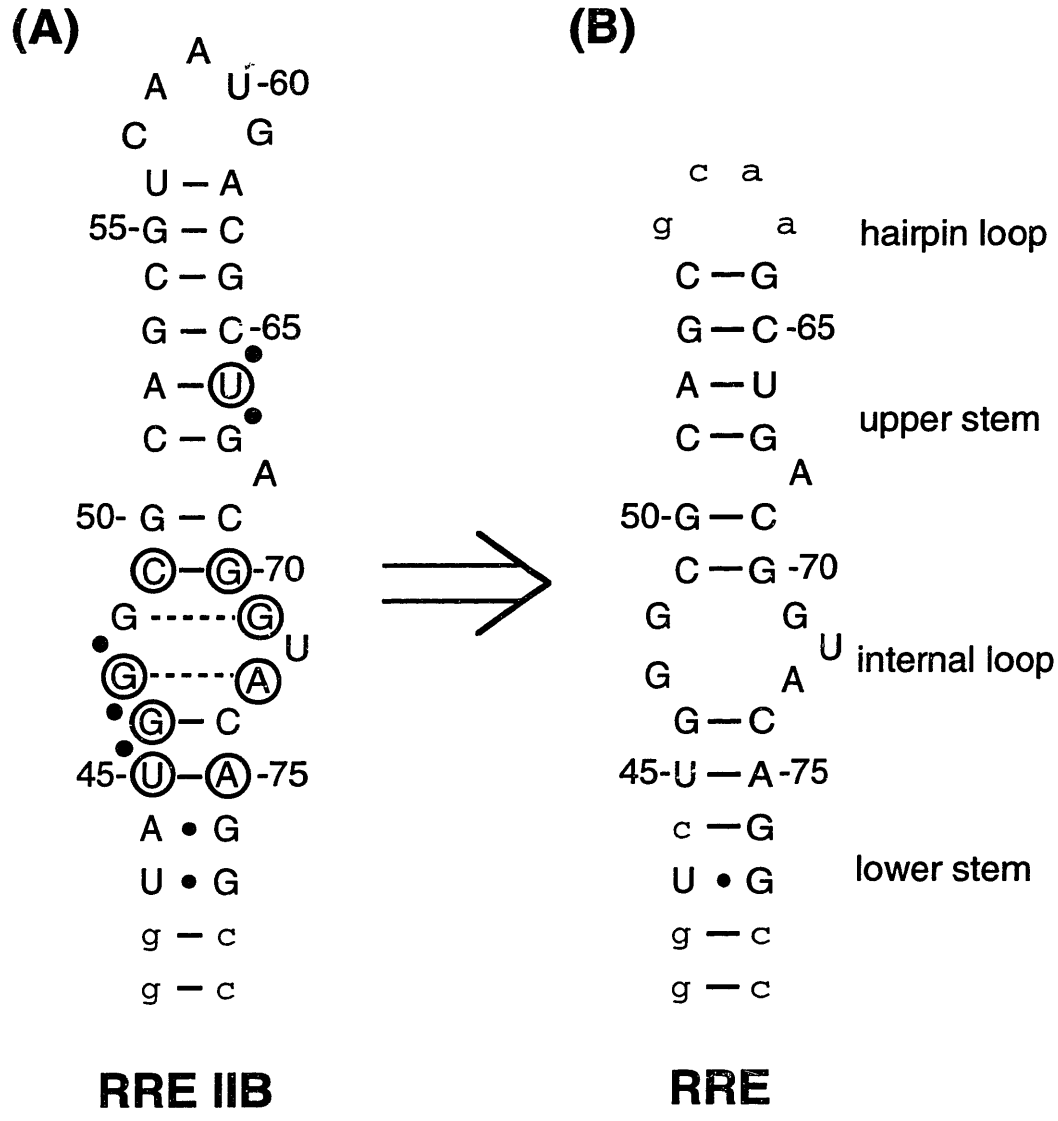


Figure 3.1: Alteration of minimal RRE RNA fragment for NMR analysis. **(A)** Minimial RRE IIB identified from deletion analysis. Open and filled circles represent chemical modification of bases and phosphates, respectively, that interfere with Rev protein binding. Lower case letters represnt non-native nucleotides (see text). **(B)** RRE RNA used for NMR analysis. Rationale for changes from RRE IIB are described in the text. For simplicity, this RNA will be referred to as "RRE" for the remainder of the thesis even though it does not encompass the entire wild-type RNA element.

tetraloop (data not shown). The use of tetraloop sequences in NMR analysis to stabilize hairpin formation of RNAs with unimportant structural elements in the hairpin loop has been recently described in the literature (Molinaro and Tinoco, 1995). To ensure that the modifications to RRE IIB RNA did not affect peptide binding, an electrophoretic mobility shift assay was performed by our collaborators Rouying Tan and Alan Frankel. These assays confirmed that the minimized RRE bound Rev peptides with the same specificity as wild-type RRE IIB (data not shown).

3.1.2 Design of peptides Two modified Rev peptides were used for the majority of the NMR analysis (Figure 3.2). One peptide, suc-Rev_{34-50}-AAAAR-am, has additional amino acids at the C-terminus of the Rev_{34-50} peptide and succinyl and amide blocking groups at the N- and C-termini, respectively (Figure 3.2). These modifications were shown to increase the α-helical content of the peptide with a concomitant increase in specificity of binding to RRE IIB (Tan et al., 1993). This peptide was used with all initial analysis of the RRE RNA conformation with labeled and unlabeled RNA, and most of the results presented in chapter 3 were obtained with this peptide. Figure 3.3 shows a titration of RRE RNA with the synthetic suc-Rev_{34-50}AAAAR-am peptide. New imino proton resonances were observed upon addition of peptide, and some resonances from the free RNA disappeared. At substoichiometric peptide concentrations, two sets of resonances were observed, indicating slow exchange between free and bound forms of the RNA on the NMR time scale, as illustrated for the imino proton resonance of U66. The formation of a discrete set of resonances for the bound RNA is consistent with the formation of a well-defined specific complex with a K_d in the nanomolar range. A similar titration with an unmodified Rev_{34-50} peptide produced quite different results. No additional imino protons were observed for complex formation, no changes in chemical shift were observed, and the linewidth of the resonances increased considerably (data not shown). The exchange

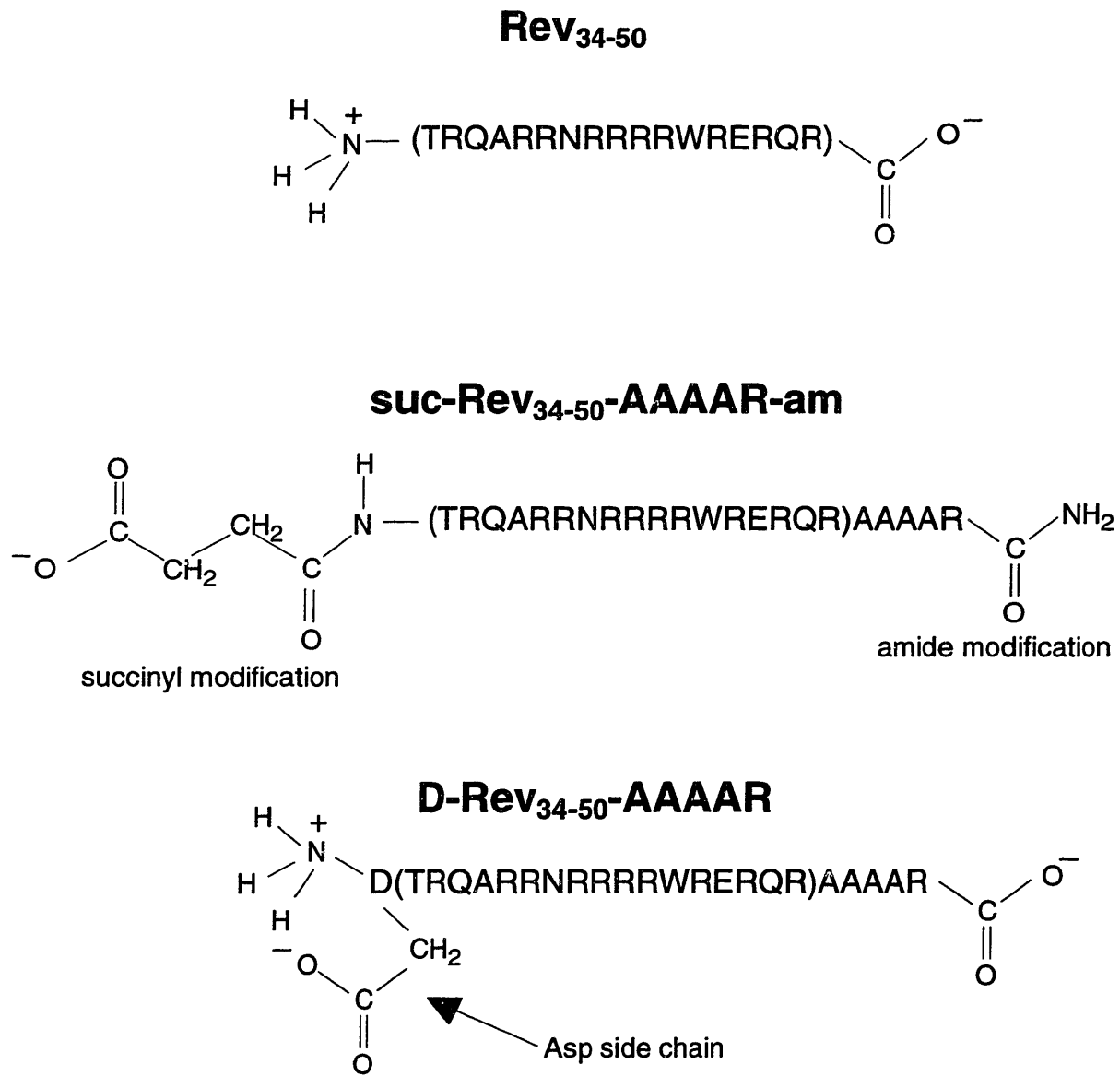


Figure 3.2: Rev peptides used for NMR analysis. The chemical structure of the termini are shown in detail. Rev_{34-50} and suc-Rev_{34-50}-AAAAR-am were made synthetically. D-Rev_{34-50}-AAAAR was expressed in E. coli as a protein fusion, which permits isotopic labeling. The sequence of the Rev peptide from amino acids 34-50 is shown in parenthesis. All other amino acids are non-wild type, yet are included to help stabilize α-helix formation.

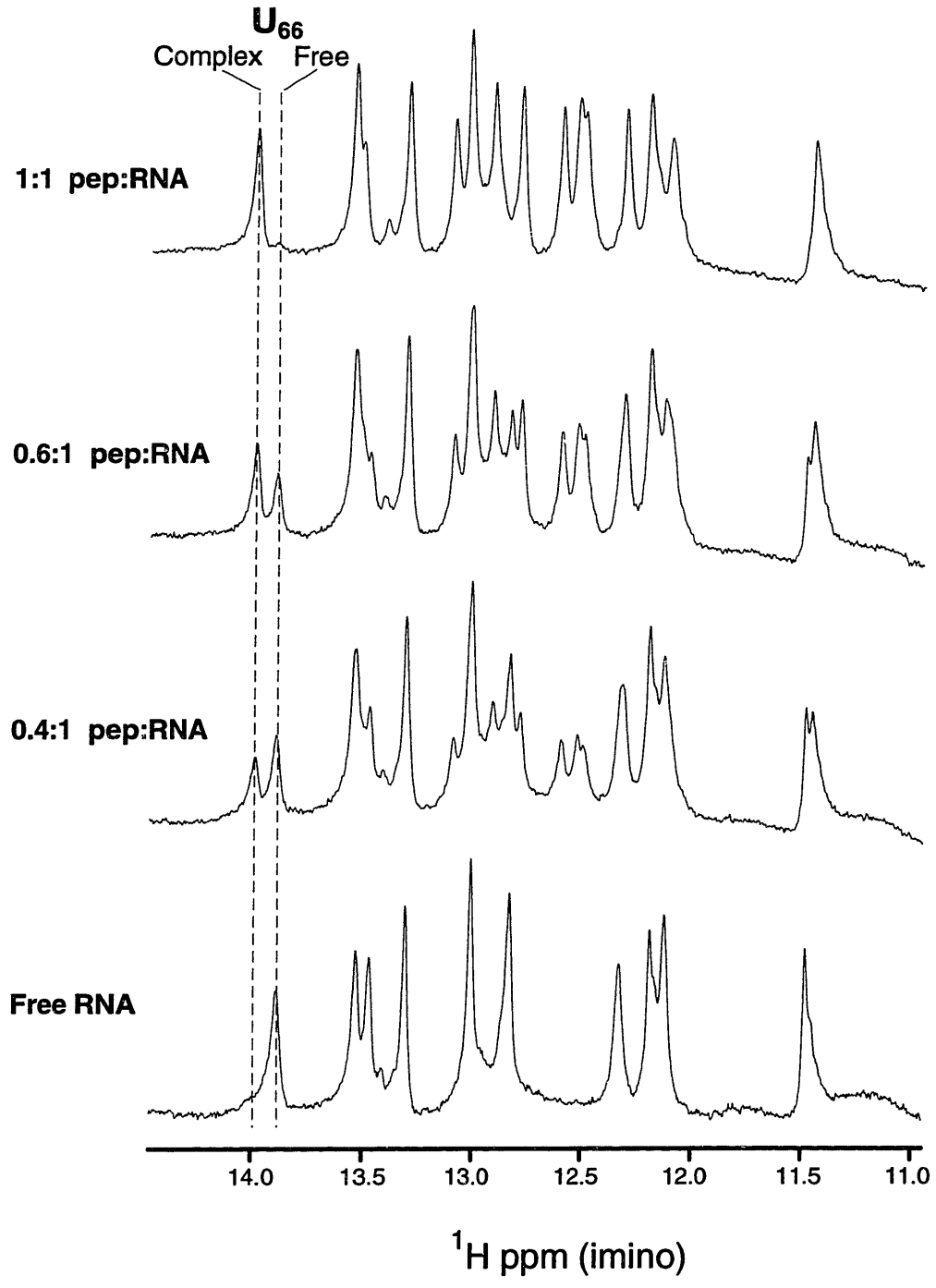


Figure 3.3: 1D NMR spectra in 90/10% H_2O/D_2O of imino protons showing the titration of RRE RNA with suc-Rev_{34-50}-AAAAR-am peptide. The imino protons of free and bound U66 are highlighted, indicating slow exchange of the complex on the NMR time scale. The stoichiometry of each spectrum is indicated to the left.

broadening suggest that the unmodified Rev peptide may bind non-specifically in multiple conformations or that the dissociation rate for this weaker interaction is in the intermediate exchange regime.

The second peptide used extensively for NMR analysis was D-Rev$_{34-50}$-AAAAR, which was produced by expression in *E. coli* (see Materials and Methods). Expression of a Rev peptide in vivo was necessary for uniform isotopic enrichment (^{15}N or ^{13}C,^{15}N), since full assignment of the NMR resonances of unlabeled suc-Rev$_{34-50}$AAAAR-am was not possible due to spectral overlap. The labeled D-Rev$_{34-50}$-AAAAR peptide was necessary for determination of the peptide structure in the complex and identification of RNA-peptide NOEs, and most of the NMR experiments with this peptide as described in more detail in chapter 4. The expressed peptide differs at the N-terminus in that a non-native aspartic acid was added to mimic the succinyl modification of the synthetic peptide. Overall, the main difference between the expressed and synthetic peptides is the charge of the termini, which can have stabilizing or destabilizing interactions with the macrodipole of the α-helix (Marqusee et al., 1989). The expressed peptide (D-Rev$_{34-50}$AAAAR) contains a zwitterion at the N-terminus (NH_3^+/Asp COO-) and a carboxy anion (COO-) at the C-terminus, compared to just a negative charge at the N-terminus (succinyl COO-) of the modified synthetic peptide (suc-Rev$_{34-50}$AAAAR-am) (Figure 3.2). Due to the extra destabilizing charges at the termini, the unbound expressed peptide has a lower α-helical content than the modified synthetic peptide. The α-helical content of the free D-Rev$_{34-50}$AAAAR peptide as measured by circular dichroism (10mM NaPO4, pH 6.5, 100mM NaF, 2°C) is ~12% (data not shown), compared to ~0% and ~50% for the unmodified Rev$_{34-50}$ and modified suc-Rev$_{34-50}$AAAAR-am peptides, respectively (Tan et al., 1993). Despite the lower α-helical content, D-Rev$_{34-50}$AAAAR gave a change in the RRE imino proton spectrum upon titration, characteristic of specific RNA binding by the suc-Rev$_{34-50}$AAAAR-am peptide, which was not observed with the completely unmodified, non-

specific Rev_{34-50} peptide (Battiste et al., 1994). Table 3.1 summarizes the biochemical and NMR data for specific RRE-binding of various Rev peptides. Together, the NMR and binding-data support the hypothesis that the peptide is binding as an α-helix, since modifications that increase helicity, by making the charges at the termini of the RNA-binding site more similar to what they would be in the intact Rev protein, increase binding-specificity and provide better defined complexes for NMR spectroscopy.

Any alteration of the RNA or peptide, however, should be done with caution, since it might perturb the interaction such that the structure observed may have no relation its function in vivo. For this reason, relation of the structural data to biochemical data for the intact protein will be made throughout the thesis. The only biochemical study discussed extensively in the thesis that was performed only with the peptides was (Tan et al., 1993), who performed an alanine-scan mutagenesis set for the Rev peptide. All other studies were done with intact Rev protein or both protein and peptide. If the structural data correlates well with the biochemical data, confidence can be taken in the validity of using the modified peptide for structural analysis. Furthermore, there is additional concern using two different peptides for structural analysis of the RNA. The NMR spectra of the RNA with both peptides were extensively compared and all chemical shifts (within experimental error) and RNA-RNA NOEs were identical for the two complexes (data not shown). In Addition, after assignments for the D-Rev_{34-50}AAAAR peptide were obtained, qualitative analysis of the bound suc-Rev_{34-50}AAAAR-am peptide NMR spectra confirmed that this peptide was forming a similar structure, including the capping box at the N-terminus (data not shown).

Table 3.1: Summary of unbound α-helical content and binding specificity for different Rev peptides. n.d. indicates not determined.

Peptide[a]	%-helix	binding-specificity[d]	NMR titration[e]
Rev_{34-50}	0.1[b]	1.5	-
Rev_{34-50}-am	3.9[b]	6.7	n.d.
ac-Rev_{34-50}-am	10.9[b]	20.0	n.d.
D-Rev_{34-50}-AAAAR	12.0[c]	n.d.	+
ac-Rev_{34-50}-AAAAR-am	27.9[b]	50	n.d
suc-Rev_{34-50}-AAAAR-am	51.2[b]	75	+

[a]All peptides except for D-Rev_{34-50}-AAAAR were produced by chemical synthesis. Chemical modifications at the termini of the synthetic peptides are as follows: am- amide, ac- acetyl, suc- succinyl. [b]determined by circular dichroism (Tan et al., 1993). [c]determined by circular dichroism (data not shown). [d]ratio of K_d's for wild-type and mutant RRE RNAs (K_{nonsp}/K_{sp}) measured by electrophoretic mobility shift assays (Tan et al., 1993). The K_d for suc-Rev_{34-50}-AAAAR-am binding to wild type RRE IIB is 20 nM [e]+ indicates pattern of imino proton resonances indicative of formation of specific complex (see text).

3.2 Analysis of Exchangeable NMR Spectra

3.2.1 Free RNA Analysis of the NMR spectra of the RRE RNA in 10/90% D_2O/H_2O indicated fewer imino proton resonances than expected from RNA secondary structure predictions. If all the base pairs in Figure 3.1B are forming, 12-15 imino protons would be expected depending on how many imino protons are hydrogen bonded in the purine-purine base pairs. Only 10 imino protons were observed in the free RNA (Figure 3.3), and these were assigned using 1D NOE difference spectroscopy (Table 3.2). The exchangeable NMR spectra of G:U wobble pairs are distinctive from Watson-Crick base pairs, since they have two hydrogen bonded imino protons that can be observed instead of one. Therefore, the U43:G77 base pair in the lower stem of the RRE was identified by a characteristic strong NOE between the two imino protons in the base pair, which was used as a starting point for sequential assignment. The remainder of the helical regions were assigned through sequential imino-imino NOEs between stacked G:C and A:U base pairs. These two types of base pairs can be distinguished by unique NOE patterns between imino and other aromatic protons (see Introduction). The pattern of sequential imino NOEs indicated the formation of stacked base pairs in the upper and lower helices of the free RRE from G41:C79 to G46:C74 and from C51:G67 to C54:G64. Previous biochemical studies have suggested the presence of purine-purine base pairs in the internal loop region of RRE IIB (Bartel et al., 1991; Iwai et al., 1992); however, no imino proton resonances were observed from the two proposed purine-purine base pairs in the internal loop. Surprisingly, two potential Watson-Crick base pairs between the bulge A_{68} and the internal loop also did not exhibit imino proton resonances. The absence of resonances indicate that these imino protons are either not hydrogen bonded or the lifetime of the base pair is not long enough to prevent "fast" exchange with water on the NMR time scale (lifetime ~<1ms at 500 MHz).

Table 3.2: Exchangeable proton chemical shifts for RRE RNA free and bound to suc-Rev$_{34-50}$-AAAAR-am peptide. Error in chemical shift ± 0.01 ppm. Proton shifts are referenced to internal TSP.

	Free RNA		Bound RNA	
Residue	imino	amino	imino	amino
G41	n.o.		n.o.	
G42	13.53		13.54	
U43	12.19	-	12.10	-
C44	-	8.50,6.91	-	8.48,7.27
U45	13.45	-	13.57	-
G46	12.31		12.62	
G47	n.o.		12.52	5.50
G48	n.o.		12.49	6.45
C49	-		-	8.31,6.47
G50	n.o.		13.11	
C51	-	8.34,6.79	-	8.32,6.77
A52	-		-	
G53	13.30		13.32	
C54	-	8.17,6.62	-	8.20,6.66
G55	n.o.		n.o.	
C56	-		-	
A57	-		-	
A58	-		-	
G64	13.00		13.03	
C65	-	8.61,6.79	-	8.65,6.80
U66	13.87	-	14.01	-
G67	12.11		12.31	
A68	-		-	
C69	-		-	7.96,6.48
G70	n.o.		12.93	
G71	n.o.		12.21	5.63
U72	n.o.	-	n.o.	-
A73	-		-	
C74	-	(8.15,6.89)	-	8.41,7.30
A75	-		-	
G76	12.83		12.81	
G77	11.49	6.32	11.46	6.39
C78	-	8.43,6.94	-	8.47,7.01
C79	-		-	

() indicates tentative assignment. n.o. indicates resonance not observed. - indicates type of proton in column heading does not exist for that nucleotide.

3.2.2 Bound RNA

Assignment of bound imino proton resonances Upon stoichiometric binding of the peptide, five additional imino proton resonances are observed for the RRE (Figure 3.3). The 15 imino resonances in the bound RRE were assigned by sequential imino-imino NOEs from 1D NOE difference spectroscopy and a 2D NOESY spectrum (Table 3.2). Figure 3.4 shows the imino-imino region of a NOESY spectrum that contains weak NOEs between stacked base pairs and strong intra-base pair imino-imino NOEs for the G:G and G:U base pairs (see below). Two of the new imino protons observed in the bound form of RRE were from the two Watson-Crick pairs below A68. Two of the other new imino protons were assigned to a G:G base pair in the internal loop. An NOE was observed from the G70 imino proton of the last Watson-Crick base pair in the upper stem to two more imino protons (peak **f** and **g**). These two imino protons had a very strong NOE to each other (peak **h**), which is consistent with the formation of a symmetric G48:G71 base pair with the H1 proton of one guanosine hydrogen bonding to the O6 carbonyl oxygen of the other. In addition, both imino protons had strong NOEs to singlet amino protons. The observation of only one peak for the two amino protons indicates they are in fast rotational exchange about the C2-N2 bond. This is consistent with the lack of hydrogen bond formation by the amino protons (note that hydrogen bonded guanosine and adenosine amino protons in Watson-Crick base pairs are typically unobservable due to intermediate rotational exchange). The assignment of these protons as guanosine amino protons was confirmed with experiments on ^{15}N-RRE, since the N2-amino nitrogen has a characteristic chemical shift (Figure 3.5). A 3D NOESY-HSQC was also acquired with the bound ^{15}N-RRE sample to resolve more NOEs between exchangeable protons for the structure determination described in chapter 4 (data not shown).

No sequential imino NOEs were observed for the fifth new imino proton observed in the complex due to spectral overlap, yet it was identified as G47 base pairing with A73

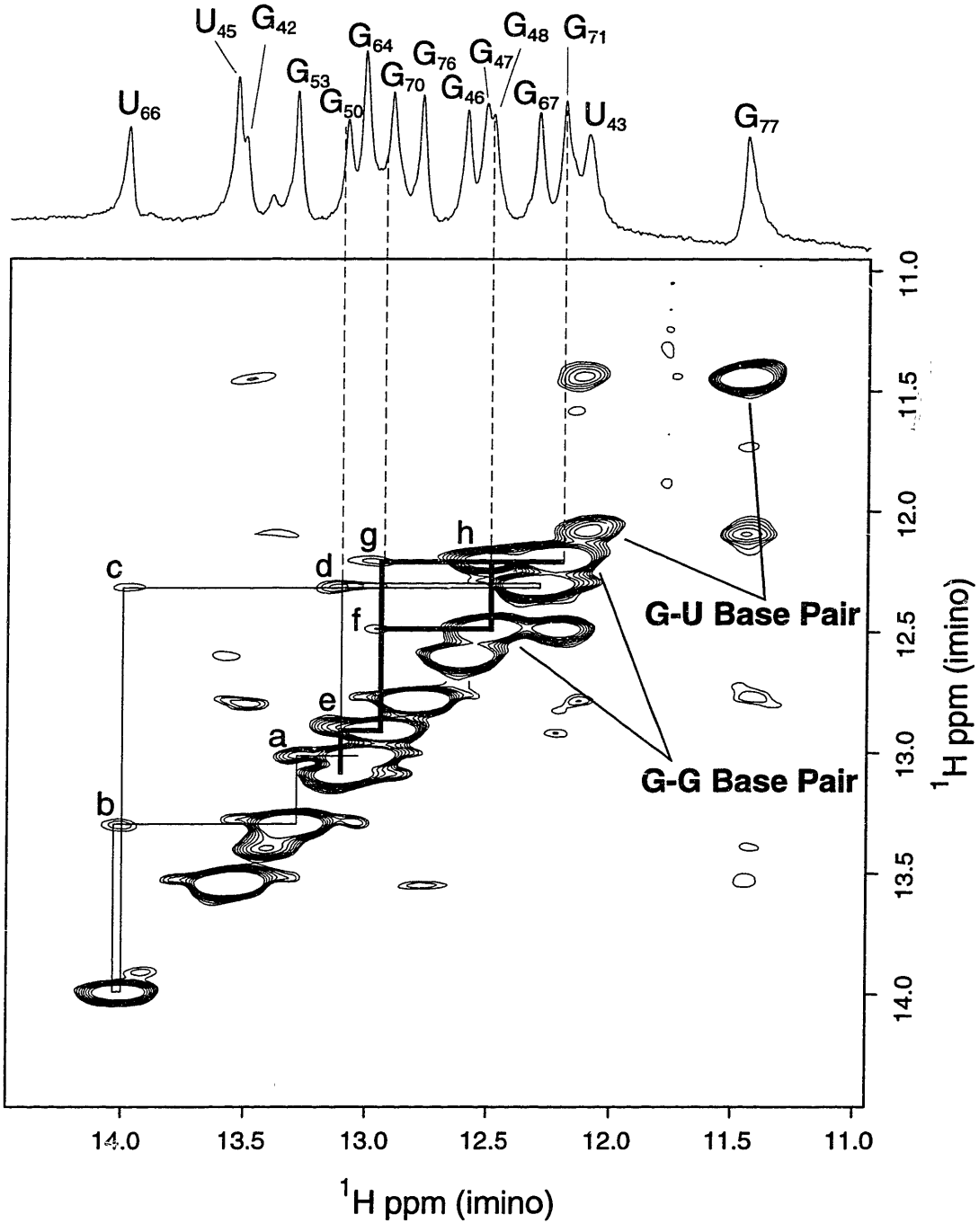


Figure 3.4: Imino-imino region of a NOESY spectrum of the RRE-suc-Rev$_{34-50}$-AAAAR-am complex in 90/10% H_2O/D_2O (mix=250ms). The sequential NOE pathway for the upper stem is traced out. The bold lines are for the additional base pairs which form only in the complex. Imino-imino NOEs are (a) G64-G53, (b) G53-U66, (c) U66-G67, (d) G67-G50, (e) G50-G70, (f) G70-G48, (g) G70-G71, and (h) G48-G71. Above the plot is a 1D imino spectrum with four of the imino protons only observed in the complex indicated by dashed lines.

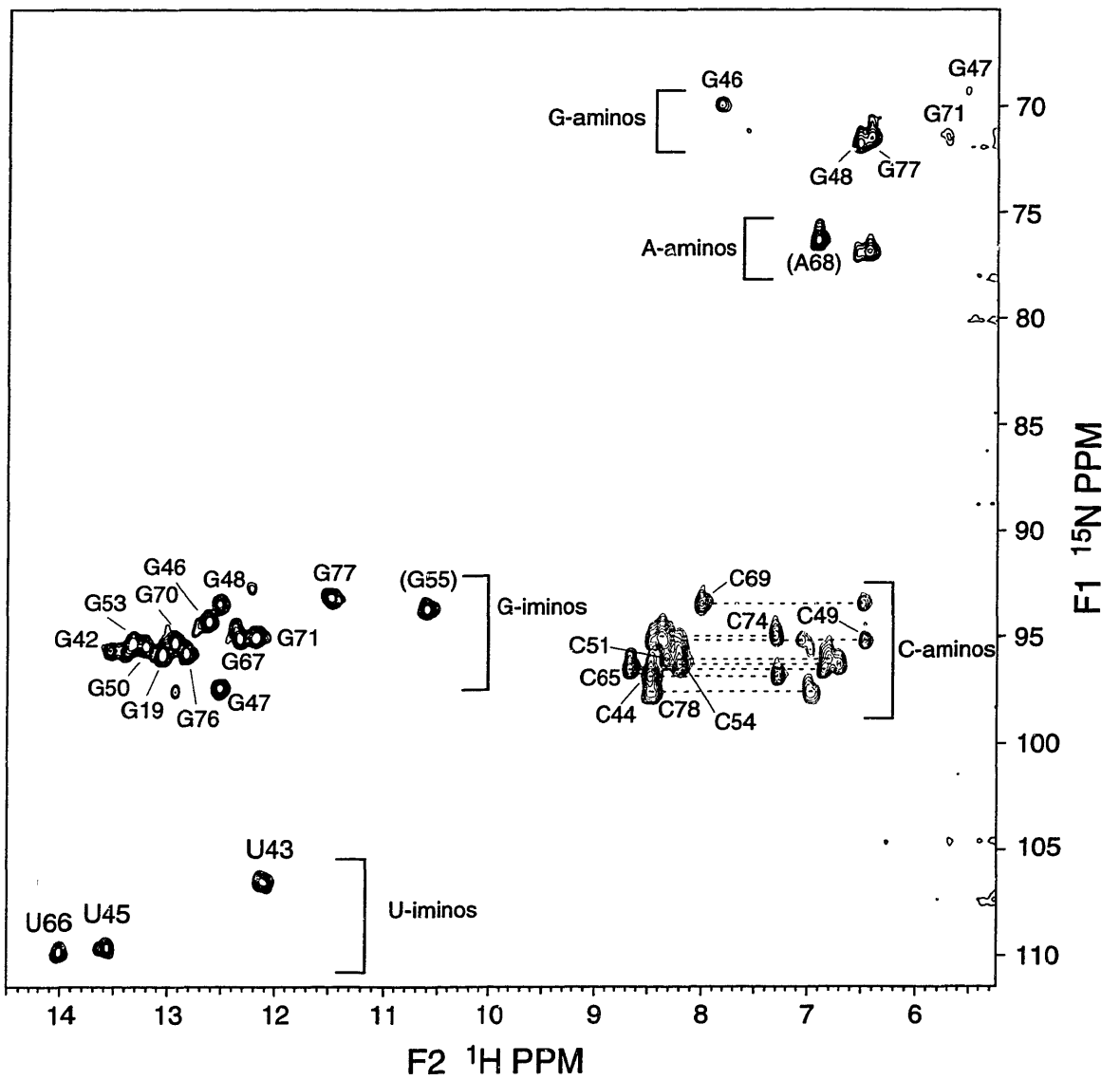


Figure 3.5: 2D G-HSQC experiment on ^{15}N-labeled RRE-D-Rev$_{34-50}$-AAAAR-am complex in 90/10% H_2O/D_2O. Assignment of peaks are given, and types of resonances characterized by chemical shift ranges are donted by brackets. Note that imino protons are folded in the ^{15}N dimension, and one sweep width (50.1 ppm) should be added to the axis value to obtain the correct chemical shift. () indicates tentative assignment.

in the internal loop on the basis of a strong NOE between an imino proton and an H2 proton. The only G:A base pair consistent with this NOE has the Watson-Crick faces of both nucleotides hydrogen bonding between the N1 and H6 positions of adenosine and the H1 and O6 positions of guanosine, respectively. In this type of base pair, the short imino-H2 distance is similar to a standard A:U pair (Gao and Patel, 1988). In 1D difference NOE and 2D NOESY spectra, three imino protons exhibited strong NOEs to aromatic protons with linewidths characteristic of adenosine H2's (Figure 3.6A). Two of the crosspeaks (**u** and **v**) corresponded to the A:U pairs in the lower and upper stem with the third (peak **y**) corresponding to the G:A pair in the internal loop. Also consistent with this base pair is a strong NOE from the imino proton to a singlet, guanosine amino proton, which is not hydrogen bonded in this type of G:A pair.

While no sequential imino-imino NOEs could be observed for the G:A pair in the internal loop due to resonance overlap, unambiguous sequential assignment of the G47:A73 pair stacked in the helix was obtained through the H2 proton of A73. Figure 3.6B and 3.6C shows the imino-aromatic region of the NOESY spectrum as well as the corresponding C2-H2 region of an HSQC-CT spectrum with uniformly ^{13}C-labeled RRE. The HSQC-CT spectrum unambiguously confirms the identities of the adenosine H2 protons, since the carbon chemical shift of C2 resonates in a region separate from all other RNA carbons with protons directly attached. The H2 proton at 7.63 ppm exhibits NOEs to both the imino protons in the G:G pair (peaks **w** and **x**) and to the imino proton of G46 (peak **z**) in the Watson-Crick pair closing the lower stem. Peak **z** was more clearly resolved in a NOESY experiment performed at 30°C (data not shown). In addition, NOESY spectra in 99.996% D_2O show that A73-H2 exhibits an NOE to C74-H1' on its 3' side, which is characteristic of helical adenosines (see Chapter 4 for complete description of bound RRE nonexchangeable proton assignments). The observation of these four new base pairs with sequential NOEs to both the upper and lower helices identified in the free

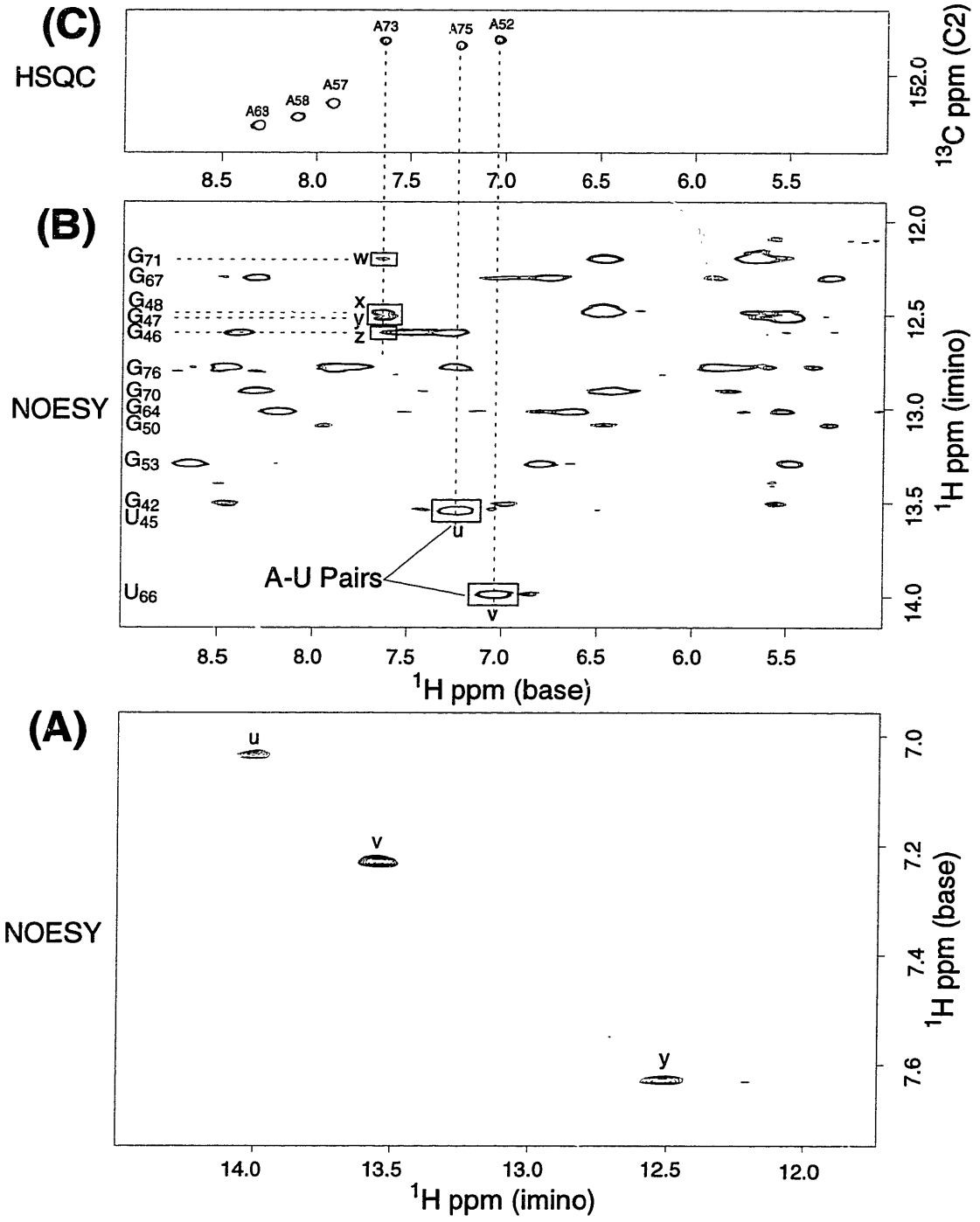


Figure 3.6: Evidence for formation of G:A base pair. **(A)** Imino-base region of NOESY spectrum of the RRE-suc-Rev_{34-50}-AAAAR-am complex in 90/10% H_2O/D_2O (mix= 250ms) showing three strong NOEs to H2 protons. Peak u is U66 H3-A52 H2, peak v is U45 H3-A75 H2, and peak y is G47 H1-A73 H2. **(B)** Base-imino region of same NOESY spectrum showing sequential base stacking NOEs to A73 H2 (boxed peaks w,x,y,z are G71,G48,G47, and G46 imino protons, respectively). **(C)** C2 region of a HSQC of ^{13}C-labeled RRE complex, identifying the six adenosine H2 proton resonances.

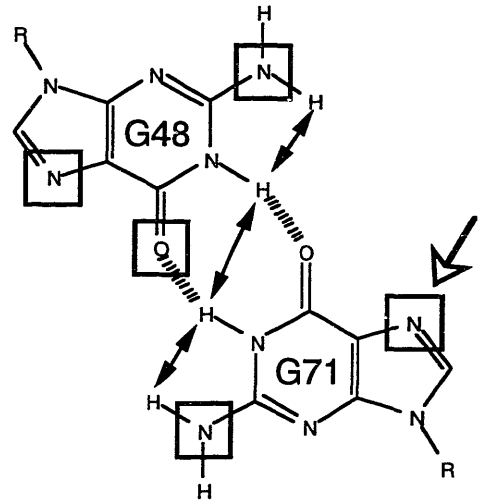
RNA suggest the formation of a continuous helix throughout the internal loop region in the peptide-bound form of the RRE.

The geometries of the purine-purine base pairs in the RNA internal loop of the complex are consistent with previous biochemical investigations on the RRE. Figure 3.7 shows the proposed hydrogen bonding pattern of the two purine-purine pairs, based on the NMR data, and the functional groups thought to be important for Rev binding based on biochemical data (Bartel et al., 1991; Iwai et al., 1992; Kjems et al., 1992; Tan et al., 1993; Pritchard et al., 1994). This symmetric G:G base pair was predicted from the results of in vitro selection of RRE mutants that bind Rev (Bartel et al., 1991). The prediction was based on the isolation of an A:A base pair which can form a base pair isosteric to the symmetric G:G base pair with hydrogen bonds between the amino and N1 positions of adenosine. Inosine substitution showed that removal of the G-amino group at both of these positions had no effect on Rev binding (Iwai et al., 1992). All of these data are consistent with the base pair scheme shown in Figure 3.7. The geometry of the proposed G:A pair is also consistent with all previous biochemical data. Each functional group involved in a hydrogen bond in this base pair has been shown to be important for Rev binding (Iwai et al., 1992; Pritchard et al., 1994). The orientation of this base pair places both N7's in the major groove. Both of the N7's give chemical modification interference of peptide binding, and it is thought that the peptide interacts in the major groove of the RRE (Kjems et al., 1992; Tiley et al., 1992).

3.3 Analysis of Nonexchangeable NMR Spectra

Analysis of the nonexchangeable NMR spectra for both free and bound RNA support the general conclusions obtained from exchangeable proton NOE analysis. Sequential base to ribose NOEs typical of an A-form helix were observed for the lower

G:G pair



G:A pair

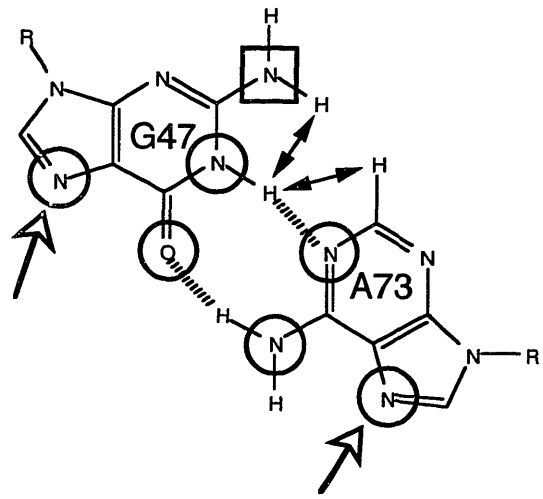


Figure 3.7: Geometry of purine-purine base pairs consistent with NMR and biochemical data. Filled double-headed arrows indicate observed NOEs. Open arrows indicate N7 positions that interfere with Rev binding upon modification with DEPC. Circles and squares represent positions where chemical substitution with a modified base does or does not interfere with Rev binding, respectively.

and upper stems of the free RNA (Figure 3.8A), with an H8-H8 NOE placing the position of the A68 base stacked below G67 (data not shown) . However, assignments for the internal loop region, including the two potential Watson-Crick base pairs below the bulge A, were incomplete due to the lack of sequential NOEs (not resonance overlap) (Table 3.3). In addition, many of these nonexchangeable resonances were broadened, suggesting intermediate exchange on the NMR time scale and possible multiple conformations. The exchange broadening of the free RNA resonances is especially clear in a ^{13}C-HSQC of the ribose region (Figure 3.9A). For example, in this spectrum one-cross peak should be observed for the H1'-C1' correlation of each nucleotide (34 for the RRE). Much less than 34 resonances are observed in the H1'-C1' region of the free RNA HSQC (compare to bound RNA). The intensity of these correlations should be independent of structure; therefore, conformational exchange broadening is the only explanation for the absence of peaks. From the assignments of the NOESY spectrum (Figure 3.8), it is clear that the missing resonances are in the internal loop of the RNA. The conformational exchange indicates that the internal loop of the free RNA is dynamically flexible with motions on the millisecond time scale. This is distinct from being completely "unstructured", which would have fast ($\leq$ microsecond time scale) motions resulting in a single, sharp time averaged resonance.

All together, the NMR data regarding the secondary structure of the RNA indicate that the internal loop is stabilized and/or undergoes a conformational change upon binding of the Rev peptide. Figure 3.10 summarizes the NMR data for the secondary structure of the RRE in the free and bound form (the bound nonexchangeable RNA data is elaborated on in Chapter 4). An exact conformational change is difficult to pin-point, since a unique structure for the internal loop was not defined in the free RNA. It is possible that the bound RNA secondary structure is present in the free RNA, yet not definable by NMR due to dynamic flexibility. A couple of pieces of evidence, however, point to a conformational

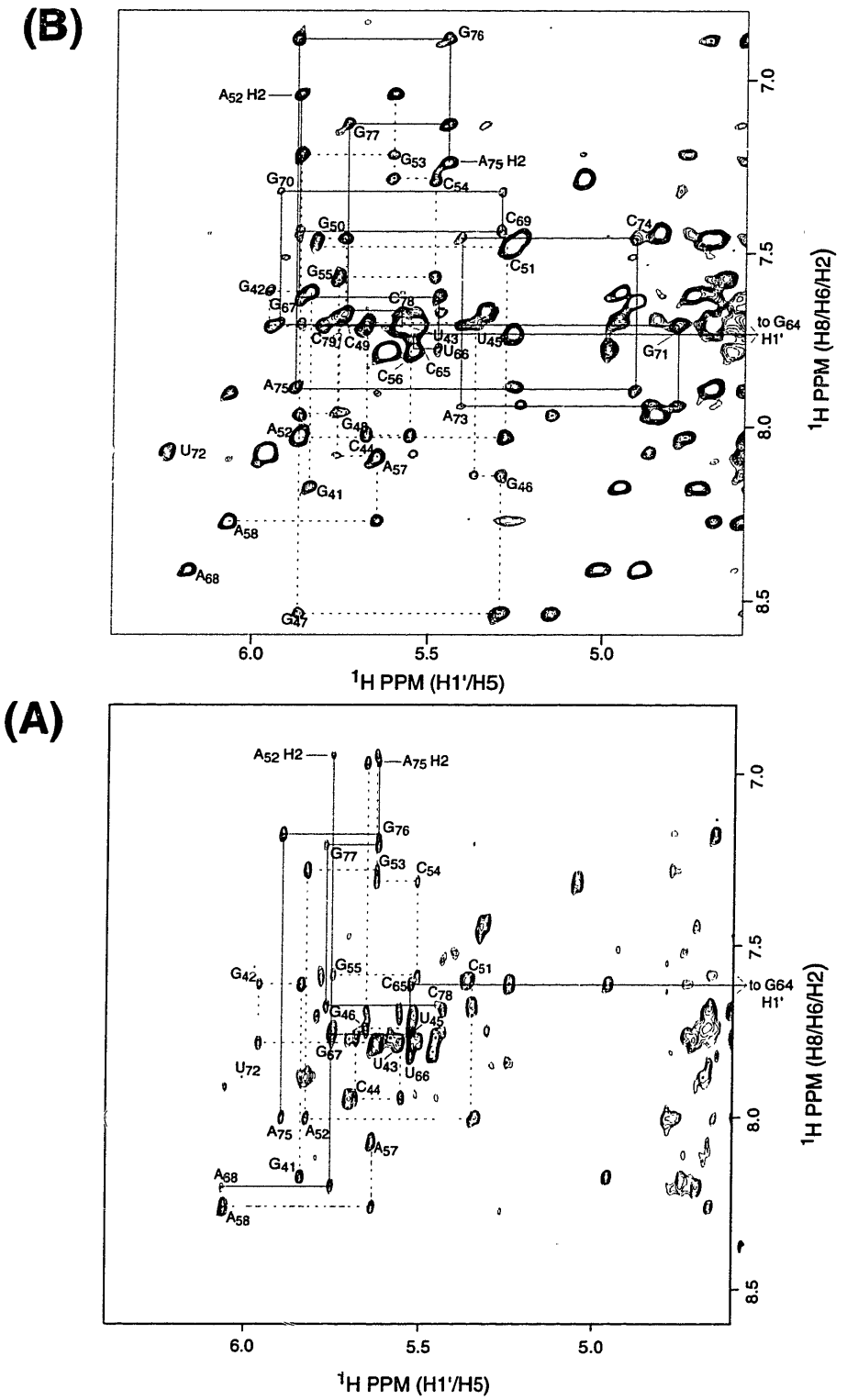


Figure 3.8: Comparison of sequential walk regions of NOESY spectra (mix=400 ms) in 99.996% D_2O for **(A)** Free RRE RNA and **(B)** RRE RNA bound to suc-Rev$_{34-50}$-AAAAR-am peptide. The dashed and solid lines denote the connectivities for nucleotides G41-A58 and G64-C79, respectively. Also denoted are the helical adenosine H2-H1' NOEs that are used as starting points for sequential assignment.

Table 3.3: Free RRE RNA nonexchangeable proton chemical shifts. Error in chemical shift ± 0.01 ppm. Proton shifts are referenced to internal TSP.

Residue	H8,H6	H5,H2	H1'	H2"
G41	8.17	-	5.84	4.96
G42	7.61	-	5.96	4.72
U43	7.78	(5.51)	5.56	4.23
C44	7.94	5.71	5.69	4.44
U45	7.76	5.45	5.58	(4.42)
G46	7.74	-	5.66	4.18
G47	7.69	-	5.44	4.62
G48	(7.54)	-		
C49				
G50		-		
C51	7.60	5.36	5.34	4.51
A52	8.00	6.95	5.83	4.56
G53	7.28	-	5.64	4.34
C54	7.32	5.05	5.52	4.50
G55	7.58	-	5.76	4.54
C56	7.78	5.63	5.57	4.42
A57	8.06	7.90	5.64	4.35
A58	8.26		6.06	4.67
G64	7.84	-	3.81	4.32
C65	7.61	5.24	5.52	4.49
U66	7.81	5.46	5.52	4.41
G67	7.76	-	5.76	4.25
A68	8.20		6.06	4.70
C69				
G70		-		
G71		-		
U72	7.88	5.83	5.99	
A73	8.12*	7.85	5.81*	
C74	7.52*	5.46*	5.34	4.51
A75	8.00	6.97	5.90	4.66
G76	7.18	-	5.63	4.44
G77	7.21	-	5.78	4.67
C78	7.68	5.36	(5.45)	
C79	7.70	5.52	5.79	4.04

() indicates tentative assignment. n.o. indicates resonance not observed. * indicates resonance observed only at 35°C.

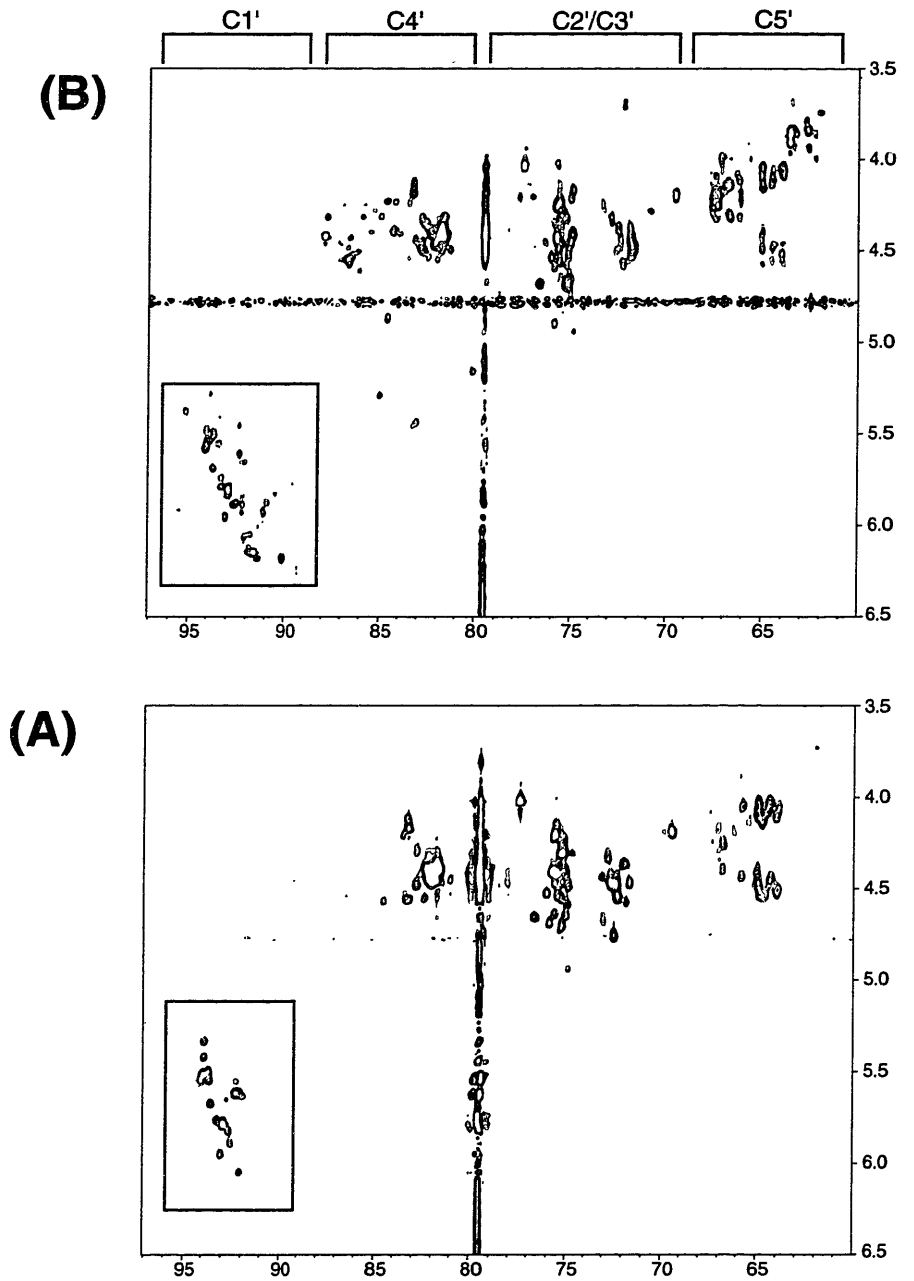


Figure 3.9: Ribose region of HSQC-CT experiments in 99.996% D_2O on ^{13}C-labeled **(A)** RRE RNA and **(B)** RRE RNA bound to suc-Rev_{34-50}-AAAAR-am peptide. Bars at top show characteristic chemical shift ranges for the ribose carbons. C1'/H1' correlations are in the lower left-hand corner of each plot (boxed).

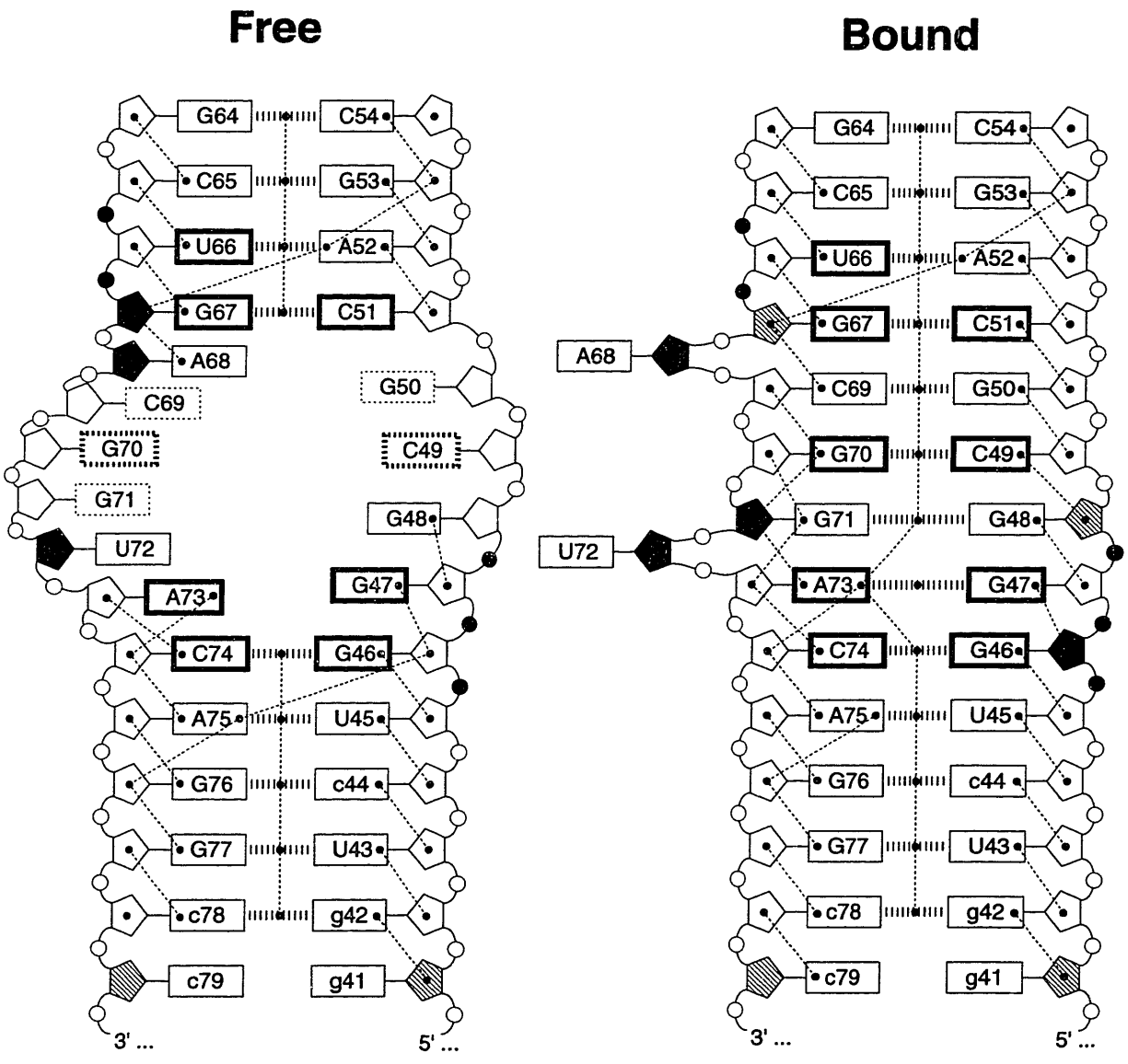


Figure 3.10: Schematic of NMR data for free and bound RRE RNA. Thick boxes indicate nucleotides invariant in RNA selection studies in vitro. Thick dashes indicate observed imino protons involved in base pairs. Thin dashes indicated observed NOEs. White, black, and hatched pentagons indicate riboses with C3'-, C2'-, and C2'/C3' mixture, respectively, sugar puckers. Important phosphates are indicated by black circles. Unassigned and/or unobservable nucleotides are indicated by dashed boxes.

change rather than just a stabilization of existing structure. NOEs indicate that A68 is stacked with the upper A-form helix in the free RNA. These NOEs are not present in the bound RNA, and A68 is bulged out of the helix with G67 and C69 stacked. In addition, the cross-strand NOE from A75-H2 to G46 H1', which is characteristic of an A-form helix, is not present in the bound RNA. The sugar pucker of G46 is also C2'-endo (non-A-form). These data qualitatively suggest a distortion of the lower A-form helix near the internal loop that is only present in the bound form of the RNA, consistent with a conformational change upon peptide-binding. Conformational changes in RNA upon protein binding have been observed in numerous cases, including TAR-arginine, $tRNA^{Gln}$-synthetase, and $tRNA^{Asp}$-synthetase complexes (Rould et al., 1989; Ruff et al., 1991; Puglisi et al., 1992). The simplest interpretation of these conformational changes would be that the affinity and specificity of binding is reduced compared to a completely preorganized binding site. Nevertheless, protein-induced conformational changes appear to be a common feature of RNA-protein complexes (see chapter 5).

CHAPTER 4: STRUCTURE DETERMINATION OF RNA-PEPTIDE COMPLEX

4.1 Assignment Procedure for RNA

The RRE RNA was predominantly assigned using a NOESY-based procedure (see Introduction) with described homonuclear and heteronuclear NMR methodologies (Varani and Tinoco, 1991; Nikonowicz and Pardi, 1993). However, a novel method of selectively labeling the RNA by nucleotide type was also utilized for assignments and NOE identification. There are three advantages to specifically labeling an RNA, which will be described in more detail below: 1) Individual resonances can be identified by nucleotide type (G, A, C, or U) in heteronuclear experiments where only protons attached to ^{13}C are observed. This helps reduce ambiguity in the sequential assignment pathway, since ribose protons cannot be identified by nucleotide type and base protons can be identified only as purine or pyrimidine. 2) Heteronuclear 2D and 3D NMR experiments have reduced spectral overlap, which allows for the identification of more NOEs for structural analysis. 3) Isotopically-filtered NOESY experiments can be performed to unambiguously identify interresidue NOEs between labeled and unlabeled nucleotides.

4.1.1 Through-bond correlation experiments

To unambiguously identify all H1' protons for assignment of the sequential H8/H6 to H1' pathway, the constant-time variant of the Heteronuclear Single Quantum Coherence (HSQC-CT) experiment was performed (Santoro and King, 1992). When the constant time interval is set to an odd integer multiple of $1/J_{CC}$ (the one bond carbon-carbon coupling constant), the sign of the cross peak for the C1' and C5' resonances is opposite to that of the C2', C3', and C4' resonances (Santoro and King, 1992). The sign alteration results from the different number of carbon-carbon bonds for each resonance, which

affects the J-coupling modulation during the constant-time interval (see Appendix A). The sign alteration is important, since ^{1}H and ^{13}C chemical shifts alone may not be sufficient in more complex RNA structures to unambiguously identify H1' proton resonances, which can be shifted upfield into the H2'/H3'/H4'/H5'/H5" region. Figure 4.1A shows the C1' and C4' region of an HSQC-CT experiment on uniformly ^{13}C-labeled RRE RNA complexed with Rev peptide. Highlighted are two C1'/H1' crosspeaks that are upfield shifted in both the proton and carbon dimensions. These two resonances were unambiguously assigned as C1'/H1' correlations by the sign of the crosspeak in the HSQC-CT spectrum. Similar experiments on the aromatic region of the RNA can resolve pyrimidine from purine base resonances due to the difference in carbon-carbon bonds between the C8/H8 and C6/H6 spin systems (data not shown).

To further aid with RRE RNA assignment, heteronuclear NMR experiments were performed on samples where only the guanosine or cytosine residues were ^{13}C-labeled. With an RNA labeled specifically by nucleotide the ribose protons for that nucleotide can be identified by the fact that they are attached to a ^{13}C-carbon (ribose ^{1}H-spin systems are isolated and therefore identical regardless of the base attached). Figure 4.1B shows the HSQC-CT spectrum for ^{13}C-RRE RNA- Rev peptide complex labeled selectively at guanosines. In this spectrum, all of the guanosine H1' protons can be identified, including the two unusually shifted H1' resonances, which helps reduce the ambiguity in assigning the through space H8/H6 to H1' NOE sequential pathway. Due to the presence of the unusual H8-H1' NOEs in the guanosine-rich internal loop, this information was important for confirming the assignments of RRE RNA.

On uniformly or specifically labeled RNA samples, 3D HCCH-COSY and HCCH-TOCSY experiments (Clore et al., 1990; Nikonowicz and Pardi, 1993) were performed to correlate the well-resolved H1'/C1' resonances identified by the HSQC-CT with the rest of the ribose spin system (H2'->H5'). Once the assignments of the H1' resonances are

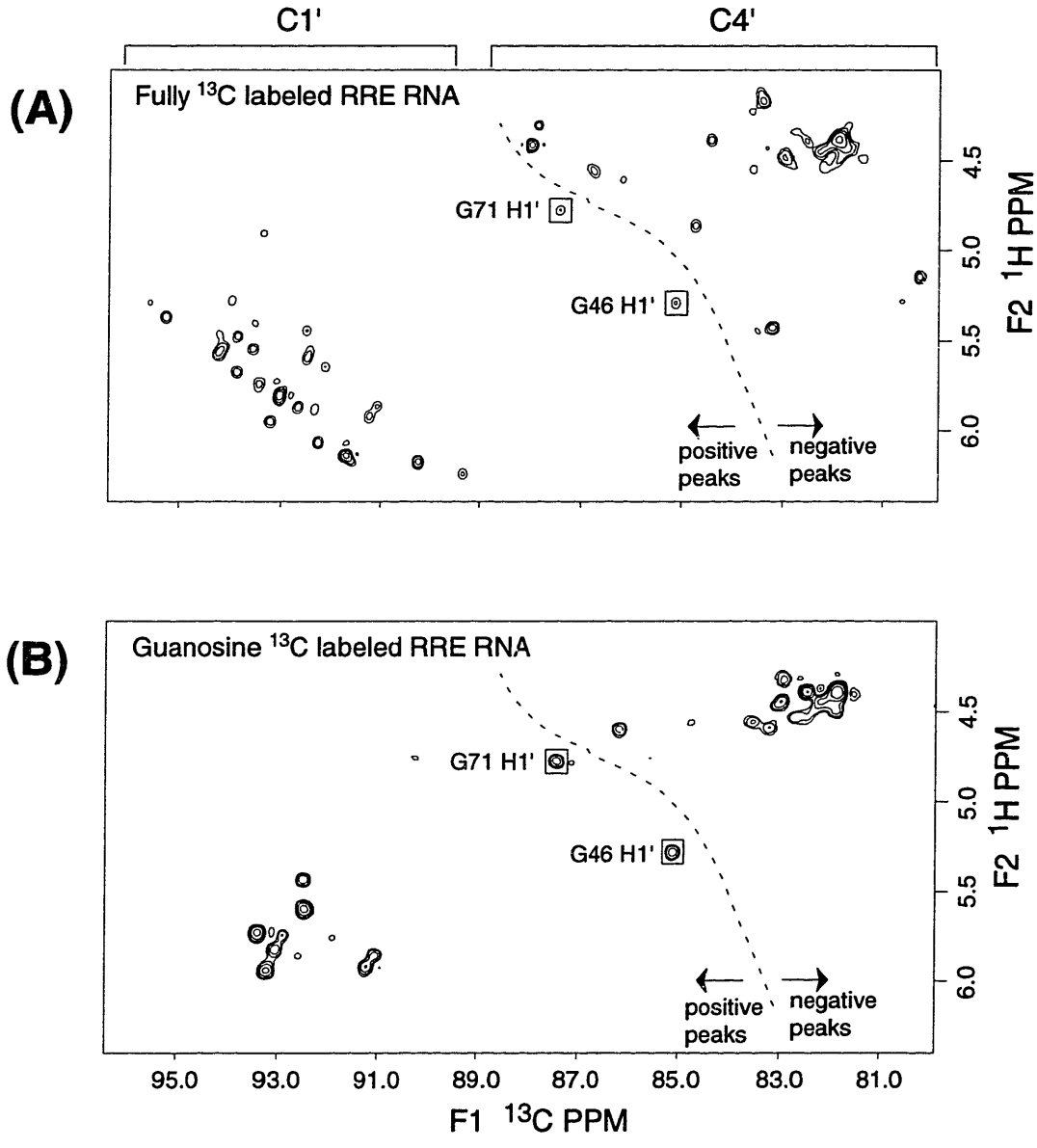


Figure 4.1: C1'/C4' region of HSQC-CT experiments in 99.996% D_2O on **(A)** Uniformly ^{13}C-labeled RRE-suc-Rev_{34-50}-AAAAR-am complex and **(B)** ^{13}C-RRE-suc-Rev_{34-50}-AAAAR-am complex selectively labeled at guanosines. The sign of the crosspeaks in the spectra are demarked by the dashed lines. Two C1'/H1' correlations with unusual carbon chemical shifts are boxed.

obtained from NOESY-experiments (see below), the rest of the ribose protons are therefore also assigned. The "HCCH" nature of these experiments means that magnetization is transferred through the large one-bond J_{HC} and J_{CC} couplings rather than the weak $^3J_{HH}$ couplings of homonuclear COSY or TOCSY experiments (see Appendix A). This is particularly important for RNA, where the C3'-endo sugar pucker of the ribose rings in A-form helices has a $^3J_{H1'-H2'}$ coupling of <3 Hz (unobservable). However, in an HCCH-COSY experiment, all 34 H1'-H2' cross-peaks of the RRE can be observed (data not shown). The assignments can be extended to the H3'->H5' protons with a HCCH-TOCSY experiment, which use a ^{13}C-spin lock to transfer magnetization throughout the ribose ring. Figure 4.2 shows one slice of a 3D HCCH-TOCSY at the carbon chemical shift of two well resolved C1' resonances.

4.1.2 NOESY experiments

Homonuclear 2D NOESY After identification of the H1' and base resonances by through-bond correlation experiments, NOESY experiments can be performed to sequentially link and assign these resonances. Sequence specific assignment of the RRE RNA in complex with the Rev peptide was not obtainable using only conventional homonuclear 2D NMR experiments. This was due to resonance overlap as well as the presence of unusual sequential NOE patterns. Sequential assignment of nucleic acids relies on a particular pattern of NOEs between base and H1' protons that is observed in Watson-Crick duplexes (see Introduction). When other types of structures are present, this pattern may be altered making it difficult to unambiguously determine the assignments. To highlight some of the ambiguities resolved by the heteronuclear experiments, the sequential NOE pathway finally determined for RRE is shown on the 2D NOESY "sequential-walk" region in Figure 4.3. The details of the heteronuclear experiments performed to obtain these assignments are described below. Sequential i,i-1 NOEs from H8/H6 protons to H1' protons on the 5'

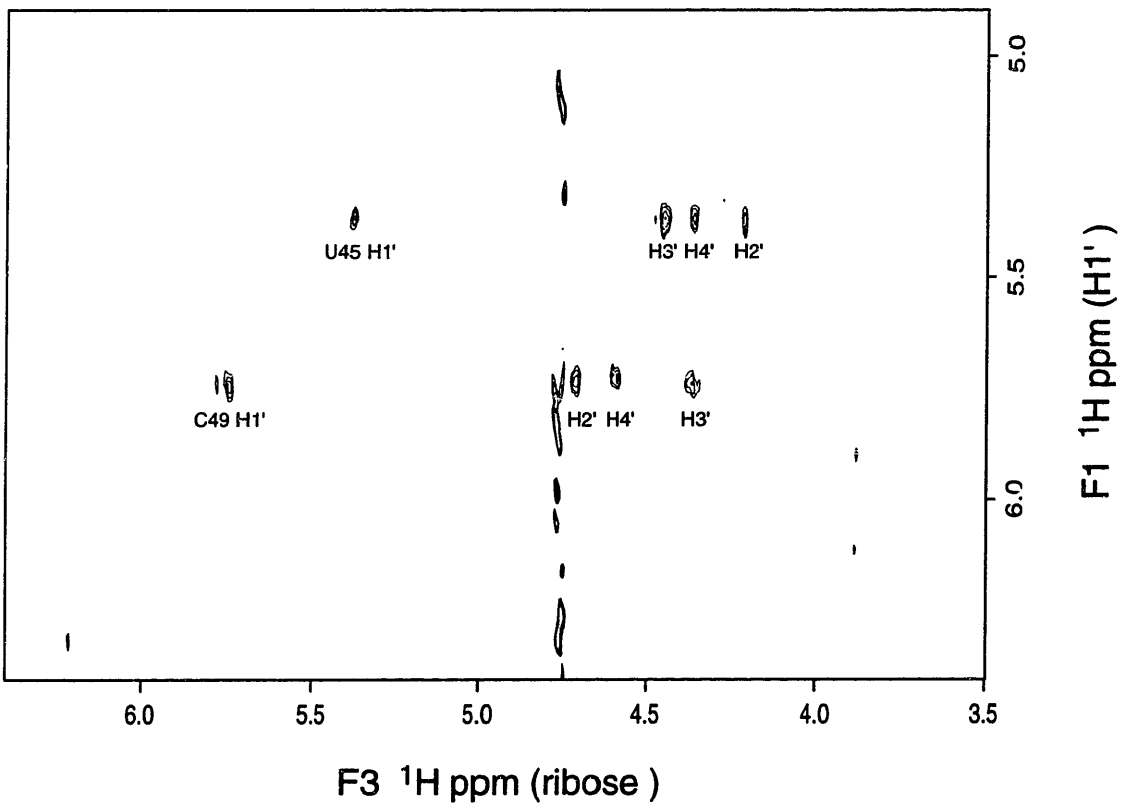


Figure 4.2: 2D slice of a 3D HCCH-TOCSY experiment in 99.996% D_2O on uniformly ^{13}C-labeled RRE-suc-Rev_{34-50}-AAAAR-am complex at a carbon chemical shift (94.7 ppm) in the C1' region. The ^{13}C-DIPSI spin lock was ~18ms resulting in correlations from C1'->C4'. The ribose assignments for U45 and C49 are given in the spectra.

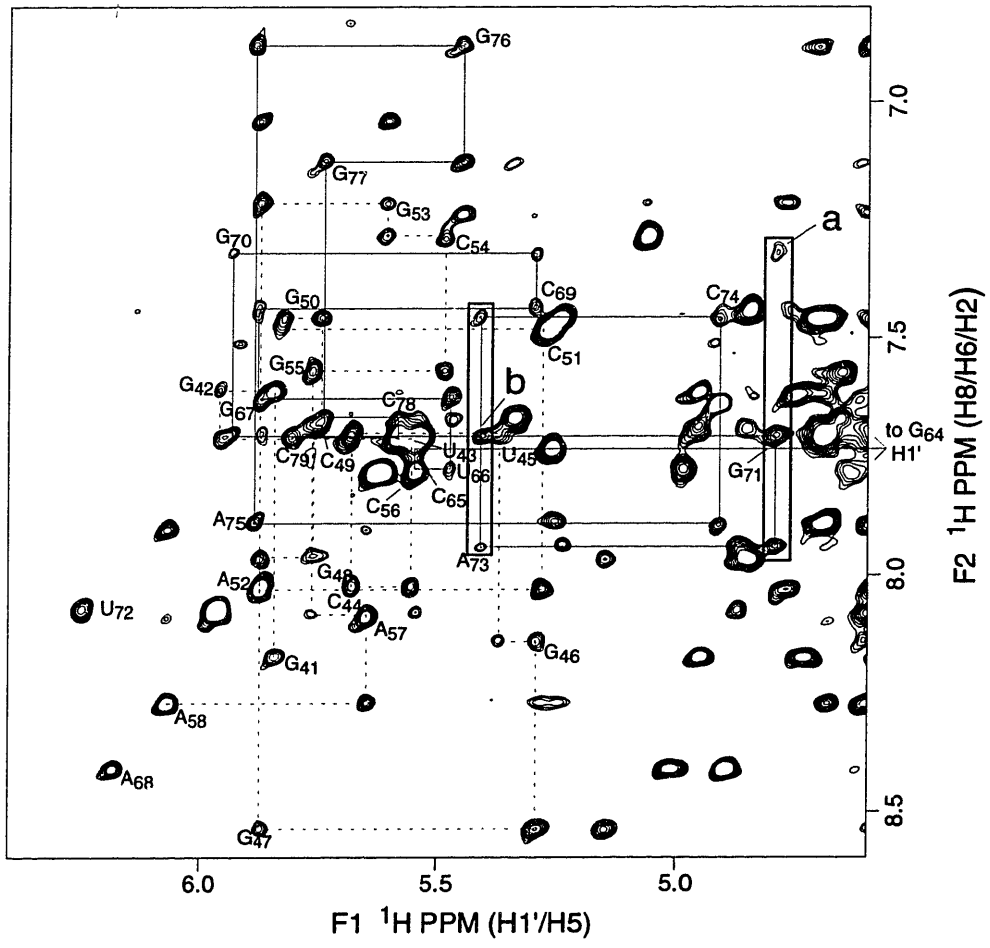


Figure 4.3: Base to H1' region of a NOESY spectrum (mix=400ms) of the RRE-suc-Rev$_{34\text{-}50}$-AAAAR-am complex in 99.996% D_2O. The sequential assignment pathway is highlighted by dashed and solid lines for nucleotides G41-A58 and G64-C79, respectively. The numbers indicate intranucleotide H8/H6 to H1' NOEs. Two H1's with NOEs to three H8/H6 protons are boxed. The nonstandard sequential H8 to H1' NOEs that were observed are indicated by peaks **a** and **b**. Peak **a** is G70 H8 to G71 H1', and peak **b** is G71 H8 to A73 H1'. The G64 H1' resonance is shifted downfield at 3.75 ppm, and is not included in this plot. This shift is a result of the tetraloop structure and has been described previously (Heus and Pardi, 1991).

neighboring nucleotide were observed along almost the entire length of both strands of the hairpin including the internal loop region. Two breaks in the sequential connectivity for U72 and A68 were observed; however, non sequential (i, i-2) NOEs were observed from A73 to G71 and C69 to G67. This pattern is consistent with unstacking of U72 and A68 as single nucleotide bulges along a continuous helix. In addition, there were two "reverse" connectivities from H8 protons on G70 and G71 to H1' protons on G71 and A73, respectively (peaks **a** and **b** in Figure 4.3). Because of the "extra" connectivities, G71 and A73 H1' had NOEs to three H8/H6 protons, rather than the normal two, increasing the ambiguity in determining the sequential assignment pathway.

Three- and four-dimensional experiments to alleviate overlap With both uniformly and specifically ^{13}C-labeled RRE RNA, 3D heteronuclear NOESY experiments were performed to alleviate spectral overlap by resolving the ^{1}H 2D NOESY spectrum in a third dimension according to ^{13}C-chemical shift. A representative plane through the C1' region from a 3D NOESY-HMQC experiment of uniformly ^{13}C-labeled RRE RNA complex is shown in Figure 4.4A. Although this is the most crowded C1' plane, there are seven fairly resolved H1' resonances. One example of an overlap problem that was alleviated by this experiment is illustrated by peak **c**, which is an NOE from C51 H6 to C51 H1' that was obscured by a strong H5-H6 crosspeak in the 2D NOESY spectrum (Figure 4.3). The presence of strong H5-H6 crosspeaks in the sequential walk region of a 2D NOESY that interfere with assignment of the base-H1' NOEs is a common problem. The 3D experiments resolve these cross peaks, since C1' and C5, unlike H1' and H5, resonate in different spectral regions.

In addition to assigning the sequential H8/H6 pathway, the multi-dimensional heteronuclear NOESY experiments were essential to identifying numerous NOEs for molecular modeling calculations. This is especially important for the non-H1' ribose

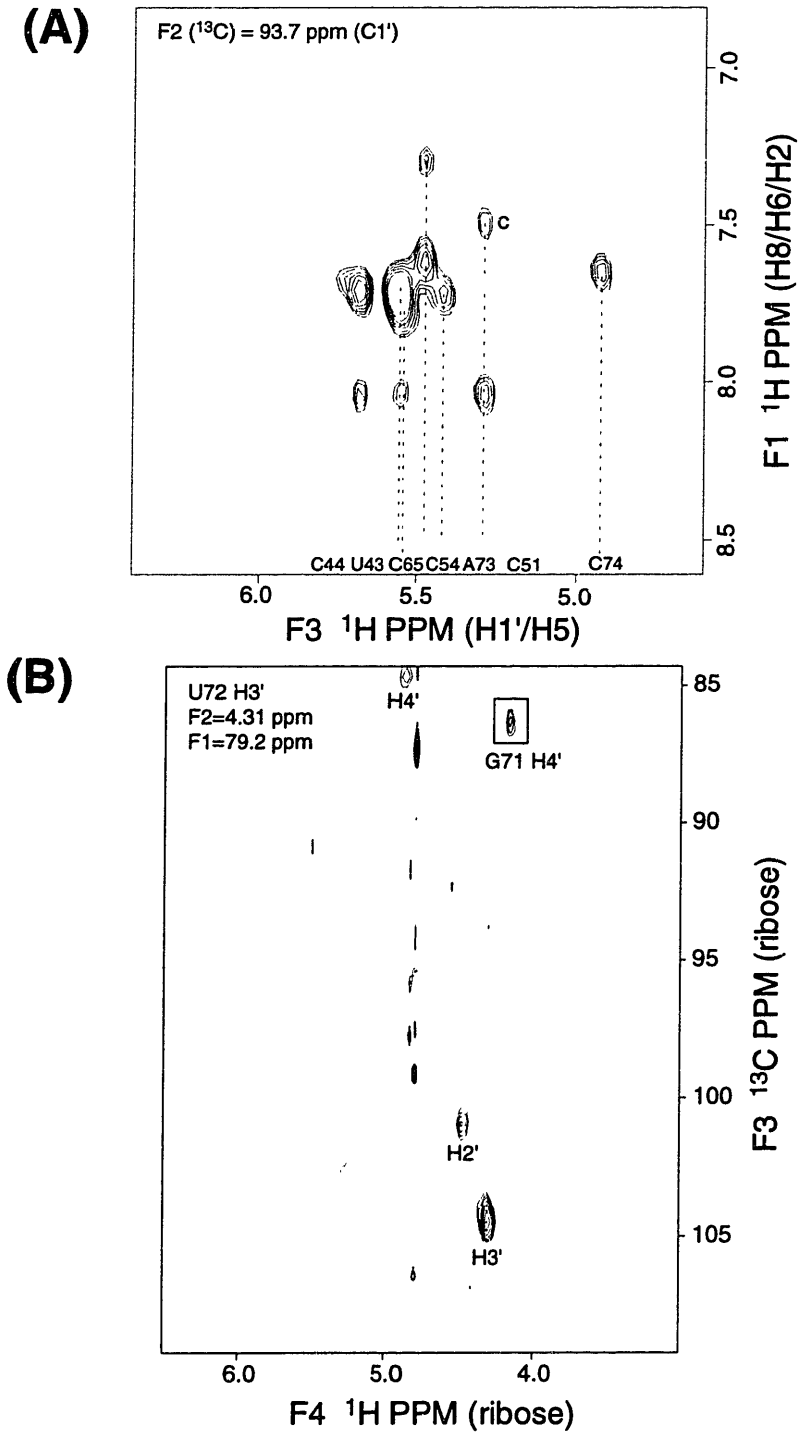


Figure 4.4: (A) 2D slice of a 3D NOESY-HMQC experiment (mix=250ms) in 99.996% D_2O on ^{13}C-RRE-suc-Rev$_{34-50}$-AAAAR-am complex. The C1' resonances in this slice are indicated with dashed lines. Peak **c** is a cross peak not resolvable in the 2D NOESY (see text). **(B)** 2D slice from a 4D HMQC-NOESY-HSQC experiment (mix= 150ms) on ^{13}C-RRE-D-Rev$_{34-50}$-AAAAR complex. This slice contains all NOEs to U72 H3' and an unusual NOE is boxed.

proton resonances that reside within a narrow 1 ppm chemical shift range. In order to identify as many NOEs as possible, a 4D HMQC-NOESY-HSQC experiment was performed on uniformly ^{13}C-labeled RRE RNA complex. This experiment not only further resolves crosspeaks from the 3D NOESY, but makes identification of the NOEs more straight-forward since the crosspeak contains the chemical shift of both carbons attached to protons involved in the NOE. Due to multiple INEPT delays and the relatively short proton and carbon T2's of the RRE RNA, the 4D experiment had poor signal to noise and only strong to medium NOEs were present in the spectrum. Nevertheless, some important NOEs were obtained that could not be obtained from other NOESY experiments due to spectral overlap. Figure 4.4B shows one HSQC plane from the 4D-experiment containing an unusual ribose-ribose NOE from U72-H3' to G71-H4'.

3D NOESY experiments were also performed on the selectively labeled RNA samples. While significantly better than the 2D homonuclear NOESY experiment, the 3D NOESY experiment with uniformly ^{13}C-labeled RNA still contains substantial overlap, particularly for the ribose resonances. With the selectively labeled samples, roughly one fourth of the resonances are selected in the F2 and F3 dimensions, significantly reducing spectral overlap and allowing for identification of more NOEs (data not shown).

Isotopic filtering NOESY experiments A significant advantage to specifically labeling RNA with only one type of ^{13}C-nucleotide is that isotopic filtering experiments can be performed to identify interresidue NOEs. In a double-half-filtered (dhf) NOESY spectrum, the conventional NOESY spectrum is divided into four different subspectra, filtering for or against protons bonded to ^{13}C in each dimension (Otting and Wuthrich, 1989; Otting and Wuthrich, 1990). One subspectrum of the sequential pathway region from a dhf-NOESY of ^{13}C-RRE complex specifically labeled at guanosines is shown in Figure 4.5A. In this subspectrum, one proton involved in the NOE is bonded to ^{12}C (F1 axis) while the other is

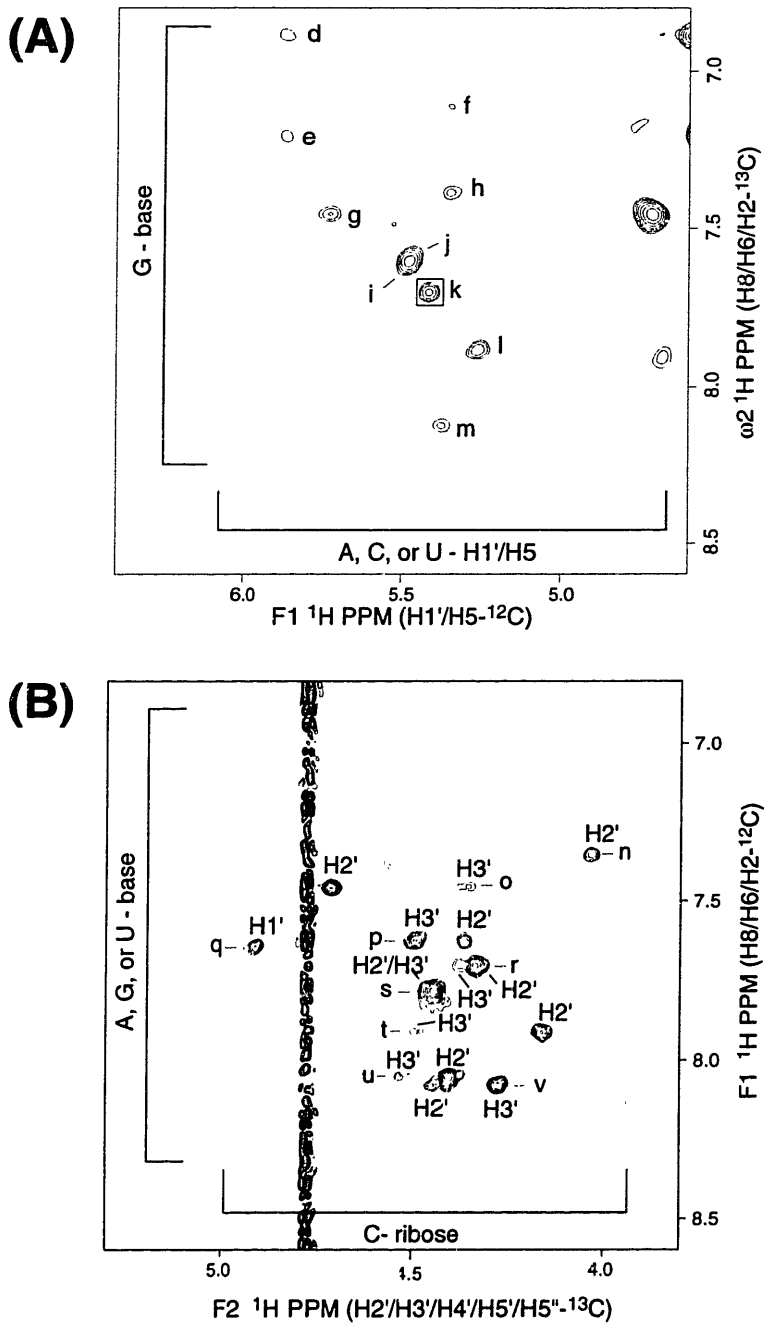


Figure 4.5: Double-half-filtered (dhf) NOESY experiments in 99.996% D_2O.
(A) Base-H1' region of the F1-edited,F2-filtered subspectrum on ^{13}C-RRE-suc-Rev34-50 -AAAAR-am complex selectively labeled at guanosines (mix=300ms). The peaks are (proton desiginated only if not H8 or H1'): (d) G76-A75, (e) G53-A52, (f) G77-C78 H5, (g) G50-C49, (h) G70-C69, (i) G67 - U66, (j) G55-C54, (k) G71-A73, (l) G64-C65 H5, (m) G46-U45. **(B)** Base-ribose region of the F1-filtered,F2-edited subspectrum on ^{13}C-RRE-D-Rev_{34-50}-AAAAR complex selectively labeled at cytosines (mix=70ms). The peaks are (H8/H6 unless noted; ribose identity in figure): (n) G70-C69, (o) G50-C49, (p) G55-C54, (q) A73 H2-C74 H1', (r) U45-C44, (s) U66-C65, (t) A75-C74, (u) A52-C51, (v) A57-C56.

bonded to ^{13}C (F2 axis). The region shown is the same as in Figure 4.3, and all crosspeaks are between guanosine H8 protons bonded to ^{13}C and H1'/H5 protons bonded to ^{12}C, which must be internucleotide NOEs. Identification of NOEs as inter- or intra-nucleotide was important for assignment of the internal loop region of the RNA where an ambiguous set of sequential NOEs were observed. For instance, A73 H1' exhibited three NOEs to H8/H6 protons, one intranucleotide and two internucleotide (see Figure 4.3). The "reverse" i,i+2 NOE between G71 H8 to A73 H1' (peak **k**) can be identified from this double-half-filtered subspectrum, since it is the only NOE of the three that involves a guanosine H8. The other two crosspeaks, which are a sequential i,i-1 internucleotide NOE from C74 H6 to A73 H1' and an intranucleotide NOE from A73 H8 to A73 H1', are absent from this double-half-filter subspectrum, since all the protons involved in the NOEs are bonded to ^{12}C. Instead, these NOEs were observed in a different double-half-filter subspectrum which only contains NOEs between protons that are both bonded to ^{12}C (data not shown).

In addition to assigning the RNA resonances, the isotopic filtering experiments can help identify more NOEs in crowded regions of the NOESY spectrum. An example of this is shown in Figure 4.5B, which is a subspectrum of a dhf-NOESY experiment performed at a short mixing time on ^{13}C-RRE complex selectively labeled at cytosines. This base to ribose region of the spectrum, which is very crowded in a regular 2D-NOESY, shows the strong to medium interresidue base-H2'/H3' NOEs characteristic of A-form helices. Again, this particular subspectrum (^{13}C-F_2,^{12}C-F_1) identifies interresidue NOEs, and the 7 helical cytosine base to 5'-non-cytosine H2'(strong) and H3'(medium) NOEs can be unambiguously assigned. In the corresponding region of the (^{12}C-F_2,^{13}C-F_1) subspectrum, the analogous non-cytosine base to 5'-cytosine H2' and H3' NOEs can also be identified (data not shown).

Identification of minor RNA conformer In the process of assigning the RRE RNA, several additional peaks were observed in NOESY spectra of the complex that did not conform with the assignments obtained for the RNA or peptide. It was determined that these "extra" peaks were due to slow exchange between conformations for a handful of the RNA resonances. Crosspeaks that are not due to NOE transfer can arise in NOESY spectra, because of conformational exchange on the millisecond time second during the mixing time. These "exchange" peaks were distinguished from NOEs by the opposite sign of the crosspeaks in a ROESY spectrum (data not shown) (Bax and Davis, 1985b; Wagner, 1989). In a ROESY experiment, NOEs are positive for large molecular weight compounds, which results in a "negative" peak (with respect to the diagonal) in the 2D spectrum. Exchange peaks, however, result in a "positive" peak (with respect to the diagonal), since they simply represent a switching of the diagonal states during the mixing time. The conformational exchange in the bound RRE was very local, including resonances A68 H2, A52 H8, A52 H1', C69 H1', and C69 H5 in the upper stem. From the intensity of one resolved diagonal peak of the minor conformation, it was estimated that the relative percentages of the major and minor conformations of the RRE-Rev complex are 90-95% and 5-10%, respectively . The chemical shifts of the minor conformation do not coincide with free RNA; therefore, the exchange must be with another RNA conformation with the peptide bound. In the major conformation, A68 is clearly unstacked, and we believe that the minor conformer may involve stacking of A68 into the upper stem due to the nature of the chemical shift changes of the exchange peaks. Mutation of A68 to U68 results in a complex whose NOESY spectra are nearly identical to wild-type except for the absence of the exchange peaks (H. Mao and J. Williamson, unpublished results). This is consistent with the hypothesis that the exchange peaks result from stacking, since single nucleotide pyrimidine bulges have a much lower propensity to stack into the helix than single nucleotide purine bulges.

4.1.3 Summary of NMR Data The non-exchangeable proton and carbon assignments for the RRE RNA bound to the Rev peptide are given in Table 4.1. All base, H1', H2' and most of the H3'/H4'/H5'/H5" protons were assigned. In the helical regions of the RRE there were NOEs typical of A-form RNA helices, except for the internal loop region. The pattern of NOEs for the internal loop region of RRE is shown schematically in Figure 4.6. A normal H8/H6 to H1' sequential pattern of NOEs was observed from residues G46 to G48 in the internal loop; however, a much more extensive network of NOEs was observed from G70 to C74. The two unusual "reverse" sequential H8 to H1' NOEs are highlighted by bold lines in Figure 4.6. In addition, there are several unusual ribose to ribose NOEs between G71-G70 and G71-U72, which are also highlighted. While there are no internucleotide NOEs involving the H6 and H1' protons of the bulged U72, there is an NOE from A73 H8 to U72 H4' which positions the ribose of U72 in the major groove. In addition, there is an unusual sequential ribose-ribose NOE from U72 H3' to G71 H4'. The structural features that give rise to these unusual NOEs will be discussed in Chapter 5.

There are no intranucleotide H8/H6 to H1' NOEs observed in a 50 ms NOESY experiment, indicating that all glycosidic torsions are in the *anti* conformation. In another study, the G71 H8 to G71 H1' NOE was identified as strong (<3.0 Å), suggesting G71 was in the *syn* conformation (Peterson et al., 1994). Our assignments for the base, H1', and H2' protons are nearly identical to their assignments, except for the G71 and A73 H1's. The unusually upfield shifted G71 H1' identified by the HSQC-CT experiment (section 4.1.1) was not identified in their homonuclear NMR analysis, and we believe that these resonances were misassigned. In addition, the putative strong intranucleotide G71 H8-G71 H1' NOE crosspeak at 7.70/5.41 ppm, indicative of a *syn* conformation, is present in a (^{13}C-F_2,^{12}C-F_1) subspectrum of the filtered NOESY spectra on ^{13}C-RRE selectively labeled at guanosines (Figure 4.5A; peak **j**). The presence in this subspectrum indicates that this strong crosspeak represents an **inter**nucleotide NOE (G71 H8 - A73 H1'

Table 4.1: Peptide-bound RRE RNA nonexchangeable proton and carbon chemical shifts. Error in chemical shift ± 0.01,0.20 ppm for proton and carbon, respectively. Proton shifts are referenced to internal TSP. Carbon chemical shifts are referenced to an external TSP sample.

residue	H8,H6 /C8,C6	H5,H2/ C5,C2	H1'/C1'	H2'/C2'	H3'/C3'	H4'/C4'	H5'/H5"/C5'
G41	8.17/139.3	-	5.84/90.5	4.96/74.8	4.73/74.5	4.55/83.5	4.26,4.40/67.1
G42	7.62/136.9	-	5.96/93.1	4.70/75.5	4.49/72.3	4.54/82.4	4.33,4.49/65.7
U43	7.71/140.1	5.56/104.6	5.56/93.4	4.27/75.7	4.56/(72.3)	4.45/(81.6)	4.05,4.51/63.6
C44	8.03/142.1	5.85/97.3	5.68/93.7	4.33/75.2	4.36/72.3	4.48/81.2	4.15,4.56/65.0
U45	7.71/140.9	4.53/102.8	5.38/95.1	4.22/75.0	4.46/71.6	4.34/82.6	4.05,4.70/63.6
G46	8.14/138.9	-	5.29/85.1	4.38/78.5	4.49/77.1	4.60/86.1	4.13,4.39/67.8
G47	8.54/141.4	-	5.87/92.5	4.15/75.7	5.15/73.9	4.44/82.9	
G48	7.97/138.4	-	5.77/87.7	4.86/77.6	4.15	4.53/84.7	
C49	7.69/142.1	4.95/97.9	5.74/94.7	4.72/75.1	4.34/73.3	4.59/82.3	4.01,4.35/65.7
G50	7.45/135.2	-	5.83/92.9	4.52/75.1	4.55/74.7	4.49/81.7	4.15,4.51/64.3
C51	7.48/139.9	5.28/97.5	5.28/93.8	4.43/74.9	4.52/71.4	4.38/81.7	4.07,4.54/63.6
A52	8.03/139.4	7.04/152.4	5.87/92.5	4.57/75.7	4.76/72.0	(4.25)	
G53	7.22/135.6	-	5.61/92.4	4.28/75.4	4.40/(76.1)	4.12/83.7	4.05,4.47/65.0
C54	7.27/139.6	5.04/97.6	5.48/93.7	4.49/75.7	4.36/75.6	(4.07)	4.05,4.44/65.0
G55	7.57/136.5	-	5.77/93.3	4.53/76.1	4.64/73.2	4.39/82.3	4.10,4.40/65.0
C56	7.78/143.4	5.63/97.8	5.54/92.4	4.43/75.7	4.24/74.8	4.11/83.2	3.91,4.17/65.3
A57	8.09/140.7	7.90/155.4	5.65/92.1	4.35/76.9	4.60/76.0	4.23/83.6	3.79,3.89/66.4
A58	8.28/141.5	8.10/154.8	6.07/92.1	4.69/77.0	5.28	4.50/82.6	4.47/66.4
G64	7.88/137.5	-	3.75/92.8	4.33/74.8	4.18/74.2	4.33/83.0	4.26,4.38/69.8
C65	7.73/135.4	5.26/96.8	5.55/94.0	4.45/75.0	(4.06)	(4.45)/(81.6)	4.04,4.59/65.0
U66	7.77/141.4	4.99/103.2	5.48/94.0	4.41/75.1	(4.35)	4.39/81.6	4.05,4.42/65.0
G67	7.63/136.8	-	5.87/91.0	4.21/78.2	4.75/74.6	(4.40)	4.09,4.62/65.0
A68	8.41/142.7	8.30/155.7	6.18/90.1	4.90/75.8	5.02/75.3	4.41/84.3	4.17,4.19/65.7
C69	7.44/143.3	4.85/96.1	5.29/95.6	4.04/75.7	4.17/(73.3)	4.43/83.2	
G70	7.32/134.6	-	5.93/91.1	4.47/76.4	(4.45)	4.36/81.7	4.17,4.40/65.0
G71	7.70/137.7	-	4.78/87.4	3.93/79.1	4.51	4.16/86.5	4.03,4.33/67.1
U72	8.07/144.4	5.97/105.5	6.24/89.3	4.46/75.4	4.29/78.8	4.86/84.5	4.01,4.26/68.5
A73	7.94/141.1	7.63/152.4	5.41/93.4	4.69/75.1	4.31/(73.3)	4.54/82.0	4.14,4.35/65.0
C74	7.45/139.6	5.24/97.1	4.91/93.2	4.18/74.9	4.47/71.2	4.31/81.5	4.05,4.38/63.6
A75	7.89/139.3	7.24/152.6	5.88/92.2	4.59/76.0	4.70/72.0	(4.40)/(81.4)	4.56/65.0
G76	6.89/135.8	-	5.45/92.4	4.32/75.7	4.25/74.2	4.38/82.0	3.98,4.33/66.4
G77	7.13/136.6	-	5.74/93.1	4.66/75.1	4.33/72.6	4.44/82.0	4.01,4.40/65.7
C78	7.67/140.8	5.34/97.1	5.57/94.1	4.24/75.7	4.46/(76.1)	(4.16)	4.05,4.17/65.7
C79	7.71/141.9	5.53/97.9	5.81/92.9	4.05/77.6	4.18/69.8	4.43/83.2	4.05,4.49/65.0

() indicates tentative assignment.

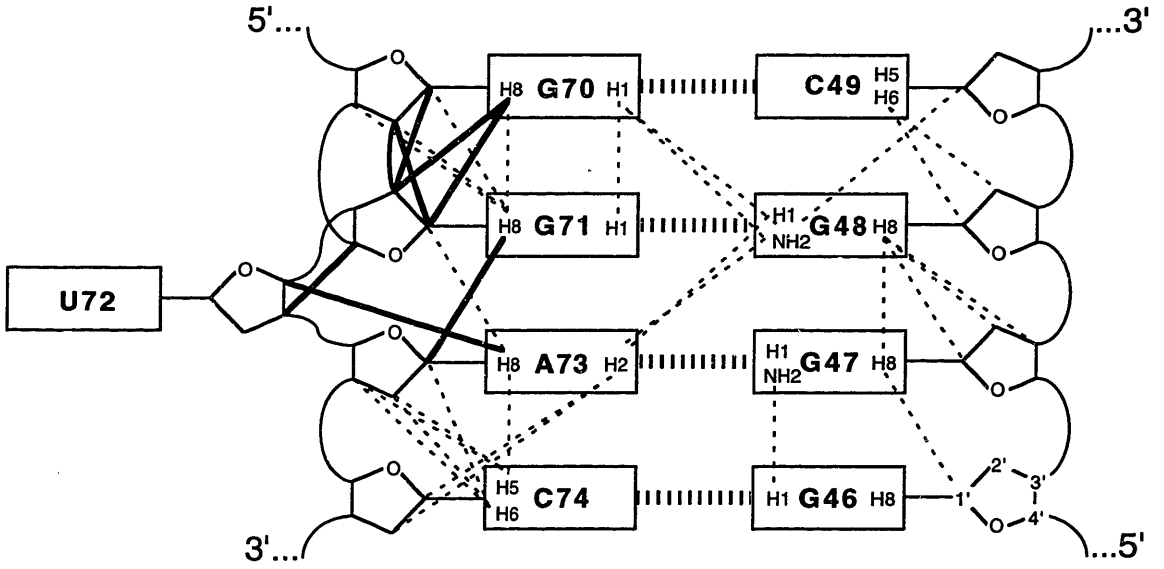


Figure 4.6: Schematic of internal loop RNA-RNA NOEs. Dashed and bold lines indicate NOEs and particularly unusual NOEs, respectively. NOEs are drawn between the labeled base proton and the corresponding position on the ribose ring, as illustrated for G46.

in our assignments). The misassignments highlight the importance of heteronuclear NMR and the selective labeling strategy for obtaining the assignments of the RNA. The intranucleotide H8-H1' crosspeak we have identified as G71 is not present at 50 ms. In addition, this nucleotide has a strong H8-H2' NOE at 50 ms, consistent with a nucleotide in an *anti*/C2'-endo conformation (see Materials and Methods). Intranucleotide base to ribose NOE intensities consistent with an *anti*/C2'-endo conformation were also observed for the other nucleotide in the G:G base pair (G48). Since the symmetric G48:G71 base pair in RRE would be expected to have one of the nucleotides in a *syn* conformation (Saenger, 1984; Bartel et al., 1991; Iwai et al., 1992), some unusual structure must be present to allow this base pair to form with both nucleotides in an *anti* conformation, and the resulting geometry might be critical for specific recognition by the peptide (see Chapter 5).

4.2 Assignment Procedure For Peptide

4.2.1 Through-bond experiments Assignment of the peptide was achieved using triple resonance through-bond correlation experiments on $^{13}C,^{15}N$-labeled D-Rev_{34-50}-AAAAR peptide bound to unlabeled RRE RNA. These include published triple resonance experiments such as H(C)(CO)NH-TOCSY (Montelione et al., 1992), (H)C(CO)NH-TOCSY (Logan et al., 1992), and HNCACB (Wittekind and Mueller, 1993; Muhandiram and Kay, 1994). The H(C)(CO)NH-TOCSY and (H)C(CO)NH-TOCSY experiments correlate the proton and carbon resonances, respectively, of each side chain to the backbone amide of the preceding (i-1) residue, providing NOE-independent sequential assignment. The HNCACB experiment correlates the α and β carbon resonances of each side chain to its own and preceding (i-1) backbone amide protons. Assignment of the long aliphatic arginine side chain resonances with these experiments was incomplete due to the overlap of

some of the arginine backbone amide proton and nitrogen frequencies. Therefore, a novel experiment (2D Arg-C(CC)-TOCSY-Nε-Hε (N.R. Sambasiva, D.R. Muhandiram & L.E. Kay, unpublished results)) was performed to correlate all of the arginine side-chain carbons and protons to the epsilon proton resonances, which are well-resolved in the Rev-RRE complex (Figure 4.7A-C). This experiment, in combination with the other triple resonance experiments, permitted full assignment of the Rev peptide, including all 11 arginine side chains (Table 4.2).

4.2.2 NOESY experiments NOEs defining the conformation of the peptide were obtained from a 3D simultaneously $^{13}C,^{15}N$-edited NOESY-HSQC (CN-NOESY) experiment (Pascal et al., 1994) on the $^{13}C,^{15}N$-peptide sample. In this experiment the NOEs for both the ^{15}N- and ^{13}C-attached protons are resolved into a third dimension in one experiment. Figure 4.8 shows one slice of the CN-NOESY with the well resolved NOEs to two β-protons and several arginine ε-protons. Figure 4.9 summarizes the NMR data obtained for the backbone of the peptide, which clearly shows NOE patterns consistent with the formation of an α-helix. α->NH(i,i+3) and α->β(i,i+3) backbone NOEs characteristic of an α-helix were observed from Gln^{36} to Arg^{55}, except for a few gaps due to overlap of α-carbon and α-proton resonances. The overlap was for the connectivities Arg^{35}->Arg^{38}, Arg^{38}->Arg^{41}, and Arg^{39}->Arg^{42}. However, other non-backbone NOEs consistent with an α-helix for all of these amino acids were observed. Small HN-Hα coupling constants (~4 Hz) typical of an α-helical φ-conformation (~-60°) were determined from an HNHA experiment for Arg^{35}-Arg^{55}, except for Arg^{39} and Arg^{41}, whose coupling constants could not be measured due to spectral overlap (data not shown). The deviations of the alpha carbon and proton chemical shifts from random coil (Figure 4.10) were also consistent with an uninterrupted α-helix from Arg^{35}-Arg^{55} (Wishart et al., 1992; Wishart and Sykes, 1994; Wishart et al., 1995).

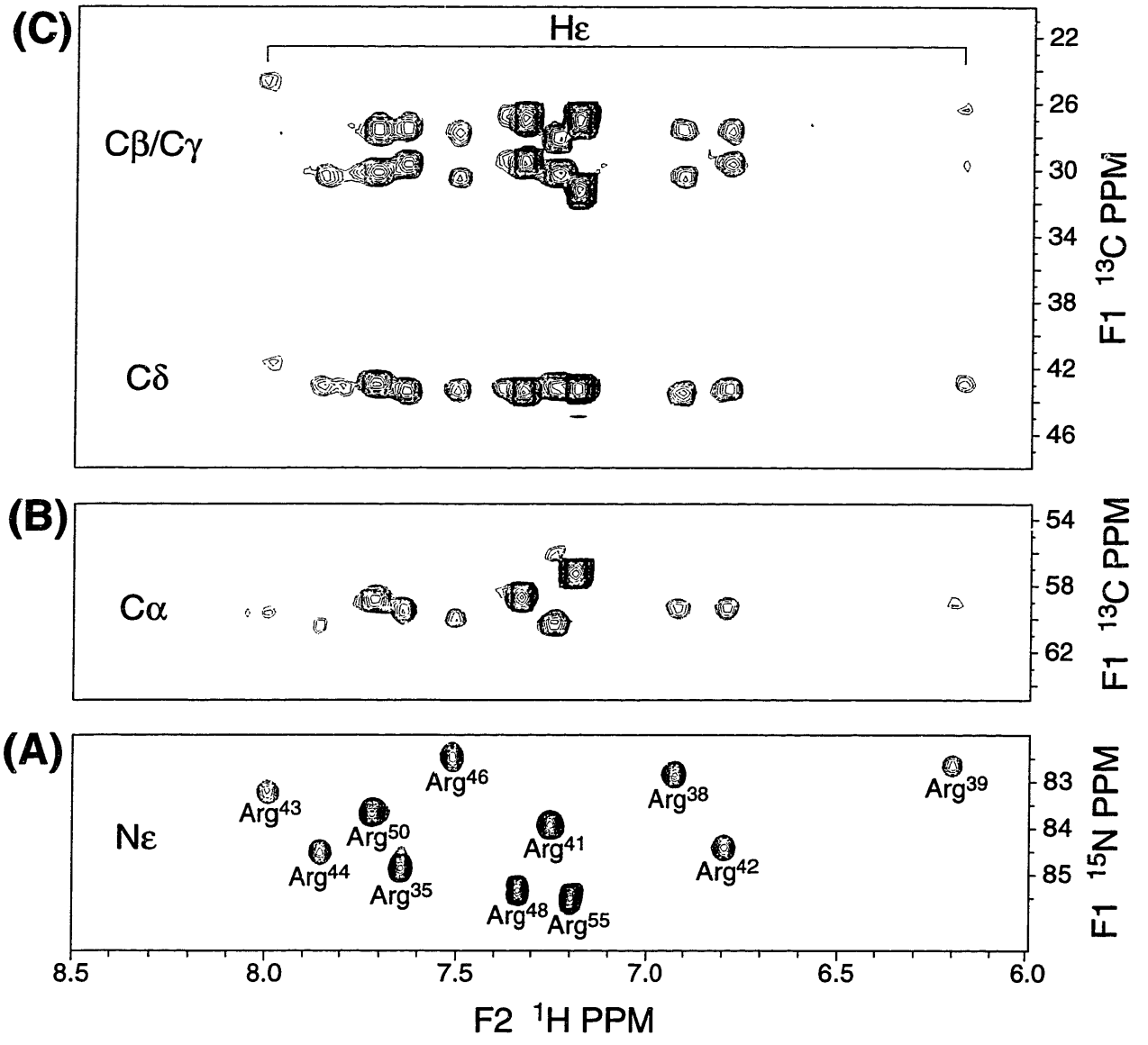


Figure 4.7: Through-bond correlation experiments with ^{13}C,^{15}N-labeled Rev-RRE complex in 90/10% H_2O/D_2O to assign arginine side chains. **(A)** Epsilon region of ^{15}N-HSQC spectrum showing the 11 well-resolved Nε-Hε correlations of the Rev peptide. **(B)** long mixing time, and **(C)** short mixing time of an Arg-C(CC)-TOCSY-Nε-Hε experiments correlating the epsilon protons to all the carbon resonances of the arginine side chain.

Table 4.2: RNA-bound peptide proton and heteronuclear chemical shifts. Error in chemical shift ± 0.01,0.10 ppm for proton and carbon/nitrogen, respectively.

Residue	NH/N	Hα/Cα	Hβ/Cβ	Hγ/Xγ	Hδ/Xδ	Hε/Xε
Asp33		4.50/52.7	2.80/40.1			
Thr34	8.12/111.3	4.70/60.1	4.76/71.7	1.39/22.4		
Arg35	9.14/121.7	3.89/59.7	1.92,1.78/29.7	1.67,1.52/27.7	3.24,3.10/43.6	7.64/84.8
Gln36	8.82/121.3	3.86/59.4	2.20/28.0	2.71,2.34/35.1		7.46,6.80/111.5
Ala37	8.24/124.6	4.02/55.1	1.73/18.7			
Arg38	8.18/118.5	3.89/59.5	1.96,1.79/30.6	1.44/27.8	3.35,3.16/43.5	6.93/82.9
Arg39	8.03/121.3	3.86/59.3	1.90,(1.77)/30.0	1.58,1.39/26.5	2.94,2.74/43.0	6.20/82.7
Asn40	8.20/118.43	4.41/56.5	2.56,2.30/38.8		7.92,6.14/116.2	
Arg41	8.03/121.2	3.87/60.4	1.94,1.80/30.4	1.78,1.55/28.3	3.40,3.16/43.4	7.26/84.0
Arg42	8.15/120.3	4.12/59.5	1.89/29.7	1.71,1.48/27.7	2.95,2.77/43.2	6.80/84.5
Arg43	7.99/121.4	4.01/59.7	2.14,1.90/28.2	1.85,1.63/24.9	3.50,3.25/41.7	7.99/83.2
Arg44	7.94/118.8	4.08/60.6	2.07,1.92/30.7	1.85,1.54/29.8	3.62,3.09/43.2	7.86/84.5
Trp45	8.22/121.6	4.36/60.8	3.61,3.40/28.5	aromatic ring:	NH-10.15/130.0 Hδ1-7.32/126.5 Hε3-7.64/120.1 Hζ2-7.48/114.3 Hζ3-7.14/121.4 Hη2-7.26/124.1	
Arg46	8.63/120.1	3.78/60.1	2.19,1.83/30.5	1.53/28.0	3.26,3.12/43.5	7.51/82.4
Glu47	8.32/120.24	3.97/58.9	2.23,2.15/29.1	2.51,2.34/35.2		
Arg48	7.93/120.9	4.13/58.9	1.95/29.6	1.85,1.69/27.1	3.29,3.21/43.6	7.34/85.5
Gln49	7.86/118.7	3.87/57.6	1.92,1.79/27.9	1.74,1.69/32.8		6.59/112.5
Arg50	7.94/120.5	4.04/59.0	1.86/30.3	1.75,1.52/27.8	3.18/43.1	7.72/83.6
Ala51	7.89/122.0	4.17/53.9	1.49/18.2			
Ala52	7.75/120.3	4.17/53.4	1.43/18.3			
Ala53	7.63/120.5	4.28/52.7	1.47/18.6			
Ala54	7.69/121.9	4.39/52.0	1.45/19.1			
Arg55	7.58/125.3	4.17/57.3	1.88,1.77/31.2	1.68/27.0	3.22/43.3	7.20/85.6

() indicates tentative assignment.

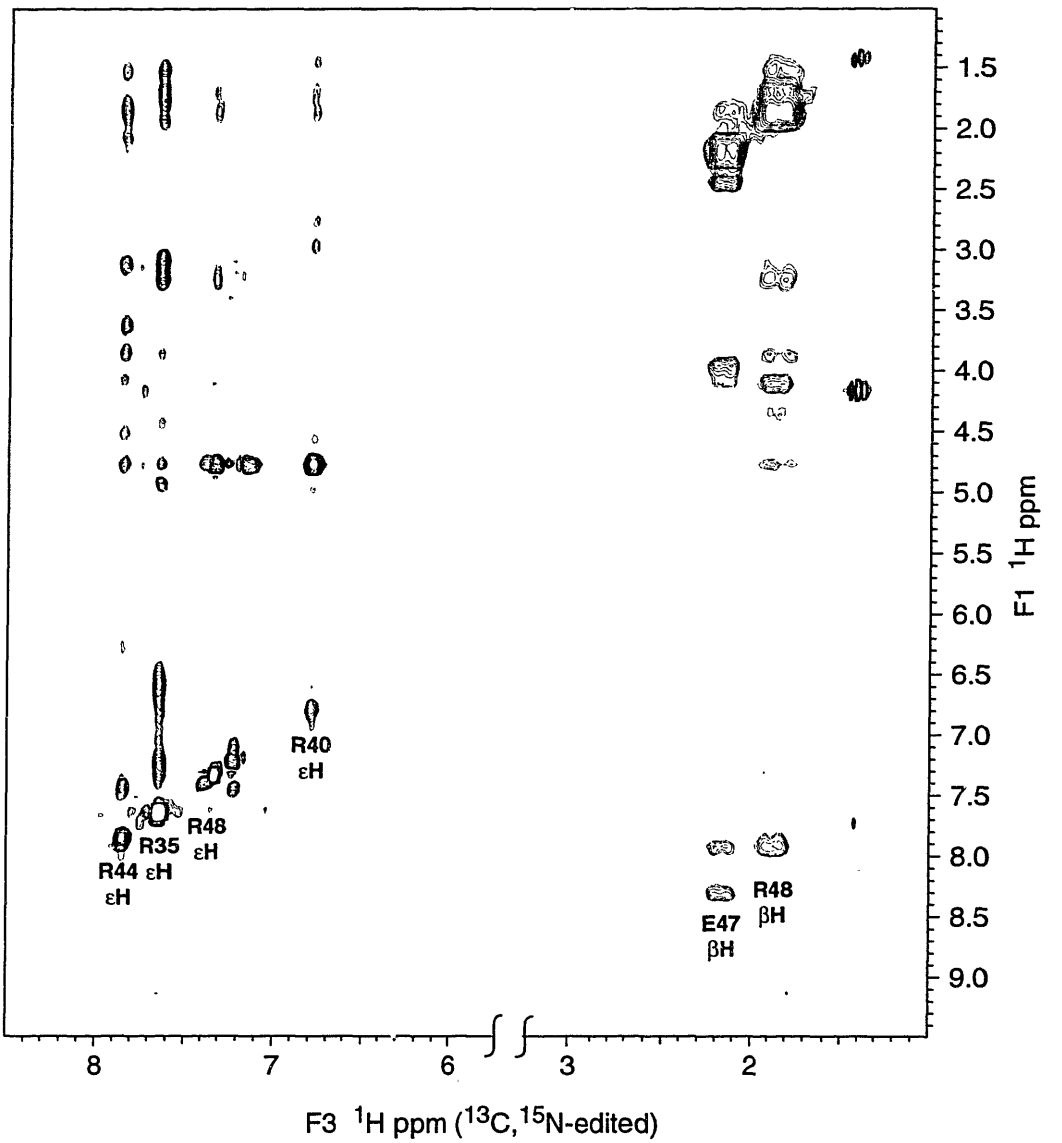


Figure 4.8: 2D slice of a CN-NOESY experiment (mix=150ms) on $^{13}C,^{15}N$-labeled Rev-RRE complex. The F3 dimension is simultaneously edited for protons attached to ^{13}C or ^{15}N. In this slice, NOEs to epsilon (^{15}N) and beta (^{13}C) peptide protons are observed.

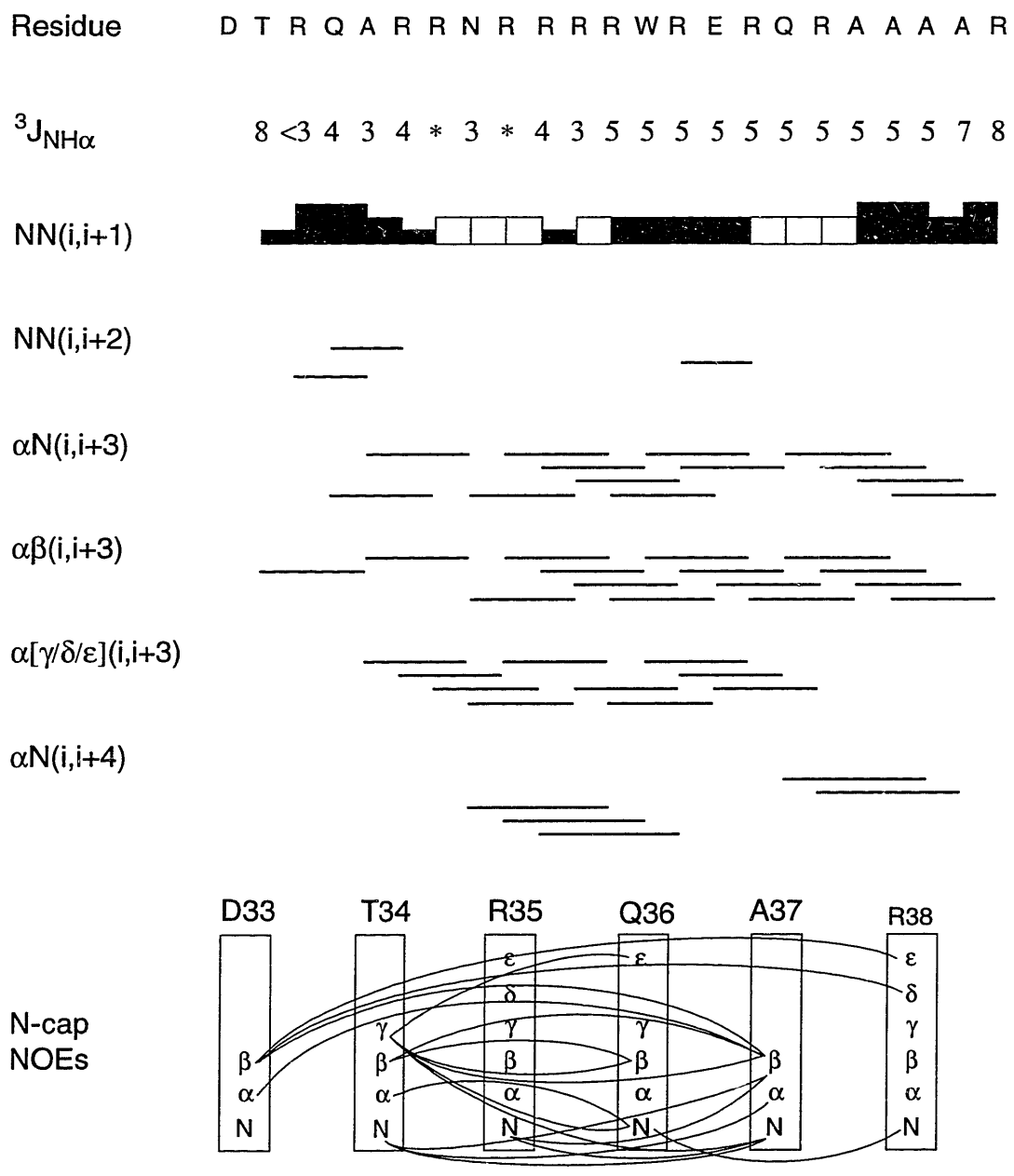


Figure 4.9: Summary of peptide NMR data. $^3J_{NH\alpha}$ is the coupling constant (Hz) for the backbone amide to alpha proton. The remainder of the figure contains NOE connectivities. N,α,β,... denote amide, alpha, beta, etc. protons of each residue, while the letters in parentheses indicate the relative position of the two protons along the peptide sequence. [γ/δ/ε] indicates any proton of the side chain other than the beta proton. For the sequential amide-amide connectivities, the size of the black boxes represent strong, medium and weak NOE intensity. The open boxes represent connectivities that could not be observed due to spectral overlap. At the bottom is a schematic of non-standard NOEs observed at the N-terminus (only non-sequential NOEs shown).

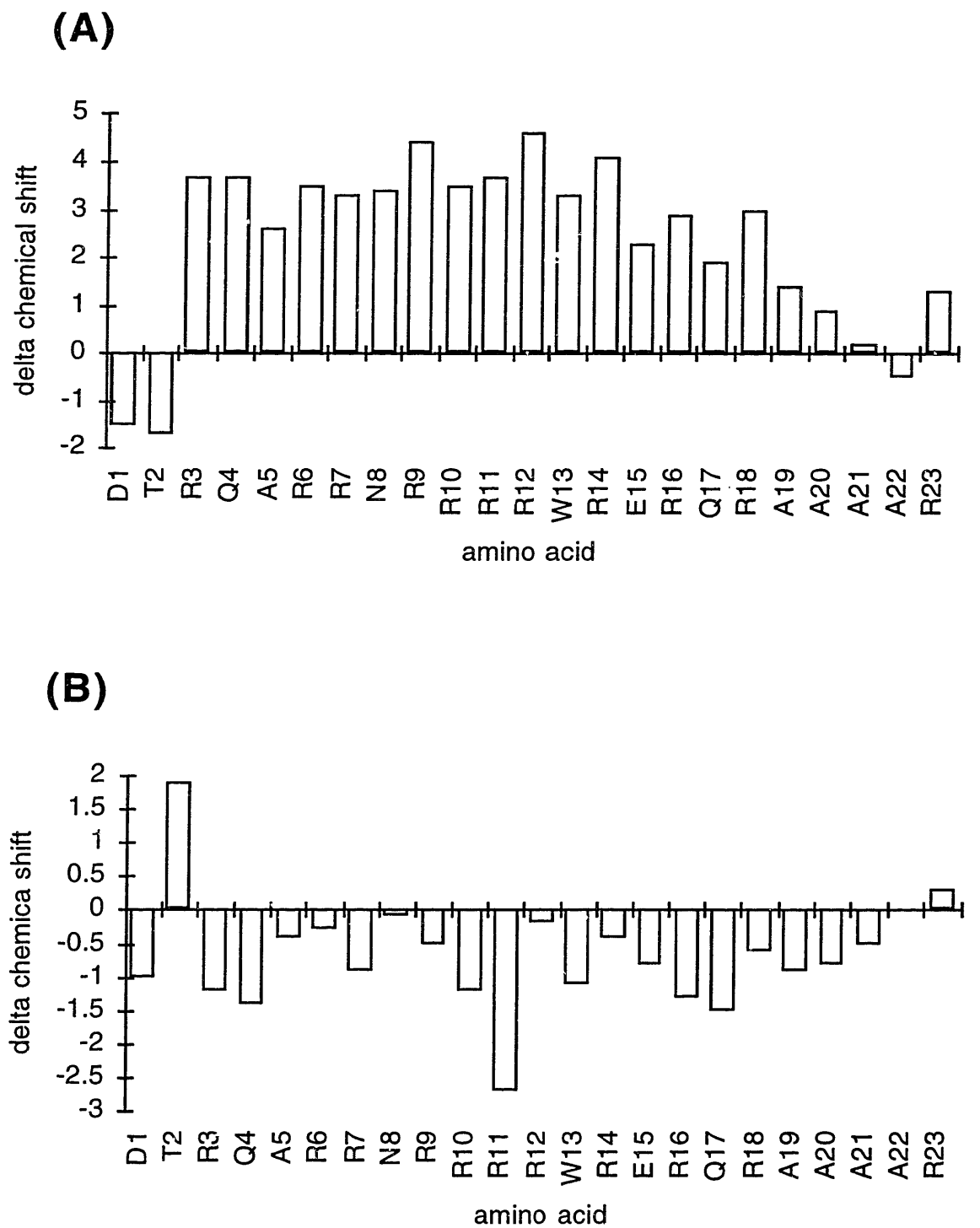


Figure 4.10: Deviation of ^{13}C chemical shift of bound Rev peptide from random coil values. **(A)** Chemical shift difference of α-carbon resonances. **(B)** Chemical shift difference of β-carbon resonances. Random coil values of carbon resonances for each amino acid taken from Wishart et al., 1994.

Many non-helical NOEs were observed near the N-terminus of the peptide (Figure 4.9). In particular, the NOEs for Thr^{34} suggested that this amino acid was involved in formation of an "N-cap" structure (Harper and Rose, 1993). N-caps are structures that help stabilize the N-terminus of α-helices by providing hydrogen bonds from side chain groups to backbone amides that would normally not be hydrogen bonded in a α-helix. The unusual NOEs at the N-terminus of Rev have been observed in other NMR studies of peptides and proteins containing N-cap structures (Lyu and Wemmer, 1993; Lee et al., 1994). Furthermore, the Cα and Cβ chemical shifts of Thr^{34} show deviations from random coil (-1.7 and +1.9 ppm, respectively) that are characteristic of an N-cap amino acid (Gronenborn and Clore, 1994) (Figure 4.10). These ^{13}C chemical shifts are sensitive to backbone dihedral angles, and the N-cap amino acid has distinct ϕ/φ angles from an α-helix (Harper and Rose, 1993). The $^{3}J_{HN\text{-}H\alpha}$ coupling constant measured for Thr^{34} (8 Hz) is consistent with a ϕ angle in the range observed for N-cap amino acids (~-150°).

4.3 Identification of RNA-Peptide NOEs

4.3.1 Assignment by unambiguous chemical shift ranges Once full assignment of both the RNA and peptide are obtained, intermolecular RNA-peptide NOEs can be identified in certain regions of the 3D NOESY experiments with both labeled peptide and RNA. For instance, there are no RNA protons with chemical shifts below 3.5 ppm; therefore, in a 3D NOESY-HMQC with ^{13}C-labeled RNA, any crosspeaks in F1 below 3.5 ppm must be NOEs to the peptide. An example of this is shown in Figure 4.11A, which contains a slice corresponding to the chemical shift of G67 H8. The crosspeak at 3.20 ppm must be a peptide proton, since it is below 3.5 ppm, and was assigned to Arg^{35} δH. This assignment can be verified in the corresponding ^{13}C-slice of the peptide proton in the CN-NOESY experiment with labeled peptide (data not shown). Analogously, no peptide

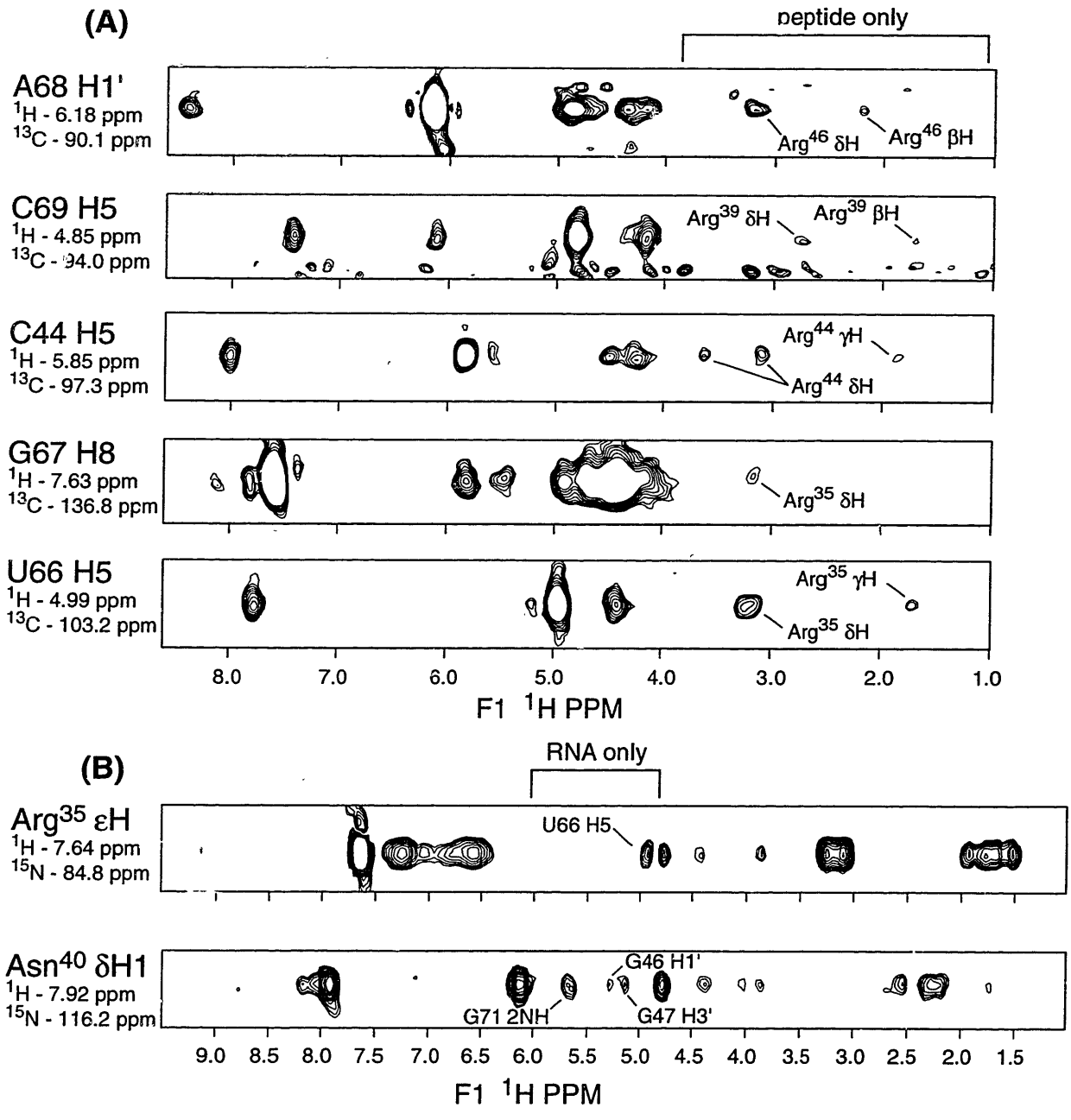


Figure 4.11: Strips from 3D NOESY experiments showing RNA-peptide NOEs identifable from non-overlapping chemical shift. **(A)** Strips (F3/F2) from a 3D NOESY-HMQC experiment (mix=250ms) on ^{13}C-labeled RRE-suc-Rev$_{34-50}$-AAAAR-am complex in 99.996% D_2O. Crosspeaks between 1.0 and 3.7 ppm arise only from peptide protons. **(B)** Strips (F3/F2) from a 3D CN-NOESY experiment (mix=150ms) on ^{13}C,^{15}N-labeled Rev-RRE complex in 90/10% H_2O/D_2O. Crosspeaks between 4.7 and 6.0 ppm only arise from RNA protons.

protons resonate between 4.8 and 6.0 ppm; therefore, resonances in this range in the CN-NOESY experiment with labeled peptide can be identified as intermolecular NOEs to the RNA. An example of this is shown in Figure 4.11B, which is the ^{15}N-slice containing the side chain amide of Asn^{40}. Three NOEs are observed at 5.15 (G47 H3'), 5.29 (G46 H1'), and 5.65 (G71 2NH). Again, all assignments of intermolecular NOEs identified by the CN-NOESY were cross-verified (spectral overlap permitting) in 3D NOESY-HSQC spectra with ^{13}C- or ^{15}N-RNA.

4.3.2 Filtered NOESY experiments The above approach only works for a limited set of NOEs, and many intermolecular NOEs are in more crowded regions. Therefore, isotopic filtering experiments that unambiguously identify intermolecular NOEs were performed. If the dhf-NOESY experiment described above is performed on uniformly ^{13}C-labeled RRE-D-Rev$_{34-50}$-AAAAR complex in 99.996% D_2O, all intermolecular NOEs between nonexchangeable protons will be in the F1-^{13}C,F2-^{12}C subspectrum (data not shown). The only drawback to this experiment is that NOEs to peptide exchangeable protons cannot be obtained; however, these can be obtained with a 3D F1-filtered,F3-edited (^{13}C/^{15}N) CN-NOESY-HSQC experiment with labeled Rev peptide. Since the peptide is double-labeled, a ^{13}C,^{15}N purge sequence (Fiura and Bax, 1992) can be added to F1 of the CN-NOESY described in section 4.2.2, removing peptide protons from this dimension. Analogously, all protons in F3 must arise through the simultaneous ^{13}C,^{15}N HSQC element (Pascal et al., 1994) for detection, removing RNA protons. The result is a 3D version of the basic dhf-NOESY experiment that contains only intermolecular NOEs. Representative strips from this spectrum are shown in Figure 4.12. Due to low sensitivity from relaxation during the filtering elements, these experiments were performed with long mixing times (200-300ms). However, these experiments provide unambiguous assignment and most RNA-peptide NOEs identified with this experiment

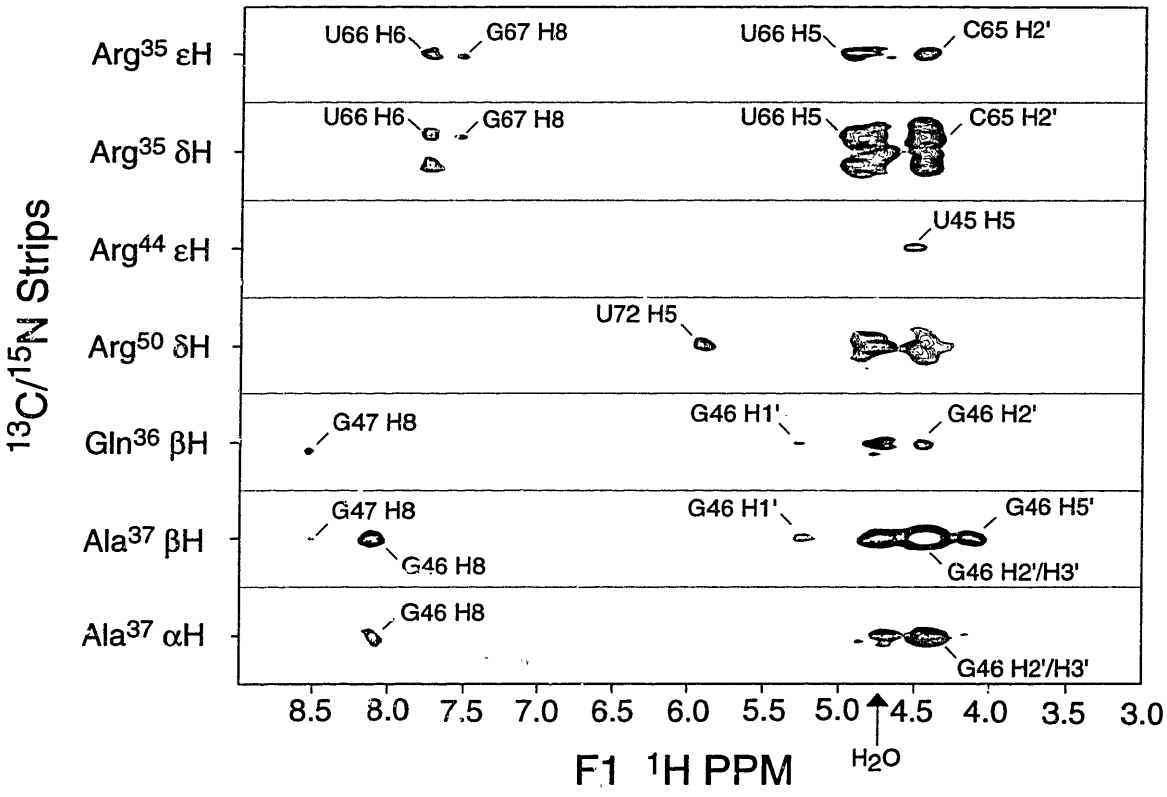


Figure 4.12: $^{13}C,^{15}N$ Strips of a F1-filtered,F3-edited($^{13}C,^{15}N$)-NOESY-HSQC experiment (mix=300ms) on $^{13}C,^{15}N$-labeld Rev peptide complex. Assignments of RNA protons in F1 are given in the figure.

could be confirmed in other NOESY experiments with labeled peptide or RNA at shorter mixing times, where the effects of spin diffusion transfer are mitigated. Overall, a total of 61 intermolecular RNA-peptide NOEs were obtained (Table 4.3), which are shown schematically in Figure 4.13. Interestingly, the intermolecular NOEs clustered into two regions of the RNA. One cluster is along nucleotides C65-U72 in the upper stem, and the other is along nucleotides C44-G48 in the lower stem.

4.4 Molecular Modeling of the Complex

The NMR data was converted into distance and dihedral restraints summarized in Table 4.4 for molecular dynamics calculations (see Appendix B for a complete listing of all structural restraints used in molecular modeling). Overall, ~1000 restraints were obtained, and the distribution of these restraints per residue is shown in Figure 4.14. The number of sequential and medium range RNA-RNA and peptide-peptide NOEs are fairly well distributed indicating no completely unstructured regions in the complex. The intermolecular RNA-peptide NOEs, however, occur in patches throughout the primary sequence as would expected for a specific interface between the RNA and the peptide.

NOE intensities were converted into four distance bound ranges (strong, medium, weak, and very weak, see section 2.3.2). The RNA base pairing hydrogen bonds were deduced from analysis of imino proton NOE patterns (Chapter 3), except that five additional hydrogen bonds not directly derived from our experiments were included for the GCAA tetraloop whose structure has been previously published (Heus and Pardi, 1991). This is conservative modeling, since all of our NOE and chemical shift data in the tetraloop region are consistent with the published tetraloop structure. Inclusion of the hydrogen bonds helps increase the convergence rates of the calculations. In addition, the tetraloop is distant from the peptide contacts and any minor inaccuracies should have no adverse effects

Table 4.3: RNA-peptide NOEs. For simplicity, NOEs are group together by residue. For instance, $Arg^{35}(\delta,\varepsilon)$ to U66-H5/H6 represents 4 NOEs: δ-H5, δ-H6, ε-H5 and ε-H6. See Appendix B for a complete listing of the entire NMR restraint set for molecular modeling.

Peptide	RNA	NOE intensity
$Arg^{35}(\delta,\varepsilon)$	G67-H8	very weak
$Arg^{35}(\delta,\varepsilon)$	U66-H5/H6	med-weak
$Arg^{35}(\delta,\varepsilon)$	C65-H2'	weak
$Gln^{36}(NH,\beta,\gamma,\varepsilon)$	G47-H3'	weak-very weak
$Gln^{36}(\alpha,\beta,\gamma)$	G47-H8	weak-very weak
$Gln^{36}(\alpha,\varepsilon)$	G71-H21/H22	very weak
$Gln^{36}(\beta)$	G46-H1'/H2'	weak-very weak
$Gln^{36}(\varepsilon)$	G48-H8	weak
$Ala^{37}(NH,\alpha,\beta)$	G46-H2'/H3	strong-med
$Ala^{37}(\alpha,\beta)$	G46-H8	weak
$Ala^{37}(\beta)$	G46-H4'/H5'/H5"	very weak
$Ala^{37}(\beta)$	G47-H8	very weak
$Arg^{39}(\delta)$	C69-H5	very weak
$Arg^{39}(\varepsilon)$	G70-H8	very weak
$Arg^{39}(\delta,\varepsilon)$	C69-H41/H42	weak-very weak
$Arg^{39}(\delta,\varepsilon)$	G71-H21/H22	very weak
$Asn^{40}(\beta)$	G46-H8	weak
$Asn^{40}(\delta)$	G46-H1'	weak
$Asn^{40}(\delta)$	G47-H3'	weak
$Asn^{40}(\delta)$	G71-H21/H22	medium
$Arg^{43}(\gamma,\delta,\varepsilon)$	U72-H5'/H5"	weak-very weak
$Arg^{43}(\gamma,\delta)$	U72-H4'	weak-very weak
$Arg^{43}(\delta)$	G71-H1'	weak
$Arg^{43}(\varepsilon)$	G71-H21/H22	weak
$Arg^{44}(\gamma,\delta,\varepsilon)$	U45-H5	med-very weak
$Arg^{44}(\gamma,\delta,\varepsilon)$	C44-H5	very weak
$Trp^{45}(\delta1,\varepsilon3,\zeta3)$	A68-H2*	very weak
$Arg^{46}(\delta,\varepsilon)$	A68-H1'	weak
$Arg^{50}(\delta,\varepsilon)$	U72-H5	weak

*These NOEs were only observed at 15°C.

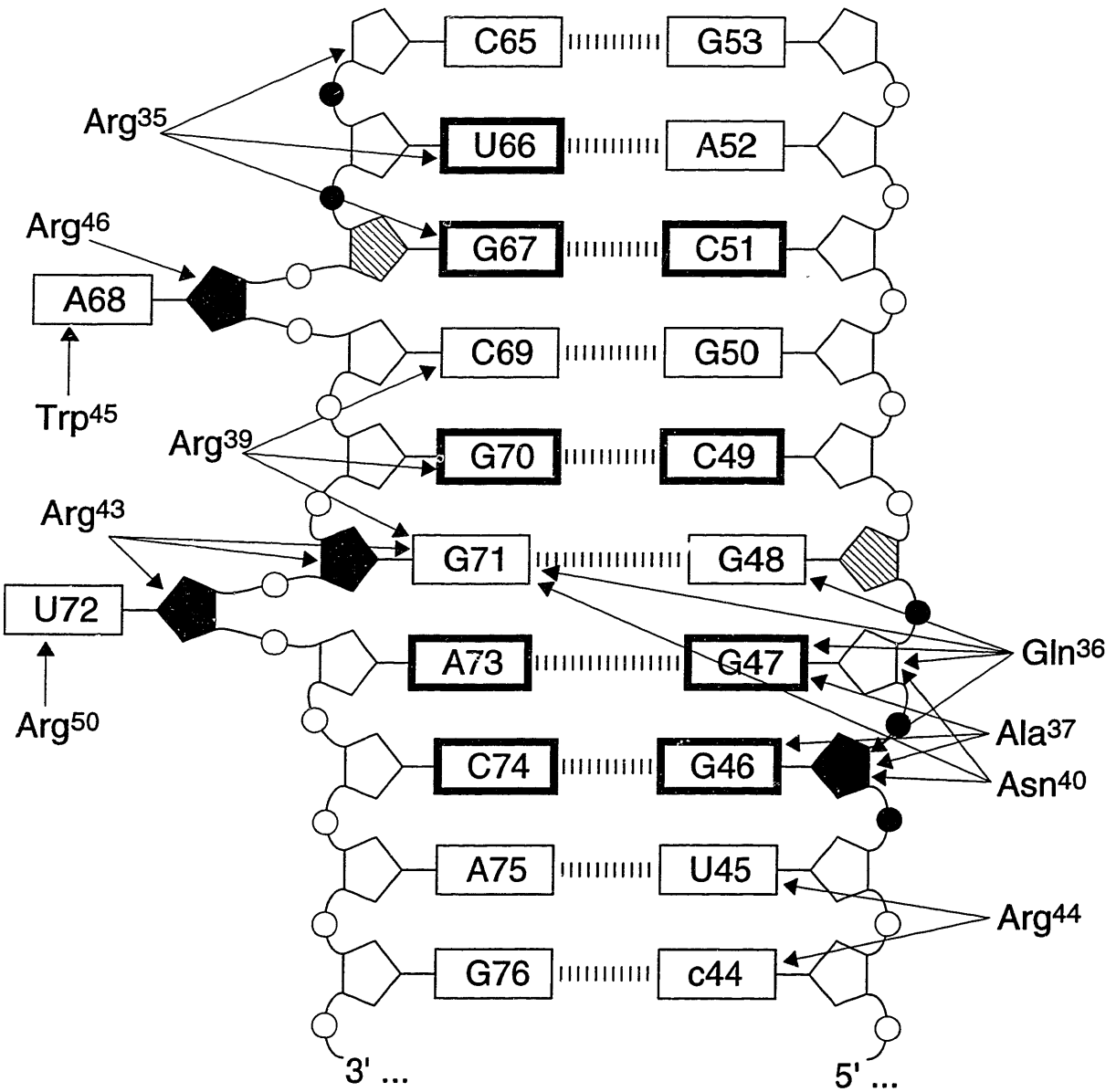


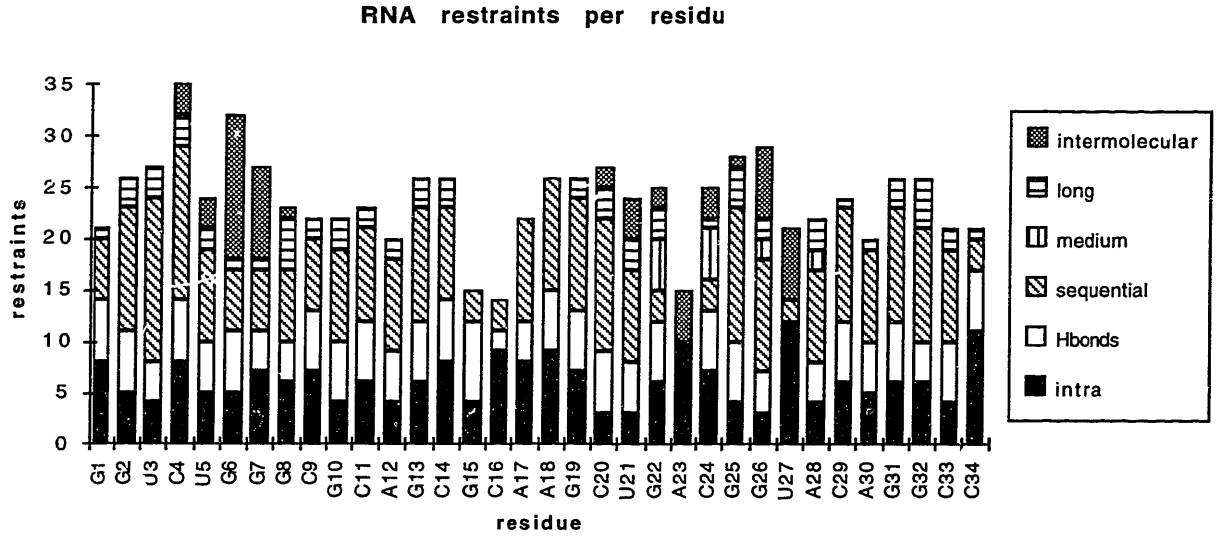
Figure 4.13: Schematic of observed RNA-peptide NOEs. Arrows depict NOEs from peptide amino acids to either the base (rectangle) or ribose (pentagon) protons of a nucleotide. RNA markings are the same as in Figure 3.10.

Table 4.4: Summary of restraints used for molecular modeling.

Type of restraint	number of restraints	restraints per residue
Distance restraints		
Hydrogen bonds-RNA	86	2.5
Intraresidue RNA	210	6.2
Interesidue RNA	184	5.4
Hydrogen bonds-peptide	34	1.5
Intraresidue peptide	158	6.7
Interresidue peptide	145	6.3
RNA-peptide	61	1.1
Total distance	**878**	**15.4**
Torsion restraints		
Glycosidic bond	34	1.0
Sugar pucker[a]	60	1.8
Peptide ϕ	17	0.7
Total torsion	**111**	**1.9**
Total all restraints	**989**	**17.3**

[a]The torsions ν_2 and ν_4 were set to restrain each sugar puckers.

(A)



(B)

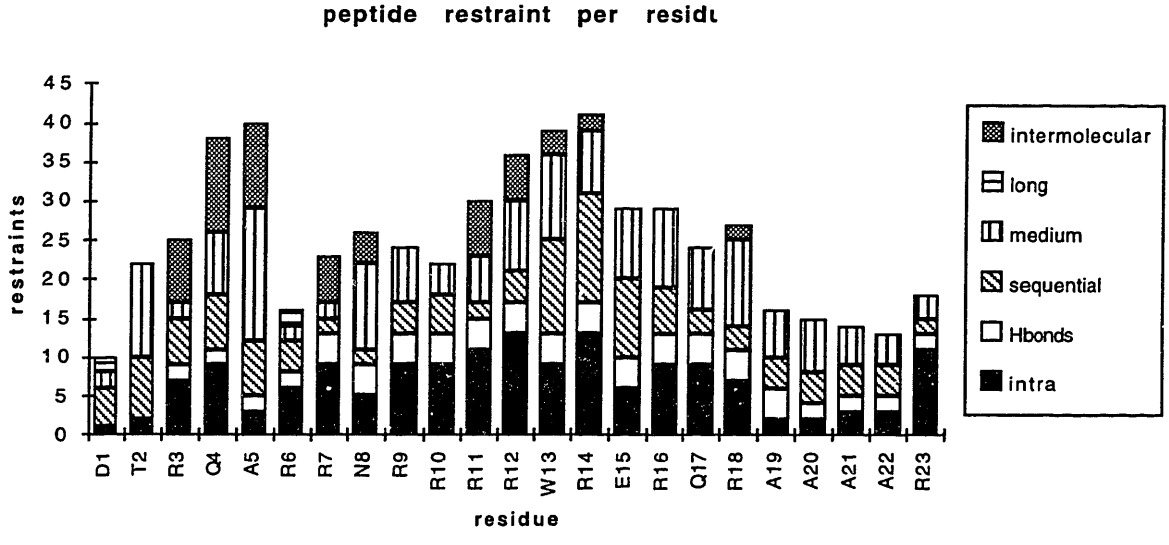


Figure 4.14: Distance restraints per residue for **(A)** RNA and **(B)** peptide. Note that in this graph, interresidue NOEs count as one for **both** residues; therefore, the total numbers will be larger than listed in Table 4.5.

on the structure of the core region of the binding site. In the later stages of modeling hydrogen bonds for the backbone amides and carbonyls were added for well-defined regions of α-helical structure.

From analysis of the $^3J_{H1'-H2'}$ couplings, the sugar puckers for nucleotides G46, A68, G71, and U72 were restrained C2'-endo. G48, C56, A57, and G67 were left unrestrained due to evidence of mixed C2'- and C3'-endo populations. All other sugar puckers were restrained to be C3'-endo. Due to weak NOEs for all intranucleotide H8/H6-H1' crosspeaks, all 34 glycosidic torsion angles were restrained as *anti.*

17 of 22 ϕ torsion angles for the peptide backbone were restrained to values deduced from the HNHA experiment. Thr^{34} had a coupling constant ~8 Hz and was restrained between -84 and -156° (10° error added to the two solutions of Karplus equation). Arg^{35} had a detectable diagonal peak, but undetectable HNHA cross peak. Therefore, its coupling constant was determined to be <3Hz and restrained between 18 and -60° (10° error add to two of four solution of the Karplus equation). Gln^{36}, Ala^{37}, Arg^{42}, and Arg^{44}-Ala^{52} (3-5 Hz) were restrained ±10° around the calculated angle (~-60° solution of the Karplus equation). Arg^{38}, Asn^{40}, and Arg^{43} were restrained ±20° of the calculated angle due to partial overlap of crosspeaks in the HNHA experiment and increased error in quantitation. The ϕ angles for Arg^{39} and Arg^{41} were not restrained due to overlap in the HNHA experiment. Three amino acids at the C-terminus (Ala^{53}, Ala^{54}, Arg^{55}), which had resolvable HNHA cross peaks, were not restrained due to evidence of motional averaging from ^{15}N-amide relaxation experiments (N.S. Rao and L.E. Kay, unpublished data).

Three dimensional models of the RRE-Rev complex were generated using a hybrid distance geometry-simulated annealing approach (Nilges et al., 1988) with the NMR-derived distance and dihedral restraints (section 2.3.2). No non-experimental restraints were used to enforce A-form geometry for any regions of the RNA. No electrostatics and a quartic repulsion function was used for non-bonded contacts in all calculations. The

Lennard_Jones van der Waals function and electrostatics both are "attractive" potentials (negative energy terms) that can have significant effects on the resulting structure. Not using these attractive non-bond potentials lets the NMR data dominate the structure calculation and avoids potential structural bias from the forcefield. A total of 140 structures were generated from distance geometry calculations and input into the high temperature simulated annealing protocol (Figure 2.3). 30 structures with the lowest violation and total energies were input into the refinement protocol. Nineteen structures with low restraint violation energies, no violations >0.3 Å, and good covalent geometry were obtained (Table 4.5). The rmsd of all heavy atom coordinates of each structure to the average structure is 1.39 Å (Table 4.5), which is moderately well-defined for an NMR structure determination . A superposition of the coordinates of the nineteen calculated structures (Figure 4.15) , shows that the entire RNA-peptide complex is well-ordered with only the C-terminus of the peptide and several nucleotides in the hairpin loop being slightly disordered.

Table 4.5: Structural statistics and coordinate root mean square deviations.

Structural Statistics

Parameter	<SA>[a] ± sd[b]	$(SA)_r$[c]
Rmsd from distance restraints (Å)		
All (878)[d]	0.015 ± 0.003	0.012
Intraresidue RNA (210)	0.013 ± 0.002	0.009
Interresidue RNA (270)[e]	0.018 ± 0.003	0.014
Intraresidue peptide (158)	0.008 ± 0.007	0.017
Interresidue peptide (179)[f]	0.013 ± 0.004	0.006
RNA-peptide (61)	0.017 ± 0.007	0.016
Rmsd from dihedral restraints (degrees)		
All (111)	0.57 ± 0.10	0.44
RNA (94)	0.62 ± 0.11	0.47
Peptide (17)	0.17 ± 0.08	0.17
Rmsd from idealized geometry		
Bonds (Å)	0.0034 ± 0.0001	0.0033
Angles (°)	0.031 ± 0.001	0.037
Impropers (°)	1.0 ± 0.2	0.65
Restraint violations		
# of distance viol. > 0.1 Å	3.7 ± 2.3	2
Largest distance violation	0.17 ± 0.05	0.18
# of dihedral viol. > 2°	1.8 ± 1.4	0
Largest dihedral violation	3.1 ± 1.3	1.9

Coordinate rms deviations

Region selected	<SA> vs. SA[g] ± sd	<SA> vs. $(SA)_r$ ± sd
All heavy	1.39 ± 0.12	1.56 ± 0.15
RNA heavy/peptide backbone	1.25 ± 0.12	1.40 ± 0.15
Core[h] heavy	1.24 ± 0.09	1.44 ± 0.11
Core: RNA heavy/peptide backbone	1.05 ± 0.10	1.21 ± 0.11

[a]<SA> is the nineteen simulated annealing structures. [b]"sd" is the standard deviation from the mean for the nineteen structures. [c]$(SA)_r$ is the structure obtained from averaging the coordinates, followed by restrained energy minimization. [d]In parentheses are the number of restraints. [e]Includes 86 hydrogen bond restraints for the base pairs. There are 2 restraints per base pair. [f]Includes 34 hydrogen bond restraints for the backbone of the α-helix. There are 2 restraints per hydrogen bond. [g]SA is the average structure before energy minimization. [h]The "core" of the complex is defined as RNA nucleotides 44-53,65-76 and peptide amino acids 33-50.

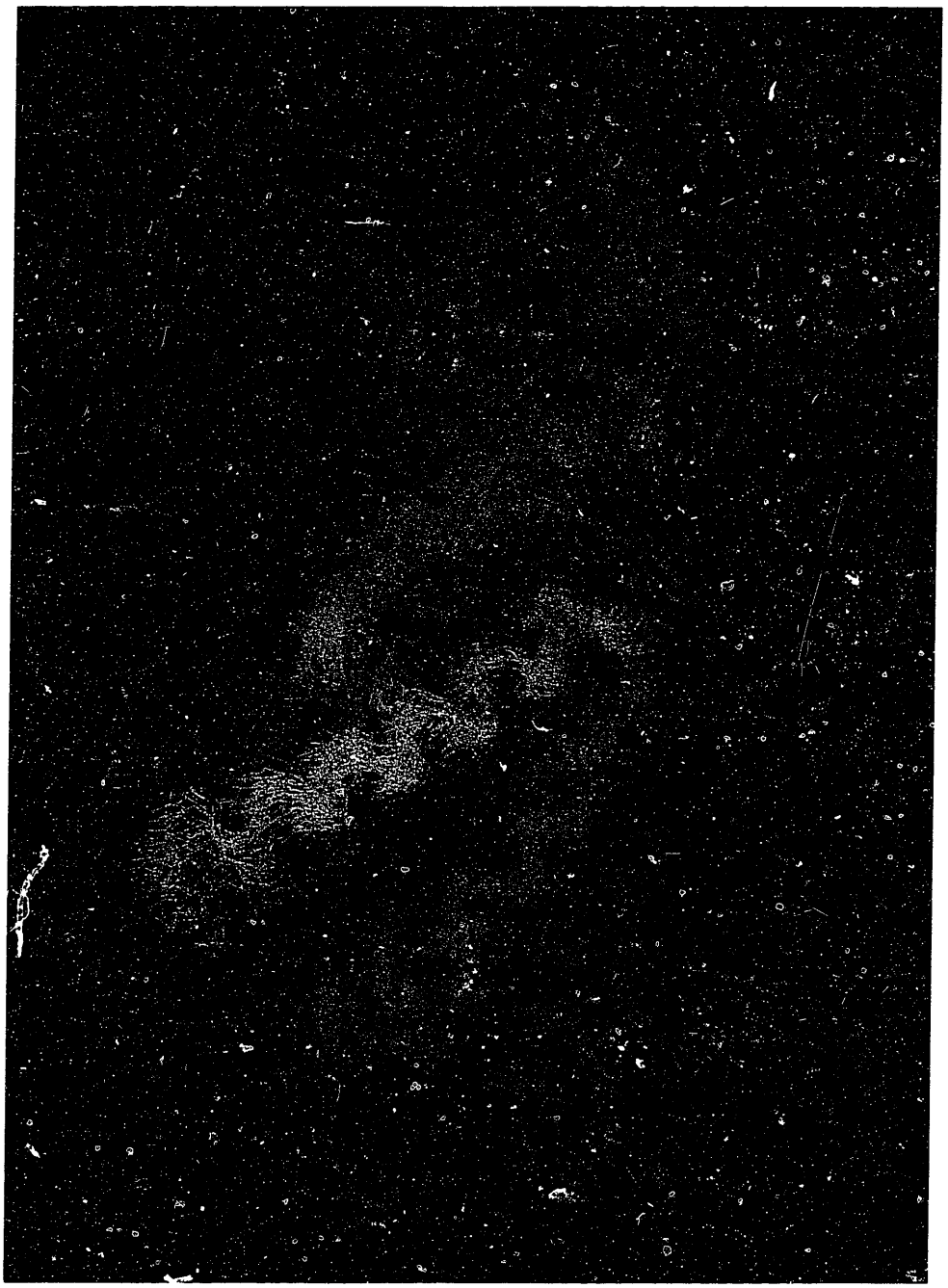


Figure 4.15: Superposition of simulated annealing structures. All heavy atoms of the RNA (blue) and backbone atoms of the peptide (magenta) of the nineteen calculated structures were superimposed on the average structure (not shown). The view shows the major groove face of the internal loop with the N-terminus of the peptide at the upper right.

CHAPTER 5: STRUCTURE OF RNA-PEPTIDE COMPLEX

5.1 Global Architecture of the RNA-Peptide Binding Site

5.1.1 RNA structure Two views of the minimized average structure of the RRE RNA in the complex are shown in Figure 5.1. The internal loop of the RNA forms a continuous helix with the flanking Watson-Crick regions and two bases (A68 and U72) unstacked. The structure of the Watson-Crick regions is predominately A-form with the only major deviation being the G46:C74 base pair adjacent to the G47:A73 base pair. The G46:C74 base pair has a roll towards the major groove, and the G46 ribose has a non-A-form C2'-endo pucker. The lack of NOEs between G47 H8 and G46 H2' or H3' is consistent with the roll towards the major groove. G46 appears to make hydrogen bonds to Arg^{44} of the peptide (see section 5.2.2), and mutational analysis has identified this base pair as a major determinant of binding specificity (Bartel et al., 1991). The most striking feature of the RNA structure is a large distortion in the backbone around the bulged U72 nucleotide (Figure 5.1B). The distortion in the phosphate backbone of RRE RNA is due to the fact that the strand polarity reverses direction for the nucleotide G71, enabling the G48:G71 base pair to form with both glycosidic torsions *anti*. A close up view of the nucleotides in the internal loop is shown in Figure 5.2. Between A73 and G71, which are stacked, the backbone goes towards the major groove for the bulged U72, then loops around to the ribose of G71. The ribose of G71 is oriented in the opposite direction from the rest of the strand, which can be easily discerned by looking at the position of the O4' oxygen. In this conformation, the positions of the H8 and H1' protons (green and magenta spheres, respectively) are reversed with respect to the major and minor groove for this one nucleotide. This results in the unusual sequential connectivity patterns observed in the NMR spectra. The sequential NOEs for G71 H1' are highlighted with white arrows to

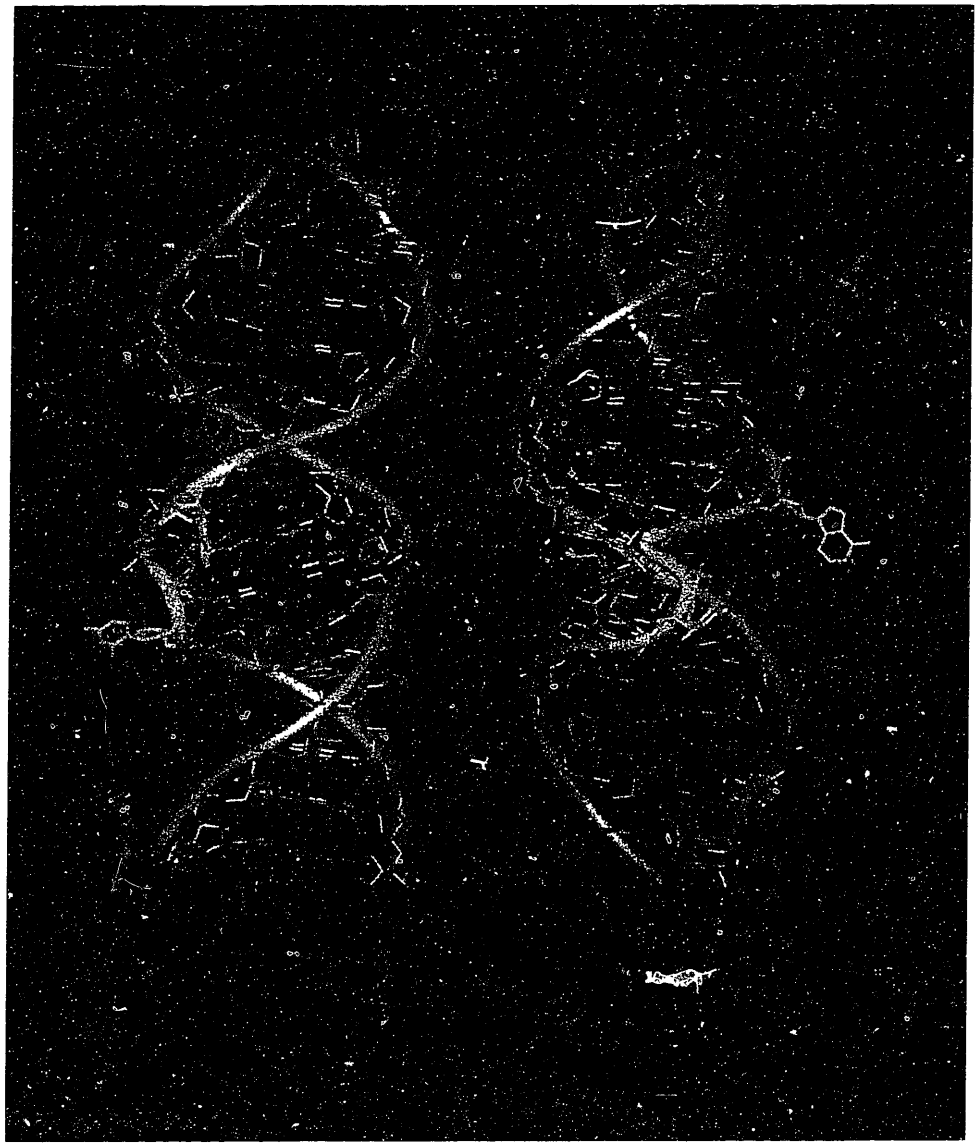


Figure 5.1: Two views of the minimized average structure of the RRE RNA with the peptide not shown. **(A)** The major groove of the internal loop region, and **(B)** the RNA rotated 90° about its helix axis, showing the phosphate backbone near the bulged U72. In both views, the purine-purine base pairs and single nucleotide bulges are colored magenta and green, respectively.

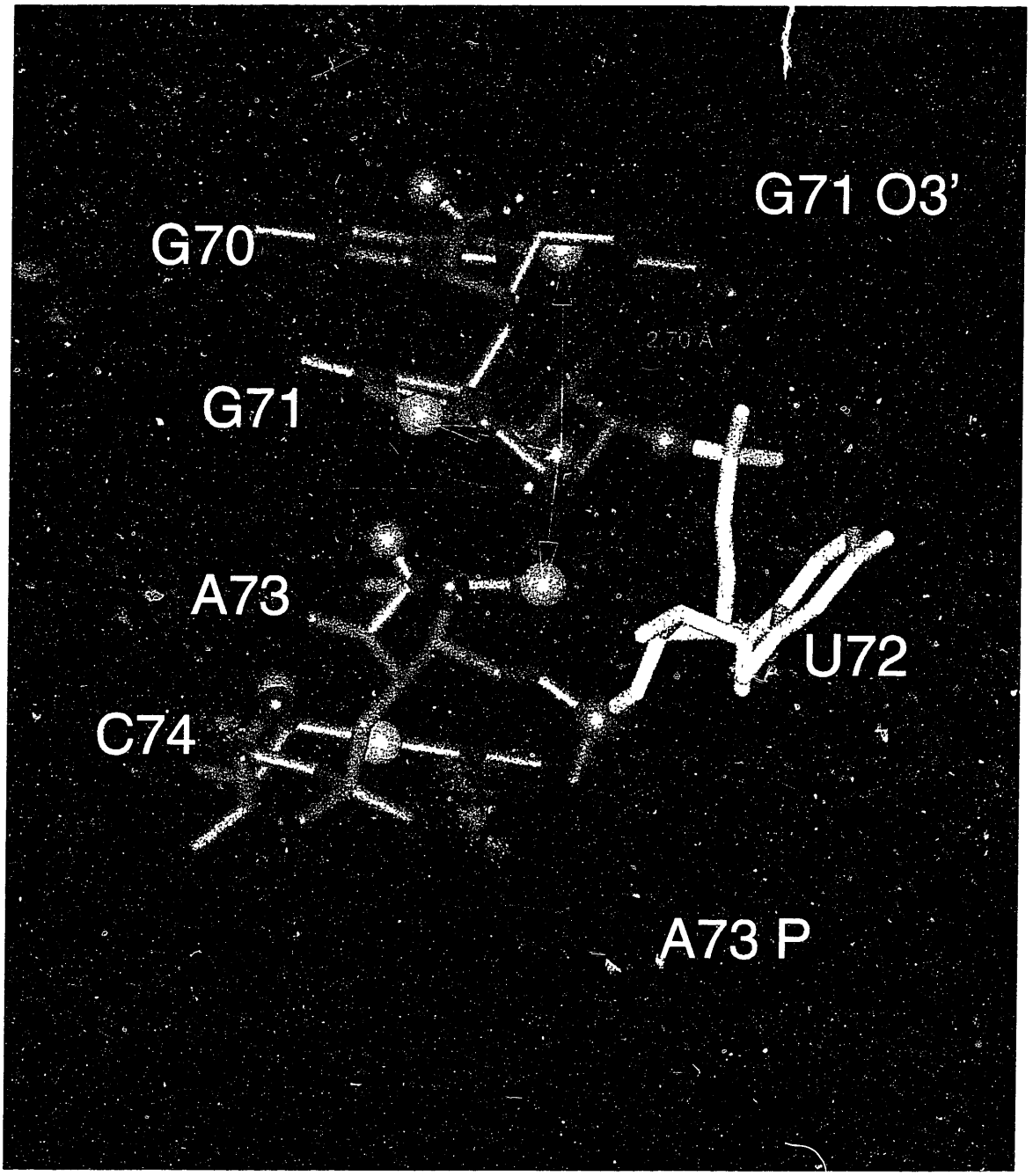


Figure 5.2: Close up view along the phosphate backbone of the nucleotides G70-C74. H6/H8 protons, H1' protons, and O4' oxygens are highlighted as green, magenta and red spheres, respectively. The red arrow highlights the orientation of the O4' in G71 compared to the rest of the riboses on that strand. The white arrows highlight the three NOEs to base protons observed for G71 H1' and show the positions of A73 P and G71 O3' atoms (see text). A potential hydrogen bond between G71 O2' and G70 O1P is shown with a green dashed line.

illustrate how the H8 protons in either direction along the strand are close enough in space to observe an NOE. The reversal of the backbone orientation for the G71 ribose also explains the unusual ribose to ribose NOEs observed. The H2' proton of G71 is positioned toward the ribose of G70 as a result of the reversal, and exhibits NOEs to the H1' and H2' protons of G70. In addition, the 2'-hydroxyl of G71 is positioned towards the phosphate oxygens of G70 (Figure 5.2) and is potentially forming a hydrogen bond that is stabilizing the unusual fold of the backbone. Substitution of G71 with a deoxyribo-guanosine (dG71) results in a ~4-10 fold decrease in Rev binding (Pritchard et al., 1994), consistent with the 2'-hydroxyl playing a role in stabilizing the critical G:G base pair. However, the G71 2'-hydroxyl is oriented towards the major groove rather than the minor groove, and interactions with the peptide could also contribute to its importance (hydrogen bonds to the Arg^{39} guanidium group are possible in a small number of the simulated annealing structures).

Interestingly, similar structures with locally parallel stranded homopurine base pairs adjacent to single nucleotide bulges have been observed in the sarcin/ricin and E loops of ribosomal RNA (Szewczak et al., 1993; Wimberly et al., 1993). In both cases, the base pair formed was an *anti-anti* purine-purine base pair where a *syn-anti* base pair would have been expected. The structure may be a common motif for internal loops of RNA (Shen et al., 1995), although it is unclear in the ribosomal RNA cases what role the motif plays in protein recognition.

The geometric constraints on formation of the G:G base pair are shown schematically in Figure 5.3. The two faces of a nucleotide base are stereochemically distinct, as represented in the figure by black and white faces. In a standard right handed anti-parallel helix, a given face is oriented in opposite directions on each strand (Figure 5.3A). In order to form the G:G base pair observed in RRE, the same face of both G residues must be oriented in the same direction. There are two possible ways to

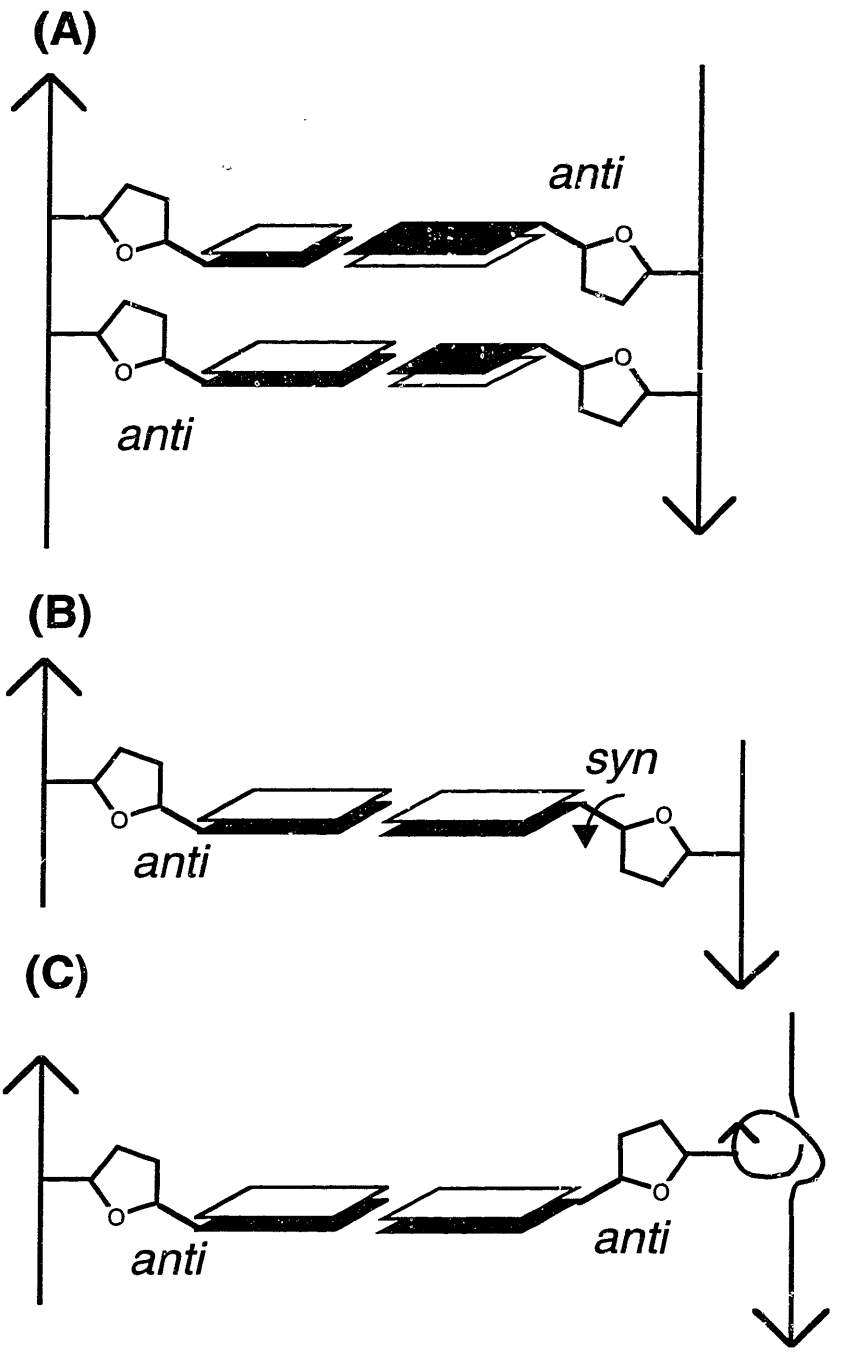


Figure 5.3: A schematic of two different ways the G:G base pair observed in RRE can be incorporated into an antiparallel helix. Each base has two distinct faces, which are represented by black and white shadowing. **(A)** Orientation of the bases in an antiparallel Watson-Crick duplex. **(B)** G:G base pair with one of the guanosine glycosidic torsions *syn*. **(C)** G:G base pair with the duplex locally parallel stranded and both glycosidic torsions *anti*.

accommodate this geometry into an antiparallel helix. One is to change the orientation of one glycosidic bond to the *syn* conformation (Figure 5.3B). Structural predictions of the RRE had assumed that one of the guanosine positions in the RRE must be *syn*. Alternatively, the orientation of the phosphate backbone can be changed such that the strands are effectively parallel for the G48:G71 base pair, while the glycosidic conformation remains *anti* (Figure 5.3C), as was illustrated in the RRE structure in Figure 5.2. Some interesting results regarding these two possible orientations have been obtained from two molecular modeling studies of the RRE RNA, where three dimensional models of the RNA were built using the available secondary structure data as restraints (Le et al., 1994; Leclerc et al., 1994). In both cases the glycosidic torsion of G71 was initially restrained as *syn*; however, during molecular dynamics calculations with no restraints the torsion flipped to the *anti* position to give structures grossly similar to the structure presented here. Potentially in the context of an adjacent single nucleotide bulge, the *anti/anti* G:G base pair may be more sterically favorable than the *syn/anti* form.

The structure of the unusual backbone geometry associated with the G48:G71 base pair is important for creating a binding pocket for the α-helical peptide, which is lying in the major groove of the RNA. α-helical recognition is typical of DNA-binding proteins, which use a variety of structural motifs to present an α-helix for recognition of specific bases in the major groove of DNA (Pabo and Sauer, 1992). In contrast to DNA, however, the major groove of a Watson-Crick RNA helix (A-form) is too deep and narrow to accommodate a protein α-helix. In general, this problem can be overcome by a widening of the major groove near bulge or loop regions of RNA structure (Weeks and Crothers, 1993). In the RRE-Rev complex, the two purine-purine base pairs provide the structural features to open the major groove. Both the G47:A73 and G48:G71 base pairs utilize the "Watson-Crick" faces of the nucleotides, resulting in an C1':C1' distances ~2 Å wider than a Watson-Crick base pair. More importantly, the widening of the major groove appears to

be facilitated by the distortion of the RNA backbone resulting from formation of the G48:G71 base pair. To visualize of how this distortion is opening the major groove, the backbone of a standard A-form helix is superimposed with the lower stem of the RRE RNA (Figure 5.4). The backbone severely deviates from the path of an A-form helix at the bulged nucleotide (U72) preceding the G48:G71 base pair, widening the major groove by ~5 Å (Figure 5.5).

The proposed role of the G:G base pair in the opening of the major groove is consistent with biochemical data on the RRE-Rev interaction. In vitro selection studies predicted that the G48:G71 base pair would play a structural rather than a sequence specific role in defining the RNA-binding site, since an A:A base pair could be substituted while maintaining the ability to bind Rev (Bartel et al., 1991). G:G and A:A can form symmetric homopurine base pairs that are isosteric, but present different hydrogen bond donors and acceptors in the major groove. The U72 bulge appears to structurally assist the formation of the G:G base pair, which is also consistent with the biochemical data. Although G71 and A73 are stacked, the unusual geometry positions the G71 O3' oxygen far from the phosphate of A73 (Figure 5.2), and the single nucleotide bulge U72 acts as a simple spacer to bridge these two nucleotides. The role of U72 as a spacer is consistent with the requirement for a bulge nucleotide at this position, but whose identity is not important for Rev binding (Bartel et al., 1991). Furthermore, chemical substitution of U72 with a three-carbon linker that has no base moiety does not effect Rev binding in vitro (Pritchard et al., 1994). Therefore, U72 should be considered part of a structural unit with the G:G base pair.

5.1.2 Peptide structure The Rev peptide forms a continuous α-helix with backbone ϕ/ψ angles ±15° of canonical values for amino acids Arg^{35}-Arg^{55} in the average structure. No side chain-side chain hydrogen bonds were observed that might stabilize the

Figure 5.4: Comparison of the major grooves of A-form RNA (red) and RRE (blue). The A-form helix was superimposed onto the lower stem of the RRE (G41-G46 and C74-C79), and only the backbone trace of the helix is shown for clarity.

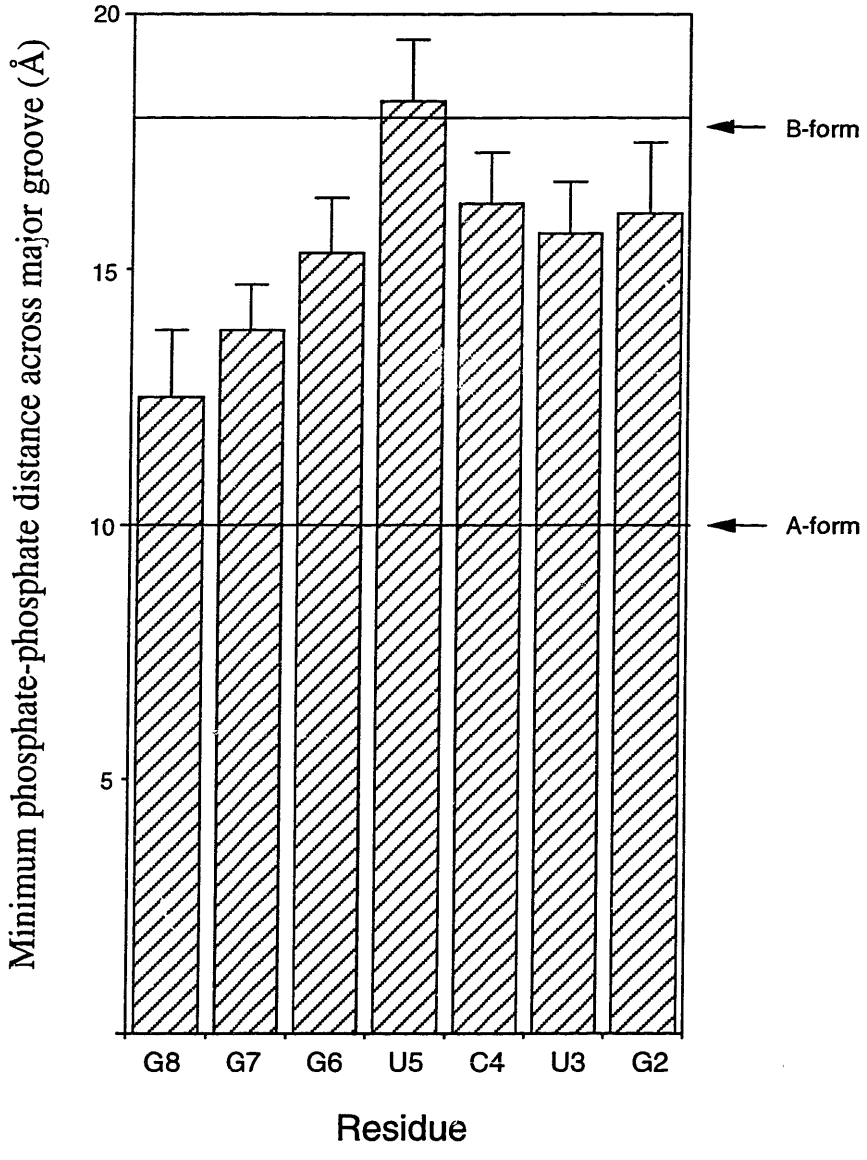


Figure 5.5: Major groove width of bound RRE RNA. The distance to the nearest phosphorous atom across the major groove from the phosphorous atom of the nucleotide listed was measured. No correction was made for the van der Waals radii of the phosphate groups (subtract ~5.8 Å). These are the only nucleotides for which a groove was definable.

α-helical conformation of this highly basic peptide. However, the bound Rev peptide contains a "capping box" structure (Harper and Rose, 1993) at its N-terminus that presumably stabilizes the α-helix. Prototypical capping boxes contain short hydrophilic amino acids at positions N and N+3 of the helix that form reciprocal side chain-backbone amide hydrogen bonds. These amide protons would otherwise not be involved in hydrogen bonds at the N-terminus of an α-helix. The N position of the helix is defined as the first N-terminal residue to have non-helical ϕ/ψ torsion angles. The observed ϕ/ψ angles for Thr^{34} are in the ranges observed for "N-cap" residues of helices in other proteins (Harper and Rose, 1993), and Thr^{34} has NOEs and carbon chemical shifts typical of an N-cap residue (see section 4.2.2). Consistent with the formation of an N-cap, the Oγ of Thr^{34} is within hydrogen bonding distance of the backbone NH of Ala^{37} in the average structure (Figure 5.6). There is some ambiguity to this hydrogen bond looking at all nineteen calculated structures. There are roughly two equally populated rotamers of the χ_1 torsion angle for Thr^{34} (~60° and ~-60°). One rotamer positions the Oγ of Thr^{34} within hydrogen bond distance, while the other rotamer has the βH proton directed towards the backbone NH of Ala^{37}. The NMR data is more consistent with the former rotamer, since a strong NOE between the Thr^{34} βH and Ala^{37} NH protons would be expected for the latter rotamer, and the observed NOE is very weak. The existence of this rotamer in the calculated structures is probably permitted by the loose van der Waals (1.8 Å) lower bound use for all NOEs ranges (see section 2.3.1). Therefore, we believe that the structure with the hydrogen bond is the most probable structure. The N3 position in the Rev helix (Ala^{37}) cannot form the reciprocal hydrogen bond to the backbone amide of Thr^{34} that is characteristic of prototypical capping boxes. However, partial N-cap structures with only one hydrogen bond have been observed (Harper and Rose, 1993), and thermodynamic studies with model peptides estimate that a threonine at position N and an alanine at position N3 contributes 0.7 kcal/mol to helix stability (Doig and Baldwin, 1995). In

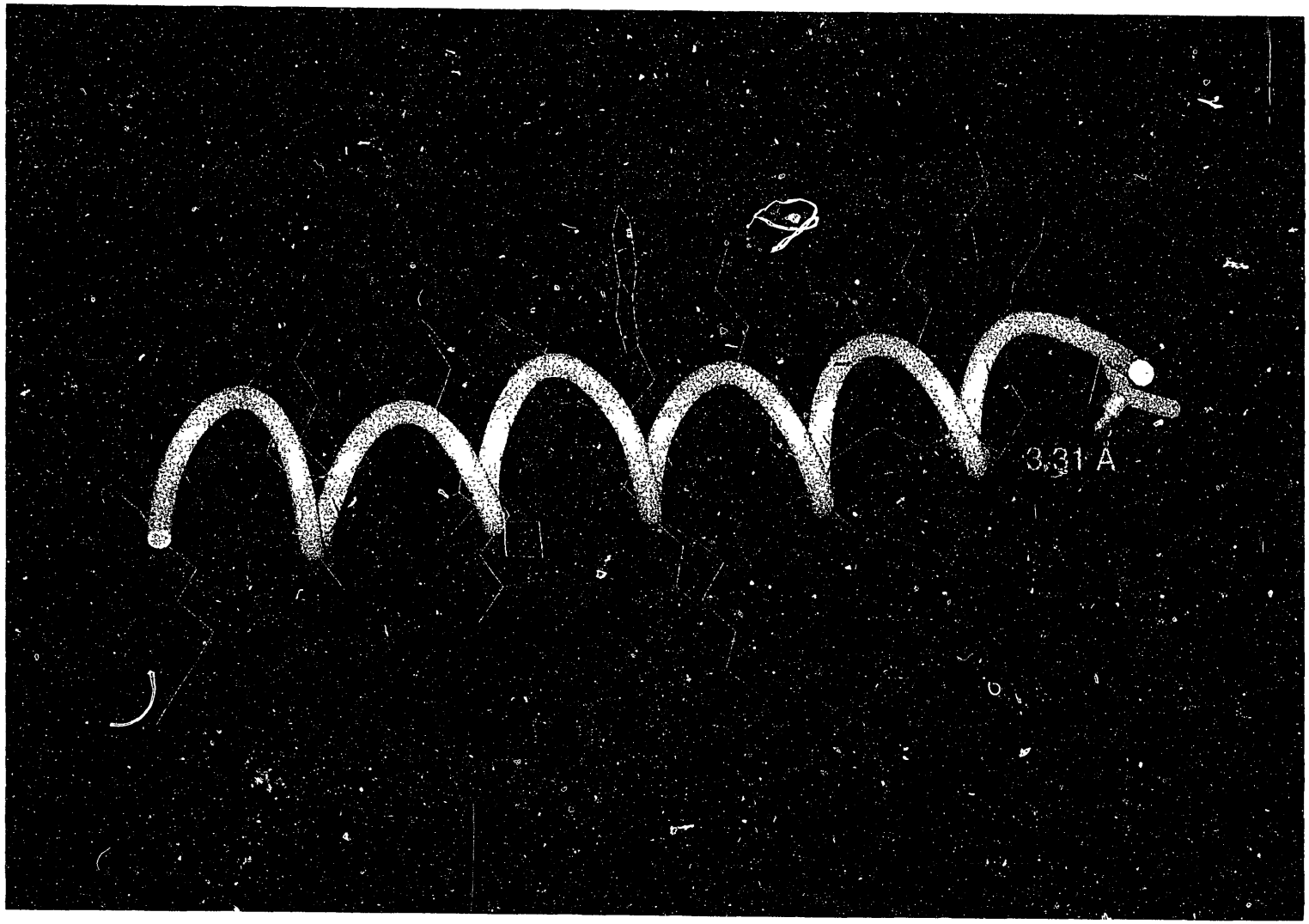


Figure 5.6: Structure of Rev peptide. The backbond oxygens and nitrogens are colored red and blue, respectively. The Thr^{34} side chain (γ oxygen colored red) and Ala^{37} backbone are shown as thick bonds with the N-cap hydrogen bond indicated with a green dashed line.

addition to the Thr^{34}-Ala^{37} interaction, NOEs were observed between Asp^{33} (N') and Arg^{38} (N4), and van der Waals interactions between these positions of the helix have been shown to be a characteristic feature of capping boxes (Seale et al., 1994). The architecture of the N-cap suggests that the N-terminal portion of the Rev protein adjacent to the arginine-rich domain may not extend completely through the RRE major groove, but may follow a course out of the major groove, perpendicular to the α-helix axis. The width of the major groove becomes progressively smaller towards the N-terminus of the peptide, making it sterically unfavorable for the α-helix to emerge through the groove.

5.1.3 Positioning of α-helix in major groove The general position of the Rev α-helix in the major groove is similar to what has been observed in many DNA-protein complexes, except that it penetrates more deeply into the groove. In contrast to the shallow major groove of B-form DNA, the RRE internal loop, despite widening the groove, retains the characteristic deep groove of an A-form RNA helix. In order for the side-chains of the Rev-peptide to reach the bases of the RNA, the peptide backbone must penetrate deeply into the groove. A comparison of the positioning of Rev and GCN4 α-helices in the major groove of their respective binding-sites (Figure 5.7) reveals that the Rev α-helix is situated ~3 Å deeper. As a result of the penetration, the Rev α-helix has an extensive interface with the RNA, burying almost twice as much surface area as any helix in a DNA-protein complex (Table 5.1). In addition to the deep penetration, however, the interaction with the RNA over 3-4 turns of the α-helix and the high density of arginines with large side-chain surface areas may also be contributing to the significantly larger buried surface area.

The surface of the bound Rev α-helix can be evenly divided into three faces based on interactions with the RRE RNA (Figure 5.8). One face interacts with one strand of the RRE from nucleotides C65-A73 (A-face), and a second face interacts with the opposite strand from nucleotides C44-G48 (B-face). These two faces segregate with the patches of

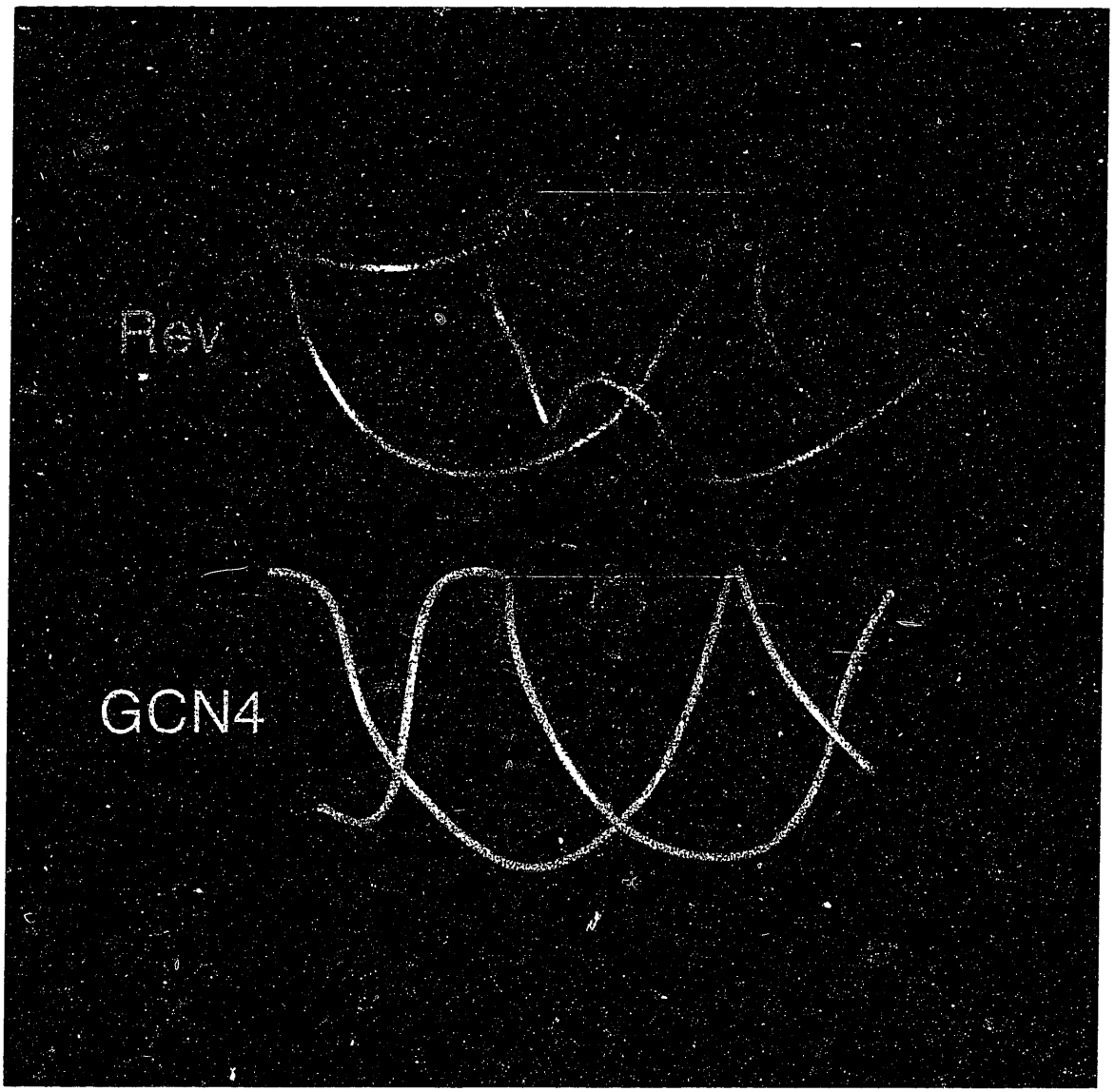
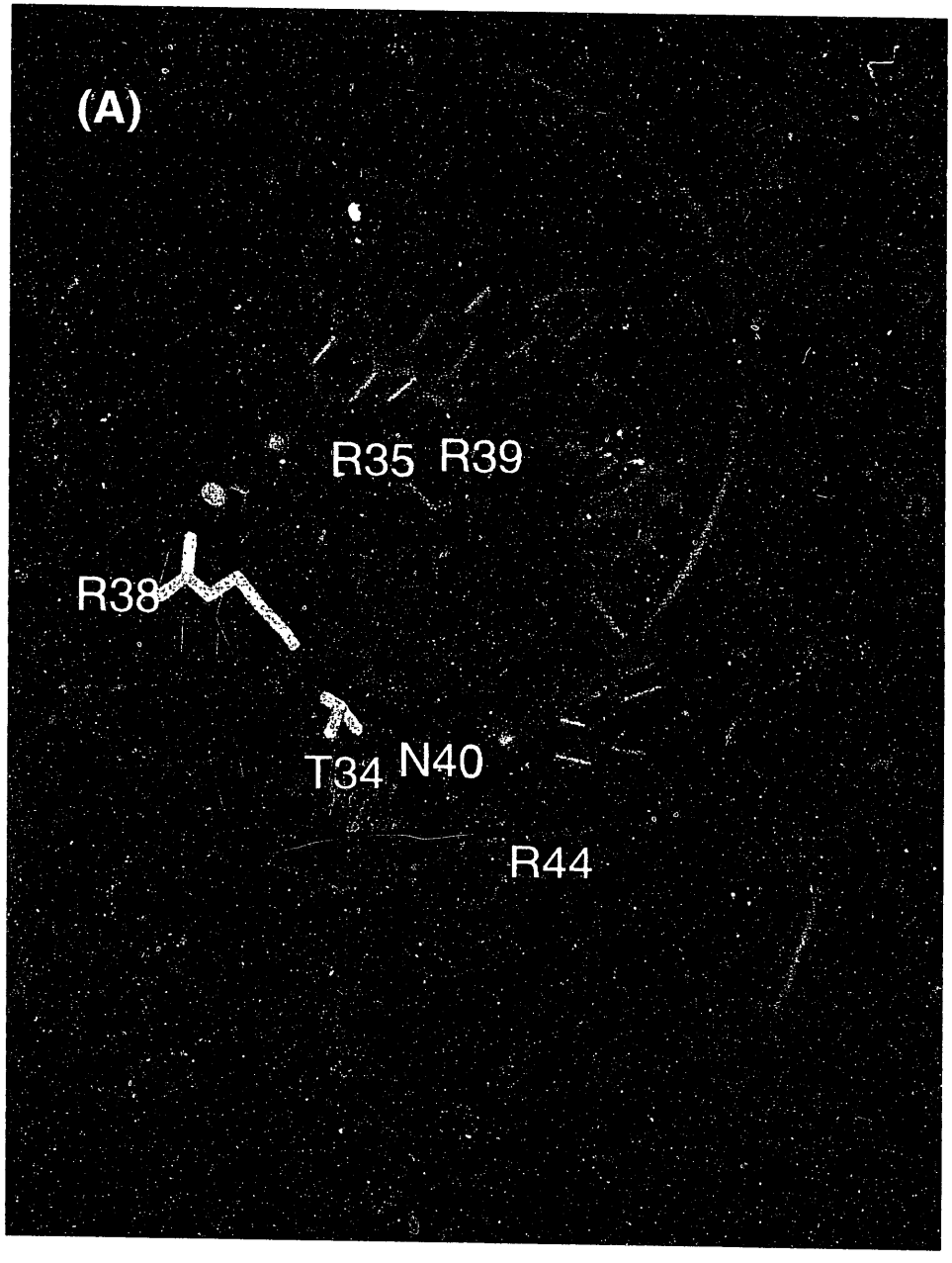


Figure 5.7: Penetration of α-helix into major groove. Shown are views down the C-terminus of the α-helix axis of Rev (blue) and GCN4 (green). White lines indicate the edges of the RNA/DNA major grooves at the helix-binding site. For comparison, two amino acids that make base contacts are shown for Rev (Arg^{35} and Arg^{39}, magenta) and GCN4 (Arg^{342} and Asn^{235}, red).

Table 5.1: Buried van der Waals surface area of the Rev α-helix and the recognition α-helices of five DNA-protein complexes. The buried area was calculated as the difference between the surface area of the complex and the sum of the individual protein and RNA/DNA components. A probe radius of 1.4 Å was used for all calculations.

Complex[a]	Buried Surface Area (Å^2)	References
Rev_{34-50}	1079	
$GCN4_{225-249}$	558	(Ellenberger et al., 1992)
$Myb_{127-142}$	535	(Ogata et al., 1994)
$Engrailed_{40-59}$	341	(Kissinger et al., 1990)
$Zif268_{45-58}$	304	(Pavletich and Pabo, 1991)
λ-$repressor_{41-54}$	289	(Beamer and Pabo, 1992)

[a]The subscripts indicate the amino acids used for the surface area calculation and do not necessarily correspond to the beginning or end of the recognition helices as would be defined by the backbone ϕ/ψ torsions.



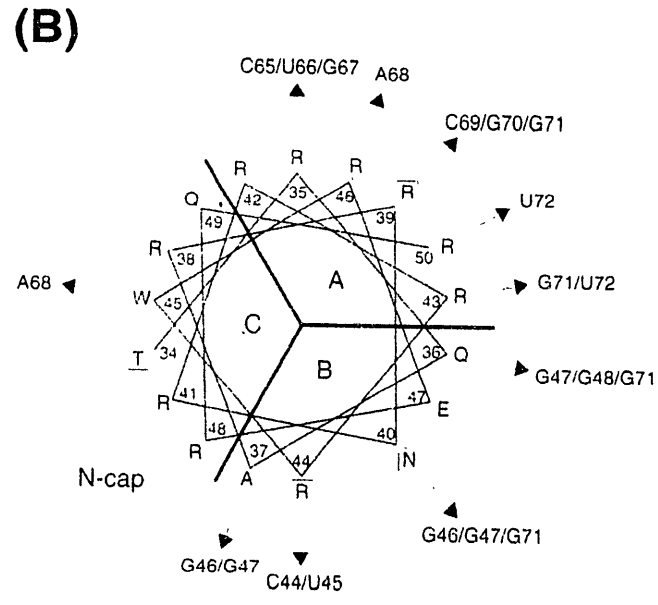


Figure 5.8: Interaction faces of Rev α-helix **(A)** View of the Rev-RRE complex down the N-terminus of the peptide. Only nucleotides 44-53, 65-76, and amino acids 33-50 are shown. The side chains of the peptide are colored magenta, red, or yellow, according to whether they are on the A-, B-, or C-face of the peptide, respectively. The six amino acids boxed in part B are indicted by thick bonds. The RNA is blue with invariant nucleotides from in vitro selection studies colored green and phosphates that interfere with binding when ethylated as blue spheres (see Figure 1.9). **(B)** Schematic of RNA-peptide NOEs on a peptide α-helical projection. The peptide is divided into 3 faces based on interactions with the RNA (A,B,C; see text). The filled and open arrowheads represent NOEs to the RNA and the N-cap interaction, respectively. The boxes indicate the six amino acids that give >10-fold decrease in binding specificity upon mutation to alanine. The circles indicate arginines that give >3-fold decrease in binding specificity upon mutation to alanine, but can be functionally substituted by lysine.

intermolecular NOEs observed for the complex (Figure 4.13 and Figure5.8B). The third face (C-face) is largely solvent-exposed, but because the α-helix penetrates the groove so deeply, amino acids on this face are still able to contact the phosphate backbone. This is in contrast to most DNA-protein interactions, where amino acids on the back side of an amphiphatic recognition helix typically make protein-protein contacts distal from the DNA-phosphate backbone. Mutational data support the buried nature of the Rev α-helix, since the six amino acids that make the largest contributions to *in vitro* binding specificity are evenly distributed around the entire circumference of the α-helix (Figure 5.8B) (Tan et al., 1993).

5.2 Specific RNA-Peptide Contacts

5.2.1 Nucleotide base contacts Structural studies of DNA-protein complexes have clearly identified the importance of "sequence-specific" hydrogen bond interactions between protein side chains and nucleotide bases for specific recognition. In the Rev-RRE complex, four amino acids make base-specific contacts in the major groove that are probably important for sequence-specific recognition. Arg^{35} and Arg^{39}, which are one turn apart on the A-face of the α-helix, interact with bases on one side of the groove, and Asn^{40} and Arg^{44}, which are one turn apart on the B-face, interact with bases on the opposite side of the groove. Mutation of any of these amino acids to alanine results in >10-fold decrease in peptide-binding specificity (Tan et al., 1993). Furthermore, mutation of any of these three arginines to lysine results in a significant loss of binding activity in vivo, suggesting that these side chains make specific hydrogen bonds rather than simple electrostatic contacts (Tan and Frankel, 1994). Almost all of the nucleotides that are close enough to make potential hydrogen bonds to these four amino acids (U45, G46, G47, U66, G67, G70, and A73) were invariant in selection experiments for binding of the Rev

protein to stem IIB RRE RNA in vitro (Bartel et al., 1991). In addition, specific chemical modifications that interfere with peptide or protein binding have been identified for most of these nucleotides.(Iwai et al., 1992; Kjems et al., 1992; Pritchard et al., 1994). The correlation of the importance of these potential base contacts in the minimal RNA-peptide complex with mutational data heavily suggest that they are critical for the wild type Rev-RRE interaction.

The precise amino acid-base hydrogen bonding arrangements of the base-specific contacts are difficult to determine given the relatively low density of available NOEs at the RNA-peptide interface. Potentially, the variability of the side-chain positions could also be attributable to intrinsic dynamics of the peptide. Relaxation experiments on the arginine epsilon protons, however, show a correlation between arginines at the RNA interface and reduced flexibility of the side chain (N.S.R., D.R.M, and L.E.K., unpublished results). Therefore, the variability of the side-chain positions is more likely correlated to the density of NOE data, especially for the longer arginine side-chains. By analyzing the distribution of side chain positions in all calculated structures (Figure 5.9) and using biochemical data as a guide, several reasonable hydrogen bonding schemes can be proposed.

The guanidium group of Arg^{35} is positioned to make bridging hydrogen bonds to the O4 and O6 oxygens of U66 and G67, respectively, though the guanidium group is within hydrogen bonding distance (<3.5 Å heavy atom-heavy atom) of U66 O6 in a lower percentage of the structures. This is consistent with the conserved nature of these bases by in vitro selection and the absence of interference by modification of the N7 position of G67 (Bartel et al., 1991; Kjems et al., 1992). In addition, chemical modification of the N3 position of U66, which disrupts the A52-U66 Watson-Crick base pair, interferes with protein binding (Kjems et al., 1992).

In the average structure, the guanidium group of Arg^{39} is positioned to make a classic arginine-guanine interaction involving hydrogen bonds to the N7 and O6 of G70

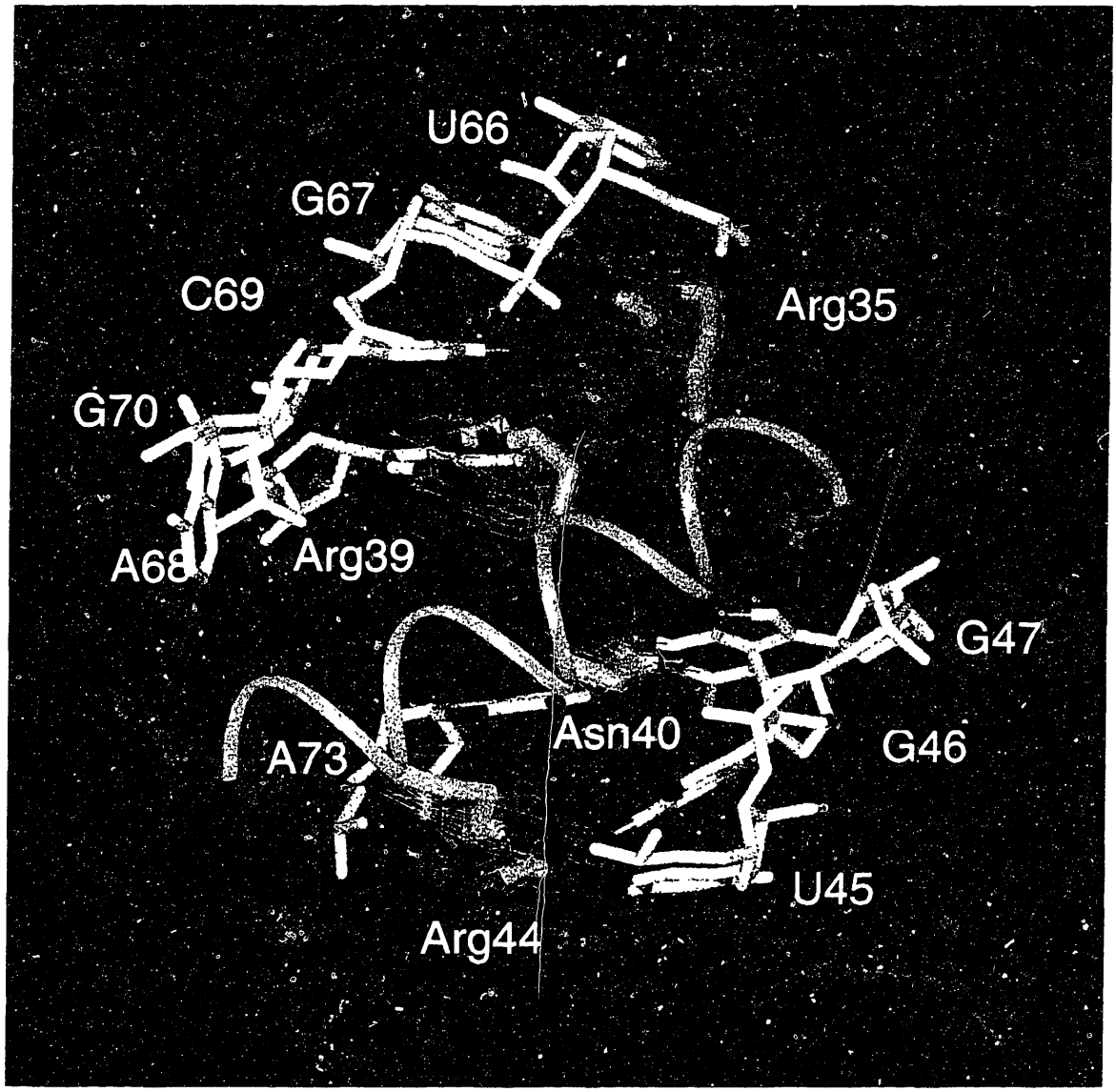


Figure 5.9: Superposition of amino acid side chains making base specific contacts. All nineteen structures were superimposed on the energy-minimized average structure, but only the side chains for Arg^{35}, Arg^{39}, Asn^{40}, and Arg^{44} are displayed. Since the variance of the positions of the amino acid side chains is much greater than the variance of the positions of the nucleotide bases, only the average RNA structure (nucleotides U45-G47, U66-G70, and A73) is shown for clarity. The average structure is shown as thick bonds. The RNA is colored gray, except for oxygens and nitrogens of the major groove edges of the bases, which are colored red and blue, respectively. For clarity, a ribbon trace of the peptide backbone for the average structure is shown.

(Seeman et al., 1976), though, in many of the simulated annealing structures only one of these two hydrogen bonds are observed. This is consistent with the conservation of the G70:C49 base pair as well as strong chemical modification interference of peptide binding for the N7 position of G70 (Bartel et al., 1991; Iwai et al., 1992; Kjems et al., 1992).

The side chain amide of Asn^{40} is positioned to make three hydrogen bonds to the G47:A73 base-pair (A73-N6, G47-O6, and G46-N7). The position of this shorter side chain is better defined than the arginines which have more torisonal degrees of freedom, and the confidence level in the proposed hydrogen bonds should be higher. The interaction with the G47:A73 base pair is consistent with its invariance in selection studies in vitro. The importance of the N7 position of G47, however, is somewhat unclear from modification data, where different chemicals and/or substitutions gave varying results (Iwai et al., 1992; Kjems et al., 1992; Pritchard et al., 1994). Nevertheless, an arrangement with Asn^{40} making hydrogen bonds to at least the A73-N6 and G47-O6 positions is consistent with both the NMR and biochemical data. A direct interaction between Asn^{40} and the G47:A73 base-pair has been recently identified using a genetic screen, where mutation of G47:A73 to a G:G base pair could be complemented by a N40Q Rev mutation (C. Jain and J. Belasco, personal communication).

The guanidium group of Arg^{44}, like Arg^{35}, is positioned to make potential bridging hydrogen bonds to the O4 and O6 of U45 and G46, respectively. The N7 position of G46, however, does not make any hydrogen bonding contacts in the average structure, which is inconsistent with chemical modification data indicating this position is critical for specific binding. In several of the simulated annealing structures, the Nε or Nη atoms of Arg^{44} are within hydrogen bonding distance of the N7 of G46; therefore, a possible interaction with the N7 position is within the range of structures defined by the NMR data.

5.2.2 Phosphate backbone contacts In addition to the base-specific interactions, the Rev peptide makes extensive contacts with the phosphate backbone. These contacts are made by Thr^{34} and several arginines. The γ-OH of Thr^{34} is both a hydrogen bond acceptor for the Ala^{37} amide (forming the N-cap) and a potential hydrogen bond donor to a phosphodiester oxygen of G47. Mutation of Thr^{34} results in a >10-fold decrease in binding specificity, and chemical modification of the G47 phosphate reduces binding specificity (Kjems et al., 1992; Tan et al., 1993; Pritchard et al., 1994). This amino acid is unique in the Rev peptide in that its side chain appears to makes both important peptide-peptide and RNA-peptide interactions. Interestingly, the N-cap structure was not observed in NMR studies of the free peptide with small amounts of the solvent trifluoroethanol, where it is 100% α-helical (Scanlon et al., 1995). Thus, the N-cap is only formed in the RNA-peptide complex, and Thr^{34} may be a critical amino acid for the costabilization of both RNA and peptide structure that is observed upon binding (Battiste et al., 1994; Tan and Frankel, 1994).

In addition to Thr^{34}, there are six arginines (Arg^{38}, Arg^{41}, Arg^{42}, Arg^{43}, Arg^{46}, and Arg^{48}) that are positioned to make either hydrogen bonding or simple electrostatic interactions with the phosphate backbone (Figure 5.10). The precision of the structure at the RNA-peptide interface is not sufficient to distinguish between arginines that make hydrogen bonds to phosphates or interact by simple electrostatic interactions, but the biochemical data can again be used as guide to interpret the structural results. Of the four arginines that cannot be substituted by lysine (Arg^{35}, Arg^{38}, Arg^{39}, and Arg^{44}) (Tan and Frankel, 1994), Arg^{38} is the only one that does not make a base specific contact. Arg^{38} is positioned near the phosphates of U66 and G67, which both give chemical modification interference (Kjems et al., 1992; Pritchard et al., 1994). It is possible that Arg^{38} makes bridging hydrogen bonds between the two η-nitrogens and the phosphodiester oxygens of U66 and G67. This would explain the inability to substitute this arginine with lysine.

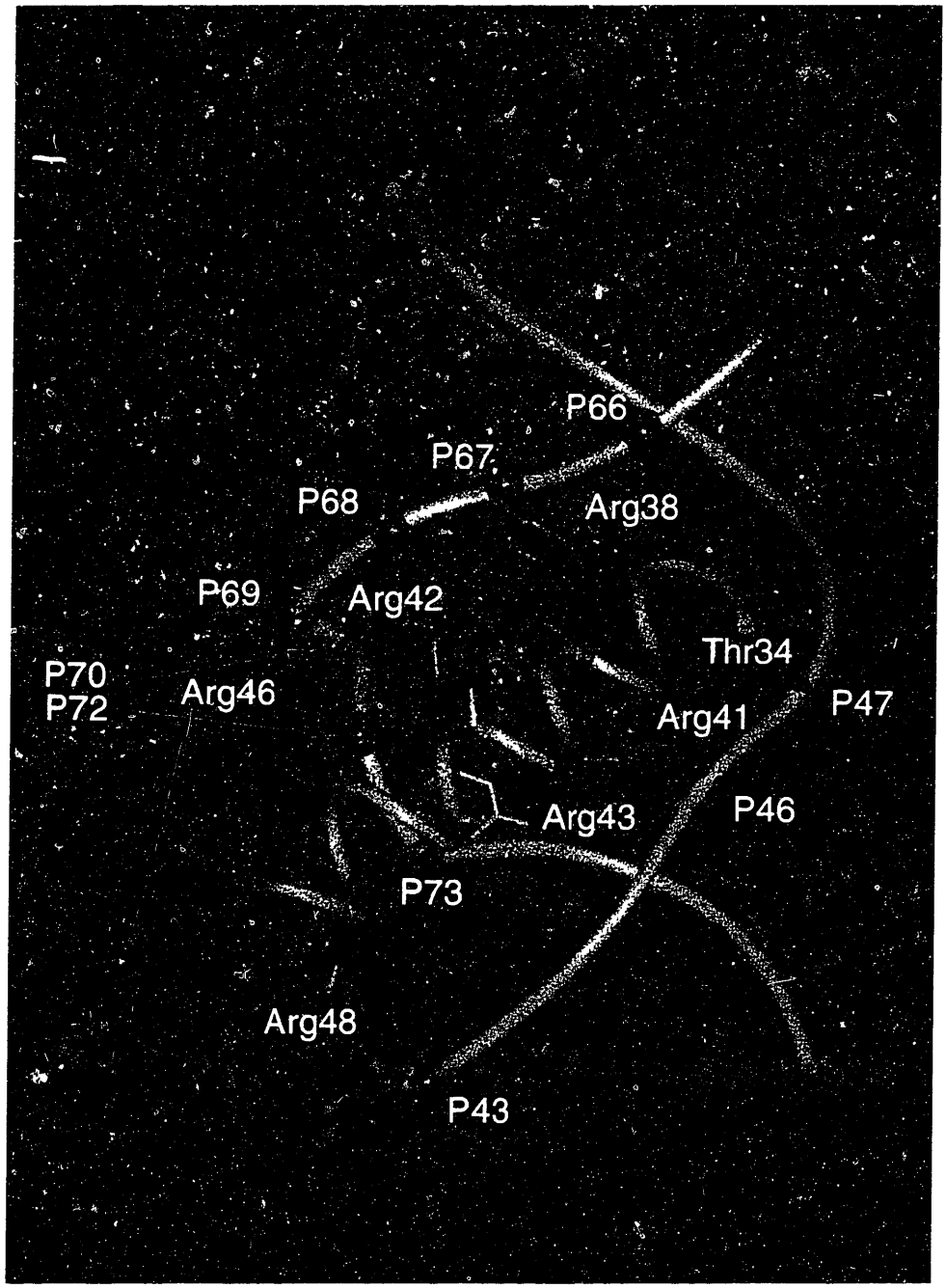


Figure 5.10: Phosphate backbone contacts. Six arginines and a threonine that are making potential electrostatic or hydrogen bond contacts with the phosphate diesteroxygens of the RNA are shown in red. The phosphorous atoms in the proximity of these amino acids are highlighted as blue spheres.

Three additional arginines (Arg^{42}, Arg^{43}, and Arg^{46}) on the A-face of the α-helix make phosphate contacts with the 3'-side of the hairpin. This especially arginine-rich face probably stabilizes the unfavorable electrostatics of the bulges on this strand, which bring many phosphates in close proximity. In particular, the unusual path of the phosphate backbone required for formation of the G48:G71 base pair positions the phosphorus atoms of G70 and U72 ~4 Å apart. The positively charged guanidium group of Arg^{46} is positioned near these two phosphates and probably helps stabilize the bound conformation of the RNA. Mutation of Arg^{46} to alanine, but not lysine, results in a 7-fold decrease in binding specificity, which is consistent with the interpretation that this arginine make a specific electrostatic contact (Tan and Frankel, 1994). Mutations that appear to remove structural elements which stabilize the structure of the G:G base pair (G71 2'-hydroxyl is another example) reduce binding specificity, consistent with the hypothesis that this structural element is essential for Rev binding.

Arg^{41} and Arg^{48} on the C-face of the peptide make contacts with phosphates on the 5'-side of the hairpin. Arg^{48} can be functionally substituted by lysine, while mutation to alanine results in a 3-fold decrease in binding specificity (Tan and Frankel, 1994), indicating that this arginine contributes a modest electrostatic interaction to binding specificity. Arg^{48} is near the phosphates of U43 and C44, and may be positioning the peptide relative to the lower stem of the RNA.

5.2.3 Van der Waals contacts Despite the very hydrophilic composition of the Rev peptide, there are several amino acids that make potentially important van der Waals interactions with the RNA (Figure 5.11). At the N-terminus of the helix, Ala^{37} and Gln^{36} anchor the B-face of the α-helix in the RNA major groove. Ala^{37} in particular makes extensive van der Waals contact with the ribose of G46, and may help stabilize the conformation of this nucleotide, which makes hydrogen bonds to Arg^{44}. The role of Ala^{37}

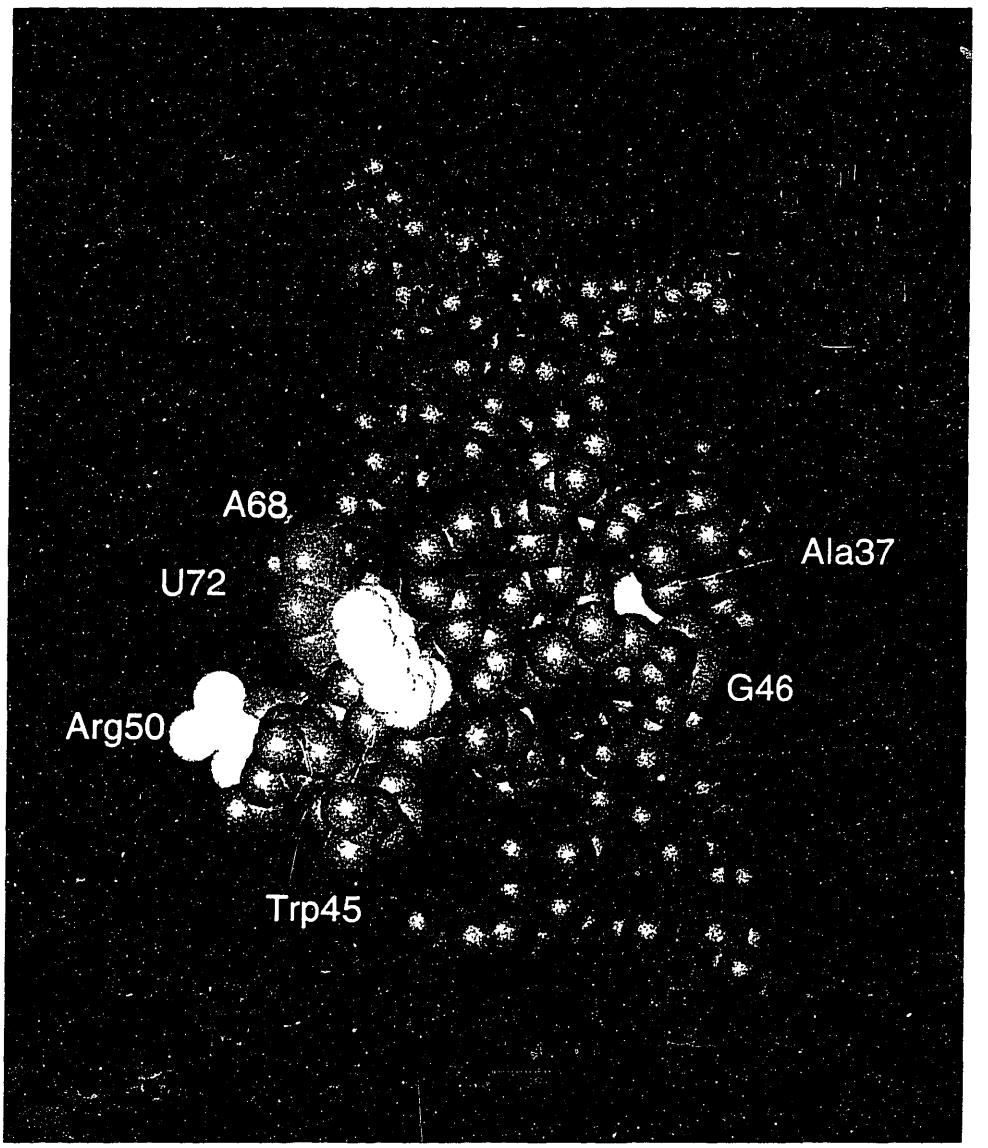


Figure 5.11: CPK representation of RRE-Rev complex with the RNA blue and peptide magenta. Van der Waals surface interactions between the peptide (yellow) and RNA (green) are highlighted. Two interactions with Arg^{43} and Gln^{36} cannot be seen in this view.

was not directly addressed by the alanine-scan mutagenesis data (Tan et al., 1993), and it is possible that this amino acid is important for specific binding through positioning of the helix in the major groove. This hydrophobic contact may explain why hydrophilic amino acids were not selected at this position (N3) to complete the N-capping box.

Although Gln^{36} has interactions between the β/γ methylene groups and the bases of G47 and G48, the side-chain amide group of Gln^{36} is not in a good orientation to make any hydrogen bonding contacts, consistent with mutagenesis data indicating that this amino-acid is not important for specific binding (Tan et al., 1993). Trp^{45} is in weak van der Waals contact with the bulged A68 base. In studies of the free RRE RNA, A68 is stacked on the G67 base (Battiste et al., 1994), and the tryptophan may help stabilize the bound conformation of the RNA where A68 is completely unstacked. Mutation of Trp45, however, has no effect on RNA-binding specificity (Tan et al., 1993). The Arg^{50} side-chain is in van der Waals contact with the unstacked base of U72, which again may help stabilize the bound conformation of the RNA and contribute to specific binding. Unlike Trp^{45}, Mutation of Arg^{50} to alanine results in a 4-fold decrease in binding specificity (Tan and Frankel, 1994). Arg^{50} is not near the phosphate backbone (>10 Å); therefore, an electrostatic interaction does not appear to account for the importance of this amino acid.

5.2.4 Summary of contacts A full summary of all the RNA-peptide contacts, discussed in detail above, is shown schematically in Figure 5.12. Overall, a role in specific binding can be assigned to all nine amino acids that mutational data identified as important for specific peptide binding, and in general these amino acids are interacting with important nucleotides or phosphates of the RNA (Figure 5.13). Three arginines (Arg^{35}, Arg^{39}, and Arg^{44}) make specific hydrogen bonds with Watson-Crick base pairs that are made accessible by the structure formed by the purine-purine base pairs. In addition, the purine-purine G47:A73 base pair is recognized by Asn^{40}, which is the first example of hydrogen

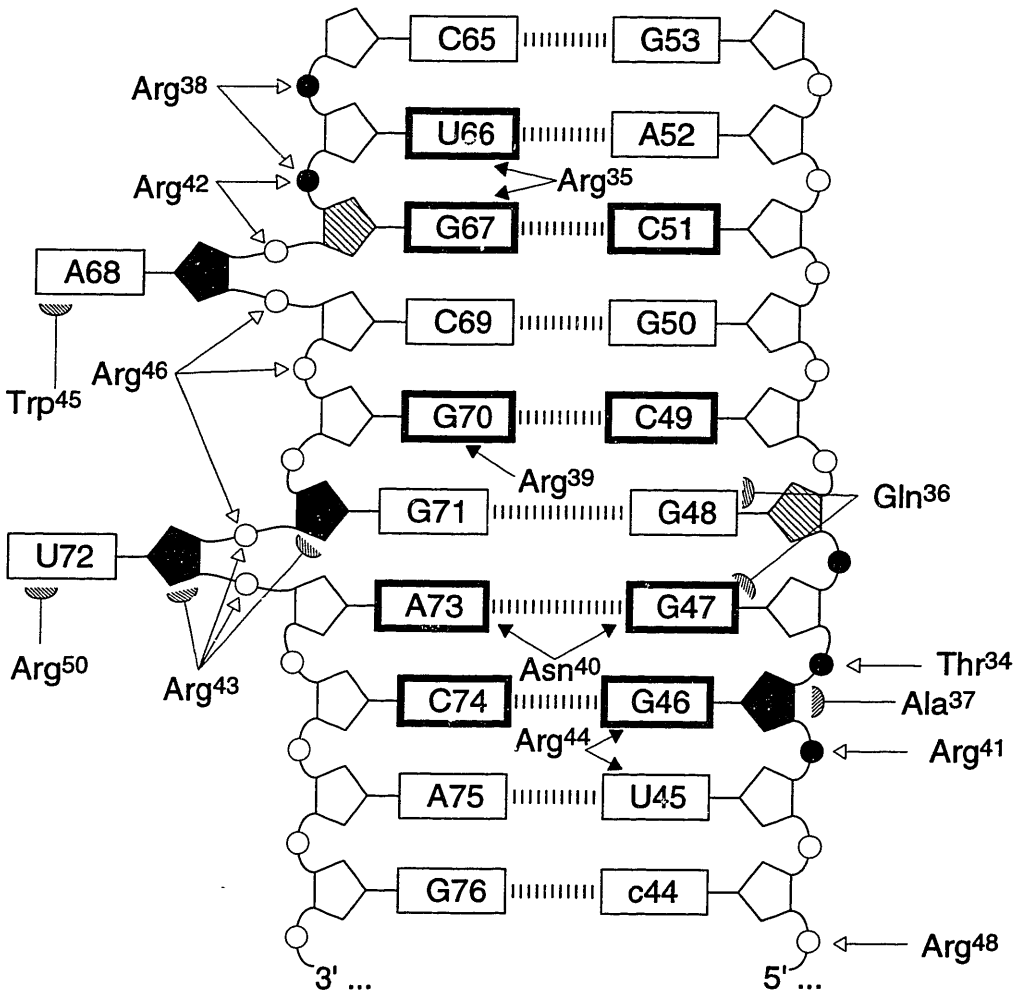


Figure 5.12: Schematic of specific RNA-peptide interactions. Black arrows, open arrows, and hatched arcs represent base-specific, phosphate backbone, and van der Waals contacts, respectively. For arginine-phosphate contacts, the arrows represent general proximity, and do not distinguish between hydrogen bonds and simple electrostatic contacts (see text). The RNA is depicted as in Figure 3.10.

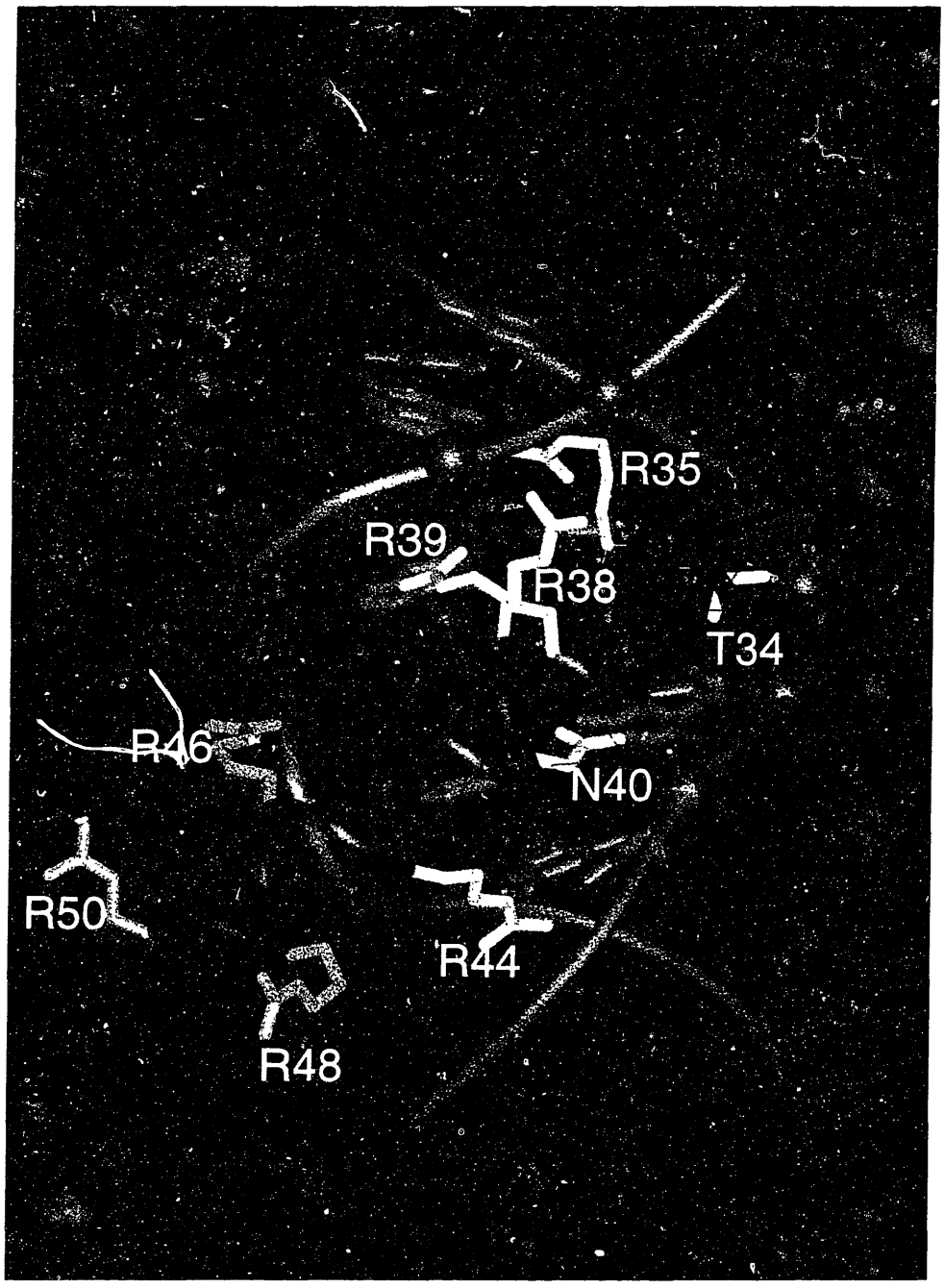


Figure 5.13: Correspondence of structure with biochemical data. Important nucleotides and amino acids are represented as thick bonds (see Figure 1.9 and 1.10). Amino acids that result in a >10-fold decrease in peptide binding specificity upon mutation to alanine are shown in yellow. Arginines that give 2 to 10-fold decrease in peptide binding specificity, but can be substituted with lysine without any decrease in binding specifity are shown in orange. The RNA is depicted as in Figure 5.8A.

bonding to a mismatch base pair in a RNA-protein complex. Thr^{34} and Arg^{38} make important phosphate contacts flanking the region of base-specific contacts. Arg^{46}, Arg^{48}, and Arg^{50} at the C-terminus of the peptide, which mutational data suggest are moderately important for specific binding (Tan et al., 1993), are making phosphate and van der Waals contacts that may help orient the C-terminus of the peptide in the groove, stabilizing the important interactions at the N-terminus of the peptide.

5.3 Comparison to Other DNA- and RNA-Protein Interactions

5.3.1 α-helix/major groove recognition The RRE-Rev complex is the first example of an RNA-binding protein that uses an α-helix for major groove recognition, and its general features are very similar to many DNA-binding proteins also containing recognition helices. Like most DNA-binding proteins, the N-terminus of the helix makes the most significant contacts with the RNA, including sequence specific contacts with the major groove edges of nucleotide bases and phosphate backbone contacts flanking the helix binding-site. These contacts are made by arginines, an asparagine and a threonine, which are common amino acids for fulfilling these roles in DNA-protein interactions. However, there are two major differences in the global structure of the RRE-Rev complex. As noted earlier, the recognition helix is more deeply embedded in the groove, which is likely a consequence of the differences in DNA/RNA helical morphology. In addition, the recognition helix itself is sufficient for specific binding, unlike most DNA-binding proteins, where amino acids outside the "recognition" helix stabilize the bound form of the protein and/or contribute important contacts with the phosphate backbone for positioning and orientation of the recognition helix (Pabo and Sauer, 1992). In this respect, however, the Rev RNA-binding domain bears the most similarity to the basic region-leucine zipper (bZIP) and basic region-helix-loop-helix (bHLH) families of DNA-binding proteins.

Besides the obvious similarity of a high density of basic residues at the RNA/DNA interface, the recognition helices of these DNA-binding proteins are also isolated from direct protein-phosphate backbone contacts from other portions of the protein. GCN4, for example, contains contacts with the DNA along almost the entire circumference of the recognition helix, similar to Rev, including several arginine-phosphate contacts along the solvent exposed side of the helix (Ellenberger et al., 1992). The high density of basic residues in these motifs is probably necessary to stabilize and position these isolated helices in the groove of the RNA/DNA. Unlike Rev, however, the basic regions of the bZIP and bHLH proteins require dimerization domains distal to the point of DNA contact to indirectly position the helices for DNA recognition (Cuenoud and Schepartz, 1993), though there is some evidence that the basic region can bind specifically in the absence of the dimerization domains (Park et al., 1996). While Rev binds as a monomer to its high-affinity site, it contains oligomerization sequences adjacent to the RNA-binding domain and requires multimerization on a much larger RRE structure for function (Malim and Cullen, 1991; Zapp et al., 1991). How multimerization occurs and whether it includes well-defined interactions with the arginine-rich motif will await structural studies of larger fragments of the Rev protein (see section 5.4.1).

5.3.2 The arginine-rich motif Proteins with the short arginine-rich motif (ARM) or the β-sheet RNP motif (RRM) form the two most prevalent "families" of RNA-binding proteins (Mattaj, 1993; Burd and Dreyfuss, 1994). Families of proteins identified by sequence comparisons typically contain a common three dimensional fold that relates to their function, as is observed for the RRM motif. In contrast, there are now two structural examples of ARM peptides bound to RNA, and each peptide is in a different secondary structure context (β-turn and α-helix for BIV Tat/TAR (Puglisi et al., 1995; Ye et al., 1995) and HIV Rev/RRE, respectively). Therefore, the ARM does not appear to contain a

common tertiary-fold that creates a scaffold for presentation of amino acids to the RNA/DNA, as is the case for most other families of nucleic acid-binding proteins (Pabo and Sauer, 1992). This confirms much biochemical data, which has suggested that the ARM does not comprise a unique structural motif (Burd and Dreyfuss, 1994; Chen and Frankel, 1994; Harada et al., 1996). Despite the differences in structure, these peptides share two common features. One is deep penetration of the peptide element into the major groove of the RNA-binding site (Figure 5.14). Both of the RNA-binding sites contain distortions that form a deep and wide major groove, creating a binding "pocket" for the peptide. The other feature, which is characteristic of most ARM proteins, is that the RNA-binding regions comprising the ARM are capable of folding and mediating specific recognition as simple peptides, separate from the remainder of the protein (Calnan et al., 1991; Weeks and Crothers, 1991; Kjems et al., 1992; Tan et al., 1993; Chen and Frankel, 1994; Tan and Frankel, 1995). Both of these features can be aided by the high density of arginines, which is only characteristic feature of the ARM. The ARM may simply provide electrostatic stabilization of deep penetration into the major groove that allows for a large and specific interface with the RNA and probably permits the 10-20 amino acid ARM to bind as a structurally independent element, possibly being advantageous for small viral proteins where this motif is commonly found. The highly distinctive RNA structures formed by the ARM binding site, however, undoubtedly also contribute to the ability of these loosely structured peptides to fold and bind RNA without the structured context of larger protein domain.

5.3.3 Induced fit of RNA-protein interactions The results in Chapter 3 showed that the structure of the RNA was stabilized and/or induced into a particular conformation by binding of the Rev peptide. Circular dichroism studies have also shown that the α-helical conformation of the Rev peptide is stabilized upon RNA-binding (Tan and

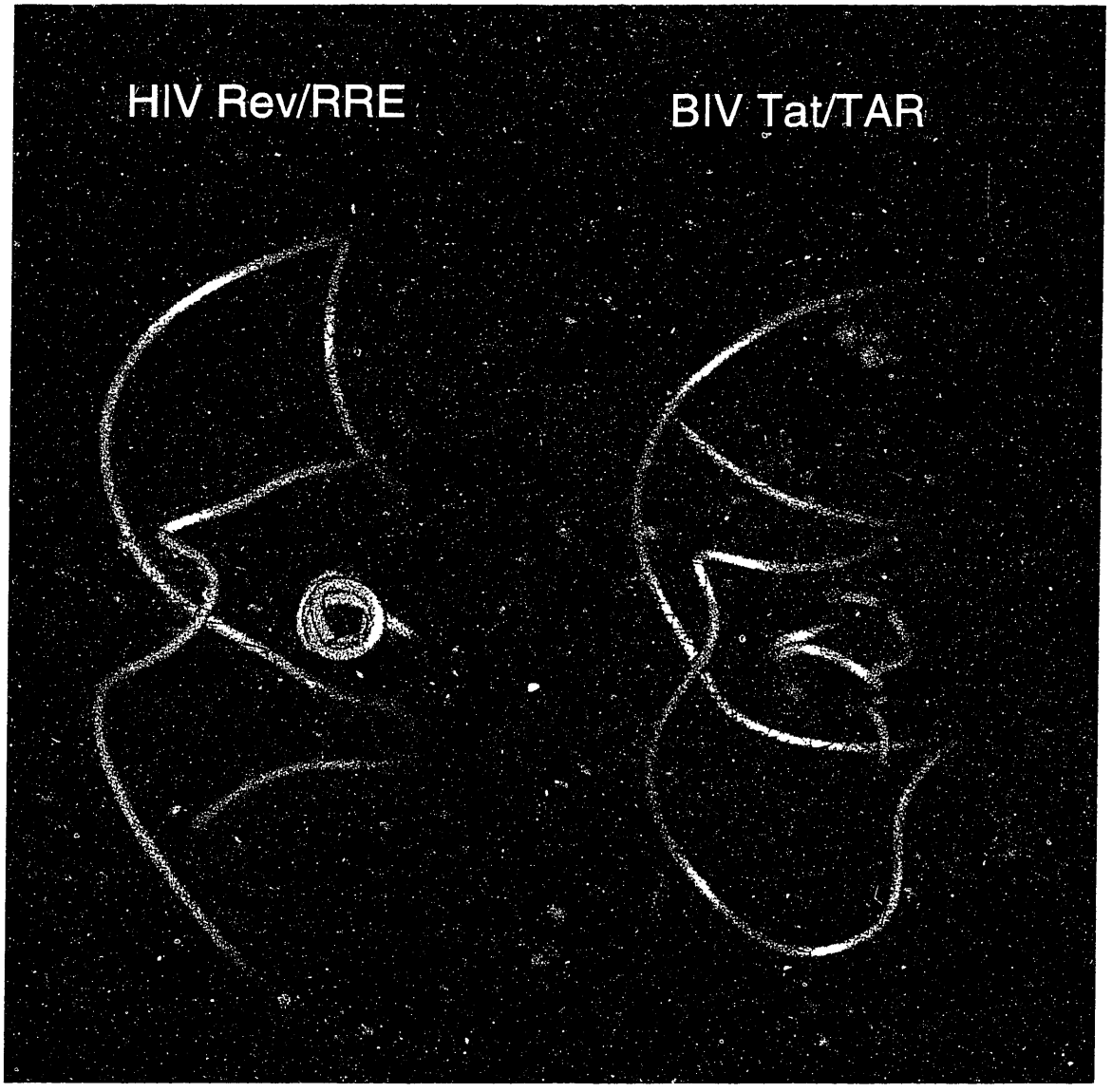


Figure 5.14: Comparison of HIV Rev-RRE complex with the BIV Tat peptide - TAR RNA complex also determined by NMR spectroscopy. The RNAs are shown in blue with the peptide backbones in magenta. The view is down the major groove axis of the peptide binding sites.

Frankel, 1994), though it is unknown whether the rest of the Rev protein stabilizes the α-helical conformation of the arginine-rich region in the unbound full-length protein. NMR studies of the free and bound peptide are also consistent with the stabilization of α-helical structure upon RNA-binding (data not shown). The RRE-Rev complex contains a large binding interface and a network of interdependent interactions that may help explain how folding of the RNA and peptide may be coupled to binding and contribute to specificity (Spolar and Record, 1994). The N-cap structure was not observed in NMR studies of the Rev peptide in trifluoroethanol (helix stabilizing solvent), and the phosphate backbone contact of Thr^{34} may assist formation of the N-cap in the bound peptide. In addition to the N-cap interaction, the arginine-rich A-face of the peptide and the 3-side of the RRE hairpin form a complementary electrostatic interface that probably stabilizes the bound form of both the peptide and RNA. The electrostatic repulsion of arginine side chains on the B-face probably destabilizes the α-helical conformation of the unbound Rev peptide. The unusual phosphate backbone structure on the 3'-side of the bound RRE hairpin contains close phosphate groups, creating an electrostatic repulsion that undoubtedly destabilizes the free RNA. These two elements interact in the complex mutually stabilizing one another (interaction of Arg^{46} and U72/G70 PO_4's is a good specific example (section 5.2.2)). This type of complementary electrostatic interaction is observed in many transcription-factor DNA-protein complexes, including the bZIP and bHLH proteins, which also undergo induced fit mechanisms with a α-helical disorder->order transitions upon DNA-binding (Frankel and Kim, 1991).

The conformational changes in the RRE-Rev complex are not unusual, and an emerging theme in the formation of RNA-protein complexes is mutually induced fit, where both the protein and RNA become ordered or change conformation upon binding. In all cases where the structures of both free and bound forms of RNAs and proteins that form specific complexes are known, there is a considerable difference in the conformations.

Binding of glutaminyl-tRNA synthetase to $tRNA^{Gln}$ results in disruption several base pairs upon binding (Rould et al., 1989). HIV TAR, BIV TAR, and HIV RRE RNA all undergo significant conformational changes upon protein binding, and at least in the case of BIV Tat and HIV Rev, the peptide conformation is considerably stabilized upon binding (Puglisi et al., 1992; Battiste et al., 1994; Tan and Frankel, 1994; Aboul-ela et al., 1995; Puglisi et al., 1995; Ye et al., 1995). In U1A protein- and MS2 coat protein-RNA interactions, rearrangements of RNA loop regions and parts of the protein structure are observed (Valegard et al., 1994; Allain et al., 1996). A conformational change upon binding necessarily entails an entropic energetic cost for changing conformation or for ordering of a disordered region. Detailed thermodynamic analysis of DNA-protein complexes where the protein undergoes a conformational change upon DNA-binding have suggested that burial of exposed hydrophobic surface upon folding is an energetic driving force (Spolar and Record, 1994). Similar interactions are likely to be occurring for RNA-protein complexes, including Rev; nevertheless, this does not explain the propensity of RNA-protein interactions to contain conformational changes. Perhaps the flexibility inherent in RNA may offer a mechanistic advantage for binding, where intimate interfaces can be made that would not be possible by rigid body docking.

5.3.4 Contribution of RNA Structure to binding specificity The wide array of structures that RNA can form is a distinguishing characteristic from DNA, which is primarily limited to Watson-Crick base pairing *in* vivo. Potentially, proteins can utilize these unique RNA structures for specific-recognition. Many examples of recognition of distinctive RNA structural features have been observed in the few structural studies of RNA-protein complexes. tRNA synthetases make contacts that recognize the L-shape of tRNA, and in the case of $tRNA^{Ser}$, also recognize the long variable arm extension (Rould et al., 1989; Ruff et al., 1991; Biou et al., 1994). In HIV TAR RNA, a base triple is

proposed to be an integral part of creating a binding pocket for an arginine side-chain, though it does not contribute any hydrogen bonds to the arginine (Puglisi et al., 1992; Puglisi et al., 1993). In RRE RNA, the G48:G71 base pair is structurally important to widen the major groove, creating a binding pocket for the peptide, but does not directly contact the peptide. Although the G47:A73 base pair makes a sequence-specific contact with the Asn^{40} side chain, it may also contribute to structure-specific recognition. The N7 position of A73 is not apparently utilized by the peptide; therefore, the remaining three hydrogen bond donors and acceptors on the G:A major groove face are indistinguishable from a G:C base pair. This suggests that groove widening by the G:A base pair (larger C1'-C1' distance) may also help to properly configure the RNA structure, since G:C substitutions are not tolerated at this position. Besides widening the major groove to permit insertion of the α-helix, the overall structure surrounding the two purine-purine base pairs may provide a distinctive shape to the major groove that the peptide can utilize for specific recognition. The width of the major groove tapers off across the Rev binding site (Figure 5.5), creating a wedge shape to the groove, with the N-terminus of the peptide tucked into the narrower portion of the groove (see Figure 5.13). The phosphate contacts at both ends of the peptide could potentially be recognizing this shape of the groove with Arg^{46} and Arg^{48} spanning the wider portion near the C-terminus and the smaller Thr^{34} and Arg^{38} at the N-terminus (see Figure 5.10 or Figure 5.13). It seems remarkable that RNA elements as small as TAR and RRE can provide such specific three dimensional frameworks for protein binding, and the Rev-RRE complex once highlights the importance of RNA structure for protein recognition.

5.4 Future Research

5.4.1 Larger fragments of RRE and Rev The minimal peptide system was useful for NMR analysis by circumventing the aggregation problems of the protein; however, many interesting and unresolved questions regarding the structure and function of the intact Rev protein obviously remain. Even though the peptide was capable of binding to the high-affinity RRE site as a fully independent unit, it will be interesting to observe what interaction the arginine-rich motif (ARM) has with the remainder of the protein both free and in complex with the RNA. Is the α-helix that recognizes the RNA fully formed in the free protein or disordered? Are there any additional contacts with the protein that stabilize the α-helix in the bound RNA? If the hypothesis that the ARM represents a completely independent recognition element is correct, then interactions with the protein directly impacting on RNA-binding should be minimal.

Regardless of the impact of the remainder of the Rev protein on binding to the high-affinity RNA site, multimerization and binding to additional RNA sites on the RRE is an important functional question. Circular dichroism studies have suggested a helix-helix interaction between hydrophobic amino acids N- and C- terminal to the RNA-binding domain (Auer et al., 1994). Since mutations in these regions effect multimerization, this putative helix-helix interaction could potentially be important for multimerization. Figure 5.15 how a hypothetical model of how this proposed interaction is reasonable given the structure of the RNA-binding domain complex. The C-terminus of Rev was extended as a α-helix until a glycine residue at position 56. In the extended helix, there are two conserved hydrophobic residues that are oriented towards the solvent exposed side of the RNA-binding region. This is consistent with the model, since the N-cap also positions the N-terminal portion of Rev perpendicular to the α-helix axis towards the solvent-exposed side. Circular dichroism and sequence analysis suggested that amino acids 17-26 form an

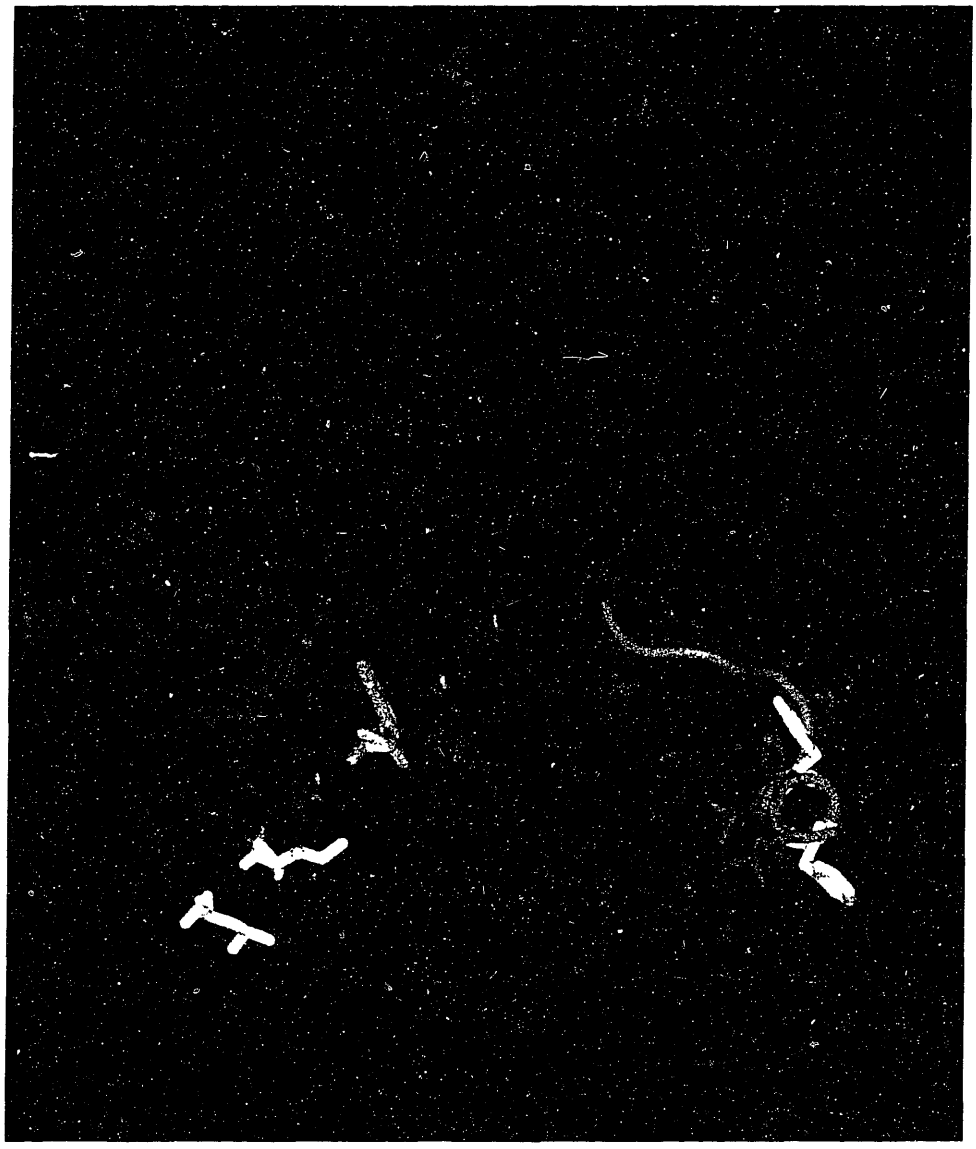


Figure 5.15: Modeling of potential helix-helix interaction in regions outside of the RNA-binding domain of Rev. The Rev α-helix was extended by 5 amino acids towards the C-terminus. Trp^{45} is shown in orange and Ile^{52} and Ile^{55} are shown in yellow. The N-terminus of the Rev protein from 17-33 is shown in green with amino acids 17-26 in an α-helical conformation (see text) and the proline-rich region from 27-33 in an extended conformation. Hydrophobic and aromatic residues in the helix are shown in purple and white, respectively. The placement and orientation of the N-terminal region is chosen to show that a possible hydrophobic patch that could interact with the C-terminal region of the RNA-binding domain exists. No other structural information is implied.

α-helix (Auer et al., 1994), which would contain hydrophobic residues clustered on one face of the helix (Figure 5.15). An interaction between these two hydrophobic patches is feasible, since the amino acids 27-33 are proline-rich and could be forming a loop connecting the two helices. Structural studies of larger fragments of the Rev protein containing more of the N-terminus that might possibly form a folding domain or multimerize on a larger RRE fragment would be very significant to understanding in vivo Rev function. The large size of any multimers precludes analysis by NMR spectroscopy, and X-ray crystallography would need to be pursued.

5.4.2 Drug binding Antibiotic aminoglycosides, such as neomycin, have been shown to specifically bind the RRE, albeit with lower affinity than the Rev peptides (Zapp et al., 1993). It has been proposed that RNA-peptide interactions in the major groove bear resemblance to protein and drug-interactions in the minor groove of DNA through deep, shape selective binding (Chen and Frankel, 1995). The observation of a small drug binding to the RRE is consistent with this view. The aminoglycosides are highly charged, consistent with necessity for electrostatic stabilization of deep major groove penetration as detailed for the Rev peptide above. Since the RNA is partially structured in the free form, it would be interesting to observe the differences in the peptide- and neomycin-bound forms of the RNA. If shape-selective recognition of the major groove is occurring, the differences should be minor On the other hand, the conformational flexibility of the RNA may allow neomycin to induce a different RNA structure suitable for recognition of the different hydrogen bond donors and acceptors presented. In vitro selection studies have shown that neomycin prefers an A:A rather than a G:G pair at positions 48 and 71, suggesting sequence-specific recognition for this base pair, which appears to play only a structure-specific role in Rev binding (Werstuck et al., 1996).

5.4.3 Site directed mutants of RRE and Rev; Understanding the general features of RNA-protein recognition will require comparison of numerous structural examples to identify trends and dissimilar features. Analogously, a complete understanding of the contributions of specific interactions in the RRE-Rev complex will require structural comparison of many mutants that test various hypothesis about functional roles of specific amino acids or nucleotides. Hypothesis about the importance of particular amino acids or nucleotide can be obtained from both existing biochemical data and analysis of the structure presented here.

Extensive mutational data on the sequences of the RRE important for the high-affinity 1:1 Rev complex have obtained by in vitro selection (Bartel et al., 1991; Giver et al., 1993; Jensen et al., 1994). The most intriguing mutations are the covariations of the G48:G71 base pair to A48:A71 and occasionally C48:A71. Both A:A and C:A can form structurally similar (isosteric) base pairs to the G:G base pair in the RRE-Rev complex described above. If the hypothesis that the phosphate backbone distortion is structurally important for formation of the Rev binding site, then the A:A and C:A base pairs should form similar structures. Initial characterization in our laboratory has indicated that induction of base pair formation in the internal loop for RRE RNAs containing these two mutations occurs upon binding of Rev peptide, similar to wild-type RRE (H. Mao and J.R. Williamson, unpublished results). However, it is not yet clear what the geometry of the backbone is for nucleotide 71 in either complex.

Another major result from the selection studies is the isolation of many different elements for the upper stem of the RRE. Non-conserved mismatch base pairs were isolated for the Watson-Crick base pair G50:C69 (Bartel et al., 1991; Giver et al., 1993) indicating that conformational flexibility may be advantageous for binding in this region. In addition, an entirely different upper stem element with a two base uridine bulge was isolated by two independent groups (Giver et al., 1993; Jensen et al., 1994). Interestingly this element

contains all of the important nucleotides for formation the HIV TAR RNA structure (J.R. Williamson, personal communication) that binds HIV Tat peptides and arginine. In the structure of the RRE-Rev complex, the Arg^{35} side chain extends to make the only base contact in this region of the RNA with U66 and G67 (Arg^{38} also makes potential phosphate contacts with U66 and G67). It is intriguing that the selected "TAR" structure is creating a binding pocket for Arg^{35} in the upper stem. Modeling studies merging the structure of TAR RNA ((Puglisi et al., 1992), A. Brodsky, unpublished results) with RRE RNA indicate that a reasonable RNA structure could be formed combining the two elements that would not require significant rearrangements of the peptide for binding of Arg^{35} in the TAR arginine pocket (data not shown). Structural studies confirming this hypothesis would be significant, suggesting that TAR RNA can be considered a modular arginine-binding element and have implications for the molecular design of RNA-protein complexes. Conversely, it is possible that the Rev peptide is binding to the selected RNA in manner proposed for HIV Tat with a single arginine providing the specificity, and the normal Rev binding mode is impaired.

In addition to the selection of RNA elements important for recognition of Rev, selection of novel peptides that bind to the RRE have also been performed (Harada et al., 1996). An interesting result from this study was that arginine/glycine-rich peptides that specifically bound to the RRE were identified. Since it is unlikely that these peptides could form α-helices, some other peptide structural motif must be recognizing the RRE. It is intriguing that this selected peptide may be binding the RRE as a β-hairpin analogously to the BIV Tat/TAR complex, which also contains numerous arginines and glycine. Deep penetration of the major groove was a common feature of both BIV-Tat and HIV-Rev structures and a glycine-rich β-hairpin peptide may be useful motif for penetration and recognition of major groove bases.

A genetic screen of non-functional RRE mutants has identified many compensatory Rev mutations that restore function (C. Jain and J. Belasco, personal communication). This is an elegant genetic approach to identifying important RNA-protein interactions at the amino acid/nucleotide level. An interesting mutation selected, which helped confirm the interaction of Asn40 with the G47:A73 base pair in the RRE-Rev complex was Asn40->Gln40. This mutation restored function to an RRE with a A73G mutation (G:A -> G:G base pair). Structural studies of this double mutant would provide additional examples of how mismatch base pairs can be recognized by protein side chains. In addition, it may shed light on the important features of the Asn40-G:A contact in wild-type Rev.

Analysis of the RRE-Rev structure presented here raises many questions that can be addressed by biochemical or structural studies. Mutational studies had not identified Ala^{37} has a critical amino acid (though alanine-scan mutagenesis would not test this position (Tan et al., 1993)), yet this amino acid make a hydrophobic contact with the ribose of a critical nucleotide, and its backbone amide is involved in the N-cap hydrogen bond. Mutational studies with a series of hydrophobic amino acids might address the potential hydrophobic role of position 37, though larger side chains may not sterically fit at this position. In addition, if the N-cap is important for stabilizing the Rev α-helix in the bound form, it is interesting that a canonical capping box structure with a hydrophilic amino acid at position 37 making a hydrogen bond to backbone amide of Thr^{34} has not evolved. A canonical capping box could be engineering into Rev to see whether increased helix stability via the N-cap could increase binding specificity. Conversely, if these substitutions impaired binding, it might provide evidence that Ala^{37} contributes to specificity in the RRE-Rev complex.

Mutational studies suggested that Arg^{50} was making an electrostatic contact with the RRE, since substitution with alanine, but not lysine, decreased binding-specificity. In the complex, however, Arg^{50} is not close to the phosphate backbone, unlike Arg^{46} and Arg^{48}

which have similar mutational properties. Instead Arg^{50} appears to make a hydrophobic contact with the aliphatic portion of its side chain and the unstacked base of U72, which may stabilize the unusual structure of the internal loop. If this hydrophobic contact is truly important for binding, substitution with large hydrophobic amino acids, such as leucine or isoleucine should have improved or unimpaired binding properties. The interaction of isoleucine with a bulged uracil stabilizing the bound RNA configuration was observed in the BIV Tat/TAR complex (Puglisi et al., 1995) and the functional importance of this interaction has been shown by mutagenesis (Chen and Frankel, 1995). Therefore, it is reasonable to assume that this contact may be potentially important for Rev-binding.

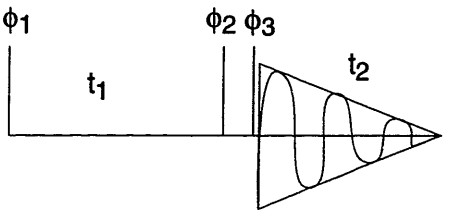
Taken as a whole, the above proposed studies should lead to a better understanding of the critical elements of specificity for the RRE-Rev interaction. With a greater understanding, it should be possible to begin engineering minimal binding elements of peptides or drugs with improved binding properties that might be useful as therapeutics to inhibit this essential HIV interaction.

APPENDIX A: NMR PULSE SEQUENCES

This appendix is intended to give a brief overview of the types of NMR experiments that were used in this thesis with more specific details of how the experiments function than is provided in the text. Descriptions of the state of magnetization at various points in each experiment will occasionally be given in product operator formalism (Sorensen et al., 1983; Shriver, 1992); however, a description of this formalism or the derivation of these states is beyond the scope of this thesis. For more detail, the original citations for the experiments should be reviewed. Many different phase cycles are possible for the same experiment. The phase cycles used are shown in the figures and in most cases correspond to the phase cycles described in the references given in Table 2.1.

Two Dimensional Homonuclear NMR Experiments Figure A.1 shows a diagram of the pulse sequences for the three basic 2D NMR experiments used for structural analysis of biological molecules (Wuthrich, 1986). These experiments share several basic features. Transverse magnetization (I_x or I_y) is created by a 90° pulse, followed by delay (t_1) which is incremented at fixed intervals to indirectly detect the chemical shift of the protons ("evolution" period). The next stage of these experiments is the "mixing" period where magnetization is transfer from one proton to another. The three experiment differ in how magnetization is transferred, which gives rise to the characteristic features of each spectrum. The COSY experiment only has a 90° pulse and cross peaks are only detected for protons which have direct J-coupling (J-coupling evolution occurs during t_1). The double-quantum-filter (DQF) adds an additional 90° editing pulse and phase cycle that removes diagonal coherences that are 90° out of phase with the crosspeaks, improving spectral quality (Muller et al., 1986). The TOCSY experiment contains a spin-lock pulse that keeps the magnetization along the x- or y-axis where through-bond transfer of

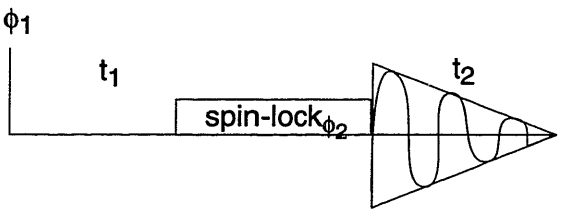
(A) DQF-COSY



ϕ_1=[4(x), 4(y)]$_2$, [4(-x),4(-y)]$_2$
ϕ_2=4(x), 4(y), 4(-x),4(-y)
ϕ_3=x, y,-x,-y,y,-x,-y,x

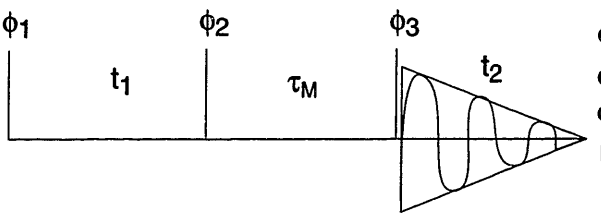
rec=(x,-y,-x,y,y,x,-y,-x),
(-x,y,x,-y,-y,-x,y,x)$_2$,
(x,-y,-x,y,y,x,-y,-x)

(B) TOCSY



ϕ_1=y,-y,-x,x
ϕ_2=[y,-x,-y,x]$_2$
rec=y,-y,-x,x,

(C) NOESY



ϕ_1=[x, -x]$_4$, [y,-y]$_4$
ϕ_2=8(x), 8(y), 8(-x),8(-y)
ϕ_3=[x, y,-x,-y]$_2$, [y,-x,-y,x]$_2$
rec=(x,-x,y,-y,-x,x,-y,y),
(y,-y,-x,x,-y,y,x,-x),
(-x,x,-y,y,x,-x,y,-y),
(-y,y,x,-x,y,-y,-x,x)

Figure A.1: 2D Homonuclear NMR pulse sequences. **(A)** Double-Quantum-Filtered (DQF) COrrelation SpectroscopY (COSY) experiment. **(B)** TOtal Correlation SpectroscopY (TOCSY). **(C)** NOE SpectroscopY (NOESY). t_1 and t_2 are the indirectly and directly detected proton evolution times, respectively. τ_M is the mixing time of the NOESY experiment. The phase cycle of each experiment is given to the right.

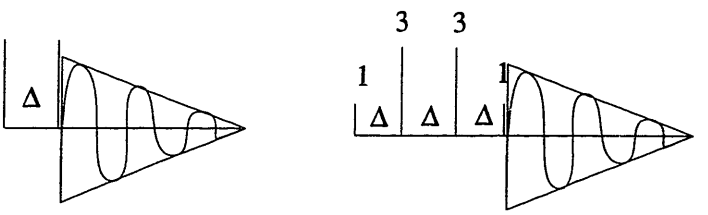
magnetization can occur through isotropic mixing (Braunschweiler and Ernst, 1983). At longer mixing times magnetization can sequentially transfer through a spin system, giving rise to cross peaks between resonances that are not directly coupled, but are indirectly connected in the spin-system. In the NOESY experiment, a 90° pulse following t_1 places the desired magnetization on the z-axis, where dipolar relaxation between spins in different energetic states (Nuclear Overhauser Effect) can occur (Kumar et al., 1980). The resonances do not need to be connected through bonds for this interaction to occur, and the efficiency of transfer has a $1/r^6$ dependence on the distance separating the two protons. The series of radiofrequencies pulses in these experiment is not sufficient to only observe the interactions described above, and phase cycling must be done to select the desired coherences. Phase cycles for each experiment in the appendix are given in the figures, yet a detailed description of how each phase cycle functions will not be given (see references).

Water Suppression Schemes Since many protons attached to nitrogen in biological molecules exchange rapidly with solvent, methodologies for detection of protons in non-deuterated solvents is essential. The concentration of the water resonance (55 M) is in great excess of most samples of biological molecules (~ mM), and the water resonance must be suppressed due to dynamic range problems associated with the analog to digital conversion (ADC) of the NMR signal. The simplest method for achieving water suppression is selective presaturation of the water resonance with a low power radiofrequency pulse for 1-2 seconds before the first pulse of the NMR experiment. There are several caveats to this approach, the most serious being loss of magnetization through exchange with saturated water resonances during the course of the experiment. This is particularly a problem for nucleic acids, where imino proton resonance intensity is completely lost due to exchange with water during the presaturation period. The intrinsic exchange rate of amide protons of proteins is slow enough to prevent complete saturation of resonances; however, a

significant loss of intensity is typically observed. Figure A.2 shows three methods of suppressing water resonances without pre-saturation. The first two methods use the same basic principle of applying a series of hard pulses and delays that result in no net excitation on resonance (H_2O), yet produce a 90° pulse at a off-resonance frequency maximum. This approach has the disadvantage of producing phase distorted spectra with uneven excitation profiles (net excitation at a given frequency off-resonance). Gradients can also be used to suppress water in a variety of fashions. The most popular method is WATERGATE, which applies a gradient to dephase all transverse magnetization, then uses a selective "3-9-19" 180° pulse train (similar to the binomial sequences) to flip the states of all spins off-resonance (H_2O receives no net excitation). A second gradient then refocuses only the magnetization that received the 180° pulse (not H_2O) for detection. The disadvantage to the WATERGATE method is that while H_2O is incoherent, it is still excited during acquisition and can diminish intensity of resonances of interest through exchange or relaxation (though the effects are not as severe as presaturation). This problem can be mitigated in some multidimensional experiments that use WATERGATE elements by appropriate phase cycling that places H_2O on the z-axis before the WATERGATE sequence, eliminating dephase transverse magnetization for the bulk of the water (e.g. FHSQC (Mori et al., 1995)). Another approach uses "soft" (low power), often shaped pulses to selectively excite the on-resonance frequency (set to H_2O), followed by a hard 90° pulse, which places water back on the z-axis while creating transverse magnetization for all other resonances. This type element has been incorporated in the HSQC experiment ("Flip-back" HSQC), which also has the important property of not saturating H_2O, increasing signal intensity of exchangeable protons (Grzesiek and Bax, 1993).

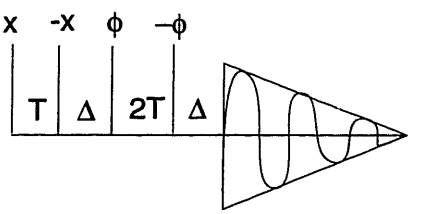
Two Dimensional Heteronuclear Correlation Experiments: The two basic experiments to correlate a proton through-bond to a heteronuclear resonance (^{13}C, ^{15}N, or ^{31}P for

(A) Binomal (11(jump return) and 1331)



$\Delta = 1/\nu_{max}$ where ν_{max} is the desired intensity maximum from the carrier in Hz.

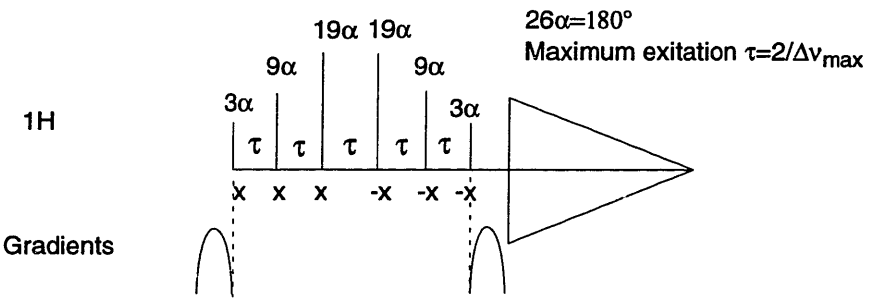
(B) Jump-return-echo



ϕ=x,y,-x,-y
rec=x,-x,x,-x

T=1/offmax Δ=10-500 μs excitation profile=$\sin^3(2\pi T\Delta)$

(C) Watergate

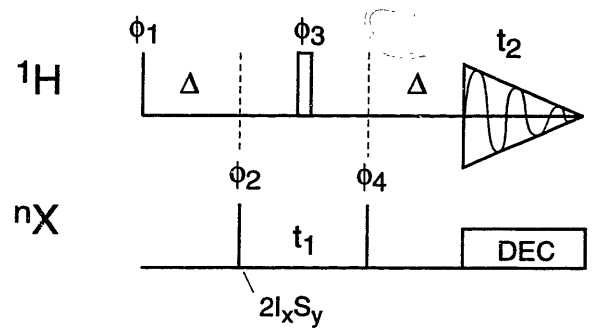


26α=180°
Maximum exitation $\tau=2/\Delta\nu_{max}$

Figure A.2: Water supression pulse schemes. **(A)** and **(B)** are non-gradient methods of water supression without presaturation. The last method **(C)** utilizes principles of the first two in combination with gradients..

biological samples) are Heteronuclear Multiple Quantum Coherence (HMQC) and Heteronuclear Single Quantum Coherence (HSQC) (Wagner, 1989). The difference in coherence (single or multiple) refers to the state of magnetization during the heteronuclear evolution time ($2I_xS_y$ for HMQC and $2I_zS_y$ HSQC; I and S refer to proton and heteronuclear magnetization, respectively). Practically, the HMQC is more sensitive than the HSQC due to fewer pulses (pulse imperfections lead to losses in intensity). However, gradients can be effectively added to the HSQC to produce high quality spectra (fewer artifacts) with a short phase cycle (see Figure A.5B). An additional advantage to the HSQC is that ^1H-^1H couplings (~3-10 Hz) are not active during t_1 (^1H magnetization is along z-axis), which removes splittings in the crosspeaks that complicate analysis. Thus, HSQC was the method of choice for obtaining ^{15}N-HSQC spectra of labeled proteins. With ^{13}C-labeled samples, however, both experiments have active J_{CC} couplings (~40-60 Hz) during t_1, which severely increase overlap of the resonances due to large multiple splittings of crosspeaks. These splittings can be removed by performing the HSQC with a constant-time interval (Figure A.3) (Santoro and King, 1992). In this experiment, carbon chemical shift evolution is detected by moving a 180° pulse in a constant time interval, rather than arraying the t_1 time. When the constant time interval is set to an integer value of k/J_{CC} (k=1,2,3,..), the couplings are refocused to a single resonance. An interesting effect of this experiment that can be utilized for resonance assignment is that the number of carbon-carbon bonds affects the sign of the refocusing. For instance, a carbon attached to one other carbon oscillates as a function of $\cos(\pi Jt)$ while a carbon with two other carbons attached oscillates as a function of $\cos(2\pi Jt)$. At a constant time interval $k=1/J_{CC}$ the refocusing terms for the different spin systems are $\cos(\pi)$ and $\cos(2\pi)$, resulting in crosspeak of negative and positive intensity respectively. In the ribose ring of RNA, a value of k=1 would give negative C1'/C5' and positive C2'/C3'/C4' resonances, which

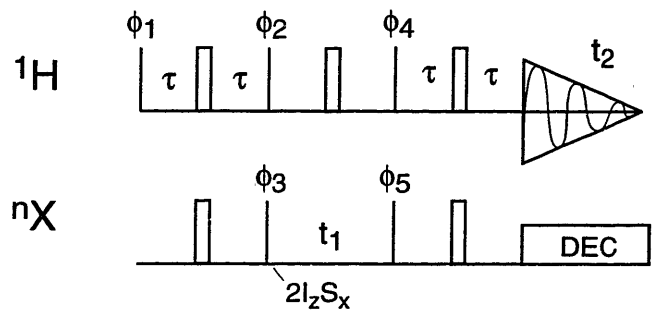
(A) HMQC



ϕ_1=2(x), 2(y)
ϕ_2=x, -x, y, -y
ϕ_3=2(x), 4(y), 2(-x)
ϕ_4=[2(x), 2(y)]$_2$, [2(-x), 2(-y)]$_2$
rec=x, -x, y, -y, 2(-x, x, -y, y), x, -x, y, -y

States-TPPI ϕ_2 - t_1

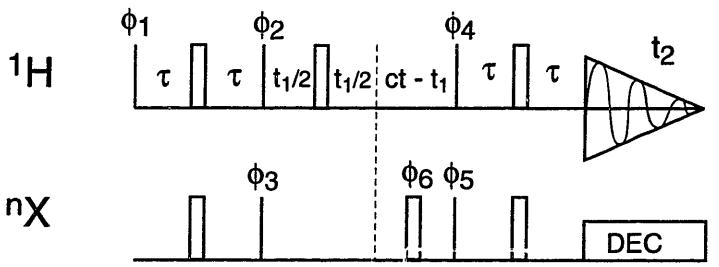
(B) HSQC



ϕ_1=x
ϕ_2=y, -y
ϕ_3=2(x), 2(-x)
ϕ_4=4(x), 4(-x)
ϕ_5=8(x), 8(-x)
rec=x, 2(-x), x, (-x, 2(x), -x)$_2$, x, 2(-x), x

States-TPPI ϕ_5 - t_1

(C) HSQC-CT



same as (B) plus:
ϕ_6=16(x), 16(-x)
States-TPPI ϕ_5 - t_1

ct = k/J_{CC}

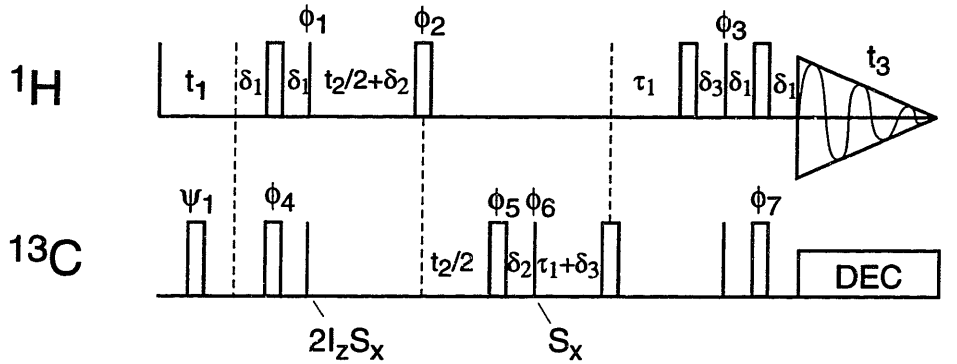
Figure A.3: Two Dimensional Heteronuclear NMR Experiments. **(A)** Heteronuclear Multiple Quantum Coherence (HMQC). Δ=1/2J. **(B)** Heteronuclear Single Quantum Coherence (HSQC). τ=1/4J. **(C)** Heternuclear Single Quantum Coherence - Constant Time (HSQC-CT). ct=k/J_{CC}, where k=1,2,3... Thin lines are 90° pulses, boxes are 180° pulses. States-TPPI indicates the method of quadrature detection for t_1 with the pulse that is phase cycled.

can be used to unambiguously identify C1' and C5' resonances since they are separated by ~30 ppm in carbon chemical shift.

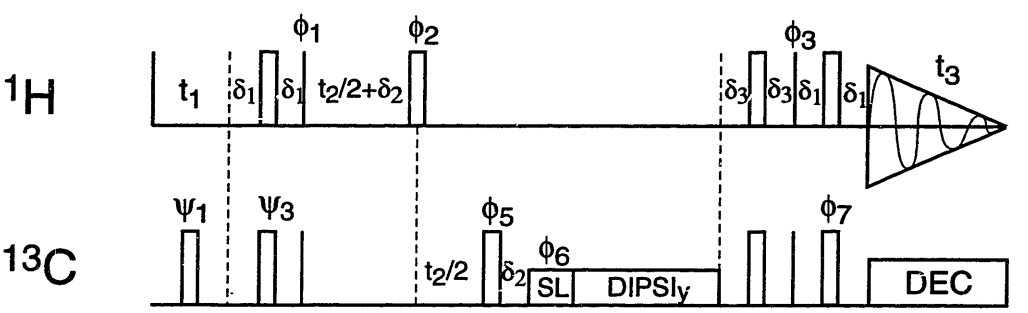
"HCCH" Experiments A class of experiments that efficiently correlates protons in a spin system through the large $^1J_{CH}$ (~160 Hz) and $^1J_{CC}$ (~50 Hz), rather than the small $^3J_{HH}$ (3-10 Hz) couplings, has been developed for through-bond assignment of isotopically labeled macromolecules (Kay et al., 1990). The "HC-" portion of the "HCCH" transfer is a refocused INEPT period that transfers magnetization to in-phase carbon magnetization (S_x after pulse $\phi 6$ in Figure A.4). Having in-phase (S_x) and not anti-phase ($2I_zS_x$) magnetization is essential for the "-CC-" transfer, which is carbon-carbon coupling delay in the HCCH-COSY experiment (transfers magnetization one bond), or a ^{13}C-spin lock in the HCCH-TOCSY experiment (transfers magnetization throughout the carbon spin system). The magnetization is then transferred back to proton ("-CH" transfer) with reverse-refocused INEPT period for detection. This mode of transfer through ^{13}C magnetization removes the conformational dependence of $^3J_{HH}$, such that all resonances can be observed regardless of the torsion angle. This is particularly important for RNA were A-form C3'-endo sugar puckers have unobservable H1'-H2' COSY crosspeaks ($^3J_{H1'-H2'}$ <3 Hz).

An interesting aspect of these experiments is that methylene (CH2) groups can be filtered from the experiment by proper choice of delays ($\delta 2 + \delta 3$) during the refocused INEPT periods. When the magnetization is anti-phase with respect to carbon (S_xI_z; point y in Figure A.4), the carbon J_{CH} coupling terms evolve differently depending on the number of protons attached ($Sx=\sin(2*2\pi J\delta_2)$ and $Sx=\sin(2*\pi J\delta_2)$ for CH2 and CH groups, respectively). Thus, a 1/4J delay for δ_2 would filter out CH2 groups ($\sin(\pi)=0$). A value of 1/8J would give optimal intensity for CH2 groups, but only partially filter CH groups. Typically an average value of 1/6J is used, which gives near optimal transfer to CH and

(A) HCCH-COSY



(B) HCCH-TOCSY



ϕ_1=y,-y
ϕ_2=4(x),4(-x),4(y), 4(-y)
ϕ_3=4(x),4(-x)
ϕ_4=2(y),2(-y)
ϕ_5=2(x),2(y),2(-x), 2(-y)
ϕ_6=4(x),4(-x)
ϕ_7=4(x),4(-x)
ψ_1=8(x),8(-x) > composite pulses
ψ_3=2(x),2(-x) > composite pulses
rec=x,-x,-x,x,-x,x,x,-x

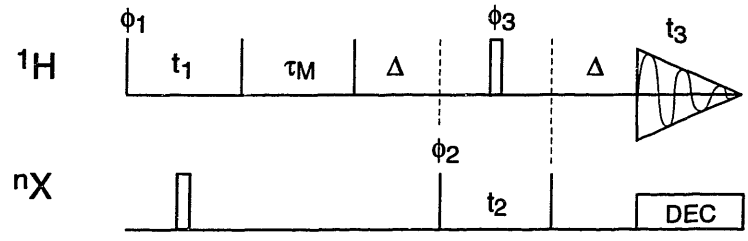
Figure A.4: "HCCH" coherence transfer experiments. δ1=1/4J, δ2/δ3=1/6J (see text). τ1=~1.6 ms. The phase cycle is the same for both experiments.

CH2 groups. Another strategy for filtering the HCCH transfer is to add or subtract a spectra taken with a 1/3J delay from the 1/6J experiment (Pardi and Nikonowicz, 1992).

Multidimensional NOESY experiments Perhaps the most simplistic, yet practically useful heteronuclear 3D experiment, is the NOESY-HMQC or NOESY-HQSC. These experiments are simple concatenations of the 2D NOESY and HMQC/HSQC experiments. They spread the 2D NOESY into a third dimension, separating crosspeaks between two protons by the chemical shift of the carbon attached to one of the protons. This is essential to resolving spectral overlap problems in larger macromolecules in order to obtain enough distance restrains for high-quality three-dimensional structures. The differences between the NOESY-HSQC and NOESY-HMQC experiments are the essentially the same as for the 2D HMQC and HSQC experiments (see above). Again, because of the suitably of adding gradients, NOESY-HSQC is the preferred experiment. The gradient NOESY-HSQC in Figure A.5B is designed for reducing the phase cycle and removing artifacts for ^{13}C-labeled samples in D_2O. For experiments with ^{15}N or ^{13}C in H_2O many modified versions of the basic NOESY-HSQC experiment have been developed for efficient water suppression using many of the elements described in Figure A.2 (Majumdar and Zuiderweg, 1993; Muhandiram et al., 1993; Sklenar et al., 1993b; Jahnke and Kessler, 1994; Jahnke et al., 1995).

Spectral overlap can be further resolved through 4D experiments, which contain the chemical shift of both carbons attached to the two protons involved in the NOE (Figure A.5C). This is important for obtaining ribose-ribose NOEs in RNA, since their chemical shifts are better resolved in carbon than in proton. Again, gradients can be utilized to shorten the phase cycle (only two scans in this case), which allows the accumulation of many more points in each dimension (increasing resolution) (Vuister et al., 1993).

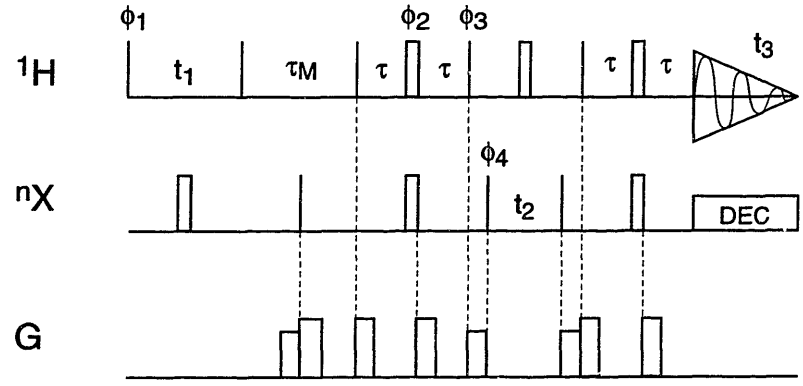
(A) 3D NOESY-HMQC



ϕ1=x, -x
ϕ2=(x)2, (-x)2
ϕ3=(x)4, (y)4
rec=x, (-x)2, x, -x, (x)2, -x

States-TPPI ϕ1 - t1, ϕ2 - t2

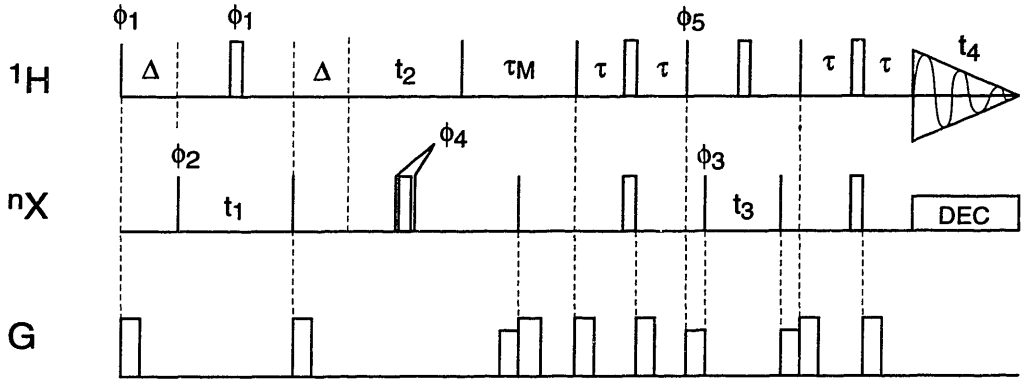
(B) 3D NOESY-HSQC (with gradients)



ϕ1=y, -y
ϕ2=y
ϕ3=-x
ϕ4=y,y,-y,-y
rec=y,-y,-y,y

States-TPPI ϕ1 - t1, ϕ4 - t2

(C) 4D HMQC-NOESY-HSQC (with gradients)



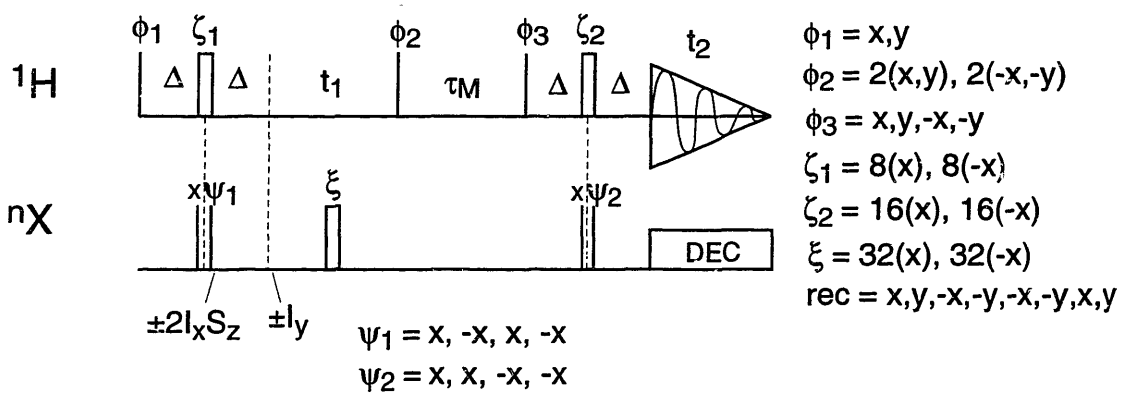
ϕ1=x
ϕ2=x, -x
ϕ3=x
ϕ4=y
ϕ5=y
rec=x, -x

States-TPPI ϕ2 - t1, ϕ1 - t2, ϕ3 - t3

Figure A.5: Multidimensional NOESY experiments. All pulse are x unless otherwise indicated.

Filtered NOESY Experiments The basic idea of a 1D filter element is to perform an HMQC experiment with no carbon evolution time. This results in consecutive 90° heteronuclear (e.g. ^{13}C) pulses with one phase cycled (ψ=x,-x) and the data stored separately. One phase (x) gives a net 180° pulse, while the other (-x) gives a net 0° pulse. The magnetization of protons attached to ^{13}C at this stage is $2I_xS_z$; therefore the two phases (180° and 0°) would give $-2I_xS_z$ and $+2I_xS_z$ terms, respectively. The second 1/2J delay (Δ) would refocus these two terms to $-I_y$ and $+I_y$, respectively, for detection. Protons not attached to ^{13}C, however, would not be effected by the phase of the ^{13}C-editing pulse (ψ), and would have the same sign in both data sets. If the two data sets are added together, the magnetization for ^{13}C-attached protons would cancel, whereas non-^{13}C-attached protons would be additive. Conversely, the opposite is true if the data set are subtracted. The double-half-filter NOESY experiment (Figure A.6A) incorporates two filters, one for each dimension (t_1 and t_2). The four different phase combinations of the editing pulses ψ_1 and ψ_2 are obtained and stored separately. They can be added or subtracted in four combinations to produce four "sub"-spectra filtering for or against observing protons attached to ^{13}C in each dimension (see (Otting and Wuthrich, 1990) for more details on the various combinations). The "HMQC-derived" filter element was the first described, however, several different types of filter and "purge" elements have been developed (see (Wand and Short, 1994) for a review). There are several practical limits to obtaining "perfect" filtering of ^{13}C-attached resonances. One is that the $^1J_{CH}$ values for different resonances vary widely in biological macromolecules (~140 Hz for aliphatic vs. ~180 for aromatic). Therefore, no matter what choice of Δ, some small amount of magnetization for ^{13}C-attached protons will still be I_y and unaffected by the editing pulse. In addition, it is difficult to obtain uniform excitement with the 180° pulse over the wide ^{13}C-chemical shift range. The artifacts that arise from these problems can be mitigated with the use of gradients (Figure A.6B). These gradients are conceptual similar to those used for a 4D

(A)



(B)

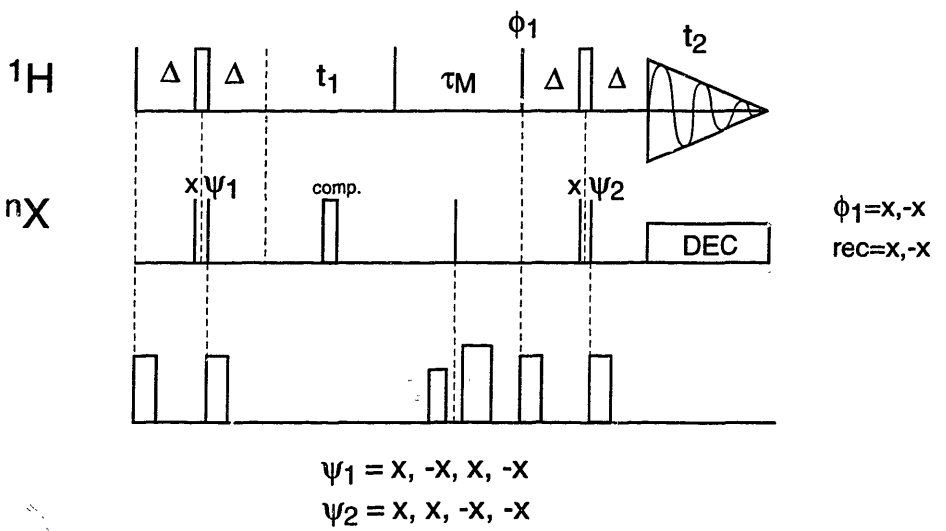


Figure A.6: Double-Half-Filtered NOESY experiments **(A)** with and **(B)** without gradients. Δ=1/4J; comp. represents a composite x-2y-x 180° pulse.

HMQC-NOESY-HMQC experiment (Vuister et al., 1993), and produce higher quality spectra in few scans.

APPENDIX B: MOLECULAR MODELING DETAILS

This appendix is merely a reference for details of some of the files used for molecular modeling calculations using the DISCOVER module of the program Insight II. Table B.1 is a list of all NMR-derived restraints used for the structure determination. The nomenclature for designation of atoms is "molecule:residue:atomname". For example "1:TRP_63:HN" is the backbone amide proton of tryptophan 63 in molecule 1 (the complex was modeled as a "single" molecule for simplicity in nomenclature and file structure). Note that the numbering of the RNA and peptide is different than that used in the text for historical and practical reasons. The two can be easily translated. The nomenclature of the atom names is similar to standard PDB format and "*" represents pseudoatoms for non-stereospecifically assigned methylene or methyl protons

Tables B.2 and B.3 show all parameters for the two simulated annealing protocols used for structure determination of the complex. A generic description of the protocols is given in Chapter II (Materials and Methods). The Insight II manual should be consulted for more specific information on any variable.

Table B.1: Molecular modeling restraint file (InsightII *.rstrnt file). The columns for distance restraints are as follows: (1) atom #1, (2) atom #2, (3) lower distance bound (Å), (4) upper distance bound (Å), (5) force constant for lower bound (kcal/mol-Å^2), (6) force constant for lower bound (kcal/mol-Å^2), (7) violation energy cutoff (kcal/mol). See Insight manual for more details. The columns for torsion restraints are identical, except two extra columns are added to specify the four atoms defining the torsion angle. Distance restraints are divided in vario classes of restraints for clarity of reading. The different sections are headed by comments "!". **Note that the numbering of the RNA and peptide in the restraint file is different than in the main body of the text.** The RNA is numbered 1-34, and the peptide 51-74.

```
!BIOSYM restraint 1
!
#distance
!peptide helix hydrogen bonds
1:ARG_73:HN     1:ALA_69:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_73:N      1:ALA_69:O    2.40  3.50  10.00  10.00  1000.00
1:ALA_72:HN     1:ARG_68:O    1.50  2.80  10.00  10.00  1000.00
1:ALA_72:N      1:ARG_68:O    2.40  3.50  10.00  10.00  1000.00
1:ALA_71:HN     1:GLN_67:O    1.50  2.80  10.00  10.00  1000.00
1:ALA_71:N      1:GLN_67:O    2.40  3.50  10.00  10.00  1000.00
1:ALA_70:HN     1:ARG_66:O    1.50  2.80  10.00  10.00  1000.00
1:ALA_70:N      1:ARG_66:O    2.40  3.50  10.00  10.00  1000.00
1:ALA_69:HN     1:GLU_65:O    1.50  2.80  10.00  10.00  1000.00
1:ALA_69:N      1:GLU_65:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_68:HN     1:ARG_64:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_68:N      1:ARG_64:O    2.40  3.50  10.00  10.00  1000.00
1:GLN_67:HN     1:TRP_63:O    1.50  2.80  10.00  10.00  1000.00
1:GLN_67:N      1:TRP_63:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_66:HN     1:ARG_62:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_66:N      1:ARG_62:O    2.40  3.50  10.00  10.00  1000.00
1:GLU_65:HN     1:ARG_61:O    1.50  2.80  10.00  10.00  1000.00
1:GLU_65:N      1:ARG_61:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_64:HN     1:ARG_60:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_64:N      1:ARG_60:O    2.40  3.50  10.00  10.00  1000.00
1:TRP_63:HN     1:ARG_59:O    1.50  2.80  10.00  10.00  1000.00
1:TRP_63:N      1:ARG_59:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_62:HN     1:ASN_58:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_62:N      1:ASN_58:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_61:HN     1:ARG_57:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_61:N      1:ARG_57:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_60:HN     1:ARG_56:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_60:N      1:ARG_56:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_59:HN     1:ALA_55:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_59:N      1:ALA_55:O    2.40  3.50  10.00  10.00  1000.00
1:ASN_58:HN     1:GLN_54:O    1.50  2.80  10.00  10.00  1000.00
1:ASN_58:N      1:GLN_54:O    2.40  3.50  10.00  10.00  1000.00
1:ARG_57:HN     1:ARG_53:O    1.50  2.80  10.00  10.00  1000.00
1:ARG_57:N      1:ARG_53:O    2.40  3.50  10.00  10.00  1000.00

!hbonding restraints - non Watson Crick
!G-G pair
1:G_8:H1        1:G_26:O6     1.70  2.20  10.00  10.00  1000.00
1:G_8:N1        1:G_26:O6     2.70  3.20  10.00  10.00  1000.00
1:G_26:H1       1:G_8:O6      1.70  2.20  10.00  10.00  1000.00
1:G_26:N1       1:G_8:O6      2.70  3.20  10.00  10.00  1000.00
!tetraloop GA pair
1:G_15:H21      1:A_18:N7     1.70  2.20  10.00  10.00  1000.00
1:G_15:N2       1:A_18:N7     2.70  3.20  10.00  10.00  1000.00
1:G_15:N3       1:A_18:H61    1.70  2.20  10.00  10.00  1000.00
1:G_15:N3       1:A_18:N6     2.70  3.20  10.00  10.00  1000.00
!tetraloop hbonds
1:G_15:HO2'     1:A_17:N7     1.70  2.20  10.00  10.00  1000.00
1:G_15:O2'      1:A_17:N7     2.70  3.20  10.00  10.00  1000.00
1:C_16:HO2'     1:A_17:O4'    1.70  2.20  10.00  10.00  1000.00
1:C_16:O2'      1:A_17:O4'    2.70  3.20  10.00  10.00  1000.00
1:G_15:H22      1:A_18:O2P    1.70  2.20  10.00  10.00  1000.00
1:G_15:N2       1:A_18:O2P    2.70  3.20  10.00  10.00  1000.00
!anti-anti GA pair
1:G_7:H1        1:A_28:N1     1.70  2.20  10.00  10.00  1000.00
1:G_7:N1        1:A_28:N1     2.70  3.20  10.00  10.00  1000.00
```

1:G_7:O6	1:A_28:H61	1.70	2.20	10.00	10.00	1000.00
1:G_7:O6	1:A_28:N6	2.70	3.20	10.00	10.00	1000.00
!G-U pair						
1:U_3:H3	1:G_32:O6	1.70	2.20	10.00	10.00	1000.00
1:U_3:N3	1:G_32:O6	2.70	3.20	10.00	10.00	1000.00
1:U_3:O2	1:G_32:H1	1.70	2.20	10.00	10.00	1000.00
1:U_3:O2	1:G_32:N1	2.70	3.20	10.00	10.00	1000.00
!hbonding restraints - Watson Crick						
1:G_1:O6	1:C_34:H42	1.70	2.20	10.00	10.00	1000.00
1:G_1:O6	1:C_34:N4	2.70	3.20	10.00	10.00	1000.00
1:G_1:H1	1:C_34:N3	1.70	2.20	10.00	10.00	1000.00
1:G_1:N1	1:C_34:N3	2.70	3.20	10.00	10.00	1000.00
1:G_1:H22	1:C_34:O2	1.70	2.20	10.00	10.00	1000.00
1:G_1:N2	1:C_34:O2	2.70	3.20	10.00	10.00	1000.00
1:G_2:O6	1:C_33:H42	1.70	2.20	10.00	10.00	1000.00
1:G_2:O6	1:C_33:N4	2.70	3.20	10.00	10.00	1000.00
1:G_2:H1	1:C_33:N3	1.70	2.20	10.00	10.00	1000.00
1:G_2:N1	1:C_33:N3	2.70	3.20	10.00	10.00	1000.00
1:G_2:H22	1:C_33:O2	1.70	2.20	10.00	10.00	1000.00
1:G_2:N2	1:C_33:O2	2.70	3.20	10.00	10.00	1000.00
1:C_4:H42	1:G_31:O6	1.70	2.20	10.00	10.00	1000.00
1:C_4:N4	1:G_31:O6	2.70	3.20	10.00	10.00	1000.00
1:C_4:N3	1:G_31:H1	1.70	2.20	10.00	10.00	1000.00
1:C_4:N3	1:G_31:N1	2.70	3.20	10.00	10.00	1000.00
1:C_4:O2	1:G_31:H22	1.70	2.20	10.00	10.00	1000.00
1:C_4:O2	1:G_31:N2	2.70	3.20	10.00	10.00	1000.00
1:U_5:H3	1:A_30:N1	1.70	2.20	10.00	10.00	1000.00
1:U_5:N3	1:A_30:N1	2.70	3.20	10.00	10.00	1000.00
1:U_5:O4	1:A_30:H61	1.70	2.20	10.00	10.00	1000.00
1:U_5:O4	1:A_30:N6	2.70	3.20	10.00	10.00	1000.00
1:U_5:O2	1:A_30:H2	2.95	3.45	10.00	10.00	1000.00
1:G_6:O6	1:C_29:H42	1.70	2.20	10.00	10.00	1000.00
1:G_6:O6	1:C_29:N4	2.70	3.20	10.00	10.00	1000.00
1:G_6:H1	1:C_29:N3	1.70	2.20	10.00	10.00	1000.00
1:G_6:N1	1:C_29:N3	2.70	3.20	10.00	10.00	1000.00
1:G_6:H22	1:C_29:O2	1.70	2.20	10.00	10.00	1000.00
1:G_6:N2	1:C_29:O2	2.70	3.20	10.00	10.00	1000.00
1:C_9:H42	1:G_25:O6	1.70	2.20	10.00	10.00	1000.00
1:C_9:N4	1:G_25:O6	2.70	3.20	10.00	10.00	1000.00
1:C_9:N3	1:G_25:H1	1.70	2.20	10.00	10.00	1000.00
1:C_9:N3	1:G_25:N1	2.70	3.20	10.00	10.00	1000.00
1:C_9:O2	1:G_25:H22	1.70	2.20	10.00	10.00	1000.00
1:C_9:O2	1:G_25:N2	2.70	3.20	10.00	10.00	1000.00
1:G_10:O6	1:C_24:H42	1.70	2.20	10.00	10.00	1000.00
1:G_10:O6	1:C_24:N4	2.70	3.20	10.00	10.00	1000.00
1:G_10:H1	1:C_24:N3	1.70	2.20	10.00	10.00	1000.00
1:G_10:N1	1:C_24:N3	2.70	3.20	10.00	10.00	1000.00
1:G_10:H22	1:C_24:O2	1.70	2.20	10.00	10.00	1000.00
1:G_10:N2	1:C_24:O2	2.70	3.20	10.00	10.00	1000.00
1:C_11:H42	1:G_22:O6	1.70	2.20	10.00	10.00	1000.00
1:C_11:N4	1:G_22:O6	2.70	3.20	10.00	10.00	1000.00
1:C_11:N3	1:G_22:H1	1.70	2.20	10.00	10.00	1000.00
1:C_11:N3	1:G_22:N1	2.70	3.20	10.00	10.00	1000.00
1:C_11:O2	1:G_22:H22	1.70	2.20	10.00	10.00	1000.00
1:C_11:O2	1:G_22:N2	2.70	3.20	10.00	10.00	1000.00
1:A_12:N1	1:U_21:H3	1.70	2.20	10.00	10.00	1000.00
1:A_12:N1	1:U_21:N3	2.70	3.20	10.00	10.00	1000.00
1:A_12:H61	1:U_21:O4	1.70	2.20	10.00	10.00	1000.00
1:A_12:N6	1:U_21:O4	2.70	3.20	10.00	10.00	1000.00
1:A_12:H2	1:U_21:O2	2.95	3.45	10.00	10.00	1000.00
1:G_13:O6	1:C_20:H42	1.70	2.20	10.00	10.00	1000.00
1:G_13:O6	1:C_20:N4	2.70	3.20	10.00	10.00	1000.00
1:G_13:H1	1:C_20:N3	1.70	2.20	10.00	10.00	1000.00
1:G_13:N1	1:C_20:N3	2.70	3.20	10.00	10.00	1000.00
1:G_13:H22	1:C_20:O2	1.70	2.20	10.00	10.00	1000.00
1:G_13:N2	1:C_20:O2	2.70	3.20	10.00	10.00	1000.00
1:C_14:H42	1:G_19:O6	1.70	2.20	10.00	10.00	1000.00
1:C_14:N4	1:G_19:O6	2.70	3.20	10.00	10.00	1000.00
1:C_14:N3	1:G_19:H1	1.70	2.20	10.00	10.00	1000.00
1:C_14:N3	1:G_19:N1	2.70	3.20	10.00	10.00	1000.00
1:C_14:O2	1:G_19:H22	1.70	2.20	10.00	10.00	1000.00
1:C_14:O2	1:G_19:N2	2.70	3.20	10.00	10.00	1000.00
!exchangeable RNA-RNA NOE restraints						
!intrabasepair						
1:U_3:H3	1:G_32:H1	1.80	3.00	10.00	10.00	1000.00
1:G_8:H1	1:G_26:H1	1.80	3.00	10.00	10.00	1000.00

!interbasepair						
1:G_1:H1	1:G_2:H1	1.80	5.00	10.00	10.00	1000.00
1:G_2:H1	1:U_3:H3	1.80	5.00	10.00	10.00	1000.00
1:G_2:H1	1:G_32:H1	1.80	5.00	10.00	10.00	1000.00
1:U_3:H3	1:G_31:H1	1.80	5.00	10.00	10.00	1000.00
1:G_32:H1	1:G_31:H1	1.80	5.00	10.00	10.00	1000.00
1:G_31:H1	1:U_5:H3	1.80	5.00	10.00	10.00	1000.00
1:U_5:H3	1:G_6:H1	1.80	5.00	10.00	10.00	1000.00
1:G_8:H1	1:A_28:H2	1.80	5.00	10.00	10.00	1000.00
1:G_8:H1	1:G_25:H1	1.80	5.00	10.00	10.00	1000.00
1:G_26:H1	1:G_25:H1	1.80	5.00	10.00	10.00	1000.00
1:G_25:H1	1:G_10:H1	1.80	5.00	10.00	10.00	1000.00
1:G_10:H1	1:G_22:H1	1.80	5.00	10.00	10.00	1000.00
1:G_22:H1	1:U_21:H3	1.80	5.00	10.00	10.00	1000.00
1:U_21:H3	1:G_13:H1	1.80	5.00	10.00	10.00	1000.00
1:G_13:H1	1:G_19:H1	1.80	5.00	10.00	10.00	1000.00
!imino-amino						
!intrabasepair						
1:G_7:H1	1:A_28:H2	1.80	3.00	10.00	10.00	1000.00
1:G_1:H1	1:C_34:H42	1.80	3.00	10.00	10.00	1000.00
1:G_2:H1	1:C_33:H42	1.80	3.00	10.00	10.00	1000.00
1:G_31:H1	1:C_4:H42	1.80	3.00	10.00	10.00	1000.00
1:U_5:H3	1:A_30:H2	1.80	3.00	10.00	10.00	1000.00
1:G_6:H1	1:C_29:H42	1.80	3.00	10.00	10.00	1000.00
1:G_25:H1	1:C_9:H42	1.80	3.00	10.00	10.00	1000.00
1:G_10:H1	1:C_24:H42	1.80	3.00	10.00	10.00	1000.00
1:G_22:H1	1:C_11:H42	1.80	3.00	10.00	10.00	1000.00
1:U_21:H3	1:A_12:H2	1.80	3.00	10.00	10.00	1000.00
1:G_13:H1	1:C_20:H42	1.80	3.00	10.00	10.00	1000.00
1:G_19:H1	1:C_14:H42	1.80	3.00	10.00	10.00	1000.00
!interbasepair						
1:G_2:H1	1:G_32:2X	1.80	6.00	10.00	10.00	1000.00
1:U_3:H3	1:C_33:H42	1.80	5.00	10.00	10.00	1000.00
1:U_3:H3	1:C_4:H42	1.80	4.00	10.00	10.00	1000.00
1:U_3:H3	1:C_4:H41	1.80	5.00	10.00	10.00	1000.00
1:G_6:H1	1:G_7:2X	1.80	5.00	10.00	10.00	1000.00
1:G_10:H1	1:C_11:H42	1.80	5.00	10.00	10.00	1000.00
1:G_13:H1	1:C_14:H42	1.80	5.00	10.00	10.00	1000.00
1:U_21:H3	1:C_20:H41	1.80	5.00	10.00	10.00	1000.00
1:U_21:H3	1:C_20:H42	1.80	5.00	10.00	10.00	1000.00
1:U_21:H3	1:C_11:H42	1.80	5.00	10.00	10.00	1000.00
1:G_19:H1	1:C_20:H42	1.80	4.00	10.00	10.00	1000.00
1:G_22:H1	1:C_24:H42	1.80	5.00	10.00	10.00	1000.00
1:G_25:H1	1:G_8:2X	1.80	6.00	10.00	10.00	1000.00
1:G_26:H1	1:C_9:H42	1.80	5.00	10.00	10.00	1000.00
1:G_32:H1	1:C_4:H42	1.80	5.00	10.00	10.00	1000.00
!amino/imino-nonexchangable						
1:U_3:H3	1:C_4:H5	1.80	5.00	10.00	10.00	1000.00
!1:G_19:H1	1:C_20:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_32:2X	1:C_4:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_32:2X	1:C_33:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_8:2X	1:C_9:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_8:2X	1:A_28:H2	1.80	5.00	10.00	10.00	1000.00
1:C_20:H42	1:C_14:H42	1.80	4.00	10.00	10.00	1000.00
1:C_20:H42	1:C_14:H41	1.80	4.00	10.00	10.00	1000.00
!non-exchangable NOE restraints						
!intranucleotide NOEs						
1:G_1:H8	1:G_1:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_1:H8	1:G_1:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_1:H8	1:G_1:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_1:H8	1:G_1:H4'	1.80	5.00	10.00	10.00	1000.00
1:G_1:H8	1:G_1:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_1:H1'	1:G_1:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_1:H1'	1:G_1:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_1:H1'	1:G_1:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_2:H8	1:G_2:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_2:H8	1:G_2:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_2:H1'	1:G_2:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_2:H1'	1:G_2:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_2:H1'	1:G_2:H4'	1.80	4.00	10.00	10.00	1000.00
1:U_3:H6	1:U_3:H1'	1.80	4.00	10.00	10.00	1000.00
1:U_3:H6	1:U_3:H5	1.80	3.00	10.00	10.00	1000.00
1:U_3:H1'	1:U_3:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_3:H1'	1:U_3:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_4:H6	1:C_4:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_4:H6	1:C_4:H2'	1.80	5.00	10.00	10.00	1000.00
1:C_4:H6	1:C_4:H3'	1.80	4.00	10.00	10.00	1000.00

1:C_4:H6	1:C_4:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_4:H6	1:C_4:H5	1.80	3.00	10.00	10.00	1000.00
1:C_4:H1'	1:C_4:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_4:H1'	1:C_4:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_4:H1'	1:C_4:H4'	1.80	4.00	10.00	10.00	1000.00
1:U_5:H6	1:U_5:H1'	1.80	4.00	10.00	10.00	1000.00
1:U_5:H6	1:U_5:H5	1.80	3.00	10.00	10.00	1000.00
1:U_5:H1'	1:U_5:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_5:H1'	1:U_5:H3'	1.80	5.00	10.00	10.00	1000.00
1:U_5:H1'	1:U_5:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_6:H8	1:G_6:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_6:H8	1:G_6:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_6:H8	1:G_6:H3'	1.80	5.00	10.00	10.00	1000.00
1:G_6:H1'	1:G_6:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_6:H1'	1:G_6:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H8	1:G_7:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H8	1:G_7:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H8	1:G_7:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H1'	1:G_7:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_7:H1'	1:G_7:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H2'	1:G_7:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_7:H3'	1:G_7:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_8:H8	1:G_8:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_8:H8	1:G_8:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_8:H1'	1:G_8:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_8:H1'	1:G_8:H3'	1.80	5.00	10.00	10.00	1000.00
1:G_8:H1'	1:G_8:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_8:H2'	1:G_8:H3'	1.80	3.00	10.00	10.00	1000.00
1:C_9:H6	1:C_9:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H6	1:C_9:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H6	1:C_9:H5	1.80	3.00	10.00	10.00	1000.00
1:C_9:H5	1:C_9:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_9:H1'	1:C_9:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_9:H1'	1:C_9:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H1'	1:C_9:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_10:H8	1:G_10:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_10:H8	1:G_10:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_10:H8	1:G_10:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_10:H1'	1:G_10:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_11:H6	1:C_11:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_11:H6	1:C_11:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_11:H6	1:C_11:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_11:H6	1:C_11:H5	1.80	3.00	10.00	10.00	1000.00
1:C_11:H1'	1:C_11:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_11:H1'	1:C_11:H5'*	1.80	6.00	10.00	10.00	1000.00
1:A_12:H8	1:A_12:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_12:H8	1:A_12:H2'	1.80	5.00	10.00	10.00	1000.00
1:A_12:H8	1:A_12:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_12:H1'	1:A_12:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_13:H8	1:G_13:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_13:H8	1:G_13:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_13:H8	1:G_13:H4'	1.80	5.00	10.00	10.00	1000.00
1:G_13:H8	1:G_13:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_13:H1'	1:G_13:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_13:H1'	1:G_13:H4'	1.80	5.00	10.00	10.00	1000.00
1:C_14:H6	1:C_14:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_14:H6	1:C_14:H2'	1.80	5.00	10.00	10.00	1000.00
1:C_14:H6	1:C_14:H3'	1.80	3.00	10.00	10.00	1000.00
1:C_14:H6	1:C_14:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_14:H6	1:C_14:H5	1.80	3.00	10.00	10.00	1000.00
1:C_14:H1'	1:C_14:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_14:H1'	1:C_14:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_14:H2'	1:C_14:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_15:H8	1:G_15:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_15:H8	1:G_15:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_15:H8	1:G_15:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_15:H1'	1:G_15:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_16:H6	1:C_16:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_16:H6	1:C_16:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_16:H6	1:C_16:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_16:H6	1:C_16:H4'	1.80	5.00	10.00	10.00	1000.00
1:C_16:H6	1:C_16:H5	1.80	3.00	10.00	10.00	1000.00
1:C_16:H1'	1:C_16:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_16:H1'	1:C_16:H4'	1.80	4.00	10.00	10.00	1000.00
1:C_16:H2'	1:C_16:H4'	1.80	4.00	10.00	10.00	1000.00
1:C_16:H2'	1:C_16:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H8	1:A_17:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_17:H8	1:A_17:H2'	1.80	4.00	10.00	10.00	1000.00

1:A_17:H8	1:A_17:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_17:H1'	1:A_17:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H2'	1:A_17:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H3'	1:A_17:H4'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H1'	1:A_17:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_17:H1'	1:A_17:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H8	1:A_18:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H8	1:A_18:H2'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H8	1:A_18:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H8	1:A_18:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_18:H1'	1:A_18:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_18:H1'	1:A_18:H3'	1.80	5.00	10.00	10.00	1000.00
1:A_18:H1'	1:A_18:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H2'	1:A_18:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H3'	1:A_18:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_19:H8	1:G_19:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_19:H8	1:G_19:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_19:H8	1:G_19:H4'	1.80	5.00	10.00	10.00	1000.00
1:G_19:H8	1:G_19:H5'*	1.80	5.00	10.00	10.00	1000.00
1:G_19:H1'	1:G_19:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_19:H1'	1:G_19:H3'	1.80	5.00	10.00	10.00	1000.00
1:G_19:H1'	1:G_19:H4'	1.80	5.00	10.00	10.00	1000.00
1:C_20:H6	1:C_20:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_20:H6	1:C_20:H5	1.80	3.00	10.00	10.00	1000.00
1:C_20:H1'	1:C_20:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_21:H6	1:U_21:H1'	1.80	5.00	10.00	10.00	1000.00
1:U_21:H6	1:U_21:H5	1.80	3.00	10.00	10.00	1000.00
1:U_21:H1'	1:U_21:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_22:H8	1:G_22:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_22:H8	1:G_22:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_22:H8	1:G_22:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_22:H1'	1:G_22:H2'	2.00	3.50	10.00	10.00	1000.00
1:G_22:H1'	1:G_22:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_22:H2'	1:G_22:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_23:H8	1:A_23:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_23:H8	1:A_23:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_23:H8	1:A_23:H3'	1.80	5.00	10.00	10.00	1000.00
1:A_23:H8	1:A_23:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_23:H1'	1:A_23:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_23:H1'	1:A_23:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_23:H1'	1:A_23:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_23:H2'	1:A_23:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_23:H2'	1:A_23:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_23:H3'	1:A_23:H4'	1.80	3.00	10.00	10.00	1000.00
1:C_24:H6	1:C_24:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_24:H6	1:C_24:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_24:H6	1:C_24:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_24:H6	1:C_24:H5	1.80	3.00	10.00	10.00	1000.00
1:C_24:H5	1:C_24:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_24:H1'	1:C_24:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_24:H1'	1:C_24:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_25:H8	1:G_25:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_25:H8	1:G_25:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_25:H8	1:G_25:H3'	1.80	3.00	10.00	10.00	1000.00
1:G_25:H1'	1:G_25:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_26:H8	1:G_26:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_26:H8	1:G_26:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_26:H1'	1:G_26:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_27:H6	1:U_27:H1'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H6	1:U_27:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_27:H6	1:U_27:H3'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H6	1:U_27:H4'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H6	1:U_27:H5	1.80	3.00	10.00	10.00	1000.00
1:U_27:H5	1:U_27:H2'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H1'	1:U_27:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_27:H1'	1:U_27:H3'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H1'	1:U_27:H4'	1.80	4.00	10.00	10.00	1000.00
1:U_27:H2'	1:U_27:H3'	1.80	3.00	10.00	10.00	1000.00
1:U_27:H2'	1:U_27:H4'	1.80	5.00	10.00	10.00	1000.00
1:U_27:H3'	1:U_27:H4'	1.80	3.00	10.00	10.00	1000.00
1:A_28:H8	1:A_28:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_28:H1'	1:A_28:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_28:H1'	1:A_28:H3'	1.80	5.00	10.00	10.00	1000.00
1:A_28:H1'	1:A_28:H4'	1.80	4.00	10.00	10.00	1000.00
1:C_29:H6	1:C_29:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_29:H6	1:C_29:H2'	1.80	5.00	10.00	10.00	1000.00
1:C_29:H6	1:C_29:H5	1.80	3.00	10.00	10.00	1000.00
1:C_29:H1'	1:C_29:H2'	1.80	3.00	10.00	10.00	1000.00

1:C_29:H1'	1:C_29:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_29:H1'	1:C_29:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_30:H8	1:A_30:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_30:H8	1:A_30:H2'	1.80	5.00	10.00	10.00	1000.00
1:A_30:H8	1:A_30:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_30:H1'	1:A_30:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_30:H1'	1:A_30:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_31:H8	1:G_31:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_31:H8	1:G_31:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_31:H8	1:G_31:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_31:H8	1:G_31:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_31:H1'	1:G_31:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_31:H1'	1:G_31:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_32:H8	1:G_32:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_32:H8	1:G_32:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_32:H8	1:G_32:H4'	1.80	5.00	10.00	10.00	1000.00
1:G_32:H8	1:G_32:H5'*	1.80	6.00	10.00	10.00	1000.00
1:G_32:H1'	1:G_32:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_32:H1'	1:G_32:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_33:H6	1:C_33:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_33:H6	1:C_33:H3'	1.80	3.00	10.00	10.00	1000.00
1:C_33:H6	1:C_33:H5	1.80	3.00	10.00	10.00	1000.00
1:C_33:H1'	1:C_33:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_34:H6	1:C_34:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_34:H6	1:C_34:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_34:H6	1:C_34:H3'	1.80	3.00	10.00	10.00	1000.00
1:C_34:H6	1:C_34:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_34:H6	1:C_34:H5	1.80	3.00	10.00	10.00	1000.00
1:C_34:H5	1:C_34:H5'*	1.80	6.00	10.00	10.00	1000.00
1:C_34:H1'	1:C_34:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_34:H1'	1:C_34:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_34:H2'	1:C_34:H3'	1.80	3.00	10.00	10.00	1000.00
1:C_34:H3'	1:C_34:H4'	1.80	3.00	10.00	10.00	1000.00
1:C_34:H3'	1:C_34:H5'*	1.80	5.00	10.00	10.00	1000.00
!internucleotide NOEs						
1:G_1:H8	1:G_2:H8	1.80	5.00	10.00	10.00	1000.00
1:G_1:H2'	1:G_2:H5'*	1.80	5.00	10.00	10.00	1000.00
1:G_2:H8	1:G_1:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_2:H8	1:U_3:H5	1.80	5.00	10.00	10.00	1000.00
1:G_2:H8	1:G_1:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_2:H8	1:G_1:H3'	1.80	4.00	10.00	10.00	1000.00
1:U_3:H6	1:G_2:H1'	1.80	5.00	10.00	10.00	1000.00
1:U_3:H6	1:G_2:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_3:H5	1:G_2:H2'	1.80	5.00	10.00	10.00	1000.00
1:U_3:H5	1:G_2:H3'	1.80	5.00	10.00	10.00	1000.00
1:U_3:H6	1:C_4:H6	1.80	5.00	10.00	10.00	1000.00
1:U_3:H5	1:C_4:H5	1.80	5.00	10.00	10.00	1000.00
1:C_4:H6	1:U_3:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_4:H6	1:U_3:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_4:H6	1:U_3:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_4:H5	1:U_3:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_4:H5	1:U_3:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_4:H5	1:U_5:H5	1.80	5.00	10.00	10.00	1000.00
1:U_5:H6	1:C_4:H1'	1.80	5.00	10.00	10.00	1000.00
1:U_5:H6	1:C_4:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_5:H6	1:C_4:H3'	1.80	4.00	10.00	10.00	1000.00
1:U_5:H5	1:C_4:H6	1.80	5.00	10.00	10.00	1000.00
1:G_6:H8	1:U_5:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_6:H8	1:U_5:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_6:H8	1:U_5:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H8	1:G_6:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_7:H8	1:G_8:H8	1.80	5.00	10.00	10.00	1000.00
1:G_8:H8	1:G_7:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_8:H8	1:G_7:H2'	1.80	5.00	10.00	10.00	1000.00
1:G_8:H8	1:G_7:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H6	1:G_8:H1'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H6	1:G_8:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_9:H2'	1:G_10:H3'	1.80	5.00	10.00	10.00	1000.00
1:G_10:H8	1:C_9:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_10:H8	1:C_9:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_10:H8	1:C_9:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_10:H8	1:C_11:H5	1.80	5.00	10.00	10.00	1000.00
1:C_11:H6	1:G_10:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_11:H6	1:G_10:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_11:H5	1:G_10:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_11:H6	1:A_12:H8	1.80	5.00	10.00	10.00	1000.00
1:A_12:H8	1:C_11:H1'	1.80	5.00	10.00	10.00	1000.00

1:A_12:H8	1:C_11:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_12:H8	1:C_11:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_12:H8	1:G_13:H8	1.80	5.00	10.00	10.00	1000.00
1:A_12:H2	1:G_13:H1'	1.80	4.00	10.00	10.00	1000.00
1:G_13:H8	1:A_12:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_13:H8	1:A_12:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_13:H8	1:A_12:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_14:H6	1:G_13:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_14:H6	1:G_13:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_14:H6	1:G_13:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_14:H5	1:G_13:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_14:H5	1:G_13:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_15:H8	1:C_14:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_15:H8	1:C_14:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_15:H8	1:C_14:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_17:H8	1:C_16:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H8	1:C_16:H3'	1.80	3.00	10.00	10.00	1000.00
1:A_17:H8	1:C_16:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_17:H2	1:A_18:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_17:H2'	1:A_18:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_18:H2	1:A_17:H2	1.80	5.00	10.00	10.00	1000.00
1:A_18:H8	1:A_17:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_18:H8	1:A_17:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_18:H8	1:A_17:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_18:H8	1:A_17:H4'	1.80	5.00	10.00	10.00	1000.00
1:A_18:H2'	1:G_19:H4'	1.80	4.00	10.00	10.00	1000.00
1:G_19:H8	1:A_18:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_19:H8	1:A_18:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_19:H1'	1:A_18:H2'	1.80	5.00	10.00	10.00	1000.00
1:G_19:H8	1:C_20:H6	1.80	5.00	10.00	10.00	1000.00
1:C_20:H6	1:G_19:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_20:H6	1:G_19:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_20:H6	1:G_19:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_20:H5	1:G_19:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_20:H5	1:G_19:H3'	1.80	5.00	10.00	10.00	1000.00
1:C_20:H5	1:U_21:H5	1.80	5.00	10.00	10.00	1000.00
1:U_21:H6	1:C_20:H1'	1.80	5.00	10.00	10.00	1000.00
1:U_21:H6	1:C_20:H2'	1.80	3.00	10.00	10.00	1000.00
1:U_21:H5	1:C_20:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_22:H8	1:U_21:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_22:H8	1:C_24:H5	1.80	5.00	10.00	10.00	1000.00
1:G_22:H8	1:U_21:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_12:H2	1:G_22:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_24:H6	1:G_22:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_24:H6	1:G_22:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_24:H6	1:G_25:H8	1.80	5.00	10.00	10.00	1000.00
1:C_24:H5	1:G_22:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_25:H8	1:C_24:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_25:H8	1:C_24:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_25:H8	1:G_26:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_25:H8	1:G_26:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_25:H8	1:G_26:H8	1.80	5.00	10.00	10.00	1000.00
1:G_25:H1'	1:G_26:H2'	1.80	5.00	10.00	10.00	1000.00
1:G_25:H2'	1:G_26:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_25:H2'	1:G_26:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_26:H8	1:G_25:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_26:H8	1:G_25:H2'	1.80	4.00	10.00	10.00	1000.00
1:G_26:H8	1:G_25:H3'	1.80	4.00	10.00	10.00	1000.00
1:G_26:H8	1:A_28:H1'	1.80	4.00	10.00	10.00	1000.00
1:U_27:H3'	1:G_26:H4'	1.80	4.00	10.00	10.00	1000.00
1:A_28:H8	1:G_26:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_28:H8	1:C_29:H5	1.80	5.00	10.00	10.00	1000.00
1:A_28:H2	1:C_29:H1'	1.80	3.00	10.00	10.00	1000.00
1:A_28:H2	1:C_29:H2'	1.80	5.00	10.00	10.00	1000.00
1:A_28:H8	1:U_27:H4'	1.80	5.00	10.00	10.00	1000.00
1:C_29:H6	1:A_28:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_29:H6	1:A_28:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_29:H6	1:A_28:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_29:H5	1:A_28:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_29:H5	1:A_28:H3'	1.80	4.00	10.00	10.00	1000.00
1:A_30:H8	1:C_29:H1'	1.80	5.00	10.00	10.00	1000.00
1:A_30:H8	1:C_29:H2'	1.80	3.00	10.00	10.00	1000.00
1:A_30:H8	1:C_29:H3'	1.80	5.00	10.00	10.00	1000.00
1:A_30:H2	1:G_31:H1'	1.80	4.00	10.00	10.00	1000.00
1:A_30:H2'	1:G_31:H5'*	1.80	5.00	10.00	10.00	1000.00
1:G_31:H8	1:A_30:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_31:H8	1:A_30:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_31:H8	1:A_30:H3'	1.80	4.00	10.00	10.00	1000.00

1:G_31:H1'	1:A_30:H2'	1.80	5.00	10.00	10.00	1000.00
1:G_31:H8	1:G_32:H8	1.80	5.00	10.00	10.00	1000.00
1:G_32:H8	1:G_31:H1'	1.80	5.00	10.00	10.00	1000.00
1:G_32:H8	1:G_31:H2'	1.80	3.00	10.00	10.00	1000.00
1:G_32:H8	1:G_31:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_33:H6	1:G_32:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_33:H6	1:G_32:H2'	1.80	3.00	10.00	10.00	1000.00
1:C_33:H6	1:G_32:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_33:H5	1:G_32:H2'	1.80	4.00	10.00	10.00	1000.00
1:C_33:H5	1:G_32:H3'	1.80	4.00	10.00	10.00	1000.00
1:C_33:H5	1:C_34:H5	1.80	5.00	10.00	10.00	1000.00
1:C_34:H6	1:C_33:H1'	1.80	5.00	10.00	10.00	1000.00
1:C_34:H6	1:C_33:H2'	1.80	3.00	10.00	10.00	1000.00
!intra residue peptide-peptide NOE restraints						
1:ASPN_51:HA	1:ASPN_51:HB*	1.80	4.00	10.00	10.00	1000.00
1:THR_52:HN	1:THR_52:HA	1.80	3.00	10.00	10.00	1000.00
1:THR_52:HN	1:THR_52:HG2*	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ARG_53:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ARG_53:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ARG_53:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ARG_53:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_53:HG*	1:ARG_53:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HD*	1:ARG_53:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_53:HA	1:ARG_53:HE	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HN	1:GLN_54:HB*	1.80	4.00	10.00	10.00	1000.00
1:GLN_54:HN	1:GLN_54:HG*	1.80	4.00	10.00	10.00	1000.00
1:GLN_54:HN	1:GLN_54:HA	1.80	3.00	10.00	10.00	1000.00
1:GLN_54:HA	1:GLN_54:HB*	1.80	4.00	10.00	10.00	1000.00
1:GLN_54:HA	1:GLN_54:HG*	1.80	4.00	10.00	10.00	1000.00
1:GLN_54:HA	1:GLN_54:HE2*	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:GLN_54:HE2*	1.80	7.00	10.00	10.00	1000.00
1:GLN_54:HG*	1:GLN_54:HE2*	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:GLN_54:HG*	1.80	5.00	10.00	10.00	1000.00
1:ALA_55:HN	1:ALA_55:HA	1.80	3.00	10.00	10.00	1000.00
1:ALA_55:HN	1:ALA_55:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HA	1:ALA_55:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_56:HN	1:ARG_56:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_56:HN	1:ARG_56:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_56:HA	1:ARG_56:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_56:HG*	1:ARG_56:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_56:HG*	1:ARG_56:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_56:HD*	1:ARG_56:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_57:HB*	1:ARG_57:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HG*	1:ARG_57:HA	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HN	1:ARG_57:HG*	1.80	6.00	10.00	10.00	1000.00
1:ARG_57:HG*	1:ARG_57:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_57:HG*	1:ARG_57:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HG*	1:ARG_57:HA	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:ARG_57:HA	1.80	6.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:ARG_57:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_57:HB*	1:ARG_57:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:ARG_57:HN	1.80	6.00	10.00	10.00	1000.00
1:ASN_58:HN	1:ASN_58:HA	1.80	3.00	10.00	10.00	1000.00
1:ASN_58:HN	1:ASN_58:HB*	1.80	3.00	10.00	10.00	1000.00
1:ASN_58:HB*	1:ASN_58:HD2*	1.80	4.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ASN_58:HB*	1.80	4.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ASN_58:HD2*	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_59:HN	1.80	3.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_59:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_59:HG*	1.80	4.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_59:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HG*	1:ARG_59:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HG*	1:ARG_59:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HB*	1:ARG_59:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HD*	1:ARG_59:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_59:HB*	1:ARG_59:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_60:HN	1:ARG_60:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_60:HN	1:ARG_60:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HN	1:ARG_60:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_60:HA	1:ARG_60:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_60:HA	1:ARG_60:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HG*	1:ARG_60:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HE	1:ARG_60:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HE	1:ARG_60:HG*	1.80	4.00	10.00	10.00	1000.00
1:ARG_60:HE	1:ARG_60:HD*	1.80	4.00	10.00	10.00	1000.00
1:ARG_61:HN	1:ARG_61:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_61:HN	1:ARG_61:HB*	1.80	4.00	10.00	10.00	1000.00

1:ARG_61:HN	1:ARG_61:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_61:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_61:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_61:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HB*	1:ARG_61:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HG*	1:ARG_61:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HB*	1:ARG_61:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HG*	1:ARG_61:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HD*	1:ARG_61:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HN	1:ARG_62:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_62:HN	1:ARG_62:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HN	1:ARG_62:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HN	1:ARG_62:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HA	1:ARG_62:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HA	1:ARG_62:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HA	1:ARG_62:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HB*	1:ARG_62:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HB*	1:ARG_62:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HB*	1:ARG_62:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HG*	1:ARG_62:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HG*	1:ARG_62:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HD*	1:ARG_62:HE	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HN	1:TRP_63:HA	1.80	3.00	10.00	10.00	1000.00
1:TRP_63:HN	1:TRP_63:HB*	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HN	1:TRP_63:HD1	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HN	1:TRP_63:HE3	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HA	1:TRP_63:HB*	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HB*	1:TRP_63:HD1	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HB*	1:TRP_63:HE3	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HA	1:TRP_63:HD1	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HA	1:TRP_63:HE3	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HN	1:ARG_64:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_64:HN	1:ARG_64:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HN	1:ARG_64:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HN	1:ARG_64:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HA	1:ARG_64:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HA	1:ARG_64:HG*	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HA	1:ARG_64:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HA	1:ARG_64:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HB*	1:ARG_64:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HB*	1:ARG_64:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HB*	1:ARG_64:HD*	1.80	7.00	10.00	10.00	1000.00
1:ARG_64:HD*	1:ARG_64:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HG*	1:ARG_64:HE	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HN	1:GLU_65:HA	1.80	3.00	10.00	10.00	1000.00
1:GLU_65:HN	1:GLU_65:HB*	1.80	4.00	10.00	10.00	1000.00
1:GLU_65:HN	1:GLU_65:HG*	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HA	1:GLU_65:HB*	1.80	4.00	10.00	10.00	1000.00
1:GLU_65:HA	1:GLU_65:HG*	1.80	4.00	10.00	10.00	1000.00
1:GLU_65:HB*	1:GLU_65:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HN	1:ARG_66:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_66:HN	1:ARG_66:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HN	1:ARG_66:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_66:HA	1:ARG_66:HG*	1.80	4.00	10.00	10.00	1000.00
1:ARG_66:HA	1:ARG_66:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_66:HB*	1:ARG_66:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HG*	1:ARG_66:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HG*	1:ARG_66:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HD*	1:ARG_66:HE	1.80	4.00	10.00	10.00	1000.00
1:GLN_67:HN	1:GLN_67:HA	1.80	3.00	10.00	10.00	1000.00
1:GLN_67:HN	1:GLN_67:HG*	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HN	1:GLN_67:HE2*	1.80	6.00	10.00	10.00	1000.00
1:GLN_67:HA	1:GLN_67:HB*	1.80	4.00	10.00	10.00	1000.00
1:GLN_67:HA	1:GLN_67:HG*	1.80	4.00	10.00	10.00	1000.00
1:GLN_67:HA	1:GLN_67:HE2*	1.80	6.00	10.00	10.00	1000.00
1:GLN_67:HB*	1:GLN_67:HG*	1.80	4.00	10.00	10.00	1000.00
1:GLN_67:HB*	1:GLN_67:HE2*	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HG*	1:GLN_67:HE2*	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HN	1:ARG_68:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_68:HN	1:ARG_68:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ARG_68:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_68:HB*	1:ARG_68:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HB*	1:ARG_68:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HB*	1:ARG_68:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HG*	1:ARG_68:HE	1.80	5.00	10.00	10.00	1000.00
1:ALA_69:HN	1:ALA_69:HA	1.80	3.00	10.00	10.00	1000.00
1:ALA_69:HN	1:ALA_69:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_70:HN	1:ALA_70:HB*	1.80	4.00	10.00	10.00	1000.00

1:ALA_70:HA	1:ALA_70:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_71:HN	1:ALA_71:HA	1.80	3.00	10.00	10.00	1000.00
1:ALA_71:HN	1:ALA_71:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_71:HA	1:ALA_71:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_72:HN	1:ALA_72:HA	1.80	3.00	10.00	10.00	1000.00
1:ALA_72:HN	1:ALA_72:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_72:HA	1:ALA_72:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_73:HN	1:ARG_73:HA	1.80	3.00	10.00	10.00	1000.00
1:ARG_73:HN	1:ARG_73:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_73:HN	1:ARG_73:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_73:HA	1:ARG_73:HB*	1.80	4.00	10.00	10.00	1000.00
1:ARG_73:HA	1:ARG_73:HG*	1.80	5.00	10.00	10.00	1000.00
1:ARG_73:HA	1:ARG_73:HD*	1.80	5.00	10.00	10.00	1000.00
1:ARG_73:HA	1:ARG_73:HE	1.80	6.00	10.00	10.00	1000.00
1:ARG_73:HB*	1:ARG_73:HD*	1.80	6.00	10.00	10.00	1000.00
1:ARG_73:HB*	1:ARG_73:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_73:HG*	1:ARG_73:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_73:HD*	1:ARG_73:HE	1.80	4.00	10.00	10.00	1000.00
!inter-residue peptide-peptide NOE restraints						
1:ASPN_51:HA	1:THR_52:HG2*	1.80	6.00	10.00	10.00	1000.00
1:ASPN_51:HA	1:ALA_55:HB*	1.80	6.00	10.00	10.00	1000.00
1:ASPN_51:HA	1:THR_52:HN	1.80	4.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:THR_52:HG2*	1.80	7.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:THR_52:HA	1.80	6.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:THR_52:HN	1.80	5.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:ALA_55:HB*	1.80	7.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:ARG_56:HD*	1.80	8.00	10.00	10.00	1000.00
1:ASPN_51:HB*	1:ARG_56:HE	1.80	7.00	10.00	10.00	1000.00
1:THR_52:HN	1:ARG_53:HN	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HN	1:ALA_55:HN	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HN	1:ALA_55:HA	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HN	1:ALA_55:HB*	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HA	1:ARG_53:HN	1.80	3.00	10.00	10.00	1000.00
1:THR_52:HA	1:GLN_54:HN	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HA	1:ALA_55:HB*	1.80	6.00	10.00	10.00	1000.00
1:THR_52:HB	1:ALA_55:HB*	1.80	7.00	10.00	10.00	1000.00
1:THR_52:HB	1:GLN_54:HB*	1.80	7.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:ARG_53:HN	1.80	5.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:GLN_54:HN	1.80	6.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:GLN_54:HB*	1.80	7.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:GLN_54:HE2*	1.80	7.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:ALA_55:HN	1.80	6.00	10.00	10.00	1000.00
1:THR_52:HG2*	1:ALA_55:HB*	1.80	7.00	10.00	10.00	1000.00
1:ARG_53:HN	1:GLN_54:HN	1.80	3.00	10.00	10.00	1000.00
1:ARG_53:HN	1:GLN_54:HB*	1.80	6.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ALA_55:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HN	1:ALA_55:HB*	1.80	7.00	10.00	10.00	1000.00
1:ARG_53:HG*	1:GLN_54:HN	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HN	1:ALA_55:HN	1.80	3.00	10.00	10.00	1000.00
1:GLN_54:HN	1:ARG_56:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_54:HA	1:ALA_55:HN	1.80	4.00	10.00	10.00	1000.00
1:GLN_54:HA	1:ARG_57:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:ALA_55:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_54:HG*	1:ALA_55:HN	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HG*	1:ASN_58:HD2*	1.80	8.00	10.00	10.00	1000.00
1:ALA_55:HN	1:ARG_56:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HA	1:ARG_56:HN	1.80	5.00	10.00	10.00	1000.00
1:ALA_55:HA	1:ASN_58:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HA	1:ASN_58:HB*	1.80	6.00	10.00	10.00	1000.00
1:ALA_55:HA	1:ASN_58:HD2*	1.80	7.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:ARG_56:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:ASN_58:HD2*	1.80	8.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:ARG_59:HD*	1.80	8.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:ARG_59:HE	1.80	7.00	10.00	10.00	1000.00
1:ARG_56:HA	1:ARG_59:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_56:HN	1:ARG_57:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_57:HA	1:ASN_58:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_57:HA	1:ARG_60:HE	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_59:HN	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_61:HN	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_62:HN	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_61:HB*	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_61:HE	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HA	1:ARG_62:HE	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HB*	1:ARG_62:HE	1.80	7.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_60:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_62:HN	1.80	4.00	10.00	10.00	1000.00

1:ARG_59:HA	1:TRP_63:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_62:HB*	1.80	6.00	10.00	10.00	1000.00
1:ARG_59:HA	1:ARG_62:HE	1.80	4.00	10.00	10.00	1000.00
1:ARG_59:HG*	1:ARG_60:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_59:HD*	1:ARG_60:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_60:HN	1:ARG_61:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HA	1:ARG_61:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_60:HA	1:TRP_63:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_60:HA	1:ARG_64:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_60:HA	1:TRP_63:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_64:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_64:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_61:HA	1:ARG_64:HE	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HN	1:TRP_63:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HA	1:TRP_63:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HA	1:GLU_65:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_62:HA	1:GLU_65:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_62:HA	1:GLU_65:HG*	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HB*	1:TRP_63:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HG*	1:TRP_63:HN	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HD1	1:ARG_66:HG*	1.80	7.00	10.00	10.00	1000.00
1:TRP_63:HE3	1:ARG_64:HA	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HE3	1:ARG_64:HB*	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HE3	1:ARG_64:HG*	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HZ3	1:ARG_64:HA	1.80	7.00	10.00	10.00	1000.00
1:TRP_63:HZ3	1:ARG_64:HB*	1.80	7.00	10.00	10.00	1000.00
1:TRP_63:HE1	1:GLN_67:HE2*	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HN	1:ARG_64:HN	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HA	1:ARG_64:HN	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HA	1:ARG_66:HN	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HA	1:ARG_66:HB*	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HA	1:ARG_66:HG*	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HA	1:ARG_66:HE	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HB*	1:ARG_64:HN	1.80	5.00	10.00	10.00	1000.00
1:TRP_63:HB*	1:GLU_65:HN	1.80	7.00	10.00	10.00	1000.00
1:TRP_63:HB*	1:ARG_66:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HN	1:GLU_65:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HN	1:ARG_66:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HA	1:GLU_65:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_64:HA	1:GLN_67:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HA	1:GLN_67:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HA	1:GLN_67:HG*	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HB*	1:GLU_65:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_64:HG*	1:GLU_65:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HD*	1:GLU_65:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HE	1:GLU_65:HN	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HN	1:ARG_66:HN	1.80	4.00	10.00	10.00	1000.00
1:GLU_65:HA	1:ARG_66:HN	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HA	1:ARG_68:HN	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HA	1:ARG_68:HB*	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HA	1:ARG_68:HE	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HB*	1:ARG_66:HN	1.80	5.00	10.00	10.00	1000.00
1:GLU_65:HB*	1:ALA_69:HB*	1.80	8.00	10.00	10.00	1000.00
1:GLU_65:HB*	1:ARG_68:HE	1.80	7.00	10.00	10.00	1000.00
1:GLU_65:HG*	1:ARG_66:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_66:HA	1:GLN_67:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HD*	1:GLN_67:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_66:HA	1:ALA_69:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_66:HB*	1:ALA_69:HB*	1.80	7.00	10.00	10.00	1000.00
1:ARG_66:HD*	1:ALA_69:HB*	1.80	7.00	10.00	10.00	1000.00
1:GLN_67:HA	1:ARG_68:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HA	1:ALA_70:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HA	1:ALA_71:HN	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HA	1:ALA_70:HB*	1.80	5.00	10.00	10.00	1000.00
1:GLN_67:HE2*	1:ALA_70:HB*	1.80	7.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ALA_69:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ALA_70:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ALA_71:HN	1.80	4.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ALA_72:HN	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HA	1:ALA_71:HB*	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HB*	1:ALA_70:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HB*	1:ALA_71:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HD*	1:ALA_69:HN	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HD*	1:ALA_72:HN	1.80	7.00	10.00	10.00	1000.00
1:ALA_69:HN	1:ALA_70:HN	1.80	3.00	10.00	10.00	1000.00
1:ALA_69:HA	1:ALA_70:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_69:HA	1:ALA_72:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_69:HA	1:ALA_72:HB*	1.80	5.00	10.00	10.00	1000.00

1:ALA_70:HN	1:ALA_71:HN	1.80	3.00	10.00	10.00	1000.00
1:ALA_70:HA	1:ALA_71:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_70:HA	1:ARG_73:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_70:HA	1:ARG_73:HB*	1.80	4.00	10.00	10.00	1000.00
1:ALA_71:HA	1:ALA_72:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_71:HA	1:ARG_73:HN	1.80	5.00	10.00	10.00	1000.00
1:ALA_71:HN	1:ALA_72:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_72:HA	1:ARG_73:HN	1.80	4.00	10.00	10.00	1000.00
1:ALA_72:HN	1:ARG_73:HN	1.80	3.00	10.00	10.00	1000.00

!intermolecular distance restraints

1:ARG_53:HD*	1:G_22:H8	1.80	7.00	10.00	10.00	1000.00
1:ARG_53:HE	1:G_22:H8	1.80	6.00	10.00	10.00	1000.00
1:ARG_53:HD*	1:U_21:H5	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HE	1:U_21:H5	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HD*	1:U_21:H6	1.80	6.00	10.00	10.00	1000.00
1:ARG_53:HE	1:U_21:H6	1.80	5.00	10.00	10.00	1000.00
1:ARG_53:HD*	1:C_20:H2'	1.80	6.00	10.00	10.00	1000.00
1:ARG_53:HE	1:C_20:H2'	1.80	5.00	10.00	10.00	1000.00
1:GLN_54:HN	1:G_7:H3'	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HA	1:G_7:H8	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HA	1:G_26:2X	1.80	7.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:G_7:H8	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:G_6:H1'	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:G_6:H2'	1.80	7.00	10.00	10.00	1000.00
1:GLN_54:HB*	1:G_7:H3'	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HG*	1:G_7:H8	1.80	7.00	10.00	10.00	1000.00
1:GLN_54:HG*	1:G_7:H3'	1.80	7.00	10.00	10.00	1000.00
1:GLN_54:HE2*	1:G_7:H3'	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HE2*	1:G_8:H8	1.80	6.00	10.00	10.00	1000.00
1:GLN_54:HE2*	1:G_26:2X	1.80	8.00	10.00	10.00	1000.00
1:ALA_55:HN	1:G_6:H2'	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HN	1:G_6:H3'	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HA	1:G_6:H8	1.80	5.00	10.00	10.00	1000.00
1:ALA_55:HA	1:G_6:H2'	1.80	3.00	10.00	10.00	1000.00
1:ALA_55:HA	1:G_6:H3'	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_7:H8	1.80	7.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_6:H8	1.80	6.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_6:H2'	1.80	5.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_6:H3'	1.80	4.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_6:H4'	1.80	7.00	10.00	10.00	1000.00
1:ALA_55:HB*	1:G_6:H5'*	1.80	8.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:C_24:H5	1.80	7.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:C_24:H41	1.80	7.00	10.00	10.00	1000.00
1:ARG_57:HD*	1:G_26:2X	1.80	7.00	10.00	10.00	1000.00
1:ARG_57:HE	1:G_26:2X	1.80	6.00	10.00	10.00	1000.00
1:ARG_57:HE	1:G_25:H8	1.80	6.00	10.00	10.00	1000.00
1:ARG_57:HE	1:C_24:H41	1.80	5.00	10.00	10.00	1000.00
1:ASN_58:HB*	1:G_6:H8	1.80	6.00	10.00	10.00	1000.00
1:ASN_58:HD2*	1:G_7:H3'	1.80	6.00	10.00	10.00	1000.00
1:ASN_58:HD2*	1:G_6:H1'	1.80	6.00	10.00	10.00	1000.00
1:ASN_58:HD2*	1:G_26:2X	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HE	1:G_26:2X	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HE	1:U_27:H5'*	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HD*	1:G_26:H1'	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HD*	1:U_27:H4'	1.80	6.00	10.00	10.00	1000.00
1:ARG_61:HD*	1:U_27:H5'*	1.80	7.00	10.00	10.00	1000.00
1:ARG_61:HG*	1:U_27:H4'	1.80	7.00	10.00	10.00	1000.00
1:ARG_61:HG*	1:U_27:H5'*	1.80	8.00	10.00	10.00	1000.00
1:ARG_62:HG*	1:C_4:H5	1.80	7.00	10.00	10.00	1000.00
1:ARG_62:HG*	1:U_5:H5	1.80	7.00	10.00	10.00	1000.00
1:ARG_62:HD*	1:C_4:H5	1.80	7.00	10.00	10.00	1000.00
1:ARG_62:HD*	1:U_5:H5	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HE	1:C_4:H5	1.80	6.00	10.00	10.00	1000.00
1:ARG_62:HE	1:U_5:H5	1.80	4.00	10.00	10.00	1000.00
1:TRP_63:HZ3	1:A_23:H2	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HD1	1:A_23:H2	1.80	6.00	10.00	10.00	1000.00
1:TRP_63:HE3	1:A_23:H2	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HD*	1:A_23:H1'	1.80	6.00	10.00	10.00	1000.00
1:ARG_64:HE	1:A_23:H1'	1.80	5.00	10.00	10.00	1000.00
1:ARG_68:HD*	1:U_27:H5	1.80	6.00	10.00	10.00	1000.00
1:ARG_68:HE	1:U_27:H5	1.80	5.00	10.00	10.00	1000.00

#NMR_dihedral

!phi torsions from HNHA experiment

1:ASPN_51:C	1:THR_52:N	1:THR_52:CA	1:THR_52:C	-156.0	-84.0	60.00	60.00	1000.00
1:THR_52:C	1:ARG_53:N	1:ARG_53:CA	1:ARG_53:C	-60.0	18.0	60.00	60.00	1000.00
1:ARG_53:C	1:GLN_54:N	1:GLN_54:CA	1:GLN_54:C	-71.0	-51.0	60.00	60.00	1000.00

1:GLN_54:C	1:ALA_55:N	1:ALA_55:CA	1:ALA_55:C	-61.0	-41.0	60.00	60.00	1000.00
1:ALA_55:C	1:ARG_56:N	1:ARG_56:CA	1:ARG_56:C	-78.0	-38.0	60.00	60.00	1000.00
1:ARG_57:C	1:ASN_58:N	1:ASN_58:CA	1:ASN_58:C	-78.0	-38.0	60.00	60.00	1000.00
1:ARG_59:C	1:ARG_60:N	1:ARG_60:CA	1:ARG_60:C	-67.0	-47.0	60.00	60.00	1000.00
1:ARG_60:C	1:ARG_61:N	1:ARG_61:CA	1:ARG_61:C	-77.0	-37.0	60.00	60.00	1000.00
1:ARG_61:C	1:ARG_62:N	1:ARG_62:CA	1:ARG_62:C	-74.0	-54.0	60.00	60.00	1000.00
1:ARG_62:C	1:TRP_63:N	1:TRP_63:CA	1:TRP_63:C	-78.0	-58.0	60.00	60.00	1000.00
1:TRP_63:C	1:ARG_64:N	1:ARG_64:CA	1:ARG_64:C	-76.0	-56.0	60.00	60.00	1000.00
1:ARG_64:C	1:GLU_65:N	1:GLU_65:CA	1:GLU_65:C	-75.0	-55.0	60.00	60.00	1000.00
1:GLU_65:C	1:ARG_66:N	1:ARG_66:CA	1:ARG_66:C	-61.0	-41.0	60.00	60.00	1000.00
1:ARG_66:C	1:GLN_67:N	1:GLN_67:CA	1:GLN_67:C	-74.0	-54.0	60.00	60.00	1000.00
1:GLN_67:C	1:ARG_68:N	1:ARG_68:CA	1:ARG_68:C	-79.0	-59.0	60.00	60.00	1000.00
1:ARG_68:C	1:ALA_69:N	1:ALA_69:CA	1:ALA_69:C	-75.0	-55.0	60.00	60.00	1000.00
1:ALA_69:C	1:ALA_70:N	1:ALA_70:CA	1:ALA_70:C	-76.0	-56.0	60.00	60.00	1000.00
!planar restraints to keep amino protons in base pairs from flipping								
1:G_1:H22	1:G_1:N2	1:G_1:C2	1:G_1:N1	-5.0	5.0	60.00	60.00	1000.00
1:G_2:H22	1:G_2:N2	1:G_2:C2	1:G_2:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_4:H42	1:C_4:N4	1:C_4:C4	1:C_4:N3	-5.0	5.0	60.00	60.00	1000.00
1:G_6:H22	1:G_6:N2	1:G_6:C2	1:G_6:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_9:H42	1:C_9:N4	1:C_9:C4	1:C_9:N3	-5.0	5.0	60.00	60.00	1000.00
1:G_10:H22	1:G_10:N2	1:G_10:C2	1:G_10:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_11:H42	1:C_11:N4	1:C_11:C4	1:C_11:N3	-5.0	5.0	60.00	60.00	1000.00
1:A_12:H61	1:A_12:N6	1:A_12:C6	1:A_12:N1	-5.0	5.0	60.00	60.00	1000.00
1:G_13:H22	1:G_13:N2	1:G_13:C2	1:G_13:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_14:H42	1:C_14:N4	1:C_14:C4	1:C_14:N3	-5.0	5.0	60.00	60.00	1000.00
1:G_15:H22	1:G_15:N2	1:G_15:C2	1:G_15:N1	-5.0	5.0	60.00	60.00	1000.00
1:A_18:H61	1:A_18:N6	1:A_18:C6	1:A_18:C5	-5.0	5.0	60.00	60.00	1000.00
1:G_19:H22	1:G_19:N2	1:G_19:C2	1:G_19:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_20:H42	1:C_20:N4	1:C_20:C4	1:C_20:N3	-5.0	5.0	60.00	60.00	1000.00
1:G_22:H22	1:G_22:N2	1:G_22:C2	1:G_22:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_24:H42	1:C_24:N4	1:C_24:C4	1:C_24:N3	-5.0	5.0	60.00	60.00	1000.00
1:G_25:H22	1:G_25:N2	1:G_25:C2	1:G_25:N1	-5.0	5.0	60.00	60.00	1000.00
1:A_28:H61	1:A_28:N6	1:A_28:C6	1:A_28:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_29:H42	1:C_29:N4	1:C_29:C4	1:C_29:N3	-5.0	5.0	60.00	60.00	1000.00
1:A_30:H61	1:A_30:N6	1:A_30:C6	1:A_30:N1	-5.0	5.0	60.00	60.00	1000.00
1:G_31:H22	1:G_31:N2	1:G_31:C2	1:G_31:N1	-5.0	5.0	60.00	60.00	1000.00
1:G_32:H22	1:G_32:N2	1:G_32:C2	1:G_32:N1	-5.0	5.0	60.00	60.00	1000.00
1:C_33:H42	1:C_33:N4	1:C_33:C4	1:C_33:N3	-5.0	5.0	60.00	60.00	1000.00
1:C_34:H42	1:C_34:N4	1:C_34:C4	1:C_34:N3	-5.0	5.0	60.00	60.00	1000.00
!glycosidic constraints								
1:G_1:O4'	1:G_1:C1'	1:G_1:N9	1:G_1:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_2:O4'	1:G_2:C1'	1:G_2:N9	1:G_2:C4	170.0	-125.0	60.00	60.00	1000.00
1:U_3:O4'	1:U_3:C1'	1:U_3:N1	1:U_3:C2	170.0	-125.0	60.00	60.00	1000.00
1:C_4:O4'	1:C_4:C1'	1:C_4:N1	1:C_4:C2	170.0	-125.0	60.00	60.00	1000.00
1:U_5:O4'	1:U_5:C1'	1:U_5:N1	1:U_5:C2	170.0	-125.0	60.00	60.00	1000.00
1:G_6:O4'	1:G_6:C1'	1:G_6:N9	1:G_6:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_7:O4'	1:G_7:C1'	1:G_7:N9	1:G_7:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_8:O4'	1:G_8:C1'	1:G_8:N9	1:G_8:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_9:O4'	1:C_9:C1'	1:C_9:N1	1:C_9:C2	170.0	-125.0	60.00	60.00	1000.00
1:G_10:O4'	1:G_10:C1'	1:G_10:N9	1:G_10:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_11:O4'	1:C_11:C1'	1:C_11:N1	1:C_11:C2	170.0	-125.0	60.00	60.00	1000.00
1:A_12:O4'	1:A_12:C1'	1:A_12:N9	1:A_12:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_13:O4'	1:G_13:C1'	1:G_13:N9	1:G_13:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_14:O4'	1:C_14:C1'	1:C_14:N1	1:C_14:C2	170.0	-125.0	60.00	60.00	1000.00
1:G_15:O4'	1:G_15:C1'	1:G_15:N9	1:G_15:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_16:O4'	1:C_16:C1'	1:C_16:N1	1:C_16:C2	170.0	-125.0	60.00	60.00	1000.00
1:A_17:O4'	1:A_17:C1'	1:A_17:N9	1:A_17:C4	170.0	-125.0	60.00	60.00	1000.00
1:A_18:O4'	1:A_18:C1'	1:A_18:N9	1:A_18:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_19:O4'	1:G_19:C1'	1:G_19:N9	1:G_19:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_20:O4'	1:C_20:C1'	1:C_20:N1	1:C_20:C2	170.0	-125.0	60.00	60.00	1000.00
1:U_21:O4'	1:U_21:C1'	1:U_21:N1	1:U_21:C2	170.0	-125.0	60.00	60.00	1000.00
1:G_22:O4'	1:G_22:C1'	1:G_22:N9	1:G_22:C4	170.0	-125.0	60.00	60.00	1000.00
1:A_23:O4'	1:A_23:C1'	1:A_23:N9	1:A_23:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_24:O4'	1:C_24:C1'	1:C_24:N1	1:C_24:C2	170.0	-125.0	60.00	60.00	1000.00
1:G_25:O4'	1:G_25:C1'	1:G_25:N9	1:G_25:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_26:O4'	1:G_26:C1'	1:G_26:N9	1:G_26:C4	170.0	-125.0	60.00	60.00	1000.00
1:U_27:O4'	1:U_27:C1'	1:U_27:N1	1:U_27:C2	170.0	-125.0	60.00	60.00	1000.00
1:A_28:O4'	1:A_28:C1'	1:A_28:N9	1:A_28:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_29:O4'	1:C_29:C1'	1:C_29:N1	1:C_29:C2	170.0	-125.0	60.00	60.00	1000.00
1:A_30:O4'	1:A_30:C1'	1:A_30:N9	1:A_30:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_31:O4'	1:G_31:C1'	1:G_31:N9	1:G_31:C4	170.0	-125.0	60.00	60.00	1000.00
1:G_32:O4'	1:G_32:C1'	1:G_32:N9	1:G_32:C4	170.0	-125.0	60.00	60.00	1000.00
1:C_33:O4'	1:C_33:C1'	1:C_33:N1	1:C_33:C2	170.0	-125.0	60.00	60.00	1000.00
1:C_34:O4'	1:C_34:C1'	1:C_34:N1	1:C_34:C2	170.0	-125.0	60.00	60.00	1000.00

!C3'-endo constraints. delta and nu2 only

1:G_1:C5'	1:G_1:C4'	1:G_1:C3'	1:G_1:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_1:C1'	1:G_1:C2'	1:G_1:C3'	1:G_1:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_2:C5'	1:G_2:C4'	1:G_2:C3'	1:G_2:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_2:C1'	1:G_2:C2'	1:G_2:C3'	1:G_2:C4'	33.00	43.00	60.00	60.00	1000.00
1:U_3:C5'	1:U_3:C4'	1:U_3:C3'	1:U_3:O3'	77.00	87.00	60.00	60.00	1000.00
1:U_3:C1'	1:U_3:C2'	1:U_3:C3'	1:U_3:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_4:C5'	1:C_4:C4'	1:C_4:C3'	1:C_4:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_4:C1'	1:C_4:C2'	1:C_4:C3'	1:C_4:C4'	33.00	43.00	60.00	60.00	1000.00
1:U_5:C5'	1:U_5:C4'	1:U_5:C3'	1:U_5:O3'	77.00	87.00	60.00	60.00	1000.00
1:U_5:C1'	1:U_5:C2'	1:U_5:C3'	1:U_5:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_7:C5'	1:G_7:C4'	1:G_7:C3'	1:G_7:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_7:C1'	1:G_7:C2'	1:G_7:C3'	1:G_7:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_9:C5'	1:C_9:C4'	1:C_9:C3'	1:C_9:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_9:C1'	1:C_9:C2'	1:C_9:C3'	1:C_9:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_10:C5'	1:G_10:C4'	1:G_10:C3'	1:G_10:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_10:C1'	1:G_10:C2'	1:G_10:C3'	1:G_10:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_11:C5'	1:C_11:C4'	1:C_11:C3'	1:C_11:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_11:C1'	1:C_11:C2'	1:C_11:C3'	1:C_11:C4'	33.00	43.00	60.00	60.00	1000.00
1:A_12:C5'	1:A_12:C4'	1:A_12:C3'	1:A_12:O3'	77.00	87.00	60.00	60.00	1000.00
1:A_12:C1'	1:A_12:C2'	1:A_12:C3'	1:A_12:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_13:C5'	1:G_13:C4'	1:G_13:C3'	1:G_13:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_13:C1'	1:G_13:C2'	1:G_13:C3'	1:G_13:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_14:C5'	1:C_14:C4'	1:C_14:C3'	1:C_14:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_14:C1'	1:C_14:C2'	1:C_14:C3'	1:C_14:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_15:C5'	1:G_15:C4'	1:G_15:C3'	1:G_15:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_15:C1'	1:G_15:C2'	1:G_15:C3'	1:G_15:C4'	33.00	43.00	60.00	60.00	1000.00
1:A_18:C5'	1:A_18:C4'	1:A_18:C3'	1:A_18:O3'	77.00	87.00	60.00	60.00	1000.00
1:A_18:C1'	1:A_18:C2'	1:A_18:C3'	1:A_18:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_19:C5'	1:G_19:C4'	1:G_19:C3'	1:G_19:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_19:C1'	1:G_19:C2'	1:G_19:C3'	1:G_19:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_20:C5'	1:C_20:C4'	1:C_20:C3'	1:C_20:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_20:C1'	1:C_20:C2'	1:C_20:C3'	1:C_20:C4'	33.00	43.00	60.00	60.00	1000.00
1:U_21:C5'	1:U_21:C4'	1:U_21:C3'	1:U_21:O3'	77.00	87.00	60.00	60.00	1000.00
1:U_21:C1'	1:U_21:C2'	1:U_21:C3'	1:U_21:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_24:C5'	1:C_24:C4'	1:C_24:C3'	1:C_24:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_24:C1'	1:C_24:C2'	1:C_24:C3'	1:C_24:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_25:C5'	1:G_25:C4'	1:G_25:C3'	1:G_25:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_25:C1'	1:G_25:C2'	1:G_25:C3'	1:G_25:C4'	33.00	43.00	60.00	60.00	1000.00
1:A_28:C5'	1:A_28:C4'	1:A_28:C3'	1:A_28:O3'	77.00	87.00	60.00	60.00	1000.00
1:A_28:C1'	1:A_28:C2'	1:A_28:C3'	1:A_28:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_29:C5'	1:C_29:C4'	1:C_29:C3'	1:C_29:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_29:C1'	1:C_29:C2'	1:C_29:C3'	1:C_29:C4'	33.00	43.00	60.00	60.00	1000.00
1:A_30:C5'	1:A_30:C4'	1:A_30:C3'	1:A_30:O3'	77.00	87.00	60.00	60.00	1000.00
1:A_30:C1'	1:A_30:C2'	1:A_30:C3'	1:A_30:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_31:C5'	1:G_31:C4'	1:G_31:C3'	1:G_31:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_31:C1'	1:G_31:C2'	1:G_31:C3'	1:G_31:C4'	33.00	43.00	60.00	60.00	1000.00
1:G_32:C5'	1:G_32:C4'	1:G_32:C3'	1:G_32:O3'	77.00	87.00	60.00	60.00	1000.00
1:G_32:C1'	1:G_32:C2'	1:G_32:C3'	1:G_32:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_33:C5'	1:C_33:C4'	1:C_33:C3'	1:C_33:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_33:C1'	1:C_33:C2'	1:C_33:C3'	1:C_33:C4'	33.00	43.00	60.00	60.00	1000.00
1:C_34:C5'	1:C_34:C4'	1:C_34:C3'	1:C_34:O3'	77.00	87.00	60.00	60.00	1000.00
1:C_34:C1'	1:C_34:C2'	1:C_34:C3'	1:C_34:C4'	33.00	43.00	60.00	60.00	1000.00
!C2'-endo constraints. nu2 and nu4 only								
1:G_6:C1'	1:G_6:C2'	1:G_6:C3'	1:G_6:C4'	-45.00	-35.00	60.00	60.00	1000.00
1:G_6:C3'	1:G_6:C4'	1:G_6:O4'	1:G_6:C1'	-5.00	5.00	60.00	60.00	1000.00
1:A_23:C1'	1:A_23:C2'	1:A_23:C3'	1:A_23:C4'	-45.00	-35.00	60.00	60.00	1000.00
1:A_23:C3'	1:A_23:C4'	1:A_23:O4'	1:A_23:C1'	-5.00	5.00	60.00	60.00	1000.00
1:G_26:C1'	1:G_26:C2'	1:G_26:C3'	1:G_26:C4'	-45.00	-35.00	60.00	60.00	1000.00
1:G_26:C3'	1:G_26:C4'	1:G_26:O4'	1:G_26:C1'	-5.00	5.00	60.00	60.00	1000.00
1:U_27:C1'	1:U_27:C2'	1:U_27:C3'	1:U_27:C4'	-45.00	-35.00	60.00	60.00	1000.00
1:U_27:C3'	1:U_27:C4'	1:U_27:O4'	1:U_27:C1'	-5.00	5.00	60.00	60.00	1000.00

Table B.2: High temperature simulated annealing protocol (Insight II *.tab file). For completeness all variables in the *.tab file are given, yet pertinent values are boxed

Phase_Num	1	2	3	4	5	6	7	8	9
Phase_Type	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics
Minimization_Alg	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest
Num_Iter	100	100	100	100	100	100	100	100	100
Max_Deriv	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Temp	1000.0	1000.0	1000.0	1000.0	500.0	200.0	100.0	50.0	10.0
Time_Step	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Equilibration_Time	0	0	0	0	0	0	0	0	0
Simulation_Time	9000	9000	18000	6000	3000	3000	3000	3000	3000
Num_Scaling_Steps	10	10	10	1	1	1	1	1	1
Demax	1e+12	1e+12	1e+12	1e+12	1e+12	1e+12	1e+12	1e+12	1e+12
Randomize_Velocity	off	off	off	off	off	off	off	off	off
History_Print_Freq	100000	100000	100000	100000	100000	100000	100000	100000	100000
Nonbond_Form	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic
Cutoff	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
R_vdW_I	1.25	1.25	1.25	0.825	0.825	0.825	0.825	0.825	0.825
R_vdW_F	1.25	1.25	0.825	0.825	0.825	0.825	0.825	0.825	0.825
Charges	off	off	off	off	off	off	off	off	off
Dielectric	Constant	Constant	Constant	Constant	Constant	Constant	Constant	Constant	Constant
Dielectric_Value	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Global_Nonbnd_Scale	on	on	on	on	on	on	on	on	on
Nonbond_I	0.01	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0
Nonbond_F	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
VDW_I	0.01	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0
VDW_F	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Coulomb_I	0.01	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0
Coulomb_F	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1_4_I	0.01	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0
1_4_F	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Use_Restraints	on	on	on	on	on	on	on	on	on
Restraint_File	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt	complexnoa.rstrnt
Fix_All_Floats	off	off	off	off	off	off	off	off	off
FKChir	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
Use_Dist_Rstrnt	on	on	on	on	on	on	on	on	on
Use_Dihed_Rstrnt	on	on	on	on	on	on	on	on	on
Use_Chiral_Rstrnt	on	on	on	on	on	on	on	on	on
Use_NOE_Vol_Rstrnt	off	off	off	off	off	off	off	off	off

Global_Rstrnt_Scale	on	on	on	on	on	on	on	on	on
Rstrnt_I	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Rstrnt_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dist_I	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dist_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dihed_I	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dihed_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Chiral_I	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Chiral_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NOE_Vol_I	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NOE_Vol_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Expont_a	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Expont_b	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Scale_By_Inv_Vol	off	off	off	off	off	off	off	off	off
NOE_Neigh	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
NOE_Gradient_Limit	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
Full_Calc_Thresh	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Randomize_Coords	off	off	off	off	off	off	off	off	off
Global_Cvlnt_Scale	on	on	on	on	on	on	on	on	on
Cvlnt_I	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Cvlnt_F	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Bonds_I	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Bonds_F	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Angle_I	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Angle_F	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Torsion_I	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Torsion_F	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Oops_I	0.01	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Oops_F	0.01	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Print_Dist_Viol	off	off	off	off	off	off	off	off	on
Dist_Viol_Thresh	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Print_Dihed_Viol	off	off	off	off	off	off	off	off	on
Dihed_Viol_Thresh	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0
Print_Chiral_Viol	off	off	off	off	off	off	off	off	off
Print_NOE_Vol_Viol	on	on	on	on	on	on	on	on	on
NOE_Vol_Viol_Thresh	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Print_Internals	None	None	None	None	None	None	None	None	None
Num_Largest_Distort	10	10	10	10	10	10	10	10	10
Spec_Freq	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Rot_Corr_Time	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T1_Leakage_From_DB	off	off	off	off	off	off	off	off	off
Default_T1_Leakage	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table B.3: Refinement simulated annealing protocol (Insight II *.tab file). For completeness all variables in the *.tab file are given, yet pertinent values are boxed.

Phase_Num	1	2	3	4	5	6	7	8	9
Phase_Type	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Dynamics	Minimization	Minimization
Minimization_Alg	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Steepest	Conjugate
Num_Iter	100	100	100	100	100	100	100	100	2000
Max_Deriv	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.1
Temp	500.0	500.0	250.0	125.0	60.0	30.0	10.0	1000	1000
Time_Step	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1	1
Equilibration_Time	0	0	0	0	0	0	0	1000	1000
Simulation_Time	6000	2000	1000	1000	1000	1000	3000	1000	1000
Num_Scaling_Steps	6	1	1	1	1	1	1	2	2
Demax	1e+10	1e+10	1e+10	1e+10	1e+10	1e+10	1e+10	800000	800000
Randomize_Velocity	off	off	off	off	off	off	off	off	off
History_Print_Freq	100000	100000	100000	100000	100000	100000	100000	100000	100000
Nonbond_Trm	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic	SA_Quartic
Cutoff	6.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
R_vdW_I	1.25	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825
R_vdW_F	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.825
Charges	off	off	off	off	off	off	off	off	off
Dielectric	Constant	Constant	Constant	Constant	Constant	Constant	Constant	Constant	Constant
Dielectric_Value	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Global_Nonbnd_Scale	on	on	on	on	on	on	on	off	off
Nonbond_I	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Nonbond_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
VDW_I	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
VDW_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Coulomb_I	0.1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Coulomb_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1_4_I	0.1	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5
1_4_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5
Use_Restraints	on	on	on	on	on	on	on	on	on
Restraint_File	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt	refine.rstrnt
Fix_All_Floats	off	off	off	off	off	off	off	off	off
FKChir	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
Use_Dist_Rstrnt	on	on	on	on	on	on	on	on	on
Use_Dihed_Rstrnt	on	on	on	on	on	on	on	on	on
Use_Chiral_Rstrnt	on	on	on	on	on	on	on	on	on
Use_NOE_Vol_Rstrnt	off	off	off	off	off	off	off	off	off
Global_Rstrnt_Scale	on	on	on	on	on	on	on	on	on

Rstrnt_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Rstrnt_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dist_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dist_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dihed_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Dihed_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Chiral_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Chiral_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NOE_Vol_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
NOE_Vol_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Expont_a	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Expont_b	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Scale_By_Inv_Vol	off	off	off	off	off	off	off	off	off
NOE_Neigh	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
NOE_Gradient_Limit	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0
Full_Calc_Thresh	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Randomize_Coords	off	off	off	off	off	off	off	off	off
Global_CvInt_Scale	on	on	on	on	on	on	on	on	on
CvInt_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
CvInt_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Bonds_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Bonds_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Angle_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Angle_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Torsion_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Torsion_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Oops_I	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Oops_F	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Print_Dist_Viol	off	off	off	off	off	off	on	off	on
Dist_Viol_Thresh	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.2
Print_Dihed_Viol	off	off	off	off	off	off	on	off	on
Dihed_Viol_Thresh	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	5.0
Print_Chiral_Viol	off	off	off	off	off	off	off	off	off
Print_NOE_Vol_Viol	on	on	on	on	on	on	on	off	off
NOE_Vol_Viol_Thresh	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Print_Internals	None	None	None	None	None	None	None	None	Largest_Distort
Num_Largest_Distort	10	10	10	10	10	10	10	10	10
Spec_Freq	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Rot_Corr_Time	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T1_Leakage_From_DB	off	off	off	off	off	off	off	off	off
Default_T1_Leakage	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

REFERENCES

Aboul-ela, F., Karn, J. and Varani, G. (1995) "The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein." *J. Mol. Biol.* **253**: 313-332.

Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D. and Varani, G. (1996) "Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation." *Nature* **380**: 646-650.

Auer, M., Gremlich, H.-U., Seifert, J.-M., Daly, T. J., Parslow, T. G., Casari, G. and Gstach, H. (1994) "Helix-Loop-Helix Motif in HIV-1 Rev." *Biochemistry* **33**: 2988-2996.

Aziz, N. and Munro, H. N. (1987) "Iron regulates ferritin mRNA translation through a segment of its 5' untranslated region." *Proc. Natl. Acad. Sci. U.S.A.* **84**: 8478-8482.

Bartel, D. P., Zapp, M. L., Green, M. R. and Szostak, J. W. (1991) "HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA." *Cell* **67**: 529-536.

Batey, R. T., Battiste, J. L. and Williamson, J. R. (1995) Preparation of Isotopically Enriched RNAs for Heteronuclear NMR. Methods in Enzymology. Vol. 261, San Diego, Academic Press. 300-322.

Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D. and Williamson, J. R. (1992) "Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA." *Nucleic Acids Res.* **20**: 4515-4523.

Battiste, J. L., Tan, R., Frankel, A. D. and Williamson, J. R. (1994) "Binding of an HIV Rev peptide to Rev Responsive Element RNA induces formation of purine-purine base pairs." *Biochemistry* **33**: 2741-2747.

Bax, A. (1994) "Multidimensional nuclear magnetic resonance methods for protein studies." *Current Opinion in Structural Biology* **4**: 738-744.

Bax, A. and Davis, D. G. (1985a) "MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy." *J. Magn. Reson.* **65**: 355-360.

Bax, A. and Davis, D. G. (1985b) "Practical Aspects of Two-Dimensional Transverse NOE Spectroscopy." *J. Magn. Reson.* **63**: 207-213.

Beamer, L. J. and Pabo, C. O. (1992) "Refined 1.8 Angstroms Crystal Structure of the Lambda Repressor-Operator Complex." *J. Mol. Biol.* **227**: 177.

Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988) "Sex-lethal, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins." *Cell* **55**: 1037-1046.

Berkhout, B., Silverman, R. H. and Jeang, K.-T. (1989) "Tat *trans*-activates the humanimmunodeficiency virus through a nascent RNA target." *Cell* **59**: 273-282.

Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) "The 2.9 Å crystal structure of T. thermophilus Seryl-tRNA synthetase complexed with $tRNA^{Ser}$." *Science* **263**: 1404-1410.

Boelens, W. C., Jansen, E. J. R., Venrooij, W. J. v., Stripecke, R., Mattaj, I. W. and Gunderson, S. I. (1993) "The Human U1 snRNP-Specific U1A Protein Inhibits Polyadenylation of Its Own Pre-mRNA." *Cell* **72**: 881-892.

Bogerd, H. P., Fridell, R. A., Madore, S. and Cullen, B. R. (1995) "Identification of Novel Cellular Cofactor for the Rev/Rex Class of Retroviral Regulatory Proteins." *Cell* **82**: 485-494.

Braunschweiler, L. and Ernst, R. R. (1983) "Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy." *J. Magn. Reson.* **53**: 521-528.

Brunger, A. T. (1992) X-PLOR User Manual. (Version 3.1). New Haven, Yale University.

Burd, C. G. and Dreyfuss, G. (1994) "Conserved Structures and Diveristy of Functions of RNA-Binding Proteins." *Science* **265**: 615-621.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. and Frankel, A. D. (1991) "Arginine-Mediated RNA Recognition: The Arginine Fork." *Science* **252**: 1167-1171.

Chen, L. and Frankel, A. D. (1994) "An RNA-Binding Peptide from Bovine Immunodeficiency Virus Tat Protein Recognizes an Unusual RNA Structure." *Biochemistry* **33**: 2708-2715.

Chen, L. and Frankel, A. D. (1995) "A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA." *Proc. Natl. Acad. Sci. U.S.A.* **92**: 5077-5081.

Cheong, C., Varani, G. and Tinoco, I., Jr. (1990) "Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC." *Nature* **346**: 680-682.

Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T. and Gronenborn, A. M. (1990) "Assignment of the Side-Chain 1H and 13C Resonances of Interleukin-1β Using Double- and Triple-Resonance Heteronuclear Three-Dimensional NMR Spectroscopy." *Biochemistry* **29**: 8172-8184.

Cole, J. L., Gehman, J. D., Shafer, J. A. and Kuo, L. C. (1993) "Solution Oligomerization of the rev Protein of HIV-1: Implications for Function." *Biochemistry* **32**: 11769-11775.

Cook, K. S., Fisk, G. J., Hauber, J., Usman, N., Daly, T. J. and Rusche, J. R. (1991) "Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate." *Nucleic Acids Res.* **19**: 1577-1582.

Cuenoud, B. and Schepartz, A. (1993) "Altered specificity of DNA-binding proteins with transition metal dimerization domains." *Science* **259**: 510-513.

Curtis, D., Lehmann, R. and Zamore, P. D. (1995) "Translational Regulation in Development." *Cell* **81**: 171-178.

Doig, A. J. and Baldwin, R. L. (1995) "N- and C-capping preferences for all 20 amino acids in α-helical peptides." *Protein Sci.* **4**: 1325-1336.

Ellenberger, T. E., Brandi, C. J., Struhl, K. and Harrison, S. C. (1992) "The GCN4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted α Helices: Crystal Structure of the Protein-DNA Complex." *Cell* **71**: 1223-1237.

Feng, S. and Holland, E. C. (1988) "HIV-1 *tat trans*-activation requires the loop sequence within *tar*." *Nature* **334**: 165-167.

Fiala, R., Jiang, F. and Patel, D. J. (1996) "Direct Correlation of Exchangeable and Nonexchageable Protons on Purine Bases in $^{13}C,^{15}N$-Labeled RNA Using a HCCNH-TOCSY Experiment." *J. Am. Chem. Soc.* **118**: 689-690.

Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W. and Luhrmann, R. (1995) "The HIV-1 Rev Activation Domain Is a Nuclear Export Signal That Accesses an Export Pathway Used by Specific Cellular RNAs." *Cell* **82**: 475-483.

Frankel, A. D. and Kim, P. S. (1991) "Modular Structure of Transciption Factors: Implications for Gene Regulation." *Cell* **65**: 717-719.

Gao, X. and Patel, D. J. (1988) "G(syn)*A(anti) Mismatch Formation in DNA Deodecamers at Acidic pH: pH-Dependent Conformational Transition of G-A Mispairs Detected by Proton NMR." *J. Am. Chem. Soc.* **110**: 5178-5182.

Gesteland, R. F. and Atkins, J. F., Ed. (1993) The RNA World. Plainview, Cold Spring Harbor Laboratory Press.

Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. and Ellington, A. D. (1993) "Selective optimization of the Rev-binding element of HIV-1." *Nucleic Acids Res.* **21**: 5509-5516.

Grodberg, J. and Dunn, J. J. (1988) "*ompT* Encodes the *Escherichia coli* Outer Membrane Protease That Cleaves T7 RNA Polymerase during Purification." *Journal of Bacteriology* **170**: 1245-1253.

Gronenborn, A. M. and Clore, G. M. (1994) "Identification of N-terminal helix capping boxes by means of 13C chemical shifts." *J. Biomol. NMR* **4**: 455-458.

Grzesiek, S. and Bax, A. (1993) "The Importance of Not Saturation H_2O in Protein NMR. Appliation to Sensitivity Enhancement and NOE Measurements." *J. Am. Chem. Soc.* **115**: 12593-12594.

Gutell, R. R., Weiser, B., Woese, C. R. and Noller, H. F. (1985) "Comparative Anatomy of 16-S-like Ribosomal RNA." *Progress in Nucleic Acid Research and Molecular Biology* **32**: 155-216.

Guthrie, C. (1991) "Messenger RNA Splicing in Yeast: Clues to Why the Spliceosome is a Ribonucleoprotein." *Science* **253**: 157-163.

Harada, K., Martin, S. A. and Frankel, A. D. (1996) "Selection of RNA-binding peptides in vivo." *Nature* **380**: 175-179.

Harper, E. T. and Rose, G. D. (1993) "Helix Stop Signals in Proteins and Peptides: The Capping Box." *Biochemistry* **32**: 7605-7609.

Heaphy, S., Finch, J. T., Gait, M. J., Karn, J. and Singh, M. (1991) "Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs." *Proc. Natl. Acad. Sci. U.S.A.* **88**: 7366-7370.

Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B. and Klausner, R. D. (1987) "Identification of the iron-responsive element for the translational regulation of human ferritin mRNA." *Science* **238**: 1570-73.

Heus, H. A. and Pardi, A. (1991) "Structural Features That Give Rise to the Unusual Stability of RNA Hairpins Containing GNRA Loops." *Science* **253**: 191-194.

Heus, H. A., Wijmenga, S. S., Ven, F. J. M. v. d. and Hilbers, C. W. (1994) "Sequential Backbone Assignment in ^{13}C-labeled RNA via Through-bond Coherence Transfer Using Three-Dimensional Triple Resonance Spectroscopy (^{1}H, ^{13}C, ^{31}P) and Two-Dimensional Hetero TOCSY." *J. Am. Chem. Soc.* **116**: 4983-4984.

Hines, J. V., Landry, S. M., Varani, G. and Tinoco, I. Jr. (1994) "Carbon-Proton Scalar Couplings in RNA: 3D Heteronuclear and 2D Isotope-Edited NMR of a 13C-Labeled Extra-Stable Hairpin." *J. Am. Chem. Soc.* **116**: 5823-5831.

Hore, P. J. (1983) "Solvent Suppression in Fourier Transform Nuclear Magnetic Resonance." *J. Magn. Reson.* **55**: 283-300.

Ikura, M. and Bax, A. (1992) "Isotope-Filtered 2D NMR of a Protein-Peptide Complex: Study of a Skeletal Muscle Myosin Light Chain Kinase Fragment Bound to Calmodulin." *J. Am. Chem. Soc.* **114**: 2433-2440.

Iwai, S., Pritchard, C., Mann, D. A., Karn, J. and Gait, M. J. (1992) "Recognition of the high affinity binding site in rev-response element RNA by the Human Immunodeficiency Virus type-1 rev protein." *Nucleic Acids Res.* **20**: 6465-6472.

Jack, A., Ladner, J. E. and Klug, A. (1976) "Crystallographic refinement of yeast phenylalanine transfer RNA at 2.5 Å resolution." *J. Mol. Biol.* **108**: 619-649.

Jahnke, W., Baur, M., Gemmecker, G. and Kessler, H. (1995) "Improved Accuracy of NMR Structures by a Modified NOESY-HSQC Experiment." *J. Magn. Reson. B* **106**: 86-88.

Jahnke, W. and Kessler, H. (1994) "Enhanced sensitivity of rapidly exchanging amide protons by improved phase cycling and the constructive use of radiation damping." *J. Biomol. NMR* **4**: 735-740.

Jensen, K. B., Green, L., MacDougal-Waugh, S. and Tuerk, C. (1994) "Characterization of an *in Vitro*-selected RNA Ligand to the HIV-1 Rev Protein." *J. Mol. Biol.* **235**: 237-247.

Kay, L. E. (1995) "Field gradient techniques in NMR spectroscopy." *Current Opinion in Structural Biology* **5**: 674-681.

Kay, L. E., Ikura, M. and Bax, A. (1990) "Proton-Proton Correlation via Carbon-Carbon Couplings: A Three-Dimensional NMR Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with Carbon-13." *J. Am. Chem. Soc.* **112**: 888-889.

Kessler, H., Griesinger, C., Kerssebaum, R., Wagner, K. and Ernst, R. R. (1987) "Separation of cross-relaxation and J cross-peaks in 2D rotating-frame NMR spectroscopy." *J. Am. Chem. Soc.* **109**: 607-609.

Kissinger, C. R., Liu, B., Martin, E., Kornberg, T. B. and Pabo, C. O. (1990) "Crystal Structure of an Engrailed Homeodomain-DNA Complex at 2.8 Angstroms Resolution: A Framework for Understanding Homeodomain-DNA Interactions." *Cell* **63**: 579.

Kjems, J., Brown, M., Chang, D. D. and Sharp, P. A. (1991) "Structural analysis of the interaction between the human immunodeficiency virus Rev protein and the Rev response element." *Proc Natl. Acad. Sci. USA* **88**: 683-687.

Kjems, J., Calnan, B. J., Frankel, A. D. and Sharp, P. A. (1992) "Specific binding of a basic peptide from HIV-1 Rev." *EMBO J.* **11**: 1119-1129.

Kjems, J., Frankel, A. D. and Sharp, P. A. (1991) "Specific Regulation of mRNA Splicing In Vitro by a Peptide from HIV-1 Rev." *Cell* **67**: 169-178.

Klausner, R. D., Rouault, T. A. and Harford, J. B. (1993) "Regulating the Fate of mRNA: The Control of Cellular Iron Metabolism." *Cell* **72**: 19-28.

Kumar, A., Ernst, R. R. and Wuthrich, K. (1980) "A Two-Dimensional Nuclear Overhauser Enhancement (2D NOE) Experiment for the Eludication of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules." *Biochemical and Biophysical Research Communications* **95**: 1-6.

Lazinski, D., Grzadzielska, E. and Das, A. (1989) "Sequence-Specific Recognition of RNA Hairpins by Bacteriophage Antiterminators Requires a Conserved Arginine-Rich Motif." *Cell* **59**: 207-218.

Le, S.-Y., Pattabiraman, N. and Jacob V. Maizel, J. (1994) "RNA tertiary structure of the HIV RRE domain II containing non-Watson-Crick base pairs GG and GA: molecular modeling studies." *Nucleic Acids Res.* **22**: 3966-3976.

Leclerc, F., Cedergren, R. and Ellington, A. D. (1994) "A three-dimensional model of the Rev-binding element of HIV-1 derived from analyses of aptamers." *Nature - Structural Biology* **1**: 293-300.

Lee, A. L., Kanaar, R., Rio, D. C. and Wemmer, D. E. (1994) "Resonance Assignments and Solution Structure of the Second RNA-Binding Domain of Sex-lethal Determined by Multidimensional Heteronuclear Magnetic Resonance." *Biochemistry* **33**: 13775-13786.

Legault, P., II, B. T. F., Mueller, L. and Pardi, A. (1994) "Through-Bond Correlation of Adenine Protons in a ^{13}C-Labeled Ribozyme." *J. Am. Chem. Soc.* **116**: 2203-2204.

Legault, P., Jucker, F. M. and Pardi, A. (1995) "Improved measurement of ^{13}C, ^{31}P J coupling constants in isotopically labeled RNA." *FEBS Lett.* **362**: 156-160.

Logan, T. M., Olejniczak, E. T., Xu, R. X. and Fesik, S. W. (1992) "Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments." *FEBS Lett.* **314**: 413-418.

Lyu, P. C. and Wemmer, D. E. (1993) "Capping Interactions in Isolated α Helices: Position-Dependent Substitution Effects and Structure of a Serine-Capped Peptide Helix." *Biochemistry* **32**: 421-425.

Majumdar, A. and Zuiderweg, E. R. P. (1993) "Improved ^{13}C-Resolved HSQC-NOESY Spectra in H_2O, Using Pulsed Field Gradients." *J. Magn. Reson. B* **102**: 242-244.

Malim, M. H. and Cullen, B. R. (1991) "HIV-1 Structural Gene Expression Requires the Binding of Multiple Rev Monomers to the Viral RRE: Implications for HIV-1 Latency." *Cell* **65**: 241-248.

Malim, M. H. and Cullen, B. R. (1993) "Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes." *Molecular and Cellular Biology* **13**: 6180-6189.

Malim, M. H., Hauber, J., Le, S.-Y., Maizel, J. V. and Cullen, B. R. (1989) "The HIV-1 *rev trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA." *Nature* **338**: 254-257.

Marino, J. P., Prestegard, J. H. and Crothers, D. M. (1993) "Correlation of Adenine H2/H8 Resonances in Uniformly ^{13}C Labeled RNAs by 2D HCCH-TOCSY: A New Tool for ^{1}H Assignment." *J. Am. Chem. Soc.* **116**: 2205-2206.

Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. and Griesinger, C. (1994) "A Three-Dimensional Triple-Resonance ^{1}H,^{13}C,^{31}P Experiment: Sequential Through-Bond Correlation of Ribose Protons and Intervening Phosphorus along the RNA Oligonucleotide Backbone." *J. Am. Chem. Soc.* **116**: 6472-6473.

Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. and Griesinger, C. (1995) "Sequential correlation of anomeric ribose protons and intervening phosphorus in RNA oligonucleotides by a ^{1}H,^{13}C,^{31}P triple resonance experiment: HCP-CCH-TOCSY." *J. Biomol. NMR* **5**: 87-92.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) "Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins." *J. Magn. Reson.* **85**: 393-399.

Marqusee, S., Robbins, V. H. and Baldwin, R. L. (1989) "Unusually stable helix formation in short alanine-base peptides." *Proc. Natl. Acad. Sci. U.S.A.* **86**: 5286-5290.

Mattaj, I. W. (1993) "RNA Recognition: A Family Matter?" *Cell* **73**: 837-840.

Molinaro, M. and Tinoco, I. Jr. (1995) "Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications." *Nucleic Acids Res.* **23**: 3056-63.

Montelione, G. T., Lyons, B. A., Emerson, S. D. and Tashiro, M. (1992) "An Efficient Triple Resonance Experiment Using Carbon-13 Isotropic Mixing for Determining Sequence-Specific Resonance Assignments of Isotopically-Enriched Proteins." *J. Am. Chem. Soc.* **114**: 10974-10975.

Mori, S., Abeygunawardana, C., Johnson, M. O. and Zijl, P. C. M. v. (1995) "Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation." *J. Magn. Reson. B* **108**: 94-98.

Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., H.Smallcombe, S. and Kay, L. E. (1993) "A Gradient ^{13}C NOESY-HSQC Experiment for Recording NOESY Spectra of ^{13}C-Labeled Proteins Dissolved in H_2O." *J. Magn. Reson. B* **102**: 317-321.

Muhandiram, D. R. and Kay, L. E. (1994) "Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity." *J. Magn. Reson. B* **103**: 203-216.

Muller, N., Ernst, R. R. and Wuthrich, K. (1986) "Multiple-Quantum-Filtered Two-Dimensional Corrleated NMR Spectroscopy of Proteins." *J. Am. Chem. Soc.* **108**: 6482-6492.

Murata, Y. and Wharton, R. P. (1995) "Binding of Pumilio to Maternal *hunchback* mRNA Is Required for Posterior Patterning in Drosphila Embryos." *Cell* **80**: 1-20.

Nikonowicz, E. P. and Pardi, A. (1993) "An Efficient Procedure for Assignment of the Proton, Carbon and Nitrogen Resonances in 13C/15N Labeled Nucleic Acids." *J. Mol. Biol.* **232**: 1141-1156.

Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. and Pardi, A. (1992) "Preparation of 13C and 15N labelled RNAs for heteronuclear multi-dimensional NMR studies." *Nucleic Acids Res.* **20**: 4507-4513.

Nilges, M., Clore, G. M. and Gronenborn, A. M. (1988) "Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations." *FEBS Lett.* **229**: 317-324.

Noller, H. F., Hoffarth, V. and Zimniak, L. (1992) "Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures." *Science* **256**: 1416-1419.

Nomura, M. (1973) "Assembly of bacterial ribosomes." *Science* **179**: 864-873.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) "Myb DNA-binding domain with cooperative recognition helices." *Cell* **79**: 639.

Olsen, H. S., Cochrane, A. W., Dillon, P. J., Nalin, C. M. and Rosen, C. A. (1990) "Interaction of the human immunodeficiency virus type 1 Rev protein with a structured region in env mRNA is dependent on multimer formation mediated through a basic stretch of amino acids." *Genes Dev.* **4**: 1357-1364.

Otting, G. and Wuthrich, K. (1989) "Extended Heteronuclear Editing of 2D 1H NMR Spectra of Isotope-Labeled Proteins, Using the X(ω1, ω2) Double Half Filter." *J. Magn. Reson.* **85**: 586-594.

Otting, G. and Wuthrich, K. (1990) "Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions." *Quarterly Reviews of Biophysics* **23**: 39-96.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. and Nagai, K. (1994) "Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin." *Nature* **372**: 432-438.

Owen, D. and Kuhn, L. C. (1987) "Noncoding 3' sequences of the transferrin receptor gene are required for mRNA regulation by iron." *EMBO J.* **6**: 1287-1293.

Pabo, C. O. and Sauer, R. T. (1992) "TRANSCRIPTION FACTORS: Structural Families and Principles of DNA Recognition." *Annu. Rev. Biochem.* **61**: 1053-95.

Pardi, A. and Nikonowicz, E. P. (1992) "Simple Procedure for Resonance Assignment of the Sugar Protons in 13C-Labeled RNAs." *J. Am. Chem. Soc.* **114**: 9202-9203.

Park, C., Campbell, J. L. and William A. Goddard, I. (1996) "Can the Monomer of the Leucine Zipper Proteins Recognized the Dimer Binding Site without Dimerization?" *J. Am. Chem. Soc.* **118**: 4235-4239.

Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D. and Kay, L. E. (1994) "Simultaneous Acquisition of 15N- and 13C-Edited NOE Spectra of Proteins Dissolved in H2O." *J. Magn. Reson. B* **103**: 197-204.

Pavletich, N. P. and Pabo, C. O. (1991) "Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å." :

Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J. and Feigon, J. (1994) "1H NMR Studies of the High-Affinity Rev Binding Site of the Rev Responsive Element of HIV-1 mRNA: Base Pairing in the Core Binding Element." *Biochemistry* **33**: 5357-5366.

Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. and Gait, M. J. (1994) "Methylphosphonate mapping of phosphate contacts critical for RNA recognition by the human immunodeficiency virus tat and rev proteins." *Nucleic Acids Res.* **22**: 2592-2600.

Puglisi, J. D., Chen, L., Blanchard, S. and Frankel, A. D. (1995) "Solution Structure of a Bovine Immunodeficiency Virus Tat-TAR Peptide-RNA Complex." *Science* **270**: 1200-1203.

Puglisi, J. D., Chen, L., Frankel, A. D. and Williamson, J. R. (1993) "Role of RNA structure in arginine recognition of TAR RNA." *Proc. Natl. Acad. Sci. U.S.A.* **90**: 3680-3684.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. and Williamson, J. R. (1992) "Conformation of TAR RNA-Arginine Complex by NMR Spectroscopy." *Science* **257**: 76-80.

Rich, A. and Kim, S. H. (1978) "The three-dimensional structure of transfer RNA." *Sci. Amer.* **238**: 52-62.

Rould, M. A., Perona, J. J., Soll, D. and Steitz, T. A. (1989) "Structure of *E. Coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution: implications for tRNA discrimination." *Science* **246**: 1135-1142.

Roy, S., Delling, U., Chen, C.-H., Rosen, C. A. and Sonenberg, N. (1990) "A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated *trans*-activation." *Genes and Development* **4**: 1365-1373.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. and Moras, D. (1991) "Class II Aminoacyl Transfer RNA Synthetases: Crystal Structure of Yeast Aspartyl-tRNA Synthetase Complexed with tRNAAsp." *Science* **252**: 1682-1689.

Ruhl, M., Himmelspach, M., Bahr, G. M., Hammerschmid, F., Jaksche, H., Wolf, B., Aschauer, H., Farrington, G. K., Probst, H., Bevec, D. and Hauber, J. (1993) "eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating *trans*-activation." *J. Cell Biol.* **123**: 1309-1320.

Saenger, W. (1984) Principles of Nucleic Acid Structure. New York, Springer-Verlag.

Santoro, J. and King, G. C. (1992) "A Constant-Time 2D Overbodenhausen Experiment for Inverse Correlation of Isotopically Enriched Species." *J. Magn. Reson.* **97**: 202-207.

Scanlon, M. J., Fairlie, D. P., Craik, D. J., Englebretsen, D. R. and West, M. L. (1995) "NMR Solution Structure of the RNA-Binding Peptide from Human Immunodeficiency Virus (Type 1) Rev." *Biochemistry* **34**: 8242-8249.

Schumacher, T. N. M., Mayr, L. M., Jr., D. L. M., Milhollen, M. A., Burgess, M. W. and Kim, P. S. (1996) "Identification of D-Peptide Ligands Through Mirror-Image Phage Display." *Science* **271**: 1854-1857.

Schwalbe, H., Marino, J. P., Glaser, S. J. and Griesinger, C. (1995) "Measurement of H,H-Coupling Constants Associated with ν1, ν2, and ν3 in Uniformly ^{13}C-labeled RNA by HCC-TOCSY-CCH-E.COSY." *J. Am. Chem. Soc.* **117**: 7251-7252.

Schwalbe, H., Marino, J. P., King, G. C., Wechselberger, R., Bermel, W. and Griesinger, C. (1994) "Determination of a complet set of coupling constants in 13C-labeled oligonucleotides." *J. Biomol. NMR* **4**: 631-644.

Seale, J. W., Srinivasan, R. and Rose, G. D. (1994) "Sequence determinants of the capping box, a stabilizing motif at the N-termini of α-helices." *Protein Sci.* **3**: 1741-1745.

Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976) "Sequence-specific recognition of double helical nucleic acids by proteins." *Proc. Natl. Acad. Sci. U.S.A.* **73**: 804-808.

Shaka, A. J., Barker, P. B. and Freeman, R. (1985) "Computer-Optimized Decoupling Scheme for Wideband Applications and Low-Level Operation." *J. Magn. Reson.* **64**: 547-552.

Shaka, A. J., Keeler, J. and Freeman, R. (1983) "Evaluation of a New Broadband Decoupling Sequence: WALTZ-16." *J. Magn. Reson.* **53**: 313-340.

Shen, L. X., Cai, Z. and Tinoco, I. Jr. (1995) "RNA structure at high resolution." *FASEB J.* **9**: 1023-1033.

Shriver, J. (1992) "Product Operators and Coherence Transfer in Multiple-Pulse NMR Experiments." *Concepts in Magnetic Resonance* **4**: 1-33.

Simorre, J.-P., Zimmermann, G. R., Pardi, A., II, B. T. F. and Mueller, L. (1995) "Triple resonance HNCCCH experiments for correlating exchangeable and nonexchangeable cytidine and uridine base protons in RNA." *J. Biomol. NMR* **6**: 427-432.

Sklenar, V., Brooks, B. R., Zon, G. and Bax, A. (1987) "Absorption mode two-dimensional NOE spectroscopy of exchangeable protons in oligonucleotides." *FEBS Lett.* **216**: 249-252.

Sklenar, V., Dieckmann, T., Butcher, S. E. and Feigon, J. (1996) "Through-bond correlation of imino and aromatic resonances in ^{13}C-,^{15}N-labeled RNA via heteronuclear TOCSY." *J. Biomol. NMR* **7**: 83-87.

Sklenar, V., Peterson, R. D., Rejante, M. R., Wang, E. and Feigon, J. (1993a) "Two-Dimensional Triple-Resonance HCNCH Experiment for Direct Correlation of Ribose H1' and Base H8, H6 Protons in ^{13}C,^{15}N-Labeled RNA Oligonucleotides." *J. Am. Chem. Soc.* **115**: 12181-12182.

Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993b) "Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity." *J. Magn. Reson. A* **102**: 241-245.

Sorensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G. and Ernst, R. R. (1983) "Product Operator Formalism for the Description of NMR Pulse Experiments." *Progress in NMR Spectroscopy* **16**: 163-192.

Spolar, R. S. and Record, M. T. J. (1994) "Coupling of Local Folding to Site-Specific Binding of Proteins to DNA." *Science* **263**: 777-784.

Staley, J. P. and Kim, P. S. (1994) "Formation of a native-like subdomain in a partially folded intermediate of bovine pancreatic trypsin inhibitor." *Protein Sci.* **3**: 1822-1832.

States, D. J., Haberkorn, R. A. and Ruben, D. J. (1982) "A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants." *J. Magn. Reson.* **48**: 286-292.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. Methods in Enzymology. Vol. 185, San Diego, Academic Press. 60-89.

Stutz, F., Neville, M. and Rosbash, M. (1995) "Identification of a Novel Nuclear Pore-Associated Protein as a Functional Target of the HIV-1 Rev Protein in Yeast." *Cell* **82**: 495-506.

Szewczak, A. A., Moore, P. B., Chan, Y.-L. and Wool, I. G. (1993) "The Solution Structure of the Sarcin/Ricin Loop from 28S Ribosomal RNA." *Proc. Natl. Acad. Sci. U.S.A.* **90**: 9581-9583.

Tan, R., Chen, L., Buettner, J. A., Hudson, D. and Frankel, A. D. (1993) "RNA Recognition by an Isolated α Helix." *Cell* **73**: 1031-1040.

Tan, R. and Frankel, A. (1994) "Costabilization of Peptide and RNA Structure in an HIV Rev Peptide-RRE Complex." *Biochemistry* **33**: 14579-14585.

Tan, R. and Frankel, A. D. (1995) "Structural variety of arginine-rich RNA-binding peptides." *Proc. Natl. Acad. Sci. U.S.A.* **92**: 5282-5286.

Tiley, L. S., Malim, M. H., Tewary, H. K., Stockley, P. G. and Cullen, B. R. (1992) "Identification of a high-affinity RNA-binding site for the human immunodificiency virus type 1 Rev protein." *Proc. Natl. Acad. Sci. U.S.A.* **89**: 758-762.

Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. and Liljas, L. (1994) "Crystal structure of an RNA bacteriophage coat protein-operator complex." *Nature* **371**: 623-626.

Varani, G., Aboul-ela, F., Allain, F. and Gubser, C. C. (1995) "Novel three-dimensional 1H-13C-31P triple resonance experiments for sequential backbone correlations in nucleic acids." *J. Biomol. NMR* **5**: 315-320.

Varani, G. and Tinoco, I., Jr. (1991) "RNA structure and NMR spectroscopy." *Quarterly Reviews of Biophysics* **24**: 479-532.

Vuister, G. W. and Bax, A. (1993) "Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HNHa) Coupling Constants in 15N-Enriched Proteins." *J. Am. Chem. Soc.* **115**: 7772-7777.

Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R. and Bax, A. (1993) "Increased Resolution and Improved Spectral Quality in Four-Dimensional 13C/13C-Separated HMQC-NOESY-HMQC Spectra Using Pulsed Field Gradients." *J. Magn. Reson. B* **101**: 210-213.

Wagner, G. (1989) Heteronuclear Nuclear Magnetic Resonance Experiments for Studies of Protein Conformation. Methods in Enzymology. Vol. 176, San Diego, Academica Press. 93-113.

Wand, A. J., and Short, J. H. (1994) Nuclear Magnetic Reonance Studies of Protein-Peptide Complexes. Methods in Enzymology. Vol. 239, San Diego, Academic Press. 700-717.

Weeks, K. M. and Crothers, D. M. (1991) "RNA Recognition by Tat-Derived Peptides: Interaction in the Major Groove." *Cell* **66**: 577-588.

Weeks, K. M. and Crothers, D. M. (1993) "Major Groove Acessibility of RNA." *Science* **261**: 1574-1577.

Werstuck, G., Zapp, M. L. and Green, M. R. (1996) "A non-canonical base pair within the human immunodeficiency virus Rev-responsive element is involved in both Rev and small molecule recognition." *Chem. Biol.* **3**: 129-137.

Wijmenga, S. S., Heus, H. A., Leeuw, H. A. E., Hoppe, H., Graaf, M. v. d. and Hilbers, C. W. (1995) "Sequential backbone assignment of uniformly 13C-labeled RNAs by a two-dimensional P(CC)H-TOCSY triple resonance NMR experiment." *J. Biomol. NMR* **5**: 82-86.

Wimberly, B., Varani, G. and Tinoco, I. Jr. (1993) "The Conformation of Loop E of Eukaryotic 5S Ribosomal RNA." *Biochemistry* **32**: 1078-1087.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. and Sykes, B. D. (1995) "1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects." *J. Biomol. NMR* **5**: 67-31.

Wishart, D. S. and Sykes, B. D. (1994) "The 13C chemical-shift index: a simplex method for the identification of protein secondary structure using 13C chemical-shift data." *J. Biomol. NMR* **4**: 171-180.

Wishart, D. S., Sykes, B. D. and Richards, F. M. (1992) "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy." *Biochemistry* **31**: 1647-1651.

Wittekind, M. and Mueller, L. (1993) "HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins." *J. Magn. Reson. B* **101**: 201-205.

Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids. New York, Wiley.

Wyatt, J. R., Chastain, M. and Puglisi, J. D. (1991) "Synthesis and Purification of Large Amounts of RNA Oligonucleotides." *BioTechniques* **11**: 764-769.

Ye, X., Kumar, R. A. and Patel, D. J. (1995) "Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex." *Chem. Biol.* **2**: 827-840.

Zapp, M. L., Hope, T. J., Parslow, T. G. and Green, M. R. (1991) "Oligomerization and RNA binding domains of the type 1 human immunodeficiency virus Rev protein: A dual function for an arginine-rich binding motif." *Proc. Natl. Acad. Sci. U.S.A.* **88**: 7734-7738.

Zapp, M. L., Stern, S. and Green, M. R. (1993) "Small molecules that selectively block RNA binding of HIV-1 Rev protein inhibit Rev function and viral production." *Cell* **74**: 969-78.

THESIS PROCESSING SLIP

FIXED FIELD: ill. _____ name _____

index _____ biblio _____

► COPIES: (Archives) Aero Dewey Eng Hum

Lindgren Music Rotch (Science)

TITLE VARIES: ► ☐ _____

NAME VARIES: ► ☑ Lawrence

IMPRINT: (COPYRIGHT) _____

► COLLATION: 200 l

► ADD DEGREE: _____ ► DEPT.: _____

SUPERVISORS: _____

NOTES:

cat'r: date:

page: 352

► DEPT: Chem

► YEAR: 1996 ► DEGREE: Ph.D

► NAME: BATTISTE, John L.